УДК 538 0:621 391

М. В. Никифорова – магистрант кафедры компьютерной математики и программирования
В. И. Хименко (д-р техн. наук, проф.) – научный руководитель

ОБЩАЯ СТРУКТУРА РАСПОЗНАВАНИЯ ЛИЧНОСТИ ПО ГОЛОСУ. ИССЛЕДОВАНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ

Среди систем анализа и распознавания речи выделяют отдельный класс задач – распознавание личности по его голосу. Данная задача была поставлена более 40 лет тому назад, и исследования в этой области все еще продолжаются. На сегодняшний день это одна из актуальнейших проблем, решение которой может найти применение в криминалистике, радиоразведке, контрразведке, антитеррористическом мониторинге, обеспечение безопасности доступа к физическим объектам, информационным и финансовым ресурсам [1].

Работа систем распознавания содержит два основных этапа: обучение системы (регистрация пользователей) и сам процесс распознавания (сравнение и принятие решения).

На первом этапе пользователи регистрируются в системе, записывая свои голоса. Образец голоса диктора обрабатывается системой с целью извлечения информативных признаков, которые описывают индивидуальность говорящего. На основе извлечённых признаков строятся модели (в некоторых случаях более подходящим термином является «шаблон») пользователей. Модель представляет собой некоторую структуру, позволяющую при данных признаках оценить степень подобия либо сразу принять решение.

На втором этапе, в зависимости от конкретной задачи, происходит принятие/отклонение заявленной личности (верификация пользователя) или определение схожести/различия входной модели с хранящимися в базе шаблонами (идентификация пользователя).

В первом случае, пользователь пытается войти в систему, предъявляя идентификатор и образец голоса. Признаки, извлечённые из предъявленного образца, сравниваются с соответствующей моделью, сохранённой в базе. Если соответствие достаточно хорошее, т.е выше порога, заявленная личность подтверждается, в противоположном случае отклоняется.

Во втором случае, во время процесса идентификации, также происходит извлечение признаков из предъявленного образца голоса неизвестного диктора, которые затем анализируется и сравнивается с речевыми моделями всех зарегистрированных в системе пользователей. Неизвестный диктор идентифицируется как пользователь, чья модель наиболее соответствует входному высказыванию [2].

Таким образом, общая схема системы распознавания диктора реализуется с помощью следующих трех компонентов (модулей).

1. Модуль обработки сигналов. На данном уровне речевой сигнал обрабатывается с целью выделения информативных признаков, существенных для задачи распознавания говорящего. На выходе мы получаем последовательность векторов признаков, которыми этот речевой сигнал описывается.
2. Модуль обучения. При регистрации пользователя данный уровень использует полученную от модуля обработки сигналов последовательность векторов признаков для построения модели(шаблона). Моделирование может заключаться как в простом копировании векторов признаков, так и в построении вероятностных моделей или других структур. Построенная модель записывает базу данных, где служит эталонной моделью для данного пользователя и используется для вычисления степени подобия.
3. Модуль принятия решений. Данный модуль можно условно разделить на два этапа: сравнение с образцом (эталоном) и принятие решения. На первом этапе вектор признаков входного сигнала сравнивается с эталонной моделью, находящейся в базе, или с шаблонами всех зарегистрированных пользователей, в зависимости от конкретной задачи. После этого происходит принятие решения о пропуске или отклонении заявленной личности в зависимости от порога(для верификации) или возврат идентификатора конкретного пользователя в зависимости от степени подобия (идентификация пользователя).

Рисунок 1 показывает общую структуру распознавания диктора по голосу.

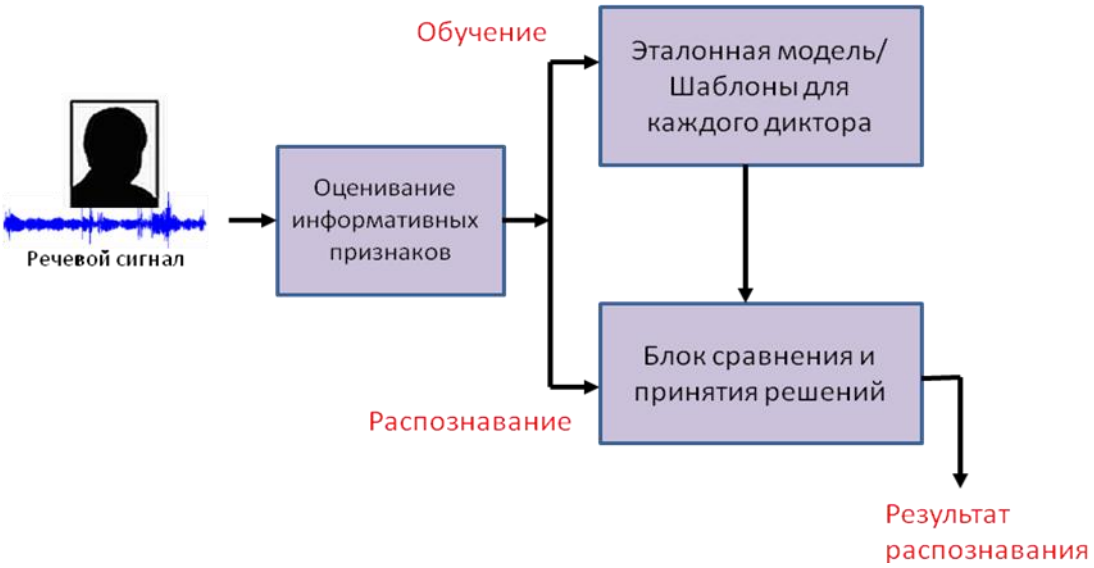
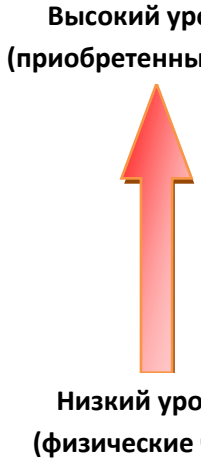


Рис. 1. Общая структура распознавания личности по голосу

Первоочередной задачей в системах распознавания личности по голосу и важнейшим элементом успешного распознавания дикторов является выбор информативных признаков (речевых параметров), способных эффективно представлять информацию об особенностях речи конкретного диктора. Индивидуальность акустических характеристик голоса определяется тремя факторами: механикой колебаний голосовых складок, анатомией речевого тракта и системой управления артикуляцией.

Анатомия тракта (геометрические размеры различных отделов речевого тракта и боковые полости) определяет спектральные характеристики звуков речи. Система управления артикуляцией формирует просодические характеристики: темп речи, скорость переходных процессов и длительность фонетических сегментов, а также эффекты коартикуляции. Механика колебания голосовых складок, которая описывается частотой колебания, формой импульсов, а так же размером, жесткость и массой самих голосовых складок, определяет частоту основного тона и тембральные характеристики речевого сигнала [3].

Все индивидуальные параметры говорящего можно разделить на два вида: низкоуровневые (обусловленные анатомическим строением речевого аппарата) и высокоуровневые (приобретённые, связанные с манерой произношения). Такое разделение связано с разными уровнями информации, представленными в речевом сигнале. В своей повседневной жизни человек полагается на совокупность совершенно разных уровней, иерархия которых представлена на рисунке 2.



Семантика, дикция, произношение, идиолект	Социально-экономический статус, образование, место рождения
Просодия, ритм, скорость интонации, громкость	Тип личности, характер, темперамент
Акустический аспект речи, нозальность, глубина, дыхание	Анатомия речевого аппарата

Рис. 2. Иерархия уровней информации в речевом сигнале

На рисунке 2 показано, что за низкоуровневые (спектральные) признаки отвечает анатомия речевого аппарата, далее начинаются высокоуровневые признаки. Просодия, ритм, скорость интонации и громкость – это индивидуальные параметры, которые человек приобретает со временем. Очень многое зависит от типа личности и характера, у уверенных в себе людей речь более громкая, у импульсивных – более быстрая, у расчетливых – более спокойная. Семантика, дикция, произношение, идиолект – все это обусловлено социально-экономическим статусом, образованием, местом рождения.

Большим плюсом высокоуровневых признаков является невосприимчивость к эффектам канала и шумам, что нельзя сказать о низкоуровневых. Однако, на этом их достоинства заканчиваются. К недостаткам можно отнести необходимость большого числа тренировочной базы, сложность в извлечении, время вычисления. Поэтому в настоящее время все практические приложения основываются именно на краткосрочных спектральных характеристиках, которые относятся к низкоуровневым признакам [4].

На сегодняшний день априори невозможно оценить, какие признаки более подходят для распознавания. Процесс определения подходящих признаков заключается в переборе возможных вариантов комбинаций признаков с последующей экспериментальной оценкой.

Библиографический список

1. В.Н.Сорокин, В.В.Вьюгин, А.А.Тананыкин. Распознавание личности по голосу: аналитический обзор. Информационные процессы, Том 12, №1, 2012 г.
2. Campbell J.P., Speaker Recognition: A Tutorial. Proceedings of the IEEE. 1997.
3. Рамишвили Г. С. Автоматическое опознавание говорящего по голосу. – М.: Радио и Связь, 1981.
4. Tomi Kinnunen, Haizhou Li. An overview of text-independent speaker recognition:From features to supervectors. - Speech Communication 52, 2010.
5. Campbell J.P., Reynolds D.A., Dunn R.B. Fusing high- and lowlevel features for speaker recognition. In: Proc. Eurospeech, 2003.