УДК 004.934.8

ПРОГРАММНЫЕ СРЕДСТВА ТЕКСТОЗАВИСИМОЙ ВЕРИФИКАЦИИ ДИКТОРА ПО ГОЛОСУ

О.Б. ЗЕЛЬМАНСКИЙ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 10 июня 2009

Верификация диктора по голосу заключается в подтверждении личности говорящего на основе индивидуальных особенностей речи. Рассмотрен метод построения алгоритма голосовой верификации, использующий отличительные признаки речи, связанные со строением и функционированием речеобразующего тракта.

Ключевые слова: текстозависимая верификация, распознавание, признаки речи, речеобразующий тракт, детектирование речи, параметры сигнала.

Введение

В настоящее время системы голосовой верификации занимают важное место при решении задач обеспечения информационной безопасности [1]. На вход таких систем подаются личные данные диктора и произнесенный им пароль, на выходе система формирует решение о том, действительно ли верифицируемый диктор является тем, кем он себя назвал. Перспективность использования уникальных особенностей голосовых технологий для биометрики обуславливается наличием ряда достоинств голосовой верификации по сравнению с верификацией на основе отпечатков пальцев, черт лица или радужной оболочки глаза [2], среди которых можно выделить следующие: естественность и привычность речевого взаимодействия между человеком и компьютерной системой; речь позволяет совместить процедуры верификации и коммуникации; личность диктора может быть определена без непосредственного контакта с ним, возможно использование телефонного канала; для верификации по голосу не требуется специальных сложных технических средств, что в свою очередь обеспечивает низкую стоимость таких систем.

Области использования голосовой биометрии сегодня постоянно расширяются. В первую очередь это организация защиты информации и контроля доступа в различных информационных системах. Кроме того, возрастающая популярность голосовой верификации обусловлена бурным развитием и распространением мобильной связи, что влечет за собой желание иметь безопасный доступ с мобильных устройств к банковским услугам, базам данных, почте и другой информации. Не менее актуальным является использование средств верификации диктора по голосу правоохранительными органами с целью криминалистической экспертизы или анализа записей переговоров, а также другими организациями в случае необходимости корректировки паролей или ключей клиентов.

Описание метода

Процесс верификации диктора по голосу может быть представлен следующим образом.

1. Выделение из аудиосигнала только речевых участков. Для этой цели используется детектор речи [3]. Характеристики детектора речи во многом определяют качество работы всей системы в целом. Поэтому алгоритмы детектирования часто являются наиболее критической

частью таких систем и одновременно, с улучшением их качества, увеличивается качество и всей системы. Таким образом, при построении системы верификации диктора по голосу в первую очередь необходимой является разработка надежного детектора речи, который позволит с большой вероятностью выделить из аудио сигнала только речевые участки. Использование детектора речи позволит уменьшить число арифметических операций при обработке сигналов, снизить загруженность информационных каналов, затраты на передачу речевых сигналов, избежать постоянного анализа эфира вычислительной техникой, приводящего к износу техники, ложным срабатываниям, повышенному энергопотреблению и прочим нежелательным последствиям.

2. Выделение различительных признаков или способа параметризации речевого сообщения. Можно выделить три группы различительных признаков.

К первой группе признаков можно отнести региональные или диалектные особенности произношения отдельных звуков, дефекты речи, социальные варианты произношения и интонирования. Необходимо отметить, что использование указанных особенностей возможно лишь в системах, обеспечивающих частый и продолжительный диалог с пользователем, в процессе которого происходит накопление статистики об артикуляционной деятельности пользователя на больших объемах речевых данных. Применение же данного подхода для реализации систем, функционирующих в режиме однократного доступа к объекту (например, на контрольно-пропускном пункте, для доступа к банковскому счету и какой-либо информации) не позволит достигнуть требуемого эффекта из-за сложности обучения системы и анализа коротких ключевых фраз.

Ко второй группе признаков, прежде всего, относятся характеристики голосового источника и характеристики акустического фильтра, т.е. артикуляторного тракта. Основным параметром голосового источника является период основного тона, использование которого при верификации является эффективным в виду его высокой информативности, легкости автоматического выделения, отсутствия большого объема вычислений при реализации [4]. Для построения систем верификации на основе использования основного тона не требуется большого объема обучающей информации. Для выделения основного тона возможно использовать метод Голда-Рабинера [4], поскольку он основан исключительно на обработке во временной области, требует малых затрат времени, а также реализует принцип параллельной обработки. Данный метод заключается в формировании нескольких импульсных последовательностей, состоящих из положительных импульсов, возникающих в месте расположения максимума или минимума сигнала. Для каждой из последовательностей вычисляется оценка периода основного тона. Путем усреднения массива частных оценок вычисляется значение периода основного тона.

Параметрами акустического фильтра являются частоты формант и среднеквадратическое значение энергии сигнала, применение которого совместно с периодом основного тона может обеспечить высокие результаты работы системы верификации.

Кратковременное среднеквадратичное значение сигнала — это наиболее простой и в то же время эффективный параметр сигнала, который вычисляется для каждого фрейма сигнала по следующей формуле:

$$RMS = \frac{1}{N}\sqrt{\sum_{n=0}^{N-1}\left|x\ \ n\right|^2},$$

где N — длина фрейма, $x(n)$ — входной речевой сигнал. Рассчитанное таким образом значение соответствует кратковременной мощности сигнала.

К третьей группе признаков относятся коэффициенты линейного предсказания, спектральные компоненты и кепстральные параметры. Следует отметить, что универсального метода параметризации, скорее всего, нет, так как все методы по-разному реагируют на специфический характер помех и реализация системы верификации на основе этой группы параметров является весьма трудоемкой и включает большой объем вычислений.

Таким образом, наиболее эффективным приемом увеличения надежности верификации согласно [5] является интегральное описание речевого сообщения. А именно, использование нескольких различительных признаков речи. В соответствие с [6], наиболее эффективным и устойчивым к условиям окружающей акустической обстановки, а также к вариациям свойств

диктора является идентификационный признак (классификационный параметр), характеризующий голосовой источник, а именно основной тон. Использование данного признака совместно со среднеквадратическим значением энергии сигнала, средним числом переходов сигнала через нуль и спектральной оценкой, которые также могут быть использованы с целью детектирования речи [3], обеспечит надежную, устойчивую, эффективную, не требующую объемных вычислений верификацию диктора по голосу.

3. Построение эталона для данного диктора. Следующим шагом после выделения отличительных признаков речи конкретного диктора является их сохранение в виде эталона в базе данных в режиме регистрации диктора. Для того, чтобы сформировать эталон пользователю предлагается несколько раз произнести ключевую фразу (пароль), ограниченную, как правило, по длительности 3–4 с. При этом обучение системы верификации в режиме регистрации диктора проводится на усредненных речевых отрезках по результатам записи нескольких произношений. Для исключения возможности подмены или уничтожения эталонов они хранятся в защищенных от записи файлах. Хранение речевых эталонов также возможно на индивидуальной карте. В этом случае доступ к системе реализуется только в случае предъявления карты. Утеря или хищение последней при обнаружении пользователем приводит к смене паролей.

Принимая во внимание тот факт, что система верификации может быть взломана, если нарушитель обладает записанным фрагментом ключевой фразы, которую он мог подслушать или получить под воздействием силы, необходимо использовать некоторую базу паролей, сформированную системой в режиме регистрации диктора. В данной концепции система верификации случайным образом выбирает пароль из этой базы и в режиме верификации предлагает пользователю каждый раз произнести новую ключевую фразу. Таким образом, наряду с биометрическими параметрами, система анализирует правильность ответа и знание пароля, что повышает надежность системы. Также для каждого диктора необходимо предусмотреть параллельный набор паролей, который будет использован в случае применения по отношению к нему насилия. При распознавании таких паролей система внешне реагирует нормально, и в то же время оповещает службу безопасности.

4. Принятие решения об индивидуальности говорящего. В режиме верификации требуется однократное сравнение совокупности измеренных значений отличительных параметров произнесенной ключевой фразы с эталонными значениями соответствующих параметров, на основании которого выносится решение о принятии или отклонении предполагаемой идентичности. Для этой цели применяются методы вычисления расстояний $D_i(j=1, 2, \ldots, M)$ в параметрическом N-мерном пространстве между двумя реализациями, где N — размерность параметрического вектора, а M — количество упорядоченных по времени векторов. Если суммарное значение расстояния, вычисленное по формуле:

$$d = \sum_{j=0}^{M-1} D_i$$

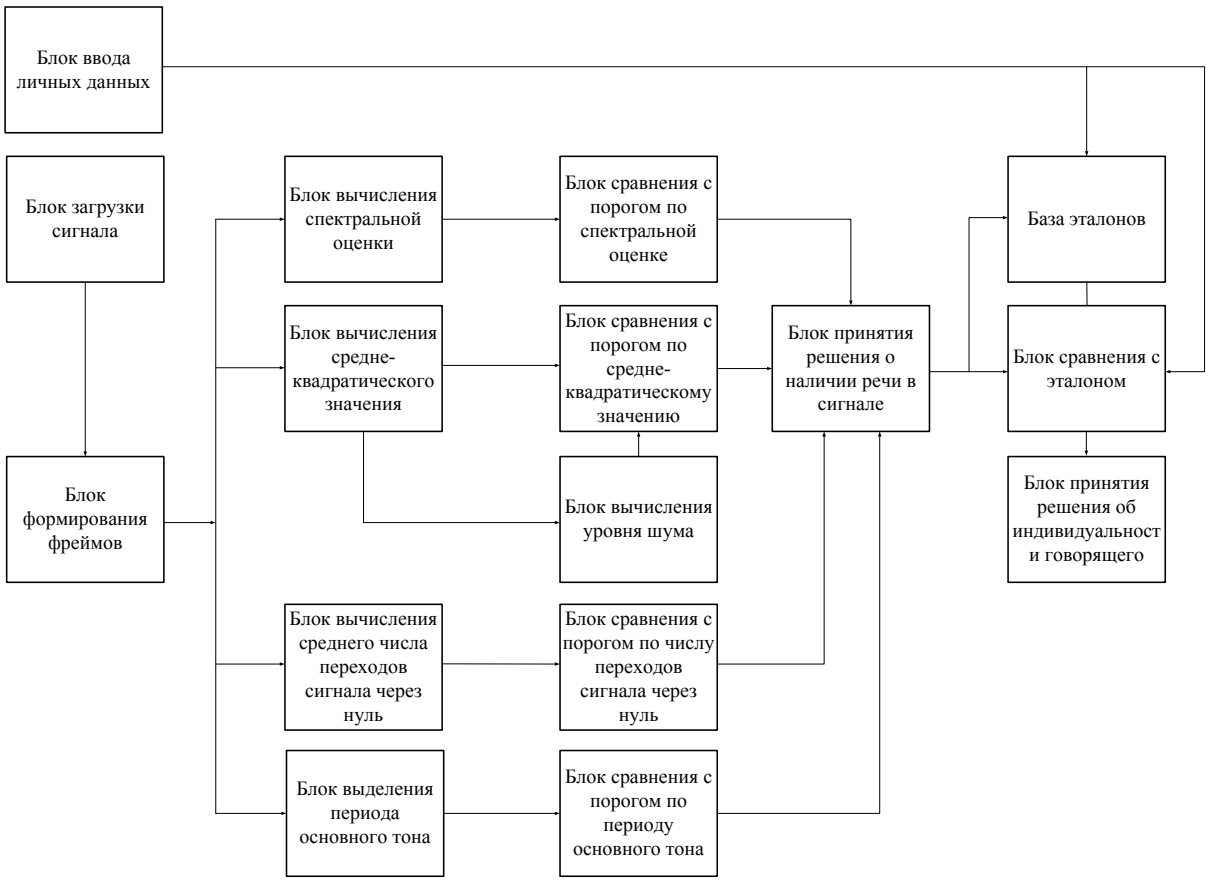
не превышает установленного порога верификации, принимается решение о положительном определении данного голоса.

Порог верификации (различимости) выражается соотношением ошибок первого и второго рода, которые характеризуют систему верификации, а его значение диктуется конкретными задачами и областью применения системы. Ошибки первого рода — это захват ложной цели или принятие злоумышленника за зарегистрированного пользователя, ошибки второго рода — это пропуск цели или отказ признать зарегистрированного пользователя. Таким образом, система верификации может перестраиваться так, 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. Вероятность перечисленных выше ошибок практически не зависит от числа эталонов, хранимых в системе.

Программная реализация метода

Для реализации программного средства верификации диктора по голосу был использован объектно-ориентированный язык программирования C++. Все математические операции, выполняемые в процессе работы программы, формализованы на языке программирования C++ непосредственно в программном коде. Данная реализация обеспечивает переносимость полученного программного модуля.

В ходе системного проектирования была разработана структурная схема верификатора, в основе которой лежит описанный выше алгоритм выделения отличительных признаков речи. Данная схема представлена на рисунке.



Структурная схема программного средства верификации диктора по голосу

Разработанное программное средство позволяет осуществлять подтверждение личности говорящего, используя такие индивидуальные особенности речи, как период основного тона, среднеквадратичное значение энергии и среднее число переходов сигнала через нуль, а также спектральную оценку сигнала, которые могу быть использованы и для детектирования речи в сигнале, поступающем на вход верификатора. Созданное программное средство характеризуется следующими особенностями.

1. Наряду с биометрическими параметрами используется личный код, что повышает надежность системы.

2. Ввод голосового пароля осуществляется в соответствие задаваемыми системой вопросами, причем пароль не является логическим ответом на вопрос. Таким образом, наряду с биометрическими параметрами система анализирует правильность ответа и знание пароля, что повышает надежность системы.

3. Предусмотрен набор паролей, который содержит "экстренные" пароли, используемые диктором в случае применения по отношению к нему насилия со стороны злоумышленника. Распознав такие пароли, система внешне реагирует нормально, но в то же время предупреждает службу безопасности.

4. В течение сеанса доступа к информации осуществляется повторная верификация, которая позволяет выявить ситуацию, когда зарегистрированный диктор подменяется после подтверждения его личности и права доступа к данным.

5. В процессе верификации необходимо повторное произнесение парольной фразы с целью выявления попытки использования записи голоса зарегистрированного диктора. В виду невозможности повторного произнесения парольной фразы совершенно одинаково такая попытка будет обнаружена.

6. В случае неудачной верификации произнесенная пользователем фраза записывается в файл для ее дальнейшего анализа.

7. В зависимости от решаемых системой задач порог верификации может перестраиваться таким образом, чтобы ошибки одного рода были уменьшены за счет увеличения ошибок другого рода или наоборот.

Выводы

Совместное применение методов спектрального и временного анализа на основе сравнения значений классификационных параметров сигнала с их пороговыми значениями позволило разработать рациональный, быстродействующий и надежный алгоритм верификации диктора, включающий детектирование речи. Разработанный алгоритм совмещает эффективность и компактность вычислений и обеспечивает результаты, обладающие высокой степенью достоверности и точности. Программное средство, созданное на основе этого алгоритма, характеризуется удобным пользовательским интерфейсом и просто в эксплуатации.

Данный алгоритм может быть использован для организации защиты информации и ограничения доступа к информационным системам.

TEXT-DEPENDENT SPEAKER VERIFICATION SYSTEM

O.B. ZELMANSKI

Abstract

One of the biometric systems — text-dependent speaker verification is described. The method for the speaker verification is displayed in the article which discusses the different stages of speaker verification in text-dependent systems. Each stage has its subparts, so those parts are described as well. Feature extraction from the raw speech data is discussed. Pre-emphasis, windowing and other parts of feature extraction are mentioned.

Литература

1. *Зельманский О.Б.* // Комплексная защита информации: Материалы XIV Междунар. конф. Минск, 2009. С. 107–108.
2. *Зубов Г.Н., Хитров М.В.* // Voice Biometrics Conference: Тезисы докл. Москва, 2007. С. 25–29.
3. *Зельманский О.Б.* // Докл. БГУИР. 2008. №7. С. 72–77.
4. *Рабинер Л.Р.* Цифровая обработка речевых сигналов: Справочник / Л.Р. Рабинер, Р.В. Шафер. М., 1981.
5. *Галунов В.И.* // Сб. тр. XIII сессии Российского акустического общества. СПб., 2003. С. 35–40.
6. *Каганов А.Ш.* // Информационные технологии процессуального доказывания. Москва, 2002.