УДК 340.6:681.327:681.5.015

ИССЛЕДОВАНИЕ ГОЛОСА ЧЕЛОВЕКА И ЕГО ИДЕНТИФИКАЦИЯ

А.А. Бацкалева
Донецкий национальный технический университет

Рассмотрены достижения в области автоматических методов идентификации личностей по голосу, которые позволили приблизить рабочие характеристики голосовой модальности к характеристикам других биометрических модальностей.

В настоящие время, в связи с широким распространением компьютерной техники и использованием их в различных областях жизни, особо остро встает вопрос ограничения доступа пользователя. При этом все чаще стандартные парольные системы защиты заменяются или дополняются биометрическими системами идентификации пользователей.

Термин "биометрия" обозначает измерение некоторых анатомических или физиологических параметров человека. Если обыкновенный пароль можно украсть или подобрать, то обмануть биометрическую систему практически невозможно. На текущий момент в качестве измеряемых параметров используют различные человеческие черты, такие как голос, отпечатки пальцев, радужная оболочка глаз, почерк и стиль нажатий на клавиши, а также внешний вид. Каждая из этих характеристик позволяет выделить конкретного человека из десятков, сотен и более людей. Также возможно комплексное использование нескольких параметров.

Основным способом персонификации пользователя является указание его сетевого имени и пароля, который пользователь может забыть или хранить в неподходящем месте или его могут украсть [1].

Биометрический подход, связанный с идентификацией голоса, удобен в применении. Однако основным и определяющим недостатком этого подхода является низкая точность идентификации. Например, человек с простудой или ларингитом может испытывать трудности при использовании данных систем. Причинами внедрения этих систем являются повсеместное распространение телефонных сетей и практика встраивания микрофонов в компьютеры и периферийные устройства. Недостатки таких систем - это факторы, влияющие на результаты распознавания: помехи в микрофонах, влияние окружающей обстановки на результаты распознавания, шум,

ошибки при произношении, различное эмоциональное состояние проверяемого в момент регистрации эталона и при каждой идентификации, использование разных устройств регистрации при записи эталонов и идентификации, помехи в низкокачественных каналах передачи данных [2].

При рассмотрении проблемы аутентификации по голосу важными вопросами с точки зрения безопасности являются следующие:

1. Как бороться против использования магнитофонных записей парольных фраз, перехваченных во время установления контакта законного пользователя с аутентификационным терминалом?

2. Как защитить систему от злоумышленников, обладающих способностью к имитации голоса, если им удастся узнать парольную фразу?

Рассмотрим пути решения поставленной задачи

1. Борьбой против использования магнитофонных записей является генерация системой псевдослучайных паролей, которые повторяются вслед за ней пользователем, а также применение комбинированных методов проверки (дополняя вводом идентификационной карточки или цифрового персонального кода).

2. Чтобы защитить систему от злоумышленников человек

вырабатывает свое мнение о специфике воспринимаемого голоса путем оценки некоторых его характерных качеств, не обращая внимание при этом на количественную сторону разнообразных мелких компонент речевого сигнала. А автомат наоборот, не обладая способностью улавливать обобщенную характеристику голоса, свой вывод делает, основываясь на конкретных параметрах (они должны быть легко измеряемы и мало зависеть от мешающих факторов окружающей среды (шумов и помех), они должны быть стабильными во времени и не должны поддаваться имитации речевого сигнала [4].

Выбор параметров речевого сигнала способных наилучшим образом описать индивидуальность голоса является самым важным этапом при построении систем автоматической аутентификации по голосу - это, что они должны быть стабильными во времени, должны быть легко измеряемы и не должны поддаваться имитации [4].

Большинство биометрических систем безопасности функционируют следующим образом: в базе данных системы хранится цифровой отпечаток пальца, радужной оболочки глаза или голоса. Человек, собирающийся получить доступ к компьютерной сети, с помощью микрофона, сканера отпечатков пальцев или других

устройств вводит информацию о себе в систему. Поступившие данные сравниваются с образцом, хранимым в базе данных [1].

При распознавании образца проводится процесс, первым шагом которого является первоначальное трансформирование вводимой информации для сокращения обрабатываемого объема так, чтобы ее можно было бы подвергнуть анализу.

Затем определяются конечные выходные параметры для варьирования голоса и производится нормализация для составления шкалы параметров, а также для определения ситуационного уровня речи. Вышеописанные измененные параметры используются затем для создания шаблона. Шаблон включается в словарь, который характеризует произнесение звуков при передаче информации говорящим, использующим эту систему. Далее в процессе распознавания новых речевых образцов (уже подвергшихся нормализации и получивших свои параметры), эти образцы сравниваются с шаблонами, уже имеющимися в базе, используя динамичное искажение и похожие метрические измерения [2].

Практическая работа используемого алгоритма.

Фильтрация шумов.

Звук, образованный колебаниями всего диапазона частот, подобный тому, спектр которого показан на рисунке 1, называется шумом.

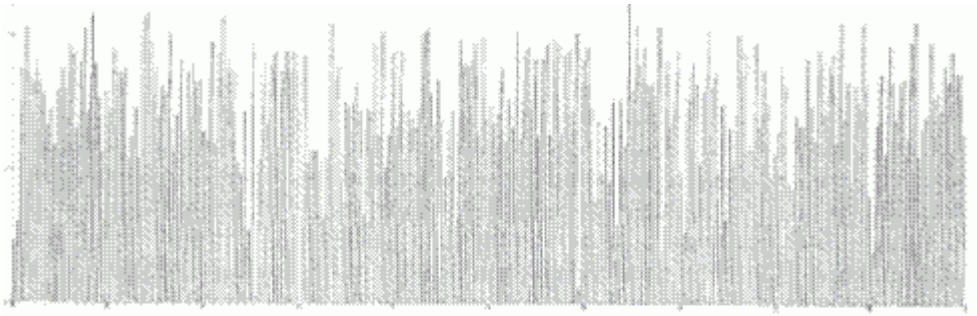

Рисунок. 1 Спектр звука, образованный колебаниями всего диапазона частот

Для того чтобы получить четкие спектральные характеристики звука их нужно отчистить от лишних шумов.

Входной дискретный звуковой сигнал обрабатывается фильтрами, для того чтобы избавится от помех, возникающих при записи по формуле [5]:

$$X_i = (X_i - 0{,}9 * X_{i-1}) \left[0{,}54 - 0{,}46 * cos(i-6) * \frac{2\pi}{180}\right]$$

После обработки в сигнале ищется начало и конец записи, а так как шумы уже отфильтрованы, то начало фрагмента будет характеризоваться всплеском сигнала, если искать с X_0. Соответственно если искать с Xn вниз, то всплеск будет характеризовать конец фрагмента.

Помимо высоты тона человек ощущает и другую характеристику звука - громкость. Основной причиной разной громкости звуков является различное давление, оказываемое ими на уши.

Так как моментальная мощность зависит от квадрата моментальной амплитуды, имеет смысл аналогичным образом подобрать похожее соотношение, связывающее среднюю амплитуду и среднюю мощность. Способ, которым это можно сделать, заключается в определении средней амплитуды.

Спектральное преобразование сигнала.

Измеряется амплитуда частоты f, которая приведена в формуле [5] в первом приближении, при вычислении следующей суммы:

$$A_f = \sum_{t=0}^{N-1} \left(e^{-\frac{2\pi t f}{N}} \right)$$

Значения t и f не соответствуют в точности времени и частоте. Более того, f - целое число, а реальная исследуемая частота - это частота дискретизации, умноженная на f/N.

Если мы знаем значения A_f мы можем восстановить отсчеты. Для восстановления сигнала необходимо сложить все значения для разных частот. Чтобы осуществлять точное обратное преобразование Фурье, помимо амплитуды и частоты необходимо измерять фазу каждой частоты.

Для этого нужны комплексные числа. Можно изменить описанный ранее метод вычислений так, что он будет давать двумерный результат. Простое комплексное число - это двумерное значение, поэтому оно одновременно, но представляет и амплитуду, и фазу, причем измеряется две амплитуды, соответствующие разным фазам. Одна из этих фаз представляется косинусом (cos), другая - синусом (sin).

Используя комплексные числа, можно проводить измерения одновременно, умножая синусную часть на -i [3]:

$$A_f = \sum_{t=0}^{N-1} \left(S_t * cos\left(\frac{2\pi t f}{N}\right) - i * S_t * sin\left(\frac{2\pi t f}{N}\right) \right)$$

Каждое значение A_f теперь представляется комплексным числом. Действительная и мнимая части задают амплитуду двух синусоидальных волн с разными фазами. Это означает, что формула дискретного преобразования Фурье может быть представлена в виде двух сумм.

Выводы

В последнее время были достигнуты значительные успехи в идентификации личности по голосу и другим модальностям.

Однако исчерпаны далеко не все резервы по повышению надежности биометрической идентификации личности. Так, перспективными направлениями развития идентификации личности являются повышение качества предварительных исходных биометрических образцов; извлечение более робастных идентификационных признаков и их комбинаций; реализация мультимодального смешивания не на уровне оценок, а на уровне признаков различной модальности.

В настоящее время на кафедре радиотехники и защиты информации ведутся работы по исследованию и модернизации существующих систем идентификации голоса человека.

Список литературы

1. Матвеев Ю.Н., Симончик К.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 / ГрафиКон’2010. Тр.20-й межд.конф. по компьютерной графике и зрению. СПб: СПбГУ ИТМО, 2010 - 315-319 с.

2. Лоханова А.И., Симончик К.К., Козлов А.В. Алгоритм детектирования музыкальных фрагментов в задачах речевой обработки / DSPA-2010. Тр. 12-й Межд. конф. «Цифровая обработка сигналов и ее применение». - М., 2010. Т. 1. - с. 210-213.

3. Идентификация дикторов на основе сравнения статистик основного тона голоса / С.Л. Коваль, П.В. Лабутин, Е.В. Малая / Информатизация, информационная безопасность правоохранительных органов. Тр.XV Межд. науч. конф., М.: Академия управления МВД России, 2006. - 324-327 с.

4. Comparison of Voice Activity Detection Algorithms for VoIP / R. Prasad et al. / ISCC'02. Proc. 7th IEEE Symposium on Computers and Communications. Washington: IEEE Computer Society, 2002. P. 530.

5. Симончик К.К., Галинина О.С., Капустина А.И. / Алгоритм обнаружения речевой активности на основе статистик основного тона в задаче распознавания диктора / Научно-технические ведомости СПбГПУ. 2010. Т. 103. № 4., - 18-23 с.