

Н. Е. Балакирев, канд. техн. наук, проф.,
"МАТИ"—РГТУ имени К. Э. Циолковского,
М. А. Малков, ст. инженер-программист,
ЗАО ИВЦ ИНСОФТ
maksi-malkov@yandex.ru

Метод идентификации голосового сообщения

Предложен новый метод идентификации, основанный на совместном использовании нескольких составляющих голосового сообщения, представлена его структурная схема, перечислены основные сферы применения реализующего его устройства.

Ключевые слова: голос, голосовое сообщение, идентификация голосового сообщения, метод идентификации голосового сообщения.

Введение

Множество существующих подходов к определению речи могут быть разделены на два больших класса [6]. К первому относятся подходы, описывающие речь с точки зрения ее информационного содержания, а ко второму — рассматривающие речь как акустический сигнал. Речевой сигнал образуется благодаря функционированию артикуляторного аппарата человека [8, 10, 5]. Этот сигнал состоит из фонем, под которыми в данной статье понимается участок акустического речевого сигнала, возникающий при произношении звуков речи русского языка.

Идентификация по голосовому сообщению, рассматриваемому как оговариваемая заранее часть речевого сигнала, имеет ряд особенностей и преимуществ по сравнению с методами, основанными на других биометрических параметрах человека:

- возможность идентификации по смысловому содержанию голосового сообщения (ССГС);
- возможность идентификации по индивидуальным характеристикам голоса (ИХГ);
- возможность совместного использования смыслового содержания сообщения и индивидуальных характеристик голоса для целей идентификации;
- возможность удаленного снятия голосового сигнала (микрофон можно разместить практически на любом расстоянии от самого устройства идентификации);
- простота снятия голосового сигнала (не требуется дорогостоящих и сложных устройств, необходим лишь стандартный микрофон);
- возможность защиты от механического воздействия на устройство снятия голосового сигнала;
- стабильность голосовых характеристик на протяжении жизни;
- возможность гибкого изменения смыслового содержания сообщения.

Голоса имеют свои индивидуальные особенности, определяемые строением артикуляторного аппарата, поэтому голосовое сообщение можно использовать как идентификационный параметр человека — действитель-

но, наш слуховой аппарат в большинстве случаев легко различает голоса разных людей. Для той же цели может быть использована и информационная составляющая речевого сообщения (пример — использование пароля при смене караула).

Однако на уровне сигналов такое различие не является тривиальным при условии автоматизации процесса распознавания их техническим устройством.

Основа идентификации голосового сообщения

Для обработки акустического голосового сообщения необходимо провести его оцифровку [7, 9]. Оцифрованное голосовое сообщение представляет собой сумму двух сигналов: гармонического сигнала [2, 3], являющегося суммой отдельных гармоник, и негармонического сигнала. Гармоническая часть сигнала определяется индивидуальными характеристиками голоса, а негармоническая — набором фонем голосового сообщения. Таким образом, в голосовом сигнале можно выделить две составляющие — индивидуальные характеристики голоса и информационную (смысловую) часть голосового сообщения, состоящую из набора фонем [1]. Первая составляющая (индивидуальные характеристики голоса) представляет собой несущую, на которую накладывается вторая составляющая (смысловая часть голосового сообщения), таким образом формируется речь.

При идентификации голосового сообщения по индивидуальным характеристикам голоса смысл сказанной фразы неважен, следовательно, идентификация может проходить по произвольной речи. Из оцифрованного голосового сообщения выделяются индивидуальные характеристики голоса, которые впоследствии сравниваются с эталонными. При идентификации голосового сообщения по смысловой составляющей из оцифрованного голосового сигнала выделяется содержание сказанной фразы, которое сравнивается с эталонным, — таким образом, смысл фразы выступает в роли обычного пароля.

При совместном использовании обеих составляющих голосового сообщения важен как голос человека, так и смысл сказанной им фразы. В роли ключа в данном случае выступает акустическое голосовое сообщение с определенным смыслом. Надежность идентификации при таком подходе значительно выше, чем в первых двух случаях.

Необходимо сделать следующее замечание относительно использования смысловой составляющей голосового сообщения для целей идентификации. Семантическая структура пароля не является актуальной, а интересным в данном случае является лишь факт совпадения смысловых составляющих двух голосовых сообщений, эталонного и текущего. Это значительно упрощает систему, следовательно, делает ее более быстрой и надежной.

В методе, описанном в данной статье, предлагается для целей идентификации голосового сообщения использовать индивидуальные характеристики голоса, смысловое содержание голосового сообщения, а также обе эти составляющие вместе.

Структурная схема метода идентификации голосового сообщения

Сформулируем основные требования к методу идентификации голосового сообщения, согласно которым сформируем в дальнейшем его общую структурную схему.

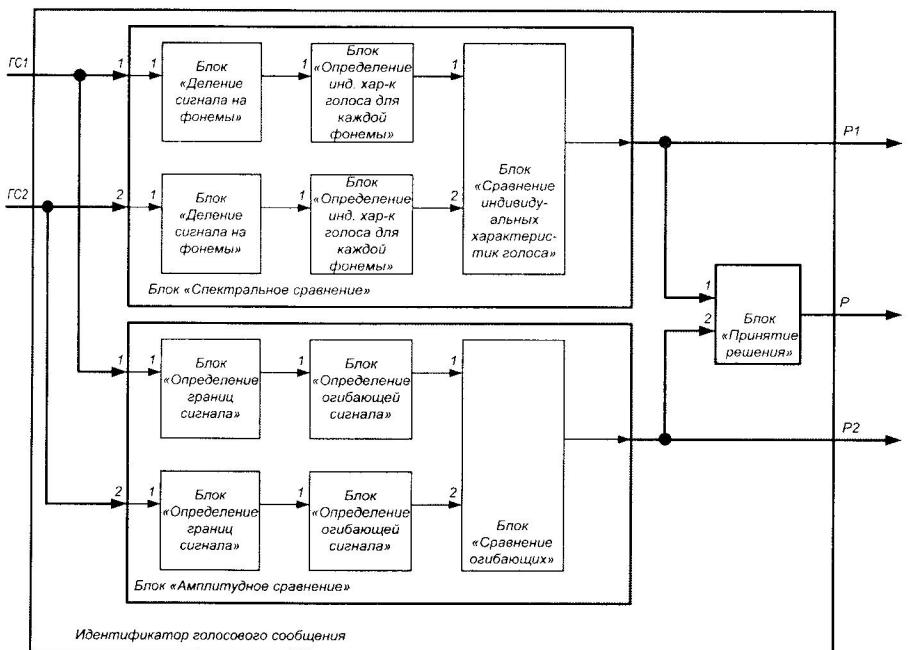


Рис. 1. Общая структурная схема идентификации голосового сообщения

му. Метод идентификации голосового сообщения должен обеспечить: низкую вероятность ошибки первого рода (неузнавание "своего"); низкую вероятность ошибки второго рода (узнавание "чужого"); высокое быстродействие; максимальную простоту реализации системы идентификации голосового сообщения.

Для идентификации голосового сообщения необходимо иметь эталонное голосовое сообщение (эталон), а при идентификации сравнивать произносимое голосовое сообщение (ключ) с эталоном по индивидуальным характеристикам голоса, по смысловой составляющей голосового сообщения либо по обеим составляющим вместе.

Предлагается следующая структурная схема идентификации голосового сообщения (рис. 1). На вход подаются два оцифрованных голосовых сообщения. Первое оцифрованное голосовое сообщение (GC_1) представляет собой заранее записанный эталон, с которым необходимо сравнивать произнесенное голосовое сообщение, а второе (GC_2) — произнесенное голосовое сообщение (ключ). На выходе схемы будут сигналы P , P_1 и P_2 . Сигнал P_1 представляет собой результат сравнения GC_1 и GC_2 по индивидуальным характеристикам голоса, сигнал P_2 — по смысловой составляющей сигналов, сигнал P — по индивидуальным характеристикам голоса и смысловым составляющим сигналов вместе. Результат (P_1 , P_2 или P) может быть либо положительным, либо отрицательным, с учетом вводимых параметров погрешности при сравнении сигналов.

После подачи на вход GC_1 и GC_2 дублируются и подаются на входы блока "Спектральное сравнение" и блока "Амплитудное сравнение". В блоке "Спектральное сравнение" GC_1 и GC_2 сравниваются по индивидуальным характеристикам голоса, а в блоке "Амплитудное сравнение" — по смысловым составляющим голосовых сообщений. Результат сравнения в обоих блоках может быть либо положительным, либо отрицательным. Если метод идентификации голосового сообщения используется для сравнения голосовых сообщений по индивидуальным ха-

рактеристикам голоса, то сигнал из блока "Спектральное сравнение" поступает на выход схемы, а если для сравнения голосовых сообщений по смысловым составляющим, то сигнал из блока "Амплитудное сравнение" поступает на выход схемы. Если метод идентификации голосового сообщения используется для сравнения по обеим составляющим, то выходные сигналы из блоков "Спектральное сравнение" и "Амплитудное сравнение" поступают в блок "Принятие решения", в котором определяется результат сравнения GC_1 и GC_2 . Результат будет положительным только в том случае, если эталон и ключ достаточно близки как по методу спектрального сравнения, так и по методу амплитудного сравнения.

Блок "Спектральное сравнение" имеет два входа и один выход. На входы 1 и 2 подаются соответственно GC_1 и GC_2 . Далее каждый из сигналов обрабатывается блоками "Деление сигнала на фонемы", и для каждой фонемы каждого сигнала определяются

в соответствующих блоках индивидуальные характеристики голоса. После этого индивидуальные характеристики голоса для GC_1 и GC_2 подаются соответственно на входы 1 и 2 блока "Сравнение индивидуальных характеристик голоса". На выходе блока будет результат сравнения, положительный либо отрицательный, по спектральному методу.

Блок "Амплитудное сравнение" имеет два входа и один выход. На входы 1 и 2 подаются соответственно GC_1 и GC_2 . Далее для каждого из сигналов определяются границы сигналов и вычисляются огибающие. На входы 1 и 2 блока "Сравнение огибающих" подаются соответственно огибающие GC_1 и GC_2 . На выходе блока будет результат сравнения GC_1 и GC_2 по амплитудному методу.

Как видно из представленной выше схемы, обработку сигналов можно распараллелить.

При реализации метода идентификации голосового сообщения используется следующий математический аппарат:

- дискретное преобразование Фурье;
- математический аппарат корреляции;
- построение интерполяционного многочлена в форме Лагранжа;
- интерполяция сплайнами;
- матричное исчисление;
- методы решения линейных уравнений;
- методы цифровой обработки сигналов.

При совместном использовании перечисленных математических методов было введено множество параметров, изменяя которые можно получать необходимое качество сравнения сигналов.

Структура комплекса идентификации голосового сообщения

Для реализации представленного выше метода наряду с основным блоком идентификации голосового сообщения необходим целый комплекс программно-аппа-

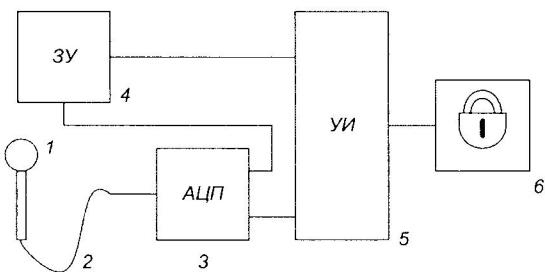


Рис. 2. Общая схема реализации комплекса идентификации голосового сообщения

ратных средств. На рис. 2 представлена общая схема комплекса идентификации голосового сообщения.

На первом этапе происходит запись эталона. Человек произносит голосовое сообщение в микрофон 1, далее по проводу 2 аналоговый сигнал с микрофона поступает в аналого-цифровой преобразователь (АЦП) 3, где преобразуется в цифровой, после чего оцифрованный сигнал записывается в память запоминающего устройства (ЗУ) 4. На втором этапе происходит запись ключа и сравнение его с эталоном. Человек произносит фразу в микрофон 1, пытаясь получить доступ, далее аналоговый сигнал от микрофона по проводу 2 поступает в АЦП 3, где преобразовывается в цифровой. Цифровой сигнал поступает на второй вход устройства идентификации (УИ) 5, на первый вход данного устройства поступает сигнал из ЗУ 4, который был записан на первом этапе. УИ работает на основе метода идентификации голосового сообщения. В данном устройстве происходит сравнение сигналов эталона и ключа. В случае равенства сигналов УИ подает сигнал на устройство 6 для открытия доступа к ресурсу, в противном случае сигнал на устройство 6 не подается и ресурс остается закрытым [4].

На практике метод идентификации голосового сообщения может быть реализован двумя способами — программно и аппаратно.

Преимуществами программной реализации являются:

- простота реализации с помощью современных средств программирования;
- отсутствие необходимости в разработке и внедрении устройств АЦП и ПЗУ, так как они присутствуют в компьютере;
- возможность использования реализованного алгоритма для разграничения доступа к информации, хранимой на компьютере.

В то же время программной реализации присущи следующие недостатки:

- необходимость наличия компьютера;
- высокая стоимость реализации в случае использования компьютера только для идентификации голосового сообщения.

Преимущества аппаратной реализации:

- высокая автономность, так как устройство не привязано к компьютеру;
- компактность устройства;
- низкая стоимость по сравнению с компьютером.

Недостатки аппаратной реализации:

- сложность по сравнению с программной реализацией;
- более сложное перепрограммирование устройства в случае изменения алгоритма.

Возможные сферы применения

Основное применение средств идентификации голосового сообщения связано с системами разграничения доступа [8], в том числе:

- к компьютерам (идентификация — ИХГ & ССГС);
- к жилым, промышленным и прочим помещениям (идентификация — ИХГ & ССГС);
- к средствам выдачи и получения денег (банкоматам) (идентификация — ИХГ & ССГС);
- к техническим устройствам, в том числе к автомобилям (идентификация — ИХГ & ССГС);
- к конфиденциальной информации по телефону (идентификация — ИХГ & ССГС);
- к системам коллективного доступа по одному паролю, известному членам коллектива (идентификация — ССГС);
- в системах идентификации по голосам (идентификация — ИХГ).

Благодаря идентификации, которая может происходить по индивидуальным характеристикам голоса, смысловой составляющей голосового сообщения и совместно по обеим этим составляющим, расширяется сфера применения и повышается качество распознавания.

Заключение

В настоящее время завершена первая версия программной реализации представленного выше метода идентификации голосового сообщения. Разработка программного комплекса велась на языке C++. Экспериментально подобраны значения необходимых параметров, обеспечивающих заданное качество сравнения сигналов. Дальнейшее направление исследований связано с разработкой и реализацией улучшенных алгоритмов обработки сигналов и определением соотношений значений параметров, повышающих качество сравнения. Предварительно проведенные эксперименты показали увеличение достоверности распознавания при совместном использовании составляющих голосового сообщения и при адаптации системы к условиям конкретного применения путем подбора параметров. Результаты этих исследований предполагается осветить в последующих публикациях.

Список литературы

1. Аграновский А. В., Ледиев Д. А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. М.: Радио и связь, 2004. 161 с.
2. Анализ и синтез речи / Под ред. Б. М. Лобанова. Минск: Изд. Института технической кибернетики АН БССР, 1991. 91 с.
3. Валиуллин А. Ф., Тюрип В. Н. Лингвистический метод определения характеристик нестационарных случайных процессов. М.: ПАИМС, 2000. 88 с.
4. Малков М. А. Выделение звуков из непрерывного речевого сигнала // Материалы Всероссийской научно-технической конференции "Новые материалы и технологии — 2006". Москва. Т. 2. М.: ИЦ МАТИ, 2006. С. 129—130.
5. Методы автоматического распознавания речи: В 2-х кн. Пер. с англ. / Под ред. У. Ли. М.: Мир, 1983. 328 с., 392 с.
6. Потапова Р. К. Речь: коммуникация, информация, кибернетика. Учеб. пос. Изд. 3-е, стереотип. М.: Эдиториал УРСС, 2003. 568 с.
7. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М. В. Назарова и Ю. Н. Прохорова. М.: Радио и связь, 1981. 496 с.
8. Рамишвили Г. С. Автоматическое опознавание говорящего по голосу. М.: Радио и связь, 1981. 224 с.
9. Соловина А. И., Улахович Д. А., Арбузов С. М., Соловьев Е. Б. Основы цифровой обработки сигналов. СПб.: БХВ-Санкт-Петербург, 2003. 576 с.
10. Сорокин В. Н. Теория речеобразования. М.: Наука, 1985. 312 с.