А.Г.Бояров, Г.М.Ваксман, Д.Н. Коновалов, С.Н.Кринов

ТЕКСТО-НЕЗАВИСИМАЯ ИДЕНТИФИКАЦИЯ ДИКТОРА

(Доклад разработан специально для Второй Биометрической Конференции «BIOMETRICS 2003 AIA RUII»)

ООО «ОТ-КОНТАКТ»
тел./факс (095) 362-49-93
e-mail:ot-kontakts@mtu-net.ru
www.ot-kontakt.webzone.ru

Параметры фиксируемого речевого сигнала являются важными биометрическими характеристиками человека. Задача измерения и анализа этих характеристик в реальных условиях осложняется наличием помех. В настоящем докладе описываются три направления работ:

1. Рабочее место эксперта – исследователя речевых сигналов. Программный комплекс измерения и анализа биометрических характеристик голоса.
2. Автоматическая оперативная идентификация личности по голосу.
3. Шумоочистка и фильтрация.

Идентификация личности по голосу и речи с участием человека – эксперта

Несмотря на достигнутые успехи в разработке автоматических систем идентификации и верификации личности по голосу и речи, потребность в экспертных системах, где окончательное право принятия решения об идентичности/неидентичности принадлежит эксперту (человеку), не снижается.

Это объясняется тем, что для автоматических систем идентификации качество используемого речевого сигнала должно удовлетворять определенным, иногда довольно жестким условиям, которые реализуются не во всех реальных ситуациях. В некоторых случаях величина ошибки, допустимая автоматическими системами, может быть недопустимо большой, например, для криминалистики. Принятие решение об идентификации зависит от множества факторов. Определение этих факторов в ряде случаев возможно только с участием человека-эксперта. Эксперту-криминалисту приходится учитывать наличие или отсутствие монтажа, физическое и эмоциональное состояние, лингвистические особенности, смысловое содержание и тому подобное, причем с учетом высоких требований к точности результата. Часто упоминаемая характеристика, как «Дружественное отношение» пользователя, например, голосового ключа, может полностью отсутствовать в случае криминалистического исследования фонограммы. В настоящее время стоит задача максимально автоматизировать труд эксперта, создать ему эффективное рабочее место. К тому же экспертная система может

оказаться хорошим инструментом для исследования и поиска индивидуальных особенностей, разработки различных алгоритмов и методик для дальнейшего использования в автоматических системах идентификации и верификации.

Отметим, что один только слуховой анализ человека не может обеспечить высокую надежность идентификации. Это объясняется предельными возможностями человеческого восприятия, способного одновременно контролировать (удерживать в памяти) ограниченное число факторов (параметров). При этом слух человека, проводящего слуховой анализ речи, должен быть тренирован именно на такую работу. Например, слуховое восприятие музыканта, диалектолога, подводника-акустика ориентировано в каждом случае на восприятие различных, сложных акустических сигналов и требует различной слуховой подготовки. Наша общая способность узнавать знакомых по голосу не должна вводить в заблуждение, поскольку мы умеем не только узнавать, но и ошибаться. При этом мы редко стремимся к количественной оценке точности своих идентификационных способностей. Кроме того, в обыденной жизни мы вряд ли сталкиваемся с необходимостью узнавания одного из нескольких сотен незнакомых нам голосов, к тому же не имеющих ярко выраженных «особых примет». Когда же мы это делаем, то становится очевидным, что точность слуховой идентификации не может быть гарантированно применена в критических случаях. Очевидно, что для проведения надежной слуховой идентификации человек должен обладать определенными способностями.

На первый взгляд кажется, что звучащая речь подобна письменной, а буквы представляют собой звуки. Тем не менее, во многих отношениях речь не соответствует письму. Например, в беглой речи не существует пробелов между словами, звуки в речи не имеют однозначного соответствия буквам и эти звуки не являются дискретными единицами, произношение слова может значительно изменяться в зависимости от фразы, в которой оно звучит. Произношение звука речи изменяется с изменением окружающих его звуков в слове и т.д.

В речи не существует специфических, индивидуальных параметров. Одни и те же параметры несут информацию:

- о смысле сказанного,
- о том кто говорит и
- о том, как он говорит.

Непонимание перечисленных и других более специальных (технических и научных) аспектов может привести к ошибочным суждениям при интерпретации параметров речевого сигнала и возможностях слухового узнавания человека.

Тем не менее, убежденность в успешном решении задачи определения личности по голосу (идентификации) основана на том, что звучащая речь, как и многие другие формы деятельности человека как созидательной, так и противоправной, являются выражением его индивидуальности (уникальности).

Слуховое восприятие с одновременным просмотром отображаемых на экране характеристик речевого сигнала, с возможностью фиксации отмечаемых экспертом событий, использование математических методов анализа набора событий, могут

расширить возможности эксперта при решении задачи идентификации по голосу. Возможности одновременного представления текстовой расшифровки, формы сигнала, графического представления параметров речевого сигнала, статистических оценок, отображение различных знаний, и т.п. повышают эффективность действий эксперта при решении им задачи идентификации личности по голосу и речи. Увязанное с фонограммой текстовое представление зафиксированного на фонограмме разговора, позволяет использовать текстовый процессор, который поможет эксперту автоматизировать поиск необходимых для идентификационного исследования фрагментов речи, оценить степень стабильности произношения говорящего, определить характеристики его речевого поведения и т.п.

Для того чтобы определить перечень отображаемых характеристик речевого сигнала, необходимо представлять себе, каким образом индивидуальность говорящего отражается в его речи.

Индивидуальность речи определяется комплексом параметров, на которых отражаются:

- размеры речевого тракта человека,
- характеристики мышц, управляющих губами, языком, челюстью, гортанью,
- длина и толщина голосовых связок,
- характеристики мышц управления голосовыми связками,
- объем легких и характеристики мышц грудной клетки, определяющие цикл речевого дыхания и ритмическую структуру речи,
- речевых навыков артикуляторных движений,
- речевое поведение, отражающееся в построении речевых высказываний (повествовательных, вопросительных, побудительных, реактивных и т.п.)
- особые приметы или дефекты речи.

Индивидуальные качества разных людей в разных ситуациях различным образом проявляются в комплексе параметров, отражающем перечисленные выше факторы. Кроме того, ряд параметров при использовании автоматических средств определяются со значительными ошибками, особенно, если исследуемый разговор происходит в присутствии акустических и/или канальных помех. Необходимость участия эксперта на этом этапе и определяет потребность в экспертных системах. В свою очередь, экспертная система может предоставлять эксперту максимальный сервис при оценке конкретной ситуации и коррекции ошибок, а также:

- автоматизировать процедуры накопления локализуемых экспертом «событий»,
- проводить анализ накопленных событий,
- оценивать противоречивости/непротиворечивости проведенного анализа для разных событий,
- предлагать эксперту на основе полученных результатов анализа и оценок конкретный сценарий продолжения исследования,
- автоматизировать процесс создания текста Отчета (Заключения эксперта).

В большинстве систем автоматической идентификации личности по голосу учитывается только небольшая часть перечисленных выше факторов. Такие системы могут работать только на ограниченном числе голосов, поскольку используемые параметры существенно пересекаются на всей человеческой популяции. Так размеры речевого тракта у многих людей могут совпадать, могут совпадать и характеристики голосовых связок, что и подтверждается многолетним опытом научных и экспертных исследований. Кроме того, голоса неравномерно распределены в рассматриваемых диапазонах признаков. Так средняя частота колебания голосовых связок у мужчин изменяется от 80 до 170 Гц. Однако мужские голоса со средней частотой 80 Гц встречаются значительно реже, чем со средней частотой 140 Гц.

Поэтому ограниченный набор параметров обеспечивает только идентификацию группы лиц, либо идентификацию личности в пределах ограниченной группы, состоящей из лиц с непересекающимся набором используемых параметров, что не всегда возможно.

К тому же в системах автоматической идентификации по короткой парольной фразе, как правило, допускается повторять попытки произнесения 3 - 5 раз, что свидетельствует о большой степени случайности используемых для идентификации параметров.

Звучащая речь возникает в результате движений артикуляторных органов говорящего (движений языка, челюсти, губ, голосовых связок). Характер этих движений определяется сформированными навыками.

Речевые навыки формируются в течение длительного времени под влиянием речевой культуры лиц, окружающих ребенка и затем подростка в период накопления его речевого опыта, который может изменяться в течение всей жизни. Как и любой другой вид движения, проявляющийся в походке, жестикуляции, почерке, вышивании, рисовании и т.п., движение артикуляторных органов определяется также антропологическими и психофизиологическими характеристиками личности. Таким образом, речь, как и любой другой вид двигательной активности конкретного человека, находит отражение в огромном количестве параметров. Среди этого огромного количества существует определенный набор устойчивых параметров, которые мало меняются со временем и с изменением состояния человека.

Система «ОТ-Expert» предназначена для проведения инструментальной части исследования при криминалистическом исследовании фонограмм. Система может быть также использована в научных целях, при проведении исследований в области экспериментальной (инструментальной) фонетики, а также в учебных целях в рамках преподавания курса экспериментальной фонетики.

Система работает с любой звуковой картой, совместимой с Windows, позволяя вводить акустический сигнал с микрофона или линейного входа, и использует принятый в Windows формат кодирования звуковых файлов.

Система позволяет получать и отображать на экране стандартный набор характеристик речевого сигнала, как в графической, так и числовой форме, делать пометки фрагментов, вызвавших интерес у исследователя, производить орфографическую или транскрипционную расшифровку фонограммы или ее фрагментов, масштабирование графического отображения и редактирование вычисляемых в процессе анализа характеристик.

Возможны следующие отображения фонограммы и ее фрагментов:

- Осциллограмма,
- спектрограмма,
- кепстрограмма,
- интонограмма,
- сглаженная спектрограмма,
- формантные траектории.

Во всех перечисленных выше отображениях пользователь имеет возможность изменять цветовую палитру и масштаб по обеим осям. В случае спектрографического, сглаженного спектрографического и кепстрографического отображения, пользователь также имеет возможность изменять размер и тип окна анализа, а также изменять освещенность (контраст графического отображения).

Любая комбинация возможных видов экранного отображения об исследуемом участке фонограммы может быть экспортирована в виде графического файла и использована для иллюстрации при составлении текстового документа (описания исследования, отчета, заключения эксперта и т.п.)

Каждое наблюдаемое исследователем явление может быть фиксировано и автоматически сохранено в виде сложного объекта в базе исследуемых явлений с

примечаниями пользователя в некоторой категориальной шкале. Накопленные в базе явления подвергаются автоматической обработке, результатом, которой может быть подтверждение либо не подтверждение выдвигаемых исследователем гипотез.

Минимальные требования к аппаратным средствам, обеспечивающим удовлетворительное функционирование OTExpert: Celeron 333, 32Mb RAM, video system 800x600.

Для эффективной работы системы необходимы: Celeron 800, 128Mb RAM, video system 1024x768

Система автоматической идентификации по голосу

Как уже отмечалось выше, при определенных условиях, можно достоверно измерять идентификационные параметры автоматическим путем. Если эти условия не изменяются в определенном технологическом процессе, то тогда возможно использование автоматической системы идентификации. Сами условия эксплуатации подобной системы могут служить дополнительным ограничением, стабилизирующим ряд параметров, влияющих на процесс идентификации и повышающим ее надежность. Например, используется один и тот же канал записи (телефонный канал, микрофон и т.п.) или идентификация (верификация) производится на ограниченном числе голосов, произносится один и тот же текст и т.п.

Разработанная фирмой «ОТ-КОНТАКТ» тексто-независимая система автоматической идентификации предназначена для работы в среде Windows 95/98/NT/2000/ME/XP. Система работает следующим образом:

- фонотека (библиотека голосов) предварительно заполняется файлами с необходимыми голосами;
- запись неизвестного голоса предъявляется на вход системы;
- система автоматически сортирует голоса фонотеки по степени их близости к неизвестному голосу;
- верификация производится с использованием пяти ближайших (исследование свой/чужой);
- система тестировалась на базе данных заказчика (100 голосов, частота дискретизации 8000Гц, отношение сигнал/шум в пределах 5 – 15 дБ).

Характеристики системы:

- точность поиска ближайшего (если он существует) более 98%;
- время создания эталона по файлу фонотеки составляет 200-300 мс (PC Athlon 1200);
- время, затраченное на идентификацию на библиотеке в 100 голосов, составляет 1500-1800 мс;
- размер эталона одного голоса составляет 24 кБ;
- для создания эталона используется фрагмент записи голоса длительностью 20-50с ;
- максимально возможный размер фонотеки – 1000 голосов.

Системы выпускается в двух версиях:

- в форме самостоятельных программных модулей с графическим интерфейсом, контекстно-зависимой помощью (встроенный help) и т.п.
- в форме модулей DLL, встраиваемых в программное обеспечение заказчика.

В настоящее время система находится в эксплуатации у заказчика. В зависимости от требований к величине ошибок первого и второго рода, система может быть настроена со следующим соотношением между обеими ошибками:

	Процент ошибок						
Классификация «своего» голоса в качестве чужого – **FRR**	0.075	0.071	0.53	0.041	0.032	0.022	0.01
Классификация «чужого» голоса как своего -**FAR**	0.0014	0.0022	0.0033	0.0068	0.032	0.044	0.06

Цифровая фильтрация

Реализованы методы проектирования следующих типов цифровых фильтров:

1) частотно-селективный (низкочастотный, высокочастотный, полосовой, режекторный) фильтр с конечной импульсной характеристикой (КИХ) и линейной фазо-частотной характеристикой (ФЧХ);
2) КИХ-фильтр с линейной ФЧХ и произвольной кусочно-линейной «целевой» амплитудно-частотной характеристикой (АЧХ);
3) КИХ-фильтр с линейной ФЧХ и АЧХ, имеющей минимальное средневзвешенное квадратичное отклонение от «целевой» АЧХ, заданной на непересекающемся наборе частотных интервалов;
4) частотно-селективный фильтр Баттерворта (фильтр с бесконечной импульсной характеристикой – БИХ-фильтр);
5) частотно-селективный БИХ-фильтр Чебышева типа I;
6) частотно-селективный БИХ-фильтр Чебышева типа II.

Коэффициенты передаточной функции КИХ-фильтров 1-3 определяются из условия минимума среднеквадратичного отклонения АЧХ фильтра заданного порядка от «целевой» АЧХ. В случае КИХ-фильтров 1, 2 полученный набор коэффициентов подвергается оконному взвешиванию с целью снижения локальных «выбросов» АЧХ в окрестности точек разрыва «целевой» АЧХ (эффект Гиббса). В распоряжении пользователя имеются следующие типы окон: прямоугольное, Ханна, Хемминга, Бартлетта-Ханна, Блэкмана, Блэкмана-Харриса, Гаусса, Кайзера, Тьюки.

Процедура построения коэффициентов числителя и знаменателя передаточной функции БИХ-фильтров 4-6 включает следующие 3 этапа:

-определение полюсов и нулей частотной характеристики (ЧХ) аналогового нормированного (с единичной частотой среза) фильтра соответствующего типа;
-преобразование ЧХ аналогового нормированного фильтра в ЧХ аналогового частотно селективного фильтра (низкочастотного, высокочастотного, полосового, режекторного);
- билинейное преобразование ЧХ аналогового фильтра в ЧХ цифрового фильтра.

При проектировании фильтра пользователь должен задать следующие входные параметры:

- тип фильтра (один из указанных выше шести типов);
- порядок фильтра;
- параметры «целевой» АЧХ:
  для частотно-селективных фильтров 1, 4, 5, 6 – частоты срезов;
  для фильтров 2, 3 – набор точек $\{\omega_i, A_0(\omega_i)\}$ «целевой» АЧХ $A_0(\omega)$;
- для фильтров 2, 3 – тип окна;
- для фильтра 3 – набор весов (по числу частотных интервалов);

- для фильтров Чебышева – параметр «волнистости» АЧХ в полосе пропускания (фильтр 5) или в полосе задержания (фильтр 6).

После построения коэффициентов передаточной функции цифрового фильтра пользователь имеет возможность произвести фильтрацию дискретного сигнала. Реализованы два варианта фильтрации:

1) прямая форма II фильтрации с произвольным набором начальных условий;
2) двунаправленная фильтрация без фазовых искажений с выбором оптимальных начальных условий.

В случае двунаправленной фильтрации фактическая частотная характеристика в точности совпадает с квадратом АЧХ исходного фильтра.

Шумоочистка речевого сигнала

Реализованные алгоритмы шумоочистки основаны на следующих стандартных допущениях о свойствах шумовой составляющей сигнала:

1) адитивность;
2) некоррелированность с речевой составляющей;
3) квазистационарность.

Общая схема шумоочистки включает два этапа:

1) оценка спектра мощности шума;
2) собственно шумоочистка речевого сигнала.

Оценка спектра мощности шума

Предусмотрены 2 варианта оценки спектра мощности шума:

1) с использованием заранее выделенных сегментов типа «Пауза»;
2) без использования сегментов типа «Пауза».

Первый вариант оценки включает следующую последовательность действий:

- каждый сегмент паузы сканируется окном анализа длительностью 20-40 мс, в каждом окне анализа вычисляется дискретный спектр мощности;
- полученные дискретные спектры последовательно усредняются по заданному числу окон;
- полученная неравномерная сетка усредненных спектров (поскольку сегменты типа «Пауза» расположены, как правило, нерегулярно) интерполируется в равномерную сетку с заданным временным шагом, зависящим от характерного времени изменения свойств шума (обычно порядка 1 секунды).

Второй вариант оценки спектра мощности шума основан на анализе полосовых спектров мощности зашумленного сигнала. При этом принимаются следующие допущения:

1) каждый из полосовых спектров мощности зашумленного сигнала $P_x^{(q)}(k)$ равен сумме соответствующих полосовых спектров мощности незашумленного сигнала $P_s^{(q)}(k)$ и шума $P_n^{(q)}(k)$ (q – номер частотной полосы, k – номер окна анализа);
2) в силу квазистационарности шума полосовой спектр его мощности $P_n^{(q)}(k)$ имеет существенно более «плавное» поведение во времени (относительно индекса k), чем соответствующий полосовой спектр незашумленного речевого сигнала $P_s^{(q)}(k)$, который в условиях нормальной речевой активности достаточно часто (со средней частотой порядка нескольких герц) принимает значение, близкое к нулевому.

При указанных допущениях в качестве нижней оценки полосового спектра мощности шума $\hat{P}_n^{(q)}(k)$ может использоваться любая «достаточно медленная» огибающая локальных минимумов функции $P_x^{(q)}(k)$. В качестве такой огибающей используется функция следующего вида :

$$\hat{P}_n^{(q)}(k)=\begin{cases} P_x^{(q)}(k), \text{если } P_x^{(q)}(k)\le \hat{P}_n^{(q)}(k-1) \\ D_\tau \hat{P}_n^{(q)}(k-1)+(1-D_\tau)P_x^{(q)}(k), \text{если } P_x^{(q)}(k)> \hat{P}_n^{(q)}(k-1) \end{cases} \quad (1)$$

где $D_\tau = \exp(-t_w/\tau)$;

t_w – временной шаг окон анализа (обычно порядка 10-20 мс);

τ – характерное время изменения огибающей, которое должно иметь порядок характерного времени изменения мощности шума (обычно 0.5-5 секунд для квазистационарного шума).

Нетрудно видеть, что эта функция действительно является нижней огибающей функции $P_x^{(q)}(k)$, т.е. $\hat{P}_n^{(q)}(k)\le P_x^{(q)}(k)$, и область ее определения состоит из чередующихся участков резкого падения (когда выполнено первое из условий (1)) и плавного роста с характерным временем изменения τ (когда выполнено второе из условий (1)).

Таким образом, во втором варианте оценки спектра мощности шума имеет место следующая последовательность действий:

- полный частотный диапазон $[0, f_N]$ (f_N – частота Найквиста) разбивается на n_f частотных полос ($[w_q f_N, w_{q+1} f_N]$ – частотная полоса с номером q, $q=0, \ldots, n_f-1$, $0\le w_q<w_{q+1}\le 1$);
- зашумленный сигнал сканируется окном анализа длительностью 20-40 мс, в каждом окне анализа вычисляются n_f значений полосового спектра мощности

$$P_x^{(q)}(k)=\sum_{i=i_q}^{i=i_{q+1}} c(i)\left|X_w(k,i)\right|^2 \quad (2)$$

где $X_w(k,i)$ – отсчеты дискретного преобразования Фурье зашумленного сигнала в k-ом окне анализа ($i=0, \ldots, [N_w/2]$; […]- целая часть);

i_q, i_{q+1} – граничные отсчеты q-ого частотного интервала (связь между безразмерной частотой w и номером отсчета спектра i имеет вид $w=2i/N_w$);

$c(i)$ – коэффициент учета симметричности амплитудного спектра:

$c(i)=1$, если $i=0$ или $2*i=N_w$;

$c(i)=2$ в противном случае.

- для каждой частотной полосы строится оценка мощности шума $\hat{P}_n^{(q)}(k)$ в виде «медленной» огибающей минимумов функции $P_x^{(q)}(k)$, одновременно производится увеличение шага оценки спектра мощности шума до требуемой величины (порядка 1 секунды);
- каждый из полученных таким образом полосовых спектров мощности шума преобразуется в дискретный спектр путем равномерного распределения мощности в каждой полосе по соответствующим отсчетам дискретного спектра.

При выборе количества частотных полос n_f необходимо принимать во внимание следующие два обстоятельства:

1) увеличение числа полос приводит к уменьшению количества отсчетов дискретного спектра в каждой из них, при этом возрастает погрешность принятого допущения об адитивности полосовых спектров мощности, а также возрастает амплитуда статистических флуктуаций компонент полосового спектра (даже в пределах сегментов типа «Пауза»). В свою очередь это приводит к занижению оценки спектра мощности шума в каждой частотной полосе, поскольку эта оценка строится, как огибающая минимумов.
2) уменьшение числа полос приводит к более грубой оценке формы спектра мощности.

Если окно анализа содержит несколько сотен отсчетов, то оптимальное значение величины n_f лежит, по-видимому, в диапазоне от 8 до 16.

Шумоочистка

Общая схема шумоочистки основана на преобразовании кратковременных амплитудных спектров зашумленного сигнала, которое может трактоваться, как нестационарная фильтрация, с последующим синтезом во временной области, используя известный метод «overlap-add». Ниже схема шумоочистки излагается более детально.

1)Зашумленный сигнал сканируется окном анализа длительностью 20-40 мс. Для текущего окна производятся вычисления следующих величин:
- дискретного спектра мощности шума $P_n(k,i)$ (k – номер окна анализа; i – номер отсчета спектра; i=0, …, [N_w/2]). При этом используется интерполяция полученных заранее оценок спектра мощности шума (см. раздел 2.1);
- дискретного спектра зашумленного сигнала $X_w(k,i)$ (используется дискретное преобразование Фурье).

2)Вычисляется оценка дискретного спектра незашумленного сигнала $S_w(k,i)$, которая в общем случае имеет следующий вид:

$$S_w(k,i) = G(k,i)X_w(k,i) \qquad (3)$$

где $G(k,i)$ – набор вещественных положительных коэффициентов усиления, который может трактоваться как частотная характеристика нестационарного фильтра, модифицирующего только амплитудный спектр.

Коэффициенты $G(k,i)$ в общем случае зависят от следующих величин:
- спектра мощности шума $P_n(k,i)$;
- амплитудного спектра зашумленного сигнала $|X_w(k,i)|$;
- коэффициентов усиления $G(k$-1,$i)$ в предыдущем окне анализа;
- амплитудного спектра $S_w(k$-1,$i)$ в предыдущем окне анализа;
- набора постоянных параметров шумоочистки (коэффициентов подавления и т.д.).

Ниже для трех алгоритмов шумоочистки приводятся выражения для коэффициентов усиления $G(k,i)$.

3) Полученная оценка дискретного спектра незашумленного сигнала $S_w(k,i)$ подвергается обратному дискретному преобразованию Фурье и используется для восстановления незашумленного сигнала во временной области.

Реализованы следующие алгоритмы шумоочистки:

1) Спектральное вычитание:

$$G(k,i)=\left[\max\left\{1-\beta\frac{[P_n(k,i)]^{\alpha/2}}{|X_w(k,i)|^{\alpha}},\ \varepsilon^{\alpha}\frac{[P_n(k,i)]^{\alpha/2}}{|X_w(k,i)|^{\alpha}}\right\}\right]^{1/\alpha} \qquad (4)$$

где α - степенной параметр ($\alpha \approx 1$ .. 2);

β - коэффициент подавления шума ($\beta \approx 1$ .. 5);

ε - коэффициент шумовой «подложки», используемой для маскировки остаточного «музыкального» шума ($\varepsilon \approx 0.1$ .. 0.3).

2) Фильтр, основанный на оценке амплитуды незашумленного сигнала по критерию максимального правдоподобия :

$$G(k,i)=G(k-1,i)+\beta\left[\hat{G}(k,i)-G(k-1,i)\right] \qquad (5)$$

$$\hat{G}(k,i)=0.5\left[1+\sqrt{\max\left\{1-\frac{P_n(k,i)}{|X_w(k,i)|^2},0\right\}}\right]\frac{\exp(-\xi)I_0\left[2\sqrt{\xi\frac{|X_w(k,i)|^2}{P_n(k,i)}}\right]}{1+\exp(-\xi)I_0\left[2\sqrt{\xi\frac{|X_w(k,i)|^2}{P_n(k,i)}}\right]} \qquad (6)$$

где ξ - коэффициент подавления шума ($\xi \approx 10 \ldots 30$);

β - коэффициент сглаживания, используемый для подавления «музыкального» шума ($\beta \approx 0.3 \ldots 0.5$);

$I_0(x)$ – модифицированная функция Бесселя 1-го рода 0-го порядка.

3) Фильтр, основанный на оценке амплитуды незашумленного сигнала по критерию минимальной среднеквадратичной ошибки:

$$G(k,i)=\frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{1+R_{post}(k,i)}\right)\left(\frac{R_{prio}(k,i)}{1+R_{prio}(k,i)}\right)}\times M\left[\left(1+R_{post}(k,i)\right)\left(\frac{R_{prio}(k,i)}{1+R_{prio}(k,i)}\right)\right] \qquad (7)$$

где $M(x)=\exp(-x/2)[(1+x)I_0(x/2)+xI_1(x/2)]$;

$I_0(x)$, $I_1(x)$ - модифицированные функции Бесселя 1-го рода 0-го и 1-го порядков;

$R_{post}(k,i)=\frac{|X_w(k,i)|^2}{P_n(k,i)}-1$ - апостериорная оценка отношения «сигнал/шум» для i-ой спектральной компоненты;

$R_{prio}(k,i)=(1-\alpha)\max\{R_{post}(k,i),0\}+\alpha\frac{|S_w(k-1,i)|^2}{P_n(k,i)}$ - априорная оценка отношения «сигнал/шум» для i-ой спектральной компоненты;

α - параметр сглаживания ($\alpha \approx 0.98$).

По-видимому, невозможно однозначно указать, какой из трех алгоритмов шумоочистки является «самым лучшим». Выбор наиболее подходящего алгоритма будет зависеть от следующих характеристик:

1) от уровня и характера шума в анализируемом сигнале;
2) от требований, предъявляемых к шумоочистке (степень оперативности, возможность интерактивной настройки параметров шумоочистки, требуемое соотношение «разборчивость-качество речи» и т.д.).

Если требуется минимальное вмешательство пользователя в процесс настройки параметров шумоочистки, то имеет смысл использовать алгоритм 3, поскольку в этом алгоритме имеется всего один параметр, причем оптимальное значение этого параметра лежит в весьма узком диапазоне значений (параметр сглаживания $\alpha \approx 0.98$). К достоинствам этого алгоритма следует отнести достаточно высокое качество шумоочистки, а также практически полное отсутствие «музыкального» шума в очищенном сигнале. Если же пользователя не устраивает качество шумоочистки с использованием этого алгоритма, то имеет смысл попытаться улучшить его, применяя алгоритмы 1, 2, имеющие большее количество настраиваемых параметров, причем в отличие от параметра сглаживания в алгоритме 3 каждый из параметров алгоритмов 1, 2 может изменяться в весьма широком диапазоне. При этом достаточно плавно изменяются

характеристики разборчивости и качества очищенного сигнала. Основной недостаток алгоритмов 1, 2 состоит в том, остаточный шум зачастую носит ярко выраженный «музыкальный» характер.

Вопросы и ответы:

1. В: Где применяется или может применяться идентификация личности по голосу?

О: Голос, как и другие биометрические характеристики, может применяться для идентификации пользователя в системах разграничения доступа. Кроме того, голос незаменим как биометрическая характеристика в системах используемых телефонный канал для сообщения конфиденциальной информации, например сообщения информации о состоянии лицевого счета в банковских системах. Кроме этого идентификация диктора по голосу и речи широко применяется в разных странах в криминалистике, тем самым, расширяя возможную доказательную базу.

2. В: Часто можно слышать термины «идентификация» и «верификация». Какая между ними разница?

О: **Верификация** – процесс установления принадлежности неизвестного речевого образца и речевого эталона одному и тому же голосу. Иными словами: «Произнесены ли образец и эталон одним и тем же человеком»?

Идентификация – процесс установления кому из ограниченной группы лиц принадлежит голос. Иными словами: «На чей эталон из группы голосов дикторов наиболее похож исследуемый образец»? Следует отметить, что в отличие от верификации, идентификация не решает вопрос о принадлежности образца и эталона одному и тому же голосу, а лишь находит самый похожий голос. Но существует также понятие «**открытой идентификации**», которое означает - процесс установления наиболее похожего из группы дикторов и после этого решения задачи верификации, либо процесс многократной верификации по каждому из группы дикторов.

3. В: Люди подчас ошибаются, не узнавая по голосу знакомых, либо ошибочно принимая чужой голос за голос знакомого человека. Возможно ли в принципе определить принадлежность речи определенному лицу?

О: Да возможно. Голос содержит достаточно индивидуализирующей информации, чтобы проводить идентификацию. Прецизионных результатов идентификации в самых разнообразных условиях на сегодняшний момент можно добиться только путем проведения экспертного человеко-машинного исследования. Автоматические системы тоже применимы, но с некоторыми ограничениями, это касается, например, длительности отрезка речи в тексто-независимых системах или фиксации фразы в тексто-зависимых системах.

4. В: Пародисты довольно хорошо имитируют голоса известных людей. Может ли система или человек-эксперт с помощью компьютера отделить голос имитатора от голоса, имитируемого им человека?

О: Пародисты имитируют тембровые характеристики и манеру речи (характерные речевые обороты, слова, речевые ошибки и т.п.) известных людей. С другой стороны

слушатели, как правило, хотят слышать в речи пародистов знакомый тембр голоса и характерные словечки, поскольку это одно из неосознаваемых условий этого жанра. Пародисты не могут имитировать голос произвольного человека, к тому же не имеющего «особых примет», заключающихся в необычном звучании и особенном речевом поведении. Проведенные эксперименты показывают, что при имитации (сознательном изменении своего голоса) в голосе имитатора сохраняется множество собственных индивидуальных характеристик.

5. В: Зачем нужно рабочее место эксперта для проведения идентификации с участием человека, если существуют системы автоматической идентификации личности?

О: Системы автоматической идентификации работают на ограниченном круге голосов. Кроме того, для успешного функционирования автоматических систем необходимо выполнение еще ряда условий, которые не всегда выполняются. Не все условия могут фиксироваться автоматически. При сравнении двух голосов длительности участков речи, сопоставимых для сравнительного исследования может быть недостаточно, что тоже пока невозможно определять автоматическим путем (например, речь в состоянии опьянения и т. п.). Автоматические системы применяются также в тех случаях, когда человек либо не подозревает, что он является участником задачи идентификации, либо «сотрудничает» с системой, стараясь произносить заданную фразу «как обычно», также как при повторении своей подписи на бумажном документе. В криминалистике ситуация часто совсем другая. «Подозреваемое» лицо не всегда «сотрудничает» и иногда сознательно меняет голос. В этих условиях также важно понимать, что критерий ошибки криминалиста должен быть значительно жестче, чем ошибки, допускаемые современными автоматическими системами.

6. В: Существуют технические средства изменения звучания голоса. При этом мужской голос может звучать как женский или детский, а детский как мужской и т.д. Какова возможность идентификации личности в таких условиях?

О: Технические системы изменения голоса существуют уже с десяток лет. Выпускаются как отдельные микросхемы, так и телефонные аппараты с функцией изменения голоса. Функция изменения голоса является атрибутом многих звуковых редакторов, например CoolEdit. Стандартная функция изменения тона входит в состав мультимедийной библиотеки программной среды Windows.

Здесь важно определить сам факт изменения голоса. По мнению авторов много уголовных дел, связанных с анонимными угрозами, ушли в разряд нераскрытых из-за того, что факт изменения голоса в этих случаях не был установлен. А не установлен он был по простой причине. Оперативные службы, не имея информации об упомянутых системах и не подозревая о возможности технического изменения голоса, сочли эти голоса, как принадлежащие не установленным лицам.

Если факт изменения установлен, необходимо определить алгоритм изменения и его параметры, после чего восстановить исходный голос. Таким образом, техническая возможность идентификации в условиях применения алгоритмов изменения голоса существует. Но не нужно забывать и о юридической стороне. Примет ли суд в качестве доказательства фонограмму, на которой зафиксирован измененный техническими средствами голос?

7. В: Голос человека изменяется с годами. Какова в таком случае ситуация с возможностью идентификации по голосу?

О: В криминалистике время между двумя моментами фиксации некоторого признака называется «идентификационным периодом». Предполагается, что на идентификационном периоде идентификационные признаки не изменяются. С годами

меняются свойства голосовых связок. Известна «ломка голоса» у подростков, к старости часто тембр голоса меняется в связи с износом голосовых связок, заболеваниями горла и т.п. Человек может потерять (удалить) ряд зубов, влияющих на его произношение, может протезировать зубы и т.п. В тоже время навыки движения артикуляторных органов в основном сохраняются.

Ответ на данный вопрос требует дополнительных исследований по представительным речевым БД. Хотя результаты, полученные при изучении единичных случаев, где «идентификационный период» превышал 20 лет, свидетельствовали о сохранении артикуляционных навыков.

8. В: Меняется ли в настоящее время ситуация в области идентификации личности по голосу в связи с ростом вычислительной мощности используемых компьютеров?

О: Современные вычислительные средства позволяют одновременно отображать на экране различные виды представления речевого сигнала, их локальные, темповые, интегральные и статистические характеристики, проводить накопление и автоматический анализ накопленных явлений. Это в свою очередь также может отображаться в интерпретируемом виде. Самые сложные и громоздкие алгоритмы шумоочистки могут выполняться в реальном времени. Все это расширяет возможности человека-эксперта при принятии решения, а также позволяет в ряде случаев использовать автоматические средства идентификации на больших (до 1000 голосов) фонотеках.

9. В: Какие признаки являются наиболее информативными при описании индивидуальных особенностей голоса?

О: Используемые в процессе идентификации признаки имеют различный идентификационный вес. Ряд признаков, такие как средний основной тон, средний спектр могут использоваться для принятия решения, к какой группе голосов относится исследуемый голос. Другие признаки, такие как динамические характеристики формантных траекторий, отражающие динамические характеристики артикуляторных органов и динамические характеристики основного тона, отражающие интонационные особенности голоса, а также особенности речевого поведения имеют высокую идентификационную значимость. Кроме того, информативность признака индивидуальна для диктора и зависит от принимаемого значения.

10. В: Есть что–нибудь общее между задачами идентификации личности по голосу и определения автора по тексту?

О: Определение характеристик речевого поведения (построение фраз, поведение в диалоге, монологе, повествовательных, вопросительных, реактивных, побудительных и т.п. высказываниях) напрямую связано с задачей определения автора текста. Обе задачи направлены на решение поиска автора, пересекаются по ряду признаков, дополняя друг друга.

11. В: Каким образом тестировалась автоматическая система оперативной идентификации?

О: Система тестировалась на базе данных заказчика (100 голосов, частота дискретизации 8000Гц, отношение сигнал/шум в пределах 5 – 15 дБ).

12. В: Чем системы идентификации, описанные в данной публикации, отличаются от систем других разработчиков?

О: Экспертная система:

- позволяет эксперту наблюдать одновременно большее число отображений сигнала и вычисляемых по сигналу или фрагменту характеристик;

- позволяет одновременно анализировать текст и любое из стандартных представлений сигнала;
- позволяет производить в двух текстах поиск одинаковых слов или близких артикуляционных событий для быстрого сравнительного анализа соответствующих звучащих фрагментов;
- содержит эффективные средства вычисления, коррекции и представления основного тона голоса;
- содержит удобный интерфейс, автоматизирующий все этапы работы эксперта, начиная от вычисления характеристик и заканчивая переносом результатов измерений в текст Заключения эксперта.

Автоматическая оперативная система:
- Большой объем фонотеки до 1000 голосов,
- точность поиска ближайшего (если он существует) более 98%,
- точность верификации свыше 93%,

время создания эталона по файлу фонотеки составляет 200-300 мс (PC Athlon 1200),
- время, затраченное на идентификацию на библиотеке в 100 голосов, составляет 1500-1800 мс,
- размер эталона одного голоса составляет 24 кБ.

13. В: Можно ли рабочее место эксперта использовать не для идентификации личности по голосу, а для идентификации звуков животного, технического устройства, автомобиля и т.п.?

О: Рабочее место эксперта предназначено для исследования любых акустических сигналов, в том числе для идентификации звуков животных, механизмов и т.п. Одной из задач, решаемых экспертом, исследующим фонограмму, является задача определения акустических условий, в которых фиксировался разговор между людьми. Здесь могут быть акустические характеристики помещения, если разговор происходил в помещении, характеристики сопутствующих разговору шумов и их источников и т.п.

14. В: Когда впервые была поведена идентификации по голосу и когда человек впервые попытался изменить свой голос?

О: Фольклор доносит до нас детскую сказку «Волк и семеро козлят», где описывается попытка слуховой идентификации. Попытка, как известно, оказалась неудачной, поскольку голос был изменен. Известно также, что в древнем Египте жрецы в храмах использовали специальные акустические приспособления для имитации «голоса богов».

15. В:Чем нужно руководствоваться при выборе типа фильтра?

О.:Если «целевая» АЧХ не принадлежит к стандартному типу частотно-селективных АЧХ, то следует выбрать КИХ-фильтр типа 2 или 3. При этом если фильтрация должна контролироваться в полном частотном диапазоне, т.е. от 0 до частоты Найквиста, то следует выбрать фильтр типа 2. Если же требования к фильтрации формулируются для набора отдельных непересекающихся частотных диапазонов, и допускается определенная степень произвола в промежутках между этими диапазонами, то следует использовать фильтр типа 3.

Если «целевая» АЧХ принадлежит к одному из четырех стандартных типов (низкочастотный, высокочастотный, полосовой, режекторный фильтры), то следует использовать КИХ-фильтр типа 1 либо один из трех БИХ-фильтров (типа 4, 5, 6). При выборе конкретного типа фильтра необходимо учитывать следующие обстоятельства:
- любой из БИХ-фильтров существенно лучше приближает «целевую» АЧХ, чем КИХ-фильтр того же порядка. Однако ФЧХ БИХ-фильтра является нелинейной, что в

принципе может являться источником нежелательных искажений сигнала при его фильтрации;
- фильтр Баттерворта обладает монотонной АЧХ в отличие от фильтров Чебышева. Колебания АЧХ фильтра Чебышева могут в принципе попасть «в резонанс» с пиками амплитудного спектра сигнала, вызывая нежелательные искажения. Указанного резонанса можно избежать путем изменения порядка фильтра, однако, для этого необходимо иметь априорную информацию о положении пиков амплитудного спектра сигнала.
- фильтры Чебышева являются регулируемыми в отличие от фильтра Баттерворта. Ширина переходных областей (в зонах разрыва «целевой» АЧХ) и амплитуда «волнистости» АЧХ могут регулироваться параметром «волнистости». При этом уменьшение амплитуды «волнистости» приводит к увеличению ширины переходных областей и наоборот. Как правило, ширина переходных областей АЧХ фильтра Чебышева (при допустимой амплитуде «волнистости») меньше ширины переходных областей фильтра Баттерворта того же порядка.

16. В.: Какой тип окна выбрать при использовании фильтров 1, 2?

О.:Оконное взвешивание коэффициентов БИХ-фильтра применяется для уменьшения «выбросов» АЧХ в окрестности разрывов «целевой» АЧХ. При этом любое оконное взвешивание (кроме взвешивания прямоугольным окном), уменьшая амплитуду «выбросов», приводит к увеличению ширины переходной области. Баланс этих двух явлений зависит от типа окна. Если «целевая» АЧХ является достаточно гладкой (при использовании фильтра 2), то следует отказаться от оконного взвешивания, что эквивалентно выбору прямоугольного окна. Если же «целевая» АЧХ имеет разрывы или участки резких переходов, то имеет смысл сначала воспользоваться одним из нерегулируемых окон (Ханна, Хемминга, Бартлетта-Ханна, Блэкмана, Блэкмана-Харриса). Если результат оказался неудовлетворительным с точки зрения соотношения «амплитуда выброса – ширина переходной области», то следует попытаться улучшить это соотношение путем использования регулируемых окон Гаусса или Кайзера с настройкой соответствующих параметров этих окон.

17. В.: Какой из двух подходов к оценке спектра мощности шума «лучше»?

О.: Естественно, подход, основанный на использовании сегментов типа «Пауза» позволяет получить более точную оценку спектра мощности шума. Однако этот подход может использоваться только при выполнении следующих двух условий:
- в речевом материале должны физически присутствовать сегменты типа «Пауза», имеющие достаточную длительность (для получения усредненных оценок спектра мощности) и частоту следования (для отслеживания изменения характеристик шума);
- пользователь должен иметь возможность выделения этих сегментов (в «ручном», автоматическом или «смешанном» режиме).

Если хотя бы одно из этих условий нарушено, то следует использовать второй подход, основанный на построении нижних огибающих полосовых спектров мощности зашумленного сигнала.

18. В.: Какой из трех алгоритмов шумоочистки дает лучшие результаты?

О.: Каждый из трех алгоритмов обладает свойствами, которые при одних условиях могут рассматриваться, как его достоинства, а при других – как его недостатки. Алгоритм 3 является наименее «регулируемым» и в то же время стабильно дает весьма высокое качество шумоочистки. В условиях оперативной шумоочистки это является его достоинством. Однако если пользователь захочет улучшить соотношение «разборчивость-качество» в очищенном сигнале путем настройки единственного параметра этого алгоритма, то в большинстве случаев это оказывается невозможным. В отличие от этого алгоритма алгоритм спектрального вычитания предоставляет широкие возможности по

регулировке соотношения «разборчивость-качество» в очищенном сигнале, поскольку предоставляет в распоряжение пользователя три параметра, позволяющие плавно регулировать качество шумоочистки. Однако процесс поиска оптимального сочетания этих параметров может потребовать значительных усилий со стороны пользователя. Алгоритм 2 является промежуточным с точки зрения регулируемости качества шумоочистки.

19. В: Какое дальнейшее развитие комплекса «Шумоочистки речевого сигнала»?

О: Развитие комплекса направлено на разработку дополнительных возможностей. Основные из них - графический редактор шумоочистки речевого сигнала и ряд автоматических алгоритмов, которые позволят осуществлять шумоочистку в реальном режиме времени. Кроме того, будет разработан удобный интерфейс, позволяющий сравнить «очищенный» речевой сигнал с исходным, как в различных отображениях на экране, так и на слух.