то условию адекватности (21) удовлетворяют три модели M_1, M_2, M_3. Поэтому по мере поступления дополнительной технологической информации следует повторить идентификацию модели M_4. В дальнейшем предполагается разработать алгоритмы предварительной обработки информации, позволяющие исключать строки с отсутствующими данными и данными, превышающими допустимые предельные значения.

Литература

1. *Дюбуа, Д.* Теория возможностей. Приложения к представлению знаний в информатике / Д. Дюбуа, А. Прад. – М., 1990.

2. *Заде, Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений. – М., 1976.

3. *Кандель, А.* Нечеткие множества, нечеткая алгебра, нечеткая статистика / А. Кандель, У.Д. Байатт // ТИИЭР. – 1978. – Т. 66. – № 12. – С. 37 – 51.

4. *Кофман, А.* Введение в теорию нечетких множеств / А. Кофман. – М., 1982.

5. *Кудинов, Ю.И.* Моделирование технологических и экологических процессов / Ю.И. Кудинов, А.Г. Венков, А.Ю. Келина. – Липецк, 2001.

6. Модели принятия решений на основе лингвистической переменной / А.Н. Борисов, А.В. Алексеев, О.А. Крумберг и др. – Рига, 1982.

7. *Zadeh, L.A.* Fuzzy – algorithmic approach to the definition of complex and imprecise concepts / L.A. Zadeh // Int. J. Man-Machine Stadies. –1976. – № 6. – P. 249 – 291.

8. *Zadeh, L.A.* Fuzzy sets / L.A. Zadeh // Inform. & Contr. – 1965. – № 8. – P. 338 – 353.

УДК 621.391

Д.Н. Чубатый
Научный руководитель: кандидат технических наук, доцент А.И. Горшков

СПОСОБ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ АБОНЕНТА В СЕТЯХ СВЯЗИ С НИЗКОСКОРОСТНЫМ КОДИРОВАНИЕМ РЕЧИ

В статье предлагается способ идентификации личности абонента в сетях сотовой связи стандарта GSM, использующих для передачи голоса алгоритм низкоскоростного кодирования речи – RPE-LTP (линейное предсказание с возбуждением регулярной последовательностью импульсов и долговременным предсказателем) на основе модели Гауссовых смесей.

Идентификация диктора, модель Гауссовых смесей, низкоскоростные кодеры речи, линейное предсказание речи.

The article suggests the method of identification of the telephone subscriber in cellular networks of GSM standard, using the algorithm of low-speed speech coding - RPE-LTP (linear prediction with the excitement of a regular sequence of pulses and long-term predictor), based on Gaussian mixture model.

Speaker's identification, Gaussian mixture model, low-speed speech coders, linear speech prediction.

В XXI в. мировое сообщество вступило в новую эру своего развития, названную глобальным информационным обществом (ГИО). Для эффективной передачи и распределения всех видов информации в структуре ГИО создана и непрерывно развивается Всемирная сеть связи (World wide communication network), представляющая из себя совокупность всех взаимосвязанных национальных сетей связи на земном шаре. Всемирная сеть связи (ВСС) обеспечивает пользователям широкий набор телекоммуникационных услуг. Основой ВСС являются волоконно-оптические линии связи и системы спутниковой связи, существенно дополняющие возможности наземных сетей.

Новый виток в развитии наземных сетей связи внесла сотовая связь. Система сотовой связи – это сложная и гибкая техническая система, допускающая большое разнообразие как по вариантам конфигурации, так и по набору выполняемых функций [3]. По данным компании *Ericsson* по состоянию на сентябрь 2011 г. общее число абонентов сотовых систем составляет 5,8 млрд, а к 2015 г. оно увеличится приблизительно до 7,8 млрд. Около 75 % всех абонентов, или 4,4 млрд, приходится на стандарт GSM и лишь 14 % составляют пользователи сетей третьего поколения. Сети стандарта GSM охватывают более 85 % населения Земли; более 40 % имеют возможность доступа к сетям WCDMA/HSPA.

На сегодняшний день в телефонных сетях очень широко используются речевые технологии, позволяющие решать все задачи обработки в единстве: распознавание и синтез речи, распознавание личности говорящего и компрессия речи.

Интерес к голосу, как к биометрическому объекту, определяется наличием широкого круга практических приложений, которые могут быть разделены на два обширных класса:

1. Проверка прав доступа к различным системам (информационным и физическим): вычислительные системы, каналы связи, банковские счета, базы данных, и т.д.

2. Криминалистическая экспертиза: анализ записей переговоров при различных аварийных ситуациях, поиск подозреваемого, доказательство в суде [1].

Преимущества установления индивидуальности по голосу при решении подобных прикладных задач

очевидны: голос не отчуждаем от человека и не требует непосредственного контакта с пропускной системой.

В системах идентификации диктора (СИД) в качестве исходного материала используется речь, передаваемая в стандартном телефонном канале в формате в соответствии с рекомендацией G.711 (импульсно-кодовая модуляция, скорость 64 кбит/с), либо в соответствии с рекомендацией G.726 (адаптивная дифференциальная ИКМ, скорость 32 кбит/с). Идентификация проводится на речевом материале, представленном в виде, характеризующим форму речевого сигнала (ИКМ и АДИКМ – кодеры формы). В том случае, если будет использоваться преобразованная низкоскоростными кодерами речь, идентификация диктора с помощью этих систем будет невозможна.

Решение этой задачи требует разработки новых методов и алгоритмов обработки речевых сигналов, преобразованных низкоскоростными кодерами. В данной статье предложен способ идентификации личности абонента в сетях сотовой связи стандарта GSM, использующих для передачи голоса алгоритм низкоскоростного кодирования речи (НКР) – RPE-LTP.

Предложенный способ идентификации состоит из трех основных этапов: выделение различительных признаков, построение эталона модели, принятие решения об индивидуальности говорящего.

1 этап. **Выделение различительных признаков**

Выделение признаков – это один из самых важных этапов при предварительной обработке исходных данных в системах идентификации дикторов. Анализ существующих СИД показал, что большинство из них используют параметры, полученные на основе метода линейного предсказания: коэффициенты логарифмического отношения площадей [6], перцептивные коэффициенты логарифмического отношения площадей [6], мел-частотные кепстральные коэффициенты [5], коэффициенты линейного предсказания, линейные предсказывающие кепстральные коэффициенты [4], линейные спектральные пары, линейные спектральные частоты [4].

Процедура получения вектора различительных признаков из оцифрованного участка речи, представленного в виде, характеризующим форму речевого сигнала сводится к следующему:

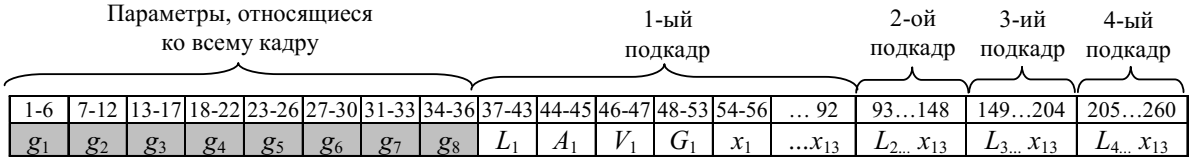
– речевой сигнал анализируется на интервалах длительностью 10 – 30 мс, на таком коротком промежутке речевой сигнал может считаться примерно стационарным;

– для каждого сегмента оцениваются параметры фильтра линейного предсказания (коэффициенты) и параметры сигнала возбуждения [2];

– для расчета необходимого вектора различительных признаков коэффициенты линейного предсказания подвергаются дополнительному преобразованию.

Учитывая тот факт, что в алгоритмах НКР используется аналогичная процедура выделения параметров, также основанная на модели линейного предсказания, для их дальнейшего кодирования и передачи по каналу связи. Структура кадра кодера RPE-LTP представлена на рис. 1 [7].

Для формирования вектора признаков были выбраны параметры $g_1 ... g_8$ – коэффициенты логарифмического отражения площадей. Чтобы выделить значения этих параметров, эталонный речевой сигнал представим в виде прямоугольного массива размером $M \times N$, где N – длина кадра, а M – количество кадров в сообщении (рис. 2).



где $g_1 ... g_8$ – коэффициенты логарифмического отражения площадей (LAR); $L_1 ... L_4$ – задержка; $A_1 ... A_4$ – коэффициент предсказания; $V_1 ... V_4$ – номер последовательности; $G_1 ... G_4$ – максимальная амплитуда; $x_1 ... x_{13}$ – нормированные амплитуды импульсов.

Рис. 1. Структура кадра кодера RPE-LTP (рекомендация GSM 6.10)



Рис. 2. Речевой сигнал в формате GSM 6.10

Формализовано данный речевой сигнал можно представить в виде выражения 1:

$$S_{[M,N]} = \begin{vmatrix} s_{11} & s_{12} & \dots & s_{1N} \\ s_{21} & s_{22} & \dots & s_{2N} \\ \dots & \dots & \dots & \dots \\ s_{M1} & s_{M2} & \dots & s_{MN} \end{vmatrix}, \quad (1)$$

где $s_{mn} \in \{0,1\}$, M – число кадров, N – длина кадра.

А каждый из параметров соответственно:

$$X(g_i)_{[M,L]} = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1L} \\ x_{21} & x_{22} & \dots & x_{2L} \\ \dots & \dots & \dots & \dots \\ x_{M1} & x_{M2} & \dots & x_{ML} \end{vmatrix},$$

$$\text{где } i = 1,\dots,8. \quad (2)$$

где g_i – коэффициент логарифмического отношения площадей, L – количество бит в параметре, а M – число кадров.

Согласно выражению 3, преобразуем значения каждого из параметров в десятичный вид:

$$K(g_i)_{[M,L]} = \begin{vmatrix} x_{11}2^{(L-1)} + x_{12}2^{(L-2)} + \dots + x_{1L}2^0 \\ x_{21}2^{(L-1)} + x_{22}2^{(L-2)} + \dots + x_{2L}2^0 \\ \dots \quad \dots \quad \dots \quad \dots \\ x_{M1}2^{(L-1)} + x_{M2}2^{(L-2)} + \dots + x_{ML}2^0 \end{vmatrix} = \begin{vmatrix} k_{1i} \\ k_{2i} \\ \dots \\ k_{Mi} \end{vmatrix},$$

$$\text{где } i = 1,\dots,8. \quad (3)$$

В итоге, вектор различительных признаков будет представлен в виде матрицы размером $M \times 8$, где 8 – количество коэффициентов логарифмического отношения площадей, а M – количество кадров в сообщении:

$$K(g)_{[M,8]} = \begin{vmatrix} x_{11}(g_1) & x_{12}(g_2) & \dots & x_{18}(g_8) \\ x_{21}(g_1) & x_{22}(g_2) & \dots & x_{28}(g_8) \\ \dots & \dots & \dots & \dots \\ x_{M1}(g_1) & x_{M2}(g_2) & \dots & x_{M8}(g_8) \end{vmatrix}. \quad (4)$$

2 этап. **Построение эталона на основе модели Гауссовых смесей (МГС)**

2.1 Расчет начальной модели

МГС представляет собой взвешенную сумму М компонента и может быть записана выражением [8]:

$$p(\bar{x}|\lambda) = \sum_{i=1}^{M} c_i b_i(\bar{x}), \quad (5)$$

где $\bar{x}$ – это D-мерный вектор случайных величин (в нашем случае это 8-ми мерный вектор коэффициентов логарифмического отношения площадей $K(g)_{[M,8]}$ – выражение 4), $b_i(\bar{x}), i = 1,\dots,M$, – функции плотности распределения составляющих модели, а $c_i, i = 1,\dots,M$, – веса компонентов в модели.

Каждый компонент является D-мерной гауссовой функцией распределения вида выражением [8]:

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2}|K_i|^{1/2}} \exp\left\{-\frac{1}{2}(\bar{x} - \bar{\mu}_i)^T K_i^{-1}(\bar{x} - \bar{\mu}_i)\right\}, \quad (6)$$

где $\bar{\mu}_i$ – вектор математического ожидания и K_i – ковариационная матрица. Веса смеси удовлетворяют выражению:

$$\sum_{i=1}^{M} c_i = 1, \quad (7)$$

Полностью модель гауссовой смеси определяется векторами математического ожидания, ковариационными матрицами и весами смесей для каждого компонента модели. Эти параметры все вместе записываются в виде выражения:

$$\lambda = \{c_i, \bar{\mu}_i, \Sigma_i\}, \quad (8)$$

В задаче распознавания голоса каждый диктор представляется моделью гауссовых смесей и ставится в соответствие со своей моделью λ.

Одна из сильных сторон смеси гауссовых моделей заключается в том, что если они могут очень точно аппроксимировать произвольные распределения.

2.2 Алгоритм оценки максимизации (обучения) параметров МГС

Цель алгоритма оценки параметров модели – при заданном обучающем высказывании диктора оценить параметры модели λ, которые наилучшим образом соответствуют распределению векторов признаков обучающего высказывания.

Существует несколько способов оценки параметров модели, но наиболее популярным и широко используемым является метод оценки максимального правдоподобия.

Цель оценки максимального правдоподобия – найти параметры модели, которые максимизируют правдоподобие ее. Для последовательности обучающих векторов $X = \{\bar{x}_1,\dots,\bar{x}_T\}$ правдоподобие модели гауссовых смесей может быть записано в виде:

$$p(X|\lambda) = \prod_{t=1}^{T} p(\bar{x}_t|\lambda), \quad (9)$$

Но это выражение представляет нелинейную функцию от параметров λ, и ее вычисление невозможно. Поэтому оценки параметров могут быть получены итерационно при помощи алгоритма оценки-максимизации [8].

Алгоритм оценки-максимизации начинается с оценки начальной модели λ. На основе вектора признаков получена начальная модель диктора λ, представляющая собой совокупность векторов математического ожидания, главной диагонали ковариационной матрицы и весами смесей для каждого компонента модели.

Затем вычисляются новые параметры модели $\bar{\lambda}$, такие, что $p(X|\bar{\lambda}) \geq p(X|\lambda)$.

Новая модель затем становится начальной моделью для следующей итерации, и процесс переоценки параметров повторяется, пока не будет достигнут некоторый порог сходимости. Этот способ используется для оценки параметров скрытых Марковских моделей при помощи алгоритма Баума-Велша [8].

В конечном виде модель эталонного высказывания диктора представляется в виде:

$$\lambda_{эт} = \left\{ \begin{vmatrix} c_1 \\ c_2 \\ \dots \\ c_M \end{vmatrix}, \begin{vmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2N} \\ \dots & \dots & \dots & \dots \\ \mu_{M1} & \mu_{M2} & \dots & \mu_{MN} \end{vmatrix}, \begin{vmatrix} D_{11} & D_{12} & \dots & D_{1N} \\ D_{21} & D_{22} & \dots & D_{2N} \\ \dots & \dots & \dots & \dots \\ D_{M1} & D_{M2} & \dots & D_{MN} \end{vmatrix} \right\}, \quad (10)$$

3 этап. **Принятие решения об индивидуальности говорящего**

Апостериорная вероятность *i*-го акустического класса задается выражением [8]:

$$p(i|\bar{x}_t, \lambda) = \frac{c_i b_i(\bar{x}_t)}{\sum_{k=1}^{M} c_k b_k(\bar{x}_t)}. \quad (11)$$

Для идентификации по голосу группа из *S* дикторов представляется набором моделей гауссовых смесей $\lambda_1, \lambda_2, ..., \lambda_S$. Задача идентификации – найти модель диктора, которая имеет наибольшее значение апостериорной вероятности для заданного высказывания:

$$S = \underset{1 \leq k \leq S}{\operatorname{argmax}} P_r(\lambda_k | X) = \underset{1 \leq k \leq S}{\operatorname{argmax}} \frac{p(X|\lambda_k) P_r(\lambda_k)}{p(X)}. \quad (12)$$

Предполагая, что все дикторы одинаково вероятны (т.е. $P_r(\lambda_k) = 1/S$), а величина $p(X)$ одинакова для моделей всех дикторов, правило классификации диктора упрощается до вида:

$$S = \underset{1 \leq k \leq S}{\operatorname{argmax}} P_r(X|\lambda). \quad (13)$$

Используем логарифм и независимость между наблюдениями:

$$S = \underset{1 \leq k \leq S}{\arg\max} \sum_{t=1}^{T} \log p(\bar{x}_t|\lambda_k), \quad (14)$$

где $p(\bar{x}_t|\lambda)$ определена в (5).

Экспериментальная оценка разработанного способа идентификации производилась на базе высказываний, смоделированных на ПЭВМ с помощью стандартной программы «Звукозапись» в формате ИКМ с частотой дискретизации 22 кГц и разрядностью АЦП 16 бит. Затем с помощью этой же программы производилось кодирование речи в формат GSM 6.10. Продолжительность обучающего высказывания составляла примерно 30 с. Экспериментальная выборка из базы голосов состояла из трех мужчин и трех женщин. Для каждого из 6 дикторов было записано по 5 речевых отрезков по 30 с.

Для определения эффективности предложенного способа идентификации личности абонента была выбрана модель с 12 компонентами, а в качестве вектора признаков для идентификации дикторов использовался восьмимерный вектор коэффициентов логарифмического отношения площадей. Первое высказывание каждого из дикторов было использовано для получения эталона на основе МГС, обучение которой проводилось с использованием одинакового числа итераций. Полученные эталонные модели дикторов хранятся в базе данных в виде отдельного файла. В результате проведенного эксперимента были правильно идентифицированы все тестируемые 24 высказывания 6 дикторов.

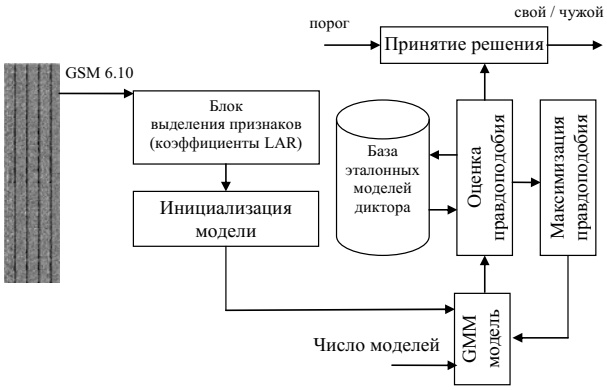


Рис. 3. Система идентификации личности абонента

На рис. 3 показаны основные компоненты СИД на основе модели Гауссовых смесей, реализующей предлагаемый способ идентификации личности абонента.

Применение разработанного способа идентификации личности абонента позволит:

1. Использовать параметры, передаваемые внутри кадра НКР для идентификации диктора.

2. Сократить количество этапов обработки; идентификация может быть произведена без предварительного декодирования сигнала.

Экспериментальная оценка способа идентификации личности абонента в сетях связи с низкоскоростным кодированием речи на основе модели Гауссовых смесей, представленная в статье, показала высокие значения вероятности правильной идентификации диктора. Поэтому дальнейшие исследования в этой области могут быть направлены на его реализацию в практических приложениях.

Литература

1. *Галунов, В.И.* Верификация и идентификация говорящего / В.И. Галунов // Доклады Санкт-Петербургского Государственного университета. – СПб., 2002. – С. 26 – 34.

2. *Маркел, Дж. Д.* Линейное предсказание речи / Дж.Д. Маркел, А.Х. Грэй. – М., 1980.

3. *Ратынский, М.В.* Основы сотовой связи / М.В. Ратынский. – М., 1998.

4. *Садыхов Р.Х.* Модели гауссовых смесей для верификации диктора по произвольной речи / Р.Х. Садыхов, В.В. Ракуш // Доклады Белорусского государственного университета информатики и радиоэлектроники. – Минск, 2003. – С. 95 – 103.

5. Campbell, J.P. Speaker recognition / J.P. Campbell // Proceedings of the IEEE. – 1997. – V. 85. – № 9 (September). – P. 1437 – 1462.

6. *Chow, D.* Robust speaker identification based on perceptual log area ratio and Gaussian mixture models / D. Chow, W.H. Abdulla // INTERSPEECH 2004, 8th International Conference on Spoken Language Processing, Jeju Island, Korea. – 4 – 8 October. – 2004. – P. 1761 – 1764.

7. ETSI/TC SMG. Recommendation GSM 6.10. GSM full rate speech transcoding. – 1992. – February.

8. *Reynolds, D.A.* Robust text-independent speaker identification using Gaussian mixture speaker models / D.A. Reynolds, R.C. Rose // IEEE transaction on speech and audio progressing. – 1995. – V. 3. – № 1 (January) – P. 72 – 83.