

## СИСТЕМА ВЕРИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ НА ОСНОВЕ АНАЛИЗА ЗВУКОВОЙ ИНФОРМАЦИИ

А. С. Кочетков

Построение системы верификации диктора является непростой задачей, не решенной окончательно вплоть до сегодняшнего дня. Основная причина заключается в сложности качественного извлечения из речевого сигнала характерных параметров, описывающих особенности произношения конкретного человека. Результат должен быть как можно менее подвержен воздействию различных внешних факторов или учитывать их влияние на оцениваемые параметры.

Не менее тяжелой для решения проблемой является создание так называемого решающего правила, позволяющего с заданной вероятностью ошибки дать ответ на вопрос о верификации диктора. Это обусловлено значительными вариациями речи каждого человека в зависимости от его настроения, физической усталости, состояния здоровья, степени сосредоточенности или, наоборот, рассеянности.

Выбор характерных параметров определяется видом модели речеобразования, принятой к рассмотрению при построении системы. В работе была использована модель линейной квазистационарной системы, в которой учитываются особенности образования вокализованных звуков [1].

При формировании вокализованной последовательности входной периодический сигнал  $p[n]$  с заданной частотой основного тона усиливается и поступает в линейную систему с постоянными параметрами, состоящую из трех частей (рис. 1).

Первая часть учитывает эффект формы источника возбуждения, вторая представляет собой систему связанных резонаторов, третья учитывает потери в речевом тракте, связанные с излучательными процессами. В качестве параметров, по которым можно идентифицировать диктора, были выбраны траектории формантных частот  $F_i$ , представляющих собой резонансные частоты системы связанных резонаторов.

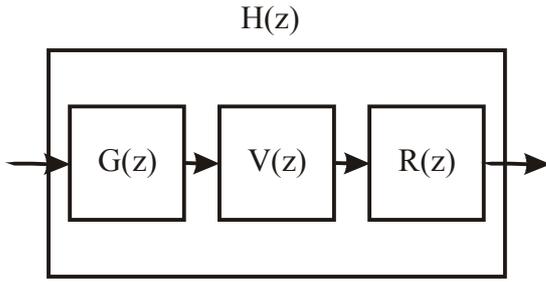


Рис. 1. Линейная квазистационарная система, описывающая речевой тракт

на составляющие, описывающие каждую часть представленной модели. Таким образом, кепстр входной последовательности  $x[n]$  можно представить следующим образом:

$$\hat{x}[n] \approx \begin{cases} \hat{gr}[n] + \sum_{i=1}^3 \hat{h}_i[n] & nT_n < T \\ \hat{p}_w[n] & nT_n \geq T \end{cases}, \quad (1)$$

где  $gr[n]$  – импульсная характеристика компонент системы, отвечающих за эффект формы источника и излучательные процессы,  $h_i[n]$  – импульсная характеристика  $i$ -того резонатора,  $p_w[n]$  – периодическая импульсная последовательность на входе системы,  $T_n$  – период дискретизации сигнала,  $T$  – период основного тона.

В формуле (1) представлены кепстры, соответствующие  $gr[n]$ ,  $h_i[n]$ ,  $p_w[n]$ .

Пропустив результат (1) через фильтр нижних частот, мы удалим из кепстра сведения о  $p_w$ , а; вычитая кепстральное представление  $gr$ , получим последовательность, содержащую информацию только о системе связанных резонаторов.

Следующий шаг – выделение максимумов результирующей характеристики, которым отвечают формантные частоты  $F_i$ . При выполнении этой процедуры мы воспользовались представленными в таблице результатами исследований в области формантного анализа речевых сигналов, согласно которым области локализации каждой из трех формант различаются для мужских и для женских (детских) голосов [2].

В случае, когда необходимые максимумы являются неразличимыми на фоне других, применяется процедура острого  $z$ -преобразования (chirp  $z$ -transform). При этом вычисления производятся по контуру, лежащему около полюсов, что позволяет существенно повысить разрешение получаемой спектральной картины.

Граничные значения формантных частот

Формантная частота	Голос	Диапазон частот, Гц ( $F_{\min}$ - $F_{\max}$ )
$F_1$	Мужской	200-900
	Женский или детский	300-1050
$F_2$	Мужской	550-2700
	Женский или детский	700-2900
$F_2$	Мужской	1100-3100
	Женский или детский	1400-3500

Заключительным шагом при реализации системы верификации является формирование процедуры принятия решения. Предлагается в качестве таковой использовать вычисление некоторой полной меры различимости и сравнение ее с выбранным соответствующим образом порогом. Например, можно применить взвешенное евклидово расстояние от траектории  $v=[v_1, \dots, v_k]$  до вычисленного ранее биометрического эталона  $m(v)=[m(v_1), \dots, m(v_k)]$ . [3, 4]

$$d_i = \sqrt{\sum_{j=1}^k (v_j - m(v_j))^2} . \quad (2)$$

На этапе обучения системы вычисляются эталонные траектории всех трех формант для каждой речевой последовательности. Кроме того, для каждой входной последовательности рассчитывается также вектор вариаций, на основе которого в дальнейшем определяются вероятности верифицировать по данной последовательности ложного диктора  $P_M$  и отклонить настоящего  $P_{FA}$ . Сравнивая вычисленные  $d_j$  с соответствующими значениями вектора вариаций, делают предварительный вывод о совпадении или несовпадении входной последовательности с эталонной.

Следующий шаг – вычисление весовых коэффициентов [3]:

$$a_i = \log \frac{1 - P_{Mi}}{P_{FAi}} \quad (3)$$

для положительного результата предварительного сравнения,

$$a_i = \log \frac{P_{Mi}}{1 - P_{FAi}} \quad (4)$$

для отрицательного.

Далее для сведения многокритериальной задачи выбора к однокритериальной производится вычисление полной меры различимости, представляющей собой взвешенную сумму элементов:

$$D = \sum_i a_i d_i. \quad (5)$$

Если результат суммирования окажется меньше нуля, то запрос на верификацию отклоняется, если больше – подтверждается.

### **Заключение**

Для создания на основе описанной модели действующей системы верификации диктора, работающей в режиме реального времени, представляется необходимым применение комплексных программно-аппаратных решений. При этом на аппаратную часть должна лечь нагрузка по непосредственной обработке входного сигнала, вычислению спектра и кепстра. Программно можно реализовать процедуру непосредственного выделения формант, решающее правило, а также пользовательский интерфейс.

### **Литература**

1. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981.
2. William D. Oliver. The Singing Tree: A Novel Interactive Musical Interface. <http://feynman.stanford.edu>.
3. Bossemeyer R.W., Ratasuk R. Speaker Verification Method: United States Patent Num. 6,070,55A.
4. Евклидова мера близости к центру биометрического эталона. [http://beda.stup.-ac.ru/biometry/Bio\\_Mon/GL-2](http://beda.stup.-ac.ru/biometry/Bio_Mon/GL-2).