

**Построение оценок достоверности результатов  
распознавания речи с использованием альтернативных  
моделей**

*Нгуен Минь Туан*

nmtuan@yahoo.com

Москва, ВЦ РАН

Один из основных подходов к выявлению ошибок при распознавании состоит в вычислении для каждого распознанного слова  $w$  числовой характеристики  $Cm(w)$  — меры достоверности. Величина  $Cm(w)$  сравнивается с некоторым порогом  $\tau_w$ . Если значение больше, то слово считается правильно распознанным. В противном случае соответствующая часть сигнала считается шумом или незнакомым словом. Существует много различных подходов к определению и оценке меры достоверности [1, 2]. В данной статье предлагается метод её вычисления, основанный на оценке достоверности составляющих слова акустических векторов.

**Оценка достоверности распознавания слова**

Пусть имеется последовательность акустических векторов  $Y = (y_1, \dots, y_N)$  сегмента сигнала, распознанного декодером как слово  $w$ . Тогда на выходе из декодера каждый акустический вектор  $y$  последовательности  $Y$  соответствует некоторому состоянию  $q$  марковской модели слова  $w$ . Определим оценку достоверности распознавания акустического вектора  $y$  как:

$$P_q(y) = \frac{P(y|\Theta_q)}{P(y|\Theta_q) + P(y|\bar{\Theta}_q)},$$

где  $P(y|\Theta_q)$  — вероятность принадлежности акустического вектора  $y$  классу акустических векторов состояния  $q$ ;  $P(y|\bar{\Theta}_q)$  — вероятность того, что  $y$  не принадлежит классу векторов состояния  $q$ .

Считаем, что вероятность  $P(y|\Theta)$  имеет вид

$$P(y|\Theta) = \left( \sum_{i=1}^M c_i N(m_i, v_i, y) \right)^\alpha,$$

где  $M$  — количество смесей;  $c_i$  — вес  $i$ -ой смеси,  $\sum_{k=1}^M c_k = 1$ ,  $c_k > 0$ ;  $N(m_i, v_i, y)$  — нормальное распределение с вектором средних  $m_i$  и диагональной матрицей ковариации  $v_i$ ;  $\alpha$  — числовой коэффициент,  $0 < \alpha \leq 1$ .

Для любого распознанного слова  $w$  с соответствующими ему последовательностями акустических векторов  $Y = (y_1, \dots, y_N)$  и состояний скрытых марковских моделей  $Q = (q_1, \dots, q_N)$  определим оценку достоверности корректного распознавания как

$$Cm(w) = \exp\left(\frac{1}{N} \sum_{i=1}^N \log P_{q_i}(y_i)\right). \quad (1)$$

Чтобы пользоваться формулой (1) для вычисления оценки достоверности слов, требуется определить значения параметров моделей  $\Theta_q$  и  $\bar{\Theta}_q$  состояний  $q$  скрытых марковских моделей звуков. Эмпирический метод оценки этих параметров описывается в следующем разделе.

### Оценка параметров моделей $\Theta_q$ и $\bar{\Theta}_q$

Предлагаемый метод оценки параметров моделей  $\Theta_q$  и  $\bar{\Theta}_q$  состояния  $q$  состоит в нахождении параметров этих моделей таким образом, чтобы увеличить значение  $P_q(y)$  для всех «корректных» акустических векторов и уменьшить значение  $P_q(y)$  в противном случае на выбранном наборе акустических векторов. Акустический вектор  $y$  будем считать «корректным», если последовательность акустических векторов  $Y = (y_1, \dots, y_N)$ , содержащая  $y$ , корректно распознана декодером.

Введем функцию стоимости акустического вектора  $y$ :

$$F(y, \Theta_q, \bar{\Theta}_q) = \begin{cases} 1 - P_q(y), & y \text{ — корректный;} \\ P_q(y), & y \text{ — некорректный.} \end{cases}$$

Тогда задача оценки параметров моделей  $\Theta_q$  и  $\bar{\Theta}_q$  сводится к решению задачи оптимизации:

$$\arg \min_{\Theta_q, \bar{\Theta}_q} \sum_y F(y, \Theta_q, \bar{\Theta}_q). \quad (2)$$

Задачу (2) предлагается решать методом градиентного спуска.

### Численные эксперименты

Численные эксперименты выполнялись на речевом корпусе FaVoR [3].

На Рис. 1 показана параметрическая кривая ошибок I и II рода, построенная для слова «два». Для всей тестовой выборки (при выборе оптимального значения порога для каждого слова из словаря системы) удалось корректно определить 92.8% неправильно распознанных слов. При этом ошибка II рода (отказ от правильно распознанного слова) составила 4.7%.

### Заключение

В докладе представлен метод выявления ошибок результатов распознавания речи, который основан на оценке достоверности для наблюдаемых акустических параметров. Его применение на представительном

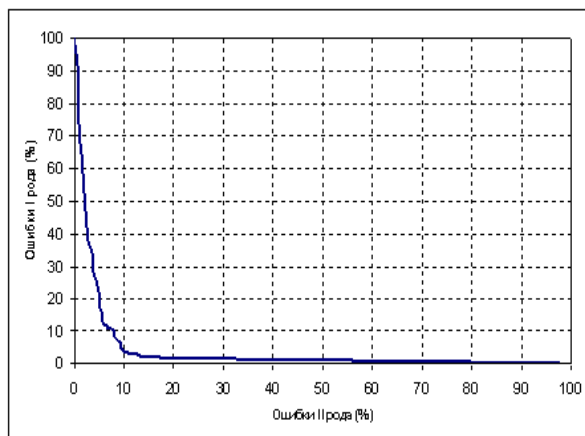


Рис. 1. Параметрическая кривая ошибок I и II рода.

корпусе речевых данных, в котором имеется большое количество различных акустических шумов, как речевой, так и неречевой природы, показало значительное сокращение количества ошибок II рода при умеренном количестве ошибок I рода.

#### Литература

- [1] *Thomas Kemp, Thomas Schaaf*. Confidence measures for spontaneous speech recognition. // Proc. of International Conference on Acoustic, Speech and Signal Processing, 1997. — Pp. 875–878.
- [2] *Manhung Siu, Herbert Gish*. Evaluation of word confidence for speech recognition systems // Computer Speech and Language, 1999. Pp. 299–319.
- [3] *Десятчиков А. А и др.* Комплекс алгоритмов для устойчивого распознавания человека. — Известия РАН. Теория и системы управления. — 2006.