

Обнаружение незнакомых слов при распознавании речи

Чучупал В. Я., Маковкин К. А., Чичагов А. В.

Москва, ВЦ РАН

chuchu@ccas.ru

Речевой сигнал, подлежащий распознаванию, как правило, содержит шумы, речь посторонних лиц, нарушения речевого потока или слова, которые не входят в словарь системы распознавания речи.

Целью настоящего исследования являлась разработка и исследование эффективности нового алгоритма обнаружения незнакомых слов на основе оценок правдоподобия для наблюдаемого речевого сигнала при заданном множестве акустико-фонетических моделей.

Акустические счета

Основные методы выявления незнакомых слов основаны на использовании величин локальных мер сходства, оценок правдоподобия или акустических счетов. Наиболее часто используются: средний счет

$$S(w) = \frac{1}{N_w} \sum_{j=1}^{N_w} s_j, \quad (1)$$

где N_w – число наблюдений, s_j – счет наблюдения j
и центрированный средний счет:

$$\tilde{S}(w) = \frac{1}{N_w} \sum_{j=1}^{N_w} (s_j - \tilde{s}) \quad (2)$$

где $\tilde{s}$ – среднее счета для длительного промежутка времени.

Обе оценки широко используются, однако несколько лучшие результаты были получены при использовании дважды нормированного счета:

$$S_1(w) = \frac{1}{N_w} \sum_{s=1}^{N_w} \frac{1}{N_s} \sum_{j=1}^{N_s} s_j \quad (3)$$

где N_w – число состояний модели слова w , N_s – длина состояния s .

По аналогии с (2) определим дважды нормированный центрированный счет:

$$\tilde{S}_1(w) = \frac{1}{N_w} \sum_{j=1}^{N_w} \frac{1}{N_s} (s_j - \tilde{s}), \quad (4)$$

На рис.1 приведены DET кривые [1] для счетов (1)–(4), вычисленные на речевом корпусе данных FaVoR [2]. Очевидно, что счет (4) – самый эффективный.

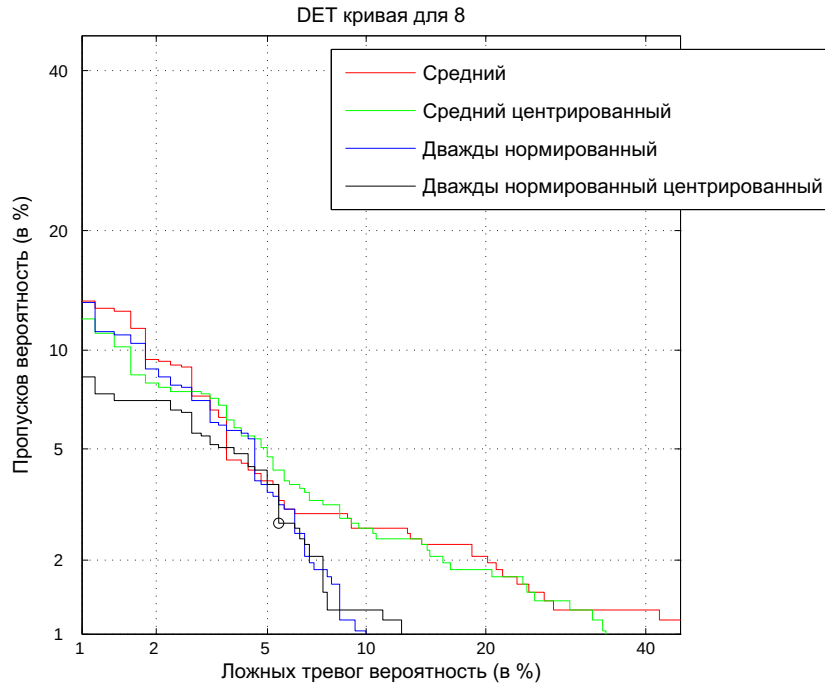


Рис. 1: DET характеристики акустических счетов (1)–(4)

Оценка счетов по параметрам моделей звуков

Существенным недостатком рассмотренных акустических счетов является то, что оценка порогового значения требует наличия настроечной выборки, т.е. какого-то количества произнесений слова в заданной акустико-фоновой обстановке.

Рассмотрим метод формирования оценки счета на основе параметров акустических моделей звуков.

Для счета (4) определим среднее, отклонение и порог принятия решения:

$$\tilde{m}(w) = \frac{1}{N_w} \sum_{s=1}^{N_w} \frac{1}{N_s} \sum_{j=1}^{N_s} m_j, \quad (5)$$

$$\sigma(w) = \sqrt{\frac{1}{N_w} \sum_{s=1}^{N_w} \frac{1}{N_s} \sum_{j=1}^{N_s} \sigma_j^2}, \quad (6)$$

$$th(w) = s(w) - m(w) - \sigma(w). \quad (7)$$

Оцениваем текущее значение счета как:

$$\tilde{m}(w, t+1) = \tilde{m}(w, t) + \alpha * (\tilde{m}(w, t) - S(w, t)), \quad (8)$$

где $S(w, t)$ – наблюдаемое значение счета, α – параметр.

На рис. 2 приведены DET – кривые для оценок (5)-(7).

Заключение

Предложен метод оценки априорных дисперсий счетов, выбора априорного порога, а также процедура адаптации средних значений и порогов в соответствии с наблюдаемым речевым сигналом. Показано, что эффективность предложенных априорных оценок счетов и метода адаптации порога практически соответствует использованию апостериорных оценок пороговых счетов на настроечной выборке. Работа выполнена при финансовой поддержке РФФИ, проект 07-01-00657а

Литература

- [1] Программа DETware. Национальный институт стандартов США, NIST, <http://www.nist.gov>.
- [2] Десятчиков А.А. и др. Комплекс алгоритмов для устойчивого распознавания человека, Известия РАН. Теория и системы управления, 2006

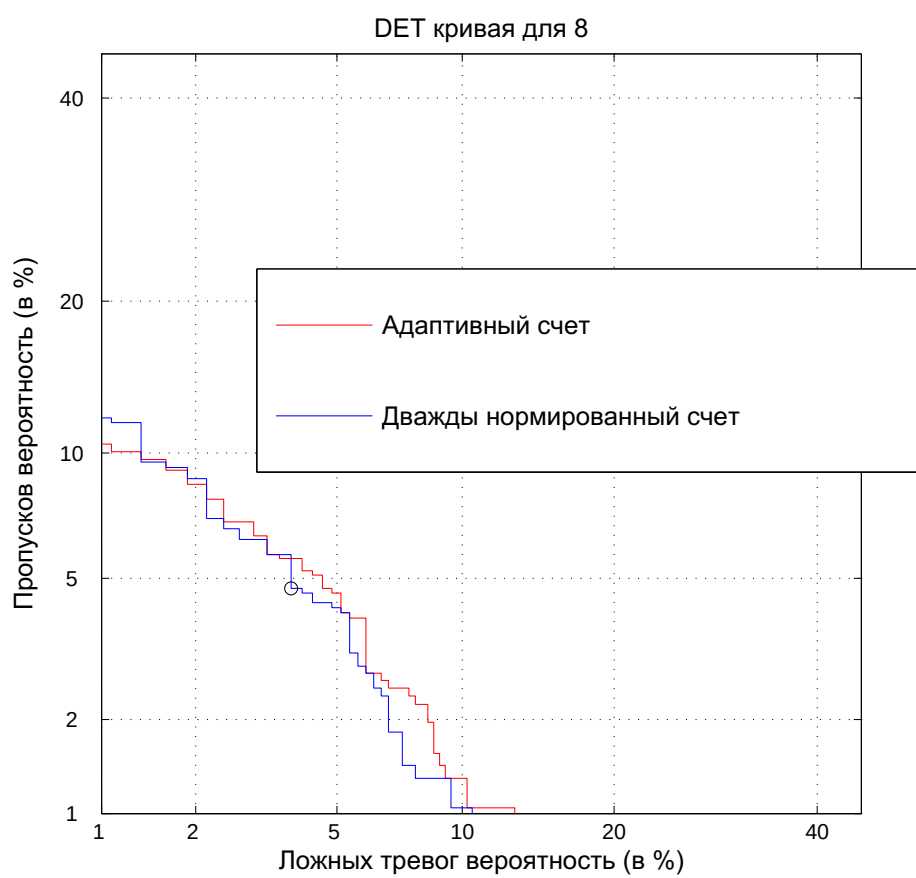


Рис. 2: DET кривая счета (5)-(7)