

НЕЧЕТКОЕ СОПОСТАВЛЕНИЕ ОБРАЗОВ НА ОСНОВЕ ВЕЙВЛЕТ-ПРЕДСТАВЛЕНИЯ РЕЧЕВЫХ СИГНАЛОВ

*Нестеренко Д.С., Федяев О.И., Бондаренко И.Ю.
Донецкий национальный технический университет*

В современных компьютерных системах все больше внимания уделяется построению интерфейса естественного ввода-вывода информации. Одним из перспективных направлений на сегодняшний день является использование систем речевого диалога, которая предполагает автоматический синтез и распознавание речи.

В настоящее время вейвлеты начинают широко применяться при решении различных прикладных задач: распознавание образов, при обработке и синтезе различных сигналов (например, речевых), при анализе изображений и во многих других случаях [1]. За относительно недолгое время существования теории вейвлетов сформировался на её основе развитый математический аппарат. Однако специалисты в этой области ещё не успели в полной мере отразить все преимущества вейвлет-преобразований на практике.

Главными направлениями вейвлет-преобразований являются непрерывный и дискретный вейвлет-анализ. В используемом методе на данном этапе разработки применяются только непрерывное вейвлет-преобразование.

Человек произносит слова, изменяя органом речи резонансную частоту, поэтому особенно важными являются резонансные частоты. Резонансные частоты для гласных звуков называют формантами, однако используют и название «локальный выброс» как расширение понятия форманта на согласные звуки. В рассматриваемом методе распознавание произносимого слова осуществляется путем определения места локального выброса и как он меняется во времени.

В созданной системе распознавания изолированных слов исходный речевой сигнал проходит процесс прямого непрерывного вейвлет-преобразования для получения вейвлет-коэффициентов в диапазоне масштабов от 1 до 100. В дальнейшем использовались только неотрицательные области значений вейвлет-коэффициентов, т.к. они в достаточной мере отражают структуру резонансных частот, определяемых по локальным выбросам энергии, с течением времени. Следующим этапом обработки вейвлет-спектрограммы является её локальное нормирование, а также приведение значений коэффициентов к диапазону от нуля до единицы.

Локальное нормирование позволяет выявить характерные особенности речевого сигнала и увеличить их значение. Данный этап обработки выполнялся путём деления значения в каждой точке вейвлет-спектрограммы на среднее значение коэффициентов в некоторой её окрестности. Такой способ нормирования позволяет избавиться от случайных резких «всплесков» и «провалов» значений коэффициентов.

На заключительном этапе применялся модифицированный метод нечёткого сопоставления образов [2].

Обозначим число исходных слов через n , множество слов через $I = \{i_1, i_2, i_3, \dots, i_n\}$, а множество образов этих слов через $X = \{x_1, x_2, x_3, \dots, x_j\}$.

Множество I – это обычное множество из n элементов, а множество X можно рассматривать как нечёткое множество, в котором x_j ($j = 1, 2, 3, \dots, n$) представляет различные образы слова i_j .

Таким образом, можно определить множество функций принадлежности $M = \{m_1, m_2, m_3, \dots, m_n\}$ подобно тому, как определяется множество образов x_{ij} слова i_j .

Рассматриваемое здесь нечёткое сопоставление образов заключается в следующем. При вводе неизвестного образа y ($y \in X$) с использованием функции принадлежности M вычисляется степень сходства S_j образов x_j и y , и результатом распознавания является слово j , такое что

$$j = \max_{j \in I} \{S_j\}$$

Для определения степени подобия используется следующее выражение:

$$S_j = \frac{P_{j0}}{P_{jd} - P_{j0}} \cdot \frac{P'_{j0}}{P_y - P'_{j0}},$$

где

$$\begin{aligned} P_{jd} &= \sum_k \sum_t m_j(k, t), \\ P_{j0} &= \sum_k \sum_t m_j(k, t) \cdot y(k, t), \\ P'_{j0} &= \sum_k \sum_t m_j(k, t) * y(k, t), \\ P_y &= \sum_f \sum_t y'(f, t), \\ y'(f, t) &= \begin{cases} 1, & y(f, t) \geq \alpha, \\ 0, & \text{иначе.} \end{cases} \end{aligned}$$

Знак « \cdot » обозначает произведение элементов m_j и y , а знак « $*$ » – логическое произведение m_j и y уровня α , то есть

$$m_j(k, t) * y(k, t) = \begin{cases} 1, & \text{если } m_j \geq \alpha, y \geq \alpha, \\ 0, & \text{иначе} \end{cases}$$

Формула для вычисления степени подобия состоит из двух множителей. Числитель и знаменатель первого множителя – это, соответственно, $y \cap m_j$ и $\bar{y} \cap \bar{m}_j$, второго - $y * m_j$ и $\bar{y} * \bar{m}_j$.

Результаты исследований свидетельствуют о том, что представление речевых сигналов при помощи вейвлет-коэффициентов, позволяет эффективно их использовать при распознавании изолированных слов. При верно распознаваемых образах, максимальная степень сходства, в большинстве случаев, отличается от остальных на несколько порядков.

Проведенные эксперименты носили предварительный характер. В настоящее время проводятся исследования по оптимизации характерных признаков речи, полученных на основе вейвлет-преобразования.

Литература

- [1] Н.М. Астафьева. Вейвлет-анализ: Основы теории и примеры применения. – Успехи физических наук, 1996, т.166, № 11, стр. 1145-1170.
- [2] Киедзи Асаи, Дзюндзо Ватада, Сокуке Иваи и др. “Распознавание речи”. – М.: «Мир», – 1993;