

МЕТОД НЕЧЁТКОГО СОПОСТАВЛЕНИЯ ОБРАЗОВ ДЛЯ РАСПОЗНАВАНИЯ РУССКОЯЗЫЧНЫХ КОМАНД УПРАВЛЕНИЯ ТЕКСТОВЫМ РЕДАКТОРОМ

Бондаренко И.Ю., Федяев О.И.
Донецкий национальный технический университет

Речевой интерфейс, как более естественный для человека, приобретает всё большую востребованность в современных человеко-машинных системах. В основе его построения лежит задача распознавания речи, для решения которой, несмотря на множество предложенных способов, не найден приемлемый метод. Намечились основные направления, базирующиеся на вероятностном, метрическом и нейросетевом подходах. Также перспективен для решения трудноформализуемых задач, к которым относится задача распознавания речи, подход на основе нечёткой логики [1]. В работе [2] описан метод нечёткого сопоставления образов и приведена высокая оценка его эффективности в распознавании английских, немецких и японских слов. Однако, вопросы применения данного метода к распознаванию русскоязычных слов не рассмотрены.

Данная работа посвящена анализу эффективности метода нечёткого сопоставления в задаче распознавания изолированных русских слов, возникающей при интеллектуализации интерфейса текстового процессора Microsoft Word. Рациональное сочетание речевого и стандартного визуального способов управления процессом ввода и редактирования текстовой информации позволит снизить нагрузку на тактильно-зрительные каналы человека и тем самым повысить эффективность его работы.

Речевой сигнал представляется в виде двумерного спектрального временного образа (СВО), получаемого с помощью оконного преобразования Фурье (рис. 1а). Такой образ отражает изменение по времени амплитуд заданных частотных составляющих речевого сигнала и хорошо выражает особенности речи, что даёт возможность его использовать для автоматического распознавания произносимых слов [2]. СВО позволяет выделить местоположение резонансных частот, т.е. локальных выбросов, что является определяющей особенностью речевого сигнала [2]. На этом основании СВО можно преобразовать к двоичному виду, не теряя указанных информативных признаков речи, с помощью следующей замены: 1 – на месте локального выброса, 0 – в других местах. Полученный образ называют двоичным спектральным временным образом (ДСВО) и используют его как отражение особенностей речевого сигнала (рис. 1б).

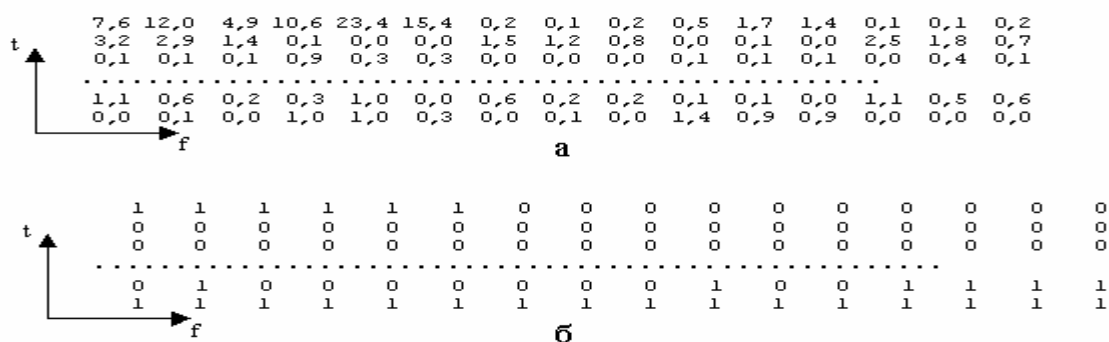


Рисунок 1 – Пример спектрально-временного представления слова «пять»:
а – СВО; б – ДСВО

В качестве единиц речи рассматриваются слова, набор которых определяет словарный состав речевого командного интерфейса с редактором Word.

Для распознавания изолированных слов, произносимых в реальном времени, применялся метод нечёткого сопоставления с эталоном [2]. Эталонные образы для каждого слова словаря формировались как среднее арифметическое ДСВО различных вариантов произношения данного слова и затем подвергались нормализации. В результате формируется бинарное нечёткое отношение между множеством F (номеров частот f) и множеством T (номеров временных интервалов t) в виде:

$$f \in F, t \in T: FR T ,$$

где R – нечёткое отношение, которое ставит каждой паре элементов $(f, t) \in F * T$ величину функции принадлежности $\mu_R(x, y) \in [0, 1]$ [1].

Обозначим число записанных слов через n, множество слов через $I = \{i_1, i_2, \dots, i_n\}$ и множество нечётких отношений, характерных для каждого слова, через $R = \{r_1, r_2, \dots, r_n\}$. Входной неизвестный образ y рассматривается как обычное (чёткое) отношение между множеством частот и множеством временных интервалов. Для него вычисляются степени сходства S_j с каждым нечётким отношением r_j . Результатом распознавания является слово j, такое, что

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

Степень подобия вычисляется по формуле

$$S_j = D_j / \overline{D}_j ,$$

где

$$D_j = \int r(f, t) \wedge y(f, t) df dt ,$$

$$\overline{D}_j = \int \overline{r(f, t)} \wedge y(f, t) df dt .$$

В результате исследований установлено, что метод нечёткого сопоставления точно распознаёт даже сильно зашумлённые образы, если они не деформированы. Однако, при распознавании речевых образов русского языка, которые характеризуются сильной изменчивостью по структуре расположения локальных выбросов, точность составила 85%. Таким образом, возможности для улучшения данной системы распознавания заключаются в разработке такого метода предварительной обработки речевого сигнала, который более устойчив к случайным искажениям СВО.

Литература

- [1] Кофман А. Введение в теорию нечетких множеств. - М.: Радио и связь, - 1982.- 432с
- [2] Киедзи Асаи, Дзюндзо Ватада, Сокуке Иваи и др. Распознавание речи // Прикладные нечеткие системы. Под редакцией Т.Тэрано, К. Асаи, М. Сугено. – М.:«Мир», – 1993. – 157-170с.