

### 3. Распознавание образов

Рассмотрим простейший случай, когда речь представляет собой последовательность ограниченного количества слов. В этом случае распознавание речи сводится к классификации двумерных векторов свойств. Каждый класс соответствует определённому слову, произнесенному различными дикторами. Конкретные двумерные векторы свойств различных слов представлены на рис. 3, где в каждом столбце представлены слова одного диктора, а в каждой строке – классы слов.

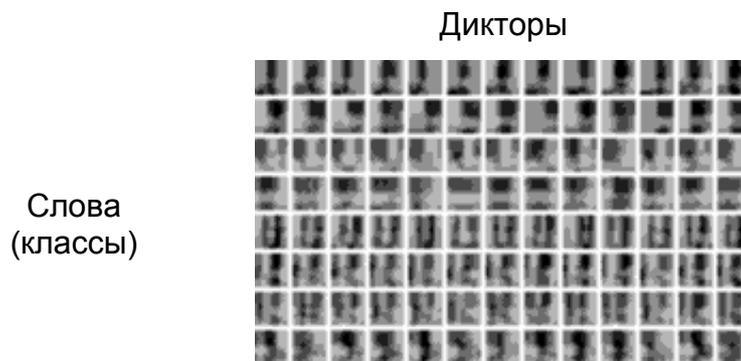


Рис. 3. Графические представления частотно-временных векторов для различных дикторов и различных слов.

Существует множество методов классификации образов с учителем. Рассмотрим основные из них: статистический, метрический и нейросетевой.

#### 3.1. Статистический метод классификации

Обозначим:

$W[j,i,x,y]$ : вектор из обучающей выборки, где  $j$  - номер класса,  $i$  - номер примера,  $(x,y)$  - координаты в векторе свойств.

$X[x,y]$ : вектор из тестовой выборки.

$P(C_j, X)$ : вероятность принадлежности  $X$  к классу  $C_j$  при распределении вероятностей.

Тогда, функция классификации  $class(X)$  определяется следующим образом:

$$M_{x,y}^j = \frac{1}{n_j} \sum_{i=1}^{n_j} W_{i,x,y}^j, \sigma_{x,y}^j = \frac{1}{n-1} \sum_{i=1}^n (M_{x,y}^j - W_{i,x,y}^j)^2, \theta_{norm} = \{M, \sigma\}$$

$$class(X) = \arg \max_j \left[ \prod_{x,y} P(C_j | X_{x,y}, \theta_{x,y}^j) \right] \quad (2)$$

### 3.2. Классификация методом наименьшего расстояния

$$class(X) = \arg \min_i \left[ \min_j \sum_{x,y} \rho(X_{x,y}, W_{j,x,y}^i) \right] \quad (4)$$

### 3.3. Нейросетевая классификация

$$class(X) = \arg \min_i \left[ \min_j \sum_{x,y} \rho(EVec(Ann(X)), Ann(X))_{j,x,y} \right], \quad (5)$$

$$EVec(x) = \left( (y_1, y_2, \dots, y_n) \mid y_i = \begin{cases} 1, & i = \arg \max_j (x_j); \\ 0. & \end{cases} \right)$$

где  $\rho$ -евклидово расстояние, функция  $Ann(x)$  - выход нейронной сети.

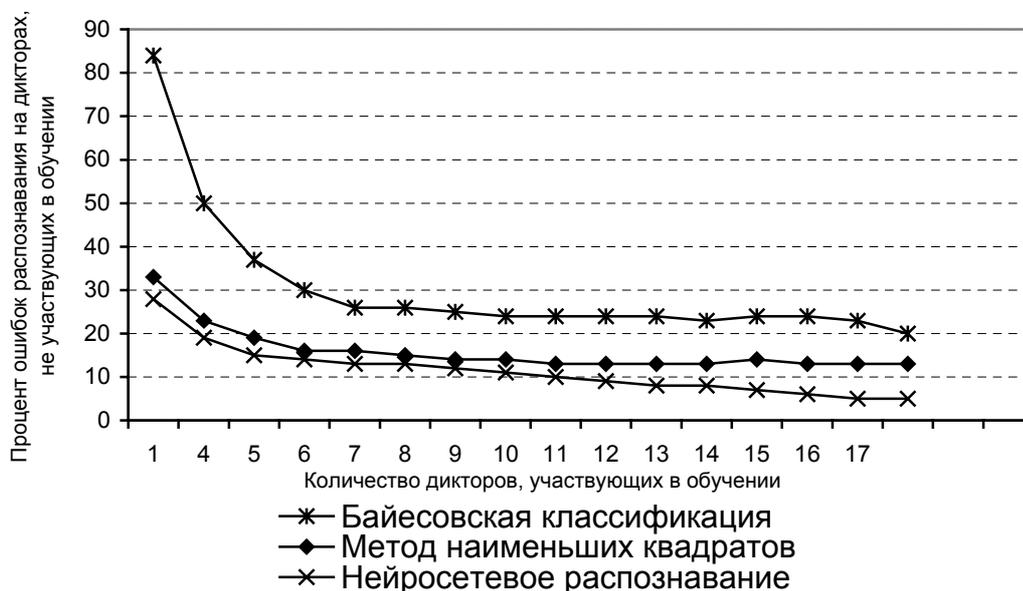


Рис 4. График зависимости процентов ошибок распознавания от метода классификации и количества дикторов, участвующих в обучении.

Из таблицы 1 и графика на рис. 4 видно, что наихудшая распознаваемость наблюдается у вероятностного подхода, основанного на двух априорных предположениях:

- 1) Нормальное распределение внутриклассовых компонент вектора
- 2) Статистическая независимость компонент вектора

Таблица 1. Процент ошибок распознавания в зависимости от метода классификации

Метод		Количество примеров в обучающей выборке		
		2	4	16
Метрический подход	Сумма квадратов	33.1%	18.1%	12.3%
	Максимум модуля	34.8%	19.5%	13.8%
	Сумма модуля	33.6%	19.2%	13.7%
Статическая классификация	Распределение Гаусса	80%	35.2%	32.0%
	Распределение Коши	49%	33.7%	32.2%
3x-слойная нейронная сеть		28%	13.3%	5.9%

Метрическое распознавание предполагает линейность вкладов каждой компоненты частотно-временного вектора в общую сумму. Единственным методом классификации, не делающим допущения о множестве векторов свойств, является нейросетевое распознавание. Известно, что 3x-слойная нейронная сеть с непрерывной сигмоидальной функцией активации может аппроксимировать любую непрерывную функцию с любой наперед заданной точностью. Отсюда следует способность классификации любых непрерывных конечных множеств. К недостаткам нейронной сети можно отнести потенциально экспоненциальное время обучения в зависимости от количества обучающих примеров. Однако такая зависимость наблюдается в случае слабой коррелированности векторов, входящих в базовый класс. Экспериментальная зависимость количества итераций при использовании реального тестового множества приведена на Рис. 5.

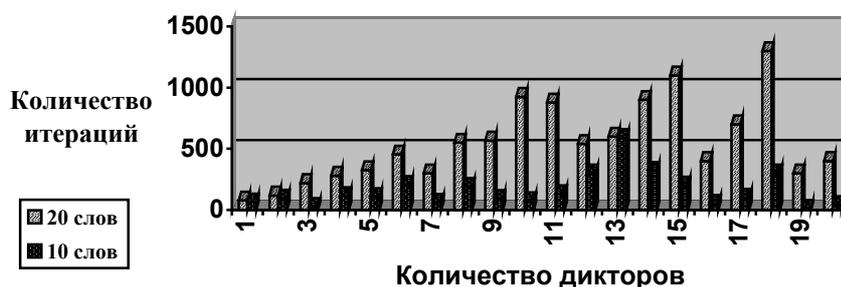


Рис 5. Зависимость количества итераций, необходимых для обучения сети, от количества дикторов и количества слов.

Основным критерием для систем распознавания речи является процент ошибки распознавания и надёжность распознавания. В таблице 2 приведены ошибки классификации нейронной сетью в зависимости от типа и интенсивности