Аннотация: в статье ставится и решается задача распознавания речевых команд в однокристальных микроконтроллерных системах. Для сокращения количества обрабатываемой речевой информации использованы законы психоакустики. Для распознавания речевых команд применен способ распознавания изображений с использованием сингулярного спектрального анализа по технологии Eigenface.

Ключевые слова: микроконтроллер, речевая команда, темп речи, быстрое преобразование Фурье, спектрограмма, мел-частотное преобразование, сингулярный спектральный анализ.

Подход к распознаванию речевых команд с использованием сингулярного спектрального анализа в системах с ограниченными ресурсами

Д.А. Вольф

1. Введение

В робототехнике однокристальный микроконтроллер (ОМК) – функционально законченное электронно-вычислительное средство (ЭВС), реализованное в виде одной БИС или СБИС. Однокристальный микроконтроллер включает в себя все устройства, необходимые для реализации цифровой системы управления. ОМК предназначен для использования в системах промышленной и бытовой автоматики. Возможности ОМК сильно ограничены аппаратными ресурсами, поэтому ОМК применяются в системах малой автоматизации в качестве полноценной системы управления либо в качестве системы сбора данных в сложных системах управления. На фоне быстро растущего интереса к любительской и бытовой робототехнике, тематика голосового управления системами на базе ОМК в последние время требует своего участия. Ввиду ограниченных аппаратных возможностей ОМК, решение задач распознавания речевых команд (РК) для таких систем является проблематичной. Так, например, на сегодняшний день самым популярным решением распознавания речи является использование программных пакетов Google Speech API [00] и Microsoft Speech API [00]. Однако, для первого и второго решения требуется взаимодействие с более мощным ЭВМ и обязательное наличие доступа в сеть Интернет, так как процедура распознавания речи осуществляется на серверах владельцев рассматриваемых компаний. Одним из интересных решений для ОМК устройств является расширение MOVI [00]. В отличии от двух предыдущих решений преимуществом MOVI является ее полная автономность, не требующая постоянного сетевого соединения с Интернет. Распознавание речи в MOVI осуществляется с помощью 1ГГц ARM-based процессора и 2-х гигабайт памяти под управлением операционной системы Debian Linux,

запускаемой с твердотельного носителя. Таким образом, само по себе MOVI является самостоятельным законченным решением и по факту является обычной ЭВМ надстройкой для ОМК устройств. Изучая рынок решений создается впечатление, что для распознавания человеческой речи нужен мощный процессор и большой объем памяти? Арджо Чакраварти (Arjo Chakravarty) [00] доказал обратное, реализовав библиотеку uSpeech [00] для устройств Arduino [00]. В данном решении, осуществляется распознавание фонем на основе вычисления мощности энергии сигнала без использования каких-либо методов анализа частотного спектра, обеспечив высокую скорость распознавания речи в ОМК (1-4 мс). Известно, что методы выделения фонем на основе вычисления энергии сигнала путем расчета средних квадратов, неустойчивы в случае низких параметров сигнал/шум. Поэтому в решении Чакраварти повышенные требования к используемому микрофону, пред-усилительному устройству и окружающей речевой обстановке.

И так, что мы наблюдаем? Для полноценного распознавания РК необходимо применение современных методов анализа и классификации. Такие методы требуют соответствующие сетевые и аппаратные ресурсы, уровня современных ЭВМ. Таким образом, очевидно, что решение задачи распознавания речи человека в системах с ограниченными ресурсами достаточна актуальна.

В представленной статье упор делается на исследование возможности применения технологии распознавания РК с помощью сингулярного спектрального анализа (ССА) в ОМК. Ограничения, установленные в настоящем исследовании на распознавание РК: уровень распознавания – слово, тип распознавания – дикторозависимое. Выбранные ограничения являются типовыми и обусловлены необходимостью отдавать речевые команды (РК) одним человеком или ограниченной группой людей.

2. Оценка размера спектрограммы для распознавания речевых команд

В психоакустике традиционно речевые сигналы рассматриваются в качестве спектрограмм (сонограмм) – зависимостей спектральных плотностей мощности речевых сигналов от времени. В практических задачах распознавания РК, актуальным вопросом является выбор размера спектрограммы, тем более для ОМК. Поэтому решим эту задачу. Согласно исследованиям [00-00], оптимальный темп речи для чтения аудиокниг на английском языке соответствует 150-160 словам в минуту (сл./мин.). Для личной беседы – 190 сл./мин. Для русского языка, в котором слова длиннее примерно на 20 – 30 %, темп равен 100 – 120 сл./мин., для личной беседы нормальный темп не превышает 150 сл./м. Таким образом, будем считать, что длительность одной речевой команды находится в диапазоне

$$\frac{150}{60} \div \frac{190}{60} \approx 2.5 \div 3 \text{ с.}$$

Речевую команду можно представить в виде дискретной функции разложения Фурье

$$x_n = \sum_{k=0}^{\frac{N}{2}-1} C_k e^{i\omega_k n}, \omega_k = \frac{2\pi k}{N}, C_k = A_k e^{\varphi_k}, n = 0,\ldots,N-1, \quad (1)$$

где

ω_k – радиальная частота k-й гармоники;

A_k – амплитуда k-й гармоники;

$e^{\varphi}{}_k$ – фаза k-й гармоники;

N – число отсчетов (для непрерывной функции (1) – период T);

n – номер отсчета (для непрерывной функции (1) – время t).

Решая обратную задачу для (1), параметрически оценивается спектр, в котором сканируется весь допустимый диапазон частот $\omega_k(\omega)$ в рамках объема N (периода T)

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-i\omega_k n}, \; k = 0,\ldots,\frac{N}{2}-1. \quad (2)$$

Результатом решения (2) является частотный спектр дискретного преобразования Фурье (ДПФ) длины $N/2$. Ввиду того, что длительность звучания фонем и дифтонгов варьируется в пределах 20-220мс, в практических задачах телефонии анализ речевых сигналов проводится с временными рядами длины 256, 512, 1024, 2048 отсчетов, что при частоте дискретизации 8кГц или 10кГц соответствует интервалам 31–250мс, 25-200мс. Таким образом, при указанных условиях, для решения задач распознавания РК в ОМК, размер спектрограмм выбирается из диапазона

$$\frac{N}{2} \times \left(\frac{2.5}{(31 \div 250) \cdot 10^{-3}} \div \frac{3}{(31 \div 250) \cdot 10^{-3}} \right) \approx \frac{N}{2} \times (10 \div 96). \quad (3)$$

Ограничимся речевыми командами с темпом в 1 с, а объёму N присвоим значение 256 (31 мс), тогда матрица $\mathbf{C}_n$:$N/2 \times L$ примет размер 128×32. Речевая команда разделяется на фреймы, делится на перекрывающиеся сегменты (преобразование исходного речевого сигнала в набор фреймов с перекрытием), после чего, во избежание эффектов Гиббса [00], каждый фрейм традиционно умножается на весовую функцию (Хэмминга, Ханна, или др.) [00], и в соответствии с (2) для него вычисляется ДПФ с накоплением частотного спектра в матрице

$$C_{k,l} = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-i\omega_k n}, \; k = 0,\ldots,\frac{N}{2}-1, \; l = 0,\ldots,L-1.$$

Таким образом, в виде исходных данных выступают матрицы $\mathbf{C}_n$:$N/2 \times L$, содержащие Фурье спектр РК (спектральные матрицы). Для ОМК размер спектральной матрицы 128×32 с

целочисленным типом байт, занимает порядка 4-х килобайт памяти. Отсюда появляется необходимость уменьшить размер спектральных матриц (уменьшить объем обрабатываемых данных в ОМК).

3. Сжатие спектрограммы с использованием частотного маскирования

Органы слуха человека обладают свойством частотного маскирования [00]. С данным понятием очень тесно связанно понятие критической полосы – ещё одна характеристика звука речи наряду с частотой. В отличие от частоты критические полосы определяются в соответствии со слуховым восприятием. Поэтому, на практике отбрасывают частотные составляющие, которые находятся за пределами этого диапазона и, соответственно, не несут информационной нагрузки. Исследования в области психофизического восприятия показали, что основная значимая информация содержится в действительном частотном спектре. Поэтому после выполнения преобразования Фурье, для дальнейшего анализа, выделяется только действительный спектр сигнала, а информация о фазе не рассматривается. Таким образом, первым действием при помощи полосового фильтра в мел-частотной области отбрасывается информация о частотных составляющих, не находящихся в диапазоне 96–3969 Гц (т.к. частота дискретизации 8кГц). Затем на полученную область (фрейм) накладываются взвешенные треугольные функции перекрывающихся окон, у которых значения центральных частот изменяются нелинейно в соответствии с мел-шкалой [00]. Например, звуковые колебания частотой 1000 Гц при эффективном звуковом давлении $2 \cdot 10^{-3}$ Па (то есть при уровне громкости 40 фон), воздействующие спереди на человека с нормальным слухом, вызывают у него восприятие высоты звука, оцениваемое по определению в 1000 мел. Звук частоты 20 Гц при уровне громкости 40 фон обладает по определению нулевой высотой (0 мел) [00]. Зависимость нелинейная, особенно при низких частотах. Далее в пределах полученных окон вычисляются средние значения действительного спектра, в результате чего получается сглаженный сильно коррелированный мел-спектр с различной детализацией диапазонов частот психофизической модели звукового восприятия (Рис. 1).

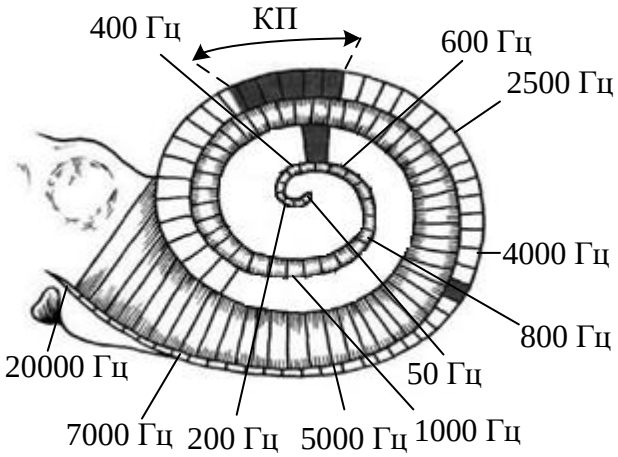

Рис. 1. Значения частот падающей звуковой волны в модели резонансной теории Гельмгольца. Анатомическая особенность улитки внутреннего уха обеспечивает нелинейную зависимость критической полосы. Толщина базальной мембраны у овального окна порядка 0,04 мм

Преобразование значений частоты (Гц) РК в значение высоты (мел) осуществляется по формуле

$$\mathrm{mel}(freq) = 1125 \ln(1 + \frac{freq}{700}),$$

соответственно обратное преобразование по формуле

$$\mathrm{freq}(mel) = 700\,(\mathrm{e}^{mel/1125} - 1).$$

Формирование банка (гребенки) мел-фильтров (Рис. 2) проводится в соответствии с системой

$$H_{m,j} = \begin{cases} 0 \forall k_j < f_{m-1}; \\ \dfrac{k_j - f_{m-1}}{f_m - f_{m-1}} \forall (f_{m-1} \le k_j) \wedge (k_j \le f_m); \\ \dfrac{f_{m+1} - k_j}{f_{m+1} - f_m} \forall (f_m \le k_j) \wedge (k_j \le f_{m+1}); \\ 0 \forall k_j > f_{m+1}. \end{cases}, k_j = \frac{j}{N} \cdot f_{\mathrm{d}},$$

$$f_m = \mathrm{freq}(\mathrm{mel}(f_{F1}) + \frac{\mathrm{mel}(f_{Fh}) - \mathrm{mel}(f_{F1})}{M} m),$$

$$j = 0,..,\frac{N}{2} - 1,\ m = 1,..,M - 2,$$

где:

$F1$ – левая граница первого мел-фильтра;

Fh – правая граница последнего мел-фильтра;

m – порядковый номер мел-фильтра;

f_{d} – частота дискретизации в Гц.

Преобразование (сжатие) спектральной матрицы (мел-спектральное преобразование) осуществляется матричным умножением

$$\Phi_{m,l} = \frac{1}{M}\sum_{k=0}^{\frac{N}{2}-1} H_{m,k} C_{k,l}\ , m = 0,..,M-1, l = 0,..,L$$

где:

m – номер канала фильтрации;

Ф – мел-спектральная матрица.

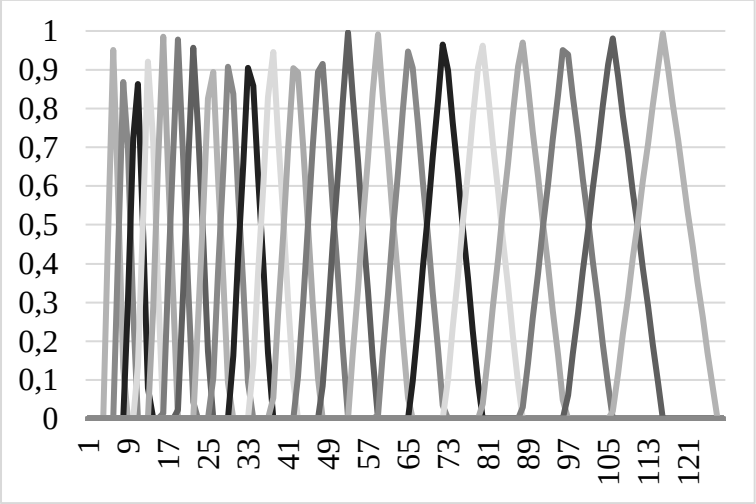

Рис.2. Взвешенные треугольные функции перекрывающихся окон в частотной области, представляющие собой банк мел-фильтров

Таким образом, обеспечивается мел-спектральное преобразование спектограммы речевой команды, содержащейся в спектральной матрице $\mathbf{C}_n$:$N/2 \times L$ до мел-спектральной матрицы $\mathbf{\Phi}_n$:$M \times L$ ($M << N/2$) меньших размеров. Это приводит к большему уплотнению количества речевой информации в области низких частот и меньшей плотности в области высоких частот, что отражает чувствительность восприятия звуковых сигналов человеческим ухом. Таким образом, основная информация речевой команды сосредотачивается (концентрируется) в области низких частот, что является наиболее релевантным признаком речевого сигнала, в результате чего обеспечивается уменьшение количества обрабатываемой информации в ОМК. Например, для N=256 и M=24, исходная матрица с размерами 128×32 сжимается до размеров 24×32, что соответствует 768 байтам занимаемой памяти ОМК, а при M=10 ($\mathbf{\Phi}_n$:10×32) – 320 байт. Такое уменьшение количества обрабатываемой информации можно рассматривать как решение одной из задач перцептуального кодирования речи, в которой эффективность сжатия исходной информации составляет 81-90% (Рис. 3).

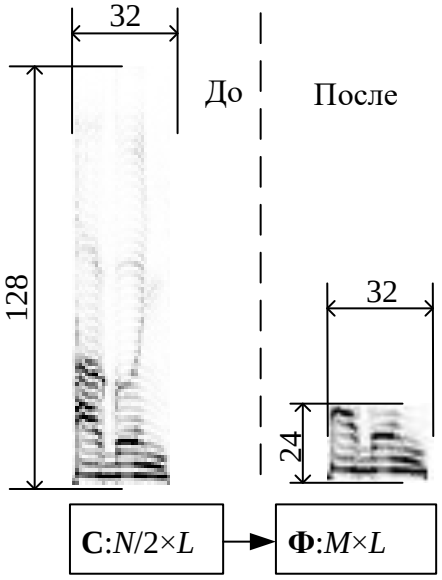

Рис. 3. Результат мел-спектрального преобразования спектограммы речевой команды «Огонь» длительностью 1 с., содержащейся в спектральной матрице **C** в мел-спектральную матрицу **Φ** меньших размеров

4. Распознавание мел-спектральных матриц с помощью сингулярного спектрального анализа

Под распознаванием РК будем понимать операцию различения слов речи в виде векторов изображений **g** и связывания их с имеющейся системой представлений **W**. Так как содержимое спектральных и мел-спектральных матриц можно рассматривать как графическое изображение с глубиной 255 бит, то к распознаванию мел-спектральных матриц с РК применим технологию распознавания лиц Eigenface [00].

Смоделируем систему представлений **W**. Из набора мел-спектральных матриц $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2,\ldots,\mathbf{\Phi}_R$:$M\times L$ (изображений) можно получить матрицу $\mathbf{\Gamma}$:$P\times R$ ($P=M\cdot L$), где $\mathbf{\Gamma}^{<n>}$ – вектор изображения $\mathbf{\Phi}_n$ в многомерном пространстве размерностью P. Будем считать, что вектор $\mathbf{\Gamma}^{<n>}$ задает n-ю точку в P-мерном пространстве **Ω**, таким образом, целевую задачу сведем к наилучшей аппроксимации конечного множества точек данного пространства. Для этого нужно снизить количество избыточной информации и выделить наиболее значимые (информативные) признаки изображений в **Ω** (снижение размерности $P\times R$ до $R\times R$). Сперва, оставим только уникальную информацию, убрав общие элементы для всех изображений в **Ω**. Для этого нормализуем все изображения в обучающей выборке $\mathbf{\Gamma}^{<n>}$, вычетом среднего вектора-изображения

$$\mathbf{G}^{<n>} = \mathbf{\Gamma}^{<n>} - \mathbf{\Psi},\ \Psi_i = \frac{1}{R}\sum_{j}^{R-1}\Gamma_{i,j},\ i=\overline{0,P-1},$$

где $\mathbf{\Psi}$ – средний вектор-изображение.

Далее, для системы векторов **G**:*P*×*R* найдем проектор **X,** состоящий из собственных чисел [00], для осуществления проекции **G**:*P*×*R* в информативный базис **W**. Для этого необходимо решить задачу сингулярного спектрального разложения матрицы **G**:*P*×*R*. Под нахождением информативного базиса понимается вычисление (нахождение) такого числа главных компонент (метод главных компонент), в котором дисперсия, определяемая собственными значениями λ_p не ниже 80% (количество информации от изображения), т.е.

$$S = \sum_p \frac{\lambda_p}{\sum \lambda} \geq 80\% . \qquad (4)$$

Остальную информацию (компоненты) отнесем к шуму. Задачу сингулярного спектрального разложения матрицы **G**:*P*×*R* сведем к задаче спектрального разложения ковариационной матрицы **A**:*R*×*R*,

$$\mathbf{A} = \mathbf{G}^{\mathrm{T}}\mathbf{G} .$$

Тем самым, большая часть вариации изображений сосредоточится в первых главных компонентах (координатах), что позволит перейти к пространству меньшей размерности.

Пусть для матрицы **A**:*R*×*R* найдено сингулярное спектральное разложение

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}},$$

тогда пространство изображений **G**:*P*×*R* может быть представлено в виде суммы взаимно ортогональных собственных подпространств

$$\mathbf{G} = \sum_{i=0}^{R-1} \tilde{\mathbf{G}}_i = \sum_{i=0}^{R-1} (\sqrt{\lambda_i}\,\mathbf{v}^{<i>})[\mathbf{u}^{<i>}]^{\mathrm{T}} \Rightarrow \mathbf{x}^{<i>} = \sqrt{\lambda_i}[\mathbf{u}^{<i>}] = \mathbf{G}\mathbf{v}^{<i>},$$

где:

$\lambda_i \in \boldsymbol{\Sigma}$ – *i*-е собственное значение матрицы **A**:*R*×*R*;

$\mathbf{v}^{<i>} \in \mathbf{V}$ – *i*-й собственный вектор матрицы **A**:*R*×*R*;

$\mathbf{u}^{<i>} \in \mathbf{U}$ – *i*-й собственный вектор матрицы **G**:*P*×*R*;

$\mathbf{x}^{<i>} \in \mathbf{X}$ – *i*-й проектор.

Проекцию **G**:*P*×*R* в информативном базисе **V** найдем как

$$\mathbf{W} = \mathbf{X}^{\mathrm{T}}\mathbf{G} .$$

Аналогично тому, как вектор $\mathbf{G}^{<n>}$ задает *n*-ю точку в *P*-мерном пространстве **Ω**, так и вектор $\mathbf{w}^{<n>} \in \mathbf{W}$:*R*×*R* – задает *n*-ю проекцию точки $\mathbf{G}^{<n>}$ в базисе **V**, а матрица (оператор) **X** – это проектор изображения **G** в пространство **W**, меньшей размерности *R*<<*P*. Проекцию **W** определим как пространство РК (Eigenspeech space), а вектор $\mathbf{w}^{<n>}$ как элемент этого пространства, задающий точку в этом пространстве, т.е. вариацию конкретной речевой команды. Тогда матрица векторов **W** – это поле вариаций РК, сосредоточенных в различных локальных областях пространства **V** (пространственная информация), поэтому матрицу **W**

будем рассматривать как пространственные данные. Таким образом, распознавание РК, обеспечивается предварительными расчетами проектора **X** и поля вариаций РК **W**. Расчет поля вариаций РК **W** (моделирование системы представлений пространственных данных) будем называть – обучением.

Теперь рассмотрим операцию различения векторов изображений **g**. Из распознаваемого вектора изображения $\mathbf{g} \notin \mathbf{\Gamma}^{<n>}$ вычитается средний вектор-изображение **Ψ**, производится проекция $\mathbf{g} \in \mathbf{\Omega}$ в **V**, вычисляется ближайшее расстояние между проекцией и вариациями РК в **W**, наименьшее значение определяет принадлежность распознаваемого изображения **g** к той или иной команде.

В качестве общего решения формализуем изложенное в виде модели Eigenspeech Recognition, позволяющей решать задачи распознавания РК с помощью ССА (Рис. 4). Имеется исполнительная система, в которую поступают приказы от подсистемы распознавания РК (ПРРК). В состав ПРРК входят: 1 – блок предварительной обработки РК, 2 – блок формирования спектрограммы, 3 – блок мел-спектрального преобразования, 4 – блок сбора обучающей выборки, 5 – блок формирования векторного пространства, 6 – блок вычисления проектора, 7 – блок формирования распознаваемого вектора изображения, 8 – блок проецирования вектора изображения в базис собственных векторов, 9 – блок с полем вариаций РК (пространственных данных), 10 – блок классификации. Работа ПРРК, обеспечивается двумя режимами – обучение и распознавание. На входе ПРКК речевая команда, а на выходе соответствующий номер приказа для системы исполнения. В некотором приближении полученная модель ПРКК напоминает ассоциативную память. Вычисление проектора **X,** обеспечивается известными алгоритмами поиска собственных чисел – Tred2 [00] и Tqli [00]. Выбор данных алгоритмов мотивирован экономичным использованием ресурсов ОМК, их программными реализациями на языке С. Как уже было обозначено, основная проблема в популяризации распознавания РК в ОМК упирается в ограниченные аппаратные возможности. Поэтому, например, для режима распознавания РК достаточно в памяти программ ОМК хранить информацию о проекторе **X** и пространственных данных **W**, которые могут быть вычислены заранее на более мощных ЭВМ. Тогда, процесс распознавания РК в ОМК сведется к обычным матричным операциям сложения и умножения. Также, актуальным будет приведение значений, содержащихся в проекторе **X** к целочисленному типу. Ошибка округления будет применена ко всем данным, и, таким образом, не повлияет на результат распознавания, а экономия памяти ОМК будет существенна.

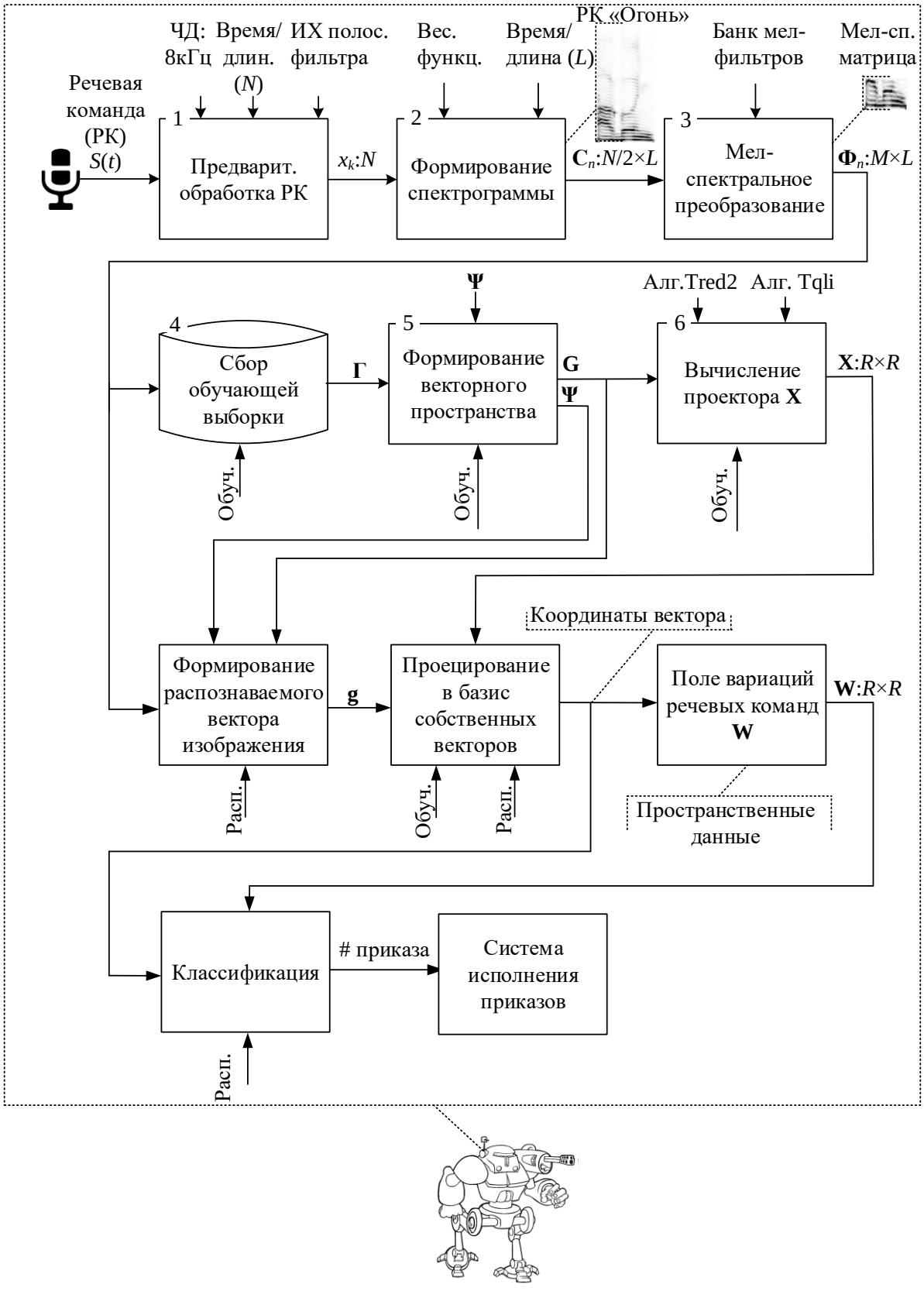

Рис. 4. Модель сингулярного распознавания РК (Eigenspeech Recognition) в исполнителе

5. Тестирование

Для тестирования использовалась экспериментальная установка для распознавания РК (Рис. 5), в составе: микрофонной ветви, состоящей из микрофона M1, сопротивления R1 (2k2); разделительного конденсатора C1 (220n0); усилительного каскада, состоящего из транзистора VT1 (КТ315) и сопротивлений R2 (180k), R3 (130k), R4 (2k2), R5 (470); средства отображения результата, состоящее из светодиодов HL1-HL3 и резисторов R6-R8 (270); ОМК DD1 серии AVR Atmega328; кварцевого резонатора BQ (16MHz); шунтирующей обвязки, состоящей из конденсаторов C2-C6 (22n0, 220n0, 100, 22p0, 22p0). Выбранный ОМК, широко применяется в любительской робототехнике и обладает следующими техническими характеристиками:

1. FLASH память, 32 Кбайт.
2. EEPROM память, 1 Кбайт.
3. SRAM память, 2 Кбайт.
4. Рабочая частота, 16 МГц.
5. Рабочее напряжение, 5В.

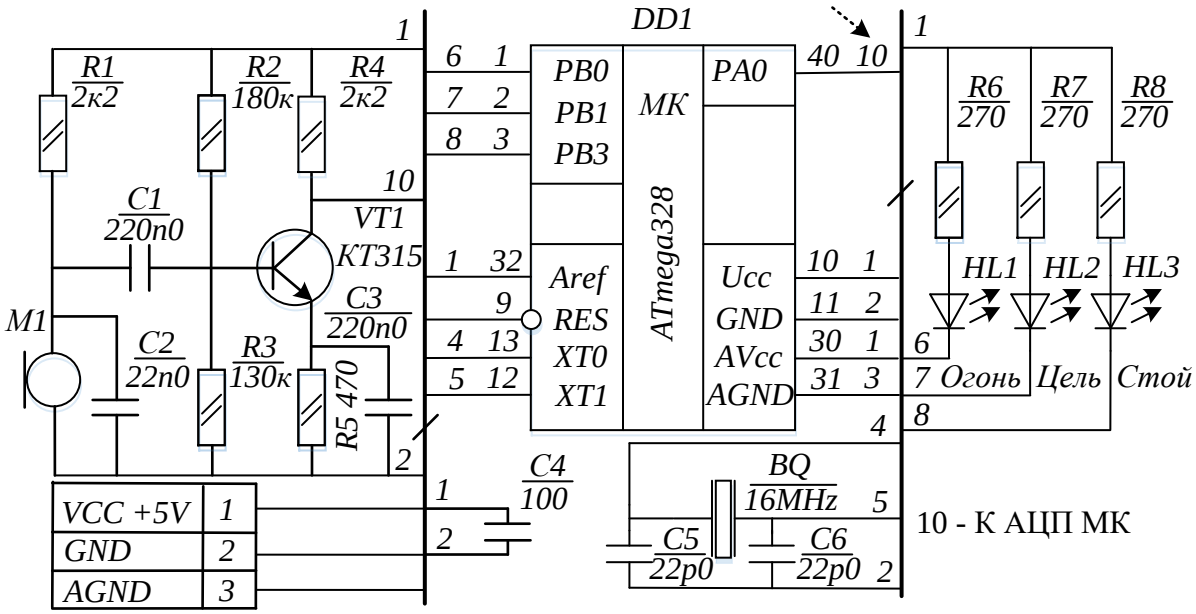

Рис. 5. Экспериментальная установка для распознавания РК. Схема электрическая

В качестве данных для обучения выбраны спектрограммы трех боевых команд «Огонь», «Цель» и «Стой» (Рис. 6). Далее, спектрограммы приведены к мел-спектральным матрицам размером 24×32 и 12×32 (Таблицы 1-2). Расчет проектора **X** и поля вариаций РК **W** осуществлялся отдельно. Резервирование постоянной памяти (памяти программ) для проектора **X**:30×678 требовало порядка 82 Кбайт данных вещественного типа (4 байт), что превышало текущий предел в ОМК – 32 Кбайт. Однако, исходя из (4), для обучающей выборки мел-спектральных матриц 24×32 (Таблица 1) выбраны первые 7-мь собственных векторов, а для матриц 12×32 (Таблица 2) – первые 11. В результате под проектор **X**:7×678

потребовалось порядка 20 Кбайт, или 16 Кбайт для **X**:11×352. Для уменшения хранимой информации в 2 раза, осуществленно приведение вещественных значений, содержащихся в проекторе **X** к целочисленным (к двухбайтным значениям – int16_t). В итоге занимаемый объем памяти программ (FLASH) матрицами **X** и **W** составил порядка 13 Кбайт (41% от общего объема FLASH ОМК).

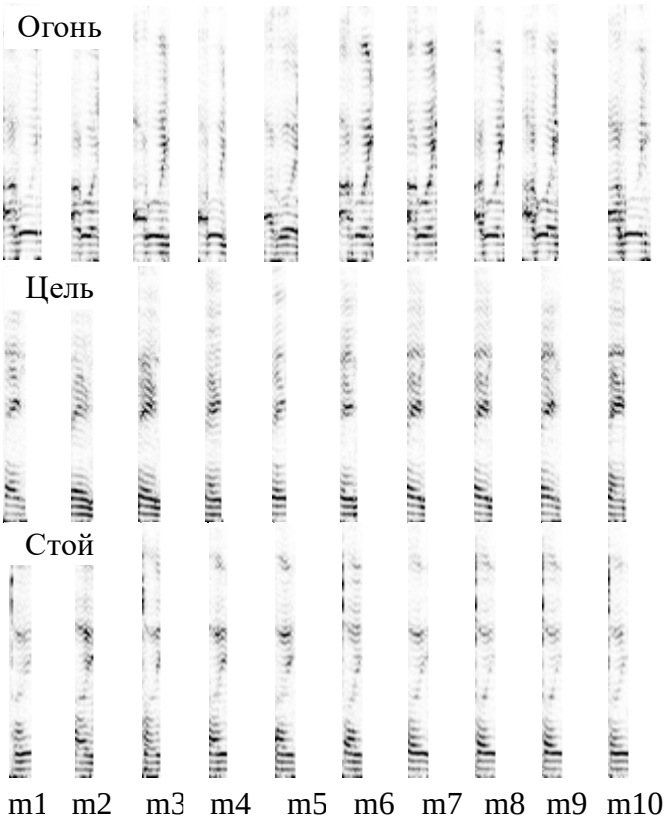

Рис. 6. Исходные данные. Спектограммы (128×32) команд «Огонь», «Цель», «Стой» каждая в 10 вариациях, произнесенная одним диктором

Таблица 1. Начальные данные для обучения – мел-спектральные матрицы 24×32

РК	Выборка для обучения									
№	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
Огонь										
Цель										
Стой										

Таблица 2. Начальные данные для обучения – мел-спектральные матрицы 12×32

РК	Выборка для обучения									
№	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
Огонь										
Цель										
Стой										

После обучения, с помощью экспериментальной установки для распознавания РК, диктором проводилось тестирование. При произнесении РК «Огонь», «Цель», «Стой» в пределах выбранного темпа 1с зажигался соответствующий светодиод из HL1- HL3. Каждая ассоциация произнесенной РК, в виде распознаваемого вектора изображения $\mathbf{g} \notin \mathbf{\Gamma}$, связывается с соответствующей вариацией из **W**, что подтверждается респределением **W** и **g** в проекциях **V**. Так, например, в проекциях $\mathbf{V}^0\mathbf{V}^1$, $\mathbf{V}^0\mathbf{V}^2$, $\mathbf{V}^0\mathbf{V}^3$ можно наблюдать кучности с выраженными границами (Рис. 7, а, б, в).

В качестве доработки выбранного метода распознавания, предложено вычислять центры масс кучностей в **W**, и, таким образом, операцию различения свести к вычислению ближайшего расстояния между проекциями **g** и математическими ожиданиями случайных векторов в пространстве **V** (Рис. 8). Такое предложение открывает стохастический подход к решению задач распознавания, а также сокращению хранимой и обрабатываемой информации в ОМК.

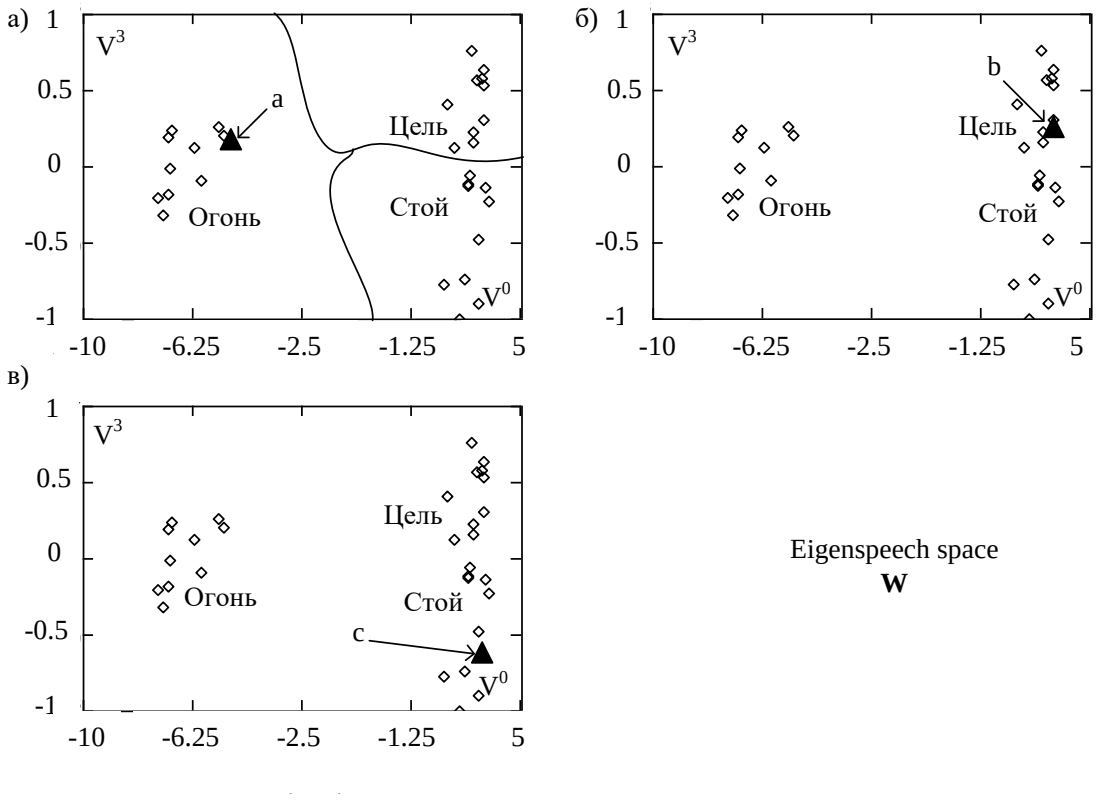

Рис. 7. Распределение вариаций РК **W** в проекции V^0V^3 (Eigenspeech space). Латинскими буквами a,b,c обозначены проекции распознаваемых образцов **g** в соотвествующих областях кучности

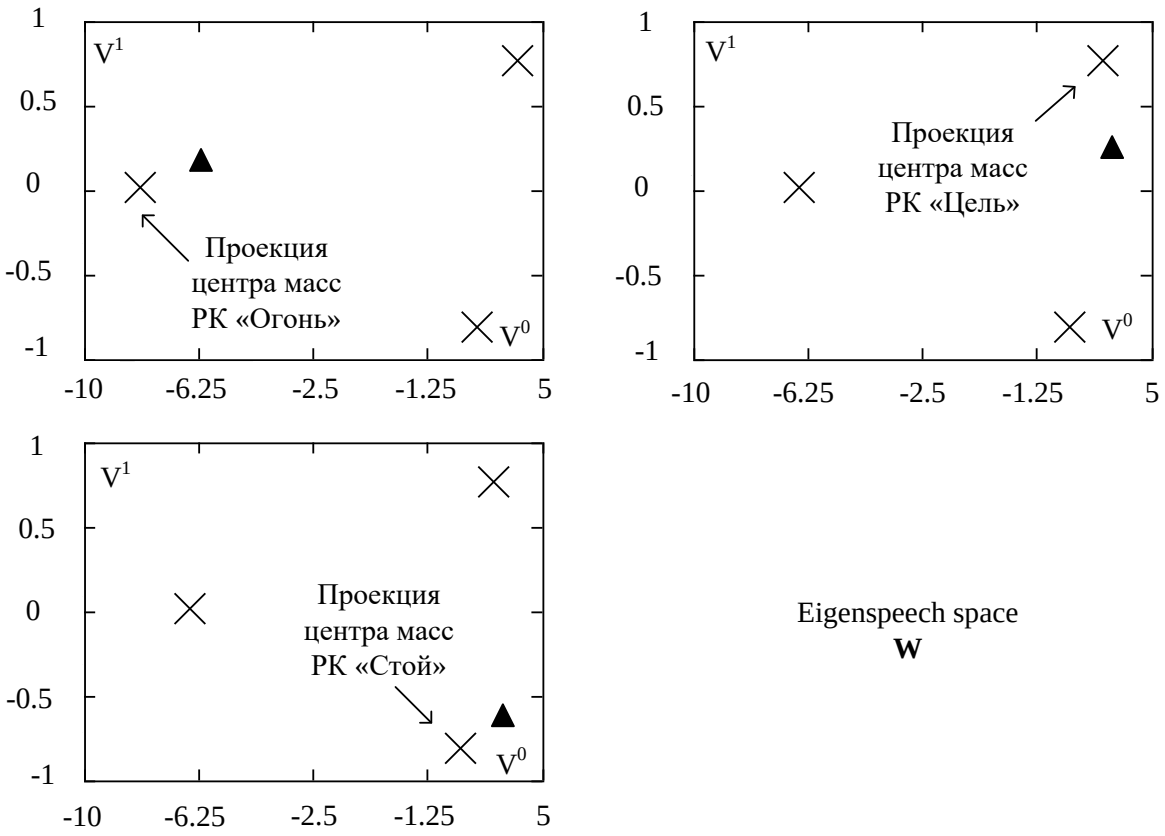

Рис. 8. Проекции центров масс вариаций РК **W** и распознаваемых образцов **g** в проекции V^0V^1

6. Заключение

Решена задача распознавания РК для систем с ограниченными ресурсами – предложены методы преобразования (сокращения) речевой информации до оптимальных объемов, позволяющих сохранить ее существенные признаки для задач распознавания. Предложена модель сингулярного распознавания РК (Eigenspeech Recognition), использующая технологию распознавания лица человека (Eigenface). Результаты проведенного исследования подтверждают возможность применения технологии Eigenface для распознавания РК в ОМК. Предложена доработка выбранного метода распознавания, которая позволяет обеспечить стохастический подход к классификации распознаваемых образов речевой информации. Достоинство модели Eigenspeech Recognition – это отсутствие в необходимости предварительной фильтрации РК, т.к. специфика применяемой технологии анализа речевой информации с помощью ССА на этапе обработки информации уже позволяет игнорировать шумовую составляющую, т.е. нечувствительна к ней.

Источники