УДК 612.789.4

А. К. Алимурадов[1], *А. Ю. Тычков*[1], *А. П. Зарецкий*[2], *А. П. Кулешов*[3], *П. П. Чураков*[1], *Ю. С. Квитка*[1]

[1]ФГБОУ ВО «Пензенский государственный университет», научно-исследовательский институт фундаментальных и прикладных исследований, 440026, г. Пенза
[2]Московский физико-технический институт (государственный университет)
[3]ФГБУ «ФНЦТИО им. ак. В. И. Шумакова» Минздрава России

Метод повышения эффективности голосового управления на основе комплементарной множественной декомпозиции на эмпирические моды

Низкая точность распознавания речевых команд является одной из основных проблем практической реализации систем голосового управления (СГУ). Это связано с использованием неэффективных и неадаптивных методов обработки речевых сигналов. В данной статье предложен усовершенствованный алгоритм распознавания речевых команд с использованием адаптивной технологии обработки – комплементарной множественной декомпозиции на эмпирические моды (КМДЭМ). Представлена блок-схема и подробное математическое описание основных блоков алгоритма. Суть предложенного алгоритма заключается в выделении из исходного речевого сигнала информации об уникальных свойствах голоса. Результаты экспериментального исследования предложенного алгоритма демонстрируют повышение точности распознавания речевых команд и эффективности голосового управления по сравнению с известными аналогами «RWTH ASR», «Julius» и «CMU Sphinx».

Ключевые слова: голосовое управление, СГУ, обработка речевых сигналов, КМДЭМ, мел-частотные кепстральные коэффициенты (МЧКК).

A. K. Alimuradov[1], *A. Yu. Tychkov*[1], *A. P. Zaretskiy*[2], *A. P. Kuleshov*[3], *P. P. Churakov*[1], *Y. S. Kvitka*[1]

[1]Penza State University, Research institute for basic and applied studies, 440026, Penza
[2]Moscow Institute of Physics and Technology (State University)
[3]Federal State Budgetary Institute V. I.Shumakov Federal Research Center of Transplantology and Artificial Organs, Ministry of Health of the Russian Federation

Method for improving efficiency of voice control based on the complementary ensemble empirical mode decomposition

The low accuracy of voice recognition is one of the main problems of practical implementation of voice control systems (VCS). It is associated with the use of inefficient and non-adaptive methods for speech signal processing. An improved algorithm for recognizing voice commands using the adaptive processing technology of the complementary ensemble empirical mode decomposition (CEEMD) is proposed. The block diagram and detailed mathematical description of basic blocks of the algorithm are given. A distinctive feature of the proposed algorithm is to extract only useful information of the unique properties of voice from the original speech signal. The experimental results of the proposed algorithm show the improved accuracy of voice commands recognition and efficiency of voice control as compared to the known RWTH ASR, Julius, and CMU Sphinx analogues.

Key words: voice control, VCS, speech signals processing, CEEMD, mel-frequency cepstral coefficients (MFCC).

1. Введение

СГУ основаны на технологии распознавания речи, которое сводится к обработке и анализу речевых команд с целью определения информативных параметров. Работа в направлении повышения эффективности голосового управления ведется достаточно активно. На сегодняшний день представлено большое количество алгоритмов, повышающих точность распознавания и эффективность голосового управления. Разнообразие алгоритмов обусловлено как важностью проблемы, так и отсутствием достаточно эффективных методов ее решения. Широкую практическую популярность получили алгоритмы распознавания с открытым исходным кодом: $RWTHASR$ [1], $Julius$ [2] и $CMUSphinx$ [3]. На рис. 1а представлена блок-схема классического алгоритма, применяемого в СГУ. Штриховой линией отмечен режим обучения, сплошной линией – рабочий режим алгоритма. Как видно из рисунка, точность распознавания зависит от предварительной обработки (блоки зеленого цвета), точности определения информативных параметров – МЧКК (блоки синего цвета) и распознавания (блоки красного цвета). Основная причина низкой точности связана с использованием неэффективных и неадаптивных методов обработки речевых сигналов. Исследования существующих методов обработки речевых сигналов, применяемых в СГУ [4], выявили перспективность использования адаптивной технологии анализа нестационарных данных – КМДЭМ [5]. Целью данной статьи является разработка алгоритма, повышающего точность распознавания речевых команд и эффективность голосового управления за счет применения КМДЭМ. Статья является развитием ранее опубликованных трудов авторов [6–8].

2. Комплементарная множественная декомпозиция на эмпирические моды

КМДЭМ представляет собой адаптивную технологию разложения сигнала на эмпирические моды (ЭМ). Адаптивность метода заключается в том, что базисные функции, используемые для разложения, извлекаются непосредственно из исходного сигнала. Аналитическое выражение [5] декомпозиции имеет следующий вид:

$$x(t) = \sum_{i=1}^{I} IMF_i(t) + r_I(t), \tag{1}$$

где $x(t)$ – исходный речевой сигнал, i – номер ЭМ, I – количество ЭМ, $IMF_i(t)$ – конечное число извлекаемых ЭМ, $r_I(t)$ – результирующий остаток.

Особенность метода КМДЭМ заключается в многократном добавлении к исходному речевому сигналу белого шума с прямыми и инверсными значениями амплитуды и вычислении среднего значения полученных мод как конечного истинного результата независимо от того, сколько сигналов белого шума использовалось (2–4):

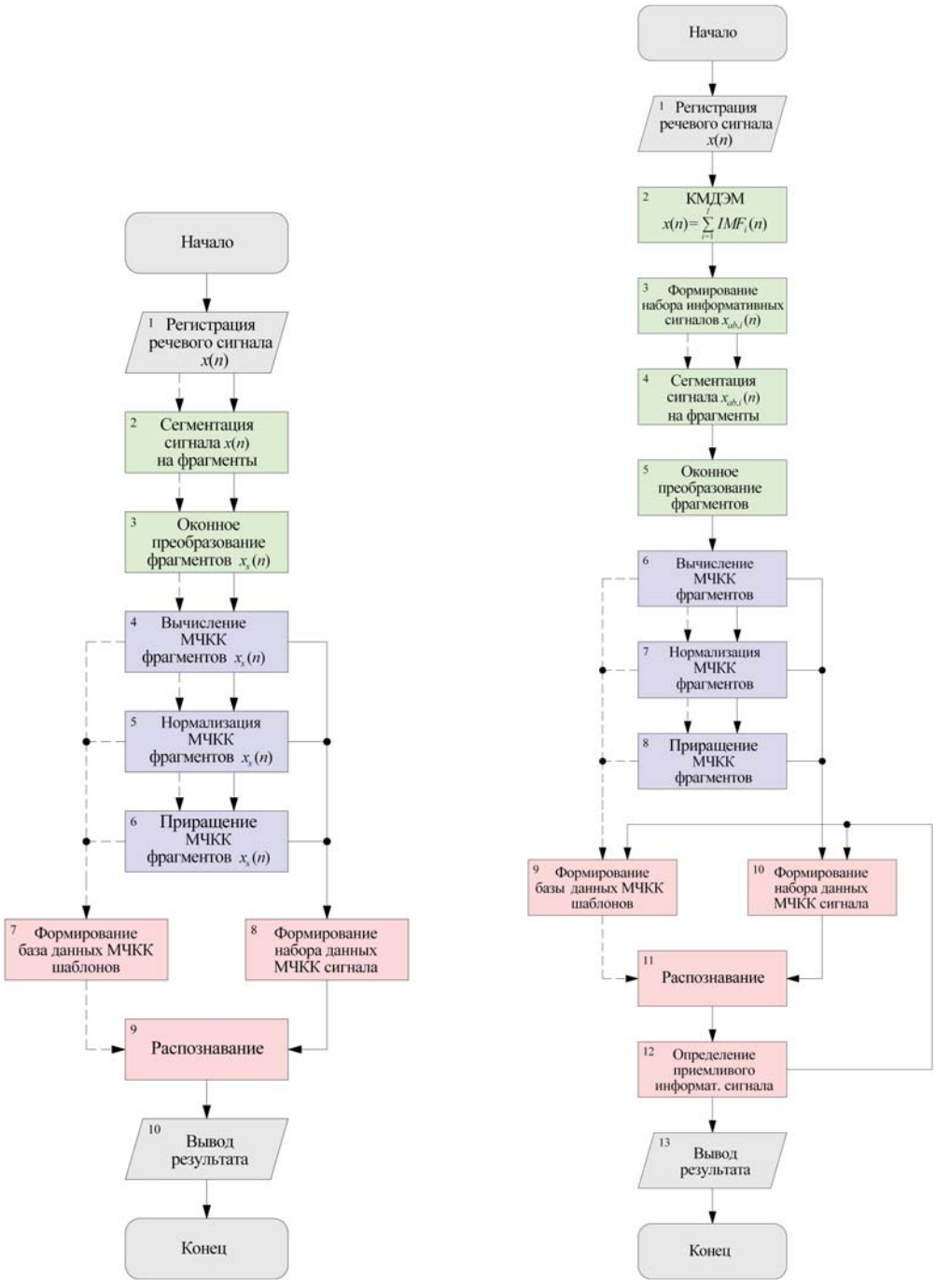
$$\begin{bmatrix} x_j(t) \\ x_j(t)^* \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} x(t) \\ w_i(t) \end{bmatrix}, \tag{2}$$

$$IMF_i(t) = \frac{\sum_{j=1}^{J} IMF_{Ji}(t)}{J}, \tag{3}$$

$$r_I(t) = \frac{\sum_{j=1}^{J} r_{Ji}(t)}{J}, \tag{4}$$

где $x_j(t)$ – зашумленный белым шумом речевой сигнал, $x_j(t)^*$ – зашумленный инверсным по знаку белым шумом речевой сигнал, $IMF_{Ji}(t), r_{Ji}(t)$ – ЭМ и результирующий остаток,

полученные при различных декомпозициях, $j = 1, 2, \ldots J$ – количество циклов декомпозиций (добавлений к сигналу белого шума). КМДЭМ в полной мере использует преимущество статистических характеристик белого шума для обнаружения слабых периодических участков речевых сигналов с минимальным значением остаточного шума.



а. Классическая б. Усовершенствованная

Рис. 1. Блок-схемы алгоритма распознавания речевых команд

3. Алгоритм распознавания речевых команд на основе КМДЭМ

На рис. 1б представлена блок-схема усовершенствованного алгоритма распознавания речевых команд. Суть предложенного алгоритма заключается в усовершенствовании этапа предварительной обработки – уменьшении разницы между поступающими в систему

речевыми командами и шаблонами (полученными в ходе обучения) посредством выделения из исходного сигнала только полезной информации об уникальных свойствах голоса для последующего распознавания. Рассмотрим подробнее основные этапы работы предложенного алгоритма.

Блок 1. Регистрация сигнала речевой команды $x(n)$ осуществляется со следующими параметрами: длительность записи – не более 3000 мс, частота дискретизации 8000 Гц, разрядность квантования 16 бит, где n – дискретный отсчет времени, $0 < n \leqslant N, N$ – количество дискретных отсчетов в сигнале.

Блок 2. Результатом КМДЭМ сигнала речевой команды будет конечное число ЭМ $IMF_i(n)$ и результирующий остаток $r_I(n)$, где i – номер ЭМ, I – количество ЭМ.

Блок 3. Ключевым понятием при формировании набора информативных сигналов является информативность ЭМ. При условии, что речевой сигнал имеет конечную энергию, число ЭМ при разложении всегда является конечным. Для абсолютно произвольного сигнала все ЭМ можно разбить на две категории [9]:

- информативные ЭМ с шумовыми и сигнальными составляющими;
- неинформативные ЭМ с трендовыми и компенсирующими составляющими.

Информативные ЭМ в разложении всегда отражают внутреннюю структуру и особенности речевого сигнала. К их числу относятся шумовые и сигнальные ЭМ. Появление в разложении первых объясняется наличием в исходном сигнале остаточного шума, а вторые связаны непосредственно с полезным сигналом и входящими в него компонентами. Неинформативные ЭМ являются медленно меняющимися функциями. Среди них выделяют трендовые ЭМ, описывающие истинную динамику среднего значения сигнала и компенсирующие ЭМ, возникающие при разложении. Трендовые ЭМ появляются, например, при разложении суммы гармонического сигнала и полиномиального тренда. Компенсирующие (ложные) ЭМ – результат несовершенства самого алгоритма декомпозиции (критериев остановки процесса отсеивания, неточностей при вычислениях, ошибок округления). Их появление не связано с какими-либо физическими или математическими особенностями рассматриваемых сигналов, а объясняется только лишь несовершенством вычислительной процедуры. Компенсирующие ЭМ обычно создают избыточность в разложении [10], а их название объясняется тем, что в сумме они дают функцию, очень близкую к нулю, и, по сути, компенсируют друг друга. Формирование набора информативных сигналов заключается в вычитании из исходного сигнала речевой команды информативных шумовых и неинформативных ЭМ. Информативными шумовыми обычно являются первые две или три ЭМ, в зависимости от интенсивности присутствующего в сигнале шума. Неинформативными являются последние три или четыре ЭМ, в зависимости от общего количества мод (число ЭМ примерно равно двоичному логарифму от числа отсчетов в сигнале). Формирование набора информативных сигналов осуществляется по формуле:

$$x_{ab,i}(n) = x(n) - (a \cdot \sum_{i=0}^{2} IMF_{i+1}(n) + b \cdot \sum_{i=0}^{2} IMF_{I-1}(n)), \tag{5}$$

где $x_{ab,i}(n)$ – информативный сигнал, $x(n)$ – исходный сигнал речевой команды, i – номер ЭМ, I – количество ЭМ, a, b – коэффициенты, определяющие участие ЭМ в формировании набора информативных сигналов. На рис. 2 представлена графическая интерпретация примера формирования набора информативных сигналов. Исходный речевой сигнал разлагается на десять ЭМ. Вычитая информативные шумовые и неинформативные ЭМ, сформирован набор, состоящий из восьми информативных сигналов (5).

Целью формирования набора информативных сигналов является возможность выбора одного сигнала, содержащего максимально большее количество информации об уникальных свойствах голоса. В последующих действиях работы алгоритма будет выбран наиболее приемлемый информативный сигнал, обеспечивающий наименьшую разницу между поступающей в систему речевой командой и шаблоном.

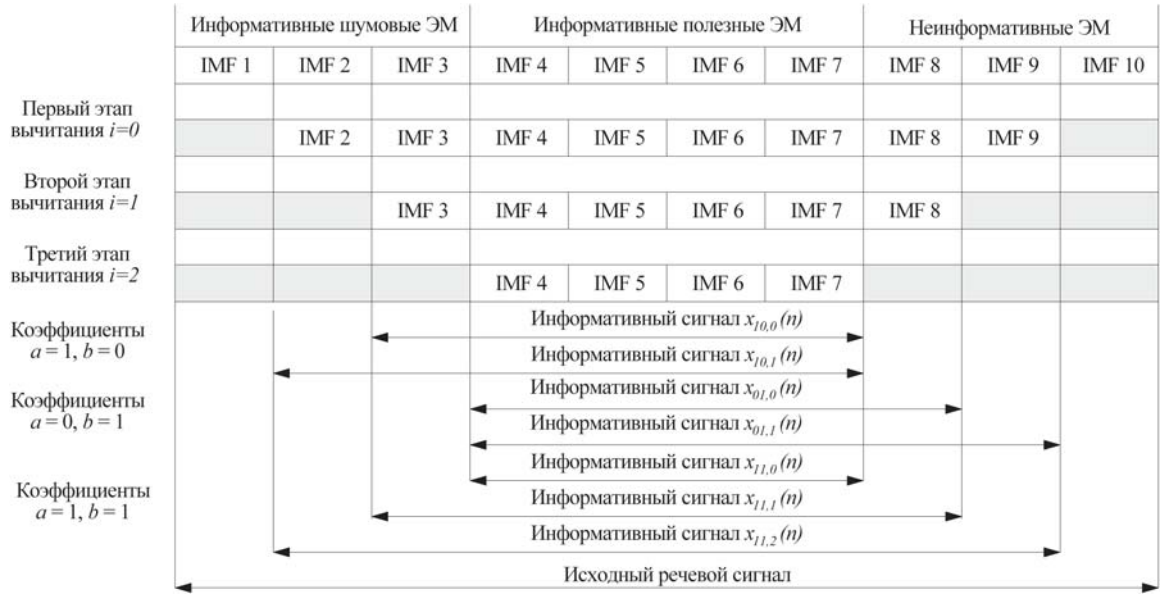


Рис. 2. Графическая интерпретация формирования набора информативных сигналов

Блок 4. Сегментация ЭМ – это линейное деление на составляющие отрезки, называемые фрагментами. Разработанный алгоритм основан на предположении о том, что свойства речевого сигнала с течением времени изменяются медленно. Это предположение приводит к кратковременному анализу, в котором фрагменты ЭМ выделяются и обрабатываются так, как если бы они были короткими участками с отличающимися свойствами. Сегментация ЭМ на фрагменты осуществляется по следующим формулам:

$$S = \frac{IMF_i(n)}{L}, \tag{6}$$

$$y_{i,s+1}(n) = IMF_i((s \cdot L) + 1; (s + 1) \cdot L), \tag{7}$$

где S – количество фрагментов в ЭМ, L – количество дискретных отсчетов в одном фрагменте, $y_{i,s+1}(n)$ – фрагмент i-й ЭМ, $s = 0, 1, 2, \ldots S$–1 – номер фрагмента.

Блок 5. Для уменьшения искажения спектра при обработке речевых сигналов используют оконное преобразование с плавно спадающими до нуля краями. Традиционно в обработке используется окно Хэмминга – вектор коэффициентов в дискретном виде, определяемый по формуле:

$$\omega(u + 1) = 0,54 - 0,46 \cdot \cos(2\pi \frac{u}{U - 1}), \tag{8}$$

где $u = 0, 1, 2, \ldots U$–1 – номер коэффициента окна Хэмминга.

Оконное преобразование фрагмента i-й ЭМ осуществляется по формуле:

$$y_{i,s+1}(n) = y_{i,s+1}(n) \otimes \omega(u + 1). \tag{9}$$

Блок 6. В качестве информативных параметров в предложенном алгоритме используются МЧКК [10], которые включают в себя два основных понятия: кепстр и мел-шкала. Кепстр – это дискретно-косинусное преобразование амплитудного спектра сигнала в логарифмическом масштабе. Кепстр сигнала определяется по формуле:

$$c(n) = DCT[\log(X(\mid x(n)^2 \mid))], \tag{10}$$

где DCT – дискретно-косинусное преобразование, X – спектральное представление сигнала $x(n)$. Понятие кепстра позволяет реализовывать преимущества спектрального представления сигнала. При преобразовании сигнала из временной области в частотную происходит

сжатие информации, которая становится более наглядной, подробной и компактной – в виде кепстра. Мел шкала – это шкала частотной восприимчивости изменений высоты звука. Мел – психофизическая единица высоты звука. Высота звука связана главным образом с частотой колебаний. По этой причине люди гораздо лучше воспринимают небольшие изменения звука на низких частотах, чем на высоких. Т.е. мел-шкала моделирует частотную чувствительность человеческого слуха.

Перевод из шкалы герц в шкалу мелов происходит по следующей формуле:

$$m(f) = 1125 \ln(1 + f/700), \tag{11}$$

где m – частота в мелах, f – частота в герцах. Начальным этапом вычисления МЧКК является спектральное представление фрагментов ЭМ по формуле:

$$Y_{i,s+1}(k) = \sum_{l=1}^{L} y_{i,s+1}(s \cdot L + l) \cdot e^{\frac{2\pi j}{L} kl}. \tag{12}$$

где $Y_{i,s+1}(k)$ – спектр фрагмента сигнала i-й ЭМ, $y_{i,s+1}$ – фрагмент сигнала i-й ЭМ, $0 \leqslant k < N$–1 – количество комплексных амплитуд синусоидальных сигналов, составляющих исходный сигнал, $l = (1, 2, \ldots L)$ – номер отсчета фрагмента сигнала ЭМ, L – количество отсчетов во фрагменте, j – мнимая единица.

Значение k определяет частоты, составляющие сигнал:

$$f_k = \frac{F_k}{N} \cdot k. \tag{13}$$

где F_s – частота дискретизации сигнала. Следующим этапом вычисления МЧКК является получение периодограммы фрагментов ЭМ по формуле:

$$P_{i,s+1}(k) = \frac{1}{L} \mid Y_{i,s+1}(k) \mid^2, \tag{14}$$

где $P_{i,s+1}(k)$ – периодограмма фрагмента i-й ЭМ.

Полученные периодограммы фрагментов содержат избыточное количество информации о частотах для задачи распознавания. По этой причине для более компактного представления информации периодограммы делятся на частотные диапазоны. К каждому диапазону применяется треугольная оконная функция – мел-фильтр, позволяющая просуммировать количество энергии каждого частотного диапазона периодограммы и определить мел-коэффициенты. Формирование набора мел-фильтров осуществляется по следующей методике [6, 10]:

- задается количество мел-фильтров G, нижняя f_l и верхняя f_h границы диапазона частот, в котором будет применяться фильтрация;

- выполняется преобразование границ диапазона из герц в мел (m_l, m_h);

- на мел-шкале отрезок $[m_l, m_h]$ разбивается на $G + 1$ непересекающихся подотрезки длиной $len = \frac{m_h - m_l}{G+1}$;

- определяются центральные частоты подотрезков по следующей формуле:

$$m_c g = m_i + g \cdot len. \tag{15}$$

где $g = 1, 2, \ldots, G$ – номер фильтра;

- центральные частоты переводятся в герцы $f_c(g)$ по следующей формуле (они соответствуют центральным частотам треугольных мел-фильтров):

$$f_{smp}(g) = \frac{L}{F_s} \cdot f_c(g), \tag{16}$$

где $f_{smp}(g)$ – частоты треугольных фильтров в дискретных отсчетах;

- для каждого мел-фильтра отсчеты периодограммы $P_{i,s+1}(k)$ умножаются на соответствующий фильтр:

$$MFCC_{i,s+1}(g) = \sum_{k+1}^{K} P_{i,s+1}(k) \cdot H_g(k), \tag{17}$$

$$H_g(k) = \begin{cases} 0, & k < f_{smp}(g-1), \\ \frac{k - f_{smp}(g-1)}{f_{smp}(g) - f_{smp}(g-1)} & f_{smp}(g-1) \leqslant k \leqslant f_{smp}(g), \\ \frac{f_{smp}(g+1) - k}{f_{smp}(g+1) - f_{smp}(g)} & f_{smp}(g) \leqslant k \leqslant f_{smp}(g+1), \\ 0 & k > f_{smp}(g+1). \end{cases} \tag{18}$$

где $g = 1, 2, \ldots, K$ – количество отсчетов в одном фрагменте. После выбора треугольных фильтров проводится логарифмирование энергии по следующей формуле:

$$MFCC_{i,s+1}(g) = \ln(MFCC_{i,s+1}(g)), \tag{19}$$

Последним этапом является вычисление дискретно-косинусного преобразования логарифма энергии набора фильтров. Так как все полосы пропускания фильтров перекрываются, энергии в наборе фильтров коррелируют друг с другом, поэтому необходимо провести декорреляцию по следующей формуле:

$$MFCC_{i,s+1}(c) = \sum_{g=1}^{G} MFCC_{i,s+1}(g) \cdot \cos(c(g - \frac{1}{2})\frac{\pi}{G}), \tag{20}$$

где $c = 1, 2, \ldots, C$ – номер МЧКК, C – желаемое количество МЧКК. Обычно для распознавания используют 12–15 МЧКК, так как чем выше индекс коэффициента, тем быстрее изменяется энергия в наборе фильтров. В результате экспериментальных исследований выяснилось, что первый МЧКК в основном несет информацию об интенсивности речевых сигналов [10]. В СГУ регистрация речевых сигналов может происходить с разными уровнями, поэтому информация первого МЧКК становится избыточной. В разработанном алгоритме в дальнейшем анализе первый МЧКК не используется.

Блок 7. Операция нормализации используется для придания равнозначности каждому МЧКК во фрагменте. Как известно, высокие частоты менее восприимчивы и МЧКК на этих частотах менее важны по сравнению с МЧКК на низких частотах. МЧКК на высоких частотах практически не влияют на результат [10]. Нормализация МЧКК – это умножение каждого коэффициента на число, которое увеличивается с номером коэффициента. Таким образом, первые коэффициенты по уровню уменьшаются, а последние коэффициенты увеличиваются. Для этой операции используется следующая формула [6, 7]:

$$MFCCN_{i,s+1}(c) = MFCC_{i,s+1}(c) \cdot (1 + \frac{Lf}{2}) \sin \frac{\pi c}{2})), \tag{21}$$

где Lf – величина, подбираемая эмпирически и равна 22 [6, 10].

Блок 8. Вычисление первого и второго приращений значений МЧКК позволяет получить динамическую информацию о коэффициентах. Вектор коэффициентов описывает фиксированную спектральную огибающую одного фрагмента, но очевидно, что речевые сигналы несут информацию и о динамике в виде незначительного изменения коэффициентов с течением времени [6, 7]:

$$MFCCD_{i,s+1}(c) = \frac{\sum_{d=1}^{d} d(MFCC_{i,s+1}(c+d) - (MFCC_{i,s+1}(c-d))}{2\sum_{d=1}^{d} d^2}, \tag{22}$$

$$MFCCDD_{i,s+1}(c) = \frac{\sum_{d=1}^{d} d(MFCCD_{i,s+1}(c+d) - (MFCCD_{i,s+1}(c-d))}{2\sum_{d=1}^{d} d^2}, \tag{23}$$

где $MFCCDi, s+1(c), MFCCDDi, s+1(c)$ – первое и второе приращения МЧКК, $MFCCi, s+1(c)$ – статические МЧКК, D – типовое значение приращения, равное 2 [6, 10].

Блоки 9, 10. Формирование базы данных шаблонов и набора данных МЧКК представляет собой объединение МЧКК (первичных, нормализованных и после приращения) в один вектор.

Блок 11. Распознавание представляет процесс сравнения поступившей в систему речевой команды с шаблоном из базы данных, полученным в ходе обучения алгоритма. Одна речевая команда может быть произнесена по-разному, так как различные части слова произносятся с разной скоростью. Для определения расхождения между поступающей в систему речевой командой и шаблоном, представленными как векторы МЧКК, должно быть выполнено выравнивание по времени. С этой целью для распознавания применяется метод динамического трансформирования времени [4], который является методикой эластичного сравнения сигнала речевой команды и шаблона в регулярных интервалах – фрагментах.

Процесс сравнения векторов МЧКК поступающей речевой команды с шаблоном начинается с расчета локальных отклонений между значениями двух векторов. В разработанном алгоритме применяются самые распространенные способы вычисления отклонений [7]:

- определение коэффициента корреляции по формуле:

$$r(x_{i,s+1}, y_{i,s+1}) = \frac{\overline{x_{i,s+1}, y_{i,s+1}} - \overline{x_{i,s+1}} * \overline{y_{i,s+1}}}{\sigma(x_{i,s+1}) * \sigma(x_{i,s+1})}, \tag{24}$$

где $r(x_{i,s+1}, y_{i,s+1})$ – элементы матрицы отклонения, x_{s+1} – вектор МЧКК фрагмента поступающей речевой команды, y_{s+1} – вектор МЧКК фрагментов шаблона, $s = 0, 1, 2, \ldots, S$–1 – номер фрагмента;

- вычисление евклидового расстояния по формуле:

$$d(x_{i,s+1}, y_{i,s+1}) = \sqrt{\sum_{J=1}^{J} (x_{i,s+1} - y_{i,s+1})^2}, \tag{25}$$

где $d(x_{i,s+1}, y_{i,s+1})$ – евклидово расстояние. Использование двух способов вычисления отклонения для определения оценки расхождения повысит точность распознавания. Результатом сравнения будет вектор, для которого было найдено минимальное расхождение между поступившей речевой командой и шаблоном. Далее вычисляется минимальная глобальная оценка расхождения (МГОР) для маршрута как сумма локальных расстояний между фрагментами речевой команды и шаблона.

Блок 12. После выполнения распознавания всех информативных сигналов, полученных из исходной речевой команды, осуществляется выбор наиболее приемлемого информативного сигнала, обеспечивающего минимальную оценку расхождения с шаблоном. Таким образом, алгоритм автоматически определяет, какие ЭМ стоит вычитать для каждой речевой команды, чтобы добиться минимальной разницы с шаблоном и максимальной точности распознавания.

4. Исследование алгоритма распознавания речевых команд

Исследование предложенного алгоритма распознавания речевых команд проводилось в экспериментально-исследовательском комплексе, реализованном в пакете прикладных программ $MATLAB$. Цель исследования: определение наиболее приемлемого информативного сигнала, обеспечивающего наименьшую разницу между поступающей в систему речевой командой и шаблоном; сравнение точности распознавания предложенного усовершенствованного и известных алгоритмов. Экспериментальное исследование проводилось с использованием разработанной базы данных речевых сигналов [11]. В качестве критериев оценки эффективности распознавания были выбраны

- точность распознавания:

$$A = \frac{SC_{true}}{SC_{total}} \cdot 100\%, \tag{26}$$

где A – точность распознавания, SC_{true} – правильно распознанные речевые сигналы, SC_{total} – общее количество речевых сигналов;

- разница МГОР между истинным и максимально близким к истинному распознаваниями:

$$\Delta = MGED_{sim.} - MGED_{appr.}, \tag{27}$$

где $MGED_{sim.}$ – МГОР истинного распознавания, $MGED_{appr.}$ – максимально близкого к истинному распознаванию.

Исходные данные для исследования: обучающая и тестовая непересекающиеся выборки речевых сигналов длительностью от 10 до 3000 мс, частота дискретизации 8000 Гц, разрядность квантования 16 бит. Обучающая выборка сформирована из 1000 чистых речевых сигналов, произнесенных 50-ю людьми (мужчинами и женщинами). Тестовая выборка сформирована из 60-ти речевых сигналов – 20 различных звуков по 3 произношения каждый. Настройки аппарата КМДЭМ: уровень амплитуды добавляемого белого шума – 0,1 мВ, количество циклов декомпозиции – 100. В табл. 1 представлены результаты определения наиболее приемлемого информативного сигнала, обеспечивающего наилучшую точность распознавания и большее значение разницы МГОР.

Таблица 1

Зависимость А и Δ от номера информативного сигнала

Параметр	Исходный речевой сигнал	Номер информативного сигнала				
		1	2	3	4	5
А%	85,1	92,3	95,2	93,9	97,1	96,2
Δ	2,47	3,88	5,22	4,55	5,84	5,74

В соответствии с полученными результатами наилучшая точность распознавания и большее значение МГОР достигается при использовании информативного сигнала № 4. Наглядно это представлено на рис. 3.

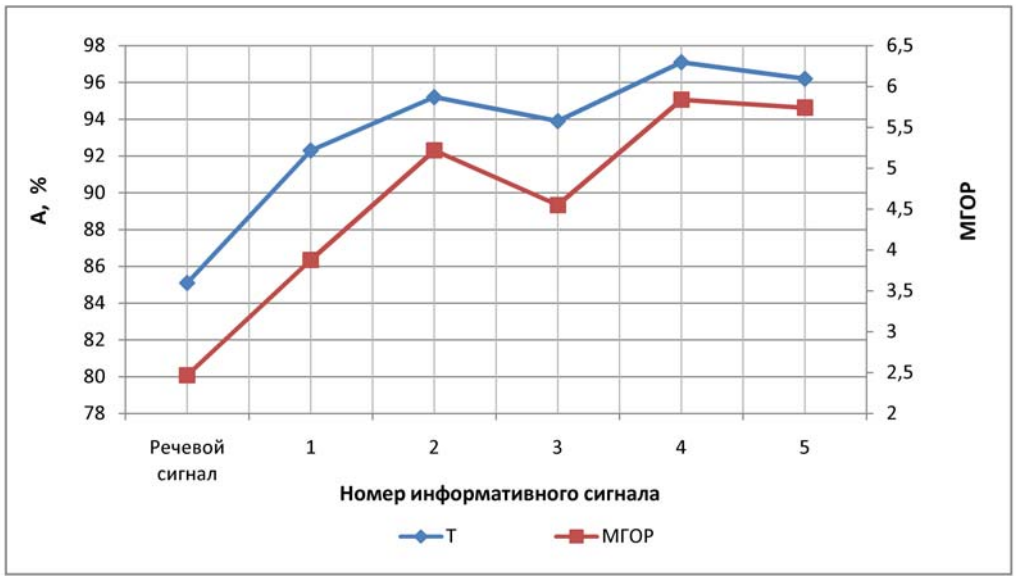
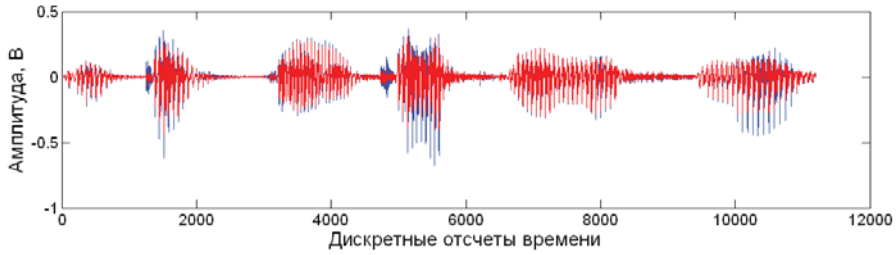
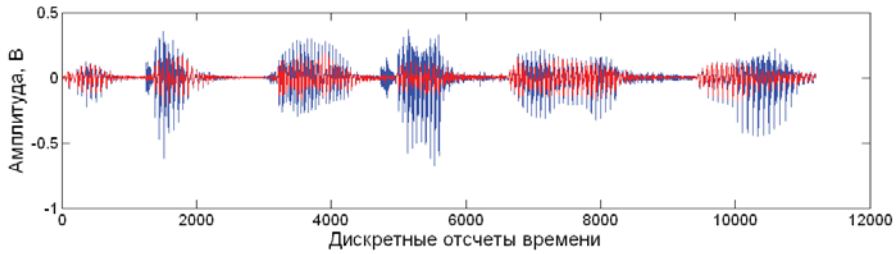


Рис. 3. Зависимость А и Δ от номера информативного сигнала

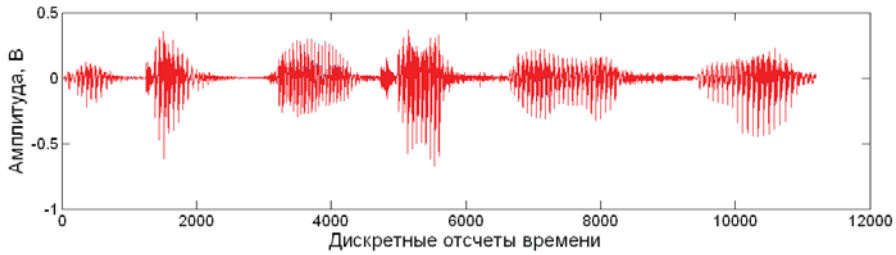
На рис. 4а – 4д представлены осциллограммы пяти информативных сигналов.



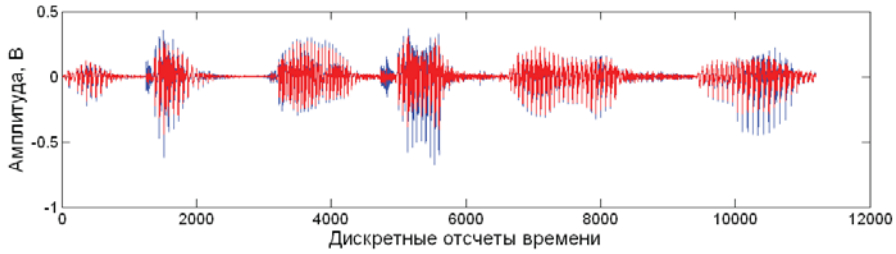
а. Информативный сигнал № 1



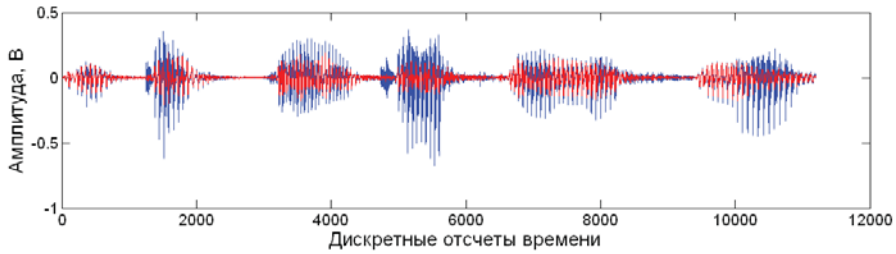
б. Информативный сигнал № 2



в. Информативный сигнал № 3



г. Информативный сигнал № 4



д. Информативный сигнал № 5

Рис. 4. Осциллограммы пяти информативных сигналов: синий цвет – исходный речевой сигнал, красный цвет – информативный сигнал

Окончательные результаты экспериментального исследования предложенного алгоритма распознавания речевых команд оценивались в сравнении с алгоритмами $RWTHASR$, $Julius$ и $CMUSphinx$ в зависимости от входного значения отношения сигнал/шум. В табл. 2 и на рис. 5 представлен сравнительный анализ точности распознавания.

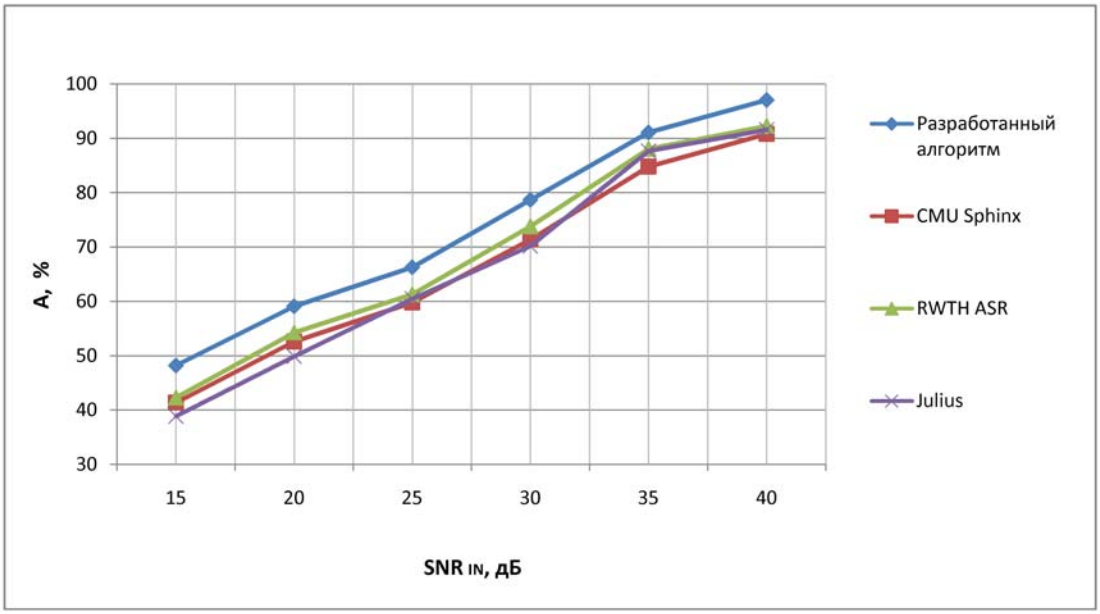


Рис. 5. Точность распознавания предложенного алгоритма распознавания и $RWTHASR$, $Julius$ и $CMUSphinx$

В соответствии с результатами, представленными в табл. 2 и на рис. 5 следует, что предложенный алгоритм распознавания эффективнее известных аналогов для всего диапазона значений входного отношения сигнал/шум:

- в среднем на 5,9 % лучше, чем метод $CMUSphinx$;
- в среднем на 10,7 % лучше, чем метод $RWTHASR$;
- в среднем на 7,7 % лучше, чем метод $Julius$.

Таблица 2

Точность распознавания предложенного алгоритма и $RWTHASR$, $Julius$ и $CMUSphinx$

Входное значение отношения сигнал/шум SNR_{IN}, дБ	A, %			
	«CMU Sphinx»	«RWTH ASR»	«Julius»	Разработанный алгоритм
15	44,1	41,4	42,3	48,2
20	55,8	52,6	54,3	59,1
25	62,9	59,8	61,3	66,3
30	74,7	71,4	73,8	78,7
35	87,2	84,8	88,1	91,1
40	92,7	90,8	92,3	97,1

Таким образом, использование разработанного алгоритма распознавания речевых команд в СГУ, усовершенствованного за счет применения КМДЭМ на этапе предварительной обработки, позволит значительно повысить эффективность голосового управления.

5. Выводы

В статье предложен алгоритм, обеспечивающий повышение точности распознавания на основе метода КМДЭМ. Результаты экспериментального исследования демонстрируют, что предложенный алгоритм распознавания точнее известных аналогов и позволит значительно повысить эффективность голосового управления. Важно отметить, что предложенный алгоритм распознавания более длителен, чем $RWTHASR$, $Julius$ и $CMUSphinx$. Это связано с особенностью комплементарной декомпозиции. Поэтому в зависимости от важности задач – быстродействие или эффективность голосового управления – необходимо искать компромисс. С целью повышения быстродействия в дальнейшем актуальным является оптимизация предложенного алгоритма.

Литература

1. *David R., Christian G., Georg H., Hermann N.* The RWTH Aachen University Open Source Speech Recognition System. Human Language Technology and Pattern Recognition Computer Science Department. RWTH Aachen University. Germany. 4 p.

2. *Lee A., Kawahara T., Shikano K.* Julius – an open source real-time large vocabulary recognition engine // Proc. European Conf. on Speech Communication and Technology. Aalborg. Denmark. Sep. 2001. P. 1691–1694.

3. *Walker W., Lamere P., Kwok P., Bhiksha Raj R.S., Gouvea E., Wolf P., Woelfel J.* Sphinx-4: A fexible open source framework for speech recognition. Sun Microsystems. Inc, Tech. Rep. SMLI TR-2004-139. Nov. 2004. 15 p.

4. *Алимурадов А.К., Чураков П.П.* Обзор и классификация методов обработки речевых сигналов в системах распознавания речи // Измерение. Мониторинг. Управление. Контроль. 2015. № 2 (12). С. 27–35.

5. *Yeh J.-R., Shieh J.-S., Huang N.E.* Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. Advances in Adaptive Data Analysis. 2010. V. 2 (2). P. 135–156.

6. *Алимурадов А.К., Чураков П.П.* Адаптивный метод повышения эффективности голосового управления // Перспективные информационные технологии (ПИТ 2016): труды Международной научно-технической конференции / под ред. С.А. Прохорова. Самара: Издательство Самарского научного центра РАН, 2016. С. 196–200.

7. *Алимурадов А.К., Муртазов Ф.Ш.* Методы повышения эффективности распознавания речевых сигналов в системах голосового управления // Измерительная техника. 2015. № 10. С. 20–24.

8. *Алимурадов А.К.* Оптимальный алгоритм обработки речевых команд для системы голосового управления // Модели, системы, сети в экономике, технике, природе и обществе. 2015. № 2 (14). С. 139–149.

9. *Клионский Д.М., Неунывакин И.В., Орешко Н.И., Геппенер В.В.* Декомпозиция на эмпирические моды и ее применение для идентификации информативных компонент и прогнозирования значений сигналов с использованием нейронных сетей // Нейрокомпьютеры. 2010. № 6. С. 69–80.

10. *Huang X., Acero A., Hon H.-W.* Spoken Language Processing. Guide to Algorithms and System Developmen. Prentice Hall, 2001. 980 p.

11. Свидетельство о государственной регистрации базы данных № 2016620597. Верифицированная база речевых команд для систем голосового управления / А.К. Алимурадов // Программы для ЭВМ, базы данных, топологии интегральных микросхем; заявл. 16.03.2016; опубл. 12.05.2016.

References

1. *David R., Christian G., Georg H., Hermann N.* The RWTH Aachen University Open Source Speech Recognition System. Human Language Technology and Pattern Recognition Computer Science Department. RWTH Aachen University. Germany. 4 p.

2. *Lee A., Kawahara T., Shikano K.* Julius – an open source real-time large vocabulary recognition engine. Proc. European Conf. on Speech Communication and Technology. Aalborg. Denmark. Sep. 2001. P. 1691–1694.

3. *Walker W., Lamere P., Kwok P., Bhiksha Raj R.S., Gouvea E., Wolf P., Woelfel J.* Sphinx-4: A fexible open source framework for speech recognition. Sun Microsystems. Inc, Tech. Rep. SMLI TR-2004-139. Nov. 2004. 15 p.

4. *Alimuradov A.K., Churakov P.P.* Review and classification of processing methods of speech signals in speech recognition systems. Measuring. Monitoring. Management. Control. 2015. V. 2 (12). P. 27–35.

5. *Yeh, J.-R., Shieh, J.-S., Huang N.E.* Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. Advances in Adaptive Data Analysis. 2010. V. 2 (2). P. 135–156.

6. *Alimuradov A.K., Churakov P.P.* An adaptive method for voice control efficiency increase. Advanced Information Technologies (AIT 2016): Proceedings of International Scientific Conference. Samara: Samara Publishing Scientific Centre. 2016. P. 196–200.

7. *Alimuradov A.K., Murtazov F.Sh.* Methods to improve the efficiency of recognition of speech signals in voice control systems. Measurement Techniques. 2015. N 10. P. 20–24.

8. *Alimuradov A.K.* Optimal algorithms of processing voice commands for voice control. Models. Systems. Networks in the Economics, Technology, Nature and Society. 2015. V. 2 (14). P. 139–149.

9. *Klionsky D.M., Neunyvakin I.V., Oreshko N.I., Geppener V.V.* Ensemble empirical mode decomposition and its application to identify the informative components and predict signal values using neural networks. Neurocomputers. 2010. N 6, 2010. P. 69–80.

10. *Huang X., Acero A., Hon H.-W.* Spoken Language Processing. Guide to Algorithms and System Developmen. Prentice Hall, 2001. 980 p.

11. *Alimuradov A.K.* State Database Registration Certificate N 2016620597. Verified speech signal database for voice control systems. Computer programs, databases, topographies of integrated microcircuits. Published 12 May 2016.

Поступила в редакцию 03.05.2017