

УДК 004.934

*А. К. Алимуратов***АЛГОРИТМ ПОМЕХОУСТОЙЧИВОЙ ОБРАБОТКИ РЕЧЕВЫХ КОМАНД ДЛЯ СИСТЕМ ГОЛОСОВОГО УПРАВЛЕНИЯ¹***А. К. Alimuradov***NOISE-ROBUST PROCESSING ALGORITHM OF VOICE COMMANDS FOR VOICE CONTROL SYSTEMS**

А н н о т а ц и я. Актуальность и цели. Повседневный ритм окружающей среды требует от систем голосового управления (СГУ) возможности функционировать в условиях интенсивных помех. Из-за шумов разница между зарегистрированными речевыми командами и шаблонами (полученными в процессе обучения системы) возрастает, а эффективность СГУ стремительно ухудшается. По этой причине актуальным является повышение помехоустойчивости СГУ. **Материалы и методы.** Для реализации помехоустойчивой обработки предложена модификация классической декомпозиции на эмпирические моды (ДЭМ) – комплементарная множественная ДЭМ. Для проведения исследований использовался пакет прикладных программ решения технических задач – *MATLAB*. **Результаты.** Представлен краткий обзор существующих алгоритмов помехоустойчивой обработки. На основе метода комплементарной множественной ДЭМ разработан алгоритм помехоустойчивой обработки речевых команд для СГУ. Проведены исследования предложенного алгоритма, результаты которых подтверждают повышение помехоустойчивости в широком диапазоне помех (от 20 до –5 дБ) в среднем на 3,1 %. **Выводы.** В соответствии с результатами исследований предложенный алгоритм помехоустойчивой обработки речевых команд рекомендуется для практического применения в СГУ, функционирующих в условиях интенсивных помех.

A b s t r a c t. Background. A daily rhythm of the environment requires a voice control system (VCS) to function in conditions of intense interference. Because of noise, a difference between the registered voice commands and templates (obtained during the system training) increases, and the efficiency of the VCS is rapidly deteriorating. For this reason, to improve the VCS noise immunity is urgent. **Materials and methods.** To implement a noise-robust processing, a modification of classical Empirical Mode Decomposition (EMD) – the Complementary Ensemble EMD (CEEMD), is proposed. To conduct the research, a software package *MATLAB* for solving technical problems was used. **Results.** A brief overview of the current noise-robust processing algorithms is presented. A noise-robust processing algorithm of voice commands for voice control systems is developed on the basis of the CEEMD method. The investigations of the proposed algorithm are carried out, the results of which have confirmed noise immunity improvement on average by 3,1 % in a wide range of noise (from 20 to –5 dB). **Conclusions.** In accordance with the research results, the proposed noise-robust processing algorithm of voice commands is recommended for practical application in VCS functioning in conditions of intense interference.

¹ Статья выполнена при финансовой поддержке РФФИ в рамках научного проекта № 16-31-00194 мол_а «Определение биомаркеров речевых и ЭЭГ сигналов и исследование их взаимосвязи для экспресс-диагностики психогенных состояний».

К л ю ч е в ы е с л о в а: обработка речевых сигналов, система голосового управления, комплементарная множественная декомпозиция на эмпирические моды, отношение сигнал/шум, интенсивные помехи

К е у w o r d s: speech signal processing, voice control system, the Complementary Ensemble Empirical Mode Decomposition, signal-to-noise ratio, intense noise

Введение

Современные системы голосового управления (СГУ) могут обеспечить высокую точность распознавания при невысоком уровне фонового шума. В ситуации, когда разница между зарегистрированными речевыми командами и шаблонами (полученными в процессе обучения СГУ) из-за посторонних шумов становится значительной, эффективность СГУ стремительно ухудшается. Повседневный ритм окружающей среды требует от СГУ возможности функционировать в условиях интенсивных помех. По этой причине актуальным является повышение помехоустойчивости СГУ.

На сегодня известны различные алгоритмы помехоустойчивой обработки:

- алгоритмы повышения четкости (корректировки) речи, хорошо известные как методы вычитания спектра [1];
- алгоритмы на основе фильтров Винера [2], используемые в предварительной обработке для удаления шумовой составляющей речевого сигнала;
- алгоритмы постобработки, основанные на кепстральном среднем вычитании (*Cepstral Mean Subtraction, CMS*) [3], нормализации кепстральных отклонений (*Cepstral Variance Normalization, CVN*) [4] и выравнивании гистограммы (*Histogram Equalization, HEQ*) [5], направленных на преобразование необработанных информативных параметров речи в форму, которая в меньшей степени подвержена искажениям в неблагоприятной окружающей среде.

В данной работе проводится исследование алгоритма помехоустойчивой обработки, основанного на комплементарной множественной декомпозиции на эмпирические моды (КМДЭМ) [6]. Статья является продолжением ранее опубликованных работ авторов [7–9].

Комплементарная множественная декомпозиция на эмпирические моды

Математический аппарат КМДЭМ основан на методе декомпозиции на эмпирические моды (ДЭМ) [10], особенность которого заключается в том, что базисные функции, используемые для разложения, извлекаются непосредственно из исходного сигнала. При разложении модель сигнала не задается заранее, эмпирические моды (ЭМ) вычисляются в ходе процедуры отсеивания с учетом локальных особенностей (таких как экстремумы и нули сигнала) и внутренней структуры каждого конкретного сигнала. Таким образом, ЭМ не имеют строгого аналитического описания, но должны удовлетворять двум условиям, гарантирующим определенную симметрию и узкополосность базисных функций [10]:

- общее число экстремумов равняется общему числу нулей с точностью до единицы;
- среднее значение двух огибающих – верхней, интерполирующей локальные максимумы, и нижней, интерполирующей локальные минимумы, должно быть приближенно равно нулю.

В результате ДЭМ из исходного сигнала $x(n)$ извлекается конечное число ЭМ и результирующий остаток:

$$x(n) = \sum_{i=1}^{I-1} IMF_i(n) + r_1(n), \quad (1)$$

где $IMF_i(n)$ – полученные после разложения ЭМ; $r_1(n)$ – остаток разложения; $i = 1, 2, \dots, I$ – номер ЭМ; n – дискретный отсчет времени.

При всех указанных преимуществах адаптивности метода ДЭМ недостатком, важным для его практического применения, является смешивание ЭМ, состоящих из различных участков сигнала несоизмеримых по частотному и амплитудному масштабам.

Для решения данной проблемы был предложен новый метод декомпозиции, основанный на многократном добавлении к сигналу белого шума с прямыми и инверсными значениями амплитуды и вычислении среднего значения ЭМ и остатка как конечного истинного результата:

$$\begin{bmatrix} y_j(n) \\ y_j(n)^* \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} x(n) \\ w_j(n) \end{bmatrix}, \quad (2)$$

где $w_j(n)$ – добавленный белый шум; $y_j(n)$ – сумма зашумленного речевого сигнала $x(n)$ с белым шумом; $y_j(n)^*$ – сумма зашумленного речевого сигнала $x(n)$ с инверсным значением амплитуды белого шума:

$$IMF_i(n) = \frac{\sum_{j=1}^J IMF_{ji}(n)}{J}, \quad (3)$$

$$r_l(n) = \frac{\sum_{j=1}^J r_{jl}(n)}{J}, \quad (4)$$

где $IMF_{ji}(n)$, $r_{jl}(n)$ – ЭМ и остаток, полученные при различных декомпозициях сигналов $y_j(n)$ и $y_j(n)^*$; $j = 1, 2, \dots, J$ – количество циклов декомпозиций (добавлений к сигналу белого шума).

Белый шум позволяет получить ЭМ, сопоставимые с участками сигнала несоизмеримых масштабов. Вычисление среднего значения полученных ЭМ как конечного истинного результата обеспечит полное удаление остаточного белого шума за счет пары прямых и инверсных значений, независимо от того, сколько сигналов шума использовалось.

Следует отметить общие правила для корректной работы метода КМДЭМ:

1. При добавлении белого шума может возникнуть эффект маскировки высокочастотной составляющей, если количество циклов декомпозиции не будет достаточно большим.
2. Уровень среднеквадратического отклонения амплитуды добавляемого белого шума должен быть на уровне присутствующего в сигнале шума (если уровень известен), в противном случае шум не должен быть больше, чем 20 % от исходного сигнала.
3. Важными исходными параметрами для КМДЭМ являются амплитудный (или энергетический) уровень добавляемого белого шума и количество циклов декомпозиции.

Алгоритм помехоустойчивой обработки

На основании вышеупомянутого и собственных исследований [7–9] разработан алгоритм помехоустойчивой обработки речевых команд для СГУ (рис. 1). Штриховой линией отмечен режим обучения алгоритма, сплошной линией – режим распознавания. Серым цветом отмечен блок 1 помехоустойчивой обработки на основе КМДЭМ. Суть обработки заключается в уменьшении разницы между поступающими в систему речевыми командами и шаблонами, полученными в процессе обучения СГУ. С помощью КМДЭМ исходная зарегистрированная речевая команда (как функции времени) разлагается на ЭМ. Основной идеей предлагаемого алгоритма является то, что ЭМ низкого порядка содержат высокочастотные шумовые составляющие, которые могут быть удалены для уменьшения разницы:

$$x'(n) = x(n) - \sum_{i=1}^{I-1} IMF_i(n), \quad (5)$$

где $x'(n)$ – отфильтрованный речевой сигнал.

На рис. 2 представлены графики, иллюстрирующие уменьшение разницы между поступающими в систему речевыми командами и шаблонами, полученными в процессе обучения. Используя верифицированную базу данных [11], из чистого речевого сигнала посредством добавления шума получили зашумленные сигналы с различным отношением сигнал/шум (*Signal Noise-to-Ratio, SNR*) от –5 до 20 дБ. Уровень зашумленности речевых сигналов хорошо иллюстрируют зависимости логарифма энергии сигнала от времени. Из рис. 2,а видно, что степень осцилляции логарифма энергии зашумленных сигналов возрастает с увеличением шума. То есть источником паразитных амплитудных всплесков является шумовая составляющая сигнала.

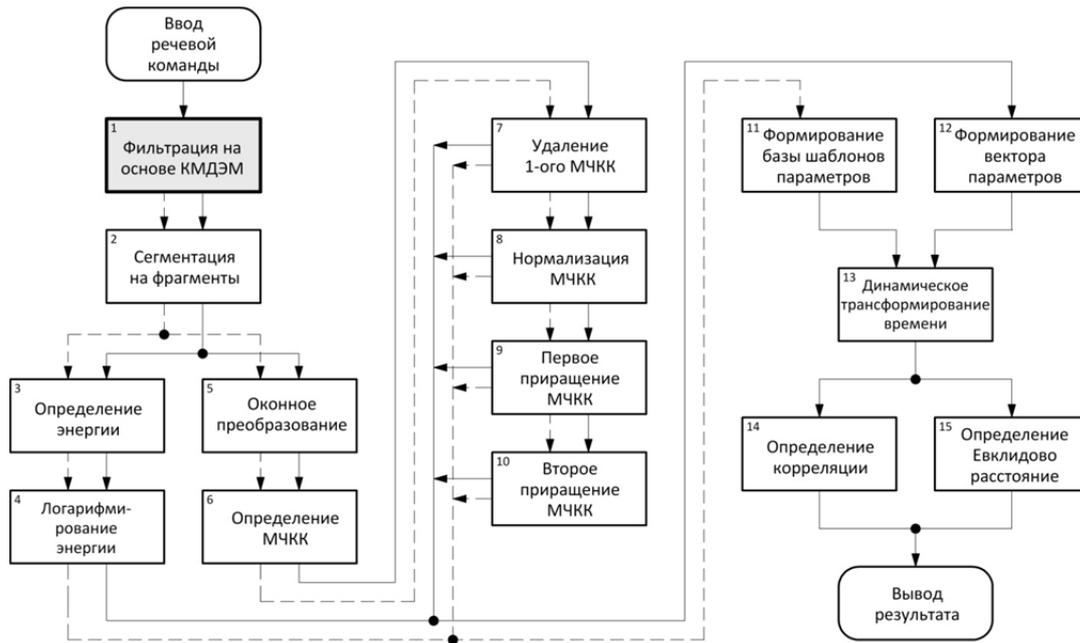


Рис. 1. Алгоритм помехоустойчивой обработки речевых команд для СГУ (МЧКК – мел-частотные кепстральные коэффициенты)

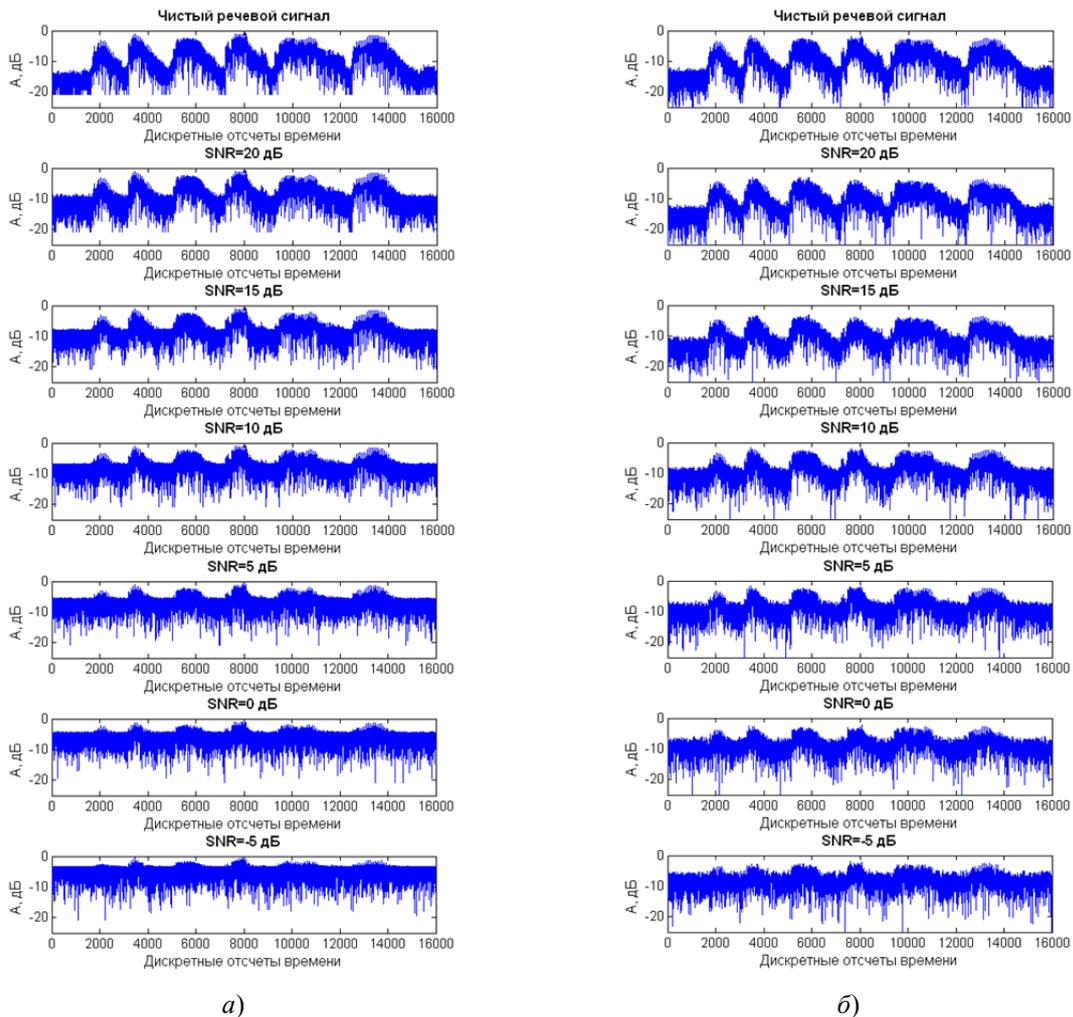


Рис. 2. Зависимости логарифма энергии речевого сигнала от времени: а – исходный речевой сигнал; б – отфильтрованный речевой сигнал

Сравнивая кривые зависимости логарифма энергии (рис. 2,б), видим, что разница между чистым речевым сигналом и зашумленными стала меньше.

Исследование алгоритма помехоустойчивой обработки

Исследование алгоритма помехоустойчивой обработки проводилось в пакете прикладных программ *MATLAB* с использованием верифицированной базы данных речевых сигналов [12]. В качестве критерия оценки эффективности использовалось значение точности распознавания – безразмерная величина, равная отношению правильно распознанных речевых команд к их общему количеству:

$$A = \frac{C_{true}}{C_{total}} 100 \%, \quad (6)$$

где A – точность распознавания; C_{true} – правильно распознанные команды; C_{total} – общее количество команд.

СГУ была обучена различными выборками, состоящими из чистых и нескольких видов зашумленных речевых сигналов. Выборка А – 90 % тестовых речевых сигналов пересекаются с обучающей выборкой. Выборка Б – 10 % тестовых речевых сигналов пересекаются с обучающей выборкой. Выборка В – ни один тестовый речевой сигнал не пересекается с обучающей выборкой.

В качестве информативных параметров для распознавания использовались статические первые 12 мел-частотных кепстральных коэффициента (МЧКК) и значения логарифмов энергии сегментированных фрагментов речевых команд. Дополнительно использовались динамические характеристики МЧКК – первое и второе приращения.

Результаты исследования оценивались в сравнении с алгоритмом помехоустойчивой обработки, получившим широкую практическую популярность, на основе нормализованного среднего отклонения (*Mean-Variance Normalized, MVN*).

В табл. 1 представлены результаты точности распознавания предлагаемого алгоритма помехоустойчивой обработки на основе КМДЭМ и алгоритма на основе *MVN* для трех выборок тестовых сигналов.

Таблица 1

Результаты точности распознавания

Тестовый сигнал	Точность распознавания, %							
	Выборка А		Выборка Б		Выборка В		Среднее значение	
	<i>MVN</i>	КМДЭМ	<i>MVN</i>	КМДЭМ	<i>MVN</i>	КМДЭМ	<i>MVN</i>	КМДЭМ
Чистый речевой сигнал	98,8	98,2	94,6	94,2	88,3	89,1	93,9	93,8
20 дБ	96,5	96,2	91,9	92,8	87,1	88,6	91,8	92,5
15 дБ	91,8	92,4	86,8	89,7	83,2	87,2	87,3	89,8
10 дБ	82,3	84,9	75,5	80,3	71,9	75,6	76,6	80,3
5 дБ	61,7	66,3	56,8	62,5	49,9	55,2	56,1	61,3
0 дБ	32,5	37,2	28,7	33,1	20,2	25,7	27,1	32,0
-5 дБ	13,4	16,9	11,1	15,6	9,2	13,7	11,2	15,4

Как видно из таблицы, предлагаемый алгоритм обеспечивает наилучшую точность распознавания для выборки тестовых сигналов В. Для выборок А и Б превосходство предлагаемого алгоритма отмечается только для зашумленных сигналов. А наименьшая разница между поступающими в систему речевыми командами и шаблонами, полученными в процессе обучения, для чистых речевых сигналов достигается алгоритмом на основе *MVN*.

По результатам исследования отметим факты превосходства предлагаемого алгоритма, имеющие большое практическое значение:

- высокая точность распознавания зашумленных речевых сигналов;
- высокая точность распознавания сигналов тестовой выборки, не пересекающейся с обучающей выборкой (дикторонезависимость СГУ).

Таким образом, предлагаемый алгоритм обработки повышает помехоустойчивость в широком диапазоне помех (от 20 до -5 дБ) в среднем на 3,1 % за счет уменьшения разницы между зарегистрированными речевыми командами и шаблонами, полученными в процессе обучения системы.

Заключение

Предложен алгоритм помехоустойчивой обработки речевых команд для СГУ, основанный на КМДЭМ. Проведены исследования алгоритма, подтверждающие повышение помехоустойчивости в широком диапазоне помех (от 20 до -5 дБ) в среднем на 3,1 %. В соответствии с результатами исследований сделан вывод, что предложенный алгоритм рекомендуется для практического применения в СГУ, функционирующих в условиях интенсивных помех.

Библиографический список

1. Boll, S. Suppression of acoustic noise in speech using spectral subtraction / S. Boll // IEEE Trans. Acoust. Speech Signal Process. – 1979. – Vol. 27 (2). – P. 113–120. DOI: 10.1109/TASSP.1979.1163209.
2. Berstein, A. A hypothesized Wiener filtering approach to noisy speech recognition, in ICASSP / A. Berstein, I. Shallom. – Canada, Toronto, 1991. – P. 913–916.
3. Furui, S. Cepstral analysis technique for automatic speaker verification / S. Furui // IEEE Trans. Acoust. Speech Signal Process. – 1981. – Vol. 29 (2). – P. 254–272. DOI:10.1109/TASSP.1981.1163530.
4. Viikki, O. A recursive feature vector normalization approach for robust speech recognition in noise / O. Viikki, D. Bye, K. Laurila // Proceedings of the ICASSP. – USA, Washington, 1998. – P. 733–736.
5. A de La Torre, A. Rubio, Histogram equalization of speech representation for robust speech recognition / A de La Torre, A. Peinado, J. Segura, J. Perez-Cordoba, M. Benitez // IEEE Trans. Speech Audio Process. – 2005. – Vol. 13 (3). – P. 355–366.
6. Zhaohua, Wu. Ensemble empirical mode decomposition: A noise – assisted data analysis method / Wu Zhaohua, N. E. Huang // Advances in Adaptive Data Analysis. – 2009. – Vol. 1 (1). – P. 1–41.
7. Алимуратов, А. К. Оптимальный алгоритм обработки речевых команд для системы голосового управления / А. К. Алимуратов // Модели, системы, сети в экономике, технике, природе и обществе. – 2015. – № 2 (14). – С. 139–149.
8. Алимуратов, А. К. Исследование оптимального алгоритма обработки речевых сигналов для системы голосового управления / А. К. Алимуратов // Модели, системы, сети в экономике, технике, природе и обществе. – 2015. – № 4 (16). – С. 120–125.
9. Алимуратов, А. К. Применение методов декомпозиции на эмпирические моды в задаче фильтрации речевых сигналов в условиях интенсивных помех / А. К. Алимуратов, П. П. Чураков // Измерение. Мониторинг. Управление. Контроль. – 2016. – № 1 (15). – С. 4–14.
10. Huang, N. E. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis / N. E. Huang, Shen Zheng, R. L. Steven // Proceedings of the Royal Society of London A. – 1998. – Vol. 454. – P. 903–995.
11. Свидетельство о государственной регистрации базы данных № 2016620597. Верифицированная база речевых команд для систем голосового управления / Алимуратов А. К. // Программы для ЭВМ, базы данных, топологии интегральных микросхем. – заявл. 16.03.2016 ; опубл. 12.05.2016.

Алимуратов Алан Казанферович

кандидат технических наук,
директор студенческого
научно-производственного бизнес-инкубатора,
инженер-исследователь
научно-исследовательского отдела,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: alansapfir@yandex.ru

Alimuradov Alan Kazanferovich

candidate of technical sciences,
director of student research
and production business incubator,
researcher of research department,
Penza State University
(40 Krasnaya street, Penza, Russia)

УДК 004.934

Алимурадов, А. К.

Алгоритм помехоустойчивой обработки речевых команд для систем голосового управления / А. К. Алимурадов // Измерение. Мониторинг. Управление. Контроль. – 2017. – № 1 (19). – С. 86–92.