

УДК 681.586  
doi:10.21685/2307-5538-2022-3-6

## ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА СИГНАЛА ПРИ РАСПОЗНАВАНИИ ГОЛОСОВЫХ КОМАНД МЕТОДОМ УЛУЧШЕННОЙ ПОЛНОЙ МНОЖЕСТВЕННОЙ ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ

В. В. Козлов<sup>1</sup>, Е. А. Фокина<sup>2</sup>, А. А. Трофимов<sup>3</sup>

<sup>1,3</sup> Пензенский государственный университет, Пенза, Россия

<sup>2</sup> Научно-исследовательский институт физических измерений, Пенза, Россия

<sup>1</sup>val369@mail.ru, <sup>2</sup>Ekaterina.isay1997@gmail.com, <sup>3</sup>iit@pnzgu.ru

**Аннотация.** *Актуальность и цели.* При распознавании речевых сигналов для работы в различных сферах жизни человека разработчику приходится решать проблемы обработки речевых сигналов, в частности проблему их нестационарности. Для решения этой проблемы существуют различные методы предварительной обработки, поэтому необходимо выбрать наилучший метод. *Материалы и методы.* Для предварительной обработки речевых сигналов был выбран наилучший метод, а именно улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом (УПМДЭМАШ). Проведено моделирование разложения речевых сигналов на составляющие с помощью УПМДЭМАШ, выделение наиболее информативной составляющей и перевод ее в частотную область с помощью преобразования Фурье. *Результаты.* Был проведен сравнительный анализ выделенных составляющих для разных команд, также был сделан вывод о правильности выбора метода и информативной составляющей.

**Ключевые слова:** речевые сигналы, предварительная обработка, декомпозиция на эмпирические моды, преобразование Фурье

**Для цитирования:** Козлов В. В., Фокина Е. А., Трофимов А. А. Предварительная обработка сигнала при распознавании голосовых команд методом улучшенной полной множественной декомпозиции на эмпирические моды // Измерения. Мониторинг. Управление. Контроль. 2022. № 3. С. 56–61. doi:10.21685/2307-5538-2022-3-6

## PRE-PROCESSING OF SIGNAL IN RECOGNITION OF VOICE COMMANDS BY METHOD OF IMPROVED COMPLETE MULTIPLE DECOMPOSITION TO EMPIRICAL MODES

V.V. Kozlov<sup>1</sup>, E.A. Fokina<sup>2</sup>, A.A. Trofimov<sup>3</sup>

<sup>1,3</sup> Penza State University, Penza, Russia

<sup>2</sup> Research and Development Institute for Physical Measurements, Penza, Russia

<sup>1</sup>val369@mail.ru, <sup>2</sup>Ekaterina.isay1997@gmail.com, <sup>3</sup>iit@pnzgu.ru

**Abstract.** *Background.* In the recognition of speech signals to work in various spheres of human life, the developer has to solve the problem of processing of speech signals, in particular the problem of non-stationarity. To solve this problem, there are various preprocessing methods and it is necessary to choose the best method. *Materials and methods.* For the preprocessing of the speech signals the best method was chosen, namely, the improved full multiple decomposition into empirical modes with adaptive noise (IFMDEMAN). We modeled the decomposition of speech signals into components using IFMDEMAN, extracted the most informative component and translated it into the frequency domain using the Fourier transform. *Results.* As a result of this work, a comparative analysis of the selected components for different teams, as well as a conclusion about the correctness of the choice of the method and the choice of the informative component.

**Keywords:** speech signals, preprocessing, decomposition into empirical modes, Fourier transform

**For citation:** Kozlov V.V., Fokina E.A., Trofimov A.A. Pre-processing of signal in recognition of voice commands by method of improved complete multiple decomposition to empirical modes. *Izmereniya. Monitoring. Upravlenie. Kontrol' = Measurements. Monitoring. Management. Control.* 2022;(3):56–61. (In Russ.). doi:10.21685/2307-5538-2022-3-6

### *Введение*

Речевой сигнал является аналоговым, поэтому для проведения каких-либо работ с ним сначала его преобразуют в дискретные сигналы. После чего речевой сигнал представлен в виде зависимости амплитуды от дискретных отсчетов времени.

Из-за сложности речевого сигнала, а именно из-за не стационарности данного сигнала, его предварительная обработка – один из важных шагов при распознавании речи. На этом этапе осуществляются фильтрация, кодирование, восстановление речи.

Основными методами предварительной обработки речевых сигналов являются:

1) преобразование Фурье является частотным методом. Недостаток – частотные компоненты не могут быть локализованы во времени;

2) вейвлет-преобразование является частотно-временным методом. Он стал альтернативой в обработке речевого сигнала преобразованию Фурье. Недостатком этого метода является большое количество предварительных вычислений;

3) в результате декомпозиции на эмпирические моды сигнал раскладывается на составляющие, называемыми модами. И достоинством этого метода является то, что эти моды вычисляются в ходе процесса обработки и не требуется никаких предварительных расчетов. Поэтому данный метод и будет далее рассматриваться.

### *Описание метода декомпозиции на эмпирические моды и выбор подметода*

Метод декомпозиции на эмпирические моды является итерационной вычислительной процедурой, после прохождения которой исходные сигналы раскладываются на эмпирические моды. Эмпирические моды не задаются заранее, а вычисляются в ходе процесса, что и выделяется в названии метода. Данный метод используется при обработке нестационарных сигналов за счет способности анализа локальных явлений.

Эмпирическая мода, внутреннее колебание или мода – эта функция, которая обладает следующими свойствами:

1) при сравнении количества максимумов и минимумов с количеством пересечений нуля, результаты не должны отличаться более, чем на единицу;

2) среднее значение, которое определяется по двум огибающим – верхней и нижней, должно быть равно нулю.

В основе механизма разложения на эмпирические моды лежит построение гладких огибающих методом сплайн аппроксимации по максимумам и минимумам сигнала и дальнейшее вычитание среднего этих огибающих из исходного сигнала.

Алгоритм действий при разложении сигнала на моды:

1. Сначала определяются точки пересечения функции в нуле и экстремумы.

2. Полученные максимумы и минимумы интерполируют. Методом интерполяции является интерполяция с помощью кубических сплайнов .

3. После вычисляется полусумма огибающих и вычитается из основного сигнала. Полученный сигнал является претендентом на то, чтобы быть первой эмпирической модой (ЭМ). Полученный сигнал проверяют на соответствие двум условиям, описанным ранее. Если сигнал удовлетворяет этим условиям, то он действительно является ЭМ. Если нет, то возвращаемся к пункту 1, только в качестве исходного сигнала будет использоваться сигнал, полученный в пункте 3.

4. После найденная ЭМ вычитается из первоначального сигнала. Остаток сигнала после этой процедуры используется для повторения пунктов 1–3 [1].

В результате разложение сигнала на эмпирические моды можно записать в следующем виде:

$$y(i) = \sum_{n=1}^N y_n''(i) + y_{res}(i),$$

где  $y(i)$  – исходный сигнал;  $y_n''(i)$  – эмпирическая мода с номером  $n$ ;  $y_{res}(i)$  – остаток, который может быть трендом или постоянной величиной;  $N$  – общее количество эмпирических мод [2].

Метод декомпозиции на эмпирические моды делится на методы, и одним из таких методов является улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом.

В этом методе для снижения уровня белого шума, который остается в модах, используются не значения мод, а локальные средние значения.

Данный метод обеспечивает:

1) адаптивное разложение, так как базисные функции, используемые при декомпозиции, извлекаются непосредственно из исходного речевого сигнала и позволяют учитывать только ему свойственные особенности (скрытые модуляции, области концентрации энергии и т.п.);

2) уменьшение значения белого шума;

3) отсутствие паразитных ЭМ, возникающих на ранних этапах декомпозиции вследствие перекрытия масштабно-энергетических пространств мод.

Параметры функционирования улучшенной ПМДЭМАШ:

1)  $Nstd$  – стандартное отклонение амплитуды добавляемого белого шума;

2)  $NR$  – число реализаций декомпозиций (добавлений белого шума);

3)  $MaxIter$  – максимально допустимое количество просеивающих итераций (критерий останова);

4)  $SNRFlag$  – отношения сигнал/шум между добавленным шумом и остатком, к которому добавляется шум (если значение равно 1, тогда отношение сигнал/шум увеличивается для каждого этапа декомпозиции; если равно 2, то отношение сигнал/шум будет одинаковым для всех этапов) [3].

Алгоритм предварительной обработки с помощью метода улучшенной ПМДЭМАШ представлен на рис. 1.

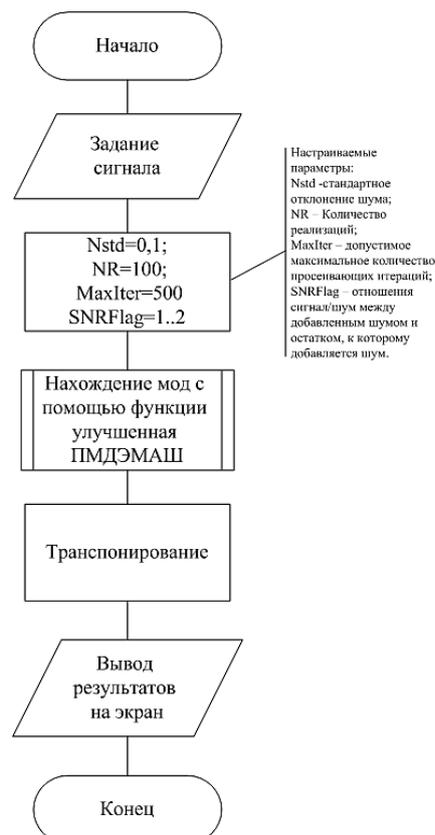


Рис. 1. Алгоритм распознавания речевых сигналов с помощью улучшенной ПМДЭМАШ

### Описание работы

Для распознавания голосовых команд требуется предварительная обработка данного сигнала.

Предварительная обработка происходит в несколько этапов. На первом этапе сигнал оцифровывается. После оцифровки сигнала происходит его разложение с помощью метода декомпозиции на эмпирические моды (ДЭМ). Данный метод описан выше и его результаты представлены на рис. 2.

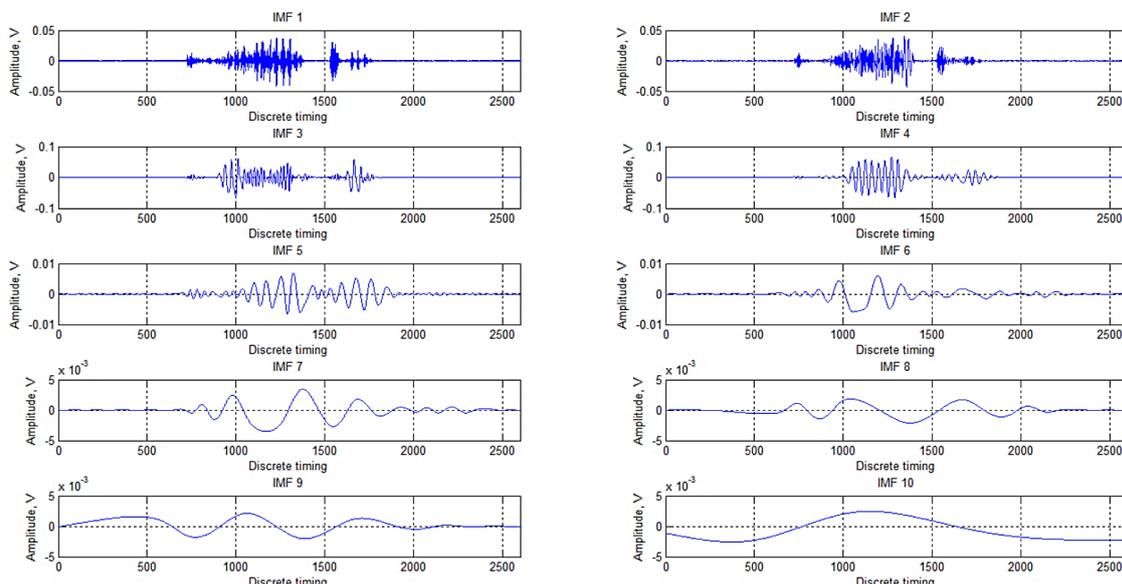


Рис. 2. Результаты разложения сигнала на эмпирические моды

После разложения сигнала на эмпирические моды требуется выделить информативную составляющую. Данная составляющая является наиболее информативной для сигнала и именно она является отличительной для сигнала. В данном сигнале информативной составляющей является четвертая мода (IMF4). Выделение информативной составляющей происходит путем сравнения нескольких разложенных сигналов (рис. 3).

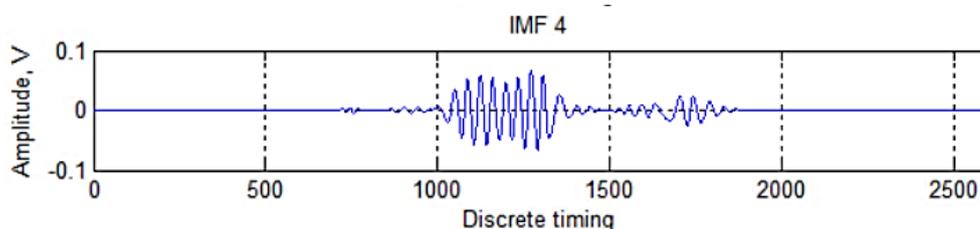


Рис. 3. Информативная мода

Для увеличения вероятности распознавания голосовых команд требуется упростить сигнал. Для этого требуется перейти в частотную область с помощью преобразования Фурье (рис. 4).

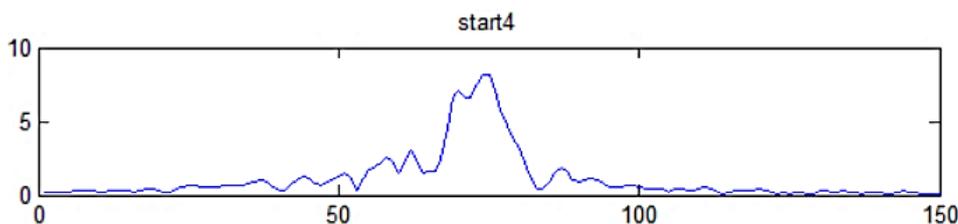


Рис. 4. Информативная мода в частотном виде

Для наглядности эксперимента строятся огибающие графика, представленного выше, и для сравнения с другими сигналами проводится нормирование графиков по оси Y (рис. 5).

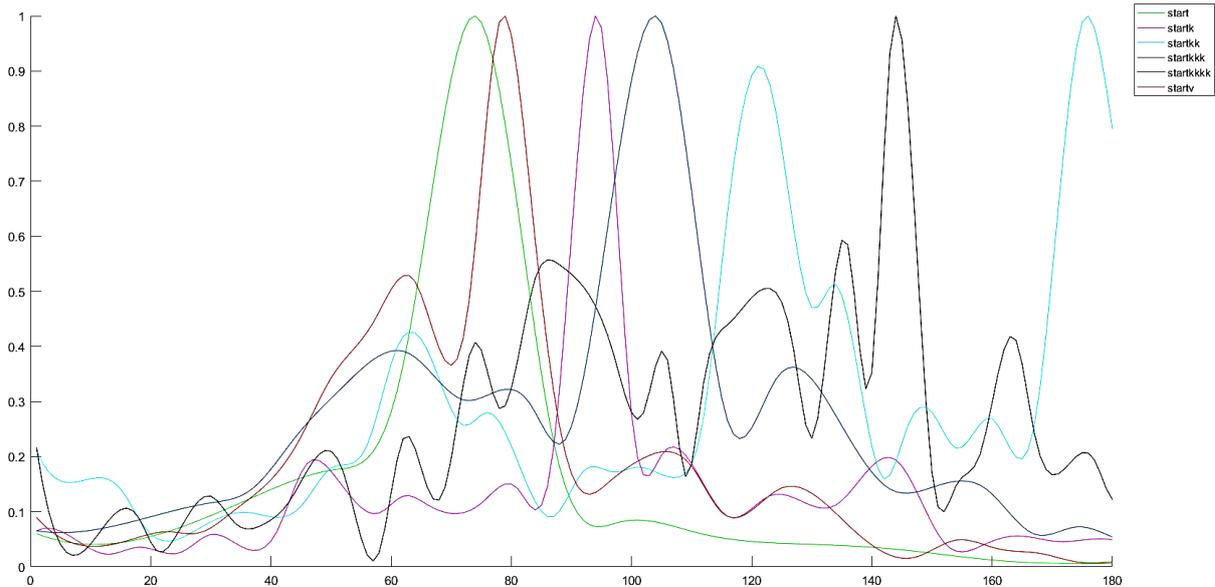


Рис. 5. Результаты обработки

### Заключение

По результатам рассмотрения разновидностей метода декомпозиции на эмпирические моды и выбран наилучший метод, а именно улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом. Произведено моделирование разложения речевого сигнала на моды и выделения информативной моды (составляющей). По полученным модам (с помощью преобразования Фурье) были построены огибающие. В результате построения огибающих информативных составляющих речевых сигналов сделан вывод о правильности выбора информативной моды. На рис. 5 видно, что речевые команды от разных людей после нормализации сигналов схожи по внешнему виду.

### Список литературы

1. Алимуратов А. К., Квитка Ю. С. Применение комплементарной множественной декомпозиции на эмпирические моды для анализа речевых сигналов // Измерение. Мониторинг. Управление. Контроль. 2014. № 4. С. 69–75.
2. Алимуратов А. К., Чураков П. П., Тычков А. Ю. Фильтрация речевых сигналов с использованием метода множественной декомпозиции и оценки энергии эмпирических мод // Известия высших учебных заведений. Поволжский регион. Технические науки. 2012. № 4. С. 50–61.
3. Омпов В. Д., Бороньев В. В. Комплементарная множественная декомпозиция на эмпирические моды с адаптивным шумом как метод решения основных проблем применения преобразования Гильберта-Хуанга // Журнал радиоэлектроники. 2016. № 9.

### References

1. Alimuradov A.K., Kvitka Yu.S. Application of complementary multiple decomposition into empirical modes for the analysis of speech signals. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control.* 2014;(4):69–75. (In Russ.)
2. Alimuradov A.K., Churakov P.P., Tychkov A.Yu. Filtering of speech signals using the method of multiple decomposition and evaluation of the energy of empirical modes. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki = Izvestia of higher educational institutions. Volga region. Technical sciences.* 2012;(4):50–61. (In Russ.)
3. Ompokov V.D., Boronoev V.V. Complementary multiple decomposition into empirical modes with adaptive noise as a method for solving the main problems of applying the Hilbert-Huang transform. *Zhurnal radioelektroniki = Journal of Radioelectronics.* 2016;(9). (In Russ.)

*Информация об авторах / Information about the authors*

**Валерий Валерьевич Козлов**

кандидат технических наук,  
доцент кафедры информационно-измерительной  
техники и метрологии,  
Пензенский государственный университет  
(Россия, г. Пенза, ул. Красная, 40)  
E-mail: val369@mail.ru

**Valeriy V. Kozlov**

Candidate of technical sciences, associate professor  
of the sub-department of information  
and measuring equipment and metrology,  
Penza State University  
(40 Krasnaya street, Penza, Russia)

**Екатерина Александровна Фокина**

инженер-электроник,  
Научно-исследовательский институт  
физических измерений  
(Россия, г. Пенза, ул. Володарского, 8/10)  
E-mail: Ekaterina.isay1997@gmail.com

**Ekaterina A. Fokina**

Electronics engineer,  
Research and Development Institute  
for Physical Measurements  
(8/10 Volodarsky street, Penza, Russia)

**Алексей Анатольевич Трофимов**

доктор технических наук, доцент,  
профессор кафедры информационно-  
измерительной техники и метрологии,  
Пензенский государственный университет  
(Россия, г. Пенза, ул. Красная, 40)  
E-mail: iit@pnzgu.ru

**Aleksey A. Trofimov**

Doctor of technical sciences, associate professor,  
professor of the sub-department of information  
and measuring equipment and metrology,  
Penza State University  
(40 Krasnaya street, Penza, Russia)

**Авторы заявляют об отсутствии конфликта интересов /  
The authors declare no conflicts of interests.**

**Поступила в редакцию/Received 16.05.2022**

**Поступила после рецензирования/Revised 14.06.2022**

**Принята к публикации/Accepted 18.07.2022**