

ПРИБОРЫ, СИСТЕМЫ И ИЗДЕЛИЯ МЕДИЦИНСКОГО НАЗНАЧЕНИЯ MEDICAL DEVICES, SYSTEMS AND PRODUCTS

УДК 004.934

doi:10.21685/2307-5538-2022-2-10

СПОСОБ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ МЕТОДА ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ

**А. К. Алимуратов¹, А. Ю. Тычков², П. П. Чураков³,
А. В. Баранова⁴, Д. С. Дудников⁵**

^{1,2,3,4,5} Пензенский государственный университет, Пенза, Россия

¹alansapfir@yandex.ru, ²tychkov-a@mail.ru, ³churakov.pp@gmail.com,

⁴anastasiyabar@mail.ru, ⁵dmitriy.s.gmpf@gmail.com

Аннотация. *Актуальность и цели.* Выбор и обоснование оптимального набора информативных параметров речевых сигналов зависит от используемых методов обработки и корректного анализа полученных результатов. Целью работы является повышение эффективности обработки речевых сигналов за счет расширения пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик речи посредством применения адаптивных методов частотно-временного анализа. *Материалы и методы.* В работе использовалась уникальная технология разложения нестационарных данных на частотные составляющие, не требующая априорной информации об анализируемом сигнале – декомпозиция на эмпирические моды. Программная реализация способа была выполнена в среде математического моделирования © Matlab (MathWorks). *Результаты.* Разработан способ обработки речевых сигналов на основе метода декомпозиции на эмпирические моды. Суть способа заключается в равномерном делении исходного речевого сигнала на фрагменты, декомпозиции фрагментов на эмпирические моды и формировании модовых речевых сигналов. Проведены исследование способа и анализ полученных результатов: количества эмпирических мод, разности между исходным и реконструированным сигналами, времени формирования модовых речевых сигналов. *Выводы.* В соответствии с полученными результатами исследования выявлено, что разработанный способ в действительности обеспечивает расширение пространства информативно-значимых характеристик за счет формирования набора новых модовых речевых сигналов с минимальной ошибкой. Обеспечено необходимое и достаточное значение разности между исходным и реконструированным сигналами – не более 0,001 В. Разработанный способ может успешно использоваться при формировании оптимального набора параметров речи для обнаружения и классификации естественно выраженных психоэмоциональных состояний человека.

Ключевые слова: обработка речевых сигналов, декомпозиция на эмпирические моды, психоэмоциональное состояние человека, эмоции

Для цитирования: Алимуратов А. К., Тычков А. Ю., Чураков П. П., Баранова А. В., Дудников Д. С. Способ обработки речевых сигналов на основе метода декомпозиции на эмпирические моды // Измерения. Мониторинг. Управление. Контроль. 2022. № 2. С. 75–89. doi:10.21685/2307-5538-2022-2-10

EMD-BASED TECHNIQUE FOR SPEECH SIGNAL PROCESSING

**A.K. Alimuradov¹, A.Yu. Tychkov², P.P. Churakov³,
A.V. Baranova⁴, D.S. Dudnikov⁵**

^{1,2,3,4,5} Penza State University, Penza, Russia

¹alansapfir@yandex.ru, ²tychkov-a@mail.ru, ³churakov.pp@gmail.com,

⁴anastasiyabar@mail.ru, ⁵dmitriy.s.gmpf@gmail.com

Abstract. *Background.* Being selected and validated, the optimal set of speech signal informative parameters depends on the used processing methods and accurate evaluation of the results obtained. The purpose of the work is to enhance the efficiency of speech signal processing by expanding the space for informatively significant amplitude, time, frequency, and energy speech characteristics via the use of adaptive time-frequency analysis methods. *Materials and methods.* A unique technology for decomposing non-stationary data into frequency components, namely, empirical mode decomposition, when no a priori information regarding the analyzed signal is needed, has been used. The software implementation of the method has been performed in ©MATLAB (The MathWorks, Inc.) mathematical modeling environment. *Results.* A technique for speech signal processing based on the empirical mode decomposition has been developed. The proposed technique is based on the uniform splitting of the original speech signal into fragments, the empirical mode decomposition of the fragments, and the formation of mode speech signals. The technique has been investigated, and the following obtained results have been analyzed: the number of empirical modes, the difference between the original and reconstructed signals, and the time duration of mode speech signal formation. *Conclusions.* Based on the obtained research results, it has been revealed that the developed technique actually provides an expansion of the space for informatively significant characteristics due to the formation of a set of new mode speech signals with a minimum error. The necessary and sufficient difference between the original and reconstructed signals of less than 0.001 V has been provided. The developed technique can be efficiently used in the formation of an optimal set of speech parameters for detecting and classifying naturally expressed human psycho-emotional states.

Keywords: speech signal processing, empirical mode decomposition, human psycho-emotional state, emotions

For citation: Alimuradov A.K., Tychkov A.Yu., Churakov P.P., Baranova A.V., Dudnikov D.S. EMD-based technique for speech signal processing. *Izmereniya. Monitoring. Upravlenie. Kontrol' = Measurements. Monitoring. Management. Control.* 2022;(2):75–89. (In Russ.). doi:10.21685/2307-5538-2022-2-10

Введение

Речь представляет собой сложный акустический сигнал, образуемый речевым аппаратом человека с целью языкового общения [1]. Цифровая обработка речевых сигналов – это область современной науки, в рамках которой решаются следующие задачи: изменение формы представления информации из звуковой в электрическую, фильтрация (линейная и адаптивная) шума, усиление, сегментация на информативные участки, извлечение информативных параметров, кодирование, сжатие, восстановление и др. [2]. Каждая задача может быть решена с применением различных способов и подходов, которые в зависимости от области обработки следует разделить на временные, частотные и частотно-временные [3].

Обработка во временной области основана на анализе только характерных дискретных отсчетов речевого сигнала, в качестве которых могут использоваться максимумы и минимумы функции речевого сигнала, а также моменты пересечения функции с нулевой осью во времени. Обработка в частотной области основана на анализе абсолютно всех дискретных отсчетов, зарегистрированных в речевом сигнале. К основным недостаткам обработки во временной области относятся зависимость от уровня шума в речевом сигнале и смещения нулевого уровня, что приводит к неоднозначности выделения характерных анализируемых дискретных отсчетов. К основным недостаткам обработки в частотной области относятся большие вычислительные затраты и низкая адаптивность к локальным свойствам функции речевого сигнала (низкое спектральное разрешение). Частотно-временная обработка включает в себя все преимущества временной и частотной обработки речевых сигналов с минимальными проявлениями их недостатков.

На сегодняшний день наибольшую популярность в решениях задач по обработке речевых сигналов получили частотно-временные способы и подходы, основанные на преобразовании Фурье и вейвлет преобразовании [4]. Преимуществом данных решений является возможность разложения исследуемых речевых сигналов на составляющие для последующего детализированного анализа.

Физический смысл преобразования Фурье заключается в представлении исходного анализируемого сигнала в виде суммы гармонических колебаний с разными частотами [5]. Параметры каждой гармонической составляющей вычисляются прямым преобразованием, а сумма гармонических составляющих – обратным преобразованием. В области цифровой обработки речевых сигналов преобразование Фурье широко используются для преобразования функции сигнала из временной области в частотную с разложением его на частотные составляющие.

Физический смысл вейвлет-преобразования заключается в разложении исходного анализируемого сигнала на масштабированные и сдвинутые по оси времени версии функции вейвлета [6]. В результате получаемый набор коэффициентов отражает, насколько поведение исходного сигнала в данный момент времени похоже на поведение функции вейвлета на данном масштабе. Таким образом, вейвлет-коэффициенты отражают близость исходного сигнала к вейвлету данного масштаба. Чем ближе вид анализируемого сигнала в окрестности данного момента времени к виду вейвлета, тем большее абсолютное значение имеет соответствующий коэффициент.

В области цифровой обработки речевых сигналов вейвлет-преобразование обладает существенными преимуществами по сравнению с преобразованием Фурье. Это следует из возможности анализировать кратковременные локальные особенности сигналов, например, короткие всплески или провалы, разрывы и ступеньки функции и т.д. Недостатком вейвлет-преобразования является необходимость априорных сведений об анализируемом сигнале для корректного подбора функции вейвлета.

В последнее время широкое практическое применение в решениях задач по обработке речевых сигналов получило преобразование Гильберта – Хуанга [7], в основе которого заложен метод декомпозиции на эмпирические моды (ДЭМ) [8]. ДЭМ – это уникальная технология разложения на частотные составляющие, не требующая априорной информации об анализируемом сигнале. Метод ДЭМ и ее модификации широко используются в различных областях, связанных с обработкой и анализом нестационарных и нелинейных сигналов: природные явления [9–11], биотехнологии [12–14], медицина [15–17], анализ данных и машинное обучение [18, 19] и др. Методы ДЭМ активно применяются в приложениях по обработке речевых и акустических сигналов:

- определение высоты звука [20];
- идентификация диктора по голосу [21];
- акустический анализ [22];
- обработка и подавление шума в речи [23] и др.

В данной статье представлен способ обработки речевых сигналов, в котором используется метод ДЭМ. Предлагаемый способ основан на равномерном делении исходного речевого сигнала на фрагменты, декомпозиции фрагментов на эмпирические моды (ЭМ) и формировании модовых речевых сигналов. Суть формирования модовых речевых сигналов заключается в расширении пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик исходного сигнала, необходимых для дальнейшей обработки.

Статья является результатом научной работы коллектива авторов [24, 25], посвященной исследованию и поиску скрытых особенностей речевых сигналов, формированию оптимального набора параметров, релевантных естественно выраженным эмоциям человека посредством применения новых адаптивных методов частотно-временного анализа. Научные исследования выполняются при финансовой поддержке Совета по грантам Президента РФ, проект «Исследование скрытых паттернов речевых сигналов и разработка способов обнаружения и классификации естественно выраженных психоэмоциональных состояний человека», № МД-1066.2022.4.

Статья состоит из шести разделов. Во втором представлено краткое описание разновидностей методов ДЭМ, их преимущества и недостатки. Третий и четвертые разделы посвящены описанию и исследованию предлагаемого способа. Последние два раздела посвящены анализу результатов исследования, выводам и перспективам.

Декомпозиция на эмпирические моды

Эффективность поиска и формирования оптимального набора параметров речевых сигналов, релевантных естественно выраженным эмоциям человека зависит от адаптивности и возможностей применяемых методов частотно-временного анализа. Применение метода ДЭМ и ее модификаций в исследованиях эмоциональной речи началось сравнительно недавно, учитывая, что ДЭМ была разработана в 1998 г. [26–29]. Большинство авторов не проводят исследований по расширению пространства информативно-значимых характеристик речи и анализируют только исходные сигналы, содержащие естественно выраженные эмоции. В настоя-

щее время наблюдается недостаточное количество научных работ, в которых прослеживается полный цикл: расширение пространства информативно-значимых характеристик речи с помощью методов ДЭМ и поисковые исследования релевантного набора амплитудных, временных, частотных и энергетических параметров естественно выраженных эмоций.

В результате разложения методом ДЭМ исходный сигнал представляет собой набор внутренних функций, называемых модами или ЭМ. Отличительной особенностью процесса разложения методом ДЭМ является то, что базисные функции, используемые при разложении, извлекаются непосредственно из исходного сигнала. Таким образом, математическая модель ЭМ не задается заранее, а вычисляется в процессе отсеивания. Такой способ разложения обеспечивает адаптивность базисным функциям, позволяющую учитывать локальные особенности (экстремумы и нули функции сигнала) и особенности внутренней структуры исходного сигнала (скрытые модуляции, области концентрации энергии и т.п.).

Подробный анализ известных методов ДЭМ, применяемых для анализа сигналов естественной природы, выявил, что наиболее адаптивными к нестационарной речи являются: множественная ДЭМ (МДЭМ) [30] и улучшенная полная МДЭМ с адаптивным шумом (ПМДЭМАШ) [31]. С точки зрения отсеивания ЭМ методы МДЭМ и улучшенной ПМДЭМАШ аналогичны. Добавление контролируемого шума малой амплитуды на каждом этапе отсеивания (для создания новых экстремумов) позволяет избежать известных недостатков декомпозиции (смешивание мод, неполнота декомпозиции, остаточный шум, неинформативные «паразитные» моды):

$$x_j(n) = x(n) + w(n), \quad (1)$$

где $x_j(n)$ – зашумленные сигналы; $w_j(n)$ – белый шум малой амплитуды;

$$x_j(n) = \sum_{i=1}^I IMF_{ji}(n) + r_{ji}(n), \quad (2)$$

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

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

где $IMF(n)$ – ЭМ; $r(n)$ – конечный неделимый остаток; $i = 1, 2, \dots, I$ – количество ЭМ.

Важными параметрами настройки методов МДЭМ и улучшенной ПМДЭМАШ, влияющими на результат разложения, являются: $Nstd$ – стандартное отклонение амплитуды добавляемого белого шума (в процентном отношении от исходного сигнала), NR – количество реализаций (для дальнейшего усреднения), $MaxIter$ – количество итераций отсеивания ЭМ, $SNRFlag$ – отношение сигнал/шум для каждого этапа разложения (только для метода улучшенной ПМДЭМАШ).

На рис. 1 представлен результат разложения вокализованного участка речи длительностью 100 мс. Разложение осуществлено посредством метода улучшенной ПМДЭМАШ ($Nstd = 20\%$, $NR = 500$, $MaxIter = 1000$, $SNRFlag = 2$ – отношение сигнал/шум одинаковое на всех этапах разложения). Для удобства визуализации значения горизонтальных (дискретные отсчеты времени) и вертикальных осей (амплитуда, В) осциллограмм в правом столбце автоматически масштабированы и не обозначены. В левом столбце значения горизонтальных (частота, Гц) и вертикальных осей (магнитуда, дБ) спектральных плотностей мощности также не обозначены и масштабированы автоматически.

В соответствии с рис. 1 в результате разложения вокализованного участка речи получено восемь мод (ЭМ1–ЭМ8). Каждой ЭМ соответствует определенный частотный диапазон. Исходя из этого, можно сделать вывод, что метод улучшенной ПМДЭМАШ функционирует как набор фильтров, осуществляя фильтрацию от высокочастотного к низкочастотному диапазону.

Из полученных мод можно выделить информативные (ЭМ1–ЭМ5) и компенсирующие (ЭМ6–ЭМ8). Первые относятся к полезной информации сигнала и характеризуются большей энергией. Вторые являются результатом неточности математического аппарата декомпозиции и характеризуются незначительной энергией (не более 5 % от общей энергии исходного сигнала).

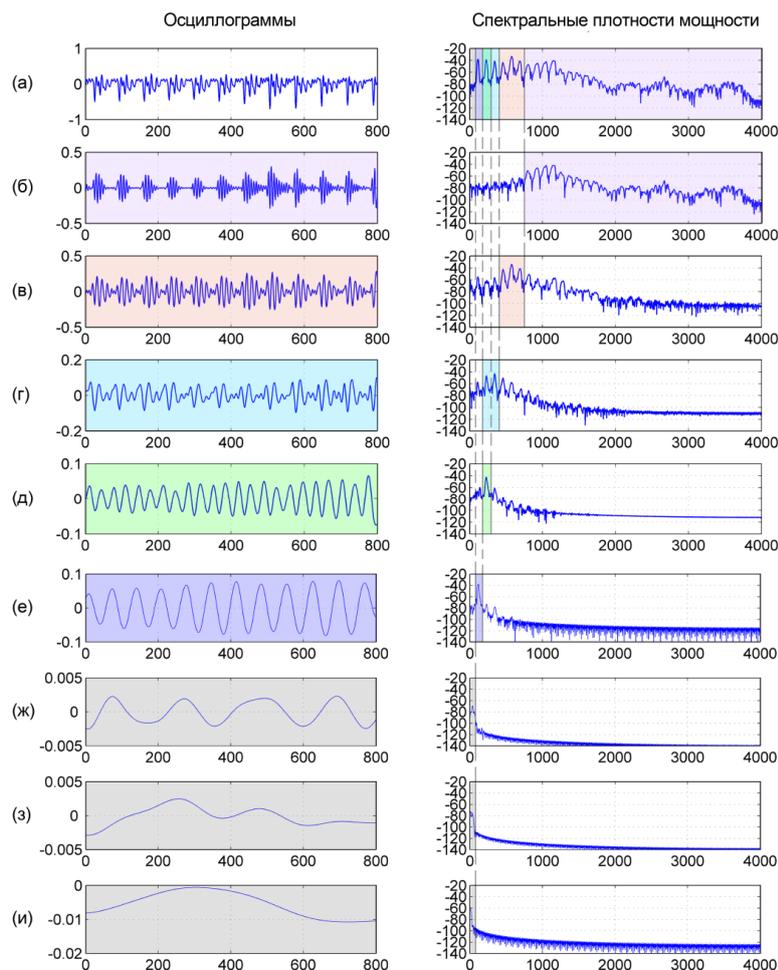


Рис. 1. Результат разложения участка вокализованной речи с помощью метода улучшенной ПМДЭМАШ

На рис. 2 представлен пример, иллюстрирующий полноту разложения и последующую реконструкцию исходного сигнала по информативным модам (ЭМ1–ЭМ5). Максимальная разница между исходным и реконструированным сигналам составляет 0,0149 В, что является незначительным.

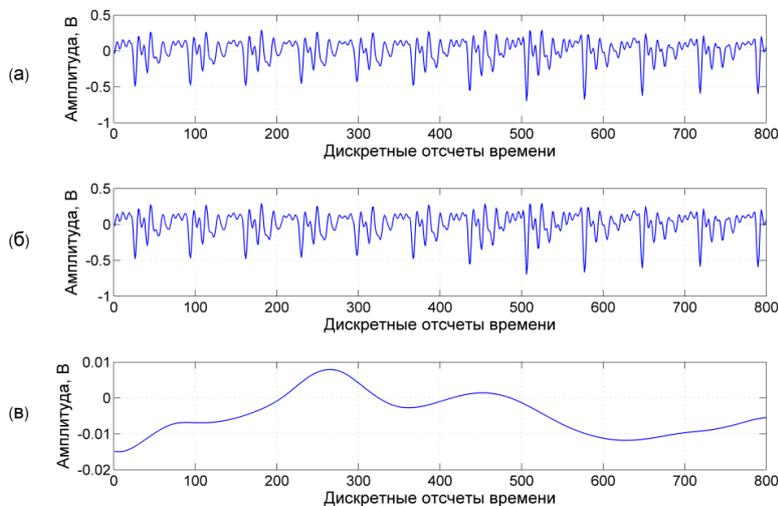


Рис. 2. Обеспечение полноты разложения и реконструкция сигнала по информативным ЭМ:
a – исходный речевой сигнал; *б* – реконструированный сигнал;
в – разность между исходным и реконструированными сигналами

Основным недостатком, ограничивающим широкое практическое применение методов ДЭМ в реальном времени, является продолжительность процедуры декомпозиции. Продолжительность процедуры и результат декомпозиции напрямую зависят от длительности анализируемого сигнала. Чем длительнее анализируемый сигнал, тем продолжительнее процедура декомпозиции, и наоборот. Также длительность анализируемого сигнала влияет на полноту разложения, то есть частотно-избирательные свойства – количество и качество получаемых информативных и компенсирующих ЭМ.

Описание способа обработки

Предлагаемый способ основан на равномерном делении исходного речевого сигнала на фрагменты, декомпозиции фрагментов на ЭМ и формировании модовых речевых сигналов. Рассмотрим подробнее этапы способа обработки речевых сигналов.

Фрагментирование представляет собой процесс линейного разделения исходного речевого сигнала $x(n)$ на отрезки одинаковой длительности, которые записываются в отдельные переменные $x_s(n)$. Далее каждый фрагмент исходного сигнала обрабатывается как отдельный кратковременный речевой сигнал. Линейное разделение речевого сигнала на фрагменты осуществляется по следующим формулам:

$$S = \frac{x(n)}{L}, \quad (5)$$

где S – количество фрагментов в исходном речевом сигнале; L – количество дискретных отсчетов времени в одном фрагменте;

$$x_{s+1}(n) = x[(sL) + 1 : (s+1)L], \quad (6)$$

где $s = 0, 1, 2, \dots, S$ – номер фрагмента.

Для экономии вычислительных ресурсов и повышения скорости обработки в предлагаемом способе фрагментирование осуществляется без перекрытия [3].

Исследование влияния длительности анализируемых фрагментов речи на частотно-избирательные свойства методов ДЭМ представлено в работе [32]. В соответствии с результатами исследования сделан вывод, что для корректного частотно-временного анализа кратковременных речевых сигналов на основе методов ДЭМ длительность анализируемых фрагментов речи должна быть не менее 30 мс. Оптимальной же длительностью, при которой частотно-избирательные свойства декомпозиции обеспечивают необходимую и достаточную полноту разложения, является 50 мс и выше. В предлагаемом способе реализована возможность изменения длительности фрагментов в диапазоне от 10 до 1000 мс.

Как отмечалось ранее, наиболее адаптивными к нестационарной речи являются методы МДЭМ и улучшенной ПМДЭМАШ. Отличительной особенностью метода улучшенной ПМДЭМАШ является возможность локального разложения белого шума на шумовые ЭМ параллельно с разложением исходного сигнала. Использование шумовых мод в качестве добавляемого контролируемого белого шума на каждом этапе декомпозиции обеспечивает полноту разложения.

В предлагаемом способе реализована возможность выбора методов ДЭМ, МДЭМ и улучшенная ПМДЭМАШ, в том числе возможность изменения параметров настройки декомпозиции: Nstd, NR, MaxIter и SNRFlag.

Суть формирования модовых речевых сигналов заключается в расширении пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик исходного сигнала. Расширение информативного пространства обеспечивается за счет формирования новых модовых речевых сигналов. Каждый модовый сигнал содержит в себе особенности внутренней структуры исходного речевого сигнала (скрытые модуляции, области концентрации энергии и т.п.).

В соответствии с результатом декомпозиции каждый фрагмент исходного речевого сигнала представлен набором ЭМ. Формирование модовых сигналов представляет собой процесс объединения ЭМ фрагментов исходного речевого сигнала:

$$xmode_i(n) = \sum_{s=1}^S IMF_{s,i}[(sL) + 1 : (s+1)L], \quad (7)$$

где $xmode_i(n)$ – модовый речевой сигнал; $i = 1, 2, \dots, I$ – количество ЭМ для каждого фрагмента.

Количество сформированных модовых речевых сигналов зависит от количества используемых информативных ЭМ, полученных для каждого фрагмента. На рис. 3 представлен пример процесса формирования четырех модовых речевых сигналов.

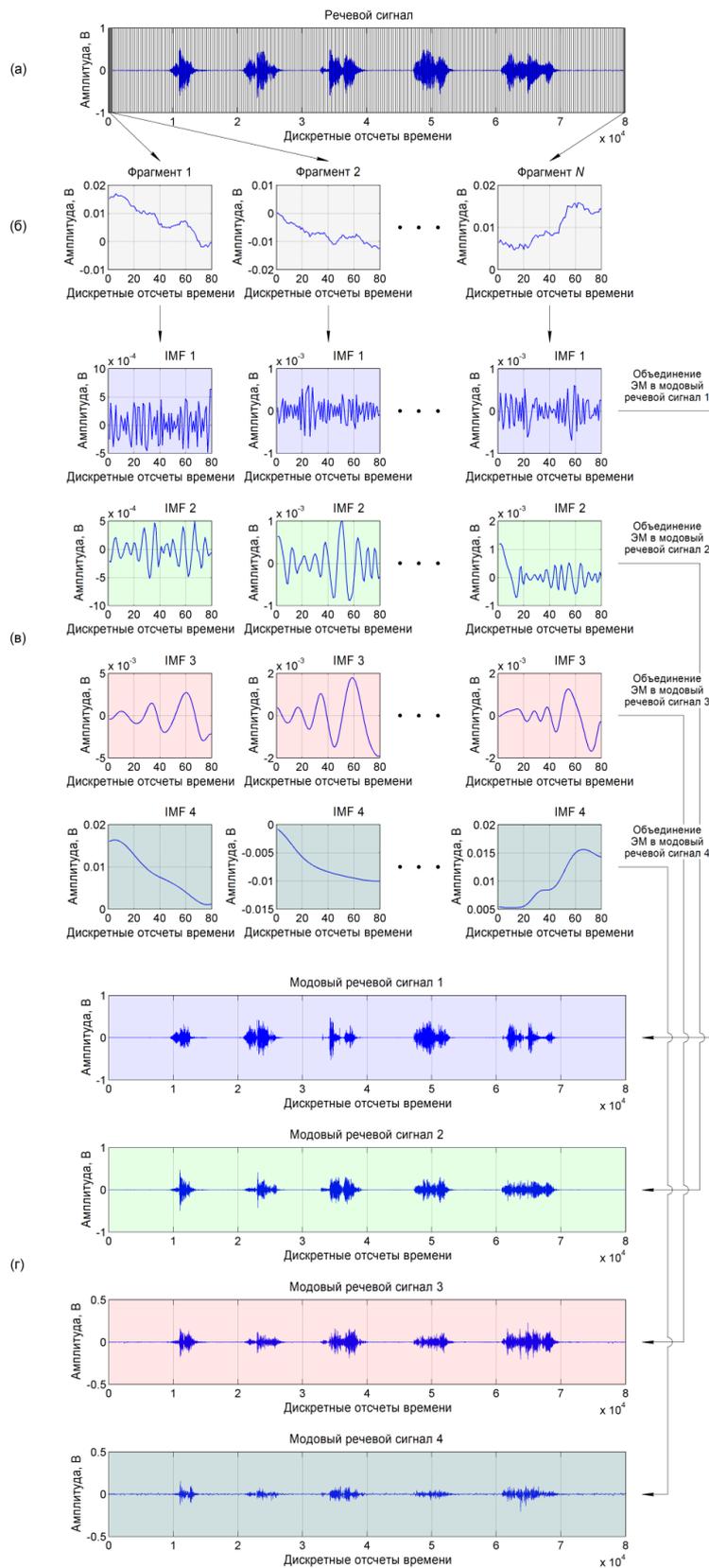


Рис. 3. Процесс формирования четырех модовых речевых сигналов: *a* – исходный речевой сигнал; *б* – фрагменты речевого сигнала; *в* – информативные ЭМ фрагментов; *г* – модовые речевые сигналы

Исследование способа обработки

В табл. 1 представлены настраиваемые и исследуемые параметры способа обработки речевых сигналов на основе методов ДЭМ.

Таблица 1

Настраиваемые и исследуемые параметры способа обработки речевых сигналов на основе методов ДЭМ

Настраиваемые параметры	Исследуемые параметры
Длительность анализируемых фрагментов (мс): 10, 20, 30, 50, 100, 300, 500, 1000, 2000	Среднее значение количества ЭМ
Метод декомпозиции: ДЭМ, МДЭМ, улучшенная ПМДЭМАШ	Разность между исходным и реконструированным сигналами (В)
Параметры МДЭМ и улучшенной ПМДЭМАШ: NR (в размах) – 5, 50; MaxIter (в размах) – 10, 100	Время формирования набора модовых речевых сигналов (сек)

Для исследования сформирована речевая база данных. Речевые сигналы длительностью не более 10 с были зарегистрированы посредством специализированных методики и технических средств. Подготовленные дикторы в количестве 20 человек воспроизводили речь на русском языке, содержащую публицистический текст (30 записей), текст из литературного произведения (30 записей) и счет чисел от 0 до 100 (10 записей).

Для разложения исходного сигнала методами МДЭМ и улучшенной ПМДЭМАШ в рамках исследования изменялись следующие параметры: NR = 5 и 50, MaxIter = 10 и 100. Параметры Nstd = 20 % и SNRFlag = 2 оставались неизменными. Реализация методов ДЭМ в программе Matlab была заимствована из работы [33].

В табл. 2–4 и на рис. 4–6 представлены усредненные результаты исследования способа обработки сигналов сформированной речевой базы данных.

Таблица 2

Среднее значение количества ЭМ

Метод декомпозиции	Длительность анализируемых фрагментов, мс								
	10	30	50	100	200	300	500	1000	2000
ДЭМ	4,55	5,92	6,67	7,69	8,90	9,67	10,45	11,50	13,80
МДЭМ (NR = 50, MaxIter = 100)	4,70	6,37	7,21	8,37	9,62	10,55	11,45	13,10	14,80
МДЭМ (NR = 5, MaxIter = 10)	4,57	5,98	6,72	7,74	8,84	9,45	10,30	11,80	13,00
Улучшенная ПМДЭМАШ (NR = 50, MaxIter = 100)	4,55	5,97	6,67	7,70	8,88	9,55	10,70	11,70	13,60
Улучшенная ПМДЭМАШ (NR = 5, MaxIter = 10)	4,50	5,84	6,53	7,54	8,54	9,24	9,90	11,00	12,20

Таблица 3

Разность между исходным и реконструированным сигналами

Метод декомпозиции	Длительность анализируемых фрагментов, мс								
	10	30	50	100	200	300	500	1000	2000
ДЭМ	0,0041731	0,0036321	0,0025682	0,001734	0,0013033	0,0010556	0,00055186	0,00030492	0,0002728
МДЭМ (NR = 50, MaxIter = 100)	0,00397	0,0014323	0,00086711	0,00032802	0,00027326	0,00013162	0,000064407	0,000014068	0,000037316
МДЭМ (NR = 5, MaxIter = 10)	0,0038898	0,0033819	0,0018001	0,0015028	0,00095801	0,00049735	0,00027791	0,0002972	0,00024455
Улучшенная ПМДЭМАШ (NR = 50, MaxIter = 100)	0,0040564	0,0037643	0,0025106	0,0016722	0,0012479	0,00081638	0,00061309	0,00037565	0,0002948
Улучшенная ПМДЭМАШ (NR = 5, MaxIter = 10)	0,0040455	0,0034429	0,0018636	0,0014704	0,0012904	0,00075916	0,00089567	0,00035437	0,00028948

Таблица 4

Время формирования набора модовых речевых сигналов

Метод декомпозиции	Длительность анализируемых фрагментов, мс								
	10	30	50	100	200	300	500	1000	2000
ДЭМ	68,99	33,43	25,41	19,03	14,89	13,38	15,48	18,26	27,45
МДЭМ (NR = 50, MaxIter = 100)	2879,84	1636,91	1237,90	897,68	737,65	665,91	696,42	918,02	1345,55
МДЭМ (NR = 5, MaxIter = 10)	261,97	125,29	96,11	70,58	49,99	43,20	35,28	30,13	30,65
Улучшенная ПМДЭМАШ (NR = 50, MaxIter = 100)	6261,58	3412,03	2562,38	1789,33	1469,13	1325,50	1387,61	1623,22	2402,96
Улучшенная ПМДЭМАШ (NR = 5, MaxIter = 10)	622,73	326,28	249,22	168,75	123,62	110,99	117,86	100,44	124,19

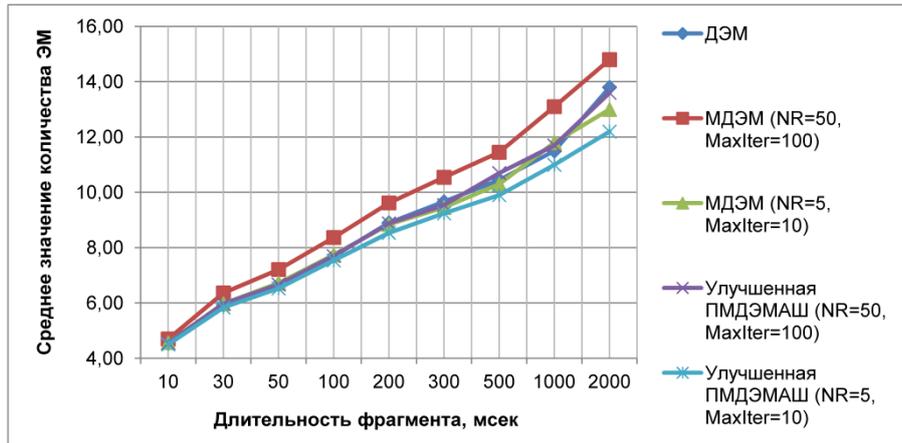


Рис. 4. Среднее значение количества ЭМ

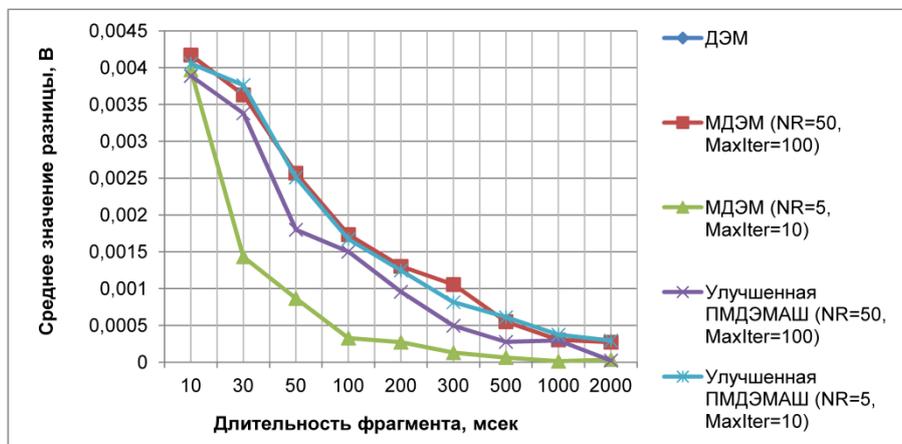


Рис. 5. Разность между исходным и реконструированным сигналами

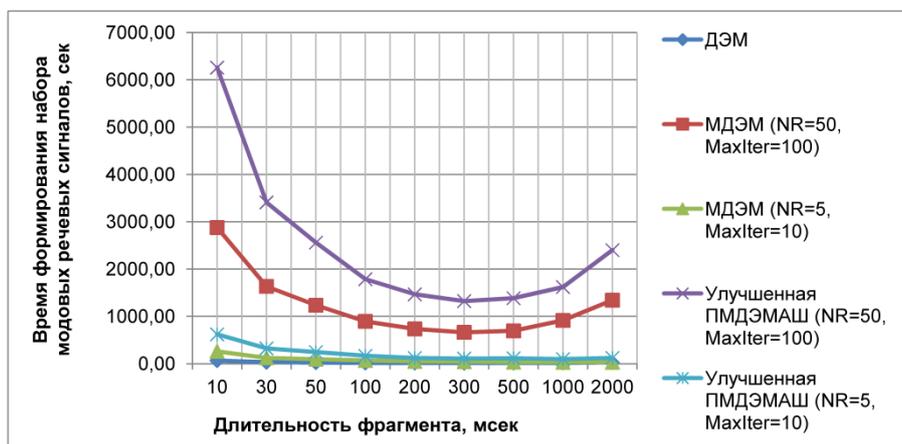


Рис. 6. Время формирования набора модовых речевых сигналов

Анализ результатов исследования

Для корректной оценки результатов исследования необходим совокупный анализ полученных значений среднего количества ЭМ, разности между исходным и реконструированным сигналами и времени формирования набора модовых речевых сигналов. Данная необходимость объясняется тем, что каждый настраиваемый и исследуемый параметры оказывают непосредственное влияние друг на друга. Например, метод и параметры декомпозиции, а также длительность анализируемых фрагментов влияют на количество получаемых ЭМ, которые

в свою очередь влияют на процесс формирования модовых сигналов и разность между исходным и реконструированным сигналами. Каждый настраиваемый параметр также влияет на итоговое время, необходимое для формирования набора модовых речевых сигналов.

В соответствии с условиями совокупного анализа и полученными результатами, представленными в табл. 2–4 на рис. 4–6, определены следующие оптимальные значения параметров:

1) длительность анализируемого фрагмента – 300 мс. В этом случае необходимо минимальное время для формирования набора модовых речевых сигналов (табл. 4, рис. 6);

2) количество ЭМ – от 8 до 10. В этом случае обеспечивается необходимая и достаточная полнота разложения (табл. 2, рис. 4);

3) разность между исходным и реконструированным сигналами – не более 0,001 В (т.е. не более 0,1 %). В этом случае обеспечивается минимальная ошибка при формировании набора модовых речевых сигналов (табл. 3, рис. 5).

Учитывая, что способ основан на методах ДЭМ, целесообразным является анализ влияния метода и параметров декомпозиции на результат формирования набора модовых речевых сигналов. Необходимая и достаточная полнота разложения (оптимальное количество ЭМ с округлением в большую сторону) обеспечивается всеми методами ДЭМ при анализе фрагментов длительностью от 50 до 500 мс. Однако, учитывая значения разности между исходным и реконструированным сигналами (0,00049735 В) и времени формирования набора модовых речевых сигналов (43,20 сек), оптимальным (но не самым быстрым) является метод МДЭМ (NR = 5, MaxIter = 10). Минимальная ошибка при формировании набора модовых речевых сигналов (оптимальное значение разности между исходным и реконструированным сигналами) обеспечивается всеми методами декомпозиции при анализе фрагментов длительностью от 500 мс. Минимальное время для формирования набора модовых речевых сигналов обеспечивается методом ДЭМ.

Сравнивая значения времени формирования набора модовых речевых сигналов, можно сделать следующие однозначные выводы:

1) для обработки речевых сигналов в реальном времени способ должен быть основан на методе ДЭМ;

2) для постобработки речевых сигналов способ может быть основан на методах МДЭМ и улучшенной ПМДЭМАШ (декомпозиции с добавлением контролируемого шума малой амплитуды);

3) для постобработки речевых сигналов, предпочтительно использовать метод улучшенной ПМДЭМАШ с оптимальными значениями NR = 50, MaxIter = 100.

Заключение

Подводя итоги анализа результатов исследований, можно сделать следующие выводы:

1) предлагаемый способ обработки речевых сигналов на основе методов ДЭМ в действительности обеспечивает расширение пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик. Это обеспечивается за счет формирования с минимальной ошибкой набора новых модовых речевых сигналов, содержащих в себе особенности внутренней структуры исходного речевого сигнала (скрытые модуляции, области концентрации энергии и т.п.);

2) способ за счет расширения пространства информативно-значимых характеристик исходных речевых сигналов может успешно использоваться при формировании оптимального набора параметров речи, релевантных естественно выраженным эмоциям человека.

В перспективе коллективом авторов планируется провести дополнительное исследование влияния смешивания ЭМ, неполноты декомпозиции, остаточного шума и неинформативных «паразитных» ЭМ на формирование набора модовых речевых сигналов.

Список литературы

1. Фант Г. К. Акустическая теория речеобразования : пер. с англ. Л. А. Варшавского, В. И. Медведева ; науч. ред. В. С. Григорьева. М. : Наука, 1964. 284 с.
2. Михайлов В. Г., Златоусова Л. В. Измерение параметров речи / под ред. М. А. Сапожникова. М. : Радио и связь, 1987. 168 с.
3. Алимуратов А. К., Чураков П. П. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи // Измерение. Мониторинг. Управление. Контроль. 2015. № 2. С. 27–35.

4. Huang X., Acero A., Hon H.-W. Spoken Language Processing. Guide to Algorithms and System Development. New Jersey : Prentice Hall, 2001. 980 p.
5. Ли У. А., Нейбург Э. П., Мартин Т. Б. [и др.] Методы автоматического распознавания речи : в 2 кн. ; пер. с англ. / под ред. У. Ли. М. : Мир, 1983. Кн. 1. 328 с.
6. Клетт Д. Х., Барнет Дж. А., Бернштейн М. И. [и др.] Методы автоматического распознавания речи : в 2 кн. ; пер. с англ. / под ред. У. Ли. М. : Мир, 1983. Кн. 2. 392 с.
7. Huang E. Huang, Samuel S.P. Shen. Hilbert-Huang Transform and its application. Interdisciplinary mathematical sciences // Interdisciplinary Mathematical Sciences. World Scientific Publishing Company. 2005. Book 5. 324 p.
8. Huang N. E., Zheng Sh., Steven R. L. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis // Proceedings of the Royal Society of London. 1998. Vol. 454. P. 903–995.
9. Mi X., Liu H., Li Y. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine // Energy Conversion and Management. 2019. Vol. 80. P. 196–205.
10. Wu Z., Schneider E., Hu Z., Cao L. The impact of global warming on enso variability in climate records // Tech. report, Cent. Ocean. Stud. 2001. Vol. 110. 25 p.
11. Lee T., Ouara D. B. M. J. Multivariate Nonstationary Oscillation Simulation of Climate Indices With Empirical Mode Decomposition // Water Resources Research. 2019. Vol. 55, № 6. P. 5033–5052.
12. Labate D., Foresta F. La, Occhiuto G. [et al.]. Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A comparison // IEEE Sensors Journal. 2013. Vol. 13, № 7. P. 2666–2674.
13. Zhang Y., Gao Y., Wang L. [et al.]. The removal of wall components in doppler ultrasound signals by using the empirical mode decomposition algorithm // IEEE Transactions on Biomedical Engineering. 2007. Vol. 54, № 9. P. 1631–1642.
14. Andrade A. O., Nasuto S., Kyberd P. [et al.]. EMG signal filtering based on Empirical Mode Decomposition // Biomedical Signal Processing and Control. 2006. Vol. 1, № 1. P. 44–55.
15. Schlotthauer G., Torres M. E., Rufiner H. L. A new algorithm for instantaneous F0 speech extraction based on ensemble empirical mode decomposition // 17th European Signal Processing Conference (Aug. 24–28, 2009). UK, Glasgow, 2009. P. 2347–2351.
16. Bouzid A., Ellouze N. Empirical mode decomposition of voiced speech signal // First International Symposium on Control, Communications and Signal Processing (March 21–24, 2004). Tunisia, Hammamet, 2004. P. 603–606.
17. Sethu V., Ambikairajah E., Epps J. Empirical mode decomposition based weighted frequency feature for speech-based emotion classification // IEEE International Conference on Acoustics, Speech and Signal Processing (31 March – 4 April, 2008). Las Vegas, NV, USA, 2008. P. 5017–5020.
18. Ali M., Prasad R. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition // Renewable and Sustainable Energy Reviews. 2019. Vol. 104. P. 281–295.
19. Prasad R., Ali M., Kwan P., Khan H. Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation // Applied Energy. 2019. Vol. 236. P. 778–792.
20. Huang H., Pan J. Speech pitch determination based on Hilbert-Huang transform // Signal Processing. 2006. Vol. 86, № 4. P. 792–803.
21. Wu J. D., Tsai Y. J. Speaker identification system using empirical mode decomposition and an artificial neural network // Expert Systems with Applications. 2011. Vol. 38, № 5. P. 6112–6117.
22. Huang N. E. Empirical mode decomposition for analyzing acoustical signals. US patent No. 6,862,558 B2, 01.03.2005.
23. Hasan T., Hasan M. K. Suppression of residual noise from speech signals using empirical mode decomposition // IEEE Signal Processing Letters. 2009. Vol. 16, № 1. P. 2–5.
24. Алимуратов А. К., Тычков А. Ю., Чураков П. П. [и др.]. Алгоритм сегментации речь/пауза на основе декомпозиции на эмпирические моды и одномерного расстояния Махаланобиса // Труды МФТИ. 2021. Т. 13, № 3. С. 4–22.
25. Алимуратов А. К. Помехоустойчивый способ сегментации речь/пауза на основе метода декомпозиции на эмпирические моды // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. 2021. № 2. С. 40–66.
26. Chu Y. Y., Xiong W. H., Chen W. Speech Emotion Recognition Based on EMD in Noisy Environments // Advanced Materials Research. 2013. Vol. 831. P. 460–466.
27. Kerkeni L., Serrestou Y., Raoof K. Kerkeni L. [et al.]. Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO // Speech Communication. 2019. Vol. 114. P. 22–35.

28. Shahnaz C., Sultana S., Fattah S.A. [et al.]. Emotion recognition based on EMD-Wavelet analysis of speech signals // IEEE International Conference on Digital Signal Processing (DSP) (July 21–24, 2015). Singapore, 2015. P. 307–310.
29. Li X., Li Xin, Zheng X., Zhang D. EMD-TEO Based Speech Emotion Recognition // Life System Modeling and Intelligent Computing. 2010. P. 180–189.
30. Zhaohua W., Huang N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method // Advances in Adaptive Data Analysis. 2009. № 1. P. 1–41.
31. Colominasa M. A., Schlotthauera G., Torres M. E. Improved complete ensemble EMD: a suitable tool for biomedical signal processing // Biomed. Signal Proces. 2014. Vol. 14. P. 19–29.
32. Алимуратов А. К., Фокина Е. А., Журина А. Е. Исследование влияния длительности анализируемых речевых сигналов на частотно-избирательные свойства декомпозиции на эмпирические моды // Новые информационные технологии и системы : сб. науч. ст. XVI Междунар. науч.-техн. конф. (г. Пенза, 27–29 ноября 2019 г.). Пенза : Изд-во ПГУ, 2019. С. 201–205.
33. National University of Entre Rios. The Laboratory of Signals and Nonlinear Dynamics, Faculty of Engineering. URL: <http://www.bioingenieria.edu.ar/grupos/ldnllys> (дата обращения: 01.05.2021).

References

1. Fant G.K. *Akusticheskaya teoriya recheobrazovaniya = Acoustic theory of speech formation*. Transl. from Engl. by L.A. Varshavskiy, V.I. Medvedev. Moscow: Nauka, 1964:284. (In Russ.)
2. Mikhaylov V.G., Zlatousova L.V. *Izmerenie parametrov rechi = Measurement of speech parameters*. Moscow: Radio i svyaz', 1987:168. (In Russ.)
3. Alimuradov A.K., Churakov P.P. Review and classification of speech signal processing methods in speech recognition systems. *Izmerenie. Monitoring. Upravlenie. Kontrol' = Measurement. Monitoring. Management. Control*. 2015;(2):27–35. (In Russ.)
4. Huang X., Acero A., Hon H.-W. *Spoken Language Processing. Guide to Algorithms and System Development*. New Jersey: Prentice Hall, 2001:980.
5. Li U.A., Neyburg E.P., Martin T.B. [et al.]. *Metody avtomaticheskogo raspoznavaniya rechi: v 2 kn.; per. s angl. = .* Moscow: Mir, 1983;(bk.1):328. (In Russ.)
6. Klett D.Kh., Barnet Dzh.A., Bernsteyn M.I. [et al.]. *Metody avtomaticheskogo raspoznavaniya rechi: v 2 kn.; per. s angl. = Methods of automatic speech recognition : in 2 books. ; translated from English*. Moscow: Mir, 1983;(bk.2):392. (In Russ.)
7. Huang E. Huang, Samuel S.P. Shen. Hilbert-Huang Transform and its application. Interdisciplinary mathematical sciences. *Interdisciplinary Mathematical Sciences*. World Scientific Publishing Company. 2005;(bk.5):324.
8. Huang N.E., Zheng Sh., Steven R.L. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London*. 1998;454:903–995.
9. Mi X., Liu H., Li Y. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. *Energy Conversion and Management*. 2019;80:196–205.
10. Wu Z., Schneider E., Hu Z., Cao L. The impact of global warming on ENSO variability in climate records. *Tech. report, Cent. Ocean. Stud.* 2001;110:25.
11. Lee T., Ouarda T.B.M.J. Multivariate Nonstationary Oscillation Simulation of Climate Indices With Empirical Mode Decomposition. *Water Resources Research*. 2019;55(6):5033–5052.
12. Labate D., Foresta F.La, Occhiuto G. [et al.]. Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A comparison. *IEEE Sensors Journal*. 2013;13(7):2666–2674.
13. Zhang Y., Gao Y., Wang L. [et al.]. The removal of wall components in doppler ultrasound signals by using the empirical mode decomposition algorithm. *IEEE Transactions on Biomedical Engineering*. 2007;54(9):1631–1642.
14. Andrade A.O., Nasuto S., Kyberd P. [et al.]. EMG signal filtering based on Empirical Mode Decomposition. *Biomedical Signal Processing and Control*. 2006;1(1):44–55.
15. Schlotthauer G., Torres M.E., Rufiner H.L. A new algorithm for instantaneous F0 speech extraction based on ensemble empirical mode decomposition. *17th European Signal Processing Conference (Aug. 24–28, 2009)*. UK, Glasgow, 2009:2347–2351.
16. Bouzid A., Ellouze N. Empirical mode decomposition of voiced speech signal. *First International Symposium on Control, Communications and Signal Processing (March 21–24, 2004)*. Tunisia, Hammamet, 2004:603–606.
17. Sethu V., Ambikairajah E., Epps J. Empirical mode decomposition based weighted frequency feature for speech-based emotion classification. *IEEE International Conference on Acoustics, Speech and Signal Processing (31 March – 4 April, 2008)*. Las Vegas, NV, USA, 2008:5017–5020.

18. Ali M., Prasad R. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustainable Energy Reviews*. 2019;104:281–295.
19. Prasad R., Ali M., Kwan P., Khan H. Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. *Applied Energy*. 2019;236:778–792.
20. Huang H., Pan J. Speech pitch determination based on Hilbert-Huang transform. *Signal Processing*. 2006;86(4):792–803.
21. Wu J.D., Tsai Y.J. Speaker identification system using empirical mode decomposition and an artificial neural network. *Expert Systems with Applications*. 2011;38(5):6112–6117.
22. Huang N.E. *Empirical mode decomposition for analyzing acoustical signals*. US patent No. 6,862,558 B2, 01.03.2005.
23. Hasan T., Hasan M.K. Suppression of residual noise from speech signals using empirical mode decomposition. *IEEE Signal Processing Letters*. 2009;16(1):2–5.
24. Alimuradov A.K., Tychkov A.Yu., Churakov P.P. [et al.]. Speech/pause segmentation algorithm based on decomposition into empirical modes and one-dimensional Mahalanobis distance. *Trudy MFTI = Proceedings of MIPT*. 2021;13(3):4–22. (In Russ.)
25. Alimuradov A.K. Noise-resistant method of segmentation speech/pause based on the method of decomposition into empirical modes. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotehnika, informatsionnye tekhnologii, sistemy upravleniya = Bulletin of Perm National Research Polytechnic University. Electrical engineering, information technology, control systems*. 2021;(2):40–66. (In Russ.)
26. Chu Y.Y., Xiong W.H., Chen W. Speech Emotion Recognition Based on EMD in Noisy Environments. *Advanced Materials Research*. 2013;831:460–466.
27. Kerkeni L., Serrestou Y., Raoof K. Kerkeni L. [et al.]. Automatic speech emotion recognition using an optimal combination of features based on EMD-TKEO. *Speech Communication*. 2019;114:22–35.
28. Shahnaz C., Sultana S., Fattah S.A. [et al.]. Emotion recognition based on EMD-Wavelet analysis of speech signals. *IEEE International Conference on Digital Signal Processing (DSP) (July 21–24, 2015)*. Singapore, 2015:307–310.
29. Li X., Li Xin, Zheng X., Zhang D. EMD-TEO Based Speech Emotion Recognition. *Life System Modeling and Intelligent Computing*. 2010:180–189.
30. Zhaohua W., Huang N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*. 2009;(1):1–41.
31. Colominasa M.A., Schlotthauera G., Torres M.E. Improved complete ensemble EMD: a suitable tool for biomedical signal processing. *Biomed. Signal Proces.* 2014;14:19–29.
32. Alimuradov A.K., Fokina E.A., Zhurina A.E. Investigation of the influence of the duration of the analyzed speech signals on the frequency-selective properties of decomposition on empirical modes. *Novye informatsionnye tekhnologii i sistemy: sb. nauch. st. XVI Mezhdunar. nauch.-tekhn. konf. (g. Penza, 27–29 noyabrya 2019 g.) = New information technologies and systems : collection of scientific Articles XVI International Scientific-technical. conf. (Penza, November 27-29, 2019)*. Penza: Izd-vo PGU, 2019:201–205. (In Russ.)
33. *National University of Entre Ríos. The Laboratory of Signals and Nonlinear Dynamics, Faculty of Engineering*. Available at: <http://www.bioingenieria.edu.ar/grupos/ldnlys> (accessed 01.05.2021).

Информация об авторах / Information about the authors

Алан Казанферович Алимуратов

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

Alan K. Alimuradov

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

Александр Юрьевич Тычков

доктор технических наук,
заведующий кафедрой радиотехники
и радиоэлектронных систем,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: tychkov-a@mail.ru

Alexander Yu. Tychkov

Doctor of technical sciences,
head of the sub-department of radio engineering
and radio electronic systems,
Penza State University
(40 Krasnaya street, Penza, Russia)

Петр Павлович Чураков

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

Petr P. Churakov

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

Анастасия Валерьевна Баранова

студентка,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: anastasiyabar@mail.ru

Anastasiya V. Baranova

Student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Дмитрий Сергеевич Дудников

студент,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: dmitriy.s.gmpf@gmail.com

Dmitriy S. Dudnikov

Student,
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 5.02.2022

Поступила после рецензирования/Revised 9.03.2022

Принята к публикации/Accepted 10.04.2022