УДК 004.934
doi:10.21685/2587-7704-2022-7-1-4

Open Access | RESEARCH ARTICLE

Новый подход к обработке речевых сигналов на основе метода декомпозиции на эмпирические моды

Алан Казанферович Алимурадов
Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40
alansapfir@yandex.ru

Богдан Андреевич Порезанов
Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40
bogdan.porezanov@yandex.ru

Илья Олегович Стешкин
Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40
ilya_steshkin@mail.ru

Кирилл Егорович Платонов
Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40
platonov.1408@mail.ru

Дмитрий Сергеевич Дудников
Пензенский государственный университет, Россия, г. Пенза, ул. Красная, 40
dmitriy.s.gmpf@gmail.com

Аннотация. Представлен новый подход к обработке речевых сигналов, основанный на адаптивном методе частотно-временного анализа – декомпозиции на эмпирические моды. Подход основан на равномерном делении исходного речевого сигнала на фрагменты, декомпозиции фрагментов на эмпирические моды и формировании новых модовых речевых сигналов. Целью разработки нового подхода является расширение пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик исходного речевого сигнала. Представлено краткое описание модификаций методов декомпозиции на эмпирические моды, их преимущества и недостатки. Подробно описан функционал предлагаемого подхода и представлены результаты исследования.

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

Для цитирования: Алимурадов А. К., Порезанов Б. А., Стешкин И. О., Платонов К. Е., Дудников Д. С. Новый подход к обработке речевых сигналов на основе метода декомпозиции на эмпирические моды // Инжиниринг и технологии. 2022. Т. 7(1). С. 1–6. doi:10.21685/2587-7704-2022-7-1-4

A novel EMD-based approach to speech signal processing

Alan K. Alimuradov
Penza State University, 40 Krasnaya Street, Penza, Russia
alansapfir@yandex.ru

Bogdan A. Porezanov
Penza State University, 40 Krasnaya Street, Penza, Russia
bogdan.porezanov@yandex.ru

Il'ya O. Steshkin
Penza State University, 40 Krasnaya Street, Penza, Russia
ilya_steshkin@mail.ru

Kirill E. Platonov
Penza State University, 40 Krasnaya Street, Penza, Russia
platonov.1408@mail.ru

Dmitriy S. Dudnikov
Penza State University, 40 Krasnaya Street, Penza, Russia
dmitriy.s.gmpf@gmail.com

Abstract. The article presents a novel approach to speech signal processing based on the empirical mode decomposition (EMD), being an adaptive time-frequency analysis method. The proposed approach is based on the uniform splitting of the original speech signal into fragments, the decomposition of fragments into empirical modes, and the formation of new mode speech signals. The goal of approach elaboration is to expand the space for informatively significant amplitude, time, frequency, and energy characteristics of the original speech signal. A brief description of various types of empirical mode decomposition has been presented, and their advantages and disadvantages have been revealed. The functionality of the proposed approach has been detailed, and the research outcomes have been reported.

Keywords: speech signal processing, empirical mode decomposition, human psychoemotional state, emotions

For citation: Alimuradov A. K., Porezanov B. A., Steshkin I. O., Platonov K. E., Dudnikov D. S. A novel EMD-based approach to speech signal processing. *Inzhiniring i tekhnologii = Engineering and Technology*. 2022;7(1):1–6. (In Russ.). doi:10.21685/2587-7704-2022-7-1-4

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: фильтрация шума (линейная и адаптивная), усиление, сегментация на информативные участки, извлечение информативных параметров, кодирование, сжатие, восстановление и др. [2].

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 [3]. Преимуществом данных решений является возможность разложения исследуемых речевых сигналов на составляющие для последующего детализированного анализа.

В последнее время широкое практическое применение в решениях задач по обработке речевых сигналов получило преобразование Гильберта – Хуанга [4], в основе которого заложен метод декомпозиции на эмпирические моды (ДЭМ) [5]. ДЭМ – это уникальная технология разложения на частотные составляющие, не требующая априорной информации об анализируемом сигнале.

В данной статье представлен новый подход к обработке речевых сигналов, в котором используется метод ДЭМ. Предлагаемый подход основан на равномерном делении исходного речевого сигнала на фрагменты, декомпозиции фрагментов на эмпирические моды (ЭМ) и формировании новых модовых речевых сигналов. Целью разработки нового подхода является расширение пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик исходного речевого сигнала. В основе расширения информативного пространства заложен принцип, что каждый новый модовый речевой сигнал содержит в себе скрытые особенности внутренней структуры исходного речевого сигнала (скрытые модуляции, области концентрации энергии и т.п.).

Статья является результатом научной работы коллектива авторов [6, 7], 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.

Подробный анализ известных методов ДЭМ, применяемых для анализа сигналов естественной природы, выявил, что наиболее адаптивными к нестационарной речи являются множественная ДЭМ (МДЭМ) [8] и улучшенная полная МДЭМ с адаптивным шумом (ПМДЭМАШ) [9].

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). Аналитические выражения методом МДЭМ и улучшенной ПМДЭМАШ представлены ниже:

$$x_j(n) = x(n) + w_j(n),$$

где $x_j(n)$ – зашумленные сигналы; n – дискретный отсчет времени; $x(n)$ – исходный речевой сигнал; $w_j(n)$ – белый шум малой амплитуды; j = 1, 2, …, J – количество реализаций белого шума;

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

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

$$r_I(n) = \sum_{j=1}^{J} \frac{r_{jI}(n)}{J},$$

где $IMF(n)$ – ЭМ; $r(n)$ – конечный неделимый остаток; i = 1, 2, …, 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).

Фрагментирование представляет собой процесс линейного разделения исходного речевого сигнала $x(n)$ на отрезки одинаковой длительности, которые записываются в отдельные переменные $x_s(n)$:

$$S = \frac{N}{L},$$

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

$$x_{s+1}(n) = x\left[(s \cdot L) + 1 : (s+1) \cdot L\right],$$

где s = 0, 1, 2, … S – номер фрагмента.

Как отмечалось ранее, наиболее адаптивными к нестационарной речи являются методы МДЭМ и улучшенной ПМДЭМАШ. Отличительной особенностью метода улучшенной ПМДЭМАШ от метода МДЭМ 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.

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:

$$xmode_i(n) = \sum_{s=1}^{S} IMF_{s,i}\left[(s \cdot L) + 1 : (s+1) \cdot L\right],$$

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

Количество сформированных модовых речевых сигналов зависит от количества используемых информативных ЭМ, полученных для каждого фрагмента.

Суть исследования заключается в изменении параметров функционирования предлагаемого подхода и анализе полученных результатов. В табл. 1 представлены наименования настраиваемых и исследуемых параметров предлагаемого подхода обработки речевых сигналов на основе методов ДЭМ.

На рис. 1–3 представлены усредненные результаты исследования нового подхода обработки речевых сигналов.

Таблица 1

Настраиваемые и исследуемые параметры нового подхода 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	Разница между исходным и реконструированным сигналами (В)
Параметры МДЭМ и улучшенной ПМДЭМАШ: NR (в разах) – 5, 50; MaxIter (в разах) – 10, 100	Время формирования набора модовых речевых сигналов (с)

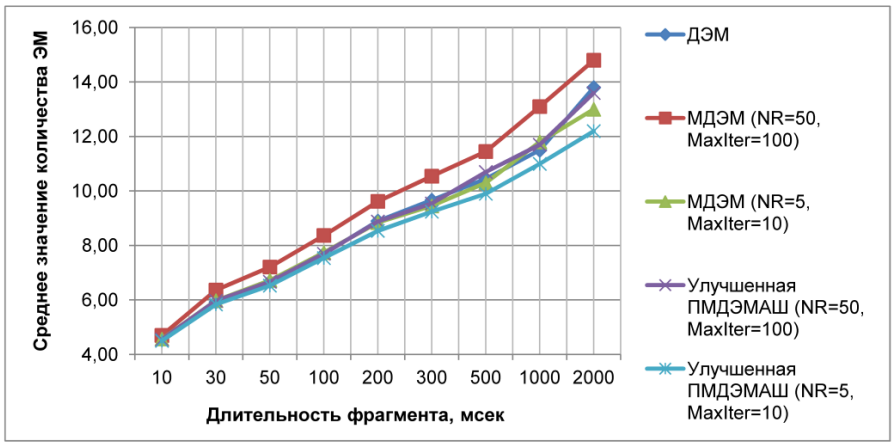

Рис. 1. Среднее значение количества ЭМ

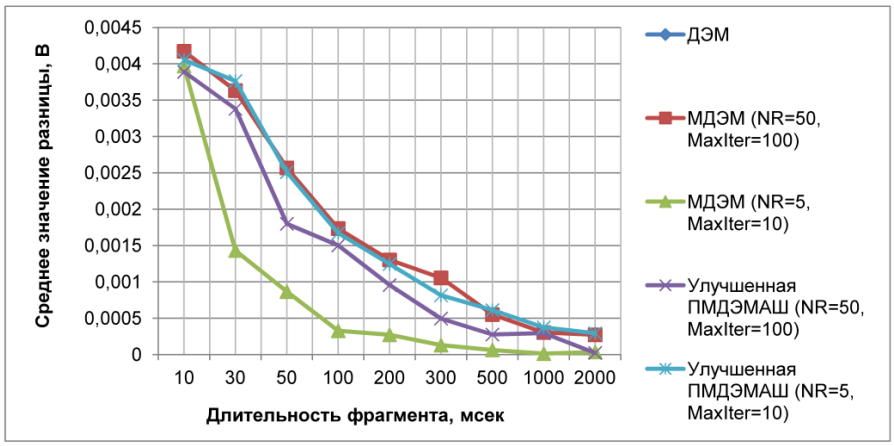

Рис. 2. Разница между исходным и реконструированным сигналами

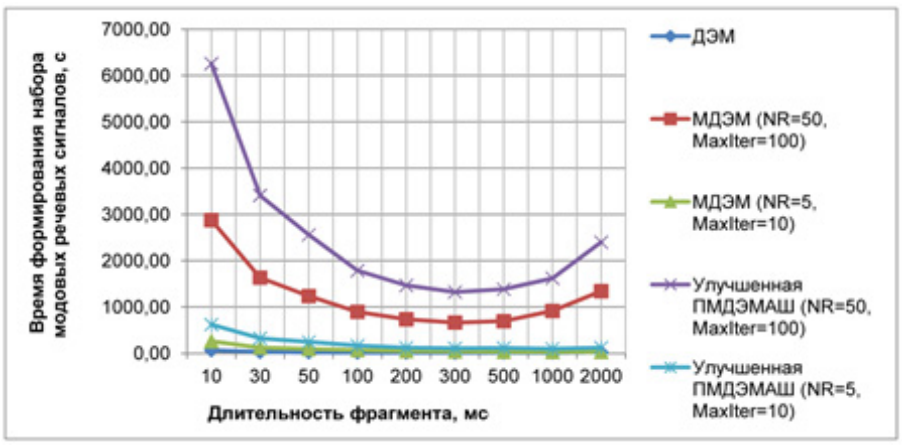

Рис. 3. Время формирования набора модовых речевых сигналов

В соответствии с результатами на рис. 1–3 выявлены закономерности и определены следующие оптимальные значения функционирования предлагаемого нового подхода.

1. Длительность анализируемого фрагмента – от 300 до 1000 мс. В этом случае необходимо минимальное время для формирования набора модовых речевых сигналов (рис. 3).

2. Количество ЭМ – от 8 до 10. В этом случае обеспечивается необходимая и достаточная полнота разложения (рис. 1). Определяется эмпирически по минимальной разнице между исходным и реконструированным сигналами.

3. Разница 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 (рис. 2).

Подводя итоги анализа результатов исследований, можно сделать основной вывод: предлагаемый новый подход обработки речевых сигналов на основе методов ДЭМ в действительности может обеспечить расширение пространства информативно-значимых амплитудных, временных, частотных и энергетических характеристик. Расширение информативного пространства обеспечивается за счет формирования набора новых модовых речевых сигналов (с минимальной ошибкой), содержащих в себе особенности внутренней структуры исходного речевого сигнала (скрытые модуляции, области концентрации энергии и т.п.).

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. Huang X., Acero A., Hon H.-W. Spoken Language Processing. Guide to Algorithms and System Development. New Jersey : Prentice Hall, 2001. 980 p.
4. Huang N. E. Shen Samuel S. P. Hilbert-Huang Transform and its application // Interdisciplinary Mathematical Sciences. Vol. 5. Singapore : World Scientific Publishing Company, 2005. 324 p.
5. 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.
6. Алимурадов А. К., Тычков А. Ю., Чураков П. П., Агейкин А. В., Кулешов А. П., Чернов И. А. Алгоритм сегментации речь/пауза на основе декомпозиции на эмпирические моды и одномерного расстояния Махаланобиса // Труды МФТИ. 2021. Т. 13, № 3 (51). С. 4–22.
7. Алимурадов 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 (38). С. 40–66.
8. Zhaohua W., Huang N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method // Advances in Adaptive Data Analysis. 2009. № 1 (1). P. 1–41.
9. 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.

References

1. Fant G.K. *Akusticheskaya teoriya recheobrazovaniya = Acoustic theory of speech production.* Translated from English 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. Huang X., Acero A., Hon H.-W. *Spoken Language Processing. Guide to Algorithms and System Development*. New Jersey: Prentice Hall, 2001:980.
4. Huang N.E., Shen Samuel S.P. *Hilbert-Huang Transform and its application*. Singapore: World Scientific Publishing Company, 2005:324. (Interdisciplinary Mathematical Sciences. Vol. 5)
5. 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.
6. Alimuradov A.K., Tychkov A.Yu., Churakov P.P., Ageykin A.V., Kuleshov A.P., Chernov I.A. Speech/pause segmentation algorithm based on empirical mode decomposition and one-dimensional Mahalanobis distance. *Trudy Moskovskogo fiziko-tekhnicheskogo instituta (Natsional'nogo issledovatel'skogo universiteta) = Proceedings of Moscow Institute of Physics and Technology*. 2021;13(3):4–22. (In Russ.)
7. Alimuradov A.K. EMD-based noise-robust method for 'speech/pause' segmentation. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniya = Bulletin of Perm National Research Polytechnic University. Electrotechnics, Informational Technologies, Control Systems*. 2021;(2):40–66. (In Russ.)

8. Zhaohua W., Huang N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*. 2009;(1):1–41.
9. 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.

Поступила в редакцию / Received 18.03.2022

Поступила после рецензирования и доработки / Revised 19.04.2022

Принята к публикации / Accepted 11.05.2022