

УДК 621.391; 519.21

*А. К. Алимуратов, Ю. С. Квитка***ПРИМЕНЕНИЕ КОМПЛЕМЕНТАРНОЙ МНОЖЕСТВЕННОЙ
ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ
ДЛЯ АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ***A. K. Alimuradov, Yu. S. Kvitka***APPLICATION COMPLEMENTARY ENSEMBLE EMPIRICAL
MODE DECOMPOSITION FOR ANALYSIS OF SPEECH SIGNALS**

А н н о т а ц и я. Рассмотрены проблемы применения адаптивных методов обработки речевых сигналов. Представлены теоретические исследования методов декомпозиции на эмпирические моды для анализа нелинейных и нестационарных сигналов, обоснование метода комплементарной множественной декомпозиции на эмпирические моды (КМДЭМ) для решения задач обработки речевых сигналов. Приведены результаты экспериментальных исследований, направленных на определение оптимальных параметров работы метода КМДЭМ для речевых сигналов.

A b s t r a c t. The paper considers the problem of applying adaptive methods of speech signal processing. Theoretical studies on the decomposition methods for the analysis of the empirical mode of nonlinear and non-stationary signals. The substantiation of the method of complementary ensemble empirical mode decomposition (CEEMD) for solving the problems of speech signal processing. The results of experimental studies aimed at determining the optimal parameters of the method KMDEM for speech signals.

К л ю ч е в ы е с л о в а: речевые технологии, адаптивная обработка, речевые сигналы, декомпозиция на эмпирические моды.

К e y w o r d s: speech technologies, adaptive processing, voice signals, empirical mode decomposition.

В настоящее время понятие «обработка речевых сигналов» имеет достаточно широкое значение, которое включает в себя вид, задачи, способы и методы, алгоритмы, средства и инструменты обработки.

Речевые сигналы представляют собой нестационарные (квазипериодические) и нелинейные сигналы сложной формы, состоящие из звуков, отождествляемых значимые единицы языка – фонемы – в предложения и слова. При обработке речевых сигналов, как правило, оперируют не исходным сигналом, а его параметрами, вычисленными с помощью его цифрового представления. Все информативные параметры можно разделить на три группы: амплитудно-частотные, спектрально-временные и кепстральные.

Для инженеров-исследователей и разработчиков наибольший интерес вызывает поиск адаптивных методов анализа и обработки речевых сигналов. На сегодняшний день для их анализа используются временная, частотная и частотно-временная области обработки [1]. Во временной области применяются методы с использованием корреляционной и автокорреляционной функций, в частотной области – методы, основанные на использовании преобразования Фурье и нахождения кепстра. Оба подхода не адаптированы ко всем задачам обработки

речевых сигналов. Так, из-за нестационарности они не позволяют фиксировать кратковременные локальные изменения в сигналах, что приводит к некорректному выявлению уникальных свойств и информативных параметров речи. Немного лучше ситуация обстоит в частотно-временной области, где применяется метод вейвлет-преобразования. Его особенность – использование базисов, с помощью которых можно обрабатывать нестационарные сигналы [2]. Недостатком вейвлет-преобразования является сложность в подборе адаптированного базиса из-за особенностей, свойственных речевому сигналу: зависимость параметров и характеристик от эмоциональной составляющей, тембра, высоты, громкости и длительности сигнала.

Все вышеперечисленное подчеркивает актуальность выбора нового метода адаптивной обработки и анализа речевых сигналов.

Декомпозиция на эмпирические моды

При обработке нестационарных и нелинейных сигналов, таких как речевые, необходимо проводить кратковременный анализ локальных явлений. Математический аппарат – декомпозиция на эмпирические моды (ДЭМ) – представляет собой итерационную вычислительную процедуру отсеивания, в результате которой сигнал раскладывается на эмпирические моды (ЭМ), или внутренние колебания [3]. Особенность ДЭМ заключается в том, что базисные функции, используемые для разложения в процедуре отсеивания, извлекаются непосредственно из исходного сигнала. Разложение на ЭМ позволяет анализировать кратковременные локальные изменения в сигнале, поэтому данный метод может быть использован при обработке нелинейных и нестационарных сигналов. При декомпозиции модель сигнала не задается заранее – ЭМ вычисляются в ходе процедуры отсеивания: по характерным точкам сигнала, например, по экстремумам, строится функция, называемая огибающей сигнала. У каждого сигнала имеются локальные экстремумы: локальные максимумы и минимумы. В результате строятся две огибающие: нижняя, построенная по точкам локального минимума, и верхняя, построенная по точкам локального максимума. При построении огибающих в качестве приближающих функций используются кубические сплайны. Сама же эмпирическая мода обладает двумя свойствами:

1. Количество экстремумов и количество пересечений нуля не должны отличаться более чем на единицу.

2. Среднее значение, которое определяется по двум огибающим, должно быть равно нулю.

В общем виде алгоритм метода выглядит следующим образом [4]:

1. Определение локальных экстремумов (максимумов и минимумов) сигнала $f_j(t_i)$:

– значение i -го отсчета $f_j(t_i)$ является локальным максимумом, если выполняется условие $f_j(t_{i-1}) < f_j(t_i) \geq f_j(t_{i+1})$;

– значение i -го отсчета $f_j(t_i)$ является локальным минимумом, если выполняется условие $f_j(t_{i-1}) > f_j(t_i) \leq f_j(t_{i+1})$, где t_i – дискретные отсчеты времени.

2. Определение верхней $e_j(t_i)$ и нижней $g_j(t_i)$ огибающих сигнала с помощью кубической сплайн-интерполяции по найденным локальным экстремумам $f_j(t_i)$:

$$e_j(t_i) = a_B t_i^3 + b_B t_i^2 + c_B t_i + d_B; \quad (1)$$

$$g_j(t_i) = a_H t_i^3 + b_H t_i^2 + c_H t_i + d_H, \quad (2)$$

где a_B, b_B, c_B, d_B – коэффициенты для каждого значения i -го отсчета верхней огибающей сигнала; a_H, b_H, c_H, d_H – коэффициенты для каждого значения i -го отсчета нижних огибающих сигнала.

3. Вычисление среднего значения огибающих сигнала в соответствии с выражением

$$h_j(t_i) = \frac{e_j(t_i) + g_j(t_i)}{2}, \quad (3)$$

где $h_j(t_i)$ – среднее значение огибающих сигнала; $e_j(t_i)$ и $g_j(t_i)$ – верхняя и нижняя огибающие исходных сигналов соответственно.

4. Вычисление остатка сигнала по формуле

$$s_j(t_i) = f_j(t_i) - h_j(t_i), \quad (4)$$

где $s_j(t_i)$ – остаток сигнала.

5. Вычисление значения критерия останова. В качестве критерия останова декомпозиции используется значение нормализованной квадратичной разности, определяемое как

$$SD = \sum_1^n \left[\frac{(|f_j(t_i) - s_j(t_i)|)^2}{f_j(t_i)^2} \right]. \quad (5)$$

6. Проверка условия останова. На этом этапе осуществляется сравнение значения остатка сигнала со значением нормализованной квадратичной разности. Если $SD > s_j(t_i)$, то переходят к выполнению действия 1; если $SD < s_j(t_i)$ и $h_j(t_i) > s_j(t_i)$, то переходят к выполнению следующего действия.

7. Вывод мод сигнала. На этом этапе осуществляется вывод мод $f_i(t)$ и остатка $r_v(t)$ сигнала.

Недостатками ДЭМ являются:

- частое смешивание нескольких мод, которые интерпретируются как одна;
- получение ЭМ, состоящих из участков сигнала несоизмеримых масштабов или участков соизмеримых масштабов, но находящихся в различных точках ЭМ.

Данные недостатки являются следствием нестационарности исходного речевого сигнала.

Множественная декомпозиция на эмпирические моды

Для решения обозначенной проблемы предложен новый метод анализа речевых сигналов, основанный на многократном добавлении к сигналу белого шума и вычислении среднего значения ЭМ как конечного истинного результата. Данный метод называется множественной декомпозицией на эмпирические моды (МДЭМ) [5]. Конечный, не бесконечно малый по амплитуде белый шум делает сигналы в ЭМ сопоставимыми со значениями участков сигнала несоизмеримых масштабов и находящихся в различных точках ЭМ для получения всех возможных решений в процессе отсеивания. Таким образом, анализируемый сигнал представляет собой объединение сигнала и шума:

$$f_j(t_i) = x_j(t_i) + n_j(t_i), \quad (6)$$

где $x_j(t_i)$ – исходный речевой сигнал; $n_j(t_i)$ – белый шум.

Данный подход в полной мере использует преимущество статистических характеристик белого шума для обнаружения слабых периодических или квазипериодических сигналов. Таким образом, МДЭМ является более точным способом анализа данных, чем ДЭМ, посредством добавления белого шума.

Комплементарная множественная декомпозиция на эмпирические моды

Использование белого шума в МДЭМ для решения проблемы смешивания ЭМ порождает в сигнале новый шумовой остаток, влияющий на реконструкцию исходного сигнала. По этой причине был предложен улучшенный метод МДЭМ с внесением дополнительного белого шума, называющегося комплементарной множественно эмпирической декомпозицией [6]. В этом методе остаток белого шума может быть удален с помощью пар дополнительного множества ЭМ с прямыми и инверсными значениями добавленного белого шума:

$$\begin{bmatrix} M_1 \\ M_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} * \begin{bmatrix} s \\ n \end{bmatrix}, \quad (7)$$

где s – первоначальные данные; n – добавленный белый шум; M_1 – сумма первоначальных данных с белым шумом; M_2 – сумма первоначальных данных с инверсным белым шумом.

В КМДЭМ множество ЭМ, полученных при добавлении белого шума, порождает ЭМ с прямым остатком добавленных белых шумов, точно так же множество ЭМ, полученных при добавлении инверсного белого шума, порождает ЭМ с инверсным остатком добавленных белых шумов. Таким образом, итоговое множество ЭМ представляет собой ЭМ с прямыми и отрицательными шумами. Сравнивая КМДЭМ с МДЭМ можно заключить, что в первом случае возможна значительная экономия времени, так как парные шумы эффективно ликвидируют остаток белого шума в ЭМ полностью независимо от того, сколько шумов использовалось.

Следует выделить общие правила в МДЭМ и КМДЭМ:

1. Добавление шума в МДЭМ может заставить высокочастотную составляющую быть «замаскированной», если количество циклов не будет достаточно большим.
2. Энергетический уровень добавленного шума должен быть на уровне встроенного шума (если уровень известен), в противном случае добавленный уровень помех не должен быть больше, чем 20 % от исходного сигнала.

Важными исходными параметрами для МДЭМ и КМДЭМ является амплитудный (или энергетический) уровень добавляемого белого шума (ϵ) и количество итераций вычисления ЭМ (N).

Применение декомпозиции для анализа речевых сигналов

Принимая во внимание преимущества и недостатки вышеизложенных методов декомпозиции, провели экспериментальное исследование метода КМДЭМ как наиболее адаптивного к нестационарным и нелинейным сигналам. Исследование направлено на определение оптимальных значений уровня добавляемого белого шума и количества итераций вычисления ЭМ для последующего эффективного анализа речевых сигналов.

Для исследования авторами рассмотрен речевой сигнал – отрезок звука согласной звонкой фонемы «а» длительностью 30 мс. Длительность одной фонемы варьируется от 10 до 25 мс, поэтому обработка речевых сигналов осуществляется последовательно отрезками в среднем по 10–30 мс. Считается, что речевые сигналы такой длительности сохраняют целостность слуховой информации фонемы. Речевой сигнал зарегистрирован с посторонними шумами с соотношением сигнал/шум 30 дБ, частота дискретизации 8000 Гц, 16 бит [7] (рис. 1).

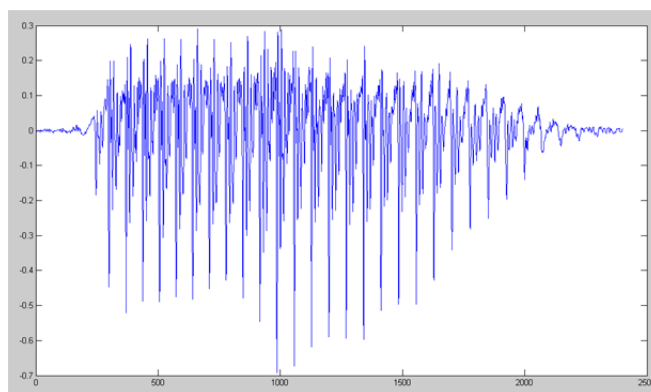


Рис. 1. Речевой сигнал – звук согласной звонкой фонемы «а»

Исследование проводилось в пакете прикладных программ для решения технических и математических задач – MATLAB 7.0.1 – и состояло из двух частей:

1. Определение оптимального значения добавляемого белого шума проводилось с уровнями шума – 0,25, 0,5, 1,0 и 3,0.
2. Определение количества итераций вычисления ЭМ проводилось со значениями – 10, 100, 1000.

На рис. 2–5 приведены результаты исследования.

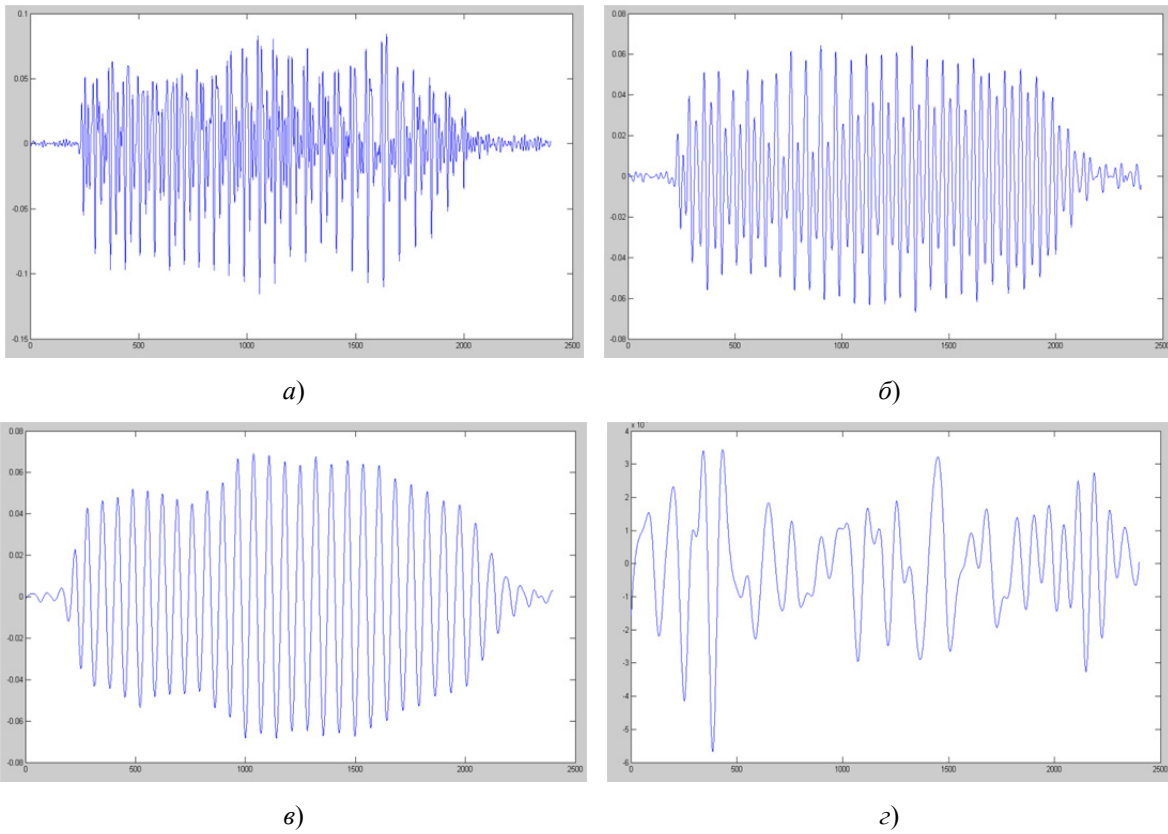


Рис. 2. Эмпирические моды 3 – 6 при $\varepsilon = 0,25$; $N = 100$:
 а – ЭМ 3; б – ЭМ 4; в – ЭМ 5; г – ЭМ 6

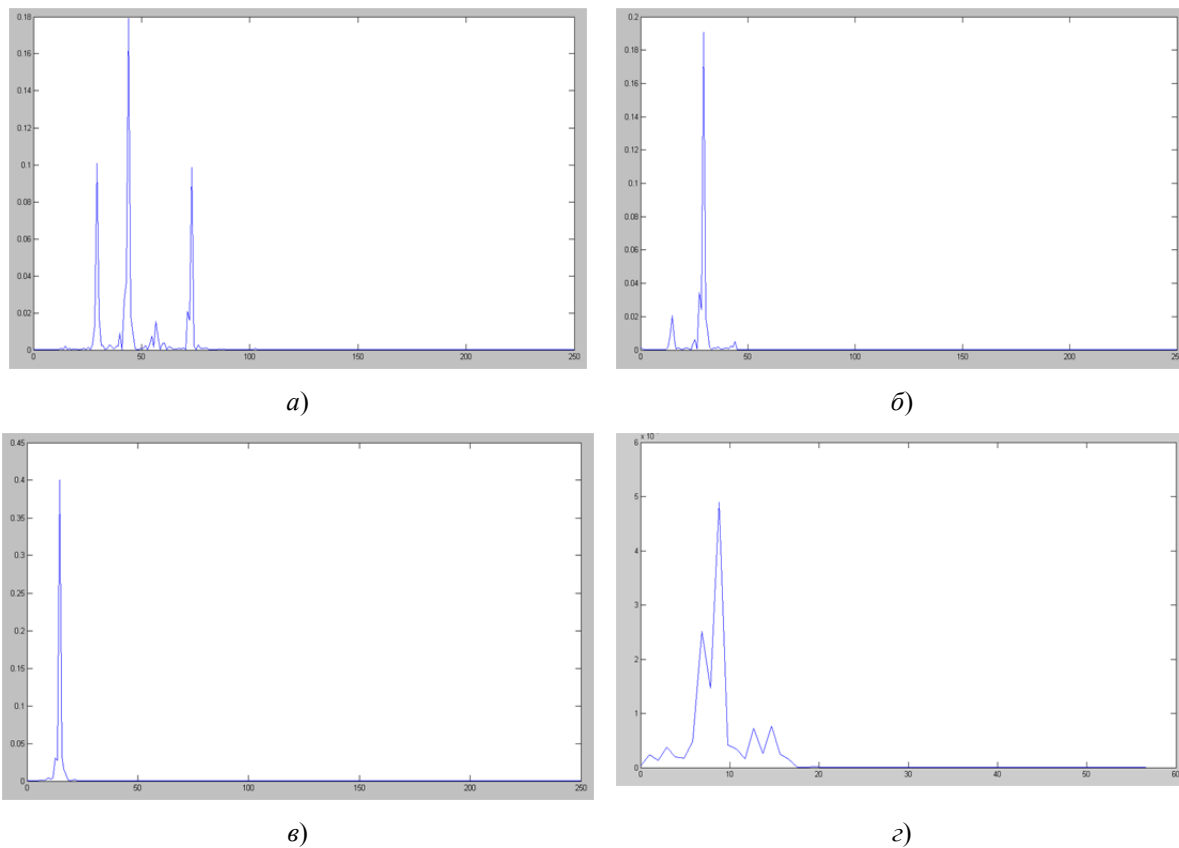


Рис. 3. Спектральное представление эмпирических мод 3 – 6 при $\varepsilon = 0,25$; $N = 100$:
 а – спектральное представление ЭМ 3; б – спектральное представление ЭМ 4;
 в – спектральное представление ЭМ 5; г – спектральное представление ЭМ 6

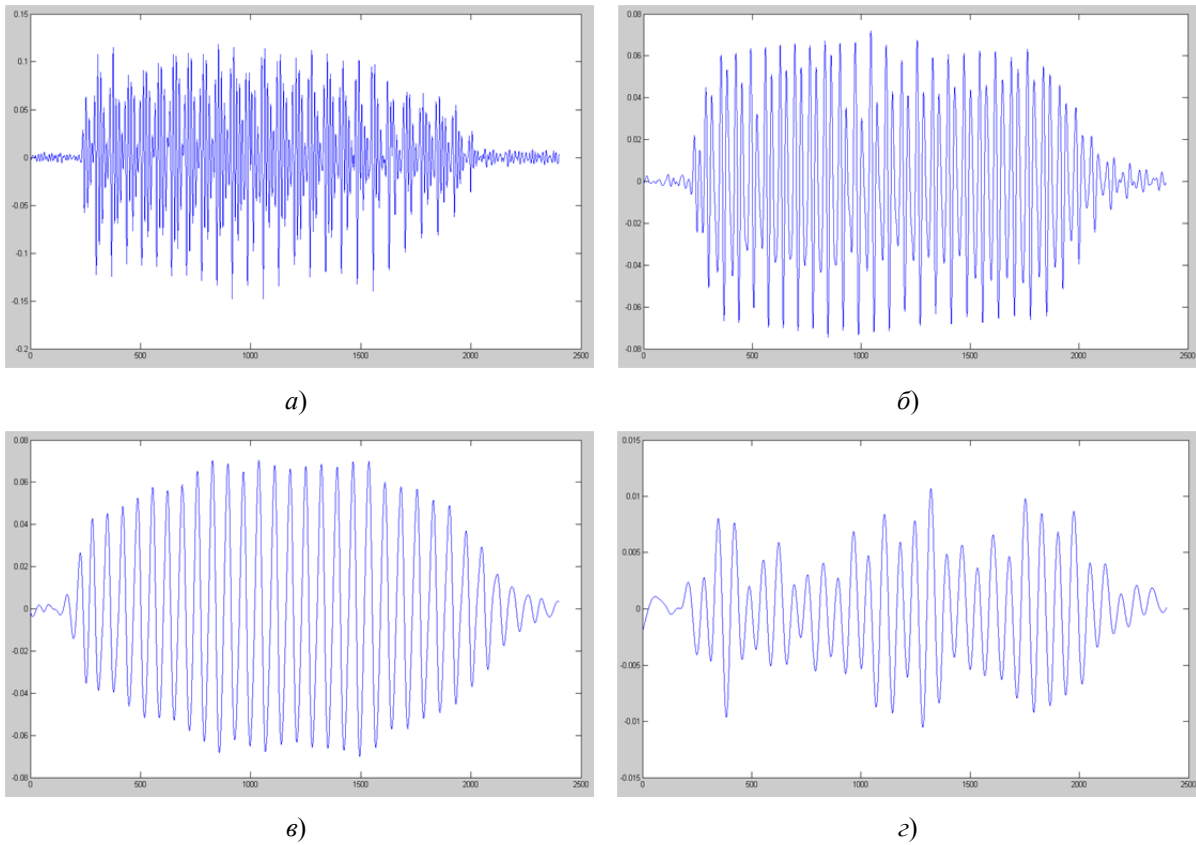


Рис. 4. Эмпирические моды 3 – 6 при $\varepsilon = 0,5$; $N = 100$:
 а – ЭМ 3; б – ЭМ 4; в – ЭМ 5; г – ЭМ 6

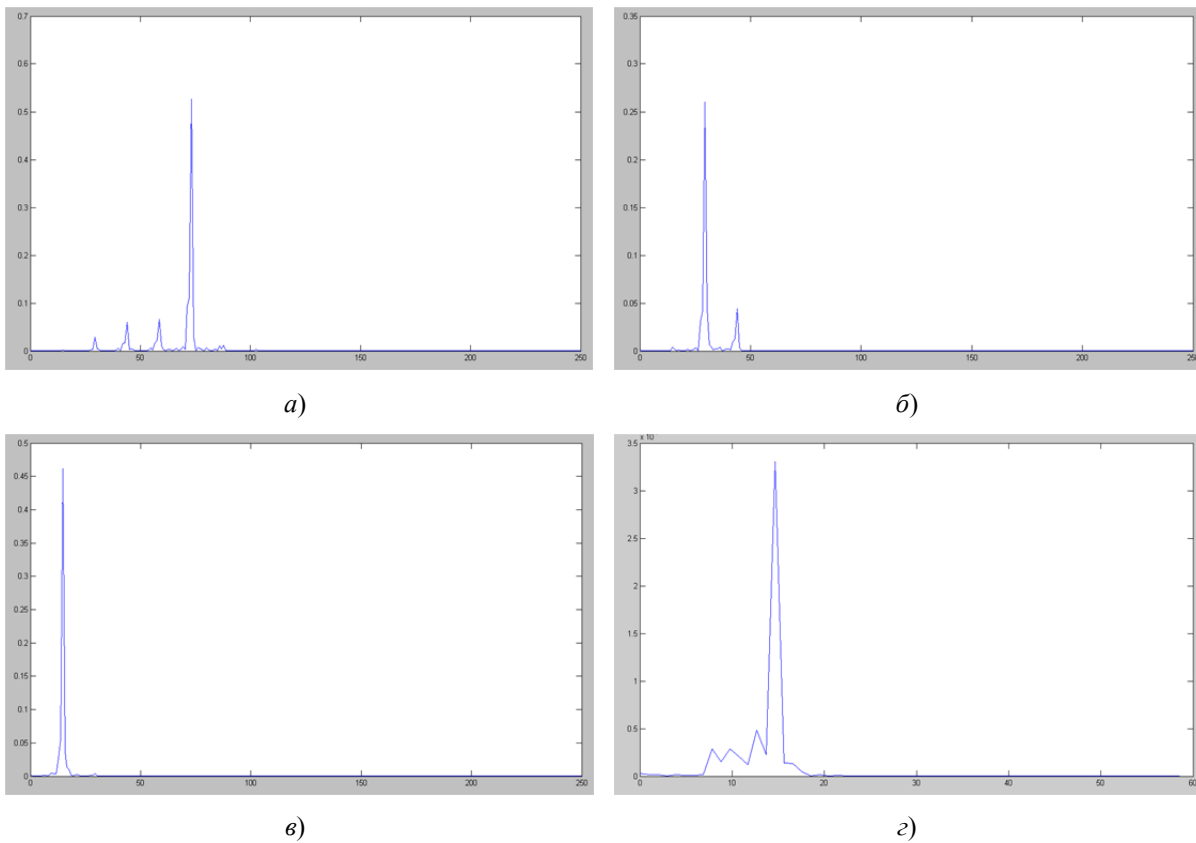


Рис. 5. Спектральное представление эмпирических мод 3 – 6 при $\varepsilon = 0,5$; $N = 100$:
 а – спектральное представление ЭМ 3; б – спектральное представление ЭМ 4;
 в – спектральное представление ЭМ 5; г – спектральное представление ЭМ 6

Согласно полученным результатам наиболее оптимальными значениями белого шума и количества итераций являются $\varepsilon = 0,25 - 0,5$ и N – не более 100. При анализе учитывались только информативные ЭМ 3 – ЭМ 6 [4, 8]. Эффективность декомпозиции метода КМДЭМ оценивалась по качеству выделения частотных составляющих исходного речевого сигнала в отдельные ЭМ, а также по наличию паразитных частотной и амплитудной модуляций в ЭМ.

Заключение

Итогом проведенной работы являются следующие теоретические и практические результаты:

- проведены теоретические исследования адаптивных методов декомпозиции на эмпирические моды для применения в задачах обработки речевых сигналов;
- проведены экспериментальные исследования, направленные на определение наиболее оптимальных значений уровня добавляемого белого шума и количества итераций вычисления ЭМ для эффективного анализа речевых сигналов.

Список литературы

1. Сергиенко, А. Б. Цифровая обработка сигналов / А. Б. Сергиенко. – СПб. : Питер, 2002. – 608 с.
2. Рабинер, Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер. – М. : Радио и связь, 1981. – 496 с.
3. Huang, N. E. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis / N. E. Huang // Proc. R. Soc. Lond. A. – 1998. – Т. 454. – С. 903–995.
4. Алимуратов, А. К. Фильтрация речевых сигналов с использованием метода множественной декомпозиции и оценки энергии эмпирических мод / А. К. Алимуратов, П. П. Чураков, А. Ю. Тычков // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2012. – № 4. – С. 50–61.
5. Wu, Z. Ensemble empirical mode decomposition a noise-assisted data analysis method / Z. Wu, N. E. Huang // Calverton, MD 20705, USA, NASA Goddard Space Flight Center Greenbelt.
6. Yeh, J.-R. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method / J.-R. Yeh, J.-S. Shieh, N. E. Huang. Adv. Adapt. Data Anal. – 2010. – № 2. – P. 135–156.
7. Алимуратов, А. К. Исследование и разработка верифицированной базы звуков речи для анализа речевых сигналов / А. К. Алимуратов, А. Ю. Тычков, Ю. С. Квитка, Д. А. Ярославцева // Современные проблемы компьютерных наук (СПКН-2013) : сб. материалов I Междунар. науч.-практ. конф., посвящ. 70-летию образования Пензенского государственного университета / под ред. Л. Р. Фионовой. – Пенза, 2013. – С. 156–159.
8. Алимуратов, А. К. Интеллектуальная обработка речевых сигналов в системах автоматического управления / А. К. Алимуратов, А. Ю. Тычков // Известия Кабардино-балкарского государственного университета. – 2012. – Т. 2. – № 5. – С. 66–67.

Алимуратов Алан Казанферович

соискатель,
кафедра информационно-измерительной техники,
Пензенский государственный университет
E-mail: alansarfir@yandex.ru

Alimuradov Alan Kazanferovich

applicant,
sub-department of information
and measuring equipment,
Penza State University

Квитка Юрий Сергеевич

аспирант,
Пензенский государственный университет
E-mail: aravikleur@gmail.com

Kvitka Yuriy Sergeevich

postgraduate student,
Penza State University

УДК 621.391; 519.21

Алимуратов, А. К.

Применение комплементарной множественной декомпозиции на эмпирические моды для анализа речевых сигналов / А. К. Алимуратов, Ю. С. Квитка // Измерение. Мониторинг. Управление. Контроль. – 2014. – № 4 (10). – С. 69–75.