УДК 621.391.8

АНАЛИЗ СИГНАЛОВ МЕТОДОМ ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ И ЕГО ПРИМЕНЕНИЕ В ОБРАБОТКЕ РЕЧЕВЫХ СИГНАЛОВ

И.А. ВОРОНЕЦКИЙ

Белорусский государственный университет информатики и радиоэлектроники
П.Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 2 мая 2012

В данной статье рассмотрены алгоритм декомпозиции сигнала на синусоидальные компоненты – эмпирические моды, способ кодирования сигнала с помощью данного метода, определения частоты основного тона сигнала, метод редактирования акустического шума в сигнале с помощью данного метода. Также проведен сравнительный анализ приложений данного алгоритма.

Ключевые слова: декомпозиция на эмпирические моды, частота основного тона, редактирование акустического шума.

Введение

Определение мгновенных параметров речевого сигнала и редактирование акустического шума – актуальная задача, и в данном направлении проводятся активные исследования. Задачи определения мгновенных параметров и редактирования шума важны в таких областях как распознавание речи, идентификация диктора, низкоскоростное кодирование речевых сигналов, улучшение и корректирование речевых сигналов с применением гармонической модели. В последнее время было предложено много методов для решения данных задач, но проблема повышения точности и стабильности их работы все еще стоит остро [1]. Главный недостаток современных методов редактирования шума – это появление различных артефактов в речевом сигнале. Вместе с шумом могут также исчезнуть некоторые важные для восприятия речи гармоники.

Декомпозиция сигнала на эмпирические моды – это новая методика обработки и анализа речевого сигнала. Метод декомпозиции на эмпирические моды учитывает локальные особенности сигнала (экстремумы и нули сигнала) и структуры сигнала (шумовые, сигнальные и трендовые компоненты). Данный метод подходит для кодирования речевого сигнала, определения мгновенных параметров сигнала, редактирования акустического шума и др. В данной статье проводится краткий анализ и обзор приложений декомпозиции сигнала на эмпирические моды.

Понятие эмпирической моды

Эмпирическая мода (ЭМ, английское название IMF – Intrinsic Mode Function) – это функция, заданная непрерывно на интервале существования сигнала или дискретно в виде вектора отсчетов, имеющая в общем случае произвольную форму и произвольную аналитическую запись (если таковая существует), но при этом строго удовлетворяющая двум условиям [2, 3].

1. Общее число максимумов и минимумов такой функции (т.е. общее число экстремумов) должно быть строго равно числу нулей функции либо отличаться от числа нулей по модулю не более чем на единицу:

$$N_{\max} + N_{\min} = N_{zero} \pm 1 \text{ или } N_{\max} + N_{\min} = N_{zero}, \quad (1)$$

где $N_{\max}$, $N_{\min}$, N_{zero} – число максимумов, минимумов и нулей функции соответственно, не считая краевые отсчеты сигнала, которые могут оказаться единственными экстремумами.

2. Локальное (мгновенное) среднее значение функции, определенное в виде полусуммы двух огибающих – верхней, полученной путем интерполяции найденных локальных максимумов, и нижней, полученной путем интерполяции найденных локальных минимумов, и должно быть меньше или равно заранее определенному пороговому значению η:

$$0,5(U(i)+L(i)) \le \eta \,, i=\overline{1,N}, \qquad (2)$$

где $U(i)$ и $L(i)$ – значения верхней и нижней огибающих сигнала в i-й момент времени; N – общее количество сигнальных отсчетов.

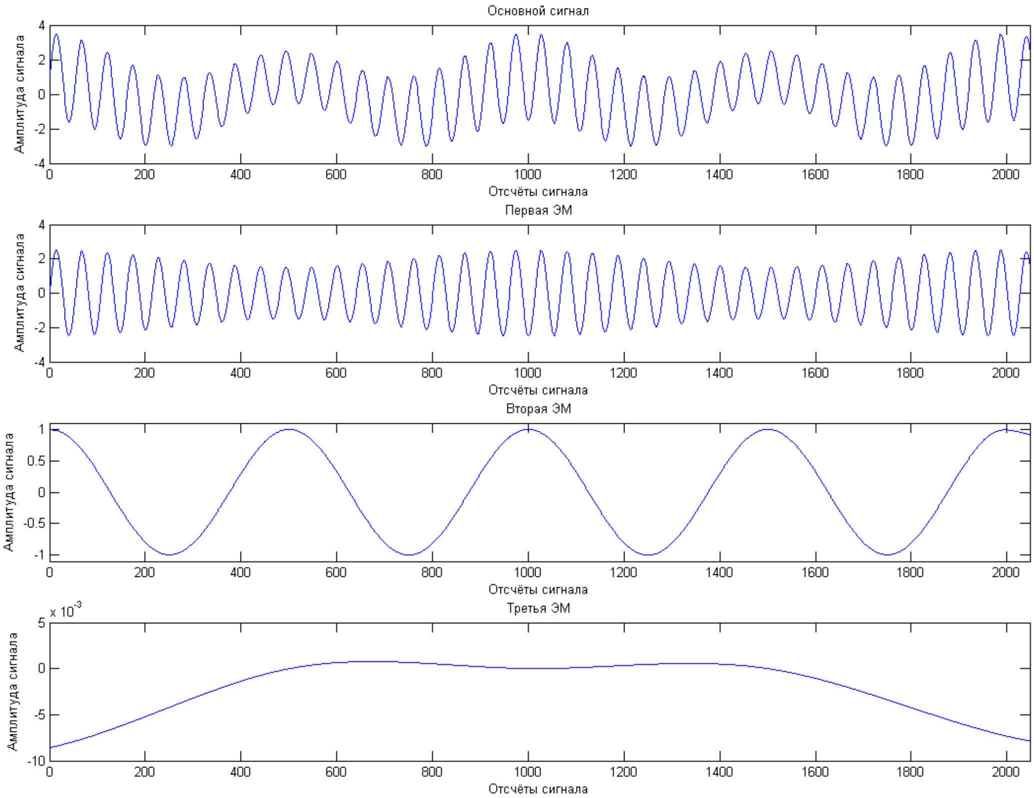

Рис 1. Пример декомпозиции сгенерированного сигнала на эмпирические моды

Алгоритм декомпозиции сигнала на эмпирические моды

При разложении сигнала на эмпирические моды (ЭМ) сначала вычисляются экстремумы исходного сигнала. Затем с помощью интерполяции данных о минимумах и максимумах сигнала, мы находим огибающие экстремумов. Далее полусумма огибающих вычитается из исходного сигнала, и получившийся сигнал проверяется на соответствие определению эмпирической моды: если функция подходит под определение, то нами получена первая ЭМ. В дальнейшем речевой сигнал можно описать как сумму всех найденных ЭМ и сигнала-остатка. Далее приведен краткий алгоритм разложения сигнала на эмпирические моды [4].

Шаг 1. Определение экстремумов и точек пересечения функцией нулевого уровня.

Шаг 2. Интерполяция полученных максимумов и минимумов. Самый распространенный способ интерполяции – интерполяция с помощью кубических сплайнов.

Шаг 3. Вычисляем полусумму огибающих, затем вычитаем ее из основного сигнала. Полученный сигнал является претендентом на то, чтобы быть первой эмпирической модой (ЭМ), поэтому необходимо проверить его на соответствие двум условиям, описанным выше. Если сигнал удовлетворяет этим условиям, то он действительно является ЭМ. Если нет, то возвращаемся к шагу 1, только в качестве исходного сигнала будет использоваться сигнал, полученный на шаге 3.

Шаг 4. Вычитая найденную ЭМ из сигнала, использованного в шаге 1, имеем сигнал-остаток, который в дальнейшем будет использоваться для повторения шагов 1–3.

Таким образом, в общем виде разложение сигнала на эмпирические моды можно записать следующим образом:

$$y(i) = \sum_{n=1}^{N} y_n^{''}(i) + y_{res}(i), \qquad (3)$$

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.

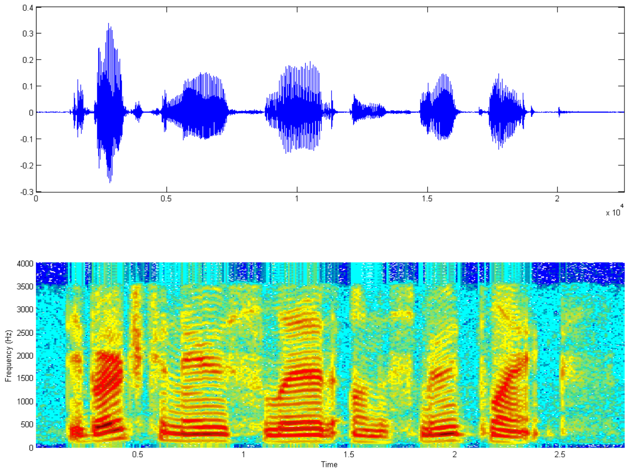

Рис 2. Пример речевого сигнала и его спектрограмма

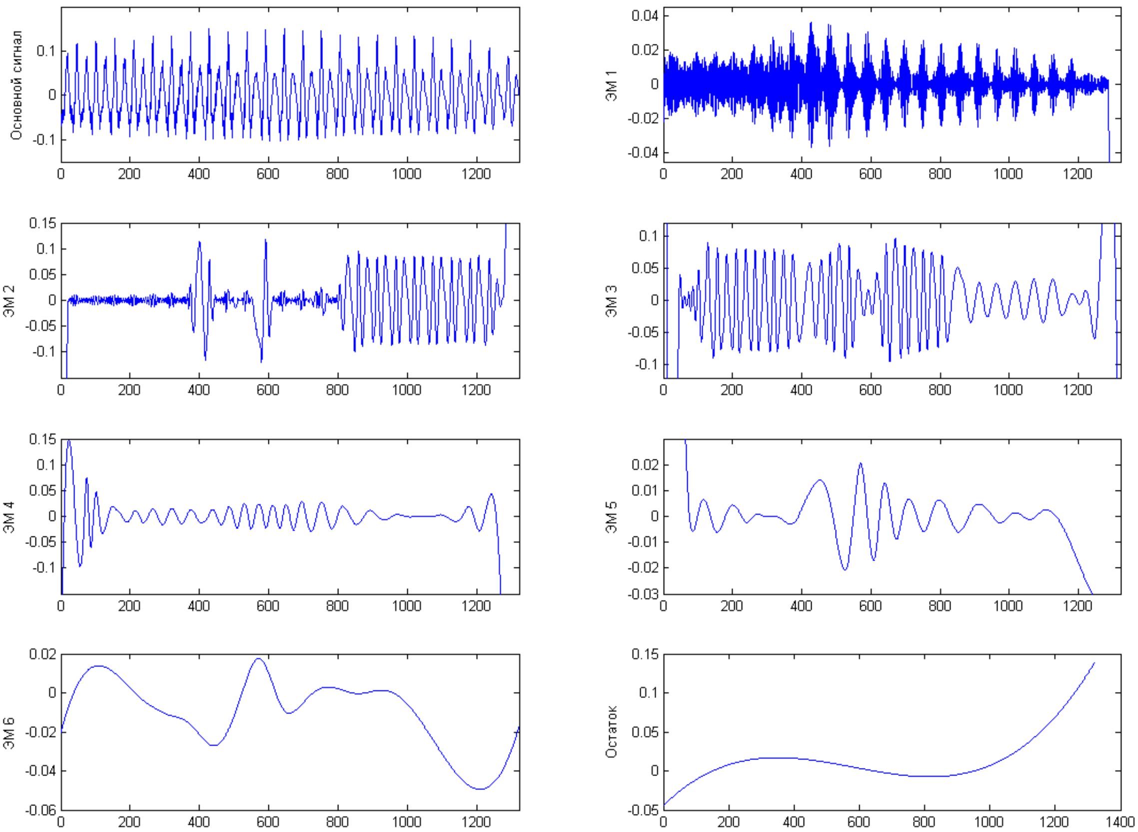

Рис 3. Пример декомпозиции речевого сигнала на эмпирические моды (ЭМ), сигнал-остаток исходного сигнала

Зададим критерии остановки процесса поиска ЭМ, чтобы компоненты ЭМ сохраняли смысл амплитудной и частотной модуляции. Процесс просеивания останавливается, когда остаток становится монотонной функцией, из которой больше нельзя извлечь ЭМ, или в случае достижения ограничения – некоторой нормированной величины, вычисляемой из двух последовательных ЭМ:

$$S = \frac{(y''_{n-1}(i) - y''_n(i))^2}{(y''_{n-1}(i))^2}, \qquad (4)$$

где S – условная величина, пределы которой задаются эмпирически (обычно устанавливается между 0,2 и 0,3); $y''_n(i)$ – ЭМ с номером n.

Разложение и реконструкция речевых сигналов на эмпирические моды

Для иллюстрации особенностей речевого сигнала фраза произнесенная мужчиной, была обработана в среде MATLAB. На рис. 2 приведены временное (вверху) и спектральное (внизу) представления речевого сигнала длиной примерно 2,6 секунды с частотой дискретизации 8 кГц.

Спектрограмма (рис. 2) описывает энергию сигнала в координатах «время-частота-яркость», затемненные участки соответствуют областям концентрации энергии. При малом спектральном разрешении из-за сглаживания точность оценки частотных параметров сигнала мала. Временные же характеристики (например, границы слов и отдельных фонем) могут быть определены достаточно точно. Большее разрешение в частотной области достижимо только за счет ухудшения разрешения по времени. На нижнем графике рис. 2 узкие горизонтальные линии на спектрограмме соответствуют траекториям гармоник основной частоты. Для невокализованных звуков подобной структуры спектра не наблюдается [5].

По своей природе речевые сигналы не являются стационарными, но на промежутке около 30 мс можно считать, что мгновенные параметры сигнала постоянны, поэтому обработка речевых сигналов выполняется по фреймам. На рис. 3 приведен пример разложения фрагмента речевого сигнала, представленного на рис. 2, на эмпирические моды; на рис. 4 – его реконструкция.

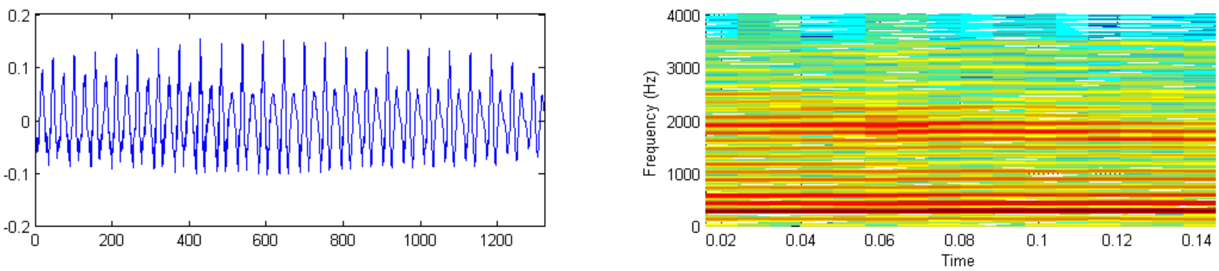

Рис. 4. Восстановленный сигнал и его спектрограмма

Как видно из рис. 4, при реконструкции сигнал не изменился.

Применение метода декомпозиции сигнала на эмпирические моды

Метод кодирования и квантования сигнала подробно описан в статье [6], далее приведены результаты экспериментальных исследований, чтобы оценить характеристики кодирования сигнала на основе метода декомпозиции на эмпирические моды. Метод был протестирован на сигналах с частотой дискретизации 44,1 кГц. В частности, следующие сигналы были взяты из базы данных SQAM [7]. Результаты были сравнены с MP3 и AAC кодеками и методом вейвлет компрессии. Были использованы вейвлеты Добеши 8 порядка, которые показывают хорошие характеристики в сравнении с другими вейвлетами. Характеристики предложенного метода анализировались с помощью соотношения сигнала к порогу маскирования (Noise to Mask Ratio – NMR) и объективного критерия разницы сигналов (Objective Difference Grade – ODG) [8]. ODG измеряет разницу между оригинальным и декодированным сигналом и возвращает значения от 0 (неощутимая разница) до 4 (достаточно большая разница). Значения NMR и ODG для четырех различных методов кодирования сигнала получены при битрэйте 64 кб/с и приведены в табл. 1.

Таблица 1. **Результаты компрессии аудио сигналов**

Тип кодирования		Сигнал					
		gspi	harp	quar	song	trpt	violin
ДЭМ	Битрейт, кб/с	64	64	64	64	64	64
	NMR	–5,37	–5,65	–5,47	–5,13	–5,32	–5,04
	ODG	–0,82	–0,73	–0,74	–0,79	–0,84	–0,83
AAC	Битрейт, кб/с	64	64	64	64	64	64
	NMR	–3,43	–6,46	–4,78	–4,23	–6,15	–4,59
	ODG	–0,85	–0,73	–0,75	–0,89	–0,88	–0,86
MP3	Битрейт, кб/с	64	64	64	64	64	64
	NMR	1,42	1,21	1,27	1,23	2,68	1,86
	ODG	–1,12	–1,87	–1,91	–1,09	–1,27	–1,34
Вейвлет	Битрейт, кб/с	65	67	64	65	66	64
	NMR	–2,30	–3,67	1,64	–3,40	–1,35	–2,52
	ODG	–0,86	–1,27	–1,74	–0,98	–0,97	–1,08

Из табл. 1 видно, что кодирование сигнала на основе метода декомпозиции на эмпирические моды дает лучшие результаты, чем остальные алгоритмы кодирования сигнала. Это происходит благодаря психоаккустической модели и свойствам симметрии эмпирических мод.

В статье [9] представлена методика оценки частоты основного тона с помощью декомпозиции сигнала на эмпирические моды. Ниже приведен краткий алгоритм метода:

– предварительно выполняется фильтрация акустического шума;

– находим нормализованную автокорреляционную функцию (НАКФ) предварительно отфильтрованного сигнала;

– применяем алгоритм ДЭМ к НАКФ сигнала. Существует одна ЭМ, которая содержит нужную нам фундаментальную частоту;

– вычисляем длину НАКФ и каждой найденной эмпирической моды;

– теперь сравниваем длину НАКФ и каждой найденной эмпирической моды: та ЭМ, длина которой наиболее близка к длине НАКФ, содержит требуемую фундаментальную частоту;

– выбранная ЭМ представляет собой затухающий синусоидальный сигнал, симметричный относительно нуля. Его частота и есть фундаментальная частота выбранного сегмента речевого сигнала.

Таблица 2. **Результаты работы алгоритмов определения частоты основного тона**

Голос	Метод	Соотношение сигнал-шум, дБ					
		–15	–5	0	10	20	30
Мужской голос	ДЭМ	29,38	5,58	2,19	0,71	0,49	0,35
	НАКФ	67,30	23,37	11,51	3,38	1,69	1,27
	WAC	69,20	24,29	12,07	3,38	1,20	0,98
Женский голос 1	ДЭМ	23,08	4,15	2,11	1,34	1,06	0,91
	НАКФ	65,65	22,94	12,87	4,71	2,11	1,90
	WAC	67,23	24,48	13,25	5,21	2,78	2,03
Женский голос 2	ДЭМ	25,81	6,23	3,09	1,93	1,54	1,49
	НАКФ	69,44	21,84	11,91	4,24	2,04	1,70
	WAC	68,56	23,05	11,58	3,97	1,93	1,59

Также в статье [9] приведена таблица, где рассмотрены результаты работы предложенного метода в сравнении со стандартным автокорреляционным методом и методом поиска частоты основного тона со взвешенной автокорреляционной функцией (WAC) [10]. Были использованы 1 сигнал с диктором мужчиной и 2 с дикторами женщинами.

Если величина частоты основного тона отклоняется от эталонной более чем на 20%, то мы классифицируем эту величину как грубую ошибку (Gross Pitch Error – GPE), в другом случае мы классифицируем величину как допустимую ошибку (Fine Pitch Error – FPE). Величина GPE выражается в процентах, эталонные значения частоты основного тона взяты из базы данных. Как видно из табл. 2, процент грубых ошибок значительно ниже у предложенного метода определения частоты основного тона, чем у стандартного автокорреляционного метода и WAC метода при любых соотношения сигнал-шум.

В статье [11] представлен метод редактирования акустического шума на основе декомпозиции на эмпирические моды. Блок-схема алгоритма представлена на рис. 5.

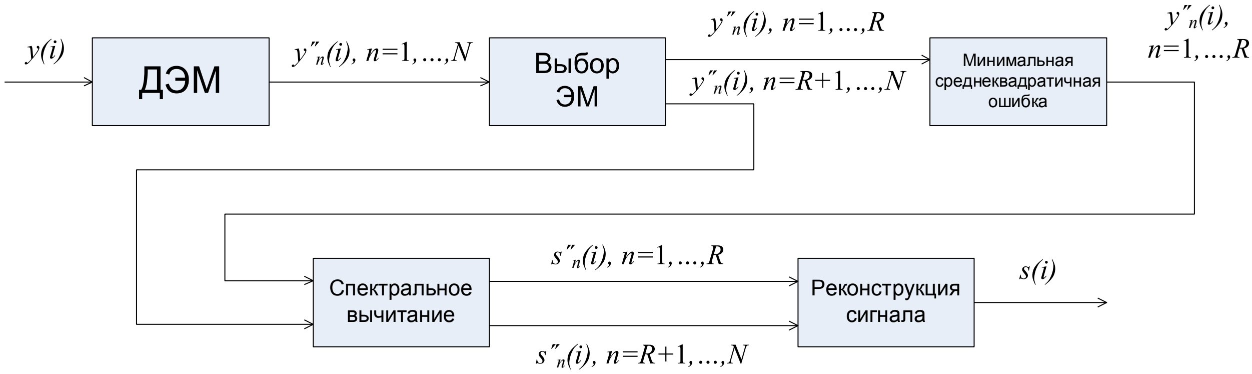

Рис. 5. Блок-схема алгоритма редактирования акустического шума с помощью декомпозиции сигнала на эмпирические моды

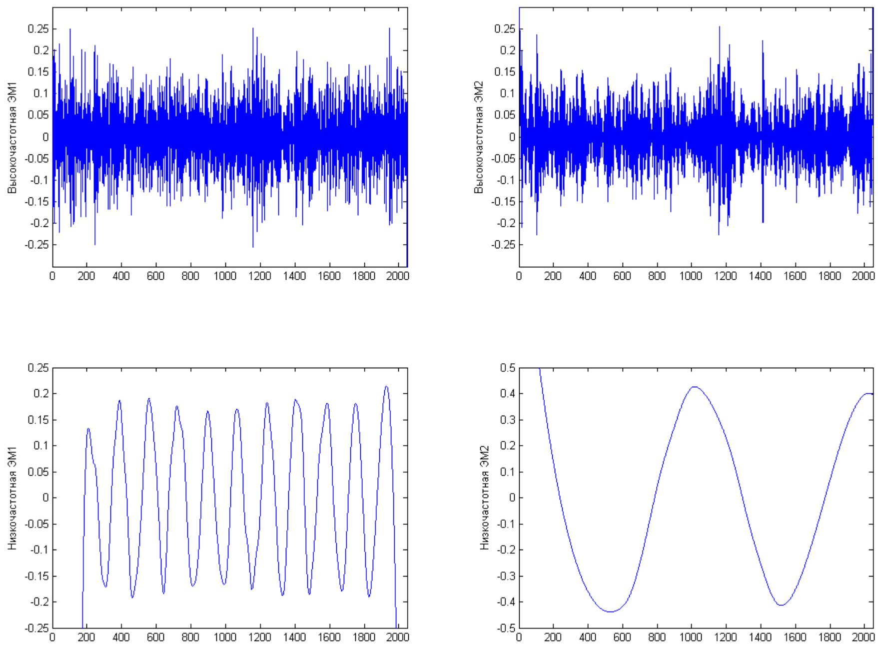

Рис. 6. Первые две эмпирические моды зашумленного сигнала (вверху) и последние две (внизу)

После первого блока декомпозиции на эмпирические моды получаем набор ЭМ, где N – количество эмпирических мод. Затем идет разделение эмпирических мод на содержащие высокочастотную часть сигнала и содержащие низкочастотную, R – это номер последней ЭМ, содержащей высокочастотную часть сигнала. Далее, применяя метод минимальной среднеквадратичной ошибки, корректируем шум в высокочастотных модах и применяем метод спектрального вычитания, чтобы минимизировать размытость спектра речевого сигнала [12] . Последним шагом реконструируем сигнал из имеющихся эмпирических мод. На рис. 6 представлены первые две ЭМ и последние две синтезированного зашумленного сигнала, которые содержат высокочастотные и низкочастотные части сигнала соответственно.

В статье [11] была проведена оценка характеристик метода редактирования акустического шума. Мерой оценки характеристик метода было соотношение сигнал-шум (SNR). Использовался зашумленный сигнал со временем реверберации 200 мс. Как видно из табл. 3, метод двухступенчатой обработки речевых сигналов с реверберацией и использованием одного микрофона [12], комбинированный с методом ДЭМ-СКО, показывает сходные результаты с предложенным методом, в то время как просто метод обработки речевых сигналов с реверберацией и использованием одного микрофона значительно отстает по показателям от остальных двух.

Таблица 3. **Результаты работы методов редактирования акустического шума**

Метод	Входной *SNR*					
	–12	–8	–4	0	2	4
Предложенный метод	–0,35	0,1	0,4	0,8	0,9	1
Метод [7] + ДЭМ-СКО	–0,37	–0,05	0,45	0,85	0,95	1,05
Метод [7]	–6,5	–4,8	–2,4	–0,6	–0,05	0,4

Заключение

В данной статье был рассмотрен метод декомпозиции сигнала на эмпирические моды и его приложения: кодирование речевого сигнала, определение частоты основного тона речевого сигнала с помощью автокорреляционной функции и редактирование акустического шума. Данная методика показывает себя как надежный и качественный метод обработки речевых сигналов, дающий качественные результаты. Были рассмотрены показатели работы предложенного алгоритма в условиях наличия шумов и реверберации в голосовом сигнале, результаты также превосходят методы, с которыми проводилось сравнение.

ANALYSIS OF SIGNALS BASED ON EMPIRICAL MODE DECOMPOSITION METHOD AND ITS APPLICATION ON SPEECH SIGNALS

I.A VORONETSKIY

Abstract

A method of signal decomposition on sinusoidal components (intrinsic mode functions), audio coding, a method of pitch tracking and a method of noise redaction are presented in this paper. The comparison of the performance of proposed method with other methods is analyzed.

Список литературы

1. *Spanias A.S.* // Proceedings of the IEEE. 1994. Vol 82, №10. P. 1539–1582.
2. *Клионский Д.М.* // Цифровая обработка сигналов. 2011. №2. С. 51–60.
3. *Клионский Д.М.* // Доклады 13-й Международной конференции «Цифровая обработка сигналов и ее применение». Москва. 2011. С. 120–123.
4. *Кан Ш.Ч., Микулович А.В., Микулович В.И.* // Информатика. 2010. №2. С. 25–35.
5. *Петровский А.А.* Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика. Минск, 2009.
6. *Khaldi K., Boudraa A.O., Turki Hadj-Alouane M., et. al.* // Signal processing. 2011.
7. Sound Quality Assessment Material recording for subjective tests // Technical Centre of the European Broadcasting Union. 1988
8. Method for Objective Measurements of Perceived Audio Quality // ITU Recommendation, ITU-R BS.1387-1, 2001.
9. *Sujan K.R., Khademul I.M., Keikichi H., et. al.*// ISCAS. 2010. P. 2658–2661.
10. *Shimamura T., Kobayashi H.* // IEEE Trans. Speech and Audio Proc. 2001. P. 727–730.
11. *Tariqullah J., Wenwu W.* // 19th European Signal Processing Conference. 2011. P. 206–210.
12. *Wu M., Wang D.L.* // IEEE Trans. Audio, Speech, Lang. Process. 2006. Vol. 14. P. 774–784.