

Л и т е р а т у р а

1. **ГОСТ 54500.1—2011.** Руководство ИСО/МЭК 98-1:2009. Неопределённость измерения. Ч. 1. Введение в руководства по неопределённости измерения.

2. **ГОСТ 54500.3—2011.** Руководство ИСО/МЭК 98-3:2008. Неопределённость измерения. Ч. 3. Руководство по выражению неопределённости измерения.

3. **Рабинович С. Г.** Навстречу новой редакции «Руководства по выражению неопределённости измерений» // Системы обработки информации. 2008. Вып. 4 (71). С. 10—14.

4. **Рабинович С. Г.** О необходимости создания новых рекомендаций по оцениванию погрешностей и неопределённостей измерений // Системы обработки информации. 2010. Вып. 4 (85). С. 23—26.

5. **ГОСТ 8.009—85.** ГСИ. Нормируемые метрологические характеристики средств измерений.

6. **Уилкс С.** Математическая статистика. М.: Наука, 1967.

7. **Большев Л. Н., Смирнов Н. В.** Таблицы математической статистики. М.: Вычислительный центр АН СССР, 1968.

Дата принятия 27.07.2015 г.

534.78

Метод адаптивной обработки сигналов для диагностики органов речевого аппарата

А. Ю. ТЫЧКОВ, А. К. АЛИМУРАДОВ, П. П. ЧУРАКОВ

Пензенский государственный университет, Пенза, Россия, e-mail: tychkov-a@mail.ru

Приведены результаты исследования разработанного метода. Показано, что среднее значение энергии спектра Гильберта речевого сигнала, полученное с использованием комплементарной множественной декомпозиции на эмпирические моды и преобразования Гильберта—Хуанга, варьируется в разных пределах в зависимости от выбора фраз, сформированных отдельными активными органами речевого аппарата.

Ключевые слова: распознавание речи, обработка речевого сигнала, преобразование Гильберта—Хуанга.

The developed method evaluation results are presented. The mean energy value of Hilbert spectrum (HS) of a speech signal received by the CEEMD and HHT is shown to vary in different ranges depending on the choice phrases formed by separate active organs of speech apparatus.

Key words: a speech signal processing, Complementary Ensemble Empirical Mode Decomposition, Hilbert-Huang transform.

В настоящее время речевые сигналы широко используются для диагностики состояния органов речевого аппарата (ОРА) [1—3]. Эффективность анализа этих сигналов зависит от точности измерений амплитудных, временных, частотных и энергетических характеристик. Основной причиной погрешностей измерений являются физиологические помехи, вызванные нестабильной работой ОРА при патологических отклонениях (ларингит, синусит, фарингит и др.). Помехи в зависимости от происхождения, интенсивности, характеристик и взаимодействия с полезным сигналом могут существенно исказить результаты анализа речевых сигналов.

Цель статьи — создание и исследование адаптивного к физиологическим особенностям человека метода обработки речевых сигналов для диагностики ОРА.

Речевой сигнал представляет акустический нестационарный сигнал сложной формы, образуемый артикуляционным отделом ОРА [4]. В процессе формирования речевых сигналов участвуют три основных отдела органов дыхания: I — лёгкие, бронхи и трахея; II — голосовой аппарат гортани; III — полость рта, глотки носа и околоносовых пазух. При использовании определённых методик (новых алгоритмов) обработки сигналов и специальных верифицированных баз ре-

чевых сигналов открывается уникальная возможность диагностировать патологию ОРА на различных стадиях развития.

Существуют разные подходы к обработке речевых сигналов, что обусловлено как важностью проблемы, так и отсутствием достаточно эффективных методов её решения. Среди известных можно отметить систему анализа спектральных характеристик и формирования изображения шума дыхательных путей [5], позволяющую регистрировать и анализировать акустические шумы тела пациента, и использовать полученную информацию для диагностики состояния дыхательных путей. Широкому использованию данного метода препятствуют следующие его недостатки: кроме акустических шумов регистрируются шумы, связанные с физиологической активностью других систем организма — сердечно-сосудистой, желудочно-кишечного тракта и др.; диагностируются заболевания только отдела I ОРА.

Способ [6] позволяет регистрировать речевой сигнал, сегментировать его на вейвлет-коэффициенты, определять информативную область вейвлет-поверхности с последующим вычислением энергии информативной области и сравнением вычисленного значения с пороговым. Недостатки

данного способа: невысокая эффективность диагностического заключения о стадии нестабильной работы ОРА, обусловленная невозможностью адаптации вейвлет-функции к структуре и локальным особенностям зарегистрированного речевого сигнала; диагностика заболеваний только отдела II ОРА, что также как и в предыдущем случае ограничивает его использование.

В [5, 6] для анализа применяются фурье- и вейвлет-преобразования. Известно, что преобразование

Фурье не подходит для обработки нестационарных речевых сигналов. Вейвлет-преобразование можно рассматривать как адаптивный инструмент разложения нестационарных данных, но при этом необходимо иметь априорную информацию — базисную функцию вейвлета. Однако её выбор на основе характеристик анализируемого сигнала не всегда оказывается однозначным. Поэтому перспективно использовать методы, связанные с анализом нестационарных данных [7—9]: метод декомпозиции на эмпирические моды (ЭМ) [10] и преобразование Гильберта—Хуанга (ПГХ) [11]. К их недостаткам относится смешивание ЭМ, возникающее при разложении речевых сигналов, содержащих кратковременные и несоизмеримые по масштабу участки.

В соответствии с изложенным и исследованиями [12—15] по устранению недостатков известных способов обработки сигналов для диагностики состояния ОРА предложен новый адаптивный метод, основанный на комплементарной множественной декомпозиции на эмпирические моды (КМДЭМ) [16] и ПГХ.

Рассмотрим особенности реализации предлагаемого метода адаптивной обработки речевых сигналов. Их регистрация осуществляется с помощью акустического приёмника и аналогово-цифрового преобразователя (АЦП). Пример зарегистрированного и оцифрованного речевого сигнала — звука «а» приведён на рис. 1.

На этом этапе предлагается фиксировать речевые сигналы конкретных фраз и выражений, состоящих из звуков, сформированных определёнными ОРА. Авторами разработана верифицированная база речевых сигналов в виде таблиц системы управления реляционной базой данных Microsoft Office Access [17], состоящей из более 1000 речевых сигналов длительностью 5—500 мс; фонемы (информативные области) речевых сигналов различных тональных и сонорных звуков записаны в интервале 10—25 мс. Использование КМДЭМ и ПГХ позволяет разложить фонемы речевого сигнала на частотные составляющие и подстроиться к исследуемому сигналу без искажения его свойств.

Алгоритм КМДЭМ фонемы речевого сигнала включает следующие этапы:

1. Добавление белого шума с прямыми и инверсными значениями к фонеме сигнала, что даёт пару фонем, соответственно $f(t_i)$ и $f^*(t_i)$:

$$\begin{bmatrix} f(t_i) \\ f^*(t_i) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \times \begin{bmatrix} x(t_i) \\ \omega(t_i) \end{bmatrix},$$

где $x(t_i)$ — речевой сигнал; $\omega(t_i)$ — добавленный белый шум.

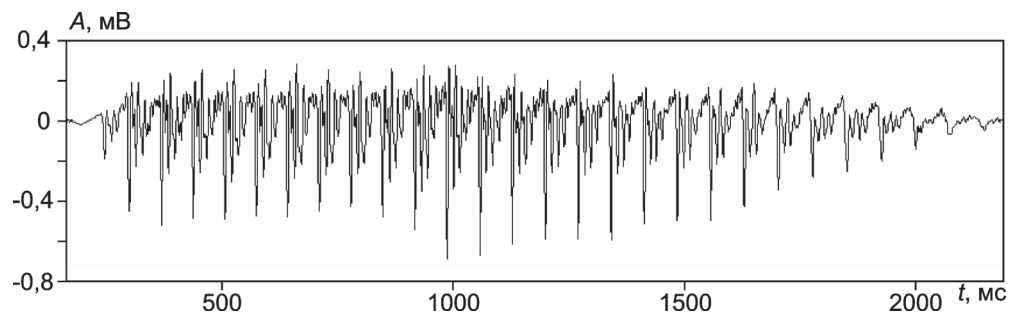


Рис. 1. Речевой сигнал — звук «а»

2. Определение локальных экстремумов фонемы речевого сигнала $f(t_i)$ — максимумов и минимумов — соответственно при выполнении условий $f(t_{i-1}) < f(t_i) \geq f(t_{i+1})$ и $f(t_{i-1}) > f(t_i) \leq f(t_{i+1})$.

3. Определение верхней $e(t_i)$ и нижней $g(t_i)$ огибающих фонемы речевого сигнала с помощью кубической сплайн-интерполяции [16] по найденным локальным экстремумам $f(t_i)$:

$$e(t_i) = a_b t_i^3 + b_b t_i^2 + c_b t_i + d_b;$$

$$g(t_i) = a_n t_i^3 + b_n t_i^2 + c_n t_i + d_n,$$

где i — дискретный отсчёт времени, $a_b, b_b, c_b, d_b, a_n, b_n, c_n, d_n$ — коэффициенты для каждого значения i -го отсчёта соответственно верхней и нижней огибающих.

4. Вычисление среднего значения огибающих фонемы речевого сигнала по формуле

$$h(t_i) = [e(t_i) + g(t_i)] / 2.$$

5. Проверка условия $h(t_i) \approx 0$. Если оно не выполняется, то вычисляется локальный остаток $c(t_i) = f(t_i) - h(t_i)$ и повторяются первые три действия.

Если условие выполняется, то локальный остаток принимает статус эмпирической моды j -го уровня декомпозиции $IMF_j(t_i) = c(t_i)$.

6. Вычисление остатка сигнала $r(t_i) = f(t_i) - IMF_j(t_i)$.

7. Вычисление критерия останова декомпозиции, в качестве которого используется нормализованная квадратичная разность

$$SD = \sum_{i=1}^n \left[\frac{(|IMF_j(t_i) - r(t_i)|)^2}{r(t_i)^2} \right],$$

где n — количество дискретных отсчётов времени в фонеме речевого сигнала.

Проверка условия останова заключается в сравнении остатка сигнала с порогом: $SD \leq \delta$. Экспериментально порог для речевых сигналов установлен $\delta = 0,25$. Если условие не выполняется, то остаток используется в качестве сигнала для последующего уровня декомпозиции ($j+1$), а если выполняется, то выводится набор ЭМ, полученных на всех уровнях декомпозиции.

В результате декомпозиции из исходного сигнала $f(t_i)$ извлекается конечное число ЭМ $IMF_j(t_i)$ и результирующий остаток $r(t_i)$:

$$f(t_i) = \sum_{j=1}^{J-1} IMF_j(t_i) - r_J(t_i),$$

где j — номер уровня декомпозиции; J — количество уровней.

После того, как получены частотные составляющие фонемы речевого сигнала с помощью декомпозиции на ЭМ, строится спектр Гильберта (СГ) [10, 11] по следующим этапам:

1. Получение сопряженного по Гильберту речевого сигнала фонемы. Преобразование Гильберта — это математический аппарат, позволяющий сформировать из анализируемого исходного речевого сигнала — действительной функции времени — аналитический (комплексный) сигнал

$$IMF_j^*(t_i) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{IMF_j(\tau)}{t_i - \tau} d\tau,$$

где τ — дискретный временной сдвиг.

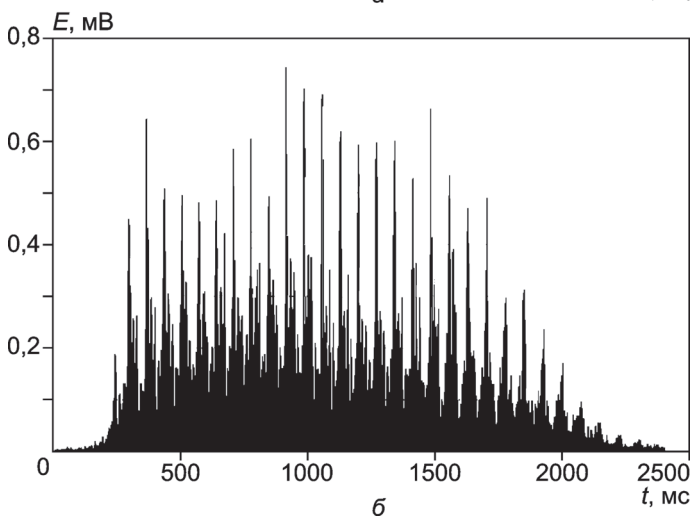
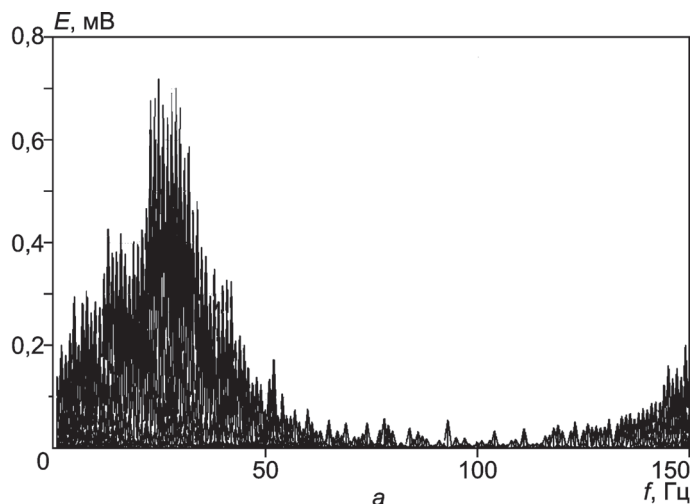


Рис. 2. Спектры Гильберта речевого сигнала в зависимости от частоты (а) и времени (б)

2. Вычисление комплексного значения для каждой моды фонемы речевого сигнала:

$$Z_j(t_i) = IMF_j(t_i) + q IMF_j^*(t_i),$$

где q — мнимая единица.

3. Определение мгновенных амплитуды огибающей аналитического сигнала

$$a_j(t_i) = |Z_j(t_i)| = \sqrt{IMF_j(t_i)^2 + q IMF_j^*(t_i)^2},$$

его фазы

$$\varphi_j(t_i) = \arg Z_j(t_i) = IMF_j^*(t_i) / IMF_j(t_i)$$

и частоты, равной скорости изменения текущей фазы: $\omega_j(t_i) = d\varphi/dt_i$.

4. Представление фонемы речевого сигнала в виде СГ — поверхности в системе координат амплитуда—частота—время [10, 11]:

$$P(t_i) = \sum_{j=1}^J a_j(t_i) \exp[q \int \omega(t_i) dt_i].$$

Для выражения энергетической плотности частотных составляющих ЭМ и построения поверхности в системе координат энергия—частота—время, используют формулу

$$E(t_i) = \sum_{j=1}^J a_j(t_i)^2 \exp[q \int \omega(t_i) dt_i]. \quad (1)$$

Визуальное изображение СГ характеризует распределение мгновенной энергии фонемы речевого сигнала в каждой точке частотно-временной плоскости и, по мнению авторов, позволяет выделить новые диагностические признаки. На рис. 2 приведён СГ речевого сигнала (звук «а»), рассчитанный по (1).

Предложенный метод адаптивной обработки экспериментально исследовали на речевых сигналах разной длительности, в частности, на звуках, словах и словосочетаниях. Обучающую выборку сформировали из 1000 чистых (соотношение сигнал—шум 50 дБ) речевых сигналов, произнесённых 50-ю мужчинами и женщинами. Тестовая выборка содержала 100 патологических речевых сигналов (40 звуков, 30 слов, 30 словосочетаний), зарегистрированных в одинаковых акустических условиях.

Результаты проведённых исследований показали, что для каждой группы ОРА среднее значение энергии СГ, полученное с помощью ПГХ, варьируется в разных пределах, как показано в таблице.

Вариации значений энергии спектра Гильберта в зависимости от выбранного отдела органов речевого аппарата и патологии

Органы речевого аппарата	E, мВ ²	
	Норма	Патология
Лёгкие, бронхи, трахея	0,07—0,09	0,12—0,24
Голосовой аппарат гортани	0,12—0,15	0,18—1,00
Полости рта, глотки носа и околоносовых пазух	0,03—0,04	0,05—0,80

Исследование предложенного метода на пациентах, находящихся в свободной двигательной активности, при внешних акустических шумах является предметом дальнейших научных изысканий. Кроме того, актуально изучение конкретных групп пациентов с патологиями в клинических условиях при большей выборке данных, а также экспериментальная проверка полученных результатов в сравнении с известными устоявшимися методиками диагностики органов речевого аппарата.

Выводы. Предложенный метод позволяет устранить недостатки известных решений — повысить достоверность диагностического заключения о стадии патологии речевого аппарата благодаря адаптивной обработке фоном речевого сигнала, учитывающей индивидуальные особенности человека. Визуальное представление СГ речевого сигнала даёт возможность выявить новые диагностические признаки в речевых сигналах для медицинских диагностических систем.

Л и т е р а т у р а

1. **Chevalier D., Mortuaire G.** Acute disorder of speech. Dysphonia // *La Revue du Practicien*. 2005. V. 11. P. 1217—1221.
2. **Аникеева З. И., Плешков И. В., Бондарева А. А.** Клинические особенности нарушений голоса у населения мегаполиса // *Вестник оториноларингологии: медицинский научно-практический журнал*. 2007. № 1. С. 14—21.
3. **Ярославцева Д. А., Тычков А. Ю.** Аналитический обзор способов и алгоритмов оценки состояния голосовых складок и гортани // *Измерение. Мониторинг. Управление. Контроль*. 2014. № 3 (1). С. 12—18.
4. **Михайлов В. Г., Златоустова Л. В.** Измерение параметров речи. М.: Радио и связь, 1987.
5. **Пат. 2314751 РФ.** Система для анализа и формирования изображения шума дыхательных путей / И. Кушнир, М. Ботбол // *Изобретения. Полезные модели*. 2008. № 5.
6. **Пат. 2313280 РФ.** Способ исследования функционального состояния голосовых складок / И. В. Плешков, С. А. Филист, Н. В. Краснова, Ю. А. Блинков // *Изобретения. Полезные модели*. 2006. № 16.
7. **Weiping H.** A novel pitch period detection algorithm bases on NHT with application to normal and pathological voice//*Proc. Eng. Med. and Bio. Soc., IEEE-EMBS 2005: 27th Ann. Int. Conf. Shanghai (China)*, 2006. P. 4541—4544.
8. **Kaleem M.** Pathological speech signal analysis and classification using empirical mode decomposition // *Med. & bio. Eng. & comp.* 2013. V. 51. N. 7. P. 811—821.
9. **Gaston S., Maria Eugenia T., Hugo L. R.** Pathological Voice Analysis and Classification Based on Empirical Mode Decomposition// *Development of Multimodal Interfaces: Active Listening and Synchrony Lecture Notes in Computer Science*. 2010. V. 5967. P. 364—381.
10. **Huang N. E., Zheng Shen, Steven R. L.** The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis // *Proc. Roy. Soc. London A*. 1998. V. 454. P. 903—995.
11. **Huang N. E., Attoh-Okine Nii O.** The Hilbert-Huang transform in engineering. Philadelphia: Taylor and Francis, 2005.
12. **Тычков А. Ю., Бодин О. Н., Кузьмин А. В., Давыдова А. А., Чураков П. П.** Информационно-измерительная система для предварительной обработки флюорографических снимков // *Измерительная техника*. 2011. № 4. С. 42—53.
13. **Alimuradov A. K.** Speech signals filtering using the ensemble empirical mode decomposition method and the intrinsic mode functions energy assessment // *Int. J. Appl. Eng. Res.* 2015. V. 10. N. 2. P. 3175—3185.
14. **Алимурадов А. К., Тычков А. Ю.** Применение преобразования Гильберта—Хуанга в задаче выделения информативных признаков речевых сигналов // *Международный научно-исследовательский журнал*. 2013. № 5-1(12). С. 57—58.
15. **Алимурадов А. К., Квитка Ю. С.** Применение комбинированной множественной декомпозиции на эмпирические моды для анализа речевых сигналов // *Измерение. Мониторинг. Управление. Контроль*. 2014. № 4 (10). С. 69—75.
16. **Yeh J.-R., Shieh J.-S., Huang N. E.** Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method // *Advances in Adaptive Data Analysis*. 2010. V. 2 (2). P. 135—156.
17. **Свид-во об офиц. Регистрации базы данных № 2013621156 «Верифицированная база речевых сигналов»** / А. Ю. Тычков, А. К. Алимурадов, Ю. С. Квитка, Д. А. Ярославцева // *Программы для ЭВМ, базы данных, топологии интегральных микросхем*. 2013.
18. **Корнейчук Н. П.** Сплаины в теории приближения. М.: Наука, 1984.
19. **Sharpley R. C., Vatchev V. D.** Analysis of the intrinsic mode functions. Industrial Mathematics Institute Department of Mathematics University of South Carolina, Columbia (USA). 2004. P. 31.

Дата принятия 03.07.2015 г.