



А.К. Алимуратов, А.Ю. Тычков, П.П. Чураков

## ПОМЕХОУСТОЙЧИВЫЙ АДАПТИВНЫЙ АЛГОРИТМ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ ДЛЯ СИСТЕМ ОБНАРУЖЕНИЯ ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ

(Пензенский государственный университет)

### Введение

Современные системы обнаружения психоэмоционального состояния обеспечивают высокую точность при невысоком уровне или полном отсутствии фонового шума (лабораторные условия). В ситуациях, когда разница между зарегистрированными речевыми сигналами и сигналами, зарегистрированными в процессе обучения системы, из-за посторонних шумов становится значительной, эффективность обнаружения психоэмоционального состояния стремительно ухудшается. Свободная двигательная активность испытуемых требует от систем обнаружения психоэмоционального состояния возможность функционировать в условиях интенсивных помех. По этой причине актуальным является повышение помехоустойчивости обработки речевых сигналов.

На сегодняшний день широкую практическую популярность получили различные алгоритмы помехоустойчивой обработки:

- алгоритмы повышения четкости (корректировки) речи, хорошо известные как методы вычитания спектра [1];
- алгоритмы на основе фильтров Винера [2], используемые в предварительной обработке для удаления шумовой составляющей речевого сигнала;
- алгоритмы постобработки, основанные на кепстральном среднем вычитании (*Cepstral Mean Subtraction, CMS*) [3], нормализации кепстральных отклонений (*Cepstral Variance Normalization, CVN*) [4] и выравниванием гистограммы (*Histogram Equalization, HEQ*) [5], направленные на преобразование необработанных информативных параметров речи в форму, которая в меньшей степени подвержена к искажениям в зашумленной окружающей среде.

В данной работе кратко представлены результаты исследования алгоритма помехоустойчивой обработки, основанного на улучшенной полной множественной декомпозиции на эмпирические моды с адаптивным шумом (ПМДЭМАШ) [6]. Работа выполнена при финансовой поддержке Совета по грантам Президента РФ в рамках проекта № СП-246.2018.5 и является продолжением ранее опубликованного труда авторов [7].

### Улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом

Декомпозиция на эмпирические моды [6] – это адаптивная технология разложения нестационарных сигналов, возникающих в нелинейных системах. ДЭМ обеспечивает локальное разложение сигнала на быстрые и медленные колебательные функции. В результате разложения исходный сигнал может быть



представлен в виде суммы амплитудных и частотных модулированных функций, называемых эмпирическими модами (ЭМ). Аналитическое выражение ДЭМ выглядит следующим образом:

$$x(n) = \sum_{i=1}^I IMF_i(n) + r_i(n)$$

где  $x(n)$  – исходный сигнал,  $IMF_i(n)$  – ЭМ,  $r_i(n)$  – конечный остаток,  $i=1, 2, \dots, I$  – номер ЭМ,  $n$  – дискретный отсчет времени ( $0 < n \leq N$ ,  $N$  – количество дискретных отсчетов в сигнале).

Среди всех разновидностей методов декомпозиции наиболее адаптивной к речевым сигналам является улучшенная ПМДЭМАШ. Математическое описание метода улучшенной ПМДЭМАШ представлен ниже:

1. С помощью аппарата ДЭМ и выражая из формулы  $\langle E_1(x_j(n)) \rangle = \langle x_j(n) \rangle - \langle M(x_j(n)) \rangle$  локальные средние значения шумовых копий исходного сигнала ( $x_j(n) = x(n) + \beta_0 E_1(w_j(n))$ ) определяется первый остаток:

$$r_1(n) = \langle M(x_j(n)) \rangle$$

где  $E_i(\cdot)$  – аппарат извлечения ЭМ методом ДЭМ ( $i$  – номер моды),  $x_j(n) = x(n) + w_j(n)$  – шумовые копии исходного сигнала ( $x(n)$  – исходный речевой сигнал,  $w_j(n)$  – реализации белого шума с нулевой средней единичной дисперсией),  $M(\cdot)$  – аппарат, создающий локальное среднее значение применяемого сигнала,  $\beta_i = \varepsilon_i \text{std}(r_i)$  – коэффициент, допускающий выбор различных значений отношения сигнал/шум.

2. На первом этапе для  $i = 1$  вычисляется первая мода:  $IMF_1(n) = x(n) - r_1(n)$ .

3. Вычисляется второй остаток как усредненное локальное среднее значение шумовых копий первого остатка  $r_1(n) + \beta_1 E_2(w_j(n))$  и определяется вторая мода:

$$\overline{IMF}_2(n) = r_1(n) - r_2(n) = r_1(n) - \langle M(r_1(n) + \beta_1 E_2(w_j(n))) \rangle$$

4. На последующих этапах для  $i = 3, \dots, I$  вычисляется  $i$ -й остаток

$$r_i(n) = \langle M(r_{i-1}(n) + \beta_{i-1} E_i(w_j(n))) \rangle$$

5. Вычисляется  $i$ -ая мода

$$\overline{IMF}_i(n) = r_{i-1}(n) - r_i(n)$$

6. Переход к шагу 4 для следующего значения  $i$ .

Константы  $\beta_i$  выбираются так, чтобы получить желаемое отношение сигнал/шум между добавленным шумом и остатком, к которому добавляется шум.

### Помехоустойчивый адаптивный алгоритм обработки речевых сигналов

На основании вышеупомянутого и собственных исследований [7] разработан помехоустойчивый адаптивный алгоритм обработки речевых сигналов для систем обнаружения психоэмоционального состояния (см. рисунок 1). За основу алгоритма взят классический способ обнаружения психоэмоциональных состояний [8]. Штриховой линией отмечен режим обучения алгоритма, сплош-



ной линией – рабочий режим. Серым цветом отмечен блок 1 помехоустойчивой обработки на основе улучшенной ПМДЭМАШ. Суть обработки заключается в уменьшении разницы между поступающими в систему речевыми сигналами и сигналами, зарегистрированными в процессе обучения. С помощью улучшенной ПМДЭМАШ, исходный зарегистрированный речевой сигнал (как функция времени) разлагается на ЭМ. Основной идеей предлагаемого алгоритма является то, что ЭМ низкого порядка содержат высокочастотные шумовые составляющие, которые могут быть удалены для уменьшения разницы:

$$x'(n) = x(n) - \sum_{i=1}^I IMF_i(n)$$

где  $x'(n)$  – отфильтрованный речевой сигнал.

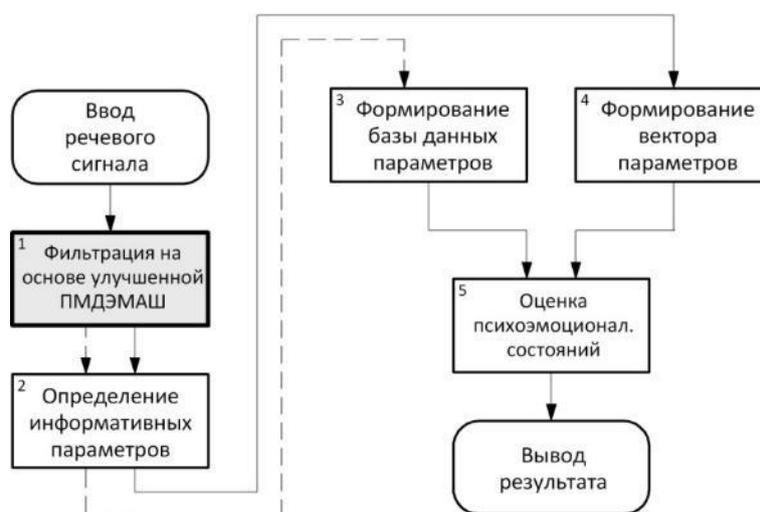


Рисунок 1 – Помехоустойчивый адаптивный алгоритм обработки речевых сигналов для систем обнаружения психоэмоционального состояния

### Исследование помехоустойчивого адаптивного алгоритма

Исследование помехоустойчивого адаптивного алгоритма обработки проводилось в пакете прикладных программ *MATLAB*. Для оценки эффективности обнаружения психоэмоциональных состояний предлагаемого авторами алгоритма при поддержке Областной клинической больницы им. К.Р. Евграфова (г. Пенза) и Пензенского государственного университета сформирована группа испытуемых и верифицированная база сигналов. В группу испытуемых отобрано 100 чел. мужского и женского пола, в возрасте от 18 до 60 лет, поступивших с явно выраженной симптоматикой психических расстройств. Для оценки эффективности, использовался параметр – ошибки первого и второго рода. Задачей исследования являлось обнаружение психоэмоциональных состояний классическим способом и предлагаемым авторами алгоритмом с фильтрацией на основе улучшенной ПМДЭМАШ. В таблице 1 представлены результаты обнаружения психоэмоциональных состояний.



В соответствии с полученными данными можно сделать вывод, что применение улучшенной ПМДЭМАШ для фильтрации в классическом способе позволяют повысить помехоустойчивость и эффективность обнаружения психоэмоциональных состояний. Лучшие результаты достигаются при вычитании из исходного речевого сигнала первой шумовой и последних трех трендовых ЭМ. Объясняется это тем, что первая ЭМ содержит большую часть остаточного шума, а три последние трендовые моды – результат разложения суммы гармонического сигнала и полиномиального тренда. Вычитание данных ЭМ избавляет исходный сигнал от шума.

Таблица 1

Результаты обнаружения пограничных психических расстройств

Прогнозируемый результат	Результат определения		Ошибки первого и второго рода, %	
	Патология	Норма		
<b>Классический алгоритм обнаружения</b>				
Патология	78 чел.	22 чел.	1-ого	22
Норма	16 чел.	84 чел.	2-ого	16
<b>Предлагаемый помехоустойчивый адаптивный алгоритм</b>				
Патология	91 чел.	9 чел.	1-ого	9
Норма	6 чел.	94 чел.	2-ого	6

Таким образом, полученные результаты позволяют сделать вывод, что предложенный авторами помехоустойчивый алгоритм, может успешно тестироваться в системах обнаружения психоэмоциональных состояний.

### Литература

1. Boll S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process.* - 1979. - V. 27 (2). - P. 113 - 120.
2. Berstein A., Shallom I. A hypothesized Wiener filtering approach to noisy speech recognition / *International Conference on Acoustics, Speech and Signal Processing // Canada, Toronto, May 14 - 17, 1991.* - P. 913 - 916.
3. Furui S. Cepstral analysis technique for automatic speaker verification. *IEEE Trans. Acoust. Speech Signal Process.* - 1981. - V. 29 (2). - P. 254 - 272.
4. Viikki O., Bye D., Laurila K. A recursive feature vector normalization approach for robust speech recognition in noise/ *International Conference on Acoustics, Speech and Signal Processing // USA, Washington, May 14 - 15, 1998.* - P. 733 - 736.
5. A de La Torre, Peinado A., Segura J., Perez-Cordoba J., Benitez M. A Rubio, Histogram equalization of speech representation for robust speech recognition. *IEEE Trans. Speech Audio Process.* - 2005. - V. 13 (3). - P. 355 - 366.
6. Marcelo A. Colominasa, Gaston Schlotthauera, Maria E. Torres. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control.* - 2014. - V. 14. - P. 19 - 29.
7. Алимуратов А.К. Разработка алгоритма обработки речевых сигналов для определения информативно-значимых параметров пограничных психиче-



ских расстройств / А.К. Алимуратов, А.Ю. Тычков, П.П. Чураков, А.В. Агейкин // Известия высших учебных заведений. Поволжский регион. Технические науки. - 2017. - № 3 (47). - С. 4 - 25.

8. Schuller B.W., Batliner A.M. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. New York: Wiley, 2013, 344 p.

А.К. Алимуратов, А.Ю. Тычков, П.П. Чураков, А.В. Агейкин

## ОЦЕНКА ПСИХОЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ НА ОСНОВЕ АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ: СОВРЕМЕННОЕ СОСТОЯНИЕ, ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

(Пензенский государственный университет)

### Введение

Оценка психоэмоционального состояния населения на сегодняшний день является социально-значимой проблемой для каждого государства, поскольку напрямую связана с формированием здорового образа жизни населения и профилактикой психических заболеваний. По данным ВОЗ современные социально-значимые заболевания, являющиеся главной причиной временной нетрудоспособности, инвалидности и смертности населения, напрямую связаны с психоэмоциональным состоянием населения.

В статье кратко представлен обзор современного состояния области оценки психоэмоционального состояния человека на основе речевых сигналов; выявлены проблемы и определены перспективы развития данной области. Работа выполнена при финансовой поддержке Совета по грантам Президента РФ в рамках проекта № СП-246.2018.5 и является продолжением ранее опубликованных трудов авторов [1, 2].

Обзорный анализ [3 - 5] выявил, что оценка психоэмоционального состояния человека на основе анализа речи сводится к выполнению трех ключевых задач: выбор/создание базы исследуемых речевых сигналов, выделение информативных параметров и построение алгоритма классификации психоэмоциональных состояний.

### База данных речевых сигналов

Проведен обзор 64 баз данных эмоционально окрашенных речевых сигналов [5]. Базы были систематизированы по следующим параметрам: язык; количество и профессия испытуемых; физиологические особенности; виды психоэмоциональных состояний (радость, нейтралитет, скука, печаль, раздражение, страх, отвращение и гнев); условия их воспроизведения (вынужденные, смоделированные и естественные). Вынужденные и смоделированные эмоции, как правило, воспроизводятся профессиональными актерами. Речь, произносимая профессионалами, является самой эмоциональной окрашенной, так как эмоции актеров имеют большой диапазон амплитуды и продолжительности. Но основ-