DOI: 10.15593/2224-9397/2020.3.01
УДК 004.934

А.К. Алимурадов, А.Ю. Тычков

Пензенский государственный университет, Пенза, Россия

ПРИМЕНЕНИЕ МЕТОДА ДЕКОМПОЗИЦИИ НА ЭМПИРИЧЕСКИЕ МОДЫ ДЛЯ ИССЛЕДОВАНИЯ ВОКАЛИЗОВАННОЙ РЕЧИ В ЗАДАЧЕ ОБНАРУЖЕНИЯ СТРЕССОВЫХ ЭМОЦИЙ ЧЕЛОВЕКА

Актуальность и цель. Эмоции человека могут выражаться по нескольким модальным системам: речь, мимика и микроэкспрессии лица, глазодвигательная активность, движение и позиции тела, биохимический анализ крови и др. Наиболее перспективным и адаптивным к современным условиям деятельности человека является способ распознавания эмоций на основе анализа речевых сигналов. Точность распознавания эмоций человека зависит от корректного выделения информативных параметров, отражающих эмоциональную составляющую речи. Целью данной работы является исследование информативных параметров вокализованной речи, релевантных нарушениям работы органов речевого аппарата вследствие эмоционального стрессового возбуждения человека. **Материалы и методы.** В рамках исследования использовалась уникальная адаптивная технология анализа нестационарных данных – улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом. Исследования были реализованы в среде математического моделирования MatLab. **Результаты.** Разработан способ обработки вокализованной речи для применения в интеллектуальных системах распознавания стрессовых состояний человека. Способ основан на принципе, что вокализованная речь в полном объеме отражает нарушение работы органов речевого аппарата вследствие эмоционального стресса. Суть способа заключается в разложении вокализованной речи на эмпирические моды с помощью улучшенной декомпозиции, в выделении мод, содержащих периодическую информацию об источнике возбуждения голосового тракта и в формировании комплексного сигнала, отражающего информацию о глоттальной активности во время вокализации речи. Приведены результаты исследования способа, представляющие собой вычисление частоты основного тона 100 мультигармонических сигналов с модуляцией в диапазоне 0–2,5 Гц/мс с шагом 0,5 Гц/мс. Модуляция имитировала нерегулярность колебаний голосовых связок (30–40 % от номинального значения), возникающих вследствие эмоционального стрессового возбуждения человека. **Практическая значимость**. В соответствии с результатами исследований предложенный способ обработки вокализованной речи обеспечивает устойчивое измерение частоты основного тона, в том числе и при наибольшем значении модуляции 2,5 Гц/мс. На основе этого сделан вывод, что предложенный способ может успешно тестироваться в интеллектуальных системах распознавания стрессовых состояний человека.

Ключевые слова: интеллектуальные системы, распознавание эмоций человека, стресс, обработка речевых сигналов, вокализованная речь, декомпозиция на эмпирические моды.

A.K. Alimuradov, A.Yu. Tychkov

Penza State University, Penza, Russian Federation

APPLICATION OF THE METHOD EMPIRICAL MODE DECOMPOSITION FOR THE STUDY OF VOICED SPEECH IN THE PROBLEM OF DETECTING HUMAN STRESS EMOTIONS

Relevance and purpose. Human emotions can be expressed in several modal systems: speech, facial expressions and micro-expressions, oculomotor activity, body movement and position, biochemical blood test, etc. A method for recognizing emotions based on the analysis of speech signals is the most promising and adaptive one to modern conditions of human activity. Recognition accuracy of human emotions depends on correct determination of informative parameters that reflect an emotional component of speech. The aim of this work is to study the informative parameters of voiced speech that are relevant to disorders in functioning of the speech apparatus organs due to emotional stress arousal in a person. **Materials and methods.** A unique adaptive technology for analyzing non-stationary data, namely, an improved complete ensemble empirical mode decomposition with adaptive noise, was used to conduct the research. The research was carried out in a mathematical modeling environment MatLab. **Results.** A method for processing voiced speech to be used in intelligent systems for recognizing human stressed states has been developed. The method is based on the principle that voiced speech fully reflects a disruption in the functioning of speech apparatus organs due to emotional stress. The method consists in decomposing voiced speech into empirical modes using improved decomposition, extracting modes containing periodic information about the excitation source of the vocal tract, and generating a composite signal reflecting information on glottal activity during speech vocalization. The results of investigating the method are presented, being the calculation of pitch frequency for 100 multiharmonic signals with modulation in the range of 0-2,5 Hz/ms, with a step of 0,5 Hz/ms. The modulation imitated the irregularity of vibrations in the vocal cords (30-40% of the nominal value) due to emotional stress arousal in a person. **Conclusions.** In accordance with the research results, the proposed method for processing voiced speech provides a stable measurement of the pitch frequency, including at the highest modulation value of 2,5 Hz/ms.It was concluded that the proposed method can be successfully tested in intelligent systems for recognizing human stressed states.

Keywords: intelligent systems, recognition of human emotions, stress, processing of speech signals, voiced speech, empirical mode decomposition.

Введение. Интеллектуальные системы обнаружения, распознавания, классификации и оценки эмоций человека получили широкое практическое применение в различных сферах человеческой деятельности: искусственный интеллект, робототехнические устройства, безопасность, цифровая медицина, коммуникации, банковская и страховая отрасли, игровая индустрия, нейромаркетинг и др.

Эмоции человека могут выражаться по нескольким модальным системам: речь, мимика и микроэкспрессии лица, глазодвигательная активность, движение и позиции тела, а также могут выражаться на уровне физиологии (сердцебиение, дыхание, пульс и др.). Точность распознавания эмоций по каждой отдельной модальности в среднем на 9,8 % ниже точности для совокупного анализа данных всех модальных систем од-

новременно [1]. Эти данные актуальны и для анализа речи [2]. Однако поскольку распознавание эмоций по речи не зависит от ментальности, национальности и культуры человека, данное направление считается достаточно перспективным. Справедливо отметить, что в сравнении с другими областями речевых технологий задача распознавания эмоций человека по речи в настоящий момент находится на самом начальном этапе ее решения.

Точность распознавания эмоций человека зависит от корректного выделения информативных параметров, отражающих эмоциональную составляющую речи. Анализ открытых источников отечественной и зарубежной литературы выявил, что вопрос разработки высокоэффективных подходов к выделению информативных параметров речи, релевантных нарушений работы органов речевого аппарата остается нерешенным и требует дальнейшей проработки [3–10]. Данная статья является продолжением научных трудов авторов [11–13], в рамках которых оценивалась возможность существующих способов и подходов к анализу и обработке речевых сигналов, а также анализировалась необходимость применения новых математических аппаратов в интеллектуальных системах распознавания эмоций.

Статья посвящена исследованию вокализованной речи с применением уникальной адаптивной технологии анализа нестационарных сигналов – декомпозиции на эмпирические моды (ДЭМ) для интеллектуальных систем распознавания стрессовых состояний человека. Работа выполнена при финансовой поддержке Совета по грантам Президента РФ в рамках проекта № МК-490.2020.8.

Структурно статья состоит из пяти разделов. Первый и второй разделы посвящены краткому обзору эмоциональной речи человека, а также адаптивной технологии разложения сигналов – ДЭМ. Третий и четвертые разделы посвящены описанию и исследованию предложенного способа обработки вокализованной речи. Последний раздел посвящен выводам и перспективам дальнейших исследований.

1. Эмоциональная речь. Речь представляет собой нестационарный акустический сигнал сложной формы. Различные изменения в вегетативной нервной системе могут изменить речь человека. Например, речь в состоянии страха, гнева или радости воспроизводится быстро, громко и четко сформулирована, с более высоким и широким диапазоном высоты основного тона (ОТ), в то время как при усталости, скуке, или печали речь воспроизводится человеком медленно и невнятно.

Наилучшие результаты достигаются при распознавании противоположных эмоций: отрицательных (гнев, страх, отвращение, грусть) и положительных (удивление, радость) с нейтральным состоянием.

Речь человека состоит из вокализованных/невокализованных участков – участков пауз и дыхания. Информативные параметры вокализованной речи в полном объеме отражают нарушения работы органов речевого аппарата вследствие эмоционального стрессового возбуждения человека [2]. Вокализованная речь образуется в результате возбуждения голосового тракта, обусловленного колебаниями голосовых связок в области голосовой щели (глоттиса). Сила возбуждения во время глоттальной активности определяется в основном скоростью смыкания голосовых связок. Периодические колебания голосовых связок во время возбуждения голосового тракта называются основным тоном (ОТ). Величина, обратная значению ОТ, называется частотой основного тона (ЧОТ) и является важным информативным параметром вокализованной речи [14].

2. Декомпозиция на эмпирические моды (ДЭМ) представляет собой уникальную адаптивную технологию анализа нестационарных данных, не требующую никакой априорной информации об исследуемом сигнале для разложения на частотные составляющие [15]. Адаптивность ДЭМ позволяет эффективно применять ее для анализа естественных сигналов. Разложение с помощью ДЭМ обеспечивает извлечение из сигнала различных колебательных функций, называемых эмпирическими модами (ЭМ), каждая из которых имеет свой частотный диапазон. Исследования показали, что ДЭМ представляет собой диадический набор фильтров, причем первая ЭМ содержит высокочастотные колебания сигнала, а последующие моды находятся в низкочастотном диапазоне [15]. Однако для естественных сигналов, например нестационарной речи, частотное разделение с помощью ДЭМ оказывается неэффективным, поскольку одна ЭМ может содержать разные диапазоны частот, или определенный частотный диапазон может присутствовать в разных модах. Это явление называется смешиванием ЭМ. Чтобы решить эту проблему, была предложена множественная декомпозиция на эмпирические моды (МДЭМ). В МДЭМ многочисленные реализации белого шума конечной амплитуды смешиваются с исходным сигналом образуя зашумленные копии. Затем осуществляется разложение зашумленных копий сигнала с помощью ДЭМ [16]. Усреднение по множеству полученных мод (сгенерированных для всех шумовых копий) представляет собой окончательный набор ЭМ. Поскольку белый шум в равной степени присутствует на всех частотах, то при

смешивании с сигналом он усиливает скрытые или подавленные частотные составляющие. Основным недостатком МДЭМ является остаточный шум. В работе [17] отмечено, что при добавлении к исходному сигналу пары сигналов белого шума, один из которых противоположной полярности, можно значительно уменьшить остаточный шум в полученных модах. Этот метод был назван комплементарной МДЭМ (КМДЭМ) [17]. Однако каждая реализация белого шума при смешивании с сигналом может привести к разному количеству получаемых мод, что создает трудности при их усреднении. Чтобы избежать этого, была предложена полная множественная декомпозиция на эмпирические моды с адаптивным шумом (ПМДЭМАШ), которая параллельно разлагает белый шум вместе с сигналом и таким образом использует белый шум для извлечения полного окончательного набора ЭМ [18]. В качестве дополнительного улучшения метода была предложена улучшенная ПМДЭМАШ, которая решает проблему паразитных мод, возникающих в ПМДЭМАШ на ранних этапах разложения [19]. Ниже представлено краткое математическое описание ДЭМ и ее модификаций с добавлением шума:

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

где $x(n)$ – анализируемый сигнал; n – дискретный отсчет времени в сигнале; $i = 1, 2, \ldots, I$ – номер ЭМ; $IMF_i(n)$ – полученные в результате разложения ЭМ; $r_i(n)$ – конечный неделимый остаток (последняя мода);

$$x_j(n) = x(n) + w(n), \quad (2)$$

где $x_j(n)$ – зашумленные сигналы; $w_j(n)$ – белый шум,

$$x_j(n) = \sum_{i=1}^{I} IMF_{ji}(n) + r_{jI}(n), \quad (3)$$

$$IMF_i(n) = \sum_{j=1}^{J} \frac{IMF_{ji}(n)}{J}, \quad (4)$$

$$r_I(n) = \sum_{j=1}^{J} \frac{r_{jI}(n)}{J}, \quad (5)$$

где J – количество реализаций белого шума, $j = 1, 2, \ldots,$.

Добавление контролируемого шума устраняет известные недостатки существующих модификаций декомпозиции: эффект смешивания ЭМ; неполноту декомпозиции (все полученные шумовые копии разлагаются независимо друг от друга без связи между собой); остаточный белый шум; неинформативные «паразитные» ЭМ, отсеиваемые на ранних этапах разложения.

На рис. 1 представлен пример разложения на ЭМ участка вокализованной речи длительностью 100 мс с помощью улучшенной

ПМДЭМАШ. В левом столбце представлены осциллограммы исходного сигнала и полученных ЭМ, в правом столбце – спектральные плотности мощности. Разными цветами обозначены частотные диапазоны соответствующих мод. Как видно из рис. 1, метод улучшенной ПМДЭМАШ в действительно функционирует как диадический набор фильтров, понижая частотный диапазон ЭМ от высокочастотного до низкочастотного.

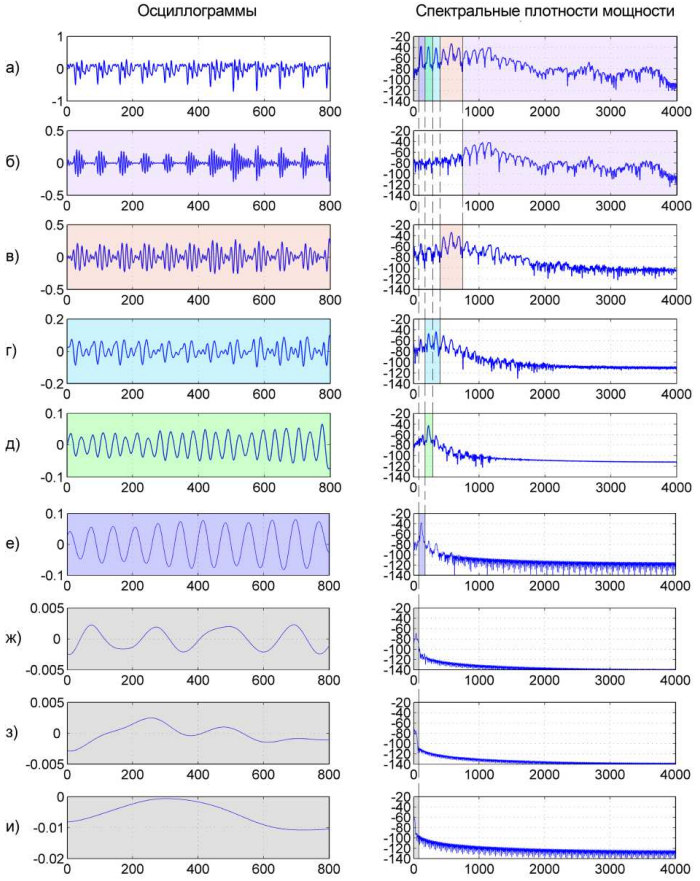

Рис. 1. Результат разложения участка вокализованной речи с помощью улучшенной ПМДЭМАШ: в левом столбце представлены осциллограммы (амплитуда, В, время в дискретных отсчетах) исходного сигнала и полученных ЭМ, в правом столбце – спектральные плотности мощности (амплитуда, дБ, частота, Гц); (*а*) – исходный речевой сигнал, (*б*)–(*и*) – ЭМ1–ЭМ8

3. Описание способа обработки вокализованной речи. На рис. 2 структурно представлен способ обработки вокализованной речи на основе ДЭМ для интеллектуальных систем распознавания стрессовых со-

стояний человека. Суть способа заключается в выполнении последовательности следующих этапов обработки:

– разложение вокализованной речи на ЭМ (блок 1);

– выделение ЭМ, содержащих периодическую информацию об источнике возбуждения голосового тракта (блоки 2–4);

– формирование комплексного сигнала, отражающего информацию о глоттальной активности (блок 5);

– определение информативных параметров вокализованной речи (блок 6).

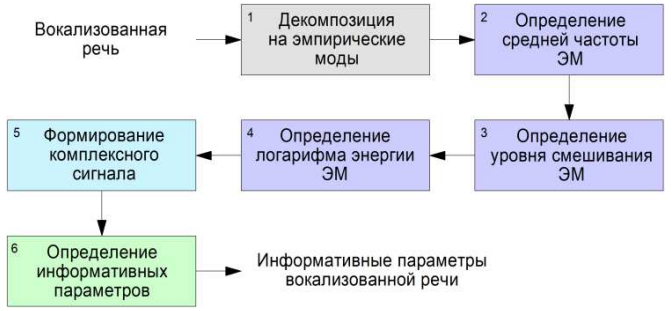

Рис. 2. Структура способа обработки вокализованной речи на основе ДЭМ

Рассмотрим некоторые этапы обработки подробнее. В соответствии с результатами анализа разных модификаций ДЭМ [19] сделан вывод, что наиболее адаптивным к речевым сигналам является метод улучшенной ПМДЭМАШ. На рис. 3 представлены результаты разложения вокализованного речевого сигнала (длительностью 2 с, частотой дискретизации 8000 Гц и квантованием 16 бит). Для простоты визуализации на рис. 3 отображены ЭМ, полученные с помощью ДЭМ, МДЭМ и улучшенной ПМДЭМАШ, так как МДЭМ и КМДЭМ, а также ПМДЭМАШ и улучшенная ПМДЭМАШ аналогичны с точки зрения просеивания ЭМ соответственно, а также для удобства визуализации амплитуды осциллограмм соответствующих ЭМ представлены в одном масштабе. При разложении использовались следующие настройки декомпозиции: стандартное отклонение шума – 20 % от стандартного отклонения сигнала; количество реализаций (добавлений шума) – 100; допустимое максимальное количество просеивающих итераций – 20; отношение стандартных отклонений сигнала и шума на всех этапах просеиваний ЭМ оставалось неизменным (для улучшенной ПМДЭМАШ).

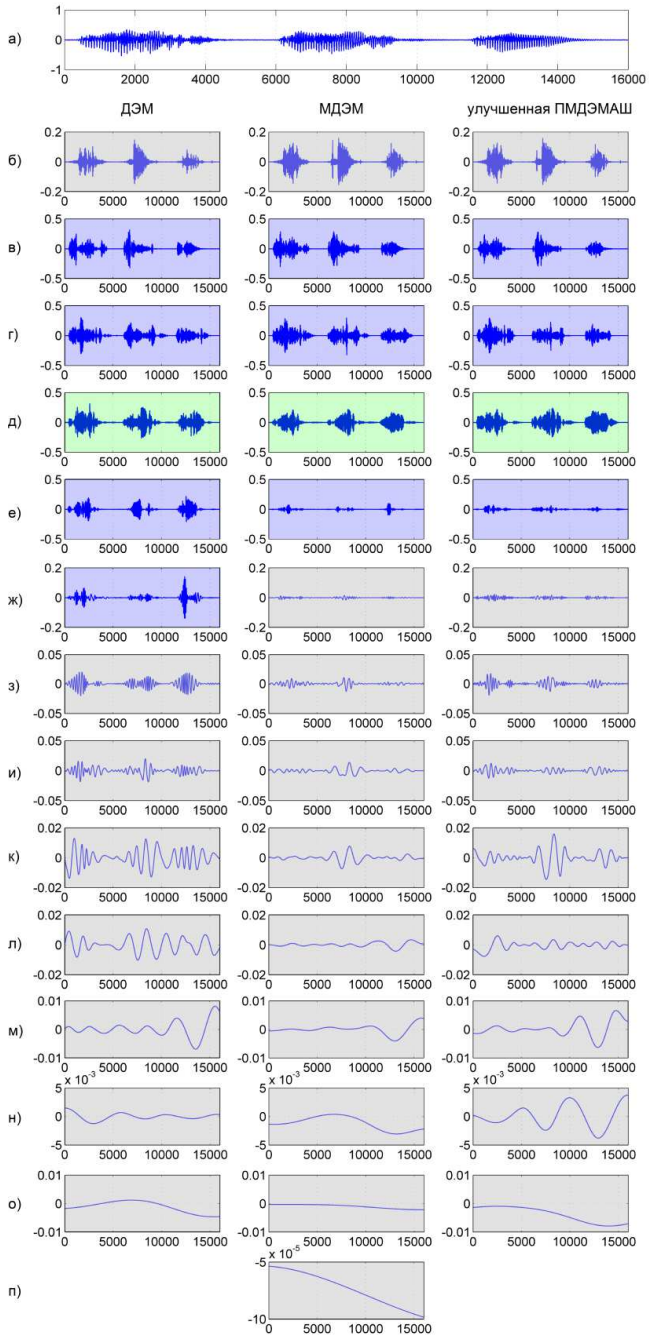

Рис. 3. Результаты разложения вокализованного речевого сигнала: в левом столбце представлены осциллограммы (амплитуда, В, время в дискретных отсчетах) ЭМ, полученных с помощью ДЭМ, в среднем столбце – с помощью МДЭМ и в правом столбце – с помощью улучшенной ПМДЭМШ; (*а*) – исходный речевой сигнал, (*б*)–(*п*) – ЭМ1–ЭМ14

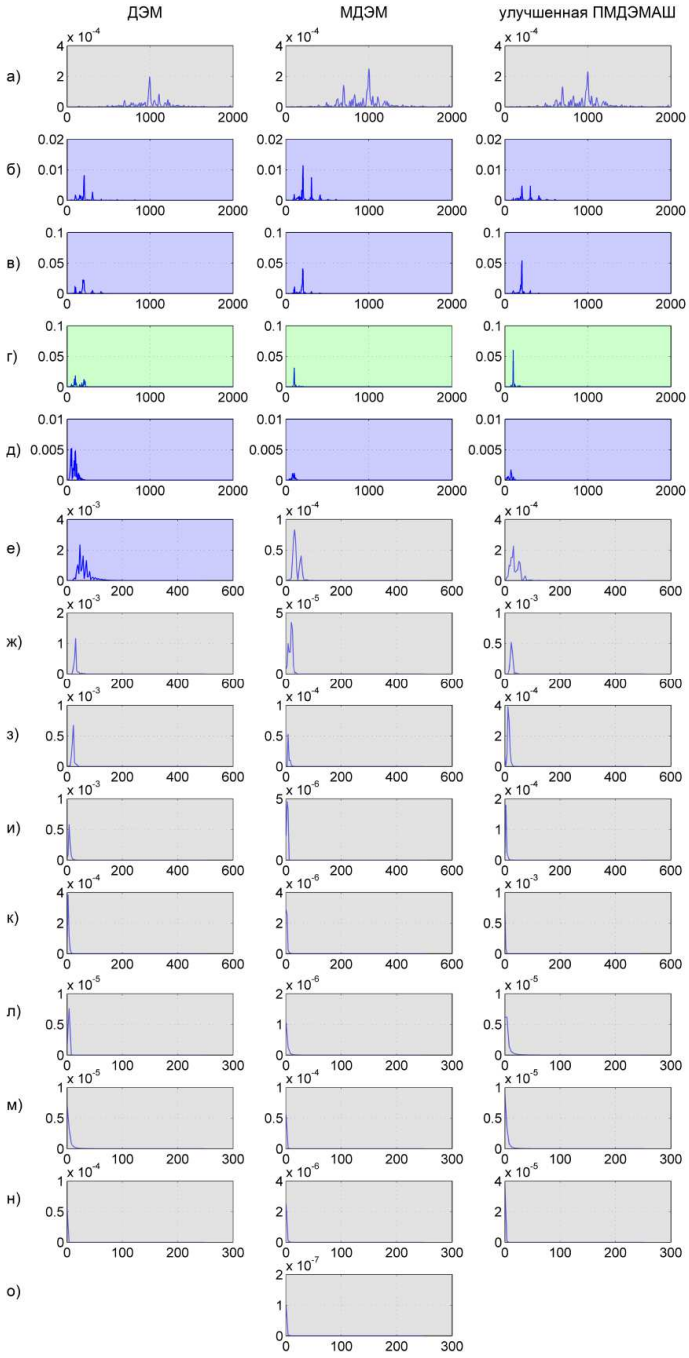

Рис. 4. Частотные спектры ЭМ: в левом столбце представлены частотные спектры (амплитуда, $В^2$, частота, Гц) ЭМ, полученных с помощью ДЭМ, в среднем столбце – с помощью МДЭМ и в правом столбце – с помощью улучшенной ПМДЭМШ; (*а*)–(*о*) – частотные спектры ЭМ1–ЭМ14

Из рис. 3 следует, что вокализованный речевой сигнал разложен с помощью ДЭМ, МДЭМ и улучшенной ПМДЭМАШ на 13, 14 и 15 мод соответственно. В каждом наборе ЭМ присутствуют моды, содержащие периодическую информацию об источнике возбуждения голосового тракта. Для оценки качества разложения сигнала на отдельные частотные диапазоны на рис. 4 представлены частотные спектры ЭМ, вычисленные с помощью быстрого преобразования Фурье.

Для удобства визуализации для первых пяти ЭМ амплитуды и частоты спектров представлены в одном масштабе, а для остальных мод – в произвольном. В соответствии с рис. 4 можно сделать предварительный вывод, что наибольший эффект смешивания частотных диапазонов наблюдается у ЭМ, полученных с помощью ДЭМ.

Процесс выделения ЭМ, содержащих периодическую информацию об источнике возбуждения голосового тракта, сводится к определению следующих параметров ЭМ: средней частоты, логарифма энергии и уровня смешивания.

Вычисление средней частоты ЭМ осуществляется по формуле

$$F_{IMF} = \frac{\sum_0^{F_s/2} f \cdot S_{IMF}(f)}{\sum_0^{F_s/2} S_{IMF}(f)}, \quad (6)$$

где F_{IMF} – средняя частота ЭМ со спектром мощности $S_{IMF}(f)$ (вычисление спектра мощности осуществляется с использованием быстрого преобразования Фурье с размерностью $N = 2048$ отсчетов); F_s – частота дискретизации речевого сигнала.

В процессе вокализации речи ЧОТ взрослого человека (мужчин и женщин) находится в диапазоне от 40 до 400 Гц. Используя эти данные, можно выделить ЭМ, которые содержат периодическую информацию об источнике возбуждения голосового тракта. На рис. 5 представлены значения средних частот ЭМ, вычисленные по формуле (6). Для удобства визуализации средние частоты первых ЭМ не представлены на рис. 5, так как имеют большие значения (1150,4 Гц – для ДЭМ, 1039,2 Гц – для МДЭМ и 1048,7 Гц – для улучшенной ПМДЭМАШ).

Из рис. 5 следует, что средние частоты ЭМ2–ЭМ6 находятся в диапазоне от 40 до 400 Гц. Остальные моды в последующей обработке

не используются. Уровень смешивания для пары последовательных ЭМ (Degree of Mode Mixing, DMM), обозначает сходство средних частот i-й и (i+1)-й мод и вычисляется как [20]:

$$DMM_i = \left[1 - \frac{F_i + F_{i+1}}{F_i/2}\right] \cdot 100\ \%, \qquad (7)$$

где DMM_i – уровень смешивания мод; F_i и F_{i+1} – средние частоты i-й и (i+1)-й ЭМ.

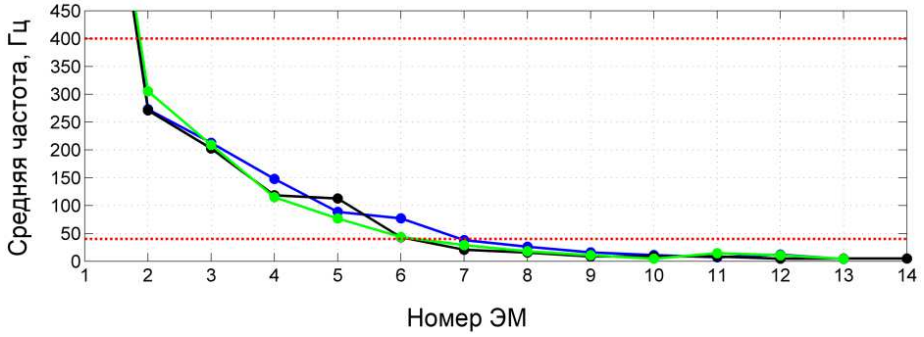

Рис. 5. Средние частоты ЭМ: синим цветом представлены значения средних частот ЭМ, полученных с помощью ДЭМ, черным цветом – с помощью МДЭМ и зеленым цветом – с помощью улучшенной ПМДЭМАШ; пунктирные линии красного цвета отображают диапазон ЧОТ от 40 до 400 Гц

Если улучшенная ПМДЭМАШ обеспечивает почти идеальный диадический набор фильтров, то значение уровня смешивания для пары последовательных ЭМ будет близко к 0 %. Если моды будут отличаться по частоте, то значение уровня смешивания будет отрицательным, что указывает на отсутствие эффекта смешивания мод. Если моды имеют одинаковую среднюю частоту, то значение уровня смешивания будет 100 %, и это означает, что информация о возбуждении голосового тракта вследствие смыкания голосовых связок распределена между ними. На рис. 6 представлены значения уровня смешивания для пар последовательных ЭМ, вычисленные по формуле (7). Для удобства визуализации значение уровня смешивания для пары ЭМ10 и ЭМ11, полученное с помощью улучшенной ПМДЭМАШ, не представлено на рис. 6, так как имеет большое значение (523,88 %).

В соответствии с рис. 6 можно сделать вывод, что во всех наборах ЭМ имеются пары, в которых информация о возбуждении голосового тракта распределена между сигналами мод:

– ДЭМ: пара ЭМ11, ЭМ12;

– МДЭМ: пары ЭМ4, ЭМ5 и ЭМ9, ЭМ10, а также ЭМ12, ЭМ13 и ЭМ13, ЭМ14;

– улучшенная ПМДЭМАШ: пара ЭМ9, ЭМ10.

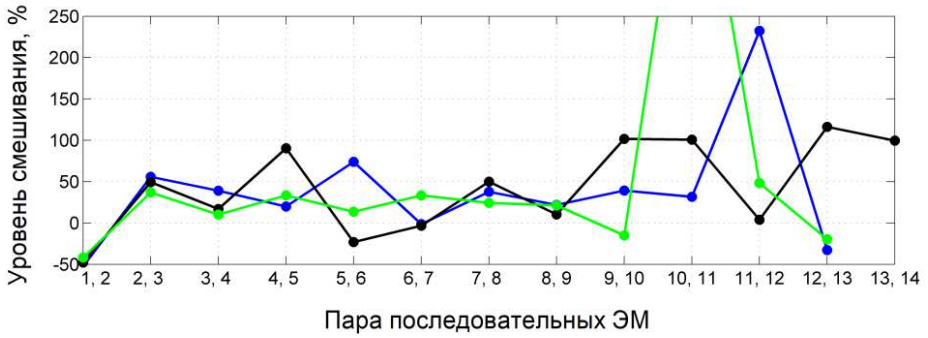

Рис. 6. Уровень смешивания для пар последовательных ЭМ: синим цветом представлены значения уровня смешивания для пар последовательных ЭМ, полученных с помощью ДЭМ, черным цветом – с помощью МДЭМ и зеленым цветом – с помощью улучшенной ПМДЭМАШ

Для последующей обработки вышеупомянутые сигналы ЭМ в парах необходимо объединить суммированием. Таким образом, формируются новые наборы мод: 12 ЭМ – для ДЭМ, 10 ЭМ – для МДЭМ и 12 ЭМ – для улучшенной ПМДЭМАШ.

Распределение амплитуды вокализованной речи во времени достаточно полно описывается с помощью функции логарифма кратковременной энергии. В соответствии с функционалом слухового аппарата человек воспринимает речь нелинейно, определяя разницу между энергиями информативных участков речи. Приближая работу способа к функционалу слухового аппарата, для сжатия амплитуды сигнала в большом динамическом диапазоне применяют логарифмирование энергии:

$$LE_i = \log_2\left(\sum_{n=1}^{N}(IMF_i(n))^2\right), \quad (8)$$

где LE_i – логарифм энергии ЭМ.

Известно, что вокализованная речь имеет большую энергию, чем невокализованная. Поэтому моды, содержащие энергию, меньшую чем 5 % от общей энергии вокализованного сигнала, в последующей обработке не используются [21]. На рис. 7 представлены значения логарифмов энергии ЭМ, вычисленные по формуле (8).

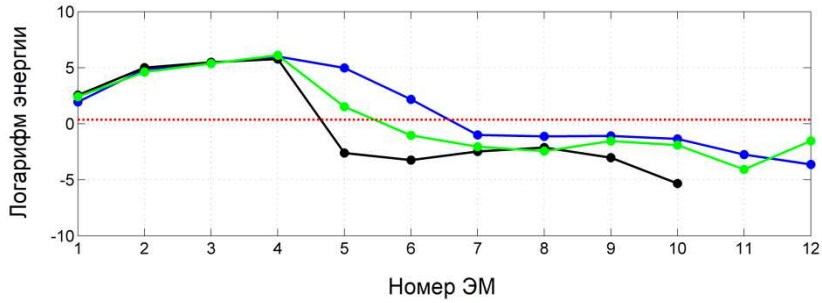

Рис. 7. Логарифмы энергии ЭМ: синим цветом представлены значения логарифмов энергии ЭМ, полученных с помощью ДЭМ, черным цветом – с помощью МДЭМ и зеленым цветом – с помощью улучшенной ПМДЭМАШ; пунктирной линией красного цвета отображается уровень, соответствующий 5 % от общей энергии вокализованного сигнала

В соответствии с полученными параметрами мод (средней частоты, логарифма энергии и уровня смешивания) осуществляется формирование комплексного сигнала, отражающего информацию о глоттальной активности. Формирование представляет собой суммирование сигналов ЭМ, содержащих периодическую информацию об источнике возбуждения голосового тракта:

- ДЭМ: суммирование ЭМ2–ЭМ6;
- МДЭМ: суммирование ЭМ2–ЭМ5;
- улучшенная ПМДЭМАШ: суммирование ЭМ2–ЭМ5.

Моды, содержащие периодическую информацию об источнике возбуждения голосового тракта, выделены на рис. 3 и 4 синим и зеленым цветом. Остальные неинформативные ЭМ на последующих этапах обработки не используются и выделены серым цветом.

Определение информативных параметров вокализованной речи представляет собой вычисление ЧОТ с помощью апробированного авторами алгоритма [13, 22]. Суть алгоритма заключается в определении моды, содержащей ОТ, и в вычислении ЧОТ с помощью функции оператора Тигера [22, 23]. Определение ЭМ, содержащей ОТ, осуществляется среди мод комплексного сигнала, содержащих периодическую информацию об источнике возбуждения голосового тракта. Суть определения моды, содержащей ОТ, заключается в вычислении разницы логарифмов энергии пары последовательных ЭМ по формуле:

$$d_i = LE_{i+1} - LE_i, \tag{9}$$

где d_i – разница логарифмов энергии пары последовательных i-й и (i+1)-й мод.

На рис. 8 представлены значения разницы логарифмов энергии пары последовательных ЭМ, вычисленные по формуле (9).

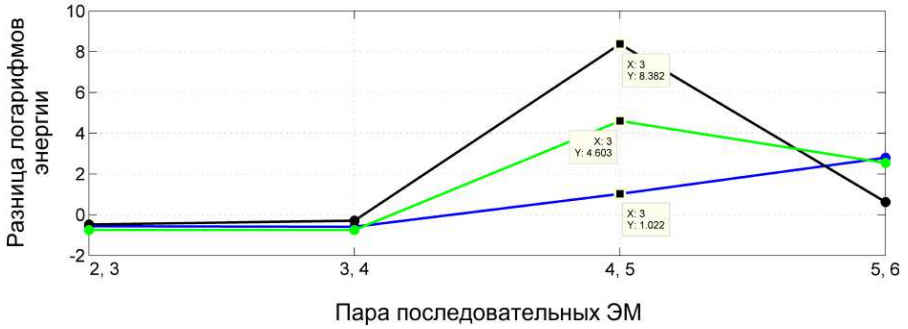

Рис. 8. Разница логарифмов энергии пары последовательных ЭМ: синим цветом представлены значения разницы логарифмов энергии пары последовательных ЭМ, полученных с помощью ДЭМ, черным цветом – с помощью МДЭМ и зеленым цветом – с помощью улучшенной ПМДЭМАШ

Большему значению разницы логарифмов энергии пары последовательных ЭМ соответствует наличие в паре моды, содержащей ОТ [22, 23]. Из рис. 8 следует, что для ЭМ, полученных с помощью МДЭМ и улучшенной ПМДЭМАШ, большее значение разницы логарифмов энергии имеет пара ЭМ4, ЭМ5. Модой, содержащей ОТ, является ЭМ4 [22, 23]. Для ЭМ, полученных с помощью ДЭМ, большее значение разницы логарифмов энергии имеет пара ЭМ5, ЭМ6, однако модой, содержащей ОТ, является ЭМ4. Это подтверждается значениями средней частоты, логарифма энергии и уровня смешивания ЭМ4. Неоднозначность в определении моды, содержащей ОТ, при использовании ДЭМ возникает из-за сильно выраженного эффекта смешивания ЭМ. Моды, содержащие ОТ, отмечены зеленым цветом на рис. 3 и 4.

На рис. 9 представлены частотные спектры вокализованного речевого сигнала и мод, содержащих ОТ, полученных с помощью ДЭМ, МДЭМ и улучшенной ПМДЭМАШ. Для удобства визуализации частотный спектр представлен в диапазоне нахождения ЧОТ от 0 до 200 Гц.

Из рис. 9 следует, что наилучшие результаты определения ЭМ, содержащей ОТ, достигаются при использовании МДЭМ и улучшенной ПМДЭМАШ. Наблюдается четко выраженный пик, соответствующий ЧОТ (101,6 Гц для рассматриваемого примера), имеющий мак-

симальную амплитуду. Наихудший результат достигается при использовании ДЭМ. Наблюдается низкоуровневый и невыраженный пик, соответствующий ЧОТ с близко расположенной и соизмеримой по амплитуде помехой на частоте 99,8 Гц.

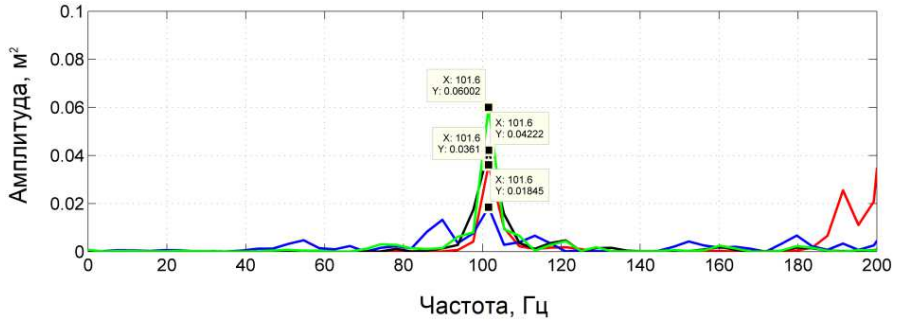

Рис. 9. Частотные спектры: красным цветом представлен частотный спектр вокализованного речевого сигнала, синим цветом – частотный спектр ЭМ, содержащей ОТ, полученной с помощью ДЭМ, черным цветом – с помощью МДЭМ, зеленым цветом – с помощью улучшенной ПМДЭМАШ

4. Исследование способа обработки вокализованной речи. Для проведения исследования сформирована тестовая выборка из 100 мультигармонических сигналов, представляющих сумму нескольких гармонических составляющих ОТ с заранее известной частотой. В рамках исследования каждый тестовый сигнал подвергался частотному модулированию в диапазоне 0–2,5 Гц/мс с шагом 0,5 Гц/мс. В качестве критериев оценки эффективностиисследования использовались: коэффициент грубых ошибок (Gross Pitch Error, GPE) – безразмерная величина, равная отношению числа фрагментов с отклонением измеренного значения ЧОТ более чем на 20 % от истинного значения к общему числу фрагментов, содержащих информацию об источнике возбуждения голосового тракта; средний коэффициент мелких ошибок (MeanFinePitchError, MFPE) – безразмерная величина, равная среднему значению отношения разности истинного и измеренного значений ЧОТ к истинному значению фрагментов без грубых ошибок. В таблице представлены усредненные результаты вычисления ЧОТ 100 комплексных сигналов, сформированных суммированием вокализованных ЭМ, полученных с помощью различных модификаций декомпозиции: ДЭМ, МДЭМ, КМДЭМ, ПМДЭМАШ и улучшенной ПМДЭМАШ.

Результаты вычисления ЧОТ комплексного сигнала, отражающего информацию о глоттальной активности

Модуляция частоты, Гц/мс	Модификации декомпозиции					Критерии оценки
	ДЭМ	МДЭМ	КМДЭМ	ПМДЭМАШ	Улучшенная ПМДЭМАШ	
0	0	0	0	0	0	GPE
	6,10	1,56	1,35	1,10	0,70	MFPE
0,5	0	0	0	0	0	GPE
	7,56	2,65	2,34	1,21	0,93	MFPE
1,0	0	0	0	0	0	GPE
	9,30	3,25	2,81	2,54	1,75	MFPE
1,5	0	0	0	0	0	GPE
	10,21	4,50	3,90	3,00	2,38	MFPE
2,0	2,05	0	0	0	0	GPE
	12,20	5,10	4,67	4,54	3,88	MFPE
2,5	7,60	5,30	4,20	3,50	2,70	GPE
	17,20	7,10	6,50	6,20	5,32	MFPE

Результаты и выводы. В соответствии с полученными данными можно сделать вывод, что наилучшие результаты вычисления ЧОТ достигаются при исследовании вокализованной речи с помощью улучшенной ПМДЭМАШ, в том числе и при больших значениях модуляции ЧОТ. Это объясняется тем, что при использовании улучшенной ПМДЭМАШ сформированный комплексный сигнал отражает максимальное количество периодической информации об источнике возбуждения голосового тракта. Справедливо отметить, что остальные методы декомпозиции с добавлением шума также обеспечивают приемлемые результаты вычисления ЧОТ. Однако значения коэффициентов грубых (GPE) и мелких ошибок, и (MFPE) больше по причине известных недостатков [15–19]. Из-за сильно выраженного эффекта смешивания ЭМ наихудшие результаты достигаются при использовании ДЭМ.

Как отмечалось ранее, глоттальная активность относится к явлению, связанному с колебаниями голосовых связок во время образования вокализованной речи. При эмоциональном возбуждении колебания голосовых связок характеризуются нерегулярностью, возникающей вследствие неполного смыкания при вокализованной речи. При крайне высоком и низком возбуждении изменение ЧОТ может достигать 30–40 % от номинального значения, соответствующего нейтральному эмоциональному состоянию. Процесс модулирования ЧОТ в диапазоне 0–2,5 Гц/мс с шагом 0,5 Гц/мс в рамках исследования имитировал нерегулярность колебаний голосовых связок. Полученные результа-

ты вычисления ЧОТ позволяют сделать вывод, что предложенный способ обработки вокализованной речи с помощью улучшенной ПМДЭМАШ может успешно тестироваться в интеллектуальных системах распознавания стрессовых состояний человека.

Для оценки корректности формирования комплексного сигнала, отражающего информацию о глоттальной активности, в перспективе планируется также определять дополнительные информативные параметры вокализованной речи: интенсивность ОТ, динамику изменения интенсивности ОТ, динамику изменения ЧОТ, девиацию ЧОТ и отношение интенсивности гармоник к интенсивности ОТ.

Коллектив авторов выражает благодарность Совету по грантам Президента РФ за финансовую поддержку проекта № МК-490.2020.8 «Разработка способов и виртуальных средств адаптивной помехозащищенной обработки и обнаружения клинически значимых параметров медицинских электрических и акустических сигналов у пациентов с пограничными психическими расстройствами».

Библиографический список

1. Проектная компания и исследовательская лаборатория в области аффективных наук и когнитивных технологий: официальный сайт [Электронный ресурс]. – URL: https://neurodatalab.com (дата обращения: 18.05.2020).

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

3. Алимурадов А.К., Фокина Е.А., Журина А.Е. Краткий обзор существующих баз данных эмоциональной речи: современное состояние, проблемы и перспективы развития // Методы, средства и технологии получения и обработки измерительной информации («Шляндинские чтения – 2020»): материалы XII Междунар. науч.-техн. конф. с элементами науч. шк. и конкурсом науч.-исслед. работ для студ., аспирант. и молод. ученых (г. Пенза, 16–18 марта 2020 г.) / под ред. д-ра техн. наук Е.А. Печерской. – Пенза: Изд-во Пенз. ГУ, 2020. – С. 292–294.

4. Мобильное приложение для сбора диагностической информации посредством регистрации речевых сигналов / А.Н. Осипов, М.М. Меженная, Т.П. Куль, Я.Ю. Бобровская, Ю.Н. Рушкевич, С.А. Лихачев // Bigdata and advanced analytics. – 2018. – № 4. – С. 346–350.

5. Куль Т.П., Рушкевич Ю.Н., Лихачев С.А. Адаптация методов цифровой обработки сигналов к задаче анализа речи при неврологических патологиях // Доклады Белорус. гос. ун-та информатики и радиоэлектроники. – 2018. – № 7(117). – С. 128–132.

6. Давыдова А.Ю. Оценка эмоционального состояния человека по параметрам речевого сигнала // Биотехнические, медицинские и экологические системы, измерительные устройства и робототехнические комплексы – Биомедсистемы-2019: сб. тр. XXXII Всерос. науч.-техн. конф. студ., молод. ученых и спец. (г. Рязань, 4–6 декабря 2019 г.) / под общ. ред. В.И. Жулева. – Рязань: ИП Коняхин А.В. (BookJet), 2019. – С. 348–351.

7. Методическое и аппаратно-программное обеспечение для регистрации и обработки речевых сигналов с целью диагностики неврологических заболеваний / Т.П. Куль, М.М. Меженная, Ю.Н. Рушкевич, А.Н. Осипов, С.А. Лихачев, И.В. Рушкевич // Информатика. – 2019. – Т. 16, № 2. – С. 27–39.

8. Прохоренко Е.И., Соколова В.С., Колесников В.А. Выделение признаков эмоционально окрашенной речи в звукозаписи // Студенчество России: век XXI: материалы VI Всерос. молодеж. науч.-практ. конф.: в 4 ч. (г. Орел, 13 декабря 2018 г.). – Орел: Изд-во Орлов. гос. аграрного ун-та им. Н.В. Парахина, 2019. – С. 184–187.

9. Гай В.Е., Утробин В.А., Поляков И.В. Система оценки психоэмоционального состояния диктора по голосу // Сложность. Разум. Постнеклассика. – 2016. – № 2. – С. 75–80.

10. Астахов Д.А., Катаев А.В. Использование современных алгоритмов машинного обучения для задачи распознавания эмоций // Cloud of Science (Москва). – 2018. – Т. 5, № 4. – С. 664–679.

11. Способ определения формантной разборчивости речи для оценки психоэмоционального состояния операторов систем управления с высокой степенью ответственности / А.К. Алимурадов, А.Ю. Тычков, П.П. Чураков, Б.В. Султанов // Измерение. Мониторинг. Управление. Контроль. – 2019. – № 4(30). – С. 58–69.

12. Помехоустойчивый алгоритм определения просодических характеристик речевых сигналов для систем оценки психоэмоционального состояния человека / А.К. Алимурадов, А.Ю. Тычков, П.П. Чураков, Д.В. Артамонов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2019. – № 3(51). – С. 3–16.

13. Алимурадов А.К., Тычков А.Ю., Чураков П.П. Способ автоматизированной сегментации речевых сигналов для определения временных паттернов естественно выраженных психоэмоциональных состояний // Измерение. Мониторинг. Управление. Контроль. – 2019. – № 3(29). – С. 48–60.

14. Выбор оптимального набора информативных признаков для классификации эмоционального состояния диктора по голосу / А.Г. Давыдов, В.В. Киселёв, Д.С. Кочетков, А.В. Ткаченя // Компьютерная лингвистика и интеллектуальные технологии: материалы ежегодной междунар. конф. «Диалог» (Бекасово, 30 мая – 3 июня 2012 г.): в 2 т. – М.: Изд-во РГГУ, 2012. – Т. 1. – № 11. – С. 122–128.

15. 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. – A 454. – P. 903–995.

16. Zhaohua W., Huang N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method // Advances in Adaptive Data Analysis. – 2009. – No. 1(1). – P. 1–41.

17. 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. – No. 2(2). – P. 135–156.

18. A complete Ensemble Empirical Mode decomposition with adaptive noise / M.E. Torres, M.A. Colominas, G. Schlotthauer, P. Flandrin // 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-11) (May 22–27, 2011). – Prague, Czech Republic, 2011. – P. 4144–4147.

19. 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.

20. Sharma R., Prasanna S.R.M. Characterizing glottal activity from speech using empirical mode decomposition // 2015 21st National Conference on Communications (NCC) (27 February–1 March, 2015). – Mumbai, India. – P. 1–6.

21. Lee H.S. Improvement of decomposing results of empirical mode decomposition and its variations for sea-level records analysis // Journal of Coastal Research: proceedings of the 15th International Coastal Symposium. – 2018. – No. 85. – P. 526–530.

22. Алимурадов А.К. Алгоритм измерения частоты основного тона речевых сигналов на основе комплементарной множественной декомпозиции на эмпирические моды // Измерительная техника. – 2016. – № 12. – С. 53–57.

23. Huang, X., Acero A., Hon H.-W. Spoken Language Processing. Guide to Algorithms and System Developmen // Prentice Hall. – New Jersey, 2001. – P. 980.

Reference

1. Proektnaia kompaniia i issledovatel'skaia laboratoriia v oblasti affektivnykh nauk i kognitivnykh tekhnologii: ofitsial'nyi sait [Design company and research laboratory in the field of affective sciences and cognitive technologies: official site], available at: https://neurodatalab.com (accessed 18 May 2020).

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

3. Alimuradov A.K., Fokina E.A., Zhurina A.E. Kratkii obzor sushchestvuiushchikh baz dannykh emotsional'noi rechi: sovremennoe sostoianie, problemy i perspektivy razvitiia [A brief overview of the existing databases of emotional speech: current state, problems and development prospects]. *Metody, sredstva i tekhnologii polucheniia i obrabotki izmeritel'noi informatsii ("Shliandinskie chteniia - 2020"). Materialy XII Mezhdunarodnoi nauchno-tekhnicheskoi konferentsii s elementami nauchnoi shkoly i konkursom nauchno-issledovatel'skikh rabot dlia studentov, aspirantov i molodykh uchenykh* (Penza, 16-18 March 2020). Ed. Doctor of Technical Sciences E.A. Pecherskaia. Penza: Penzenskii gosudarstvennyi tekhnologicheskii universitet, 2020, pp. 292-294.

4. Osipov A.N., Mezhennaia M.M., Kul' T.P., Bobrovskaia Ia.Iu., Rushkevich Iu.N., Likhachev S.A. Mobil'noe prilozhenie dlia sbora diagnosticheskoi informatsii posredstvom registratsii rechevykh signalov [Mobile app to collect diagnostic information through the registration of speech signals]. *Bigdata and advanced analytics*, 2018, no. 4, pp. 346-350.

5. Kul' T.P., Rushkevich Iu.N., Likhachev S.A. Adaptatsiia metodov tsifrovoi obrabotki signalov k zadache analiza rechi pri nevrologicheskikh patologiiakh [Adaptation of digital signal processing methods to the analysis of speech in neurological pathologies]. *Doklady*

Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki, 2018, no. 7(117), pp. 128-132.

6. Davydova A.Iu. Otsenka emotsional'nogo sostoianiia cheloveka po parametram rechevogo signala [Assessment of the emotional state of a person by the parameters of the speech signal]. *Biotekhnicheskie, meditsinskie i ekologicheskie sistemy, izmeritel'nye ustroistva i robototekhnicheskie kompleksy - Biomedsistemy-2019. Sbornik trudov XXXII Vserossiiskaia nauchno-tekhnicheskaia konferentsiia studentov, molodykh uchenykh i spetsialistov* (Ryazan, 4-6 December 2019). Ed. V.I. Zhulev. Ryazan: IP Koniakhin A.V. (BookJet), 2019, pp. 348-351.

7. Kul' T.P., Mezhennaia M.M., Rushkevich Iu.N., Osipov A.N., Likhachev S.A., Rushkevich I.V. Metodicheskoe i apparatno-programmnoe obespechenie dlia registratsii i obrabotki rechevykh signalov s tsel'iu diagnostiki nevrologicheskikh zabolevanii [Methodical and hardware-software for recording and processing speech signals for diagnosis of neurological diseases]. *Informatika*, 2019, vol. 16, no. 2, pp. 27-39.

8. Prokhorenko E.I., Sokolova V.S., Kolesnikov V.A. Vydelenie priznakov emotsional'no okrashennoi rechi v zvukozapisi [Isolation of signs of emotionally colored speech in sound recording]. *Studenchestvo Rossii: vek XXI. Materialy VI Vserossiiskoi molodezhnoi nauchno-prakticheskoi konferentsii* (Orel, 13 December 2018). Orel: Orlovskii gosudarstvennyi agrarnyi universitet imeni N.V. Parakhina, 2019, pp. 184-187.

9. Gai V.E., Utrobin V.A., Poliakov I.V. Sistema otsenki psikhoemotsional'nogo sostoianiia diktora po golosu [The evaluation system emotional the state of the speaker's voice]. *Slozhnost'. Razum. Postneklassika*, 2016, no. 2, pp. 75-80.

10. Astakhov D.A., Kataev A.V. Ispol'zovanie sovremennykh algoritmov mashinnogo obucheniia dlia zadachi raspoznavaniia emotsii [Use of modern machine training algorithms for the task of recognition of emotions]. *Cloud of Science (Moskva)*, 2018, vol. 5, no. 4, pp. 664-679.

11. Alimuradov A.K., Tychkov A.Iu., Churakov P.P., Sultanov B.V. Sposob opredeleniia formantnoi razborchivosti rechi dlia otsenki psikhoemotsional'nogo sostoianiia operatorov sistem upravleniia s vysokoi stepen'iu otvetstvennosti [Method to determine formant speech intelligibility for estimating psycho-emotional state of control system operators with a high degree of responsibility]. *Izmerenie. Monitoring. Upravlenie. Kontrol'*, 2019, no. 4(30), pp. 58-69.

12. Alimuradov A.K., Tychkov A.Iu., Churakov P.P., Artamonov D.V. Pomekhoustoichivyi algoritm opredeleniia prosodicheskikh kharakteristik rechevykh signalov dlia sistem otsenki psikhoemotsional'nogo sostoianiia cheloveka [A noise-robust algorithm to determine prosodic characteristics of speech signals for systems of human psycho-emotional state assessment]. *Izvestiia vysshikh uchebnykh zavedenii. Povolzhskii region. Tekhnicheskie nauki*, 2019, no. 3(51), pp. 3-16.

13. Alimuradov A.K., Tychkov A.Iu., Churakov P.P. Sposob avtomatizirovannoi segmentatsii rechevykh signalov dlia opredeleniia vremennykh patternov estestvenno vyrazhennykh psikhoemotsional'nykh sostoianii [A method for automated segmentation of speech signals to determine temporal patterns of naturally expressed psycho-emotional states]. *Izmerenie. Monitoring. Upravlenie. Kontrol'*, 2019, no. 3(29), pp. 48-60.

14. Davydov A.G., Kiselev V.V., Kochetkov D.S., Tkachenia A.V. Vybor optimal'nogo nabora informativnykh priznakov dlia klassifikatsii emotsional'nogo sostoianiia diktora po golosu [Selection of the optimal set of informative features for classifying the emotional state of the speaker by voice]. *Komp'iuternaia lingvistika i intellektual'nye tekhnologii. Materialy ezhegodnoi mezhdunarodnoi konferentsii "Dialog'* (Bekasovo, 30 May - 3 June 2012). Moscow: Rossiiskii gosudarstvennyi gumanitarnyi universitet, 2012, vol. 1, no. 11, pp. 122-128.

15. 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, A 454, pp. 903-995.

16. Zhaohua W., Huang N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 2009, no. 1(1), pp. 1-41.

17. 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, no. 2(2), pp. 135-156.

18. Torres M.E., Colominas M.A., Schlotthauer G., Flandrin P. A complete Ensemble Empirical Mode decomposition with adaptive noise. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-11)* (May 22-27, 2011). Prague, Czech Republic, 2011, pp. 4144-4147.

19. 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, pp. 19-29.

20. Sharma R., Prasanna S.R.M. Characterizing glottal activity from speech using empirical mode decomposition. *2015 21st National Conference on Communications (NCC)* (27 February-1 March, 2015). Mumbai, India, pp. 1-6.

21. Lee H.S. Improvement of decomposing results of empirical mode decomposition and its variations for sea-level records analysis. *Journal of Coastal Research: proceedings of the 15th International Coastal Symposium*, 2018, no. 85, pp. 526-530.

22. Alimuradov A.K. Algoritm izmereniia chastoty osnovnogo tona rechevykh signalov na osnove komplementarnoi mnozhestvennoi dekompozitsii na empiricheskie mody [An algorithm for measurement of the pitch frequency of speech signals based on complementary ensemble decomposition into empirical modes]. *Izmeritel'naia tekhnika*, 2016, no. 12, pp. 53-57.

23. Huang, X., Acero A., Hon H.-W. Spoken Language Processing. Guide to Algorithms and System Developmen. *Prentice Hall*. New Jersey, 2001, 980 p.

Сведения об авторах

Алимурадов Алан Казанферович (Пенза, Россия) – кандидат технических наук, доцент кафедры «Радиотехника и радиоэлектронные системы» Пензенского государственного университета (440026, Пенза, ул. Красная, 40, e-mail: alansapfir@yandex.ru).

Тычков Александр Юрьевич (Пенза, Россия) – кандидат технических наук, доцент кафедры «Радиотехника и радиоэлектронные системы» Пензенского государственного университета (440026, Пенза, ул. Красная, 40, e-mail: tychkov-a@mail.ru).

About the authors

Alimuradov Alan Kazanferovich (Penza, Russian Federation) is a Ph. D. in Technical Sciences, Associate Professor of the Department of Radio Engineering and Radioelectronic Systems Penza State University (440026, Penza, 40, Krasnaya str., e-mail: alansapfir@yandex.ru).

Tychkov Alexander Yurievich (Penza, Russian Federation) is a Ph. D. in Technical Sciences, Associate Professor of the Department of Radio Engineering and Radioelectronic Systems Penza State University (440026, Penza, 40, Krasnaya str., e-mail: tychkov-a@mail.ru).

Получено 17.08.2020