

ПРИБОРЫ, СИСТЕМЫ И ИЗДЕЛИЯ  
МЕДИЦИНСКОГО НАЗНАЧЕНИЯ

УДК 004.934

DOI 10.21685/2307-5538-2019-3-6

*А. К. Алимуратов, А. Ю. Тычков, П. П. Чураков*СПОСОБ АВТОМАТИЗИРОВАННОЙ СЕГМЕНТАЦИИ  
РЕЧЕВЫХ СИГНАЛОВ ДЛЯ ОПРЕДЕЛЕНИЯ ВРЕМЕННЫХ  
ПАТТЕРНОВ ЕСТЕСТВЕННО ВЫРАЖЕННЫХ  
ПСИХОЭМОЦИОНАЛЬНЫХ СОСТОЯНИЙ*A. K. Alimuradov, A. Yu. Tychkov, P. P. Churakov*A METHOD FOR AUTOMATED SEGMENTATION OF SPEECH  
SIGNALS TO DETERMINE TEMPORAL PATTERNS OF  
NATURALLY EXPRESSED PSYCHO-EMOTIONAL STATES

**А н н о т а ц и я.** *Актуальность и цели.* Оценка психоэмоционального состояния человека в отраслях профессиональной деятельности, сопряженных с повышенным риском возникновения техногенных и биогенных аварий, является важной социально значимой проблемой для государства. Целью исследования является разработка способа автоматизированной сегментации речевых сигналов для повышения эффективности определения временных паттернов речи, релевантных естественно выраженным психоэмоциональным состояниям человека. *Материалы и методы.* В разработке способа использовались: уникальная технология адаптивной декомпозиции нестационарных сигналов – улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом, а также правило разграничения на основе физиологического аспекта формирования речи. Программная реализация способа была выполнена в среде математического моделирования © Matlab (MathWorks). *Результаты.* Разработан способ автоматизированной сегментации речевых сигналов на вокализованные, невокализованные участки и участки пауз для определения временных паттернов речи, отражающих естественно выраженные психоэмоциональные состояния человека. Проведено исследование способа с использованием базы речевых сигналов, зарегистрированных с группы людей, переживаемых естественные положительные и отрицательные эмоции. *Выводы.* Результаты исследования выявили, что в условиях нестабильности моторики речевого аппарата разработанный способ сегментации позволяет точнее определять границы вокализованных, невокализованных участков и участков пауз, тем самым повышая эффективность вычисления временных паттернов речи и определения психоэмоциональных состояний человека.

**A b s t r a c t.** *Background.* An assessment of human psycho-emotional state in the fields of professional activity associated with an increased risk of man-made and biogenic accidents is an important socially significant problem for the state. The aim of the study is to develop a meth-

od for automated segmentation of speech signals to improve the efficiency of determining temporal patterns of speech relevant to naturally expressed psycho-emotional states of a person. **Materials and methods.** To develop the method, a unique technology for adaptive decomposition of non-stationary signals, namely, the improved ensemble empirical mode decomposition with adaptive noise, as well as the rule of differentiation based on the physiological aspect of speech formation, have been used. Software implementation of the method was performed in ©Matlab (MathWorks) mathematical modeling environment. **Results.** A method for automated segmentation of speech signals into voiced, unvoiced, and pause sections to determine temporal patterns of speech reflecting naturally expressed human psycho-emotional states, has been developed. A study was conducted using a base of speech signals recorded from a group of subjects experiencing natural positive and negative emotions. **Conclusions.** The results of the study have revealed that in conditions of instability of the speech apparatus motility, the developed method for segmentation makes it possible to more accurately determine the boundaries of voiced, unvoiced, and pause sections, thereby increasing the efficiency of calculating temporal patterns of speech, and determining psycho-emotional states of a person.

**К л ю ч е в ы е с л о в а:** обработка речевых сигналов, автоматизация процесса обработки, сегментация, адаптивная декомпозиция, временные паттерны речи, естественно выраженные психоэмоциональные состояния.

**К e y w o r d s:** speech signal processing, automation of processing, segmentation, adaptive decomposition, temporal speech patterns, naturally expressed psycho-emotional states.

### *Введение*

Воспроизведение речи является одним из самых сложных приобретаемых навыков человека в течение жизни. Речевой аппарат чрезвычайно чувствителен к нарушениям работы нервной системы [1]. На протяжении долгих лет оценка нестабильности моторики речевого аппарата при естественно выраженных психоэмоциональных состояниях ограничивалась лабораторными анализами и тестами непосредственного психического восприятия. На сегодняшний день эта задача успешно решается методами на основе анализа речевых сигналов [2].

Психоэмоциональное состояние человека шифруется в определенные паттерны речи, отражающие вид и степень выраженности состояния. Основные понятия, которые характеризуют паттерны речи, связанные с формой, размерами, динамикой изменения речевого аппарата и описывающие психоэмоциональные состояния человека, можно разделить на три группы объективных признаков: спектрально-временные, кепстральные и амплитудно-частотные [3].

Спектрально-временные паттерны характеризуют речевой сигнал в его физико-математической сущности исходя из наличия компонентов трех видов: вокализованных и невокализованных участков и участков пауз. Спектрально-временные паттерны достаточно хорошо отображают особенности формы временного ряда и спектра голосовых импульсов, а также уникальность фильтрующих функций речевого аппарата при психоэмоциональном возбуждении человека.

Временные паттерны характеризуют особенности потока речи, связанные с динамикой перестройки артикуляционных органов говорящего и являются интегральными характеристиками, отражающими взаимосвязь движения артикуляционных органов говорящего [4]. В настоящее время известно большое количество разнообразных методик определения временных паттернов речи [5–7]. Каждая из методик имеет ряд преимуществ и недостатков. Точность определения временных паттернов в представленных подходах зависит от эффективности сегментации речи на информативные участки, характеризующиеся определенной длительностью.

В данной работе представлен способ автоматизированной сегментации речи на вокализованные, невокализованные участки и участки пауз для определения временных паттернов при естественно выраженных психоэмоциональных состояниях. Суть способа заключается в применении автоматизированной адаптивной декомпозиции для сегментации речевых сигналов.

Статья является продолжением опубликованных трудов авторов [8, 9] и построена следующим образом. В первом разделе статьи кратко представлена информация об адаптивной декомпозиции и о временных паттернах речи. Во втором и третьем разделах кратко описывается и исследуется разработанный способ. В четвертом разделе представлены результаты исследований и сделаны выводы. Авторский коллектив благодарит Российский фонд фундаментальных исследований за финансовую поддержку проекта № 18-37-00256 мол\_а «Исследование информативных параметров речевых сигналов и разработка алгоритмов обнаружения скрытых признаков психоэмоционального состояния операторов критичных по безопасности систем управления».

### *Материалы и методы*

#### *Адаптивная декомпозиция*

Важным условием автоматизированной адаптивной декомпозиции является формирование базисной функции разложения, функционально зависимой от внутренней структуры исходного речевого сигнала. Данное условие может выполняться при использовании декомпозиции на эмпирические моды (ДЭМ). Классический метод ДЭМ был разработан Norden Huang в 1998 г. [10] и предназначался для разложения нестационарных сигналов, возникающих в нелинейных системах. ДЭМ обеспечивает разложение нестационарного сигнала на высокочастотные и низкочастотные составляющие, называемые эмпирическими модами (ЭМ).

При разложении модель сигнала не задается заранее, ЭМ вычисляются в ходе процедуры отсеивания с учетом локальных особенностей (таких как экстремумы и нули сигнала) и внутренней структуры каждого конкретного сигнала. Таким образом, ЭМ не имеют строгого аналитического описания, но должны удовлетворять двум условиям, гарантирующим автоматизацию процесса, а также определенную симметричность и узкополосность базисных функций:

- общее количество экстремумов функции сигнала должно быть равным количеству нулей функции с точностью до единицы;
- среднее значение верхней и нижней огибающих, интерполирующих локальные максимумы и минимумы функции сигнала, должно стремиться к нулю.

Аналитически ДЭМ представляет собой следующее выражение:

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

где  $x(n)$  – исходный сигнал;  $IMF_i(n)$  – ЭМ;  $r_l(n)$  – конечный остаток;  $i = 1, 2, \dots, I$  – номер ЭМ;  $n$  – дискретный отсчет времени.

Кроме адаптивности, технология ДЭМ обладает и другими важными свойствами:

- локальностью – возможностью учета локальных особенностей сигнала;
- ортогональностью – обеспечением восстановления сигнала с определенной точностью;
- полнотой – гарантией конечного числа базисных функций при конечной длительности сигнала.

Уникальность декомпозиции заключается в том, что функции базиса разложения извлекаются непосредственно из внутренней структуры исходного сигнала. Это позволяет учитывать такие особенности исходного сигнала, как частотную и амплитудную модуляции, концентрацию энергии в определенном диапазоне частот и др.

С момента появления классической ДЭМ разработано большое количество разновидностей декомпозиций. Наиболее адаптивной к нестационарным речевым сигналам сложной формы является улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом (ПМДЭМАШ). Появление улучшенной ПМДЭМАШ позволило решить ряд проблем, присущих другим видам декомпозиции [11]:

- смешивание ЭМ вследствие перекрытия масштабно-энергетических пространств мод;
- зашумление ЭМ;
- наличие паразитных ЭМ, возникающих на ранних этапах декомпозиции.

Отличительной особенностью улучшенной ПМДЭМАШ является добавление к исходному сигналу контролируемого адаптивного шума для создания новых локальных экстремумов [11]:

$$x_j(n) = x(n) + w(n),$$

где  $x_j(n)$  – шумовые копии речевого сигнала;  $w_j(n)$  – реализации белого шума,

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

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

$$r_i(n) = \sum_{j=1}^J \frac{r_{ji}(n)}{J},$$

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

На рис. 1 представлен пример разложения фрагментов вокализованной и невокализованной речи длительностью 100 мс. Вокализованный и невокализованный фрагменты были разложены на восемь ЭМ. Первые ЭМ для обоих видов фрагментов являются высокочастотными, а последующие моды – низкочастотными (по убыванию). Для вокализованного фрагмента первые пять ЭМ являются информативными (концентрирующие основную энергию сигнала), последние четыре моды – компенсирующими и трендовыми [12]. Для невокализованного фрагмента однозначно определить информативные, компенсирующие и трендовые моды невозможно. Это связано с тем, что невокализованный речевой сигнал по характеристикам подобен шуму.

Подробно особенности видов декомпозиции, их преимущества и недостатки при обработке речи описаны в работах [13–15].

#### *Временные паттерны*

Обзор информативных параметров речевых сигналов [1, 3, 16] выявил следующие временные паттерны релевантные к естественно выраженным психоэмоциональным состояниям:

- скорость распределения временных интервалов вокализованных, невокализованных участков и участков пауз (Rate of Speech Timing, RST);
- ускорение распределения временных интервалов вокализованных, невокализованных участков и участков пауз (Acceleration of Speech Timing, AST);
- энтропия распределения временных интервалов вокализованных, невокализованных участков и участков пауз (Entropy of Speech Timing, EST);
- продолжительность интервалов пауз (Duration of Pause Intervals, DPI).

#### *Описание способа*

На рис. 2 структурно представлена блок-схема способа автоматизированного сегментирования речи для определения временных паттернов при естественно выраженных психоэмоциональных состояниях. Способ включает в себя два этапа обработки. Первый этап (блоки 1–4) заключается в автоматизированной сегментации речи на информативные участки с помощью адаптивной декомпозиции и энергетического анализа ЭМ. Второй этап (блок 5) заключается в определении временных паттернов вокализованных, невокализованных участков и участков пауз, отражающих нарушения моторики речевого аппарата, вызванных психоэмоциональными расстройствами. Блоки 6 и 7 не относятся к разработанному способу и используются на этапе исследования. Рассмотрим некоторые этапы способа подробнее.

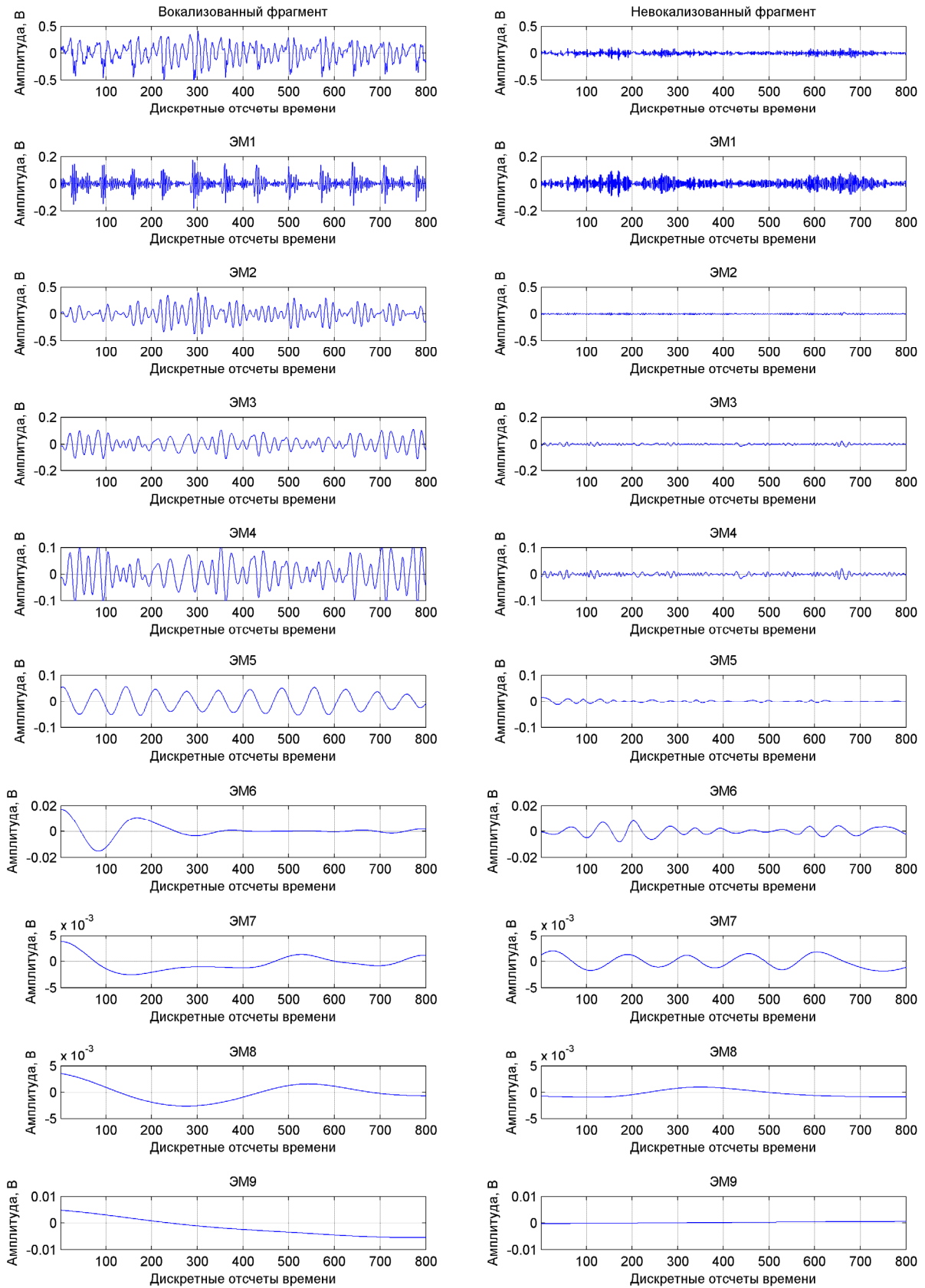


Рис. 1. Результат разложения фрагментов вокализованного (левый столбец) и невокализованного (правый столбец) фрагментов речи методом улучшенной ПМДЭМАШ

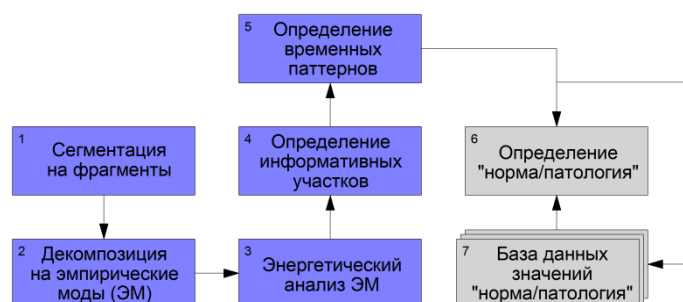


Рис. 2. Упрощенная блок-схема способа автоматизированного сегментирования речи для определения временных паттернов

Речь представляет собой процесс, спектр которого остается относительно неизменным в течение короткого интервала времени. Это позволяет разделить речевой сигнал на равные кратковременные фрагменты по 10 мс, в пределах которых можно считать сигнал условно стационарным. После сегментации сигнал представляет собой набор фрагментов, а дальнейшая работа способа осуществляется с каждым фрагментом в отдельности.

Декомпозиция сигнала осуществляется методом улучшенной ПМДЭМАШ. Параметры декомпозиции задавались в соответствии с автоматизированной методикой определения оптимальных значений настройки [12].

Изменение уровня речевого сигнала во времени характеризуется важным информативным параметром – амплитудным распределением. Распределение амплитуды сигнала во времени достаточно полно описывается с помощью кратковременной энергии. В соответствии с функционалом слухового аппарата человек воспринимает речь нелинейно, определяя разницу между энергиями различных информативных участков речи.

В разработанном способе сегментация осуществляется на основе энергетического анализа ЭМ фрагментов речевого сигнала в скользящем окне длительностью 10 мс. На рис. 3 представлен пример разложения кратковременных фрагментов вокализованной, невокализованной речи и паузы на ЭМ, полученных методом улучшенной ПМДЭМАШ. Особенностью способа является то, что соотнесение анализируемого фрагмента сигнала к вокализованной, невокализованной речи или к паузе осуществляется, исследуя энергетические свойства каждой ЭМ фрагмента в отдельности. Учитывая, что каждая ЭМ обладает определенными параметрами, сравнительный энергетический анализ мод по отдельности значительно повышает эффективность определения границ вокализованных, невокализованных участков и участков пауз при нестабильной моторике речевого аппарата.

К исследуемым энергетическим параметрам ЭМ фрагментов речевого сигнала относятся:

- логарифм энергии:

$$LE_{s,i} = \log_2 \left( \sum_{n=1}^N (IMF_{s,i}(n))^2 \right),$$

где  $LE_{s,i}$  – логарифм энергии ЭМ фрагмента речевого сигнала;  $s$  – номер фрагмента;

- отношение кратковременной энергии ЭМ к скорости пересечения сигнала ЭМ через нулевое значение (СПСН):

$$Z_{s,i} = \frac{\sum_{n=1}^N (IMF_{s,i}(n))^2}{ZCR_{s,i}},$$

$$ZCR_{s,i} = 0,5 \sum_{n=1}^{N-1} \left| \operatorname{sgn}(IMF_{s,i}((n-1)N + n + 1)) - \operatorname{sgn}(IMF_{s,i}((n-1)N + n)) \right|,$$

где  $ZCR_{s,i}$  – скорость пересечения сигнала ЭМ через нулевое значение;  $\operatorname{sgn}$  – знаковая функция ( $\operatorname{sgn}(x) = 1$ , если  $x \geq 0$  и  $-1$  при  $x < 0$ ).

Определение информативных участков заключается в обнаружении точных границ вокализованных, невокализованных участков и участков пауз. Сегментация всех участков одно-

временно неэффективна. Для повышения точности сегментации необходимо последовательное разделение речевого сигнала на паузы, затем на вокализованные и невокализованные участки.

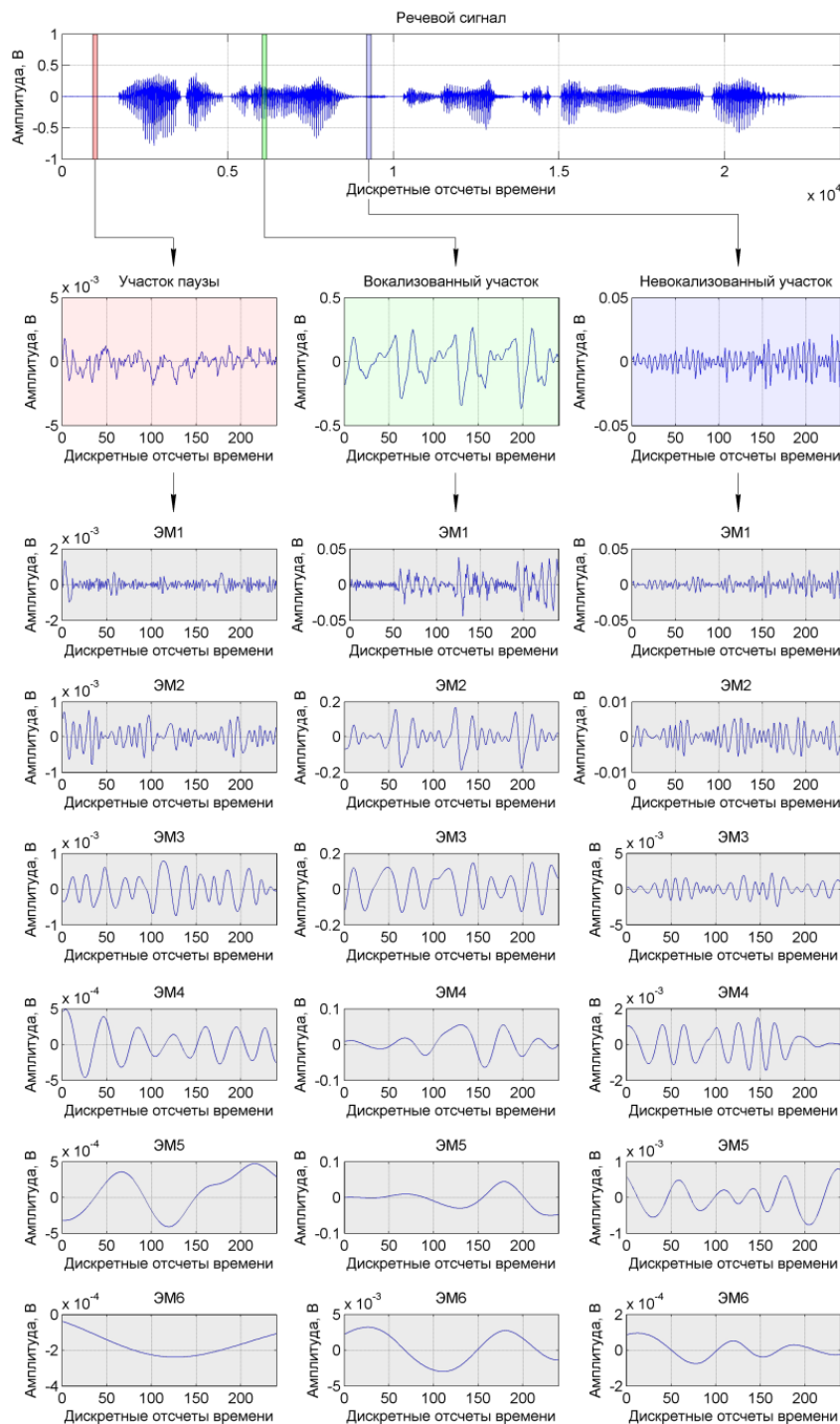


Рис. 3. Разложение кратковременных фрагментов вокализованной, невокализованной речи и паузы на ЭМ

В соответствии с физиологическим аспектом формирования речи человек перед произношением делает начальную кратковременную паузу, не содержащую речь и соответствующую тишине. Обычно длительность начальной паузы составляет 200–500 мс. Применяя усредненные значения параметров  $LE$  и  $Z$  ЭМ для фрагментов начальной паузы, можно определить пороговые значения  $LE_{resh}$  и  $Z_{resh}$ . На рис. 4,а представлена графическая интерпретация

формирования пороговых значений по шести первым модам анализируемых фрагментов. Пунктирными линиями красного цвета отмечены значения энергетических параметров ЭМ фрагментов начальной паузы, утолщенной сплошной линией красного цвета отмечены усредненные пороговые значения.

На рис. 4,б представлена графическая интерпретация пороговой обработки. Утолщенной сплошной линией зеленого цвета отмечены значения энергетических параметров ЭМ вокализованного фрагмента речи. Утолщенной линией синего цвета – для невокализованного фрагмента.

На рис. 5 представлена интерпретация процесса автоматизированной сегментации на информативные участки в разработанном способе.

Определение временных паттернов речевых сигналов заключается в вычислении значений  $RST$ ,  $AST$ ,  $DPI$  и  $EST$ .

$RST$  обеспечивает более точную оценку ухудшения скорости речи, чем простое измерение длительности пауз, поскольку данный паттерн учитывает не только паузы, но и вокализованные, невокализованные участки. Вокализованные участки предоставляют дополнительную информацию об ухудшении фонации, тогда как невокализованные участки предоставляют информацию о нечеткой артикуляции органов говорящего. Значение  $RST$  в комплексе приблизительно равно скорости речи, поскольку ухудшение скорости речи связано с недостатками во всех элементах речевых сигналов. Каждый вокализованный, невокализованный участок и участки пауз описываются временем возникновения, определяемым как среднее значение времени между началом участка и его окончанием.

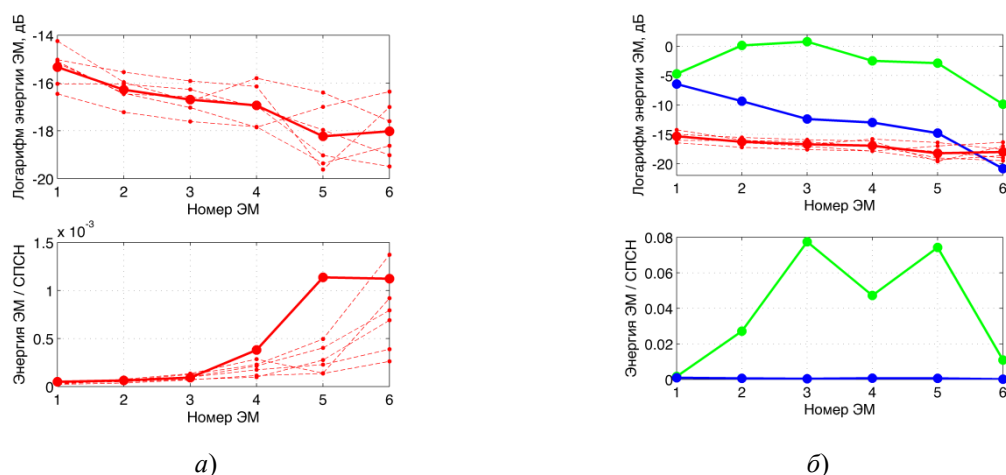


Рис. 4. Анализ энергетических параметров ЭМ фрагментов речевого сигнала:  
а – формирование пороговых значений; б – пороговая обработка

$AST$  определяет степень ускорения времени. Каждый анализируемый фрагмент речевого сигнала делится на две части с перекрытием 25 %, что обеспечивает плавный переход между частями. Значение  $AST$  рассчитывается как разница между значениями  $RST$  обеих частей, разделенное на общую продолжительность фрагмента речевого сигнала.

$EST$  описывает упорядоченность (предсказуемость) речевых сигналов, в том числе вокализованные, невокализованные участки и участки пауз. Соответственно, уменьшение энтропии равнозначно нарушению моторики речи. Для определения  $EST$  рассчитывается количество всех интервалов вокализованных и невокализованных участков, участков паузы и общее количество участков. Значение  $EST$  определяется следующим образом:

$$EST = -\frac{kv}{kt} \log_2 \left( \frac{kv}{kt} \right) - \frac{ku}{kt} \log_2 \left( \frac{ku}{kt} \right) - \frac{kp}{kt} \log_2 \left( \frac{kp}{kt} \right),$$

$kv$  – вокализованные фрагменты;  $ku$  – невокализованные фрагменты;  $kp$  – фрагменты пауз;  $kt$  – общее количество фрагментов.

$DPI$  определяет способность говорящего начать воспроизведение речи. Нестабильность моторики речевого аппарата при естественно выраженных психоэмоциональных состояниях



может вызвать трудности при воспроизведении, которые порождают увеличение длительности пауз. Значение  $DPI$  вычисляется как средняя длительность всех участков пауз.

### Исследование способа

Для тестирования разработанного способа автоматизированного сегментирования речи сформирована группа испытуемых и зарегистрирована база речевых сигналов, состоящая из 1000 записей.

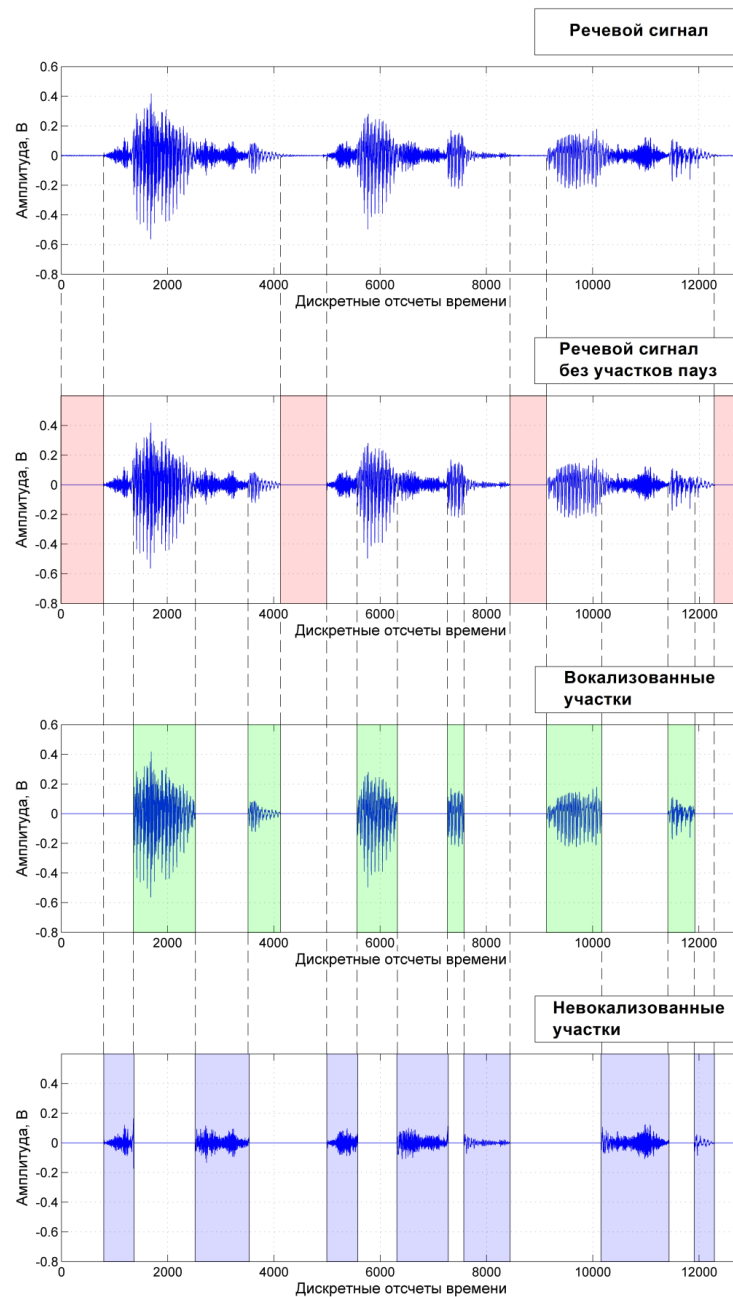


Рис. 5. Процесс автоматизированной сегментации речи на информативные участки

Формирование группы испытуемых осуществлялось при поддержке областной психиатрической больницы им. К. Р. Евграфова (г. Пенза) и Медицинского института Пензенского государственного университета. В группу испытуемых было отобрано 100 человек в возрасте от 18 до 79 лет мужского и женского пола, переживающих естественные положительные и отрицательные эмоции. Эффективность сегментации оценивалась в сравнении с популярными на практике способами сегментации с последующим измерением временных паттернов:

- основанные на анализе кратковременной энергии (Short-time Energy, STE) и скорости пересечения сигнала через нулевое значение  $ZCR$  [17, 18];
- основанные на использовании статистических свойств фонового шума и одномерного расстояния Махаланобиса [19].

Для определения эффективности обнаружения психоэмоциональных положительных и отрицательных состояний использовался параметр – ошибки первого и второго рода [8]. Ошибкой первого рода является ложное присваивание статуса «норма» речевому сигналу, произнесенному человеком с психоэмоциональным расстройством, а ошибкой второго рода – ложное присваивание статуса «патология» речевому сигналу, произнесенному здоровым человеком. Программная реализация способа выполнена в пакете прикладных программ для решения задач технических вычислений © Matlab (MathWorks).

В табл. 1 представлены результаты определения психоэмоциональных положительных и отрицательных состояний.

Таблица 1

## Результаты определения психоэмоциональных состояний

| Прогнозируемый результат                                    | Результат определения, чел. |       | Ошибка, % |    |
|---|-----------------------------|-------|-----------|----|
|   | Патология                   | Норма |           |    |
| <b>Способ на основе STE и ZCR</b>                           |                             |       |           |    |
| <i>Положительное психоэмоциональное состояние</i>           |                             |       |           |    |
| Патология   | 63                          | 37    | 1-го      | 37 |
| Норма   | 12                          | 88    | 2-го      | 12 |
| <i>Отрицательное психоэмоциональное состояние</i>           |                             |       |           |    |
| Патология   | 81                          | 19    | 1-го      | 19 |
| Норма   | 7                           | 93    | 2-го      | 7  |
| <b>Способ на основе одномерного расстояния Махаланобиса</b> |                             |       |           |    |
| <i>Положительное психоэмоциональное состояние</i>           |                             |       |           |    |
| Патология   | 71                          | 29    | 1-го      | 29 |
| Норма   | 8                           | 92    | 2-го      | 8  |
| <i>Отрицательное психоэмоциональное состояние</i>           |                             |       |           |    |
| Патология   | 84                          | 16    | 1-го      | 16 |
| Норма   | 6                           | 94    | 2-го      | 6  |
| <b>Разработанный способ</b>                                 |                             |       |           |    |
| <i>Положительное психоэмоциональное состояние</i>           |                             |       |           |    |
| Патология   | 88                          | 12    | 1-го      | 12 |
| Норма   | 5                           | 95    | 2-го      | 5  |
| <i>Отрицательное психоэмоциональное состояние</i>           |                             |       |           |    |
| Патология   | 89                          | 11    | 1-го      | 11 |
| Норма   | 4                           | 96    | 2-го      | 4  |

**Результаты и выводы**

Из таблицы видно, что процент ложных присваиваний статуса «норма» речевым сигналам, произнесенным пациентами, находящимися в состоянии психоэмоционального возбуждения, у способов, основанных на  $STE$ ,  $ZCR$ , и одномерного расстояния Махаланобиса больше, чем у разработанного способа: на 25 и 17 % для отрицательного психоэмоционального состояния, 7 и 4 % для положительного психоэмоционального состояния соответственно. То же самое можно сказать о ложных присваиваниях статуса «патология» речевым сигналам, произнесенным пациентами в нейтральном состоянии: на 7 и 3 %, 3 и 2 % соответственно. Наименьшие значения ошибок 1-го и 2-го родов были достигнуты разработанным способом: всего лишь 12 и 5 % для положительных эмоций, 11 и 4 % для отрицательных эмоций.

Таким образом, можно сделать вывод, что в условиях нестабильности моторики речевого аппарата разработанный способ сегментации позволяет точнее определять границы вокализованных, невокализованных участков и участков пауз, тем самым повышая эффективность вычисления временных паттернов речи и определения психоэмоциональных состояний человека. Данные результаты достигаются исключительно за счет преимуществ энергетического анализа параметров ЭМ, полученных методом улучшенной ПМДЭМАШ для каждого анализируемого фрагмента.

## Библиографический список

1. *Schuller, B. W.* Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing / B. W. Schuller, A. M. Batliner. – New York : Wiley, 2013. – 344 p.
2. *Trigeorgis, G.* Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network / G. Trigeorgis, F. Ringeval, R. Brueckner, E. Marchi, M.A. Nicolaou, B. Schuller, S. Zafeiriou // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Shanghai, China, 20–25 March 2016). – Shanghai, China : IEEE, 2016. – P. 5200–5204.
3. *Huang, X.* Spoken Language Processing. Guide to Algorithms and System Development / X. Huang, A. Acero, H.-W. Hon // Prentice Hall. – New Jersey, 2001. – 980 p.
4. *Фант, Г. К.* Акустическая теория речеобразования / Г. К. Фант ; пер. с англ. Л. А. Варшавского и В. И. Медведева ; науч. ред. В. С. Григорьева. – Москва : Наука, 1964. – 284 с.
5. *Whitehead, R. L.* Temporal characteristics of speech produced by inexperienced signers during simultaneous communication / R. L. Whitehead, N. Schiavetti, D. E. Metz, T. Farrell // Journal of Communication Disorders. – 1999. – Vol. 32, № 2. – P. 79–95.
6. *Bóna, J.* Temporal characteristics of speech: The effect of age and speech style / J. Bóna // The Journal of the Acoustical Society of America. – 2014. – Vol. 136, № 2. – P. 116–121.
7. *Bakaev, A. V.* Spectral and temporal characteristics of vocal speech in the emotional aspect / A. V. Bakaev // Almanac of Modern Science and Education. – 2014. – Vol. 4, № 83. – P. 28–32.
8. *Алимурадов, А. К.* Способ сегментации речевого сигнала для систем оценки психогенных состояний / А. К. Алимурадов, А. Ю. Тычков, П. П. Чураков, С. И. Торгашин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 3 (47). – С. 81–93.
9. *Алимурадов, А. К.* Повышение точности измерения частоты основного тона на основе оптимизации процесса декомпозиции речевых сигналов на эмпирические моды / А. К. Алимурадов, Ю. С. Квитка, П. П. Чураков, А. Ю. Тычков // Измерение. Мониторинг. Управление. Контроль. – 2018. – № 4 (26). – С. 53–65.
10. *Huang, N. E.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis / N. E. Huang, Sh. Zheng, R. L. Steven // Proceedings of the Royal Society of London. – 1998. – Vol. A454. – P. 903–995.
11. *Colominasa, M. A.* Improved complete ensemble EMD: A suitable tool for biomedical signal processing / M. A. Colominasa, G. Schlotthauer, M. E. Torres // Biomedical Signal Processing and Control. – 2014. – Vol. 14. – P. 19–29.
12. *Zhaohua, Wu.* Ensemble empirical mode decomposition: A noise – assisted data analysis method / Wu Zhaohua, N. E. Huang // Advances in Adaptive Data Analysis. – 2009. – Vol. 1, № 1. – P. 1–41.
13. *Yeh, J.-R.* Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method / J.-R. Yeh, J.-S. Shieh, N. E. Huang // Advances in Adaptive Data Analysis. – 2010. – Vol. 2, № 2. – P. 135–156.
14. *Torres, M. E.* A complete Ensemble Empirical Mode decomposition with adaptive noise / M. E. Torres, M. A. Colominas, G. Schlotthauer, P. Flandrin // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Prague, Czech Republic, 22–27 May 2011). – Prague, Czech Republic : IEEE, 2011. – P. 4144–4147.
15. Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder / J. Hlavnička, R. Čmejla, T. Tykalová, K. Šonka, E. Růžička, J. Rusz // Scientific Reports. – 2017. – Vol. 7, № 12. – 13 p.
16. *Alimuradov, A. K.* Automation of empirical mode decomposition to increase efficiency of speech signal processing / A. K. Alimuradov, A. Yu. Tychkov, Yu. S. Kvitka // International Russian Automation Conference (RusAutoCon) (Sochi, Russia, 9–16 Sept. 2018). – Sochi, Russia : IEEE, 2018. – 6 p.
17. *Bachu, R. G.* Separation of Voiced and Unvoiced Using Zero Crossing Rate and Energy of the Speech Signal / R. G. Bachu, S. Kopparthi, B. Adapa, B. D. Barkana // American Society for Engineering Education (ASEE) Zone Conference Proceedings (Pittsburgh, USA, 22–25 June 2008). – Pittsburgh, USA, 2008. – P. 1–7.
18. *Moattar, M. H.* A simple but efficient real-time voice activity detection algorithm / M. H. Moattar, M. M. Homayounpour // 17th European Signal Processing Conference (EUSIPCO 2009) (Glasgow, Scotland, August 24–28 2009). – Glasgow, Scotland : Eusipco, 2009. – P. 2549–2553.
19. *Saha, G.* A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications / G. Saha, Ch. Sandipan, S. Suman // The National Conference on Communications (28–30 January 2005). – New Delhi : Allied Publishers, 2005. – 5 p.

## References

1. Schuller B. W., Batliner A. M. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. New York: Wiley, 2013, 344 p.
2. Trigeorgis G., Ringeval F., Brueckner R., Marchi E., Nicolaou M. A., Schuller B., Zafeiriou S. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Shanghai, China, 20–25 March 2016)*. Shanghai, China: IEEE, 2016, pp. 5200–5204.

3. Huang X., Acero A., Hon H.-W. *Prentice Hall*. New Jersey, 2001, 980 p.
4. Fant G. K. *Akusticheskaya teoriya recheobrazovaniya* [Acoustic theory of speech formation]. Transl. from Engl. by L. A. Varshavskiy and V. I. Medvedev. Moscow: Nauka, 1964, 284 p. [In Russian]
5. Whitehead R. L., Schiavetti N., Metz D. E., Farrell T. *Journal of Communication Disorders*. 1999, vol. 32, no. 2, pp. 79–95.
6. Bóna J. *The Journal of the Acoustical Society of America*. 2014, vol. 136, no. 2, pp. 116–121.
7. Bakaev A. V. *Almanac of Modern Science and Education*. 2014, vol. 4, no. 83, pp. 28–32.
8. Alimuradov A. K., Tychkov A. Yu., Churakov P. P., Torgashin S. I. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki* [University proceedings. Volga region. Engineering sciences]. 2017, no. 3 (47), pp. 81–93. [In Russian]
9. Alimuradov A. K., Kvitka Yu. S., Churakov P. P., Tychkov A. Yu. *Izmerenie. Monitoring. Upravlenie. Kontrol'* [Measurement. Monitoring. Management. Control]. 2018, no. 4 (26), pp. 53–65. [In Russian]
10. Huang N. E., Zheng Sh., Steven R. L. *Proceedings of the Royal Society of London*. 1998, vol. A454, pp. 903–995.
11. Colominasa M. A., Schlotthauera G., Torres M. E. *Biomedical Signal Processing and Control*. 2014, vol. 14, pp. 19–29.
12. Zhaohua Wu., Huang N. E. *Advances in Adaptive Data Analysis*. 2009, vol. 1, no. 1, pp. 1–41.
13. Yeh J.-R., Shieh J.-S., Huang N. E. *Advances in Adaptive Data Analysis*. 2010, vol. 2, no. 2, pp. 135–156.
14. Torres M. E., Colominas M. A., Schlotthauer G., Flandrin P. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Prague, Czech Republic, 22–27 May 2011)*. Prague, Czech Republic: IEEE, 2011, pp. 4144–4147.
15. Hlavnička J., Čmejla R., Tykalová T., Šonka K., Růžička E., Ruzs J. *Scientific Reports*. 2017, vol. 7, no. 12, 13 p.
16. Alimuradov A. K., Tychkov A. Yu., Kvitka Yu. S. *International Russian Automation Conference (RusAutoCon) (Sochi, Russia, 9–16 Sept. 2018)*. Sochi, Russia: IEEE, 2018, 6 p.
17. Bachu R. G., Kopparthi S., Adapa B., Barkana B. D. *American Society for Engineering Education (ASEE) Zone Conference Proceedings (Pittsburgh, USA, 22–25 June 2008)*. Pittsburgh, USA, 2008, pp. 1–7.
18. Moattar M. H., Homayounpour M. M. *17th European Signal Processing Conference (EUSIPCO 2009) (Glasgow, Scotland, August 24–28 2009)*. Glasgow, Scotland: Eusipco, 2009, pp. 2549–2553.
19. Saha G., Sandipan Ch., Suman S. *The National Conference on Communications (28–30 January 2005)*. New Delhi: Allied Publishers, 2005, 5 p.

**Алимурадов Алан Казанферович**

кандидат технических наук,  
директор студенческого научно-производственного  
бизнес-инкубатора,  
Пензенский государственный университет  
(Россия, г. Пенза, ул. Красная, 40)  
E-mail: alansapfir@yandex.ru

**Alimuradov Alan Kazanferovich**

candidate of technical sciences,  
director of student research and production business  
incubator,  
Penza State University  
(40 Krasnaya street, Penza, Russia)

**Тычков Александр Юрьевич**

кандидат технических наук, заместитель директора  
научно-исследовательского института  
фундаментальных и прикладных исследований,  
Пензенский государственный университет  
(Россия, г. Пенза, ул. Красная, 40)  
E-mail: tychkov-a@mail.ru

**Tychkov Aleksandr Yur'evich**

candidate of technical sciences,  
deputy director at the Research Institute  
for Basic and Applied Studies,  
Penza State University  
(40 Krasnaya street, Penza, Russia)

**Чураков Петр Павлович**

доктор технических наук, профессор,  
кафедра информационно-измерительной техники  
и метрологии,  
Пензенский государственный университет  
(Россия, г. Пенза, ул. Красная, 40)  
E-mail: churakov-pp@mail.ru

**Churakov Petr Pavlovich**

doctor of technical sciences, professor,  
sub-department of information  
and measuring equipment and metrology,  
Penza State University  
(40 Krasnaya street, Penza, Russia)

**Образец цитирования:**

Алимурадов, А. К. Способ автоматизированной сегментации речевых сигналов для определения временных паттернов естественно выраженных психоэмоциональных состояний / А. К. Алимурадов, А. Ю. Тычков, П. П. Чураков // Измерение. Мониторинг. Управление. Контроль. – 2019. – № 3 (29). – С. 48–60. – DOI 10.21685/2307-5538-2019-3-6.