

**ПРИБОРЫ, СИСТЕМЫ И ИЗДЕЛИЯ
МЕДИЦИНСКОГО НАЗНАЧЕНИЯ**

УДК 004.934

DOI 10.21685/2307-5538-2019-4-7

*А. К. Алимуратов, А. Ю. Тычков, П. П. Чураков, Б. В. Султанов***СПОСОБ ОПРЕДЕЛЕНИЯ ФОРМАНТНОЙ РАЗБОРЧИВОСТИ
РЕЧИ ДЛЯ ОЦЕНКИ ПСИХОЭМОЦИОНАЛЬНОГО
СОСТОЯНИЯ ОПЕРАТОРОВ СИСТЕМ УПРАВЛЕНИЯ
С ВЫСОКОЙ СТЕПЕНЬЮ ОТВЕТСТВЕННОСТИ***A. K. Alimuradov, A. Yu. Tychkov, P. P. Churakov, B. V. Sultanov***METHOD TO DETERMINE FORMANT SPEECH
INTELLIGIBILITY FOR ESTIMATING PSYCHO-EMOTIONAL
STATE OF CONTROL SYSTEM OPERATORS
WITH A HIGH DEGREE OF RESPONSIBILITY**

А н н о т а ц и я. Актуальность и цели. На сегодняшний день для оценки психоэмоционального состояния применяются различные способы, основанные на анализе: мимики и жестикуляции, параметров физиологической активности организма, биохимических параметров крови, параметров рукописного и клавиатурного написания текстов, окулографии и др. Наиболее адаптивными являются способы на основе анализа речи. Целью работы является разработка способа определения формантной разборчивости речи для повышения эффективности оценки психоэмоционального состояния операторов систем управления с высокой степенью ответственности. **Материалы и методы.** В разработке способа использовались: уникальная технология адаптивной декомпозиции нестационарных сигналов – улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом; формантный анализ. Программная реализация способа была выполнена в среде математического моделирования © Matlab (MathWorks). **Результаты.** Разработан способ оценки психоэмоционального состояния, применимый для операторов систем управления с высокой степенью ответственности. Суть способа заключается в адаптивном разложении речевого сигнала на равноартикуляционные полосы частот, с помощью улучшенной полной множественной декомпозиции на эмпирические моды с адаптивным шумом и определении формантной разборчивости, а также частоты, ширины, уровня и добротности формантных областей в данных полосах с помощью формантного анализа. Кратко представлены математическое описание декомпозиции и упрощенная блок-схема предложенного способа. Проведено исследование способа с использованием сформированной базы фонетически сбалансированных речевых сигналов, зарегистрированных у здоровых людей и людей с психоэмоциональными расстройствами. **Выводы.** В соответствии с результатами исследования разработанный способ обеспечивает высокую точность определения психоэмоциональных расстройств (ошибка первого рода 8,16 %,

ошибка второго рода 5,66 %) и может успешно тестироваться для оценки психоэмоционального состояния операторов систем управления с высокой степенью ответственности.

A b s t r a c t. Background. At present, various methods based on analysis of facial expressions and gesticulation, parameters of physiological activity of the organism, blood biochemical parameters, parameters of text handwriting and keyboarding, oculography, etc. are used to assess psycho-emotional state. Methods based on speech analysis are the most adaptive. The purpose of the work is to develop a method for determining formant intelligibility of speech to increase the assessment efficiency of psycho-emotional state of control system operators with a high degree of responsibility. **Materials and methods.** A unique technology for adaptive decomposition of non-stationary signals, namely, the improved complete ensemble empirical mode decomposition with adaptive noise, and formant analysis were used in the development of the method. Software implementation of the method was performed in the mathematical modeling environment © Matlab (MathWorks). **Results.** A method for assessing psycho-emotional state, applicable to control system operators with a high degree of responsibility, has been developed. The nature of the method consists in adaptive decomposition of a speech signal into frequency bands of equal articulation, using the improved complete ensemble empirical mode decomposition with adaptive noise, and determining formant intelligibility, as well as frequency, width, level and quality factor of formant regions in these bands using formant analysis. A mathematical description of the decomposition and a simplified block diagram for the proposed method are briefly presented. A study of the method was conducted using a formed base of phonetically balanced speech signals recorded in healthy people and people with psycho-emotional disorders. **Conclusions.** In accordance with the research results, the developed method provides high determination accuracy of psycho-emotional disorders (the first kind error is 8.16 %, the second kind error is 5.66 %), and can be successfully tested to assess psycho-emotional state of control system operators with a high degree of responsibility.

К л ю ч е в ы е с л о в а: речевой сигнал, улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом, формантная разборчивость, формантный анализ, психоэмоциональное расстройство.

К e y w o r d s: speech signal, improved complete ensemble empirical mode decomposition with adaptive noise, formant intelligibility, format analysis, psycho-emotional disorder.

Введение

Оценка психоэмоционального состояния человека на сегодняшний день является важной социально значимой проблемой [1]. Особенно это важно в тех отраслях человеческой деятельности, которые сопряжены с повышенным риском возникновения техногенных аварий.

На сегодняшний день для оценки психоэмоционального состояния применяются различные способы, основанные на анализе: мимики и жестуляции, параметров физиологической активности организма, биохимических параметров крови, параметров рукописного и клавиатурного написания текстов и окулографии [2]. Недостатком указанных способов является контакт с человеком (регистрация, взятие пробы, написание и т.п.), что, безусловно, влияет на психоэмоциональное состояние, эффективно оценить которое уже не представляется возможным. Наиболее адаптивным является способ на основе анализа речевых сигналов [3].

В последние два десятилетия исследования в области оценки психоэмоционального состояния по речи активно поддерживаются международными фондами и грантами организаций:

– «Deep Learning Speech Enhancement», Industry Cooperation with HUAWEI TECHNOLOGIES, 12.11.2016 – 11.11.2018;

– «Training network on Automatic Processing of Pathological Speech» (#766287), EU H2020 Marie Skłodowska-Curie Innovative Training Networks European Training Networks (MSCA-ITN-ETN: ENG), 01.11.2017 – 31.10.2021;

– «Sentiment Analyses», Industry Cooperation with BMW AG, 01.05.2018 – 31.05.2021;

– «Поиск скрытых паттернов пограничных психических расстройств и разработка системы экспресс оценки состояния психического здоровья человека», Российский научный фонд (№ 17-71-20029), 01.09.2017 – 31.07.2020.

На рынке речевых технологий представлены коммерческие системы оценки психоэмоционального состояния человека: «*lingWAVES*» [4] и «*Sense*» [5]. Наибольший интерес в данных системах представляют способы обработки речевых сигналов и возможность их применения для операторов систем управления с высокой степенью ответственности. Однако в силу коммерческой тайны производители такую информацию не предоставляют. По этой причине модернизация существующих и разработка новых способов обработки речевых сигналов остаются в центре внимания специалистов при создании систем «человек – машина».

Целью данного исследования является определение формантной разборчивости речи операторов систем управления с высокой степенью ответственности для оценки психоэмоционального состояния на основе адаптивных технологий обработки речевых сигналов. Работа является развитием ранее опубликованных трудов авторов [2, 6] и выполнена при финансовой поддержке совета по грантам Президента РФ, проект «Исследование информативно-значимых параметров речевых сигналов: поиск уникально новых признаков естественно выраженных эмоций для повышения точности оценки психоэмоционального состояния операторов систем управления с высокой степенью ответственности» № СП-246.2018.5, 2018–2020 гг.

Речеобразующая система и речь

Речеобразующая система человека состоит из легких, трахеи, полости глотки (горла), языка, ротовой и носовой полостей (рис. 1,а). Воздушный поток, выдавливаемый из легких, проходит через голосовые связки и речевой тракт, колеблясь на разных частотах.

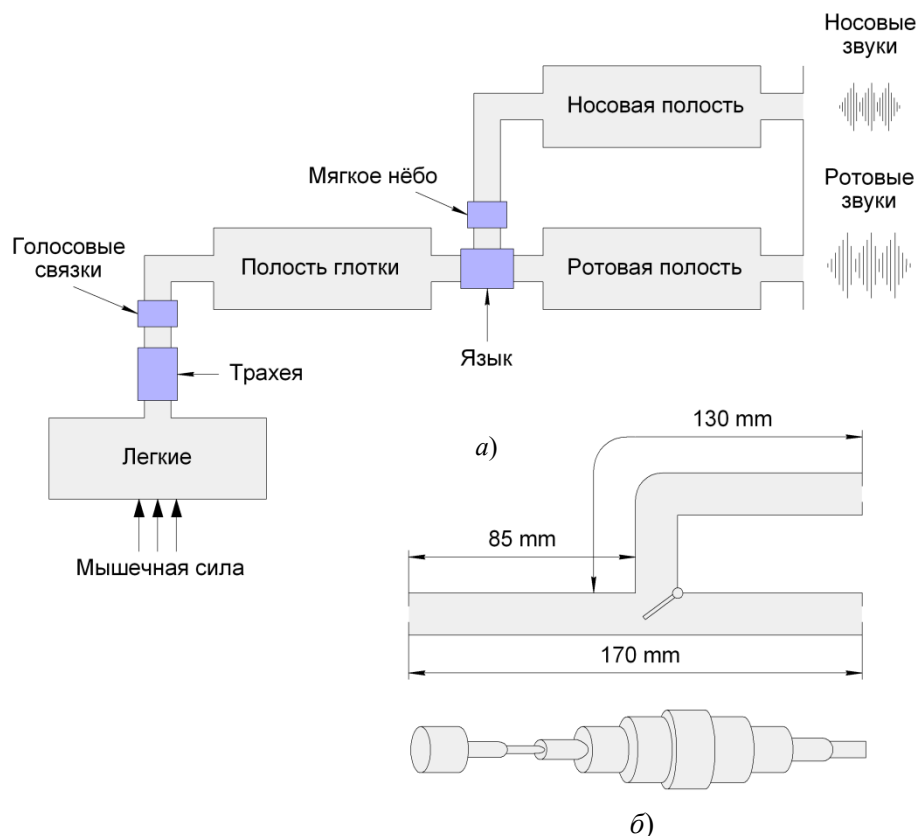


Рис. 1. Речеобразующая система человека

Речевой сигнал представляет собой нестационарный акустический сигнал сложной формы, амплитудные и частотные характеристики которого быстро изменяются во времени. Речь состоит из вокализованных и невокализованных участков, образующихся соответственно в результате периодических и непериодических колебаний голосовых связок. Периодические

колебания голосовых связок называется основным тоном. Частота колебаний связок называется частотой основного тона.

Известно, что звуки являются неделимыми элементами речи, характеризуются уникальным распределением энергии в спектральной области. В спектре каждого звука речи присутствуют области повышенной концентрации энергии, называемые формантами. Расположение формант, их число и параметры определяются самим процессом воспроизведения речеобразующей системой человека и позволяют однозначно установить звук речи. Изменяемый объем полостей гортани, рта, движения артикуляционного аппарата говорящего всегда подобны при создании каждого из звуков речи. Анатомические отличия людей друг от друга порождают отличия в высоте и тембре голоса, но характер распределения формант остается постоянным.

Формантная структура звуков, формантная разборчивость

Речевой тракт (рис. 1,б) представляет собой трубку неоднородного сечения, длиной 170 мм (у взрослого человека). Формантная структура звуков описывается резонансными частотами речевого тракта. Артикуляционный отдел речевого тракта очень чувствителен к нарушениям работы нервной системы. В работах [3, 7, 8] представлены результаты исследований частотной локализации формантных областей при патологиях в моторике артикуляционного отдела речевого тракта. Различные патологии (физиологические или психологические) вызывают смещения формантных областей по частоте и приводят к деформации и трансформированию звуков при воспроизведении. В соответствии с этим следует, что параметры формант (частота, ширина, уровень и др.) способны служить маркерами формантной разборчивости речи при психоэмоциональных расстройствах человека.

Разборчивость речи представляет собой меру понятности, выраженную относительным числом правильно понятых элементов речи: звуков, слогов, слов и фраз. Формантная разборчивость определяется произведением частотного F и динамического D диапазонов тракта связи, т.е. $A = kFD$, где k – коэффициент пропорциональности. Учитывая, что объем информации, передаваемой по речевому тракту за время T , также пропорционален произведению F и D , получаем $I = AT$, т.е. формантная разборчивость пропорциональна объему информации в единицу времени.

Несмотря на разнообразие применяемых обозначений, определений, большинство предложенных методов определения формантной разборчивости представляют собой развитие идей Коларда [9]. Теория разборчивости и методы определения основаны на представлении речевого сигнала в виде совокупности формант. Формантная теория расчета принимает форманты за элементы звуков речи, несущие информацию и определяющие разборчивость. Понятие «форманта» распространяется на все звуки речи. Нужно иметь в виду различие между смыслом, вкладываемым в это понятие формантной теории, и акустическим определением спектров звуков. Спектральный анализ звуков обнаруживает до десяти формант в отдельных случаях. Однако исследования речи показывают, что слуховое восприятие (по крайней мере для гласных звуков) полностью определяется четырьмя формантами, поэтому формантная теория относит к формантам только те области концентрации энергии в спектре, которые определяют звук.

Формантная теория разборчивости основана на предположении, что в пределах некоторой полосы частот речевого сигнала Δf_s вероятность встречаемости формант $\Delta A_{smax} = n_s/N$, где N – число формант во всей полосе частот речевого сигнала; n_s – число формант в полосе Δf_s .

Статистическое распределение формант речевого сигнала может отличаться от распределения формант, определяющих разборчивость речи, а уровень спектрального распределения речи может отличаться от спектрального распределения формант. Для оценки вероятности встречаемости формант в некоторой полосе частот речевого сигнала необходим анализ слухового восприятия звуков речи.

Адаптивная обработка

Формантный анализ успешно применяется для обнаружения и оценки психоэмоциональной окраски речи человека [3, 7–9]. Представленное разнообразие способов анализа обусловлено как важностью проблемы, так и отсутствием достаточно эффективных способов ее реше-

ния. Учитывая нерегулярность моторики речеобразующей системы при психоэмоциональных расстройствах, возможности этих способов существенно ограничены. Это ограничение обусловлено использованием неадаптивных методов обработки сложных нестационарных речевых сигналов. По этой причине актуальным является разработка новых способов, основанных на адаптированных методах обработки речевых сигналов, повышающих эффективность оценки психоэмоционального состояния операторов систем управления с высокой степенью ответственности.

Исследования методов обработки речевых сигналов выявили перспективность использования адаптивной технологии анализа нестационарных сигналов – улучшенной полной множественной декомпозиции на эмпирические моды с адаптивным шумом (ПМДЭМАШ) [10]. В настоящее время технология улучшенной ПМДЭМАШ широко используется в различных научных приложениях, в том числе и при решении задач оценки психоэмоционального состояния [11, 12].

Классическая декомпозиция на эмпирические моды (ДЭМ) – это адаптивный метод анализа нестационарных сигналов, возникающих в нелинейных системах, который обеспечивает локальное разложение сигнала на быстрые и медленные колебательные функции [13]. В результате разложения исходный сигнал представляется в виде суммы амплитудно- и частотно-модулированных функций – эмпирических мод (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).

Отличительной особенностью улучшенной ПМДЭМАШ является добавление к исходному сигналу контролируемого шума для создания новых экстремумов. Это позволяет устранить недостатки существующих разновидностей декомпозиций [10, 13]: смешивание ЭМ; остаточный шум в ЭМ; «паразитных» ЭМ на ранних этапах декомпозиции.

Алгоритм и математическое описание метода улучшенной ПМДЭМАШ:

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

$$r_1(n) = M(x_j(n)),$$

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:

$$\widetilde{IMF}_2(n) = r_1(n) - r_2(n) = r_1(n) - M(r_1(n) + \beta_1 E_2(w_j(n))).$$

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

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

5. Вычисляется i -я мода

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

6. Переход к шагу 4 для следующего значения i .

Константы β_i выбираются так, чтобы получить желаемое отношение сигнал/шум между добавленным шумом и остатком, к которому добавляется шум. Для получения шумовых реализаций с меньшими амплитудами для последних стадий декомпозиции в остальных модах будем воспринимать шум как результат его предварительной обработки ДЭМ, т.е. без их нормализации стандартным отклонением.

Описание способа

На рис. 2 представлена упрощенная блок-схема способа определения формантной разборчивости речи для оценки психоэмоционального состояния человека. Суть способа заключается в адаптивном разложении речевого сигнала на равноартикуляционные полосы частот, с помощью улучшенной ПМДЭМАШ и определении разборчивости в данных полосах с помощью формантного анализа. Структурно способ делится на четыре этапа: предварительная обработка (блоки 2–4), формантный анализ (блок 5/8), декомпозиция (блоки 6, 7) и непосредственное определение формантной разборчивости (блок 9). Блоки 11 и 12 применяются только для исследования предложенного способа. На рис. 3 представлена краткая иллюстрация, поясняющая функционирование способа. Рассмотрим вкратце этапы обработки предложенного способа.

Блок 1. Ввод речевого сигнала осуществляется со следующими параметрами: частота дискретизации 8000 Гц, разрядность квантования 16 бит. На рис. 3,а представлена осциллограмма речевого сигнала – предложение «Лара врала, что была королевой Урала».

Блок 2. Обнаружение вокализованных участков представляет собой определение точных границ звуков речи, образованных с участием голосовых связок в общем речевом потоке. Корректное обнаружение границ вокализованных участков не только повышает эффективность форматного анализа, но и уменьшает количество вычислительных операций. На рис. 3,б вокализованные участки отмечены прямоугольниками зеленого цвета. Неиспользуемые в дальнейшей работе способа невокализованные участки удалены.

Блок 3. Сегментация представляет собой линейное деление вокализованных участков речевого сигнала на составляющие отрезки, называемые фрагментами. В основе сегментации заложен принцип: свойства речевого сигнала с течением времени изменяются медленно. Это приводит к кратковременному анализу, в котором фрагменты выделяются и обрабатываются так, как если бы они были короткими участками отдельных звуков с отличающимися свойствами. На рис. 3,в представлен фрагмент вокализованного участка длительность 100 мс. Длительность фрагмента в 100 мс определена эмпирически для повышения эффективности формантного анализа и декомпозиции.

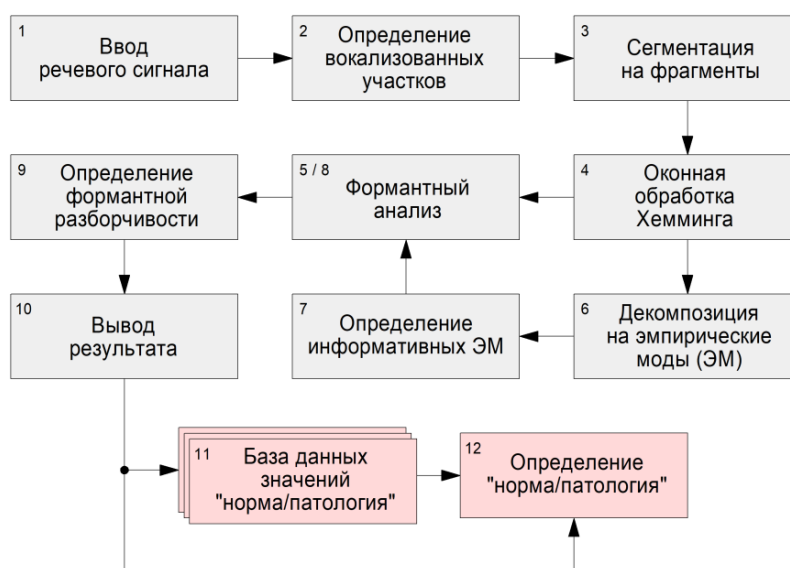


Рис. 2. Упрощенная блок-схема способа определения формантной разборчивости для оценки психоэмоционального состояния

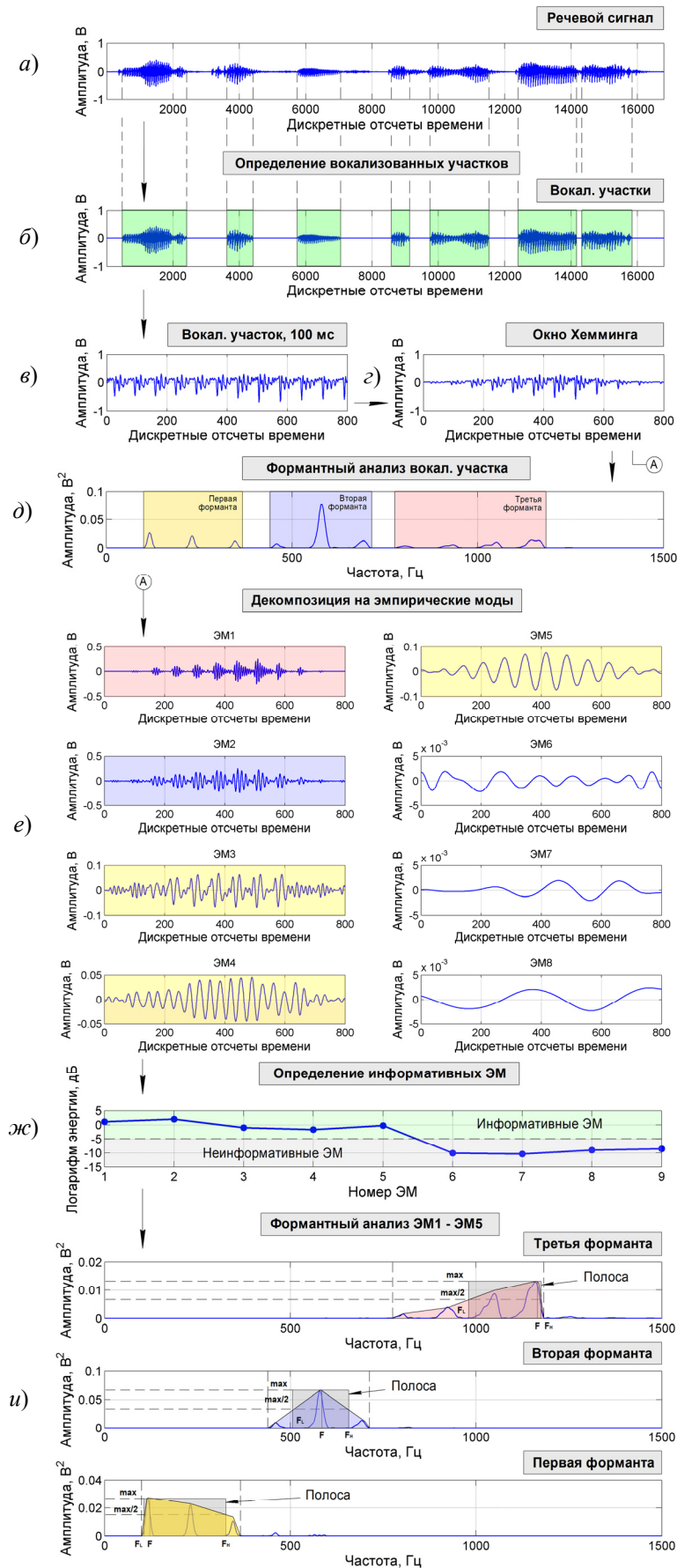


Рис. 3. Функционирование способа определения формантной разборчивости для оценки психоэмоционального состояния

Блок 4. Сегментация на фрагменты во временной области эквивалентна умножению сигнала на прямоугольное окно. В частотной области эквивалентно свертке. Из-за того, что передаточная функция прямоугольного окна имеет боковые лепестки, уровень которых достигает 20 %, спектр сигнала искажается (эффект Гиббса) [14]. Для уменьшения подобных искажений при обработке речевых сигналов применяются окна с плавно спадающими до нуля краями – окно Хемминга. На рис. 3,з представлен фрагмент вокализованного участка после оконной обработки.

Блок 5. Форматный анализ для задачи оценки психоэмоционального состояния человека представляет собой определение числовых значений частоты, ширины, уровня и добротности полосы формантных областей. Частота формантной области – F определяется как частотная составляющая, имеющая максимальный по амплитуде уровень. Ширина формантной области ΔF определяется как разность нижней F_L и верхней F_H частотной составляющей, на которых мощность спектральной огибающей снижается вдвое [15]. Добротность определяется как отношение частоты и ширины формантной области $Q = F/\Delta F$. На рис. 3,з представлен пример формантного анализа вокализованного участка. Для упрощения формы спектра формант строились с использованием прямых линий, соединяющих вершины частотных составляющих.

Блок 6. Важным условием адаптивного разложения речевого сигнала на равноартикуляционные полосы частот, является возможность формирования адаптивного базиса, функционально зависящего от структуры самого сигнала. Это условие выполняется в методе улучшенной ПМДЭМАШ, частотно-избирательные свойства которого обеспечивают адаптивность при разложении и позволяют учитывать внутренние особенности сигнала (скрытые модуляции, области концентрации энергии и т.п.).

Результат разложения приведен на рис. 3,е. Вокализованный участок речевого сигнала был разложен на 9 частотных составляющих – равноартикуляционных частотных диапазонов. Однако, для удобства отображения ЭМ9 не отображена на рис. 3,е.

Блок 7. Задача определения информативных ЭМ сводится к обнаружению частотных диапазонов, отражающих информацию о частотной локализации формантных областей. В основе определения заложен принцип выделения информативных мод имеющих большую энергию чем трендовые моды [2]. Амплитудное распределение ЭМ хорошо описывается с помощью функции кратковременной энергии. Для сжатия амплитуды сигнала в большом динамическом диапазоне применяется логарифмирование энергии, максимально приближая работу способа к работе слухового аппарата человека. На рис. 3,ж представлена зависимость логарифма энергии ЭМ от ее номера. Как видно из рис. 3,е и 3,ж наиболее информативными, отражающими информацию о частотной локализации формантных областей являются ЭМ1 – ЭМ5.

Блок 8. Иллюстрация результата формантного анализа для информативных ЭМ представлена на рис. 3,и. Локализации формантных областей:

- первая форманта – сумма ЭМ3, ЭМ4 и ЭМ5 (выделена желтым цветом);
- вторая форманта – ЭМ2 (выделена синим цветом);
- третья форманта – ЭМ1 (выделена красным цветом).

Блок 9. Формантная разборчивость речи представляет собой величину смещения частотной локализации формантных резонаторов вследствие патологий моторики речеобразующей системы. В реальных условиях определению формант будут мешать различные искажения, помехи и акустические шумы, создающие маскировку. Поэтому формантная разборчивость уменьшится и будет определяться по следующей формуле [15]:

$$\Delta A_s = \Delta A_{smax} P_s(E'),$$

где $P_s(E')$ – коэффициент, показывающий, какая часть формант в полосе будет воспринята в реальных условиях; E' – уровень восприятия формант.

Считая значения ΔA_s в смежных полосах спектра речи независимыми, получаем свойство аддитивности формант. Формантная разборчивость для всей полосы частот определяется по формуле

$$A = \sum_1 \Delta A_s = \sum_1 \Delta A_{smax} P_s(E').$$

При разложении с помощью улучшенной ПМДЭМАШ на пять равноартикуляционных частотных диапазонов при условии, что $\Delta A_{1\max} = \Delta A_{2\max} = \dots = 0,25$ и $\sum_{s=1}^5 \Delta A_{s\max} = 1$, формантная разборчивость

$$A = \frac{1}{5} \sum_{s=1}^5 P_s(E').$$

Уровень восприятия формант $E' = B'_p - b - B_{noise}$, где B'_p – уровень спектра формант; b – затухание при передаче речи; B_{noise} – спектральный уровень шума (с интегральным уровнем ощущения не менее 10дБ). Данные значения b взяты из работы [15].

Блок 10. Вывод результата заключается в формировании полученных значений формантной разборчивости, частоты, ширины, уровня и добротности полосы формантных областей в удобный для дальнейшего определения «норма/патология» вида. Это осуществляется посредством распространенного метода, используемого в приложениях по обработке речевых сигналов – смешанных гауссовских моделей (*Gaussian mixture model, GMM*) [16].

Исследование способа

Для тестирования предложенного способа сформирована группа исследуемых людей при поддержке центра психологической диагностики федерального казенного учреждения здравоохранения «Медицинская санитарная часть Министерства внутренних дел Российской Федерации по Пензенской области» (ЦПД ФКУЗ «МСЧ МВД РФ по Пензенской области»). Группа сформирована из наблюдаемых врачом-психиатром ЦПД ФКУЗ «МСЧ МВД РФ по Пензенской области» пациентов с эмоциональной нагрузкой, соответствующей 3-му классу по степени напряженности. В группу отобрано 53 человека мужского и женского пола, в возрасте от 18 до 79 лет. В том же количестве сформирована контрольная группа условно здоровых людей (без признаков психоэмоциональных расстройств и дефектов речи) из числа сотрудников и преподавателей Пензенского государственного университета.

В соответствии с разработанной авторами методикой была зарегистрирована база данных речевых сигналов. В подготовленном помещении с помощью специально настроенных параметров и расположения микрофона осуществлялась регистрация речи, воспроизводимой участниками обеих групп, при чтении трех видов текстового материала:

- фонетически сбалансированный текст публицистической статьи из газеты, состоящий из 100 слов;
- фонетически сбалансированный текст литературного произведения, состоящий из 100 слов;
- специально подобранные слова, составленные из них словосочетаний и предложения, включающие в себя звуки, максимально отражающие нарушения при воспроизведении речи.

Для оценки эффективности разработанного способа использовался параметр – ошибки первого и второго рода [16]. Анализируя значение формантной разборчивости, частоты, ширины, уровня и добротности полосы формантных областей, определялось наличие/отсутствие психоэмоционального расстройства. Ошибкой первого рода будет ложное присваивания статуса «норма» речевому сигналу, произнесенному человеком с эмоциональным расстройством, а ошибкой второго рода – ложное присваивание статуса «патология» речевому сигналу, произнесенному здоровым человеком. Исследование проводилось в два этапа: обучение и тестирование. Для определения ошибки первого рода обучение осуществлялось на группе здоровых людей и представляло собой формирование базы данных значений информативных параметров (формантной разборчивости, частоты, ширины, уровня и добротности полосы формантных областей), соответствующих статусу «норма». На этапе обучения также автоматически формируется пороговое значение статуса «норма» как среднее арифметическое значение данных информативных параметров. Тестирование на группе людей с психоэмоциональными расстройствами заключается в последовательном сравнении значений информативных параметров исследуемых речевых сигналов с пороговыми значениями.

Для определения ошибки второго рода последовательность действия исследования сохраняется, изменяются только группы. Обучение осуществляется на группе людей с психоэмоциональными расстройствами, а тестирование – на группе условно здоровых людей.

Все этапы обработки сигналов и анализа данных были выполнены в среде математического моделирования © *Matlab (MathWorks)*.

Исследование проводилось для трех видов речевых сигналов: слоги, слова и предложения. Сигналы длительностью не менее 100 мс были специально подготовлены с помощью аудиоредактора из зарегистрированной базы данных речевых сигналов. Для увеличения анализируемой информации слоги и слова в сигналах повторялись несколько раз (30 раз минимум). В табл. 1 представлены результаты определения психоэмоциональных расстройств.

Таблица 1

Результаты определения психоэмоционального расстройства

Прогнозируемый результат	Результат обнаружения		Ошибки первого и второго рода, %	
	Патология	Норма		
Слоги				
Патология	31 чел.	22 чел.	1st, α	41,5
Норма	13 чел.	40 чел.	2nd, β	24,52
Слова				
Патология	35 чел.	18 чел.	1st, α	33,96
Норма	9 чел.	44 чел.	2nd, β	16,98
Предложения				
Патология	49 чел.	4 чел.	1st, α	8,16
Норма	3 чел.	50 чел.	2nd, β	5,66

Результаты исследования и выводы

В сравнении между тремя видами сигналов наименьшее количество ошибок первого и второго родов ($\alpha = 8,16\%$, $\beta = 5,66\%$) достигается при анализе предложений. Объясняется это большим количеством анализируемой информации о формантной разборчивости, частоте, ширине, уровне и добротности полос формантных областей вокализованных участков в предложениях. Худшие результаты наблюдаются в анализе более коротких сигналов: слоги $\alpha = 41,5\%$, $\beta = 24,52\%$ и слова $\alpha = 33,96\%$, $\beta = 16,98\%$, которые никак не применимы в реальной практике.

В соответствии с результатами исследований можно сделать выводы:

- психоэмоциональные расстройства человека достаточно серьезно влияют на вокализованные характеристики речевого тракта;
- значения формантной разборчивости, частоты, ширины, уровня и добротности полосы формантных областей вокализованных участков анализируемых сигналов достаточно полно отображают нарушения работы артикуляционного отдела речевого тракта;
- предложенный способ обеспечивает точное определение психоэмоционального расстройства при анализе более длительных сигналов;
- разработанный способ на основе улучшенной ПМДЭМАШ и формантного анализа предварительно может тестироваться на реальных речевых сигналах, зарегистрированных у операторов систем управления с высокой степенью ответственности для оценки их психоэмоционального состояния.

Библиографический список

1. Здравоохранение в России 2017. Статистический сборник. – URL: http://www.gks.ru/free_doc/doc_2017/zdrav17.pdf. (дата обращения: 20.06.2019).
2. Improved CEEMDAN based speech signal analysis algorithm for mental disorders diagnostic system. Pitch frequency detection and measurement / A. K. Alimuradov, A. Yu. Tychkov, A. V. Kuzmin, P. P. Churakov, A. V. Ageykin, G. V. Vishnevskaya // International Journal of Embedded and Real-Time Communication Systems (IJERTCS). – 2019. – Vol. 10, № 2. – P. 22–47.
3. Schuller, B. W. Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing / B. W. Schuller, A. M. Batliner. – New York : Wiley, 2013. – P. 344.

4. Официальный сайт компании «WEVOSYS». – URL: <http://www.wevosys.com> (дата обращения: 20.06.2019).
5. Официальный сайт компании «NEMESYSKO». – URL: <http://www.nemesysco.com> (дата обращения: 20.06.2019).
6. Alimuradov, A. K. A method to determine cepstral markers of speech signals under psychogenic disorders / A. K. Alimuradov, A. Yu. Tychkov, A. V. Ageykin, P. P. Churakov // Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT) (Yekaterinburg, Russia, 7–8 May 2018). – Yekaterinburg, Russia : IEEE, 2018. – P. 128–131.
7. Schwartz, T. L. Depression: Treatment Strategies and Management (Medical Psychiatry Series) / T. L. Schwartz and T. J. Petersen. – Boca Raton : Informa Healthcare, 2009. – 339 p.
8. Williamson, J. R. Vocal and facial biomarkers of depression based on motor incoordination and timing / J. R. Williamson, Th. F. Quatieri, B. S. Helfer, G. Ciccarelli, D. D. Mehta // 4th International Workshop on Audio/Visual Emotional Challenge (AVEC) (Orlando, USA, 3–7 November 2014). – Orlando, USA : New York, NY ACM, 2014. – P. 65–72.
9. 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.
10. Colominasa M. A. Improved complete ensemble EMD: a suitable tool for biomedical signal processing / M. A. Colominasa, G. Schlotthauer, M. E. Torres // Biomed. Signal Proces. – 2014. – Vol. 14. – P. 19–29.
11. Stassi, A. E. Vocal fold activity detection from speech related biomedical signals: a preliminary study / A. E. Stassi, G. A. Alzamendi, G. Schlotthauer, M. E. Torres // VI Latin American Congress on Biomedical Engineering CLAIB 2014 (Parana, Argentina, 29 – 31 October 2014). IFMBE Proceedings. – Springer, 2014. – Vol. 49. – P. 520–523.
12. Torres, M. E. Empirical mode decomposition. Spectral properties in normal and pathological voices / M. E. Torres, G. Schlotthauer, H. L. Rufiner, M. C. Jackson-Menaldi // 4th European Conference of the International Federation for Medical and Biological Engineering, ECIFMBE. (Antwerp, Belgium, 23–27 November 2008). – Antwerp, Belgium : Springer, 2008. – P. 252–255.
13. 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. – A 454. – P. 903–995.
14. Афонский, А. А. Цифровые анализаторы спектра, сигналов и логики / А. А. Афонский, В. П. Дьяконов. – Москва : Солон-Пресс, 2005. – 576 с.
15. Михайлов, В. Г. Измерение параметров речи / В. Г. Михайлов, Л. В. Златоусова ; под ред. М. А. Сапожникова. – Москва : Радио и связь, 1987. – 168 с.
16. Huang, X. Spoken Language Processing. Guide to Algorithms and System Developmen / X. Huang, A. Acero, H.-W. Hon. – New Jersey : Prentice Hall, 2001. – 980 p.

References

1. Zdravookhranenie v Rossii 2017. Statisticheskiiy sbornik [Healthcare in Russia 2017. Statistical compendium]. Available at: http://www.gks.ru/free_doc/doc_2017/zdrav17.pdf. (accessed Jun. 20, 2019). [In Russian]
2. Alimuradov A. K., Tychkov A. Yu., Kuzmin A. V., Churakov P. P., Ageykin A. V., Vishnevskaya G. V. *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)*. 2019, vol. 10, no. 2, pp. 22–47.
3. Schuller B. W., Batliner A. M. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. New York: Wiley, 2013, p. 344.
4. *Ofitsial'nyy sayt kompanii «WEVOSYS»* [Official website of "WEVOSYS"]. Available at: <http://www.wevosys.com> (accessed Jun. 20, 2019).
5. *Ofitsial'nyy sayt kompanii «NEMESYSKO»* [Official website of NEMESYSKO]. Available at: <http://www.nemesysco.com> (accessed Jun. 20, 2019).
6. Alimuradov A. K., Tychkov A. Yu., Ageykin A. V., Churakov P. P. *Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT) (Yekaterinburg, Russia, 7–8 May 2018)*. Yekaterinburg, Russia: IEEE, 2018, pp. 128–131.
7. Schwartz T. L., Petersen T. J. *Depression: Treatment Strategies and Management (Medical Psychiatry Series)*. Boca Raton: Informa Healthcare, 2009, 339 p.
8. Williamson J. R., Quatieri Th. F., Helfer B. S., Ciccarelli G., Mehta D. D. *4th International Workshop on Audio/Visual Emotional Challenge (AVEC) (Orlando, USA, 3–7 November 2014)*. Orlando, USA: New York, NY ACM, 2014, pp. 65–72.
9. Fant G. K. *Akusticheskaya teoriya recheobrazovaniya* [Acoustic theory of speech formation]; transl. from Engl. L. A. Varshavskiy and V. I. Medvedev. Moscow: Nauka, 1964, 284 p. [In Russian]
10. Colominasa M. A., Schlotthauer G., Torres M. E. *Biomed. Signal Proces.* 2014, vol. 14, pp. 19–29.

11. Stassi A. E., Alzamendi G. A., Schlotthauer G., Torres M. E. *VI Latin American Congress on Biomedical Engineering CLAIB 2014 (Parana, Argentina, 29 – 31 October 2014). IFMBE Proceedings*. Springer, 2014, vol. 49, pp. 520–523.
12. Torres M. E., Schlotthauer G., Rufiner H. L., Jackson-Menaldi M. C. *4th European Conference of the International Federation for Medical and Biological Engineering, ECIFMBE. (Antwerp, Belgium, 23–27 November 2008)*. Antwerp, Belgium: Springer, 2008, pp. 252–255.
13. Huang N. E., Zheng Sh., Steven R. L. *Proceedings of the Royal Society of London*. 1998, A 454, pp. 903–995.
14. Afonskiy A. A., Dyakonov V. P. *Tsifrovyye analizatory spektra, signalov i logiki* [Digital spectrum, signal and logic analyzers]. Moscow: Solon-Press, 2005, 576 p. [In Russian]
15. Mikhaylov V. G., Zlatousova L. V. *Izmerenie parametrov rechi* [Measurement of parameters of speech]. Moscow: Radio i svyaz', 1987, 168 p. [In Russian]
16. Huang X., Acero A., Hon H.-W. *Spoken Language Processing. Guide to Algorithms and System Development*. New Jersey: Prentice Hall, 2001, 980 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)

Султанов Борис Владимирович

доктор технических наук, профессор,
кафедра информационной безопасности систем
и технологий,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: ibst@pnzgu.ru

Sultanov Boris Vladimirovich

doctor of technical sciences, professor,
sub-department of information security systems
and technologies,
Penza State University
(40 Krasnaya street, Penza, Russia)

Образец цитирования:

Алимурадов, А. К. 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. – DOI 10.21685/2307-5538-2019-4-7.