

## **ВОЗМОЖНОСТИ КОМПЬЮТЕРНОГО АНАЛИЗА АКУСТИЧЕСКИХ ХАРАКТЕРИСТИК РЕЧИ ЧЕЛОВЕКА-ОПЕРАТОРА В УСЛОВИЯХ КОСМИЧЕСКОГО ПОЛЕТА**

С.А. Лебедева, Д.М. Швед, В.И. Гущин

С.А. Лебедева; канд. мед. наук, ст.н.с. Д.М. Швед;  
докт. мед. наук, вед.н.с. В.И. Гущин (ГНЦ РФ–ИМБП РАН)

Работа посвящена обзору современных подходов к оценке психофизиологического состояния человека-оператора с помощью анализа его устной речи. Представлен исторический обзор применения акустического анализа речи в сферах военной авиации и пилотируемых полетов в космос, а также в условиях наземных модельных экспериментов. Рассмотрены различные методы математического анализа речевого сигнала при решении задач определения эмоционального статуса и функциональных состояний человека, а также очистки аудиозаписей от шумов. Проведена оценка применимости данных методов в условиях космических полетов в рамках неинвазивного психофизиологического мониторинга космонавтов.

**Ключевые слова:** анализ речи, речевой сигнал, акустические характеристики, человек-оператор, эмоциональные состояния, функциональное состояние, шумоподавление.

### **The Potentials of Computer Analyzing the Acoustic Characteristics of a Human Speech under Space Flight Conditions. S.A. Lebedeva, D.M. Shved, V.I. Gushchin**

The paper presents the current approaches to the assessment of the psycho-physiological status of a human operator by analyzing his/her speaking and gives a historical review of an application of speech acoustic analysis in the field of military aviation and manned spaceflight as well as under conditions of on-ground model experiments. Various methods of mathematical analysis of a speech signal when determining the human emotional and functional status and techniques of noise suppression in audio recordings are also considered. The applicability of the said methods for non-invasive psycho-physiological monitoring of cosmonauts under space flight conditions was assessed.

**Keywords:** speech analysis, acoustic characteristics, human operator, emotional status, functional status, noise suppression.

Изучение и использование устной речи человека-оператора, находящегося в экстремальных условиях жизнедеятельности, имеет ряд прикладных значений. С точки зрения психологии и медицины, исследование речи может служить целям мониторинга психофизиологического состояния человека. Так, анализ переговоров между космонавтами и Центром управления полетами

(ЦУП), включающий изучение речи каждого члена экипажа, является штатной процедурой и традиционным способом исследования психологического и функционального состояния человека-оператора на борту космического аппарата (КА) [22]. Основным преимуществом используемого специалистами ЦУПа метода является то, что он, в отличие от традиционных физиологических методов мониторинга, таких, как анализ частоты сердечных сокращений (ЧСС), кожно-гальванической реакции (КГР), электромиографии (ЭМГ) и проч., является неинвазивным. Он осуществляется непосредственно в ходе выполнения космонавтом его реальной рабочей деятельности без использования дополнительного оборудования или дополнительных затрат времени. Тем не менее, речь изучается преимущественно с помощью контент-анализа, то есть с позиции ее содержания [7].

Иной способ использования устной речи в условиях, приближенных к космическому полету, реализуется в виде речевых командных систем (РКС), позволяющих пилотам военных самолетов и вертолетов отдавать устные команды бортовому оборудованию летательного аппарата. Применение РКС повышает эффективность и скорость взаимодействия в контуре «человек–машина», но не ставит своей целью мониторинг функционального состояния оператора. Несмотря на широкое применение данного типа интерфейса, разработчики РКС не используют анализ проявлений эмоциональных состояний в речи пилотов, хотя признают, что такая возможность существует [21].

В силу ограничений, связанных с отсутствием достаточно эффективных систем шумоподавления, акустический анализ речи оказался малоприменим в условиях экстремального воздействия. Тем не менее, было показано, что акустические характеристики речи в меньшей степени подвержены сознательному контролю и точнее отражают глубинные особенности функционального состояния человека, чем ее содержание [1, 5, 12, 19]. В рамках наземных исследований данные параметры уже достаточно давно применяются для оценки психофизиологического статуса человека в стрессогенных условиях. Так, в ГНЦ РФ–ИМБП РАН акустические характеристики речи изучались экспертным методом во время переговоров между экипажем и Центром управления в рамках проектов SFINCSS'99, «Марс-500» и в других изоляционных экспериментах [7]. В космических полетах на борту станций «Мир» и МКС в рамках эксперимента «Пилот-М» с помощью системы «Нейролаб-Б», помимо основных психофизиологических параметров, в рамках стресс-диагностики изучалась частота основного тона речи [26]. Непроизвольная речь исследовалась также в условиях 30-суточной гиподинамии в эксперименте «Аккорд», где изучалось влияние гиподинамии на амплитудно-временные и частотные структуры речевых сигналов [14]. Основной целью применения данных методик анализа речи при оценке психофизиологических аспектов взаимодействия оператора со сложной техникой, в том числе и аэрокосмической, является объективизация психологических

и психофизиологических методов исследования. Ограничением проведенных исследований акустических составляющих речи можно считать вынужденный, произвольный характер речеобразования и ограниченность изучаемых параметров. Во многом это можно объяснить недостатками технического обеспечения на момент проведения исследований – помимо использования в качестве изучаемого материала заранее заготовленных, стандартных фраз и звуков, которые человек-оператор повторял через заданные временные интервалы (произвольная речь), для чистоты эксперимента исследователям приходилось создавать специальные условия, чтобы избежать зашумления аудиозаписей и изолировать отдельные параметры речевого сигнала (использование ларингофонов и т.п.). Таким образом, применявшиеся методы являются достаточно инвазивными и не подходят для исследования произвольной речи операторов в неадаптированных акустических условиях.

В настоящее время, в связи с развитием компьютерного анализа речи и совершенствованием способов очистки аудиозаписей от постороннего шума, акустический анализ речи может стать одним из наиболее прогрессивных способов оценки функционального состояния космонавтов в ходе их профессиональной деятельности. В данной статье представлен обзор существующих методов анализа речи и способов шумоподавления, которые проходят апробацию в наземных экспериментах, моделирующих условия космического полета, с целью подготовки к проведению полетных исследований речи космонавтов на российском сегменте Международной космической станции (МКС).

### **Акустические характеристики речи и их интерпретация**

Речевое общение начинается с формирования сообщения в соответствующих отделах центральной нервной системы говорящего. В процессе речеобразования это сообщение преобразуется в акустические речевые колебания, характер которых зависит от физических параметров речевого аппарата, а также особенностей функционального, психологического и эмоционального состояния человека. Наиболее стабильные во времени акустические характеристики речевого сигнала связаны с индивидуальными чертами человека: его антропометрическими (строение речевого аппарата, масса тела, рост) и физиологическими (пол, возраст, функциональное состояние, наличие хронических заболеваний, состояние слухового анализатора) особенностями. Также на речевую экспрессию оказывают влияние тип эмоционально-волевой регуляции и интеллектуальные особенности человека (общительность, стилистика речи, общий когнитивный уровень). На акустических характеристиках речи также отражаются натренированность и сила голоса, актерские навыки, акцент, речевые дефекты [12, 15].

Как правило, для оценки стабильных и переменных характеристик звучащей речи используют такие физические параметры речевого сигнала, как средний спектр, формантные (квазиформантные) области, ширину формант,

локализацию формантных максимумов, суммарную спектральную энергию. Под формантой подразумевается акустическая характеристика звуков речи (прежде всего гласных), связанная с частотой голосового тона и образующая тембр звука [12]. Далее на физические основы речи, своего рода «речевой портрет» говорящего, накладываются проявления функционального и эмоционального состояний человека – в совокупности они представляют собой невербальные характеристики речи. Прикладной и исследовательский интерес представляет собой изучение следующих типов состояний человека-оператора, отражающихся в акустических характеристиках его речи [10]:

1. Эмоциональное возбуждение и напряжение, зачастую являющиеся проявлениями психологического стресса. Для сотрудников служб медицинского и психологического сопровождения полетов обнаружение таких состояний представляет особый интерес, так как является показателем отклонения состояния оператора от нормы эмоционального или психического равновесия.

2. Функциональные состояния утомления, снижения концентрации внимания и т.п., представляющие интерес с позиций разработки средств диагностики функционального состояния космонавтов, летчиков, водителей поездов, операторов военной техники и т.д. Выявление подобных состояний может стать поводом для рекомендаций по изменению режима труда и отдыха, либо приостановки текущей деятельности.

3. Широкий спектр эмоциональных состояний, количество которых может достигать нескольких десятков. Мониторинг эмоционального состояния оператора имеет существенное значение для служб психологической поддержки.

Определение функционального и эмоционального состояний человека-оператора осуществляется при помощи анализа следующих основных акустических параметров его речи [1, 12, 19]:

- а) тембра голоса (или спектра звука – графического отражения частотного (обертонового) состава голоса);
- б) мелодики речи (изменение высоты голоса во времени);
- в) энергетических характеристик (сила голоса и ее изменение);
- г) темпо-ритмических особенностей речи;
- д) атипичных индивидуальных особенностей произношения (смех, покашливание, заикание и т.п.).

Техническое исследование звукового сигнала, включающего в себя речь и фоновые шумы, начинается с приема акустического колебания с помощью микрофона; полученный сигнал подается на усилитель, затем на полосовой фильтр. Далее звуковой сигнал преобразуется в цифровой код с помощью аналого-цифрового преобразователя и подается через интерфейс на шину данных ЭВМ. На следующем этапе посредством специализированного программного обеспечения происходит преобразование оцифрованного сигнала в форму, удобную для изучения и интерпретации речевого сигнала [20].

Важно отметить, что в связи с разнообразием электронной и вычислительной техники, вышеописанная схема преобразования аудиосигнала может варьироваться, и ее особенности могут оказывать существенное влияние на качество звука и результаты анализа речи.

Сообщение, передаваемое с помощью речевого сигнала, – дискретно, т.е. может быть представлено в виде последовательности символов из конечного их числа [3]. При решении задач обработки речи первичным является выделение наиболее существенных признаков – это может быть амплитудный спектр, корреляционная функция и т.д. Выбор базисных функций и последующий анализ параметров этих функций позволяет сформировать численный образ речевого сигнала.

Интерпретация получаемых данных происходит с помощью большой группы математических алгоритмов, позволяющих представить речь как динамическую дискретную структуру. Интерпретация акустических характеристик речи говорящего служит цели получения информации, интересующей исследователей – этим занимаются такие разделы психологии, как психофизика и психоакустика. Математический аппарат, применяемый для анализа акустических характеристик речи, является достаточно трудоемким. В связи с этим применение такого рода алгоритмов на практике стало возможным лишь в процессе развития вычислительной техники.

## **Разработка алгоритмов акустического анализа речи**

Одним из первых исследуемых и до сих пор наиболее востребованных параметров речевого сигнала является частота основного тона речи (ЧОТ). Ее вычисление позволило решать многие практические задачи развивающихся речевых технологий, такие, как идентификация пола говорящего, определение его или ее эмоционального состояния, расшифровка сказанного, а также выявление функционального состояния голосового аппарата [33]. Таким образом, начиная с 1930-х годов, определение ЧОТ стало основой для развития различных подходов к акустическому анализу речи [24].

Математические алгоритмы вычисления ЧОТ разделялись по особенностям работы в «трудных» условиях: с искаженным акустическим сигналом, в зашумленных или необычных условиях звукозаписи. К 1970-м годам уже были известны основные подходы к решению данных проблем: алгоритмы временной и частотной областей вычисления, а также их комбинации [17].

Ранние подходы, основанные на методах автокорреляционной функции, разностной функции и пересечения с нулем, показали себя как достаточно простые в использовании, но недостаточно надежные для анализа таких сложных звуковых сигналов, как человеческая речь. Для анализа речи автокорреляционный метод прошел ряд модификаций для повышения точности – этот алгоритм стал известен как YIN (Инь, балансирующий слабые места классического автокорреляционного метода с помощью понижения

чувствительности к амплитудным модуляциям, параболической интерполяции минимума и предотвращением резких скачков значений). Долгое время он считался наиболее надежным подходом к изучению акустических особенностей речевого сигнала [23].

В настоящее время существует несколько способов компьютерной обработки звучащей речи – это методы, позволяющие различать не только речевые образцы, но и различные эмоциональные состояния дикторов. К ним относятся спектрально-временные, кепстральные, амплитудно-частотные, на основе нелинейной динамики, вейвлет-преобразования и методы на основе опорных векторов.

Спектрально-временные признаки отражают своеобразие формы временного ряда, спектра голоса у разных людей и специфику фильтрующих функций их голосовых трактов. Метод спектрально-временных признаков включает анализ трех компонентов речевого сигнала: периодических (тональных) участков звуковой волны, непериодических участков звуковой волны (шумовых, взрывных) и участков, не содержащих речевых пауз. Изменение этих характеристик во времени отражает особенности речевого потока, связанные с динамикой перестройки артикуляционных органов речи говорящего [4]. В группе спектрально-временных признаков были выделены параметры, инвариантные к действию повышенного уровня сигнала. Являясь интегральными характеристиками речевого потока, они описывают статистические характеристики речевого сигнала и основного тона, особенности спектральной структуры. Однако Н.Н. Филатова и К.В. Сидоров показали, что, основываясь только на простых спектральных характеристиках звукового сигнала, невозможно правильно распознавать и идентифицировать различные эмоции [18]. Тем не менее, спектральные и кепстральные признаки широко используют для быстрого и эффективного подавления постороннего шума на аудиозаписях для дальнейшей работы с чистым речевым сигналом [27].

Большинство исследований в качестве наиболее информативных акустических коррелятов эмоциональных и функциональных состояний рассматривают динамику частотных, временных и мощностных характеристик голосового сигнала. Так, Г.А. Адашинская указывает на взаимосвязь акустических параметров речи, эмоциональных и функциональных состояний, обусловленную индивидуальными особенностями говорящих, которая выражается в разнонаправленности изменений ряда временных и мощностных параметров речи [2].

Метод кепстральных коэффициентов (Mel Frequency Cepstral Coefficient, MFCC) является одним из наиболее распространенных в системах распознавания речи и вычисления ее признаков и характеристик, поскольку он учитывает психоакустические принципы восприятия речи и использует мел-шкалу, связанную с критическими полосами слуха. Вычисляются мел-частотные коэффициенты, коэффициенты линейного предсказания с коррекцией на не-

равномерную чувствительность человеческого слуха, коэффициенты мощности частоты, кепстра и спектра. Группа признаков эмоционально окрашенной речи по кепстральным коэффициентам позволяет отделить сигнал эмоционального возбуждения от характерных сигналов речевого тракта [4]. Исследования, однако, показывают, что достоверность определения базовых эмоций методом мел-частотных коэффициентов может быть недостаточно высокой – от 64,47 % до 80 % [29, 31].

Для группы признаков нелинейной динамики речевой сигнал рассматривается как скалярная величина, наблюдаемая в системе голосового тракта человека. В настоящее время методы нелинейной динамики и нелинейной авторегрессии позволяют восстанавливать фазовый портрет аттрактора по временному ряду или по одной его координате. Экспериментально подтверждено, что выявленные отличия в форме аттракторов можно использовать для формирования диагностических правил и признаков, позволяющих распознать и правильно идентифицировать различные эмоции в речевом сигнале. Н.Н. Филатова и К.В. Сидоров предлагают модель интерпретации знака эмоции по правилу объединения нечетких множеств, характеризующих значения усредненного максимального вектора реконструкции аттрактора по четырем квадрантам [18]. А.Ф. Хроматиди в своем исследовании с помощью метода нелинейной динамики определила размерность фазового пространства и показатели Ляпунова для отдельных количественных характеристик речи, которые позволили классифицировать не только эмоции дикторов, но и психофизиологическое состояние железнодорожных диспетчеров. Автор отмечает повышение ЧОТ, а также формант (на 50–150 %) всех ударных гласных в речи операторов, подверженных стрессовому воздействию [20]. В работе китайских математиков Ying Sun, Xue-Ying Zhang и др. было проведено сравнение метода на основе нелинейной динамики (Nonlinear Dynamics, NLD) с реконструкцией фазового пространства с методом мел-частотных коэффициентов (MFCC). Для двух исследованных международных речевых баз результаты получились схожими: усовершенствованный авторами метод NLD давал 61,9–82,3 % правильных оценок шести базовых эмоций, тогда как для MFCC достоверность составляла 36,5–77 % [34].

Еще один способ классификации эмоционально окрашенной речи – метод опорных векторов (Support Vector Machines, SVM). Так, находящаяся в открытом доступе библиотека libSVM использует метод опорных векторов для построения классификатора, позволяющего выделить свыше 6500 характеристик звукового сигнала. На данный момент существуют модификации метода опорных векторов для идентификации диктора и особенностей изменения различных его эмоциональных состояний. А.Г. Давыдов с соавторами показали, что, добавив определенные информативные признаки в алгоритм на основе опорных векторов с квадратичной разделяющей функцией, можно добиться 96 % точности классификации эмоциональных состояний говорящего [8]. Взвешенный по значимости метод на основе опорных векторов

(Importance-Weighted Support Vector Machines, IW-SVM) использовался исследователями Huang, Ao и Zhang для повышения устойчивости программного алгоритма к шумам, в то время как сам вес значимости применялся для компенсации ковариатного сдвига между набором тестовых данных и данными, на которых проводилось обучение. Дополненный метод опорных векторов показал высокую устойчивость к гауссовскому белому шуму при распознавании эмоций речи по сравнению с обычными программами анализа речи [25]. Объединив метод опорных векторов с пакетом программ вейвлет-преобразования, авторы получили высокоадаптивную программу для анализа нестационарных сигналов речи в разнообразных условиях звукозаписи.

Вейвлет-анализ считается одним из наиболее передовых методов работы с речевыми сигналами. Вейвлет-спектрограммы отчетливо выделяют такие особенности сигналов, как разрывы, изменение знаков первой и второй производных, изменение частоты нестационарного сигнала и т.д., то есть те особенности сигналов, которые плохо выделяются с помощью преобразования Фурье. Большие уровни при вейвлет-анализе характерны для коэффициентов тех вейвлетов, которые смещены вблизи той или иной локальной особенности функции [9]. Таким образом, вейвлет-анализ открывает принципиально новые возможности в анализе нестационарных сигналов, к которым относятся и речевые. Было показано, что программные алгоритмы на основе многоуровневого вейвлет-анализа речевых сигналов, помимо эффективного шумоподавления, позволяют идентифицировать говорящего и проводить оценку его эмоционального состояния [6].

А.М. Сорока с соавторами создали методику объединенных векторов на основе вейвлет-преобразования с адаптивными базовыми функциями, что позволило параллельно локализовать и обобщить физические признаки речевого сигнала. В сравнении с другими методами на основе преобразования Фурье (в частности, мел-частотных кепстральных коэффициентов, MFCC), данная разработка повысила точность распознавания эмоций на 3,2 %, достигнув точности 81–91,5 % для семи различных эмоциональных состояний [16]. Похожее сравнение провела Kunxia Wang с соавторами: используя статические и динамические коэффициенты программного пакета на основе вейвлетов, они проанализировали две базы аудиозаписей на разных языках. Как и в предыдущих исследованиях, применение вейвлет-преобразования показало большую эффективность при распознавании эмоций по сравнению с анализом на основе MFCC на 14,9 % и 4,3 % для международных речевых баз EMODB и EESDB соответственно [30].

Таким образом, уже более десяти лет алгоритмы акустического анализа речи являются достаточно совершенными, чтобы позволить исследователям качественно анализировать особенности эмоционального и функционального состояния человека по характеристикам его устной речи. Практически все наиболее часто используемые способы расчета акустических параметров речевого сигнала реализованы в известных и общедоступных

математических компьютерных пакетах обработки сигналов, таких, как MATLAB, SPL и IPPS. Компьютерные системы анализа речи, такие, как PRAAT, SpeechAnalyzer, WinCecil, MacSpeechLab, PhonologyAssistant и др., в большинстве своем находятся в открытом доступе и применяются для фонетического и фонологического анализа звукового материала. Эти программы визуализируют звуковой сигнал, представляя его в виде спектрограммы, графического изображения формы волны, диаграммы изменения частоты тона, диаграмм интенсивности, длительности и формантной составляющей. Развитием речевых технологий, а также фундаментальными и клиническими исследованиями речевых функций занимаются ведущие мировые научно-образовательные учреждения: Гарвардский университет, Массачусетский технологический институт, Бостонский университет и проч. В России технологии автоматической обработки речи на данный момент разрабатываются, в частности, на профильных кафедрах МФТИ (с 2006 года) и ТМО (с 2011 года).

### **Алгоритмы шумоподавления**

Существенной проблемой, ограничивающей применение автоматизированных методов анализа частотных характеристик речи в космической медицине, является не до конца решенный вопрос подавления технологических шумов внутренней среды пилотируемых космических аппаратов и орбитальных станций. Решение этой проблемы во многом зависит от успешности создания комплекса инженерно-технических решений, направленных на снижение уровня шумов различного характера (фоновых, периодических, взрывных). Альтернативный путь решения данной проблемы заключается в разработке и оптимизации программных средств шумоподавления, применяемых при звукозаписи и анализе аудиосигнала. Важными представляются оба подхода, поскольку часть шумов ввиду своих акустических и частотных характеристик может быть устранена лишь за счет внесения изменений в их источники либо среду распространения, однако подобные изменения конфигурации КА и расположенных на их борту технических средств не всегда представляются возможными, что обуславливает значимость программных методов шумоподавления при работе с полученным аудиосигналом [3, 27].

Следует отметить, что данная проблема касается не только анализа речи в особых условиях жизнедеятельности. Речевые сигналы, зарегистрированные в естественной обстановке, зачастую имеют недостаточно высокое качество ввиду воздействия окружающих шумов, посторонней речи и прочих акустических помех. Это значительно усложняет процесс автоматической обработки речи в задачах ее распознавания, идентификации диктора, кодирования и т.д. Для достижения эффективности такой обработки зашумленный речевой сигнал необходимо очищать от посторонних звуков алгоритмами шумоподавления, разработка и совершенствование которых является одной из наиболее актуальных задач в исследованиях акустических парамет-

ров речи [31, 32]. При этом не все разрабатываемые и применяемые алгоритмы устранения шума подходят для исследовательских задач, поскольку основной задачей шумоподавления является увеличение разборчивости речи в условиях шума повышенной интенсивности, а не сохранение ее спектрального профиля [3, 11, 21, 28].

А.А. Борискевич и С.И. Киркоров предлагают выделить следующие типы алгоритмического способа шумоподавления и фильтрации аддитивного шума [11]:

- методы цифровой обработки зашумленных речевых сигналов;
- методы адаптивной компенсации помех (например, с помощью выделения опорных сигналов, коррелированных и некоррелированных с шумовым отрезком аудиозаписи);
- методы на основе статистических моделей речевых сигналов во временной области (такие, как авторегрессионная модель, линейный фильтр Калмана, рекуррентные алгоритмы оценки параметров речевого сигнала);
- методы на основе статистических моделей речевых сигналов в частотной области (такие, как моделирование шума или полезного речевого сигнала с помощью скрытой Марковской цепи, а также оценка минимальной среднеквадратической ошибки, что особенно актуально при использовании одного микрофона);
- методы, основанные на использовании спектральных характеристик шума (классическое вычитание амплитудных спектров и фильтрация Винера);
- методы, основанные на использовании моделей глубоких нейронных сетей;
- методы, основанные на моделях восприятия речи человеком.

Многообразие представленных методов отражает значимость проблемы активного шумоподавления и шумоочистки в анализе речевых сигналов. На данный момент не существует надежного и универсального алгоритма, адаптивно подстраивающегося к различным условиям звукозаписи речи. Два различных подхода к обработке речевого сигнала ведут к двум разным приоритетам в его коррекции: улучшение разборчивости речи (с возможными значительными искажениями акустических характеристик голоса), либо улучшение качества речевого сигнала с сохранением его натурального обертонового состава (с возможным ухудшением разборчивости речи) [32]. Отчасти недостаток в адаптивности вышеописанных методов можно нивелировать, используя информацию об условиях звукозаписи, характере шумового воздействия и содержании произносимой речи, а также «эталонные» записи голоса диктора [13].

По нашему мнению, представленные методы шумоподавления должны дополнять друг друга и служить взаимной оценке эффективности, чтобы обеспечить выбор необходимого для определенных условий звукозаписи набора методов.

## Выводы

Распознавание различных психологических (эмоциональных) и функциональных состояний человека по его естественной речи является весьма сложной задачей как с точки зрения математической формализации, так и в плане поиска способов четкой конкретизации этих состояний. В настоящее время отсутствует универсальная теоретическая модель описания речевых образцов в условиях проявления разных видов эмоций. Вместе с тем, остается не до конца разработанной тема комплексной оценки функциональных состояний оператора непосредственно в ходе выполнения профессиональной деятельности. Таким образом, при диагностике психофизиологического состояния человека необходимо использовать комплексную систему оценки множества параметров организма, в которой особую роль играет речевое поведение. Являясь индикатором индивидуальной адаптации к внешним воздействиям, речь позволяет оценить успешность преодоления кризисных ситуаций и психофизиологическую «цену» такого преодоления.

На сегодняшний день существует несколько подходов к анализу эмоциональной составляющей речи: спектрально-временные, кепстральные и амплитудно-частотные алгоритмы, признаки на основе нелинейной динамики, вейвлет-преобразования и методы на основе опорных векторов. Данные методы используются в зависимости от исследовательских целей, совершенствуются дополнительными адаптивными функциями и применяются совместно для выявления метода анализа речи, наиболее адекватного с позиций задач и условий проводимой работы. Ключевая задача – определить, что для исследования является более значимым: содержательная или частотная целостность речевого сигнала.

Долгое время в космической медицине и при разработке бортового оборудования последняя задача практически не рассматривалась в связи с трудностью извлечения качественного сигнала. Развитие систем шумоподавления и алгоритмов анализа эмоциональных состояний по характеристикам речевого сигнала сделало возможным решение данной задачи. По нашему мнению, для достоверного определения эмоциональных и психофизиологических состояний человека-оператора наиболее эффективными могут оказаться решения на основе опорных векторов и вейвлет-преобразования. При работе со стационарными, аддитивными и мультипликативными шумами, являющимися неотъемлемой частью рабочего пространства МКС, представляется рациональным применение систем шумоподавления на основе глубоких нейронных сетей, обученных на специфических для космических полетов аудиозаписях.

Перспективы практического применения результатов акустического анализа речи могут заключаться не только в создании более совершенных методов мониторинга функционального состояния человека-оператора, но и во включении их как элемента автономного психофизиологического сопровож-

дения экипажей при разработке систем бортового искусственного интеллекта для длительных космических полетов.

*Работа выполнена в рамках темы № 63.2 программы фундаментальных научных исследований РАН.*

#### ЛИТЕРАТУРА

- [1] Агарков Ю.Н., Козина М.Д., Косырев В.Н. Об использовании акустических показателей речи в диагностике психоэмоциональных состояний // Вестник ТГУ. – 2002. – № 2.
- [2] Адашинская Г.А. Акустические корреляты индивидуальных особенностей функциональных и эмоциональных состояний / Г.А. Адашинская, Д.Н. Чернов // Авиакосмическая и экологическая медицина. – 2007. – Т. 41. – № 2. – С. 3–13.
- [3] Ахмад Х.М., Жирков В.Ф. Введение в цифровую обработку речевых сигналов: учеб. пособие // Владим. гос. ун-т. – Владимир: Изд-во Владим. гос. ун-та, 2007. – 192 с.
- [4] Вартанов А.В. Антропоморфный метод распознавания эмоций в звучащей речи // Национальный психологический журнал. – 2013. – № 2(10). – С. 69–79.
- [5] Воронцова Ю.А., Хорошко Е.Ю. Невербальные знаки как неотъемлемая часть речевого общения // Вестник БелЮИ МВД России. – 2006. – № 2.
- [6] Горшков Ю.Г. Обработка речевых сигналов на основе вейвлетов // Т-Comm. – 2015. – № 2.
- [7] Гуцин В.И., Виноходова А.Г., Комиссарова Д.В., Белаковский М.С., Орлов О.И. Эксперименты с изоляцией: прошлое, настоящее, будущее // Авиационная и экологическая медицина. – 2018. – Т. 52. – № 4. – С. 5.
- [8] Давыдов А.Г., Киселев В.В., Кочетков Д.С. Классификация эмоционального состояния диктора по голосу: проблемы и решения // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». – Вып. 11 (18): В 2 т. – Т. 1: Основная программа конференции. – М.: Изд-во РГГУ, 2012.
- [9] Зубаков А.П. Фурье и вейвлет-преобразования в проблеме распознавания речи // Вестник Тамбовского университета. – Серия: Естественные и технические науки. – 2010. – № 6.
- [10] Картавенко М.В. Об использовании акустических характеристик речи для диагностики психических состояний человека // Известия ЮФУ. Технические науки. 2005. № 5.
- [11] Киркоров С.И., Борискевич А.А. Методы и средства восстановления разборчивости зашумленной речи // Курс: цифровая обработка речи и изображения, на основе лекции Борискевича А.А. – Минск, 2010.
- [12] Морозов В.П. Психологический портрет человека по невербальным особенностям его речи // Психологический журнал. – 2001. – Т. 22. – № 6. – С. 48.
- [13] Мясников В.И., Замалетдинов И.С. Психическое состояние и групповое взаимодействие космонавтов в полете // Космическая биология и медицина. – Т. 3. Человек в космическом полете. – Кн. 2. – М.: Наука, 1997.
- [14] Никонов А.В. Психологические проблемы акустической диагностики функциональных состояний оператора // Психологические проблемы деятельности в особых условиях. – М.: Наука, 1985. – С. 136–153.

- [15] Потапова Р.К., Потапов В.В. Перспективы развития концепции речевой портрет говорящего // Сб. трудов XVII Международной научной конференции «Информатизация и информационная безопасность правоохранительных органов». – Москва, 2008. – С. 381–382.
- [16] Сорока А.М.; Семенченко А.В.; Хейдоров И.Э. Распознавание эмоций в речи на основе вейвлетного анализа с адаптируемыми базовыми функциями // Международный конгресс по информатике: Информационные системы и технологии. Материалы Международного научного конгресса, Республика Беларусь, Минск, 2016. – С. 219.
- [17] Тиунов С.Д. Модель, численные методы и комплекс программ для акустического анализа голоса в задачах диагностики голосовых расстройств: Дис. канд. тех. наук. – Томск, 2014.
- [18] Филатова Н.Н., Сидоров К.В. Модель интерпретации знака эмоций по естественной речи / Н.Н. Филатова, К.В. Сидоров // Известия ЮФУ. Технические науки. Тематический выпуск. – 2012. – Т. 134. – № 9 – С. 39–45.
- [19] Фролов М.В., Милованова Г.Б. Особенности контроля состояния человека-оператора по показателям основного тона и спектра его речи // Физиология человека. – 2009. – Т. 35. – № 2. – С. 136–138.
- [20] Хроматиди А.Ф. Исследование психофизиологического состояния человека на основе эмоциональных признаков речи: Дис. ... канд. техн. наук: 05.11.17, 05.13.01. – Таганрог, 2005. – 154 с. РГБ ОД, 61:05-5/3659.
- [21] Чепурных И.В., Чепурных С.А. Системы бортового оборудования самолетов и вертолетов. Топливная система и кабинное оборудование: учеб. пособие. – Комсомольск-на-Амуре: ФГБОУ ВПО «КнАГТУ», 2015. – 169 с.
- [22] Юсупова А.К., Гущин В.И., Ушаков И.Б. Коммуникации космических экипажей в реальных и моделируемых космических полетах: монография. – М.: ГИЦ РФ–ИМБП РАН, 2011. – 199 с.
- [23] de Cheveign'e A., Kawahara H. YIN, a fundamental frequency estimator for speech and music // J. Acoust. Soc. Am, 2002.
- [24] Flanagan James L., Allen Jont B., Hasegawa-Johnson Mark A. Speech Analysis Synthesis and Perception. Third ed. изд. Springer-Verlag, 2008. – pp. 1–486.
- [25] Huang Y., Ao, W. & Zhang, G. Novel Sub-band Spectral Centroid Weighted Wavelet Packet Features with Importance-Weighted Support Vector Machines for Robust Speech Emotion Recognition // Wireless Pers Commun 95, 2223–2238 (2017).
- [26] Johannes B., Salnitski V., Soll H., Rauch M., Hoermann H.-J. De-individualized psychophysiological strain assessment during a flight simulation test – validation of a space methodology // Acta astronautica. – 2008. – Vol. 63. – pp. 791–799.
- [27] Jouni Pohjalainen, Fabien Ringeval, Zixing Zhang and Björn Schuller. Spectral and Cepstral Audio Noise Reduction Techniques in Speech Emotion Recognition // In Proceedings of the 24<sup>th</sup> ACM International Conference on Multimedia, MM 2016, pp. 670–674, Amsterdam, The Netherlands, October, 2016.
- [28] Kim Gibak, Philipos Loizou. Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms // The Journal of the Acoustical Society of America, 2011. 130. 1581-96. 10.1121/1.3619790.
- [29] Koo H., Jeong S., Yoon S. and Kim W. Development of Speech Emotion Recognition Algorithm using MFCC and Prosody // 2020 International Conference on Electronics, Information, and Communication (ICEIC), Barcelona, Spain, 2020, pp. 1–4.

- [30] Kunxia Wang, Guoxin Su, Li Liu, Shu Wang. Wavelet packet analysis for speaker-independent emotion recognition // *Neurocomputing*, 2020.
- [31] Likitha M.S., Gupta S.R.R., Hasitha K., Raju A.U. Speech based human emotion recognition using MFCC // 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2017, pp. 2257–2260.
- [32] Loizou Philipos, Kim Gibak. Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions, 2011 // *Audio, Speech, and Language Processing*, IEEE Transactions on. 19. 47–56. 10.1109/TASL.2010.2045180.
- [33] Speyer R. Effects of voice therapy: a systematic review // *J Voice*. 2008. Sep. T. 22. – pp. 565–580.
- [34] Ying Sun, Xue-Ying Zhang, Jiang-He Ma, Chun-Xiao Song, Hui-Fen. Nonlinear Dynamic Feature Extraction Based on Phase Space Reconstruction for the Classification of Speech and Emotion // *Mathematical Problems in Engineering*, 2020.

## REFERENCES

- [1] Agarkov Yu.N., Kozina M.D., Kosyrev V.N. The use of acoustic speech indicators in the diagnosis of psychoemotional states // *Tomsk State University Journal*. – 2002. – No 2.
- [2] G.A. Adashinskaya. Acoustic correlates of individual characteristics of functional and emotional states / G.A. Adashinskaya, D.N. Chernov // *Aerospace and Environmental Medicine* – 2007. – V. 41. – No 2. – pp. 3–13.
- [3] Ahmad H.M., Zhirkov V.F. Introduction to digital voice processing: learning guide // Vladimir State University. – Vladimir: Publishing House of Vladimir State University, 2007. – p. 192.
- [4] Vartanov A.V. Anthropomorphic method of emotion recognition in sounding speech // *National Psychological Journal [Natsionalny Psikhologichesky Zhurnal]* – 2013. – No 2(10). – pp. 69–79.
- [5] Vorontsova Yu.A., Khoroshko E.Yu. Non-verbal signs as an essential part of verbal communication // *Vestnik of Belgorod Law Institute of Ministry of Interior of Russia [Vestnik Bel YuI MVD RF]*. – 2006. – No 2.
- [6] Gorshkov Yu.G. Wavelet-based speech processing // *T-Comm*. – 2015. – No 2.
- [7] Gushchin V.I., Vinokhodova A.G., Komissarova D.V., Belakovsky M.S., Orlov O.I. Experiments with isolation: the past, present and future // *Aerospace and Environmental Medicine*. – 2018. – V. 52. – No 4. – p. 5.
- [8] Davydov A.G., Kiselev V.V., Kochetkov D.S. Voice classification of a speaker's emotional state: problems and solutions // *Computer Linguistics and Intelligent Technologies: adapted from the materials of the annual International Conference "Dialogue"*. – Issue. 11(18): In 2 vol. – V. 1: Conference main program. – Moscow: Publishing House of RSU for Humanities (RGGU), 2012.
- [9] Zubakov A.P. Fourier and wavelet transforms in the problem of speech recognition // *Journal "Tambov University Review. Series Humanities"*. – 2010. – No 6.
- [10] Kartavenko M.V. The use of acoustic characteristics of speech for the diagnosis of mental status // *"Izvestiya SFedU. Engineering Sciences"*. 2005. No 5.
- [11] Kirkorov S.I., Boriskevich A.A. Methods and means of restoring intelligibility of noisy speech // *Course: Digital processing of speech and images, based on a lecture by A. Boriskevich* – Minsk, 2010.

- [12] Morozov V.P. Psychological profile of a person based on non-verbal features of his speech // *Psychological Journal*. – 2001. – V. 22. – No 6. – p. 48.
- [13] Myasnikov V.I., Zamaletdinov I.S. Mental state and group interaction of cosmonauts in flight // *Space Biomedicine*. – V. 3. Man in Space Flight. – Book 2. – Moscow: Nauka, 1997.
- [14] Nikonov A.V. Psychological problems of acoustic diagnostics of an operator's functional status // *Psychological problems of working under special conditions*. – Moscow: Nauka, 1985. – pp. 136–153.
- [15] Potapova R.K., Potapov V.V. Prospects for the concept of a speaker's speech portrait // *Proceedings of the XVII International Scientific Conference "Informatization and Information Security of Law Enforcement Agencies"*. – Moscow, 2008. – pp. 381–382.
- [16] Forty A.M., Semenchenko A.V., Kheydorov I.E. Recognition of emotions in speech based on wavelet analysis with adaptable basic function // *International Congress on Informatics: Information Systems and Technologies. Proceedings of the International scientific congress, Republic of Belarus, Minsk, 2016*. – p. 219.
- [17] Tiunov S.D. Model, numerical methods and software package for acoustic voice analysis in the diagnosis of voice disorders. Dis. ...PhD in Engineering. – Tomsk, 2014.
- [18] Filatova N.N., Sidorov K.V. The model for the interpretation of sign of emotions on natural speech / N.N. Filatova, K.V. Sidorov // "Izvestiya SFedU. Engineering Sciences". – 2012. – V. 134. – No 9. – pp. 39–45.
- [19] Frolov M.V., Milovanova G.B. Features of monitoring the state of a human operator in terms of his fundamental frequency speech spectrum // *Human physiology*. – 2009. – V. 35. – No 2. – pp. 136–138.
- [20] Chromatidi A.F. The study of the psychophysiological state of a person based on the emotional signs of speech: Dis. ...PhD in Engineering: 05.11.17, 05.13.01. – Taganrog, 2005. – p. 154. Russian State Library (RGB OD), 61:05-5/3659.
- [21] Chepurnykh I.V., Chepurnykh S.A. Systems of on-board equipment for aircraft and helicopters. Fuel system and cab equipment: learning guide. – Komsomolsk-on-Amur: FSBE "Komsomolsky-On-Amur State University" (FGBO VPO "KiAGTU"), 2015. – p. 169.
- [22] Yusupova A.K., Gushchin V.I., Ushakov I.B. Communications of space crews in real and simulated space flights: a monograph. – Moscow: SRC RF IMBP RAS, 2011. – p. 199.
- [23] de Cheveign'e A., Kawahara H. YIN, a fundamental frequency estimator for speech and music // *J. Acoust. Soc. Am*, 2002.
- [24] Flanagan James L., Allen Jont B., Hasegawa-Johnson Mark A. *Speech Analysis Synthesis and Perception*. Third ed. Publ. Springer-Verlag, 2008. – pp. 1–486.
- [25] Huang, Y., Ao, W. & Zhang, G. Novel Sub-band Spectral Centroid Weighted Wavelet Packet Features with Importance-Weighted Support Vector Machines for Robust Speech Emotion Recognition // *Wireless Pers Commun* 95, 2223–2238 (2017).
- [26] Johannes B., Salnitski V., Soll H., Rauch M., Hoermann H.-J. De-individualized psychophysiological strain assessment during a flight simulation test – Validation of a space methodology // *Acta astronautica*. – 2008. – Vol. 63. – pp. 791–799.
- [27] Jouni Pohjalainen, Fabien Ringeval, Zixing Zhang, and Björn Schuller. Spectral and Cepstral Audio Noise Reduction Techniques in Speech Emotion Recognition // *In Proceedings of the 24<sup>th</sup> ACM International Conference on Multimedia, MM 2016*, pp. 670–674, Amsterdam, The Netherlands, October, 2016.
- [28] Kim Gibak, Philipos Loizou. Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms // *The Journal of the Acoustical Society of America*, 2011. 130. 1581-96. 10.1121/1.3619790.

- [29] Koo H., Jeong S., Yoon S. and Kim W. Development of Speech Emotion Recognition Algorithm using MFCC and Prosody // 2020 International Conference on Electronics, Information, and Communication (ICEIC), Barcelona, Spain, 2020, pp. 1–4.
- [30] Kunxia Wang, Guoxin Su, Li Liu, Shu Wang. Wavelet packet analysis for speaker-independent emotion recognition // Neurocomputing, 2020.
- [31] Likitha M.S., Gupta S.R.R., Hasitha K., Raju A.U. Speech based human emotion recognition using MFCC // 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2017, pp. 2257–2260.
- [32] Loizou Philipos, Kim Gibak. Reasons why Current Speech-Enhancement Algorithms do not Improve Speech Intelligibility and Suggested Solutions, 2011 // Audio, Speech, and Language Processing, IEEE Transactions on. 19. 47–56. 10.1109/TASL.2010.2045180.
- [33] Speyer R. Effects of voice therapy: a systematic review // J Voice. 2008. Sep. T. 22. – pp. 565–580.
- [34] Ying Sun, Xue-Ying Zhang, Jiang-He Ma, Chun-Xiao Song, Hui-Fen. Nonlinear Dynamic Feature Extraction Based on Phase Space Reconstruction for the Classification of Speech and Emotion // Mathematical Problems in Engineering, 2020.