

УДК 159.9:629.7, 629.7.069:159.9

**АНАЛИЗ ФУНКЦИОНИРОВАНИЯ  
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ  
ГОЛОСОВОГО АССИСТЕНТА КОСМОНАВТА  
В УСЛОВИЯХ МОДЕЛИРУЕМОЙ  
ШУМОВОЙ ОБСТАНОВКИ РС МКС**

Д.Н. Рулев, С.И. Ашманов, В.С. Крайнов, Е.В. Попова, Л.А. Умнова

Канд. физ.-мат. наук Д.Н. Рулев (ПАО «РКК «Энергия» им. С.П. Королёва») С.И. Ашманов; В.С. Крайнов (ООО «Лаборатория Наносемантика») Канд. пед. наук Е.В. Попова; Л.А. Умнова (ФГБУ «НИИ ЦПК имени Ю.А. Гагарина»)

В статье представлен анализ функционирования разрабатываемого программного обеспечения голосового ассистента космонавта в условиях моделируемой шумовой обстановки РС МКС на тренажерах модулей РС МКС. Анализ полученных аудиозаписей выполнялся по акустическим параметрам (чистота и ясность звука; искажения и артефакты; громкость и динамический диапазон; отношение сигнал/шум; позиционирование и направленность; субъективное восприятие) и по параметрам автоматического распознавания речи. **Ключевые слова:** голосовой ассистент, условия космического полета, частотно-акустические характеристики речи, параметры громкости звука, распознавание речи

**Analysis of the Software Operation of a Voice Assistant for  
Cosmonauts under Simulated Noise Conditions on the ISS RS.**

**D.N. Rulev, S.I. Ashmanov, V.S. Krainov, E.V. Popova,  
L.A. Umnova**

The paper gives an analysis of the software operation being developed for a cosmonaut's voice assistant under conditions of a simulated noise environment on the ISS RS using the simulators of the modules of the ISS RS. Audio recordings obtained were analyzed according to acoustic parameters (sound purity and clarity; distortions and artifacts; volume and dynamic range; signal-to-noise ratio; positioning and directionality; subjective perception) and automatic speech recognition parameters.

**Keywords:** voice assistant, spaceflight environment, frequency-acoustic characteristics of speech, sound volume parameters, speech recognition

В рамках подготовки к реализации целевой работы (космического эксперимента) «Ассистент», целью которой является отработка технологии создания и сопровождения в полете виртуального ассистента космонавта и выполнение экспериментальных исследований влияния условий космического полета на эффективность технологий распознавания и идентификации речи космонавта [2, 4], получены экспериментальные данные для анализа

функционирования разрабатываемого программного обеспечения (ПО) голосового ассистента космонавта [1] в условиях моделируемой шумовой обстановки РС МКС в ФГБУ «НИИ ЦПК имени Ю.А. Гагарина» на тренажерах служебного модуля (СМ) «Звезда», многоцелевого лабораторного модуля (МЛМ) «Наука» и малого исследовательского модуля № 1 (МИМ1) «Рассвет» [3].

В испытаниях использовался ноутбук MacBook Air M2 с установленным ПО голосового ассистента. Использовался микрофон, встроенный в ноутбук. В тренажерах модулей СМ, МЛМ и МИМ1 выполнялось включение бортовых приборов для создания (моделирования) возможной шумовой обстановки, на фоне которой выполнялись сеансы разговора оператора с голосовым ассистентом (с получением аудиозаписи разговора на диск ноутбука).

Сеансы выполнялись при нахождении оператора с ноутбуком в не менее трех местах каждого из модулей.

Шумовая обстановка включала следующие элементы:

- постоянная работа системы вентиляции;
- периодические включения аварийного сигнала;
- периодический фоновый сторонний разговор (на удалении не менее двух метров от оператора в этом же или в соседнем модуле).

Оператор располагался непосредственно рядом с ноутбуком (расстояние до микрофона менее полуметра) и выполнял разговор с вариациями уровня громкости голоса, включающими субъективно оцениваемое незначительное понижение и повышение громкости голоса от нормального среднего.

Местоположения оператора выбирались исходя из различных условий расположения относительно работающего вентилятора, включая нахождение оператора в непосредственной близости к вентилятору (менее полуметра) и при различной ориентации оператора с находящимся перед ним ноутбуком (в руках или на коленях) относительно вентилятора – спиной или лицом.

Решалась задача оценки аудиозаписей, полученных в различных позициях в модулях РС МКС, с различной удаленностью оператора от микрофона, громкостью речи и позицией относительно источников шума.

Анализ аудиозаписей проводился по акустическим параметрам (чистота и ясность звука; искажения и артефакты; громкость и динамический диапазон; отношение сигнал/шум; позиционирование и направленность; субъективное восприятие) и по параметрам автоматического распознавания речи.

Отметим, что субъективная оценка аудиозаписей с точки зрения восприятия речевой информации слушателем включала оценку разборчивости речи (понимается как относительное количество правильно принятых элементов речи (звук, слог, слово, фраза), выраженное в процентах от общего числа переданных элементов), качества речи (параметр, характеризующий субъективную оценку звучания речи в испытываемой системе передачи речи, выраженную в баллах по пятибалльной шкале или в процентах предпочтения

при сравнении с эталонным трактом) и в целом оценку интегрального качества как показателя, характеризующего общее впечатление слушателя от принимаемой речи.

Например, указанные оценки качества передачи речи через каналы связи регламентируются ГОСТами [5, 6], при этом в настоящей работе метод измерений/испытаний реализовывался одним диктором и одним слушателем (аудитором), а в качестве речевого теста использовался текст по космической тематике без специальных профессиональных терминов.

Оценка аудиозаписей по параметрам автоматического распознавания речи опиралась на использование такого показателя качества системы распознавания слитной речи как показатель спутывания WER (Word Error Rate – процент ошибок на уровне слов: среднее количество удалений, вставок и замен слов в общем количестве слов в рассмотренных тестовых данных), регламентируемого (с учетом контекста задач, решаемых голосовым ассистентом космонавта, используемая голосовым ассистентом система распознавания слитной речи обоснованно может рассматриваться как система распознавания голосовых команд управления) [7].

### Анализ полученных аудиозаписей

Получено 27 аудиозаписей (образцов), которые оценивались по представленным в табл. 1 параметрам. В ходе анализа аудиозаписей выделены следующие основные группы:

- *аудиозаписи с высокочастотным шумом* – записи с нормальным уровнем полезного сигнала, но присутствует высокочастотный шум, который частотно схож с голосом. Такие аудиозаписи характеризуются накладыванием шума на речь;

- *аудиозаписи с низкочастотным шумом* – аудиозаписи с частотой шума ниже частоты голоса, что позволяет голосу звучать более отчетливо относительно шума. Эти записи, несмотря на низкие технические показатели, обычно хорошо воспринимаются человеком (оператором) и хорошо поддаются обработке;

- *аудиозаписи с низким уровнем полезного сигнала* – записи, где полезный сигнал имеет очень низкий уровень по сравнению с шумом. К этой группе также относятся образцы с постоянным однотонным шумом. Такие записи обычно хуже поддаются обработке.

Таблица 1

Перечень используемых параметров оценки аудиозаписей

Оценка	Запись 1
1. Чистота и ясность звука: оценка уровня шума в аудиозаписи; определение четкости и различимости речи на разной удаленности от микрофона	М: –33 LUFS Речь четкая, различимая

Окончание табл. 1

Оценка	Запись 1
2. Искажения и артефакты: выявление и оценка искажений, таких как дрожание, искажение голоса, эхо и т. д.; определение наличия артефактов, вызванных низким качеством микрофона или компрессией	Дефектов нет Дефектов нет
3. Громкость и динамический диапазон: измерение уровня громкости речи при разной интенсивности произнесения; анализ динамического диапазона – отношения между самыми тихими и самыми громкими звуками	M: –19,3/–14,7; I: –20,9 LUFS LRA: 4,6 LU
4. Отношение сигнал/шум: расчет отношения между сигналом (речью) и фоновым шумом; сравнение этого отношения при разной удаленности от источника шума	PEAK: –23/–4,6 dB Шумы слышно хуже, чем речь
5. Позиционирование и направленность: исследование влияния позиции микрофона относительно источника шума на качество записи; оценка эффективности направленных микрофонов при разной ориентации	Дефектов нет Дефектов нет
6. Субъективное восприятие: опрос участников эксперимента (в виде комментариев участников команды)	Запись хорошая, но есть шумы. На восприятие не влияет
Оценка качества речи	80 %

## Технические параметры громкости

В рамках данного тестирования при анализе аудиозаписей использовано несколько технических параметров громкости, каждый из которых, в частности, различным образом отражает специфику, как аудио воспринимается человеческим слухом. Ниже приведены основные параметры громкости и их объяснение:

1. PEAK (dB) – определяет пиковую громкость в каждый момент времени, отражая максимальную амплитуду звука. Этот показатель часто используется в повседневной жизни для оценки громкости.

2. RMS и LUFS (dB/LUFS) – определяют среднюю громкость на протяжении определенного времени. RMS более устаревший метод, в то время как LUFS представляет собой более современный и точный формат измерения средней громкости. LUFS измеряется в LUFS, аналогично dB. Он предоставляет показатель средней громкости аудио за определенный период времени. Этот параметр более точно отражает то, как аудио воспринимается человеческим слухом, учитывая изменения громкости во времени.

3. LRA (LU) – определяет разницу между самым тихим и самым громким сигналом на протяжении всего отрезка времени. Он предоставляет информацию о динамическом диапазоне средней громкости в аудиозаписи.

4. Momentary (LUFS) – показывает максимальную среднюю громкость в данный момент времени и близок к показателю пиковой громкости.

Параметр LUFS, в отличие от PEAK, ближе к тому, как человеческий слух воспринимает громкость в целом. Например, музыка может начинаться тихо, затем стать громче и затем вернуться к тишине. LUFS отражает этот средний уровень громкости по результатам прослушивания всей аудиозаписи. Он особенно важен при обработке аудио.

Используем следующие типы параметра LUFS:

- Long (I – Intergated) – определяет общую среднюю громкость за длительный период времени. Он является основным параметром, отображающим среднюю громкость на протяжении всей аудиозаписи;
- Short Term – работает аналогично Long, но отображает более короткий период времени и используется чаще для обработки, чем для анализа;
- Range (LRA, LU) – показывает разницу между самым тихим и самым громким сигналом на всем отрезке времени. Он измеряется в LU и предоставляет информацию о динамическом диапазоне средней громкости;
- Momentary – отражает максимальную среднюю громкость в текущий момент времени и ближе к показателю пиковой громкости.

В табл. 2 представлены ключевые параметры и оценки, используемые для анализа аудиозаписей. Эти параметры помогают в понимании уровня шума, громкости речи при разной интенсивности произношения, динамического диапазона и отношения между сигналом и фоновым шумом.

Таблица 2

Ключевые параметры и оценки, используемые для анализа аудиозаписей

Оценка	Описание
Оценка уровня шума в аудиозаписи	S (Short Term): ... LUFS. Показывает средний уровень шума на всей записи. Шум и голос оцениваются вместе
Измерение уровня громкости речи при разной интенсивности произнесения	M (Momentary): ... / ...; I (Intergated): ... LUFS. M – показывает самый тихий момент полезного сигнала и (через /) самый громкий момент полезного сигнала, оценивая голос и шум вместе; I – показывает среднюю общую громкость записи за весь ее период
Анализ динамического диапазона – отношения между самыми тихими и самыми громкими звуками	LRA: ... LU. Показатель разницы между самым громким и самым тихим моментом записи, исключая начальную и конечную тишину
Расчет отношения между сигналом (речью) и фоновым шумом	PEAK: .../... dB. Показывает самый громкий уровень чистого шума и (через /) самый громкий, пиковый уровень громкости сигнала, оценивая голос и шум вместе. Если вычесть одно из другого, можно получить примерный уровень громкости полезного сигнала

На рис. 1 приведено типовое графическое представление параметров громкости аудиозаписи (на примере аудиозаписи 1), на котором показаны графики.



Рис. 1. Графическое представление параметров громкости аудиозаписи 1:  
*серая линия* (PEAK, dB) отображает пиковую громкость в каждый момент времени. На графике видны пики громкости, клиппинг (ограничение амплитуды), паузы в речи и уровень шума в начале и конце записи;  
*белая линия* (Short Term) отображает среднюю громкость на коротких отрезках времени. Этот график помогает определить, как человеческий слух воспринимает громкость на различных участках записи;  
*зеленая линия* (Integrated) отображает среднюю общую громкость на протяжении всей записи

Далее приведена сводная таблица полученных значений параметров громкости.

Таблица 3

Сведенные количественные параметры громкости

Loudness				
Short Term	Integrated	Momentary	LRA	True Peak
-22,3	-20,9	-14,7	4,6	-4,6
-20,3	-20,3	-13,6	6,3	-4,5
-25,4	-17,8	-10,6	5,9	-0,5
-22,9	-21,2	-14	3,9	-4,2

Окончание табл. 3

Loudness				
Short Term	Integrated	Momentary	LRA	True Peak
-22,1	-21,8	-15,8	4	-6,5
-22,6	-22,5	-17,4	3,9	-6,2
-21,1	-19,4	-13	3,9	-3,3
-22,5	-21	-13,2	4,7	-2,3
-20,8	-21,5	-13,7	5	-3,4
-27	-27,3	-21,3	7,3	-8,8
-26,9	-25,1	-19,8	10,1	-9,2
-29,3	-27,7	-22,4	9,6	-10,3
-21,2	-18	-11,6	3,5	0,4
-15,7	-15,6	-10	6,8	-2
-19,6	-17	-7,5	10,9	0,2
-17,4	-15,3	-6,8	5,4	0,3
-13,2	-12,4	-3,7	7,5	0,5
-27,7	-20,7	-5,3	19,6	3,1
-7,1	-7	-3,8	3,3	0,6
-19	-17,6	-8	6,5	0,2
-20,3	-18,8	-12,4	10,9	-2,2
-23	-21,3	-13,4	6,6	-3,9
-22,6	-19,6	-11,6	5	-1,6
-15,6	-14,9	-7,4	3	0,4
-21	-20,8	-11,6	12,4	-1
-7,9	-7,8	-5,4	2,5	0,5
-6,4	-6,8	-3,8	2,6	0,5

Таким образом, получены следующие общие значения акустических параметров:

- *уровень громкости* (оценка среднего уровня громкости шум + голос в аудиозаписи, Loudness Short Term): все образцы имеют уровни громкости в диапазоне от -29,3 до -6,4 LUFS, что свидетельствует о значительной разнице в громкости между образцами;
- *динамический диапазон* (разница между самым тихим и самым громким сигналом на протяжении записи, Loudness Range (LRA)): значения LRA варьируются от 2,5 до 19,6 LU, что указывает на существенные различия в динамическом диапазоне громкости между образцами;
- *пиковая громкость* (True Peak): значения True Peak колеблются от -10,3 до 3,1 dB, что указывает на наличие пиковых моментов с разной интенсивностью в образцах.

Представленный анализ акустических параметров показывает существенное различие аудиозаписей по уровню громкости и динамическому диапазону при наличии пиковых значений в широком диапазоне.

### Результаты распознавания речи

В рамках данного исследования получены текстовые транскрипции из аудиофайлов с использованием системы автоматического распознавания речи (ASR – Automatic Speech Recognition). Были получены следующие средние показатели:

– CER (Character Error Rate) – процент ошибок (отличий от оригинала) на уровне символов в распознанном тексте составляет 2,7 %;

– WER (Word Error Rate) – процент ошибок (отличий от оригинала) на уровне слов в распознанном тексте – 9,8 %.

CER = 2,7 % указывает на относительно высокую точность распознавания символов, где около 2,7 % символов в распознанном тексте отличаются от оригинала.

WER = 9,8 % является низким показателем и указывает на относительно высокую точность распознавания слов, где около 9,8 % слов в распознанном тексте отличаются от оригинала.

Наряду с абсолютными нулевыми эталонными значениями данных показателей (которые соответствуют теоретически идеальному случаю максимально эффективного распознавания речи при отсутствии воздействия шумовой обстановки) методически обоснованным представляется рассматривать в качестве фактического эталонного образца запись с наилучшими из фактически полученных показателей. В настоящем исследовании такими значениями являются CER = 0,8 % и WER = 2,6 %.

### Выводы

На основании полученных показателей можно сделать вывод о хорошей производительности системы автоматического распознавания в задаче получения текстовых транскрипций из аудиофайлов. Даже при наличии факторов, таких как уровень шума и изменчивость громкости речи, система автоматического распознавания продемонстрировала достаточно высокую точность распознавания как символов, так и слов.

Точность распознавания на уровне символов и слов является приемлемой для практического использования системы распознавания на борту РС МКС. Широкий спектр условий и вариантов аудиозаписей позволяет судить о практической применимости полученных результатов распознавания.

Представленные результаты полностью применимы к использованному оборудованию (MacBook) и необходимо продолжить тестирование с использованием экземпляра фактического ноутбука, который планируется использовать для бортовой реализации эксперимента (HP ZBook).

## ЛИТЕРАТУРА

- [1] Рулев, Д.Н. Разработка виртуального ассистента космонавта. Космический эксперимент «Ассистент» / Д.Н. Рулев, С.И. Ашманов. – DOI: 10.21046/ISS-2023 // Наука на МКС: Третья международная конференция, посвященная 25-летию Международной космической станции: сб. тез. докл. – Москва: ИКИ РАН, 2023. – С. 235–238.
- [2] Рулев, Д.Н. Подход к разработке программного обеспечения виртуального ассистента космонавта / Д.Н. Рулев, С.И. Ашманов // Современные проблемы ракетной и космической техники. Сборник статей, 2023. – Казань: РИЦ «Школа», 2023. – С. 182–190.
- [3] Оценка функционирования программного обеспечения голосового ассистента космонавта в условиях моделируемой шумовой обстановки РС МКС / Д.Н. Рулев, С.И. Ашманов, В.С. Крайнов, Е.В. Попова [и др.] // Пилотируемые полеты в космос. Материалы XV Международной научно-практической конференции. – Звездный городок, 2023. – С. 167–169.
- [4] Рулев, Д.Н. Разработка виртуального ассистента космонавта / Д.Н. Рулев, Д.М. Швед, С.И. Ашманов // XLVII Академические чтения по космонавтике, посвященные памяти академика С.П. Королёва и других выдающихся отечественных ученых – пионеров освоения космического пространства (Москва, 24–27 января 2023 года): сборник тезисов: в 4 т. – Москва: Изд-во МГТУ им. Н. Э. Баумана, 2023. – ISBN 978-5-7038-6053-3. – Т. 2. – ISBN 978-5-7038-6122-6. – С. 368–370.
- [5] ГОСТ Р 51061-97. Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений = Low bitrate speech transmission systems. Speech quality characteristics and their evaluation: государственный стандарт Российской Федерации: издание официальное: утвержден и введен Постановлением Госстандарта России от 24 июня 1997 г. N 223: введен впервые: дата введения 1998-01-01 / разработан Министерством связи Российской Федерации. – Москва: Стандартинформ, 1998. – 24 с.
- [6] ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости = Speech transmission over varies communication channels. Techniques for measurements of speech quality, intelligibility and voice identification: государственный стандарт Российской Федерации: издание официальное: утвержден и введен Постановлением Госстандарта России от 21.11.95 N 579: введен впервые: дата введения 1997-01-01 / разработан Московским Государственным университетом и войсковой частью 93901. – Москва: Стандартинформ, 1997. – 234 с.
- [7] ГОСТ Р 59879-2021. Эргономика. Проектирование и применение испытаний речевых технологий. Методика определения показателей качества распознавания голосовых команд управления = Ergonomics. Design and application of speech technology tests. Methodology for determining quality indicators of recognition of voice control commands: национальный стандарт Российской Федерации: издание официальное: утвержден и введен Приказом Федерального агентства по техническому регулированию и метрологии от 25 ноября 2021 г. N 1582-ст: введен впервые: дата введения 01-03-2022 / разработан Федеральным государственным унитарным предприятием «Научно-исследовательский институт «Квант». – Москва: Стандартинформ, 2022. – 20 с.

## REFERENCES

- [1] Rulev, D.N. Development of a Cosmonaut's Virtual Assistant. Development of a Voice Assistant for Cosmonauts. "Assistant" Space Experiment / D.N. Rulev, S.I. Ashmanov. – DOI: 10.21046/ISS-2023 // Science on the ISS: The 3d International Conference Dedicated to the 25th Anniversary of the International Space Station: Collection of Abstracts, Reports. – Moscow: IKI RAS, 2023. – P. 235–238.
- [2] Rulev, D.N. An Approach to the Development of the Software for a Virtual Assistant for Cosmonauts / D.N. Rulev, S.I. Ashmanov // Current Problems of Rocket and Space Technology. Collection of Papers, 2023. – Kazan: Editorial and Publishing Center "Shkola", 2023. – P. 182–190.
- [3] Evaluation of the Software Operation for a Voice Assistant for Cosmonauts Under Simulated Noise Conditions on the ISS RS / D.N. Rulev, S.I. Ashmanov, V.S. Krainov, E.V. Popova [et al.] // Manned Flights to Space. Proceedings of the 15th International Scientific and Practical Conference. – Star City, 2023. – P. 167–169.
- [4] Rulev, D.N. Development of a Virtual Assistant for Cosmonauts / D.N. Rulev, D.M. Shved, S.I. Ashmanov // XLVII Academic Readings on Cosmonautics Dedicated to the Memory of Academician S.P. Korolev and Other Outstanding Domestic Scientists – Pioneers of Space Exploration (Moscow, January 24–27, 2023): Collection of Abstracts: in 4 Volumes. – Moscow: Publishing House of N.E. Bauman MSTU, 2023. – ISBN 978-5-7038-6053-3. – V. 2. – ISBN 978-5-7038-6122-6. – P. 368–370.
- [5] GOST R 51061-97. Low-speed Speech Transmission Via Digital Channels. Parameters of Speech Quality and Measurement Methods: the State Standard of the Russian Federation: Official Publication: Approved and Introduced by Resolution of the State Standard of Russia Dated June 24, 1997 N 223: Introduced for the First Time: Date of Introduction 1998-01-01 / Developed by the Ministry of Communications of the Russian Federation. – Moscow: Standartinform, 1998. – 24 p.
- [6] GOST R 50840-95. Speech Transmission Via Communication Paths. Methods for Evaluating the Quality, Intelligibility, and Recognizability: the State Standard of the Russian Federation: Official Publication: Approved and Introduced by Resolution of the State Standard of Russia Dated 11-21-95 No 579: Introduced for the First Time: Date of Introduction 1997-01-01 / Developed by Moscow State University and Military Unit 93901. – Moscow: Standartinform, 1997. – 234 p.
- [7] GOST R 59879-2021. Ergonomics. Designing and Application of Speech Technologies Trials. Methods for Determining the Quality Indicators of the Recognition of Voice Control Commands: the National Standard of the Russian Federation: Official Publication: Approved and Introduced by Order of the Federal Agency for Technical Regulation and Metrology Dated November 25, 2021 N 1582-st: introduced for the First Time: Date of Introduction 01-03-2022 / Developed by the Federal State Unitary Enterprise "Scientific Research Institute "Kvant". – Moscow: Standartinform, 2022. – 20 p.