

# Operational and Experimental Tasks, Performance, and Voice in Space

Bernd Johannes; Sergey V. Bronnikov; Juri A. Bubeev; Tatyana I. Kotrovskaya; Daria V. Shastlivtseva; Sarah Piechowski; Hans-Juergen Hoermann; Joern Rittweger; Jens Jordan

- BACKGROUND:** Voice analysis offers an unobtrusive approach for psychological monitoring. We demonstrate the relationship between voice parameters and cognitive performance in: 1) a task with psychological test character, and 2) performance in an operational, mission-relevant task. The central methodological aim was to verify the usefulness of voice commands and counting in providing anchor values for the step-function model of voice pitch.
- METHODS:** During a 22-yr period, 42 cosmonauts participated in the Russian space experiment "Pilot", which was a hand-controlled docking maneuver. As reference the experiment included the cognitive task "Manometer." This task was controlled through voice commands. These voice commands were stored and are the basis for the present analysis.
- RESULTS:** Cosmonauts differed in their working style and respective performance during the Manometer task. Clustered groups can be assumed to represent different effort. Importantly, these groups differed in the changes of voice pitch among mission phases and among task repetitions. However, there were no differences between these motivation groups and performance in the professional task.
- DISCUSSION:** The differing effort is the effect of different motivation of cosmonauts for experimental test tasks vs. mission-relevant professional tasks. Latter ones provide a more reliable chance to assess the real actual state and skills of a cosmonaut. Voice pitch measurement seems to be reliable and useful under space conditions for monitoring this volitional effort.
- KEYWORDS:** fundamental voice frequency, spaceflight, docking maneuver, mental load, monitoring.

Johannes B, Bronnikov SV, Bubeev JA, Kotrovskaya TI, Shastlivtseva DV, Piechowski S, Hoermann H-J, Rittweger J, Jordan J. *Operational and experimental tasks, performance, and voice in space. Aerosp Med Hum Perform.* 2019; 90(7):624–631.

For more than 50 years, voice analysis<sup>22,23</sup> has been investigated as a tool for monitoring mental state and operator readiness during mission-relevant activities. In military aviation as well as in spaceflight, research has been conducted to verify the information content of voice parameters with respect to cognitive, emotional, and behavioral aspects of pilots<sup>30,40</sup> or astronauts. In civil aviation, other transport systems, or power stations, voice analysis played an important role in analyzing accidents and disasters.<sup>3</sup> In space, the first experimental studies were conducted in the late 1960s by Sulc and Remek;<sup>35</sup> Friedrich and Vaic;<sup>5</sup> and Vaic, Friedrich, and Kolinchenko.<sup>39</sup> In 1965, voice analysis helped monitor the very first extravehicular activity (EVA) of the cosmonaut Leonov outside the spacecraft "Voschod-2."<sup>27</sup>

The complete frequency spectrum of the human voice (70–4000 Hz) displays ranges of higher or lower intensity, which are known as formants, the intensity of a particular frequency range in the speech.<sup>36</sup> The sound spectrum of the voice varies

enormously during speaking, singing, or screaming. In our work, we focus on voice pitch, the fundamental frequency (F0) of the voice. The air stream produced by intrathoracic pressure and initially pitched by the glottis is the source of voice production. F0 is the lowest frequency in the spectrum, pitched by the vibrations of the glottis. Lower frequencies are impossible. The innervation of the vocal chords by the recurrent nerve, a side branch of the vagal nerve,<sup>9</sup> gives rise to speculations that voice pitch is related to the autonomic nervous system and, thus,

From the German Aerospace Center (DLR), Institute of Aerospace Medicine, Cologne, Germany; S. P. Korolev Rocket and Space Corporation "Energia", Korolev City, Russia; and the Russian Federation State Research Center, Institute for Biomedical Problems (IBMP) of the Russian Academy of Sciences, Moscow, Russia.

This manuscript was received for review in February 2019. It was accepted for publication in April 2019.

Address correspondence to: Bernd Johannes, Ph.D., Institute of Aerospace Medicine, Linder Hoehe, D-51147; bernd.johannes@dlr.de.

Reprint & Copyright © by the Aerospace Medical Association, Alexandria, VA.

DOI: <https://doi.org/10.3357/AMHP.5351.2019>

involved in all emotional reactions.<sup>21,34</sup> In addition, the vocal sounds that we perceive and hear are further influenced by individual anatomical features of the acoustic tract.

The literature (for review see Scherer<sup>34</sup> and Giddens *et al.*<sup>6</sup>) suggests a general correlation between voice pitch and emotional load. Already the early studies suggested that voice parameters indicate general emotional states.<sup>20,21,33</sup> In particular, higher fundamental voice frequencies seem to indicate that the speaker is emotionally excited.<sup>7,29,37</sup> A significant increase in the F0 was found under extreme natural stress conditions.<sup>40,41</sup> In critical life events and under experimental stress, F0 depends more on psychological than on physical load.<sup>10,38</sup>

More recent studies<sup>2,18</sup> support these findings, specifically in research on emotion recognition.<sup>4,11,19</sup> Moreover, the approach was clinically tested as a diagnostic tool for depression.<sup>26,28</sup> However, acoustic emotion correlates in the human voice exhibit large interindividual differences. Therefore, calibration for personal voice parameters is required for the evaluation of individual psychological states by means of voice analysis. Some investigators assumed a linear relationship between voice pitch and physical load.<sup>25</sup> Yet our experiments on voice pitch<sup>14,16</sup> yielded a hypothetical step function model.<sup>17</sup> The modal value of voice pitch remains relatively stable across certain small frequency ranges. However, these ranges differ between rest, mental or physical activity which is well tolerated by the individual, and individually borderline (physical) load. Anchor values at rest and during mental loading were assessed using an experimental approach also used in our space experiment and described in detail below. Therefore, we tested the hypothesis that these anchor values are also applicable under spaceflight conditions. The standardized voice commands in the experiment “Pilot” provided quality material for this analysis. We standardized the mental load using the cognitive task Manometer. Furthermore, we compared the unique conditions during spaceflight such as weightlessness with terrestrial conditions.

In the present work, we investigated the relationship between voice pitch as a dependent variable and the independent variables mission phases and performance in: 1) a cognitive task with psychological test character (Manometer task, **Fig. 1**); and 2) the performance in an operational, mission-relevant task (hand-controlled docking maneuver). The analysis suggests that voice pitch indicates volitional effort.

## METHODS

### Subjects

The experiment “Pilot” was approved both by the local institutional review board of the Institute for Biomedical Problems in Moscow and the Human Research Multilateral Review Board (for ISS experiments). In the period between October 1996 and December 2018, Russian cosmonauts participated in the experiment “Regulation” as a part of the Russian Long-Term-Program experiment “Pilot.” The experiment was conducted in three epochs: from 1996 to 2000 on the Mir station and from 2008–2011 and 2015–2018 on the International Space Station (ISS).

Overall, 42 cosmonauts ages  $45.5 \pm 5.6$  yr participated in the experiments in all three mission phases: preflight, in flight, and postflight. Flight duration varied from 13 to 381 ( $185 \pm 60$ ) d. Cosmonauts underwent three preflight (–1 mo, –10 d, –3 d prior to launch) and three postflight (+3 d, +10 d, +2 to 3 mo post-landing) experimental sessions. In flight, cosmonauts executed the experiment at irregular intervals on Mir and at regular monthly intervals on the ISS.

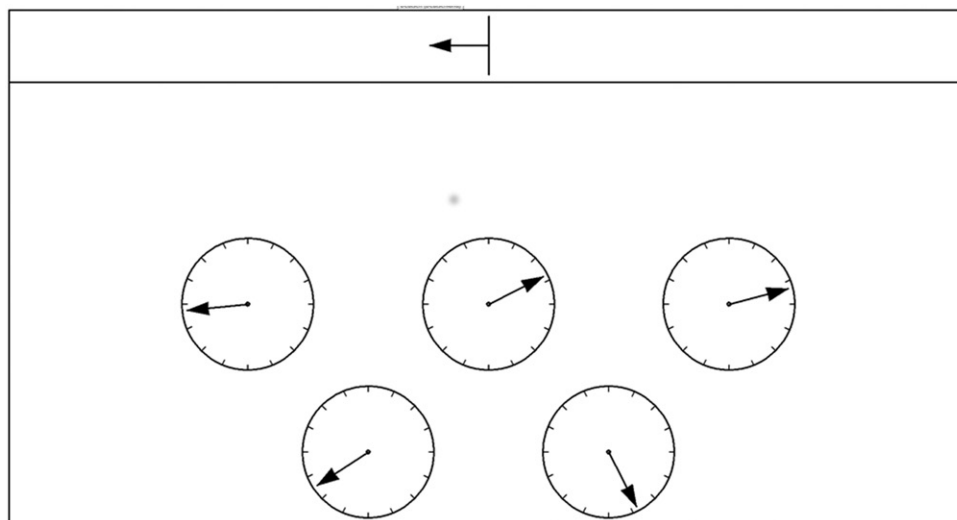
### Equipment

During the first two epochs, original spacecraft controls were used for the experiment. For the actual simulator and ground studies, laboratory hand controls were manufactured by Koralewski Industrie Elektronik oHG (Hambühren, Germany). Functionally, they are equivalent to the original controls. Psychophysiological and voice parameters were registered using different generations of the Neurolab system (Neurolab-B, Neurolab-2000M, Neurolab-2010). Neurolab-B was assembled by the Bulgarian Academy of Science, and the two later device generations were developed and produced by Koralewski Industrie Elektronik oHG. The experimental computer software and the firmware controlling all measurement systems were developed by SpaceBit GmbH (Eberswalde, Germany). For the first two device generations, all sensors and measurement modules were integrated into body vests. The actual polygraph is used as an on-table application. These three polygraph generations all featured specific additional measurement channels but were comparable in the main channels, described below.

### Material

For speech recordings, a commercially available, space certified head-set microphone (Sennheiser HMD 25-1, Wedemark, Germany; last space series Beyerdynamic MMX300, Heilbronn, Germany), was used. The microphone was positioned approximately 1 cm in front and beside the speaker's mouth. The voice commands were sampled and stored at 8 kHz for a 2-s interval and were twice verified off-line by the first author. The first analysis verified the cosmonaut's final decision. Sometimes, cosmonauts started with one voice command and then changed their mind. The second verification served to remove unclearly spoken voice commands and other noises from the detailed voice analysis. Sometimes cosmonauts used voice commands different from “Okay” and “Error”, or they talked to other crewmembers. Volitional changes of intonation such as yawning or singing or statements indicating doubts were also excluded. All word samples were analyzed separately. For a robust detection of the “averaged” modal voice pitch per experiment, we lumped histograms of single F0 values together for both voice commands separately. The results were also sorted for each experiment into four histograms: per voice command and correctness of the response.

There is a plethora of algorithms for voice pitch detection.<sup>8</sup> We used an algorithm developed by Lüdge and Gips.<sup>24</sup> The approach includes sliding calculations of the autocorrelation function with a fixed 20-ms time window and 10-ms overlap. Thus, the fundamental frequency is computed in 10-ms increments.

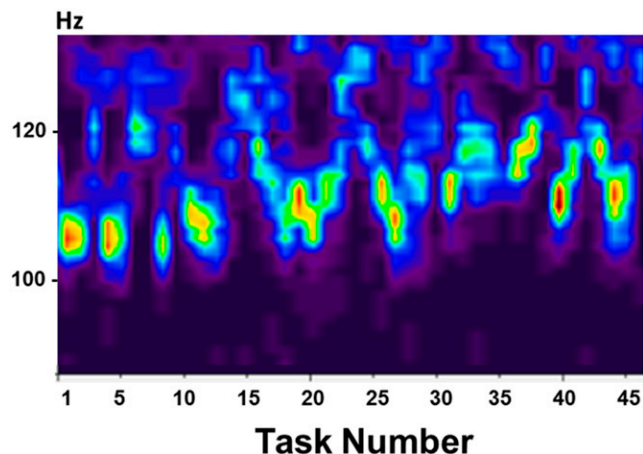


**Fig. 1.** Manometer task with difficulty level 5 (five gauges).

For male speakers, the detection range is limited between 70 Hz and 180 Hz. The reliability can be further increased by calculating the short-time histogram and the mode of F0 values (F0m), the frequency occurring most often in the histogram (Fig. 2). The position of the F0m is, to a large extent, devoid of specific intonations or external occasional noise disturbances and depends on an individual's state.

As mentioned above, our primary goal was to verify the usefulness of voice commands and counting to provide anchor values for the voice pitch step-function model. While useful as anchor frequencies, the magnitude of deviation among these anchor values is also of importance. The performance in the Manometer task was analyzed with respect to the performance in the mission-relevant task—the docking maneuver.

The registered voice samples were of high acoustic quality, thus allowing for sophisticated voice analysis, including jitter, shimmer, and formants. We present these results, which are beyond the scope of the manuscript, in **Appendix A** (online; <https://doi.org/10.3357/AMHP.5351sd.2019>).



**Fig. 2.** Voice pitch histograms as time series over tasks; the mode is visualized by the bright color and indicates relative stability within a small frequency range.

## Procedure

Within the Russian Long-Term-Program, the experiment “Pilot” was developed to examine cosmonauts’ performance in a simulated training task of hand-controlled approximation and docking of the spacecrafts Soyuz and Progress at the present space station (Mir and ISS) during different stages of long-term spaceflights.<sup>13,31,32</sup> The dynamic and informational demands on cosmonauts during the docking simulation is based on mathematical models of the realistic dynamics of spacecraft movements. The view on the space station during the docking

maneuver is near-photographically displayed on the computer screen.

The experiment “Regulation” assessed psychophysiological response patterns for individualized classification and calibration of physiological arousal and energetic effort.<sup>12</sup> These integrated physiological scores have been applied for the evaluation of mental load during the docking training experiment Pilot. The mentally challenging tasks were executed using a word command recognition system. The voice commands were recorded and provide, along with the test performance information, the data material for the present analysis. The protocols of the Regulation experiment differed slightly over time with respect to other additional tasks. However, the Manometer task was always applied first.

The Manometer task induced time pressure adaptively at the individual performance level and varies the information complexity. When all needles of the gauges pointed to the same semi-circular arch (right, left, upper, lower half) as displayed in the upper part of the screen, the subjects had to respond with “Okay!” (see Fig. 1). When at least one of the needles differed more than 90° from the predetermined direction subjects had to state “Error.” The pace of the presentation varied in such a way that the individual error rate remained stable between 20 and 25%. Information complexity was escalated by increasing the number of gauges from five to seven, and then to nine. However, the cosmonauts always had the possibility of abbreviating the Manometer procedure by pressing the “escape” button and continuing with the docking training. We excluded experiments with less than 50% of Manometer tasks (8%) from the present analysis.

The voice pitch of the voice commands in the Manometer task served as reference (anchor value) for the mental load state. The neutral counting from 1 to 10—a commonly accepted no-load condition—indicated resting conditions.

## Statistical Analysis

For the statistical analysis, we used the SPSS IBM® package (V21). Generally, we averaged voice parameters for each

cosmonaut and respective test conditions. This approach was necessary given the substantial interindividual variability in voice data and the fact that the number of voice samples differed between cosmonauts. Thereby, statistical power resembled that of the averaged histograms for the voice commands. We developed linear mixed effect models to test the statistical significances of the independent variables as fixed effects. Cosmonauts were included as random effects. For the analyses including all voice samples, we visually inspected a q-q-plot to accept an LME model. For averaged data, the normal distribution of residuals was tested by means of the Kolmogorov-Smirnov test. Additionally, for all relevant comparisons nonparametric tests were applied. A  $P$ -value of 0.05 was accepted as significance level. The performance data of the Manometer task were analyzed by means of the single linkage method of Ward to determine the number of existing clusters.<sup>1</sup>

## RESULTS

Data sets were accepted from 42 cosmonauts having run the experiment in all three mission phases: preflight, in flight, and postflight. Voice data's provenience was verified by acoustic speaker identification. **Table I** provides an overview of experimental data during the three space study epochs and the mission phases.

In several experiments, the Manometer part was aborted and in 9 experiments the performance was below 40% (> 66% is higher guess level) such that 355 experiments of 413 (86%) could be included in the analysis. In the included experiments, 42,927 voice commands were registered. The first word verification with respect to cosmonaut's final decision in a task left 36,810 voice commands for the performance estimation. After selecting the word samples without acoustic or semantic disturbances, 34,798 single word samples and 5653 histograms remained for voice analysis. The task distribution of Okay tasks and Error tasks was planned to be equal, and the observed distribution of both task types was 51.6 vs. 48.4. However, the responses were differently distributed: 52.2% "Okay" and 47.8% "Error"; therefore 54.3% of the responses were correct and 45.7% were wrong.

**Table I.** Number of Experiments per Space Study Epochs and Mission Phases.

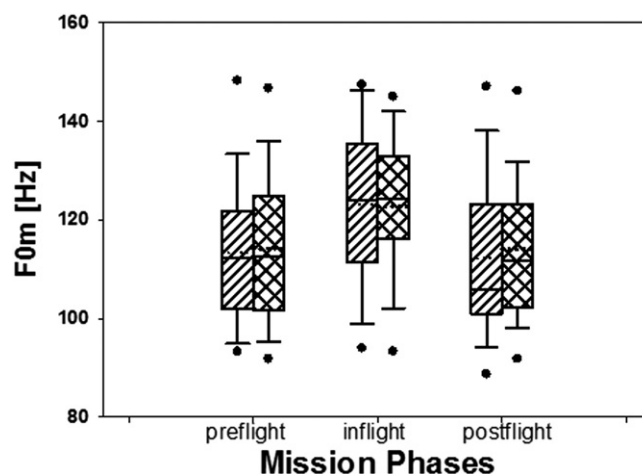
SPACE STATION	SPACE STUDY EPOCHS	MISSION PHASE	EXPERIMENTS
1 (Mir)	1 (1996–2000)	Preflight	40
		In flight	45
		Postflight	24
2 (ISS)	2 (2008–2011)	Preflight	22
		In flight	63
		Postflight	24
	3 (2015–2018)	Preflight	28
		In flight	139
		Postflight	28
Sum			413

The first analysis tested frequency differences between "Okay" and "Error" commands [Russian: "впорядке" (vporядke), and "ошибка" (oshipka)]. We reasoned whether data from both commands could be lumped to define one common anchor value representing the mental load level. The first LME model analyzed the fixed effects of the two words [ $F(\text{num: } 1, \text{denum: } 172,019) = 0.281, P = 0.595$ ; MW-U  $P = 0.610$ ], the three mission phases (preflight, in flight, postflight) [ $F(\text{num: } 2, \text{denum: } 172,30131.606) = 28,685, P < 0.001$ ; MW-U  $P < 0.001$ ], and the respective interactions with the words. No difference was found for F0m between the voice commands "Okay" and "Error", but there was a difference among mission phases (**Fig. 3**), without interaction between both factors. The residuals were normally distributed (K-S-Z = 0.767,  $P = 0.599$ ).

The second LME analyzed the fixed effects of the two anchor frequencies voice commands (Manometer task) vs. neutral counting [ $F(\text{num: } 1, \text{denum: } 142,370) = 15.611, P > 0.000$ ; MW-U  $P = 0.225$ ], the three mission phases (preflight, in flight, postflight) [ $F(\text{num: } 2, \text{denum: } 136,737; P < 0.001)$ ; MW-U  $P = 0.075$ ], and the respective interactions with the voice responses. F0m differed significantly between both anchor frequencies and also among mission phases. No interaction occurred between both factors. The residuals were normally distributed (K-S-Z = 0.511,  $P = 0.957$ ).

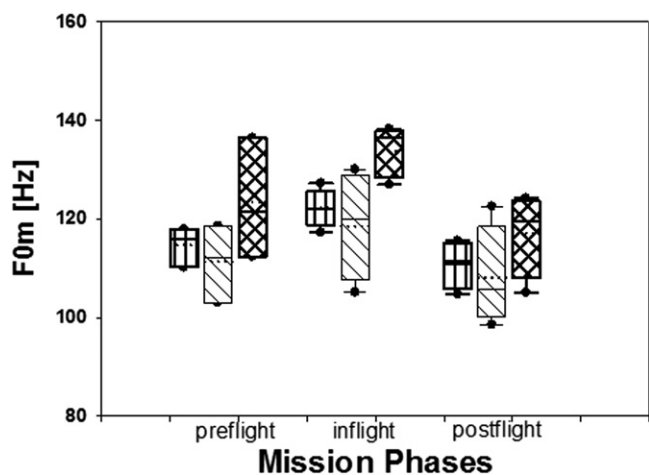
During the early docking training sessions on Mir, but not during later experiments, cosmonauts had to loudly describe what they were doing for ground control. They described the actual situation between the spacecraft and the space station and their active actions. During these experiments, the voice pitch was assessed in real-time by the same procedure as during the counting phases. **Fig. 4** illustrates that the voices of the cosmonauts remained on the level of neutral counting and did not show extraordinary excitation during the docking training.

In this paper, we only present the performance values assessed after acoustically corrected word recognition, but disregarded other available information (reaction times, presentation times). We did not observe significant performance



**Fig. 3.** Voice pitch mode of the two voice commands ("Okay" – sloped lines; "Error" – crossed lines) during different mission phases. Means and medians are represented by dotted and solid lines, respectively.



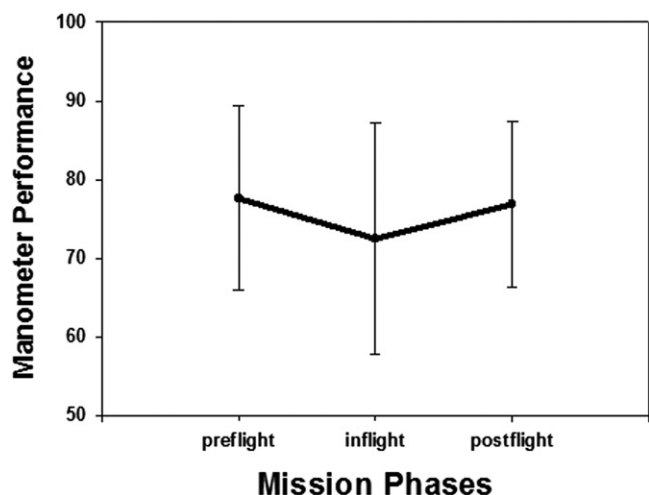


**Fig. 4.** Voice pitch mode of counting (vertical lines) and voice commands during Manometer (crossed lines), which served as anchor values compared to voice pitch during the docking training (sloped lines). Note that the voice pitch during docking was elevated in flight as compared to preflight and postflight, but lower than the in-flight anchor value for “mentally loaded” state.

differences depending on the difficulty of the Manometer task. Therefore, we averaged common performance values across the three degrees of difficulty.

The general performance in the Manometer task did not differ among the mission phases [Fig. 5;  $F(\text{num}: 2, \text{denum}: 319,850) = 0.113, P = 0.893$ ; KW:  $P < 0.420$ ]. Residuals were normally distributed (K-S-Z = 1.347,  $P = 0.053$ ). A more detailed analysis showed that True tasks and Error tasks were solved differently in different experiments. A cluster analysis (Ward method) based on the performance in both task types differentiated three groups (Fig. 6).

We chose the three-cluster solution ( $N1 = 30, N2 = 191, N3 = 134$ ) given the clear structure of the smallest one (group 1), which was first separated from the others. The three groups mainly differed in their performance in Error tasks. Whereas Okay tasks were solved over 80% in all groups (blank bars in Fig. 6), the Error tasks were solved well (77%) in the third group



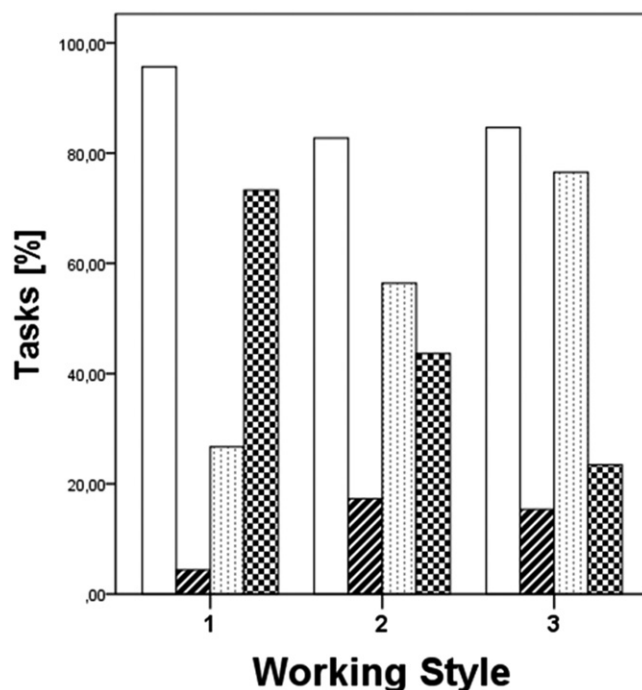
**Fig. 5.** The averaged performance values of the Manometer task did not differ between the three mission phases.

only (dotted bars in Fig. 6). In the second group, tasks were solved at the guess level (56%) and very poorly in the first group (27%). Tautologically, averaged performance differed between groups [ANOVA- $F(2) = 68.509, P < 0.001$ ; KW:  $P < 0.001$ ]. These groups of different “working styles” were not significantly related to the mission phases ( $cc = 0.146, P = 0.104$ ) but to the cosmonauts ( $cc = 0.633, P < 0.001$ ).

The finding suggests that cosmonauts generally adhered to one of these clustered groups during their space mission. Note that clusters were arranged in a way that a higher group number indicates a higher performance in the Error tasks, which we took as an indicator of higher effort in the Manometer task. We constructed an effort score by the mean of a subject’s class assignment and split the cohort into two effort groups: lower effort = group 0; higher effort = group 1. Group affiliation was not related to mission phases ( $cc = 0.016, P = 0.959$ ).

The effort groups did not differ in F0m values during the baseline counting [ $F(\text{num}: 1, \text{denum}: 38,540) = 1.313, P = 0.259$ ]. Counting vs. voice commands [ $F(\text{num}: 1, \text{denum}: 128,481) = 16.809, P < 0.001$ ] as well as the mission phases [ $F(\text{num}: 2, \text{denum}: 123,182) = 27.307, P < 0.001$ ] had significant effects on the F0m. Residuals were normally distributed (K-S-Z = 0.508,  $P = 0.959$ ).

The better performing higher effort group 1 generally showed higher voice pitch values. During the single tasks F0m remained on a higher frequency level in group 1 compared with group 0. There was no direct interaction between the number of items and the effort group. Yet we observed a specific significant three-way interaction between effort groups, item number, and mission phases [ $F(\text{num}: 58, \text{denum}: 22,192,800) = 1.977, P < 0.001$ ].



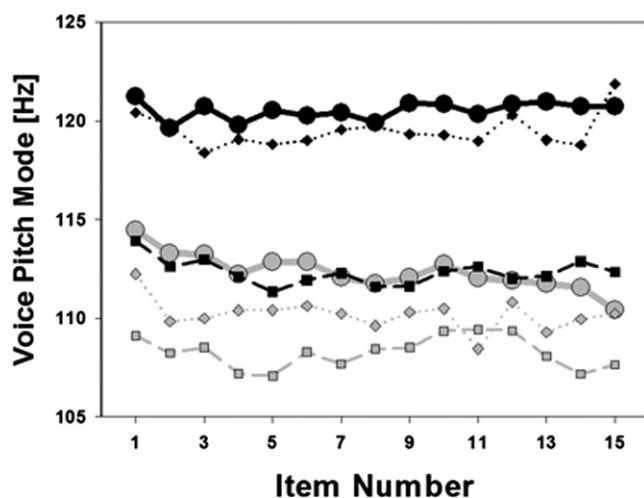
**Fig. 6.** Performance values of three clustered working style groups. White = OK, task correct; black with white stripes = OK, task wrong; white with grey dots = error, task correct; and checkered = error, task wrong.

The findings suggest that different time curves only occurred in flight (Fig. 7). Pre- and postflight these voice pitch differences were diminished.

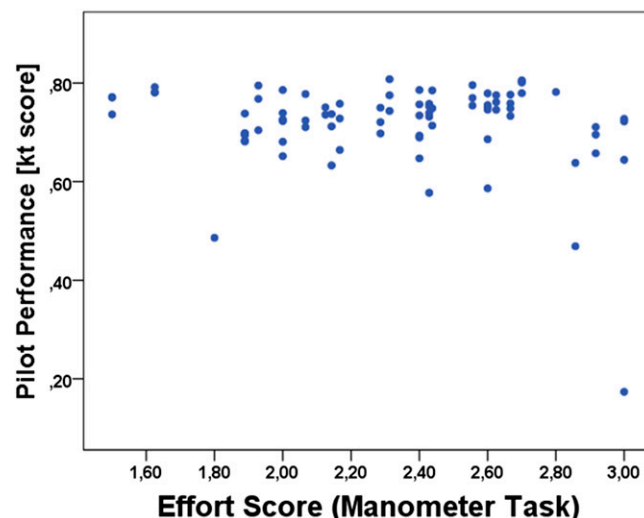
During the experiments, cosmonauts rarely switched between effort groups. An averaged value of the individual's group number can be taken as the strength of their affiliation to these groups, and thus as an indicator of effort during the Manometer task. Whereas Manometer performance was lower in the second group, the performance in the professional task, the docking maneuver, did not differ between groups. The effort score was tautologically highly correlated to the performance in the Manometer task ( $r = 0.474$ ;  $P < 0.001$ ), but not with performance in the docking task ( $r = -0.190$ ;  $P = 0.088$ ) as shown in Fig. 8. There was no mean difference in the pilot performance between effort groups.

## DISCUSSION

Over more than 20 yr, voice commands have been used in space experiments to control mentally challenging tasks that served as reference to the training of a mission-relevant operation, namely the hand-controlled docking of a spacecraft on a space station. The voice data contains scientifically valuable information. In the very first voice experiments in space, the investigators noted increases in voice pitch.<sup>35,39</sup> Our study, which comprises the largest dataset to date, confirms and extends the observation. A detailed analysis verified the validity of a nonlinear level model of voice changes<sup>17</sup> in space. The mentally nonchallenging counting from 1 to 10 provided anchor frequencies for a relaxed and awake state. Voice commands were used under mentally challenging conditions to provide respective anchor frequencies. The two voice commands did not differ in voice pitch frequency on the ground or in space. However, nearly all other voice



**Fig. 7.** Voice pitch changes in two effort groups. Black = higher motivated group, rhombus with dotted line = preflight, circle with solid line = in flight, square with dashed line = postflight; gray = less motivated group, rhombus with dotted line = preflight, circle with solid line = in flight, square with dashed line = postflight.



**Fig. 8.** Scatter plot of pilot performance vs. effort groups (assessed during the Manometer task).

parameters, including mean, median of pitch, jitter, shimmer, and formants (presented in Appendix A online; <https://doi.org/10.3357/AMHP.5351sd.2019>) differed and varied significantly. The finding was expected because both voice commands consist of three syllables (“vpo-ryad-ke” and “o-ship-ka”), with the longest and emphasized syllable with an “a” as main vocal in the first voice command and an “i” in the second one. These differences between the words were even more pronounced than the variations across the mission phases and insofar did not provide additional information. Even if these words were well standardized over the experiment, the different three syllables could still be considered as “free speech.” Therefore, we focused on the fundamental frequency mode.

The difference between the anchor frequencies at rest and during mental load decreased in space. The difference, while significant, was only modest (less 10%, except on Mir). In addition, the Mir experiments provided F0m data during docking training and demonstrated possible application for the anchor frequency model for the evaluation of “free” voice data during various activities. Similar results were obtained under space simulation conditions.<sup>15</sup> However, sporadically and occasionally, registrations of “free talks” with other crewmembers instead of the experimental voice commands suggest that “real” free talk among crewmembers provided higher frequency values than the voice commands during the Manometer task. The mentally challenging procedure of the Manometer task seems to only provide the lower voice pitch range during acceptable loads. Therefore, for the assessment of anchor frequencies of the upper level of normal and acceptable workloads, we recommend using the standardized ergometer physical stress test as previously done.<sup>17</sup>

In conclusion, our results are compatible with the step function hypothesis. Moreover, we speculate that voice pitch difference between ground and space may result from physiological changes such as altered autonomic nervous activation of the vocal tract rather than increased psychological load. The observed

decrease in the range between both anchor frequencies may point to hitherto unknown physiological mechanisms that deserve to be studied in more detail.

Performance analyses of the Manometer task differentiate cosmonauts with respect to their effort. The finding is supported by voice data. Approximately half of the cosmonauts tended to answer the dichotomous task stereotypically with “Okay.” This resulted in “good” performance in Okay tasks where the response was correct. The performance in Error tasks, however, was at the guess level and below, indicating that only a small amount of errors in the Error tasks was recognized. By contrast, the group with higher effort scores showed higher F0m values during all single tasks under in-flight conditions, whereas the other group reacted only to the first tasks of a new difficulty level. Postflight these differences were diminished.

Notably, we observed this effort effect in the Manometer task only, but not during the docking tasks. The discrepancy may be explained by the personal relevance attributed to these tasks (i.e., “game” vs. “professionally important task”). We interpret the higher effort in cosmonauts as a result of their higher willingness to fulfill the given task which, in turn, impacted the voice. The fact that the Manometer task was relatively often prematurely terminated supports this interpretation. Nevertheless, Error tasks were performed poorly. A similar response while monitoring instruments in real life could be fatal.

Summarizing, the analysis of voice pitch during a mentally loading experiment provided reliable and systematic data. Many space activities could be monitored with voice pitch analysis provided that relevant anchor frequencies are reliably assessed. For the accepted mentally or physically loaded state, another assessment method is required such as bicycle ergometer testing. The discrepancy in performance results between the Manometer task and the docking task are illuminating. We suggest that cosmonaut’s skills and state evaluations should be based on professional and mission-relevant tasks.

## ACKNOWLEDGMENTS

First we have to thank all the cosmonauts who participated in the experiment. We are thankful to the German Aerospace Center (DLR), acting as space agency, for the continuous support of the project (to the first author: DARA-Grants 50WB9128, 50WB93401, 50WB93401-ZA, DLR-Grants 50WB 96220; to KIE: 50WP0306, 50WP0501, 50WP0602, 50WP0603, 50WP1104, 50WP1304, 50WP1609).

Greatest thanks go in memory to Vyatcheslav P. Salnitski (deceased 2016), who pioneered the complex IBMP research on docking training.

*Authors and affiliations:* Bernd Johannes, Dr. Dipl. Psych., Sarah Piechowski, B.Sc., M.Sc. Psychology, Hans-Juergen Hoermann, Dipl.-Psych., Dr., Joern Rittweger, Prof. Dr. med., and Jens Jordan, Prof. Dr. med., Director, Institute of Aerospace Medicine, German Aerospace Center (DLR), Cologne, Germany; Sergey Vasilievich Bronnikov, Dr. eng., Rocket & Space Corporation “Energia”, Korolev City, Russia; and Yuri Akardevich Bubeev, Prof. Dr. med., Tatyana Ivanovna Kotrovskaya, Dr. med., and Daria Vladimirovna Schastlivtseva, Dr. eng., Division of Psychophysiology and Neurophysiology of Operator’s Activity, State Research Centre of the Russian Federation, Institute for Biomedical Problems (SCR IBMP), Moscow, Russia.

## REFERENCES

1. Backhaus K, Erichson B, Plinke W, Weiber R. *Multivariate Analysemethoden*, 8th ed. Berlin: Springer; 1996.
2. Borchert M, Dusterhoft A. Emotions in speech—experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments. *Proceedings of 2005 IEEE International Conference on Natural Language Processing and Knowledge Engineering, IEEE NLP-KE’05 2005*. Piscataway (NJ): IEEE; 2005:147–151.
3. Brenner M, Cash JR. Speech analysis as an index of alcohol intoxication - the Exxon Valdez accident. *Aviat Space Environ Med*. 1991; 62(9, Pt. 1): 893–898.
4. Busso C, Lee S, Narayanan S. Analysis of emotionally salient aspects of fundamental frequency for emotion detection. *IEEE Trans Audio Speech Lang Process*. 2009; 17(4):582–596.
5. Friedrich J, Vaic H. Sprachanalyse in der Luft-und Raumfahrtmedizin. *Z Militärmed*. 1978; 6:259–262.
6. Giddens CL, Barron KW, Byrd-Craven J, Clark KF, Winter AS. Vocal indices of stress: a review. *J Voice*. 2013; 27(3):390.e21–390.e29.
7. Hecker MHL, Stevens KN, Bismark GV, Williams CE. Manifestations of task-induced stress in the acoustical speech signal. *J Acoust Soc Am*. 1968; 44(4):993–1001.
8. Hess W. *Pitch determination of speech signals*. New York: Springer; 1983.
9. Hoehn-Saric R, McLeod DR. The peripheral sympathetic nervous system. Its role in normal and pathologic anxiety. *Psychiatr Clin North Am*. 1988; 11(2):375–386.
10. Hollien H. Vocal indicators of psychological stress. *Ann N Y Acad Sci*. 1980; 347(1):47–72.
11. Huckvale M. Data processing: analysis of speech audio signals. In: Muller B, editor. *Research methods in clinical linguistics and phonetics*. Hoboken (NJ): Wiley-Blackwell; 2012.
12. Johannes B, Gaillard A. A methodology to compensate for individual differences in psychophysiological assessment. *Biol Psychol*. 2014; 96:77–85.
13. Johannes B, Salnitski VP, Dudukin AV, Shevchenko LG, Shebuev AE, Bronnikov SV. Performance assessment in the experiment PILOT on-board space stations MIR and ISS. *Aerosp Med Hum Perform*. 2016; 87(6):534–544.
14. Johannes B, Salnitski VP, Gunga H-G, Kirsch K. Voice stress monitoring in space- possibilities and limits. *Aviat Space Environ Med*. 2000; 71 (9, Suppl.):A58–A65.
15. Johannes B, Salnitski VP, Haller H, Wilke D, Fischer F, Schlykova L. Comparison of voice stress reactivity under psychological stress test and simulated MIR docking maneuver. *J Gravit Physiol*. 1995; 2(1):P107–P108.
16. Johannes B, Salnitski VP, Petsch J, Karashtin VV, Kirsch K. Continuous voice-frequency monitoring of vocal outgoing communication during long-term confinement. *World Space Congress*, Oct. 10–19, 2002; Houston, TX, USA. Daytona Beach (FL): Embry-Riddle Aeronautical University; 2002.
17. Johannes B, Wittels P, Enne R, Eisinger G, Castro C, et al. Non-linear function model of voice pitch dependency on physical and mental load. *Eur J Appl Physiol*. 2007; 101(3):267–276.
18. Kirchhübel C, Howard D, Stedmon A. Acoustic correlates of speech when under stress: Research, methods and future directions. *Int J Speech Lang Law*. 2011; 18(1):75–98.
19. Koolagudi S, Sreenivasa Rao K. Emotion recognition from speech: a review. *Int J Speech Technol*. 2012; 15(2):99–117.
20. Kuroda I, Fujiwara O, Okamura N, Utsuki N. Method for determining pilot stress through analysis of voice communication. *Aviat Space Environ Med*. 1976; 47(5):528–533.
21. Levin H, Lord W. Speech pitch frequency as an emotional state indicator. *IEEE Trans Syst Man Cybern*. 1975; SMC-5(2):259–273.
22. Lieberman P. Perturbation in vocal pitch. *J Acoust Soc Am*. 1961; 33(5):597–603.
23. Lieberman P, Michaels SB. Some aspects of fundamental frequency, envelope amplitude and the emotional content of speech. *J Acoust Soc Am*. 1962; 34(7):922–927.

24. Lüdge W, Gips P. A new method for microprocessor-based real-time measurement of pitch detection on speech signals. In: Hofmann D, editor. *Intelligent measurement*. Budapest (Hungary): International Measurement Confederation; 1986.
25. Mohler JG. Quantification of dyspnea confirmed by voice pitch analysis. *Bull Eur Pathophysiol Respir*. 1982; 18(6):837–850.
26. Mundt JC, Snyder PJ, Cannizzaro MS, Chappie K, Gerals DS. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *J Neurolinguistics*. 2007; 20(1):50–64.
27. Nikonov AV, Vaic H, Johannes B. Historical aspects of voice analysis researches for estimation of the state of cosmonauts during flight. 10th International Symposium of Aerospace History Proceedings. Moscow: Iiet Ran; 1995:64–65.
28. Nilsson A. Speech characteristics as indicators of depressive illness. *Acta Psychiatr Scand*. 1988; 77(3):253–263.
29. Rubenstein L. Electro-acoustical measurement of vocal responses to limited stress. *Behav Res Ther*. 1966; 4(1-2):135–138.
30. Ruiz R, Legros C, Guell A. Voice analysis to predict the psychological or physical state of a speaker. *Aviat Space Environ Med*. 1990; 61(3): 266–271.
31. Salnitski VP, Bronnikov SV, Gorodetzki IG. Psychodiagnostic trainings complex for evaluation and prediction of the cosmonauts's professional activities reliability. *Pribori*. 1994; 4(94):23–28.
32. Salnitski VP, Bronnikov SV, Shevchenko LG. Investigation of questions of professional skill maintenance during space flight. XXXI Scientific letters, Internal works of K.E. Ziolkovski's creative labor, Section 9. Kaluga (Russia): State Museum of Cosmonautics History; 17–20 Sept. 1996:69–70.
33. Scherer KR. Vocal indicators of stress. In: Darby J, editor. *Speech evaluation in psychiatry*. New York: Grune & Stratton; 1982.
34. Scherer KR. Vocal communication of emotion: a review of research paradigms. *Speech Commun*. 2003; 40(1-2):227–256.
35. Sulc J, Remek V. Possibilities and limits of using speech signals in aviation and space psychophysiology. *Acta Neurobiol Exp (Warsz)*. 1986; 46(5–6): 347–352.
36. Titze IR, Baken RJ, Bozeman KW, Granqvist S, Henrich N, et al. Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *J Acoust Soc Am*. 2015; 137(5):3005–3007.
37. Tolkmitt FJ, Scherer KR. Effect of experimentally induced stress on vocal parameters. *J Exp Psychol Hum Percept Perform*. 1986; 12(3):302–313.
38. Vaic H, Friedrich J. Der Einfluß von physischer und mental-konzentrativer Belastung auf die Grundfrequenz der Sprache von Operateuren. Ein Beitrag zur Sprachanalyse in der Luft- und Raumfahrtmedizin. *Z Militärmed*. 1982; 1:26–31.
39. Vaic H, Friedrich J, Kolinichenko TB. Analyse des Flugfunkverkehrs als Beitrag zur Beurteilung der Arbeitsfähigkeit von Kosmonauten. *Z Militärmed*. 1981; 2:73–76.
40. Williams CE, Stevens KN. On determining the emotional state of pilots during flight. An exploratory study. *Aerosp Med*. 1969; 40(12):1369–1372.
41. Wittels P, Johannes B, Enne R, Kirsch K, Gunga HC. Voice monitoring to measure emotional load during short-time stress. *Eur J Appl Physiol*. 2002; 87(3):278–282.