Manned space vehicle multimodal dialog control: application of speech technologies and knowledge-based systems

V.A. Bondarev, Yu.N. Zhygulyovtsev, V.A. Soukhanov

Department of Automatic Control Systems for Flight Vehicles
Bauman Moscow State Technical University
ynzh@mail.ru

Abstract

The paper deals with introduction of additional channels of man-machine interaction (voice and gesture interaction) into manned space vehicle control systems. A concept is proposed for constructing an integrated intelligent control system with a multimodal human-machine interface using uniform models and mechanisms of presentation and use of knowledge about speech and control processes.

1. Introduction

An important factor in ensuring the safety and efficiency of manned space vehicle (MSV) flight control is a rational organization of interaction between crews and technical facilities. This determines the relevance of research and development works on creation of so-called multi-modal human-machine interfaces offering a more rational use of not only motor and visual abilities of a human, but also speech and hearing abilities, as well as sensory systems perceiving eye and head positions, gestures, etc.

Another way to improve control efficiency is due to intellectualization of control systems providing expansion of tasks the system can solve in an automatic or semi-automatic, supervisory mode, where the user can simply select and specify system behavior or maneuver, which is then performed automatically. To do this, the control system must have special knowledge of the domain and rules for its use, which, on the basis of analysis of the control object state allows to generate control actions or to provide advice on possible actions for the spaceman. The theoretical basis for creation of such systems is the theory of automatic control and artificial intelligence techniques, in particular, technologies of knowledge-based systems.

2. Speech technologies for control purposes

Traditionally, control actions are formed by means of manual control – handles, joysticks, switches, buttons, etc. Information on the object state and control processes is provided to spacemen through a visual analyzer using visual surveillance data and various indicators, the last become integrated textual and graphic ones in recent years. The volume of transmitted and received data is often close to the limit, and sometimes exceeds the physiological capabilities of the spaceman. Therefore, the involvement of additional channels to redistribute information flows will reduce this overload and improve the comfort and efficiency of control processes.

The basis for a more effective use of acoustic channels are speech technologies, a complex of scientific and technical methods and tools that implement all the variety of options of man-machine communication using speech signals. Almost the entire range of speech technologies can be applied to ensure the interoperability of the space vehicle crew with the onboard systems.

Speech recognition allows generating voice commands and help requests, to document experiment observations and control processes converted to a textual form.

Speech synthesis ensures a speech dialog mode that allows to control the recognition results, to receive information and warning voice messages.

An important advantage of the speech dialogue is lack of necessity to move around remote controls of various subsystems for access to controls and displays.

Speaker recognition provides the possibility to determine who of the crew member is speaking (speaker identification), which is important to maintain the dialogue, as well as to control the right to access to certain subsystems, possibly with password confirmation and speaker verification.

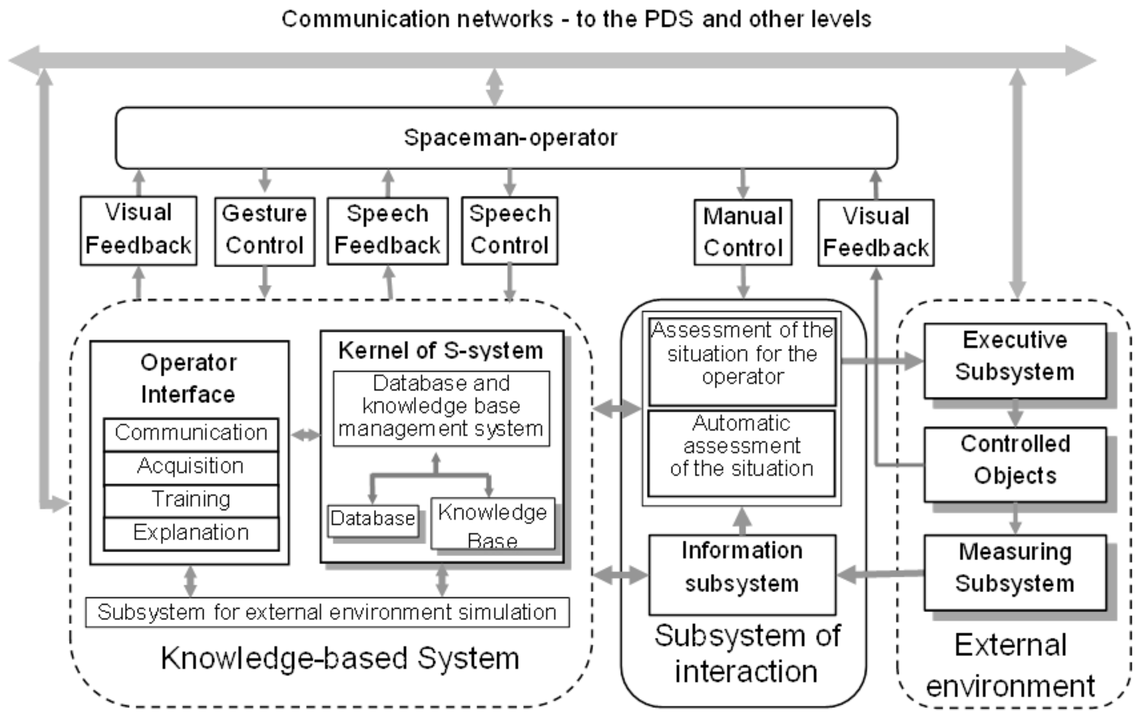
Information on the psychophysiological and emotional state of the spacemen and their performance can be obtained by analyzing voice characteristics.

An important component in the organization of verbal interaction of the crew with the onboard systems is the speech signals input system. When constructing it, it is advisable to apply noise reduction methods based on multi-microphone spatial configurations providing the possibility of speaker detection and localization, as well as tracking their movements. This allows to form the direction pattern providing suppression of background noise, including speech interference, coming from other crew members and loudspeaker communication.

These technologies are elaborated to the extent that suggests the possibility of their introduction into applied systems. However, some issues of the integrated application of several technologies in complex control systems require further theoretical and experimental studies.

3. Architecture of the onboard knowledge-based control system

The introduction of additional modalities (e.g. speech and gesture ones) to the control system requires their integration into the system with traditional (manual and automatic) control. The figure presents a general diagram showing the control system architecture built on technologies of knowledge-based systems (S-systems).



The S-system is comprised of the following: a kernel; spaceman-operator interface and environment simulation subsystem. The kernel of the S-system is a database, a knowledge base and a database and knowledge control system.

Current data on the state of the object and control processes are contained in the database. The knowledge base is a part of the S-system, which includes formal sources of knowledge presented in the form of production rules.

General information on how to apply this knowledge, special mechanisms of knowledge formation, processing and control are localized in the database and knowledge control system. It uses Petri nets controlled by productions. This unit allows to create a hierarchical/heterarchical system for heterogeneous sources of knowledge (including knowledge of speech, objects and control processes, etc.) in a uniform representation with the possibility of its verification, later updating and acquisition.

The spaceman-operator interface includes a range of sub-systems (some of which are intended for use in the mode of constructing a system of knowledge sources) that provide interaction of the spaceman with S-system components and through this system with the control objects.

The subsystem of interaction with the environment implements functions of collection, primary processing and conversion of information to a form suitable for further processing. Its structure includes the components describing the situation for the spaceman-operator, and automatic description and assessment of the control object state for automatic control loops. In general, these two components are part of the executive subsystem, ensuring both automatic and manual gesture and speech control modes.

The environment is defined here as a manned space vehicle with its many onboard systems, each of which can be represented as a loop including the executive subsystem, the control object and measurement subsystem. The loop is closed through the interaction subsystem, the S-system and the operator.

4. Establishment of the speech dialogue control

The dialog interaction implies a two-way exchange of voice messages between the operator and the control system. The nature of these messages is determined by the tasks solved in the control process.

In addition to voice commands prescribing the system to perform certain actions, and voice messages confirming receipt and execution of commands, the system can generate information and diagnostic messages displaying the state of the object subsystems and control processes. The messages can be initiated both by the operator’s and the system request, for example, in cases of deviations of the object subsystem state from the values provided by the domain model. Therefore, the voice subsystem can be called “Speech Information Management System”, abbreviated RIUS.

Consider operation modes of the speech control subsystem

Modes implemented by the operator’s voice messages:

1) Activation and deactivation of the voice subsystem

2) Selection of the operation mode

3) Recognition learning and additional learning

4) Control of the recognition subsystem

5) Formation of speech commands

6) Validation of the output commands (based on the analysis of the system’s understanding of commands)

7) Output of information requests

Modes implemented by the system:

1) Recognition of voice messages

2) Understanding of received messages taking into account the knowledge sources system

3) Execution of recognized commands and requests from the operator

4) Analysis of the control object state based on the domain model

5) Formation and synthesis of voice information and diagnostic messages

6) Changing the control modes: - speech - gesture - manual

7) Control together with other modalities

The vocabulary of the speech subsystem must contain names of commands, modes, parameters, and values used in the implementation of control processes. Words are distributed by subgroups according to the order of commands formation.

At each step of the dialog control in accordance with the domain model and dialogue scenarios, only part of the words are recognized (from one to twenty) which increases the reliability of recognition. The reliability of control is further provided by syntactic and pragmatic control of commands formation implemented by the network model of the dialog control.

Consider possible scenarios of voice commands application with a few examples. The RIUS starts in the activation waiting mode simultaneously with the other control object subsystems. The signal for activation is the utterance of the key phrase “СИРИУС начало” (“SIRIUS beginning”) recognized with high reliability and for any speaker, as well as a group of commands to select the operation mode. The sign of the system functioning is a voice message confirming its working condition. In addition, the RIUS messages are duplicated on the display of visual feedback.

After the activation, the system switches to selection of the operation mode, waiting for the operator’s utterance of the standard format “Mode <mode name> Operator<operator number/name>”. In the process of pronunciation, the recognized words are displayed on the display of visual feedback, and at the end of pronunciation the utterances are repeated by voice, and the operator can give the command “execute”. If there are no standards for the named operator in the system, the system offers to switch to a learn mode, which starts after the operator’s confirmation.

Switching to the system deactivation mode, in which the only command expected is the one to exit from the deactivation mode, is executed by the command “СИРИУС стоп” (“SIRIUS stop”). Resuming the mode, from which the RIUS was switched from to the deactivation mode, occurs after the utterance “СИРИУС старт” (“SIIRIUS start”). No actions can be performed in the deactivation mode.

“Control” mode

Depending on the complexity of the control object, this mode may have several submodes. For example, for extra-vehicular activity in a suit, it can be modes for parameters setting or requesting parameter values for life-supporting systems, lighting on/off status, communications, videocameras (if any), etc.

The formats of control commands are determined by the importance level of appropriate actions. For example, requests for information, switching non-critical subsystems on/off are performed without confirmation, while potentially dangerous actions must be executed with informing and confirmation.

When giving commands consisting of several words, at any time, one can cancel the process of command formation and return to the original status using the “cancel” command.

Confirmation mode

In the process of recognition the speech recognition subsystem assesses the credibility of recognition, and, moreover, can provide a signal to abandon the recognition. This assessment can be used to determine the need for confirmation of the command or its retry. Responsible commands must be executed with a confirmation and request-repeating, maybe even more than once. The support for disputable situations arising from possible errors of recognition is provided by the RIUS in the validation mode for generated commands.

Switching to the confirmation mode occurs automatically when a signal is received on the refusal of the speech recognition subsystem, with a lack of recognition reliability, while giving a potentially dangerous command. The RIUS plays a voice message informing about its switching to the confirmation mode. At any time, one may exit from this mode using the “cancel” command; at this the command confirmed is not executed.

If the credibility level of the spoken word is below the threshold value or when a potentially dangerous command must be confirmed, the system reproduces the operator’s word or command, which has been hypothetically pronounced, and the operator confirms the hypothesis by the “execute” command or rejects it by the “cancel” command.

If a signal was received on the refusal of recognition, then the RIUS requests of the operator to repeat the word; if the operator cancels the repetition by the “cancel” command, then the RIUS returns to the original mode in which it has been before the word was pronounced.

5. Discussion

For general considerations, there are a number of modes in which efficient use of resources is possible, expanding the range of human operator's sensory-effector resources used. In the first place these are modes when traditional control resources involved are close to the limits of human capabilities, i.e. the vision is focused on monitoring the control process, and the motor system is focused on the control actions using the controls. This includes piloting control, robotic manipulator control, tracking moving objects in the tracking and guidance mode, etc.

In scientific experiments conducted aboard manned space vehicles, optical observation facilities are often used to investigate processes; in this case, the voice channel may be very effectively used for recording and documenting the results of observations, as well as requests for help and other information that is not available without diverting from the observations.

The results of the research of subsystems' operation speed with various modalities allow to determine the possibility of using certain subsystems for specific tasks. The time of motor reactions is several times less than the reaction time of the speech recognition subsystem, since the latter cannot be less than the duration of command pronunciation plus the maximum duration of a voiceless closure, which is a sign of the command end – this is about 1..1,5 seconds. Therefore, manual or gesture control must be applied in modes requiring fast (fraction of a second) reaction, while speech data input may be used for input of rather complicated tasks in the supervisory mode to control automated subsystems. However, in the case where values of time constants of the controlled process equal to a second or more, the mode of direct motion control is also possible, which was confirmed experimentally [1].

In any case, gesture and verbal subsystems both individually and together provide additional channels of interaction between the human operator and the technical facilities, which enables the construction of more efficient man-machine control systems. Manual, voice and gesture control mutually overlap allowing to address the full range of tasks attributed to them in a manner most convenient or available at any given moment. But it is necessary to ensure the consistency of the control actions of all three interoperable systems, that is, to establish an arbitration process, taking into account the priorities of the subsystems and correctness of generated control commands.

To identify possible areas of application of emerging technologies, it is necessary to conduct a comprehensive survey of the domain area in spacemen's operator activity aboard the MSV and extravehicular activity, as well as in the development of MSV control systems. This requires comprehensive consulting with experts in these fields with demonstration of the capabilities offered by multimodal control systems, and collection of advice (recommendations) and opinions on the feasibility and usages of the proposed solutions.

It is necessary to conduct complex research and large-scale development to improve technical and ergonomic solutions of sensory and effector subsystems regarding improvement of their accuracy, stability, manufacturability, and use of modern components. These include the problem of improving the design and scheme-technical solutions for sensor gloves, electroacoustic input-output subsystems, and subsystems displaying visual information.

Improvement of the RIUS algorithms and software involves the development of methods for recognition of connected words, which requires the formation of a speech database consisting of a representative sample of speech signals implementations in Russian and English languages obtained from many speakers in conditions closest possible to real ones, as well as phonetic and prosodic marking needed to implement the methods of automatic formation of the speech elements vocabulary and restoration of the system of knowledge sources related to speech. The most promising option is conduction of an experiment to collect speech material aboard a real space vehicle, such as the International Space Station. At the preliminary stages it is possible to accumulate experimental data during spacemen's preflight training.

In the future, it is advisable to conduct more extensive research, which require creation of a complex hardware-in-the-loop (HIL) simulation of spacemen's intra-and extravehicular activity.

6. Conclusions

As a result, the proposed approaches increase the reliability of the control system performance due to the monitoring of the operator's actions, performed by the system of knowledge sources embedded in the domain model, and also due to the possibility to correct generated commands and mechanisms of their correctness confirmation implemented in interoperability scenarios.

The joint or alternate use of interface subsystems of various modalities increases the flexibility and variety of the control and information modes' implementation that improves the efficiency of the operator's activity due to a more efficient load distribution in sensor and effector channels of the human operator.

7. References

[1] Soukhanov V.A., Zhygulyovtsev Yu.N. The study of speech dialogue control methods in MHTS –Bauman Moscow State Technical University. Speech Technologies, №1, 2008, pp.78-82 (in Russian).