Computing and Informatics, Vol. 36, 2017, 908–924, doi: 10.4149/cai_2017_4_908

MULTIMODAL SUBSCRIBER INTERFACES FOR INFOCOMMUNICATION SYSTEMS

Oleg Basov, Irina Kipyatkova, Anton Saveliev

St. Petersburg Institute of Informatics and Automation
of the Russian Academy of Sciences
14th *Line 39*
199178 St. Petersburg, Russia
e-mail: oobasov@mail.ru, {kipyatkova, saveliev}@iias.spb.su

Abstract. The quality enhancement of infocommunication systems based on integration of semantically different information transmitted over different communication channels is discussed. The increase of the informational efficiency of the multimodal communication systems is provided by using multimodal subscriber interfaces, compared to traditional telecommunications. The proposed approach is new in the subject area and requires a variety of research studies to develop innovative elements of the theory of multimodal information transmission. A notion of information value is introduced in order to estimate and chose functional structures and parameters that possess the most favorable informational characteristics during development of multimodal communication systems.

Keywords: Multimodal interfaces, information processing, trust communication, signal processing, telecommunication

Mathematics Subject Classification 2010: 68U35

1 INTRODUCTION

During their development the national information infrastructure of developed countries has evolved from a highly specialized communication networks (telegraph networks, telephone networks, etc.) into, first, telecommunication network, and then into infocommunication systems. In most classical bibliographic resources of the subject area and in the directive documents of the branch "Communication and Telecommunications", communication service or information service is understood as a result of functioning of information infrastructure of any level [1, 2, 3]. However, in recent years, there is a tendency to depersonalize certain infocommunication services a subscriber often requests a type of service called "a connection to network". This service implies an opportunity for a subscriber to obtain available or most convenient, in subscriber's opinion, way of interaction. The concept of "convenience" when realizing the service "a connection to network" is defined by users in accordance with their personal preferences and physical requirements, along with the properties of the environment in which communication occurs. Therefore, it is vital to consider the process of interaction between subscribers through infocommunication system (ICS) from the standpoint of maximizing convenience of traditional interpersonal communication when people communicate directly with each other [4, 5].

There is a wide spectrum of communication services, which used audio channel for speech transmission and processing [6]. However, during traditional interpersonal communication, people are almost always interacting multimodally, using verbal and nonverbal channels. The analysis of the existing multimodal interfaces, their main characteristics and fields of application, as well as the results of common research in the field of multimodal interaction and interfaces design enabled us to draw a conclusion about the possibility and necessity to make interaction between subscribers through infocommunication system closer to traditional interpersonal communication. Even nowadays, there are objective prerequisites to reject principles of communicated information division into communication services and to implement polymodal infocommunication systems (PICS). The latter refers to an interconnected set of processing (multimodal user terminals) and data storage systems, telecommunication networks, uniting them, operating under single control in order to collect, process, store, protect, transmit and distribute, display and use multimodal information that takes into account the meaning of reported messages, subscribers (users) identity, their mood, physiological and emotional state [7, 8]. Obviously, the realization of such revolutionary ideas is impossible without comprehensive theoretical and experimental research. The aim of this work is to theoretically prove the necessity of transition to PICS.

The paper below describes modern approaches to development of multimodal/polymodal telecommunication systems and services in Section 2, as well as presents main principles of PICS synthesis in Section 3. Increasing the informational efficiency of PICS with using several modalities is discussed in Section 4. The criterion of value of multimodal information is introduced and used for estimation the proposed approach in Section 5.

2 RELATED WORKS

Functioning of any ICS is significantly related to receiving and processing information, because without this it is impossible to take the necessary decision about the realization of a certain service, and therefore to carry out the action needed (transmission of information to the interlocutor), which is the ultimate goal of such systems functioning. During this research, the analytical review of modern approaches to development of multimodal/polymodal telecommunication systems and services was carried out.

In [9] a multimodal contact list is proposed. Multimodal contact list is an application allowing user to display certain context information (for example, happiness or availability) to people which are in users contact list and communicate both verbally and by touch. The application can prevent undesirable calls by determining if the contact is available for a phone call. Also the application can filter the contacts according to some parameters, for example, it allows the user to choose the person in contact list which is happy and not busy. In [10] a crisis based human like multimodal system is proposed. The aim of the system is to help people in crisis situation. For interaction with the system the user can use visual set of crisis icons, text messages, pen input, photos and direct speech.

Multimodal platform for e-learning which allows multimodal access to an educational social network is presented in [11]. The platform enables the retrieval, editing and sharing of multimedia contents in an educational social network using different modalities. The platform is based on client-server architecture.

In [12] an approach for developing multimodal mobile user interface using thin clients for building high modular client interfaces, managing asynchronous interactions. Authors introduced the concept of multimodal objects which are able to collect vocal, sketch and stylus inputs and to render multimedia output (text, graphics, audio and video). An approach based on the multimodal objects that are able to use the benefits of both telecommunication and web protocols to manage input/output processes is proposed for building mobile multimodal clients.

World Wide Web Consortiums (W3C) Multimodal Architecture and Interfaces standard, an architecture and communications protocol for integrating different modalities into multimodal application are presented in [13]. This architecture simplifies integrating components from multiple sources.

In [14] an approach of integrating session initial protocol (SIP) in Converged Multimodal/Multimedia Communication Services is presented. The paper describes the structure of the VoIPTeleserver for VoIP in SIP environment which is based on dialogue system and Web convergence.

Automatic systems for coding verbal and gestural communication of TV interviews and debates in [15] were used for analysis of communicative competences of politicians.

3 MAIN PRINCIPLES OF POLYMODAL INFOCOMMUNICATION SYSTEM

Traditional tasks of telecommunication systems to increase the efficiency of using the communication channels, as well as to increase the speed and quality of messages transmission via these channels, can be solved with the help of information theory [16]. As for the newly emerging challenges related to multimodal representation and processing of information, they become the subject matter of the applied information theory [17, 18]. This theory shows the effectiveness (informational aspect) of PICS [19].

Prerequisites for the application of information theory to the synthesis (designing) of PICS are due to the fact that in some cases, systems work can be represented in terms of the concept of choice. In fact, the set of values of the parameters describing the state of PICS is represented by the set of points (Performance, Speed, Integrity), which occupies in the multidimensional parameter space volume V^{PICS} (Figure 1). The movement of displaying points inside this hypothetical volume, in this case, will express a change in system parameters (performance, speed, integrity). The task of the design system ultimately comes to limiting a diversity of the possible parameters values up to those shown by some private volume V^{PICS}, lying within the volume V^{PICS}.

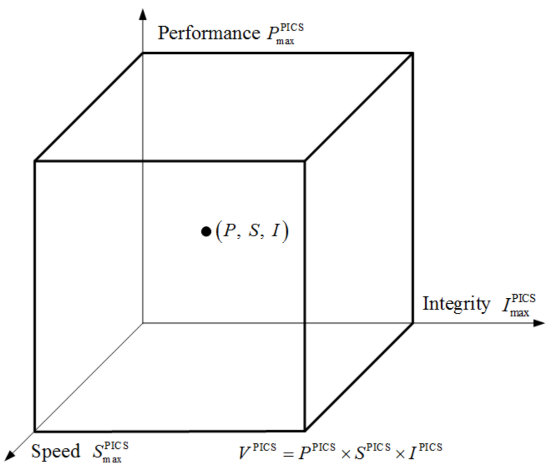


Figure 1. Effectiveness of multimodal infocommunication systems

The minimum necessary amount of information I_0 which must be received and processed by the system to select particular volume V_*^{PICS} from the whole a priori volume V^{PICS}, is equal to entropy of removing corresponding uncertainty (with equiprobable distribution of private volumes):

$$I_0 = \log_2 \left(V^{PICS} / V_*^{PICS} \right) . \tag{1}$$

To accomplish tasks by a system, it is necessary to process the amount of information in a time not exceeding some value. The minimum required system capacity (speed of obtaining and processing information) must be

$$C_0 = \frac{I_0}{T} = \frac{1}{T} \log_2 \left(V^{PICS} / V_*^{PICS} \right). \tag{2}$$

An efficient system must have redundancy

$$\delta = \frac{C - C_0}{C_0} = \frac{CT}{\log_2 \left(V^{PICS} / V_*^{PICS} \right)} - 1 = \frac{I_{max}}{I_0} - 1 \tag{3}$$

where I_{max} is the maximum amount of information that the system asymptotically can process in a time T. The expression (3) contains fundamental for systems operation indicators such as the complexity of a task (to some extent being expressed by the value of the a priori volume V^{PICS}) *the quality of the PICS functioning* (volume V^{PICS}) and *the speed* (time T). Thus, the redundancy represents *informational expression of the PICS effectiveness* and allows investigating such systems, based on their informational aspect.

4 INCREASING THE INFORMATIONAL EFFICIENCY OF POLYMODAL INFOCOMMUNICATION SYSTEMS

In polymodal ICS, user terminals perform functions for obtaining and processing information, so there are two ways to increase the information efficiency criterion:

- use of interrelations between subsystems of a multimodal ICS user terminal, the so-called integration;
- practical user terminal architecture.

The basis of the first way is that subsystems of a complex, interacting between themselves and with a given system (a user terminal), receive and provide information being a priori for the latter, i.e., reducing the entropy of a task that needs to be solved. So, a consistent decrease of a priori uncertainty can be seen (Figure 2) in multimodal user terminals [20, 21, 22]:

- during the transition from a multitude of natural signals $NS = \{ns_a\}$ to a multitude of artificial signals $AS = \{as_q\}$ that is associated with entry procedures (limitation of dynamic and frequency range of signals, their discretization and quantization) and signal preprocessing (localization, filtering, denoising);
- when separating artificial signals $AS = \{as_q\}$ into a given set of input modalities (modality signals) $IM = \{IM_1, IM_2, \ldots, IM_{N_{IM}}\}$;
- when extracting from input modalities IM their parameters $F = \{f_1, f_2, \ldots, f_{n_p}\}$;
- during the implementation of polymodal services based on the fusion of the given (for each particular service) parameters of the modalities.

At each stage, there is a reduction of the choice uncertainty of the volume V_δ^{INF} within which values of the subsystem parameters are being expected. Resulting

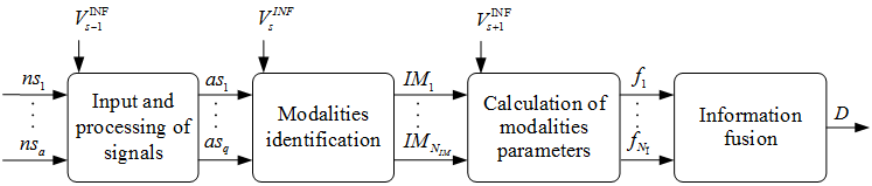


Figure 2. Use of integration when constructing polymodal ICS subscriber terminals

minimization of the value C_0 up to the value $C_{0\Delta} < C_0$ gives optional redundancy. Further decrease in the required capacity of a subscriber terminal can be achieved with a multi-channel subscriber terminal architecture.

In the context of information processing in separate communication channels of multimodal interfaces, such rational construction is achieved by the fission of signals $AS = \{as_q\}$ into modalities (Figure 3), and extraction of features F from the modalities.

In fact, if for normal functioning of the system it is necessary to receive information about N_p its parameters, then the total amount of information needed to be processed in all N_p informative channels should not be less than the joint entropy of all these parameters.

Since these parameters are related to the same physical object (modality), signals carrying information about the parameters are, to a greater or lesser extent, correlated. Therefore, the joint entropy H_0 of the signals, carrying the information about combination of parameters, is less than the sum of the entropies of these signals H_Σ.

Hence, the minimum required system capacity

$$C_0 = I_0/T = H_0/T \tag{4}$$

will be less than

$$C_\Sigma = I_\Sigma/T. \tag{5}$$

The difference between H_0 and $H\Sigma$ is due to the cross-correlation relationships between the various parameters that define operation of the multimodal user terminal. Minimum required capacity value of a N_p-channel system [23]:

$$C_{0N_\Pi} = C_\Sigma + C_\Delta \tag{6}$$

where the value

$$C_\Delta = \frac{1}{2T}\log_2 |r_c| \tag{7}$$

represents that particular decrease in the necessary user terminal capacity, which is obtained by a terminal rational architecture that takes into account the statistics of the signal ensemble, compared to a multi-channel system built without considering this statistics.

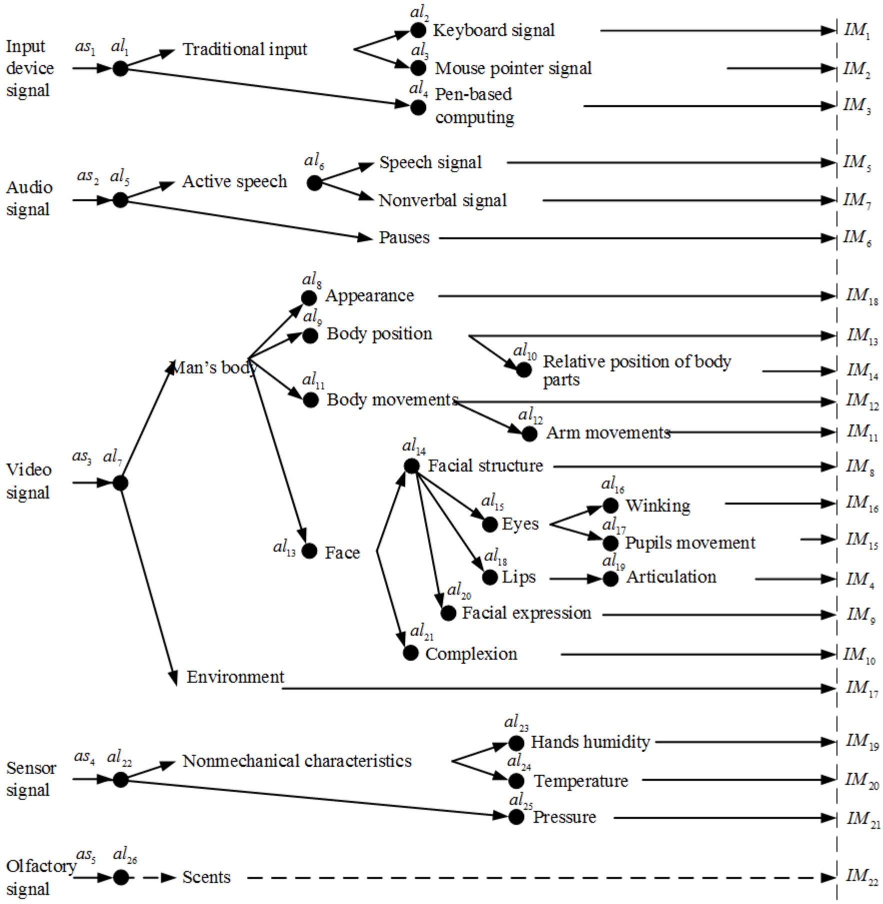


Figure 3. Graphs of the modalities of different communication channels: $\{al_1, \ldots, al_{26}\}$ – a subset of the algorithms for identifying modalities

Decrease in the required capacity C_{0m}, i.e. increase in C_Δ, is determined by the value of the correlation signal determinant $|r_c|$. Therefore, the analysis of the correlation structure of multimodal interfaces is reduced to analyzing this determinant and determining the conditions under which its value will be the lowest. The elements of the correlation matrices determinants are the cross-correlation moments between all the sources (signals AS, modalities IM or their parameters F), which are a basis of a solution to the problem (problems) of multimodal ICS synthesis.

Let us consider two independent discrete sources with the given ensembles of messages in the form of random variables A_1 and A_2, each of which takes values from the ensembles $\{A_{11}, A_{12}, \ldots, A_{1i}, \ldots, A_{1N}\}$ and $\{A_{21}, A_{22}, \ldots, A_{2i}, \ldots, A_{2N}\}$ with probabilities $p(A_1)$ and $p(A_2)$, respectively (N is the number of possible messages at the outputs of the sources. We assume that the random variables A_1 and A_2 are estimated on finite intervals, so we define for them some k^{th} interval $(k, k+N_{INT}-1)$, where the indices i, j take N_{INT} values: $k, k+1, \dots, k+N_{INT}-1$.

On the analogy of [24], by random correlation time of independent discrete random variables A_1 and A_2 on the interval $(k, k+N_{INT}-1)$ or by the k^{th} interval correlation coefficient, we will understand the value

$$r_{ckA_1A_2} = \frac{\left(\overrightarrow{m}^*_{kA_1} - \overrightarrow{m}_{A_1}\right)\left(\overrightarrow{m}^*_{kA_2} - \overrightarrow{m}_{A_2}\right)}{\sigma^*_{kA_1}\sigma^*_{kA_2}} \tag{8}$$

where $\overrightarrow{m}_{A_1}$, $\overrightarrow{m}_{A_2}$ are the mathematical expectation of variables A_1 and A_2, respectively; $\overrightarrow{m}^*_{kA_1}$, $\overrightarrow{m}^*_{kA_2}$ and $\sigma^*_{kA_1}\sigma^*_{kA_2}$ are sample mean and sample mean quadratic deviations of random variables A_1 and A_2 on the interval $(k, k+N_{INT}-1)$, respectively.

The expression (8) shows that if for each random variable its sample mean at the specified interval does not coincide with its mathematical expectation, correlation coefficient of these random variables on the given interval is nonzero. Therefore, on relatively small sets of values, which independent random values A_1 and A_2 can take, there may occur correlation dependencies between them. Similarly, on a limited set of messages generated by independent sources, there may be correlation dependencies between them.

In [24], experimentally, (under the assumption that each source message appears equally probable with the further averaging of results), graphical representation of dependencies of the average sample correlation coefficient for two independent sources of video and voice information on the parameter N_{INT} were obtained (Figure 4). As the objects of correlation analysis, the sources of speech messages $as_2\left(\overrightarrow{m}_{A_1} = \overrightarrow{m}_{A_2} = \overrightarrow{0}\right)$ and static model images $as_3\left(\overrightarrow{m}_{A_2} = \overrightarrow{m}_{A_2} = 128 \cdot \overrightarrow{1}\right)$ in RGB format were used.

The analysis of the represented dependences (Figure 4) shows that with the increase of N_{INT}, correlation time of the sources decreases exponentially, and at $N_{INT} = N$ – the cross-correlation between messages from the sources is practically absent (the value $\overline{r}_{kA_1A_2}$ is practically equal to zero). The maximum slope corresponds to sources with maximum redundancy, in the case under consideration, – to video images sources.

During the transition from artificial signals to their modalities IM, there is further decrease in correlation moments of the sources, and hence decrease in the required capacity C_0 of the user terminal.

Thus, multimodal representation and processing of input information can improve the information criterion of the polymodal ICS effectiveness. The decorrelation of the input signals inevitably leads to the decrease of gain in the total transmission speed, with joint encoding of the sources, between which there are correlation relationships [16]. Therefore, in multimodal subscriber terminals, it is advisable to use separate coding of parameters of single modalities and/or results of the multimodal services implementation.

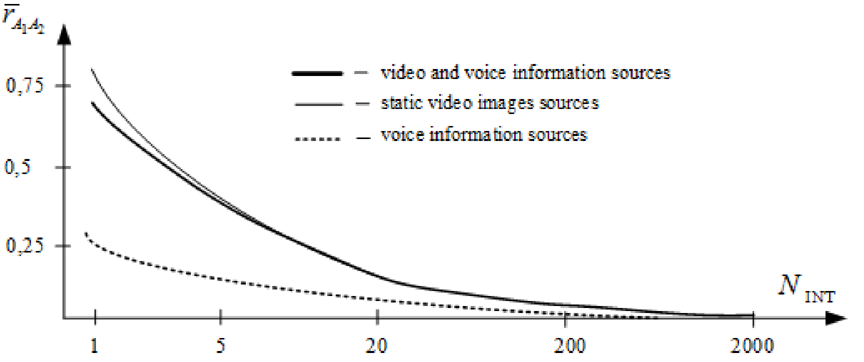


Figure 4. The N_{INT} dependencies of the average sample correlation coefficient for two independent sources [24]

5 MULTIMODAL INFORMATION VALUE ESTIMATION

Processing of a certain minimum necessary amount of information, received from a subscriber, is a prerequisite for obtaining the given results. One of the simplest models it may happen that one (Figure 5), in which the objective X (service realization of X) can be attained in various ways x_j ($1 \leq j \leq n$), where total number n is known; at this, the probabilities of achieving the aim by different ways $\varphi(x_j)$ are a priori unknown.

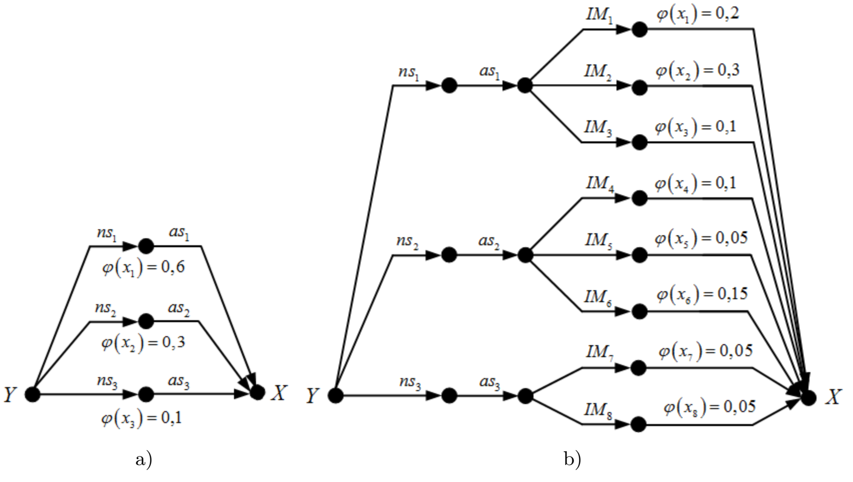


Figure 5. Ways of achieving the objective of functioning in the a) existing and b) multimodal ICS

In order to solve the problem in the most efficient way, the information obtained should allow determining the values $\varphi(x_j)$ on the paths x_j and choosing the path (paths), where the value $\varphi(x_j)$ is the largest. The more valuable the information I_i from the source Y is, the greater the likelihood of attaining the objective. It should be noted that even without receiving the information, the likelihood of achieving the objective $\Phi(X)$ is not equal to zero, because even for a random (arbitrary) choice of the path x_j, the objective can be achieved. Based on the above stated, as a *measure of sensitivity*, we take the value of difference Z_i between the probabilities of achieving the objective in the absence of information $\Phi(X)$ and upon receipt of the i^{th} message:

$$Z_i = \Phi_i(X) - \Phi(X). \tag{9}$$

The value Z_i can theoretically be negative, which reflects the receipt of false information (disinformation). Hereafter, we assume that the information received is true, and then always $Z_i > 0$.

The probability of attaining the objective at a random choice of one of the paths x_j is

$$\Phi(X) = \sum_{j=1}^{n} P_j \varphi(x_j) \tag{10}$$

where P_j is the probability of choosing the j^{th} path x_j.

Before receiving information, this value has some a priori distribution. It is natural to assume that in the total absence of a priori information, the choice of different ways is equiprobable, then

$$P_j = \frac{1}{n} = \text{const}; \quad \Phi(X) = \frac{1}{n} \sum_{j=1}^{n} \varphi(x_j), \tag{11}$$

i.e., the probability of achieving the objective in the absence of information equals the average value $\varphi(x_j)$ over all n paths.

Receiving information I_{ij} from the i^{th} message about j^{th} path changes the probability distribution P_j. This posterior distribution is largely determined by subjective factors that appear in the decision taken by a man or put into the behavior of the technical system. Therefore, complete formalization of the system behavior at obtaining information is difficult. However, in some typical cases one can observe some general trends of changing distribution P_j.

1. The message I_{ij} provides information about the value $\varphi(x_j)$ only at one j^{th} path and does not contain any information about the other $(n-1)$ paths; therefore, the choice of any of these $(n-1)$ paths is equiprobable. Then

$$\Phi_i^I(X) = P_j \varphi(x_j) + \frac{1 - P_j}{n-1} \left[\sum_{k=1}^{j-1} \varphi(x_k) + \sum_{k=j+1}^{n} \varphi(x_k) \right]. \tag{12}$$

It is natural to assume that, in this case, the higher the probability $\varphi(x_j)$, which can be expected to achieve the objective X, based on the information received from the source y_i, the more probable the choice of the path x_j is. P_j – dependence on $\varphi(x_j)$ is a non-decreasing function, which lies within the range $[0, 1]$. If we take for its approximation a function convenient for calculations

$$P_j = [\varphi(x_j)]^r \, (r > 0), \tag{13}$$

a characteristic of the values of the i^{th} message about the j^{th} path is obtained in the form

$$Z_{ij}^I = [\varphi(x_j)]^{r+1} + \frac{1 - [\varphi(x_j)]^r}{n-1} \left[\sum_{k=1}^{j-1} \varphi(x_k) + \sum_{k=j+1}^{n} \varphi(x_k) \right] - \frac{1}{n} \sum_{k=1}^{n} \varphi(x_k). \tag{14}$$

Even if on the j^{th} path $\varphi(x_j) = 0$, then even in this case $Z_i > 0$, since the probability of choosing this path $P_j = 0$, and, therefore, this a priori unproductive path is excluded from the future behavior of the system. The case (Figure 6 a)) is specific for the existing unimodal ICS, where service implementation involves processing and transmitting messages of one type (audio, video, or data)[1].

2. The message I_i provides information about the values $\varphi(x_j)$ on all n paths. In this case, it is necessary to choose that path for which $\varphi(x_j)$ s a maximum value, then $\Phi_i^{II}(X) = \varphi_m ax(x_j)$.

   In this respect, a characteristic of the information value will be

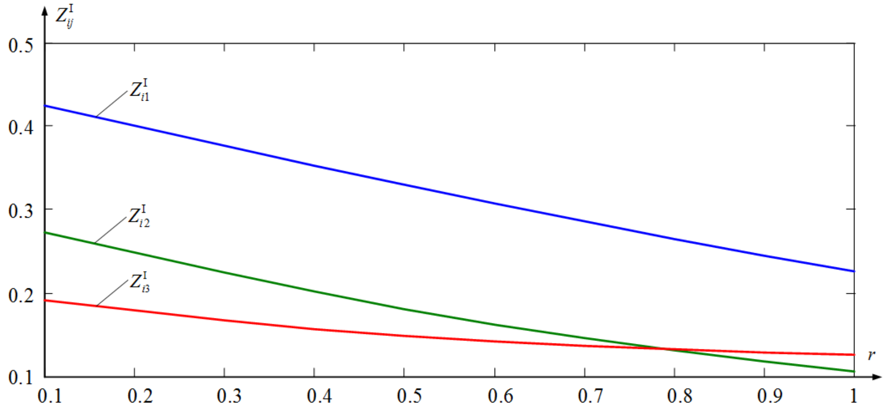
$$Z_i^{II}(X) = \varphi_{max}(x_j) - \frac{1}{n} \sum_{k=1}^{n} \varphi(x_k). \tag{15}$$

   The case (Figure 6 b)) is common to multiservice ICS, in which heterogeneous data stream is processed and transmitted, however, there are no mechanisms that can clearly define values $\varphi(x_j)$, even with a small number of paths.

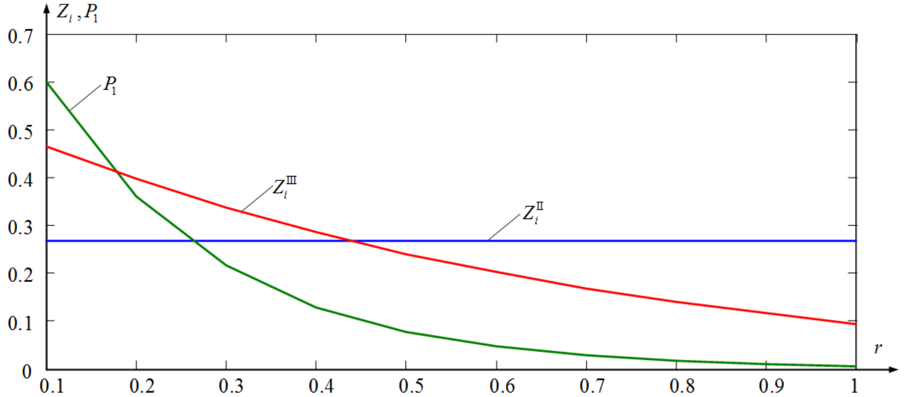
3. The message I_i provides information about the values $\varphi(x_j)$ on M $(M < n)$ from n paths and does not contain any information about the other $(n - M)$ paths; the choice of any of these $(n - M)$ paths is equiprobable. Then

$$\Phi_i^{III}(X) = \sum_{j=1}^{M} P_j \varphi(x_j) + \frac{1}{n-M} \sum_{k=M+1}^{n} P_k \varphi(x_k). \tag{16}$$

[1] The probabilities of achieving the objective on different paths are defined in a graph (Figure 4)



a)



b)

Figure 6. The value of information being processed by a) unimodal and b) multimodal subscriber terminals

Taking into account (5), a characteristic of the values of the i^{th} message about the j^{th} path is obtained in the form

$$Z_i^{III} = \sum_{j=1}^{M} [\varphi(x_j)]^{r+1} + \frac{1}{n-M} \sum_{k=M+1}^{n} [\varphi(x_k)]^{r+1} - \frac{1}{n} \sum_{k=1}^{n} \varphi(x_k). \qquad (17)$$

This case (Figure 6 b)) is common to polymodal ICS, in which the signals of single modalities (their parameters) are processed and transmitted. In this respect, given the presence of the developed algorithms $\{al_1, \ldots, al_{26}\}$ for single modalities identification, it is possible to uniquely determine the values $\varphi(x_j)$.

Analysis of the dependence of the probability of achieving the objective X on a number of "used" by the subscriber modalities N_{IM} (Figure 7) allows us to conclude that for a given probability of choosing P_j $(r = 0, 1)$ communication channel (the j^{th} path (paths)), the probability $\Phi_i(X)$ of achieving the purpose of multimodal ICS grows with the increase of N_{IM}. This system should be built so that the sources and ways of information processing communicated messages about those paths x_j that ensure the maximization $\Phi_i(X)$. Thus, the algebraic expression () clearly links information value with multimodal ICS effectiveness probability of reaching the purpose X (service implementation) – and allows one to determine the minimum required number of introduced objects $M = minN_{IM}$ that is used to assess the multimodal information integrity:

$$I^{ICS} = \frac{1}{M} \sum_{m=1}^{M} x_m^{I0} C_m \tag{18}$$

where C_m is the credibility of the assessment of the m^{th} introduced object's state; the m^{th} index of the completeness of information is defined as

$$x_m^{I0} = \begin{cases} 1, & \text{if there is introduced object;} \\ 0, & \text{– in the contrary case.} \end{cases}$$

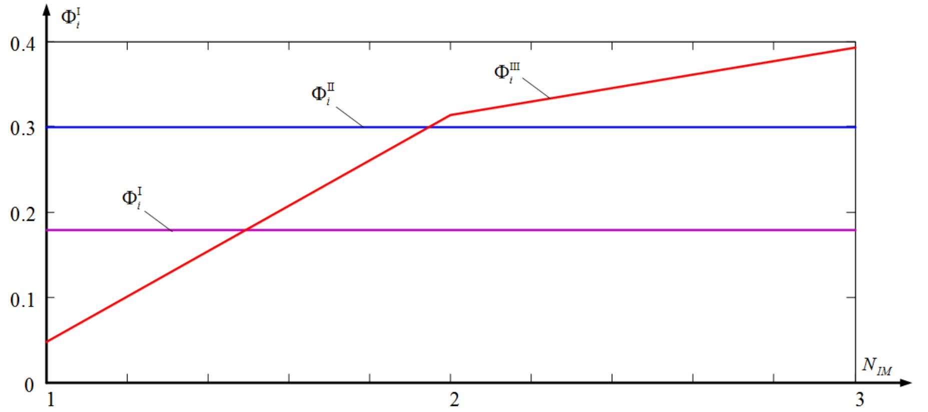


Figure 7. The dependence of the probability of attaining the objective on a number of modalities

In case of the traditional infocommunications, introduced objects at the data transmission network level are blocks of data corresponding to telecommunication services. As for multimodal ICS, introduced objects are blocks of data that carry information about different aspects of the communicative act (communication parties). This corresponds to rejection of the traditional principles to provide communication services to the user, stated in this study, in favour of ensuring polymodality of communication via technical means of communication.

6 CONCLUSIONS

The review of the latest achievements in science and technology regarding the development of multimodal user terminals showed that the transition from providing traditional communication services to transmitting and receiving signals of different modalities will contribute to reduction or elimination of technical barriers of interpersonal communication. At the level of decision-making, integration of semantically different information transmitted over different communication channels allows for offering higher quality services. This approach is new in the subject area and requires a variety of research studies to develop innovative elements of the theory of multimodal information transmission. This paper shows the increase of the informational efficiency of the multimodal communication systems based on multimodal subscriber interfaces, compared to traditional telecommunications. Taking into account the introduced notion of information value, the expediency of transition to multimodal communication systems is grounded. There are prerequisites for laying down the general principles of constructing multimodal communication systems, and in particular, for choosing functional structures and signals (parameters) that possess the most favorable informational characteristics. The proposed approach for constructing multimodal communication systems is implemented in several developing application for videoconferencing and distributed E-meetings system [25, 26, 27].

Acknowledgements

This work is supported by state research No. 0073-2014-0005 and the Russian Foundation for Basic Research (grant No. 15-07-06774).

REFERENCES

[1] Zhang, D.—Huang, H.—Lai, C. F.—Liang, X.—Zou, Q.—Guo, M.: Survey on Context-Awareness in Ubiquitous Media. Multimedia Tools and Applications. Vol. 67, 2013, No. 1, pp. 179–211, doi: 10.1007/s11042-011-0940-9.

[2] Schmalstieg, D.—Gervautz, M.—Stieglecker, P.: Optimizing Communication in Distributed Virtual Environments by Specialized Protocols. Virtual Environments and Scientific Visualization 96. Part of the Series Eurographics. 1996, pp. 1–10.

[3] Juhár, J.—Čižmár, A.—Rusko, M.—Trnka, M.—Rozinaj, G.—Jarina, R.: Voice Operated Information System in Slovak. Computing and Informatics, Vol. 26, 2007, No. 6, pp. 577–603.

[4] Basov, O. O.—Struev, D. A.—Ronzhin, A. L.: Synthesis of Multi-Service Infocommunication Systems with Multimodal Interfaces. Internet of Things, Smart Spaces, and Next Generation Networks and Systems (NEW2AN 2015/ruSMART

2015), St. Petersburg, Russia. Springer, Lecture Notes in Computer Science, Vol. 9247, 2015, pp. 128–139, doi: 10.1007/978-3-319-23126-6_12.

[5] Saveliev, A. I.—Ronzhin, A. L.: Algorithms and Software Tools for Distribution of Multimedia Data Streams in Client Server Videoconferencing Applications. Pattern Recognition and Image Analysis, Springer, Vol. 25, 2015, No. 3, pp. 517–525, doi: 10.1134/S1054661815030207.

[6] Rusko, M.—Darjaa, S.—Trnka, M.—Sabo, R.—Ritomský, M.: Expressive Speech Synthesis for Critical Situations. Computing and Informatics, Vol. 33, 2014, No. 6, pp. 1312–1332.

[7] Basov, O. O.—Saitov, I. A.: Methods of Polymodal Information Transmission. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, Vol. 15, 2015, No. 2, pp. 293–299.

[8] Karpov, A.—Ronzhin, A.—Kipyatkova, I.: An Assistive Bi-Modal User Interface Integrating Multi-Channel Speech Recognition and Computer Vision. Proceedings 14th International Conference on Human-Computer Interaction (HCII 2011), Orlando, FL, USA. Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science, Vol. 6762, 2011, pp. 454–463, doi: 10.1007/978-3-642-21605-3_50.

[9] Crossan, A.—Lefebvre, G.—Zijp-Rouzier, S.—Murray-Smith, R.: A Multimodal Contact List to Enhance Remote Communication. First International Workshop on Mobile Social Signal Processing (MSSP 2010). Lecture Notes in Computer Science, Vol. 8045, 2014, pp. 84–100, doi: 10.1007/978-3-642-54325-8_9.

[10] Datcu, D.—Yang, Zh.—Rothkrantz, L.: A System for Adaptive Multimodal Interaction in Crisis Environments. International Conference on Usability and Internationalization (UI-HCII 2007), Part II. Lecture Notes in Computer Science, Vol. 4560, 2007, pp. 334–343.

[11] Caschera, M. C.—D'Ulizia, A.—Ferri, F.—Grifoni, P.: An Advanced Multimodal Platform for Educational Social Networks. On the Move to Meaningful Internet Systems: OTM 2010 Workshops (OTM 2010). Lecture Notes in Computer Science, Vol. 6428, 2010, pp. 339–348, doi: 10.1007/978-3-642-16961-8_53.

[12] Frattini, G. et al.: A New Approach Toward a Modular Multimodal Interface for PDAs and Smartphones. Visual Information Systems. Web-Based Visual Information Search and Management (VISUAL 2008). Lecture Notes in Computer Science, Vol. 5188, 2008, pp. 179–191, doi: 10.1007/978-3-540-85891-1_21.

[13] Dahl, D. A.: The W3C Multimodal Architecture and Interfaces Standard. Journal on Multimodal User Interfaces. Vol. 7, 2008, No. 3, pp. 171–182.

[14] Liu, F.—Chou, W.—Guo, W.—Wong, W. E.: An Approach to Integrating SIP in Converged Multimodal/Multimedia Communication Services Springer. Telecommunication Systems, Vol. 28, 2005, No. 3, pp. 387–405.

[15] Maricchiolo, F.—Gnisci, A.—Bonaiuto, M.: Political Leaders Communicative Style and Audience Evaluation in an Italian General Election Debate. International Workshop on Multimodal Communication in Political Speech. Shaping Minds and Social Action (Political Speech 2010). Lecture Notes in Computer Science, Vol. 7688, 2013, pp. 114–132, doi: 10.1007/978-3-642-41545-6_10.

[16] Csiszar, I.: Information Theory: Coding Theorems for Discrete Memoryless Systems. Mir, Moscow, 1985, 400 pp.

[17] Kogan, I. M.: Applied Information Theory. Radio and Communication, Moscow, 1981, 216 pp.

[18] Basov, O. O.: Principles of Construction of Polymodal Info-Communication Systems Based on Multimodal Architectures of Subscribers Terminals. SPIIRAS Proceedings, 2015, No. 39, pp. 109–122.

[19] Basov, O. O.—Saitov, I. A.: Functioning Quality and Effectiveness of Polymodal Infocommunication Systems. SPIIRAS Proceedings, Vol. 1, 2014, No. 32, pp. 152–170.

[20] Ronzhin, A. L.: Design of Interactive Applications with Multimodal Interface. In: Ronzhin, A. L., Karpov, A. A. (Eds.): Doklady TUSUR, Vol. 1, 2010, No. 21, Part 1, pp. 124–127.

[21] Ronzhin, A. L.—Karpov, A. A.: Russian Voice Interface. Pattern Recognition and Image Analysis, Vol. 17, 2007, No. 2, pp. 321–336, doi: 10.1134/S1054661807020216.

[22] Ronzhin, A. L.—Budkov, V. Yu.—Karpov, A.: Multichannel System of Audio-Visual Support of Remote Mobile Participant at E-Meeting. Third Conference on Smart Spaces, ruSMART 2010, and 10th International Conference, NEW2AN 2010. Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science, Vol. 6294, 2010, pp. 62–71, doi: 10.1007/978-3-642-14891-0_6.

[23] Kogan, I. .: Applied Information Theory: Results of Science and Technology. Radiotechnica, VINITI, Moscow, 1972, Vol. 4.

[24] Ustinov, A. A.: Stochastic Coding of Video and Voice Information. Monogr. 2 parts. MAC, SPb, 2005.

[25] Saveliev, A. I.—Vatamaniuk, I. V.—Ronzhin, A. L.: Architecture of Data Exchange with Minimal Client-Server Interaction at Multipoint Video Conferencing. Internet of Things, Smart Spaces, and Next Generation Networks and Systems. 14th International Conference, NEW2AN 2014, and 7th Conference, ruSMART 2014. Lecture Notes in Computer Science, Vol. 8638, 2014, pp. 164–176, doi: 10.1007/978-3-319-10353-2_15.

[26] Ronzhin, A. L.—Budkov, V. Yu.—Karpov, A.: Multichannel System of Audio-Visual Support of Remote Mobile Participant at E-Meeting. Third Conference on Smart Spaces, ruSMART 2010, and 10th International Conference, NEW2AN 2010. Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science, Vol. 6294, 2010, pp. 62–71, doi: 10.1007/978-3-642-14891-0_6.

[27] Budkov, V. Yu.—Ronzhin A. L.—Glazkov, S. V.—Ronzhin, A. L.: Event-Driven Content Management System for Smart Meeting Room. 11th International Conference, NEW2AN 2011, and 4th Conference on Smart Spaces, ruSMART 2011. Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science, Vol. 6869, 2011, pp. 550–560, doi: 10.1007/978-3-642-22875-9_50.

Oleg Basov works as Senior Reseacher at the Laboratory of Speech and Multimodal Interfaces at the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. His research interests are processing and coding of speech and iconic signals, polymodal infocommunicational systems design. The number of publications: 165.

Irina Kipyatkova works as Senior Reseacher at the Laboratory of Speech and Multimodal Interfaces at the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. She defended her Ph.D. thesis in 2011. Her research interests are automatic speech recognition, statistical language models, neural networks. The number of publications: 60.

Anton Saveliev works as Reseacher at the Laboratory of Speech and Multimodal Interfaces at the St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences. His research interests are human-computer interaction and robotics. The number of publications: 15.