Multimodal Interface for Smart Home

Norbert Gyurian, Ivan Drozd, Gregor Rozinaj
Institute of Telecommunications, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia
norbertgyurian@gmail.com, ivan.drozd@ktl.elf.stuba.sk, gregor.rozinaj@gmail.com

***Abstract* – This paper deals with multimodal interface for smart living and its design. The first part of the article describes the topic of multimodal interface in general. The second part is focused on its architecture and design. This topic is nowadays very actual, because number of intelligent devices in our daily life increases very fast. They connect to the network and they are able to communicate between each other. It's important to design and create well defined architecture able to manage its extensions and provide access to common resources.**

***Keywords* – Multimodal Interface; Smart living; Smartphone; MMI**

I. Introduction

The science and technology in recent years come a long way. Things that happen to us a few years ago seemed unimaginable is now reality. People have been trying to invent a mechanism, whereby it would be possible to communicate with computer for a long time. Man as a human being feels the need to communicate and share information with other people. We meet mainly with communication between human beings in daily life. Science and technology develop and people feel need to communicate also with computers or more generally with machines. Communication using keyboard, mouse or other peripheral devices is no longer sufficient. Human feels the need to communicate with computer as an equal partner. This issue came to the forefront with invention of the first computers. As an example, it could be used as access to information technologies for people with visual or other disabilities (reading websites, etc.).

Human being is able to collect, evaluate and reproduce the information using large amount of input and output channels. Senses can be considered as input channels like for example vision, taste, smell, hearing and touch. These information are collected and evaluated using human brain. Output channels are for example speech, mimic, motoric functions and another. Human body can by therefore considered as special type of the multimodal intelligent system.

There are already a number of home automation systems. For example Tecomat FOXTROT. They are small extendable machines with a modular architecture, which increases the versatility of the system. In addition to management of intelligent home they can be used also in different sectors of industrial automation. [4] Another one is control system called Nikobus. It is a system usable in households, smaller buildings or hotels. The system is aimed at apartment complexes and focuses mainly on the requirements in this area [7]. And many others.

II. Multimodal Interface

Currently, the most widely used interfaces for human–computer communication a computer are peripheral devices, such as keyboard, mouse, etc. Human is able to control many devices connected to the computer using these peripheral devices. This type of modality has become a popular and practical to preserve the very beginning of evolution of computer technology, due to the complex anatomy of the hands, with which one can express a large set of gestures / commands. However, as we mentioned above, communication with the computer using the keyboard and mouse is currently insufficient and human wants to be an equal partner with computer. Speech, gestures, face and many others are modalities that can make from computer an equal partner. Under multimodal interfaces meant an interface enabling entry of two or more modalities. This exchange of information is more natural than just using of one modality. When we imagine communication human-human, human receives information from many sources and not just using speech. For example mimic, gesticulation and many others.

Multimodal interface became more and more popular with development of the sensors and digital signal processing. Nowadays we know lots of implementations of the multimodal interface but none of them is taken off significantly. Their future potential is significant when we consider systems for people with disabilities, smart living etc. Multimodal interface brings several advantages[1], [2]:

- Input overloading – in the case of using only one modality, we would have every command transmitted only through this modality and it can lead to the "overloading". [1], [2]
- Collection of information – another major advantage of a multimodal system is a collection of information. This advantage can be explained using the example of the authentication system. Imagine a computer security using biometric features. In the case of using only one biometric feature such as speech (speaker recognition) which we know that the percentage is more than 80%, we would have to rely on it. This can in many cases, lead to a false identification. This disadvantage can be removed by an additional modality, such as the identification by the face. [1], [2]
- Redundancy – it is an advantage, which is closely linked to the collection of information from the previous example. Advantage can be explained on the same example of the authentication system. Provided that the system is used to identify the speaker in an environment, where successrate

around 95% can be achieved under ideal circumstances. But in the real environment conditions, the environment can be adversely affected by various factors (e.g. noise, etc.). In this case, we can use redundant information from a system of identification using face. [1], [2]

A. *Input and output channels*

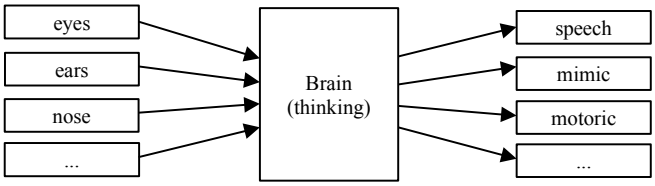

Figure 1. Human input and output channels.

Human being is able to collect, evaluate and reproduce the information using large amount of input and output channels. Senses can be considered as input channels like for example vision, taste, smell, hearing and touch. These information are collected and evaluated using human brain. Output channels are for example speech, mimic, motoric functions and furthers. Human body can be considered as special type of the multimodal intelligent system. The block diagram of this system is shown in figure above (see Fig. 1)[1], [2].

The communication between human being and computer is far away from this vision. The reason is imperfection of the software and hardware. The problem is for example limited capabilities of sensors and algorithms unable to emulate functions of the human brain.

III. DESIGN AND IMPLEMENTATION

Because during our research in area of multimodal interfaces and recognition of biometric functions, we met with several problems like flexibility, scalability etc., design of multimodal interface was one of the key concepts on which we decided to focus on.

Requirements for our multimodal interface:

- Extensibility – the functionality of multimodal interface can be easily extended using additional applications and modalities with well defined interfaces. There is no need to modify the core logic or get into the touch with core application of multimodal interface.
- Centralized user profiles – user should be able to access his own user data or configurations even if he is not at home. For example his favorite TV channels and many other application data.
- Integration of portable devices – Possibility to control the multimodal interface or authentication using a mobile phone or another portable device.

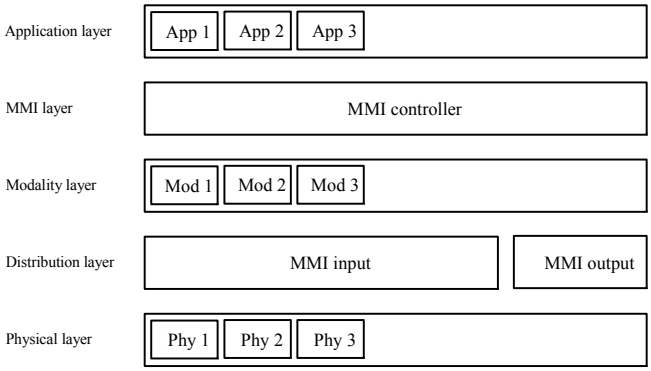

Figure 2. Layered model of multimodal interface.

As it can be seen in the Fig. 2, for the purpose of our requirements I mentioned before, we decided to split model of our multimodal interface into the several layers. Each of these are independent with well defined interfaces between each other.

Application layer is the first layer laying on the top of the stack using services of the lower layer (database access, modality output, etc.). The role of this layer is to cover every single application running on our multimodal interface.

MMI layer is the main part of the multimodal interface covering the core logic of the whole application. This layer is responsible for run-time loading of every single application from application layer., management of these applications, information exchange, etc. This layer acts as service provider for application layer. Except of application layer, MMI layer is also responsible for runtime loading of modalities available for multimodal interface, management of these modalities, information exchange, etc. MMI layer also manages list of running applications and makes link between application and modalities whose are needed in particular time.

Modality layer is third layer covering modalities available for multimodal interface. Same as with application layer, this layer covers every single modality running on multimodal interface. A modality can be considered gesture recognition, face recognition, voice recognition system and many others.

Distribution layer is layer which fulfill the role of the abstract layer between physical and modality layer. As it's maybe obvious the main role of this layer is to ensure the compatibility between physical device connected to the multimodal interface and specific modality.

Physical layer is defined with the hardware used by the multimodal interface (Kinect sensor, special camera, etc.)

II. *Extensibility and modularity*

Because the MMI is application on which we plan to work also in the future, during the development of the MMI we decided to design it as modular as possible. On our department we are working on several systems based on analysis of different biometric functions (iris recognition, voice recognition and many others). We can say that the rule is that

each working group is working on different modality. This practice leads to the development of systems which are often incompatible, so it was necessary to define an interface that would standardize them and create an architecture that would cover them and offers lots of another options.

As it is shown in the Fig. 2, we split our architecture to five layers. This separation brings us opportunity to define standardized interfaces and create architecture which is more extensive than just one dedicated application. In conjunction with our system, we consider the most important advantage of this system to be adding applications and modalities in the form of plugins. For this purpose we use DLL libraries. In Figure 3 you can see a block diagram of the multimodal interface and applications implementing a common interface. Similar philosophy we have tried to adhere to the modalities. Using this standardization MMI controller doesn't need to know lots of details about used applications and modules. It behaves the same way to each application, of course, depending on priorities.

Dynamic library loading (DLL) is often used in a modular software architectures. This feature provides the flexibility of the software and allows the addition of plug-ins, to extend the functionality of the program. Plugins may be developed in parallel / independently and without the need to re-build the kernel.

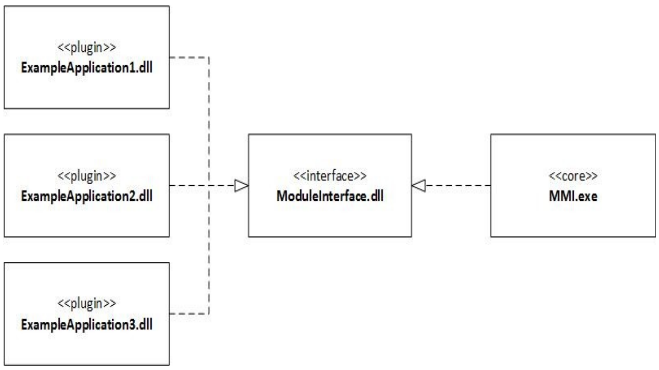

Figure 3. Human input and output channels.

As you can see in the Fig. 2, the logic of the MMI is separated from logic of the application/plugin. MMI core cares only about loading and management of applications and modalities.

The Fig. 4 shows the more detailed architecture of the MMI. This separation can be achieved using well defined interfaces shared among all applications and modalities and by logic separation. MMI controller doesn't have a clue about the modality type or the application type. His job is only to manage the life cycle of individual applications and modalities. It loads applications, runs them, pause them or stops them. Also depending on type of running applications, it pause appropriate modality to make better use of computing power (idle mode). As shown in the figure, not every application needs access to each modality.

Because the architecture of our multimodal interface is modular, the support of mobile devices is implemented using application/plugin philosophy and is loaded during initialization process of our MMI interface.

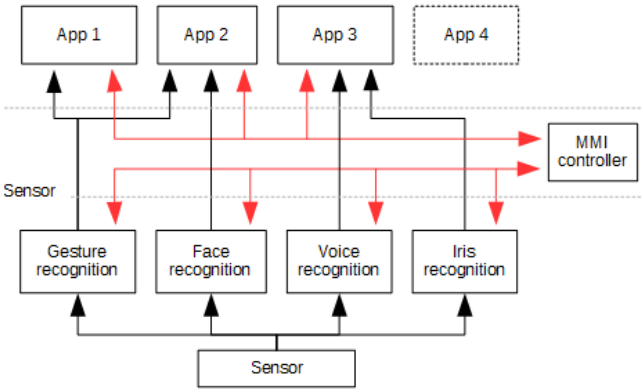

Figure 4. System architecture diagram.

III. Control using smart-phone

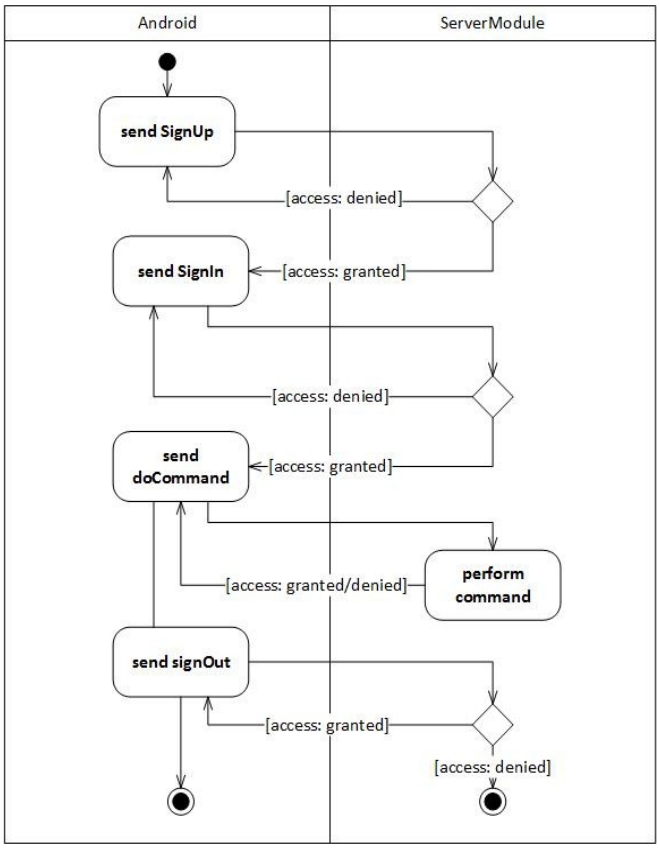

Figure 5. Life-cycle of the connection.

Because mobile technologies spread very fast and almost everybody of us owns his own smartphone, we decided to implement also authentication and control using smart device. User can confirm his identity using smartphone (he must sing

in using a password authentication) and after successful authentication he is able to control MMI using his smart device.

IV. Protocol

The protocol designed by us is based on the JSON, because it's faster to process and we don't need a large number of characters to create a valid JSON document like it's with XML documents. For more effective communication, we decided for persistent TCP connection, it means that device doesn't create the connection every time it wants to send a message, but the connection is established, until device sing out.

V. Communication messages

We tried to create simple and logical communication messages. We had in mind that massages should be brief and contain only what is strictly essential for communication. It follows that all entities are required to report. Messages can be divided into requests and responses. Requirements sends the client, in this case the Android device and the server sends responses to requests. The server response can be positive or negative. A positive response is sent if the request was syntactically valid, the client has the permission for the command, if the structure corresponds to the requirements defined by us or if the statement was meaningful. A negative reply is sent if any of the requirements was not met. The negative response contains an array of exceptions, refers to the entity "exception".

TABLE I. COMMUNICATION MESSAGES.

action	**description**
signUp	Sends client to create the registration
signUpResponse	Server response to request
signIn	Sends client to sign in
signInResponse	Server response to sign in request
signOut	Sens client to sign out
signOutResponse	Server response to sign out request
doCommand	Sends client to execute the command
doCommandResponse	Server response to command execution request

Below two examples of communication messages are introduced:

{"action":"signIn","userName":"Ivan","token":" … "}<EOF>

{"action":"signInResponse","access":"granted"}

ACKNOWLEDGMENT

The authors would like to thank for financial contribution from the STU Grant scheme for Support of Young Researchers. This paper presents also some of the results and acquired experience from the following projects: H2020 project NEWTON, No. 688503, VEGA project INOMET, No. 1/0800/16 and APVV project MUFLON, No. APVV-0258-12.

CONCLUSION

We designed and developed architecture for multimodal interface which is fully modular and easily expendable using applications and modalities in form of plugins (dll files) with well-defined interfaces. We successfully integrated gesture recognition and speaker identifications system and developed several applications with help of which we successfully demonstrated functionality of our proposed system.

In the future we plan to integrate also many other modalities developed on our department and integrate them with system of smart living.

REFERENCES

[1] Roope Raisamo: Multimodal Human-Computer Interaction: a constructive and empirical study, Academic dissertation, University of Tampere ,1999 ,85 s.

[2] Lai Wei, Huosheng Hu: Towards Multimodal Human-Machine Interface for Hands-free Control: A survey, Technical Report: CES-510, 2011, 26 s.

[3] VACHÁLEK, J. 2014 Využitie senzorického systému Microsoft Kinect pre potreby inteligentných domov a budov (3). In iDB Journal. ISSN 1338-3337, 2014, roč. 4, č. 1, s. 18-19.

[4] STRAŠIFTÁK, A. 2014. Automatizácia procesov v riadení inteligentného domu: dizertačná práca. Trnava: Trnava STU, 2014. 85 s.

[5] WAKEFIELD, J. 2016. Tomorrow's Buildings: Help! My building has bennhacke. In BBC News [online]. 2016 [cit. 2016-04- 25]. Dostupné na internete: <http://www.bbc.com/news/technology-35746649>.

[6] When to use dynamic linking and static linking. In IBM Knowledge Center [online]. [cit. 2016-04- 28]. Dostupné na internete: < https://www.ibm.com/support/knowledgecenter/ssw_aix_61/com.ibm.aix.performance/when_dyn_linking_static_linking.htm >.

[7] HAMERNÍK, P. 2012. Návrh informačných a riadiacich systémov pre inteligentné domy: dizertačná práca. Trnava: Trnava STU, 2012. 113 s.