Proceedings of the Federated Conference on Computer Science and Information Systems pp. 1075–1080
ISBN 978-83-60810-51-4

Voice controlled environment for the assistive tools and living space control

Vytautas Rudzionis
Vilnius university, Kaunas faculty
Muitines str. 8, Kaunas, Lithuania
Email:
vytautas.rudzionis@vukhf.lt

Rytis Maskeliunas
Kaunas university of technology
Institute of Automation and
Control Systems, Studentu str.
48A-303 Kaunas, Lithuania
Email: rytis.maskeliunas@ktu.lt

Kestutis Driaunys
Vilnius university, Kaunas faculty
Muitines str. 8, Kaunas, Lithuania
Email: kestutis.driaumys@vukhf.lt

***Abstract*—This paper describes our efforts developing the smart home environment for the assistive living. The key element of the smart environment is the ubiquitous voice user interface with several additional capabilities (such as the recognition of several gestures). This work is a further development of voice controlled devices. The presence of the commercial speech recognition engines and our experience of adapting foreign language engine to recognize Lithuanian voice commands suggested expansion of the platform including the possibilities to control various devices in the living space. The key element of the proposed platform is its universal nature, the possibility to adapt the platform for the personal needs and the economical solutions used. Platform was developed using inexpensive hardware and software elements available on the market. The field tests with several sets of voice commands used by people with motoric disabilities showed high robustness of proposed platform.**

I. Introduction

VOICE technology is potentially of an enormous benefit for people with physical disabilities. People with different kinds of disabilities may benefit from different kinds of speech and voice processing technologies but it is really hard to find the impaired person that can't benefit form voice technology more than ordinary people may benefit in the same situation. Since people with various types of impairments have the problems using one or more channels of communication used by the ordinary people they rely more on the voice based interfaces. This is true for the blind or visually impaired people: since they have difficulties using GUI type interfaces they are more willing to use voice based interfaces. And this is especially true for the motor-handicapped persons. The impossibility or difficulties changing their location requires the means to control the devices in the living environment by distance and often with minimal use of hands or fingers. To be successful, applications implementing voice processing should effectively take into account the specific needs of user groups and have the ability to adapt to the needs of individual person. One of the most desirable needs of the voice controlled system is the flexibility of the platform. Under the flexibility we mean the possibility to construct a personal vocabulary, to adapt the voice user interface to the voice of particular speaker and the possibility to introduce new devices or change the existing ones with minimum efforts and complication.

Recent years witnessed at least several important events when developing various speech controlled applications for the different fields of applications. Microsoft, Nuance and some other companies began to distribute speech application programming tools (usually known as SAPI – speech application programming interface; the example of such type of tools could be Microsoft speech server (MSS) or Microsoft Windows SAPI). Such tools provided the developers with a whole set of new and useful tools easing, speeding up and making more economically affordable the development of voice user based interfaces. Researchers usually have their proprietary tools to implement voice user interfaces into the specific types of applications but their implementation into the more universal platforms often makes troubles. Another very important factor was the increased recognition accuracy of available speech recognition engines that made voice user interfaces acceptable even in the wide area of applications intended to the use of ordinary users not only for the impaired persons.

But all of this is true only for the countries where the widely spoken languages (such as English or Spanish) are used as a primary mean of communication. The providers of SAPI tools distribute speech recognition engines only for such languages and aren't interested in the development of speech engines for such rarely spoken languages as Lithuanian is. For users of such languages other approaches are necessary to be found. Our experience with the adaptation of foreign language speech engines to recognize the Lithuanian voice commands showed that it is possible to achieve very high recognition accuracy for many of the Lithuanian voice commands using only the proper selection of phonetic transcriptions for the Lithuanian voice commands. Such approach enabled to make the development of some limited vocabulary applications easier and more economically viable. From another point of view the experience showed that not all Lithuanian voice commands may be adapted to be recognized good enough and the implementation of proprietary recognition tools may be necessary in some situations. It means that hybrid recognition approach may be applied: a) to use two or more recognizers working in parallel and; b) combining their responses to get the final decision.

Parts of this work were done under research project No.: 20101216-90 funded by EU SF project "Postdoctoral Fellowship Implementation in Lithuania" (VP1-3.1-ŠMM-01).

All these considerations enabled us to propose and to develop the platform for the voice user interface targeted at the people with motoric impairments (and some other categories of disabled persons too). The platform include speech processing server, tablet PC or a smartphone which is used by a person, and various assistive and home tools connected to the executive devices controlled via speech server. Such approach enabled to achieve high flexibility, robust performance and relatively economical realization. The further paper is organized as follows: chapter 2 presents the related and similar work, chapter 3 introduces the platform in detail, while chapter 4 presents some results evaluating platform's performance. Finally some conclusions and suggestions are presented.

II. Related Work

The ability to control the living space is an essential component of independence and e-inclusion. Environmental Control Systems (ECS) or Smart home control interfaces are available which address many elements of home management for disabled people, such as control of audio-visual equipment, telephones, household appliances, doors and curtains as well as the ability to summon assistance. Most ECSs utilize switch-scanning or keypad interfaces for control. More recently, ECSs with speech recognition have been introduced and a number of such systems are available on the market. Their success depends on a number of factors. Most important of them is the maturity of voice processing technology used.

One group of the devices used in the smart home environment is the devices with the embedded voice recognition. Modern houses are equipped with a set of household appliances ranging from simple lights to feature-rich hi-fi systems, DVD players and TV sets. Typically they are controlled by keypads but more and more often various devices with voice control capabilities are present on the market. Unfortunately since these devices are developed and manufactured by various producers and typical modern household has many devices, serious problem of matching of their interfaces and avoiding of conflicting commands arises. Another serious inconvenience of embedded systems is that such systems often are very difficult to personalize and to change the command used for the control. This is particularly important for the Lithuanian speakers since major household appliances manufacturers do not provide Lithuanian voice commands recognition as an option.

The idea to implement voice command recognition capabilities in the smart home environments isn't new. There were also the attempts to implement voice command recognition capabilities for the specific needs of the impaired people. An example of adaptation and integration of various communication technologies could be the e-wheelchair project [1]. E-wheelchair is an electronic wheelchair with integrated communication equipment based on IPv6 protocol. This protocol enables mobile communication using an internet data transfer. Among the most important benefits for the impaired people in this case is the possibility to be in touch with the caregiver nearly all the time and to increase the level of independency having the possibility to obtain the information which previously was inaccessible independently.

Various activities to develop a speech operated smart-home control, systems with a focus on people with special needs by providing a unified speech-controlled interface were observed in recent years [2]. The ultimate goal of home-environments is to achieve the level when the user has to say what he/she desires and the system will orchestrate the home appliances in order to fulfill specific wish. Unified speech-oriented smart home interface will enable easy replacement of old household devices by the new ones as well as an easy installment of completely new devices. Such interfaces are often designed taking into account that disabled people should not be confined in any way to the vicinity of their homes, so the telephony interface is implemented to enable full remote access to the system.

In [3] even the privacy issues when developing smart home environments are discussed and methods to avoid the loss of privacy in some situations were suggested. Wu and Fu [4] showed that it is important to achieve mixed-initiative when designing the algorithms for smart home systems. In this way services could be subdivided to the levels called resources allocation, privacy, attention and priority. Despite that their paper isn't specifically aimed to the problems implementing voice based interfaces but many insights could be successfully applied to the development of VUI for smart home environments. Ye and Huang [5] investigated cloud-based framework for the smart home. Cloud-based approach could be a valuable framework when designing universal and ubiquitous platform for the smart home environment. But it should be noted that cloud-based frameworks are still in the early stages of development and lots to be done remain to achieve wider applicability and functionality of such platforms. Alam and colleagues [6] presented multi-agent based human activity classification model for the smart home environment. Agents are designed using hierarchical approach with task oriented architecture and the results obtained from human psychology studies. The authors showed that during single communication episode people typically are using 2-6 activities and the bigger number of activities and what is very important – the single activity – are rare. Manchon and colleagues [7] presented the analysis of multimodal interaction in smart home environment. The experimental investigation showed that speech input and often output is the preferred way of interaction between human and the devices comprising smart home ecosystem. It is important that preference given to the speech based input grows in time when the users become more familiar with the voice interface.

Further we will present our platform for the smart home aimed to serve the impaired people and the motivation behind some architectural and technical aspects of this platform. Later several experiments evaluating the efficiency and the accuracy of the presented platform will be described. The primary focus of interest will be related to the speech recognition issues since speech is the main mode of communication while other modes serves more as supplementary ways for communication.

III. Voice Controlled Platform For The Assisted Living

The voice controlled platform for the assisted living is the further development of our voice controlled devices [8], which could be used by the motor-handicapped people in their daily lives. Mobility devices, such as a wheelchair provides the opportunity to overcome the main limitation of this type of people – the inability to move independently. Modern wheelchairs often uses battery powered electric vehicle and small keypad to control the movement. But for a wide group of motor handicapped persons the voice control could be most desirable and convenient mode of control.

There were successful attempts to manufacture modern devices controlled by voice commands. Such applications typically have an embedded voice command recognition system and control unit designed to recognize and process pre-specified set of commands. Usually the task to change this set of commands is complicated task from a user's point of view. From a Lithuanian speaker's perspective it is very important that such device recognizes only English (or some other language) commands as it is difficult to embed the recognition of Lithuanian commands.

These considerations suggested us to propose a client-server based architecture for smart device control: the user is provided with the PDA type device which serves as the recorder and does some initial processing of speech signal and transmits it to the server. Server runs a speech recognition engine, receives voice commands, recognizes them and makes appropriate turns. The PDA client and server are linked using Bluetooth or wireless connection. In the same way server is linked using wireless connection with the controllers and executive devices used to control the movements of the assistive tool. Later fuzzy sets theory based control methods were implemented [9] to control the trajectory of the device. This approach implemented elements of smart interface and intelligent control for the assistive device.

The experience with this development led to the idea to expand this approach to universal voice control platform for the smart home for the assisted living. It should be noted that main advantage of such approach is its flexibility: since nowadays houses are equipped with a set of different household appliances ranging from simple lights to feature-rich hi-fi systems, DVD players and TV sets then it is possible to use a voice command recognition system to control this big variety of home appliances and not only the specialized devices. Another advantage is the possibility to implement other than voice commands control modalities or to use them as the additional channel to transfer the commands.

Other primary requirements for the smart home VUI platform were the affordability and the robustness.

The proposed voice controlled platform for the assistive living is presented in Fig. 1.The user is provided with the remote control unit which could be PDA device, or a smartphone. This unit is used to record the speech signal, to perform some initial speech processing (trimming the silence, reducing the noise) and to transfer the speech signal to the server for the further processing and recognition. The server runs a commercial speech engine adapted to recognize the

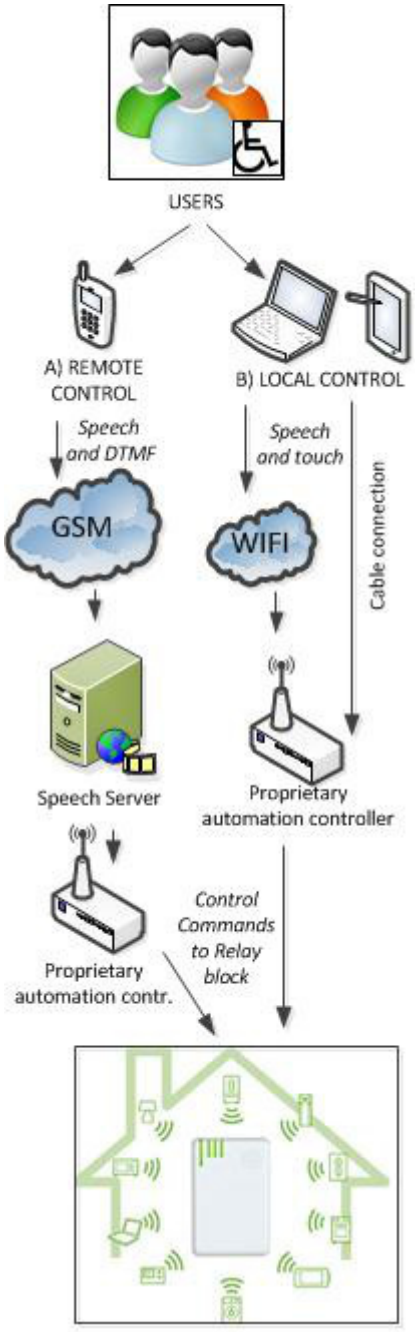

Fig. 1 Voice controlled platform for the control of home environment for the assistive living

Lithuanian voice commands and also a proprietary speech recognizer used to recognize the same voice commands. The additional analyzer is used to combine the output and to make the final decision. Then a server transmits the necessary command to the appropriate controller and the execution unit to achieve the desired user action. Also a server is equipped with the software tools necessary to construct the library of voice commands, to train and to adapt the recognizers, to prepare script files for the control of home appliances, etc. In principle all of these steps could be perform by the skilled user of the system but more often the help of trained consultant will be used to configure the system for

the later use. Though such principles aren't very new but their combination enabled us to achieve all design requirements set for this platform. The main advantage of the platform is that it allows to reconfigure the system and to adapt it to some particular environment in the easiest and most flexible way (to set particular vocabulary and devices to be controlled).

From the technical point of view the platform is developed from the widely available, not expensive, easily scalable, configurable and programmable elements such as the Phidgets, ATMega controller chips, etc. In this way we were able to build economical (costs of the hardware were less than 150 euro) and at the same time capable system: the user could select own set of household devices. The set of devices used for laboratory tests is shown in Fig. 2.

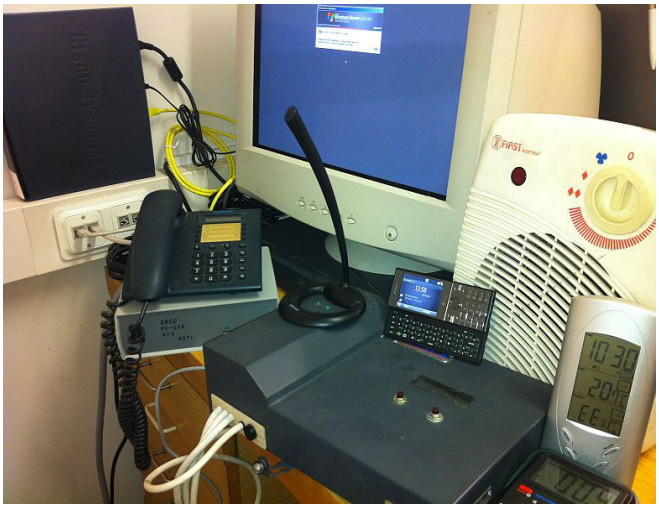

Fig. 2 Household appliances used in the testing of the smart environment for the assisted living

The control capabilities for different devices depend on their interfaces and some design peculiarities: in all cases it is possible to switch on or off any device while for some devices (e.g. the computer) exist a wider set of control capabilities (open the particular program, open the website of interest, etc.).

IV. Voice Command Recognition for the Assistive Living Environment

The crucial element of the successful implementation of the platform is the robust recognition of voice commands. Beside the high accuracy it should operate successfully in noisy environments or be able to react only to the commands pronounced by the intended speaker and to recognize only the commands from the vocabulary. The successful solution of all of these tasks in complex is very difficult and in many aspects still unsolved problem.

Several ways were used to adapt foreign language recognizers for the resources of another language. Some of them take a small amount of data in native language for retraining the foreign language acoustic models for the language of interest [10]. In this paper we are focused on the adaptation via phonetic transcriptions proper selection since our earlier experiments showed that it is possible to achieve high recognition accuracy for the wide class of Lithuanian voice commands in this way, the fastest [11]. In our experiments Microsoft Speech server and English or Spanish speech engines from this software package were used. For many voice commands very high accuracy (more than 95% recognition accuracy) was observed for a wide range of different speakers and different acoustical conditions. At the same time it was seen that some voice commands can't be adapted to be recognized good enough and proprietary speech recognizer needed to be implemented.

In the proposed platform we implemented a hybrid recognition approach: two recognizers are working in parallel and their outputs are combined to get the final response of the system. One of the recognizers is the Microsoft Speech server with English engine. The second recognizer is the proprietary CD-HMM based voice command recognizer trained using a rather limited set of training material. Similarity values from both recognizers were used to train linear discriminant classifier:

$$F(x)=(\mu_i-\dot{\mu})\Sigma^{-1}(\mu_i-\dot{\mu})^T \qquad (1)$$

where μ is used to denote the mean vector of parameters and Σ is used to denote the covariance matrix. Then classification and regression tree-like procedure is applied to select the most likely discriminant function value from a set of discriminant function values for different classes. In principle this approach is similar to the linear heteroscedastic analysis [12] which was successfully used to combine the recognition results of several differently trained speech recognizers [13]. While the adaptation of phonetic transcriptions was used in some of our experiments here we tried to use the hybrid approach for the first time to evaluate the possibilities to improve voice commands recognition accuracy.

In order to evaluate the efficiency and robustness of the recognition system, several groups of experiments were carried on. First group of experiments contained 12 often used voice commands to control assistive tools and household appliances. Recordings of 200 female speakers and 110 male speakers were used to train the recognizer or to find the optimal phonetic transcriptions for the foreign language recognizer (the male and female utterances were treated as distinct commands to be recognized). The recordings of 20 male and 20 female speakers were used to test the performance. In the first experiment each voice command was uttered in strictly isolated manner. In the second experiment voice commands were pronounced in word strings imitating the possibility to utter several voice commands in a row to speed up the control of the devices. Table 1 summarizes the results of this experiment.

The main conclusion that could be drawn from these results is that while the error rates varied insignificantly for different kinds of recognizers and the recognition accuracy was very high but the confidence level was higher using hybrid recognition approach. This observation allowed us to expect that the recognition accuracy will be higher even when acoustic conditions will be more complicated or

TABLE 1.
THE RECOGNITION ACCURACY AND AVERAGE CONFIDENCE FOR DIFFERENT TYPES OF RECOGNIZERS

Recognizer	Isolated commands		Groups of words	
	Error %	Average confidence	Error, %	Average confidence
adapted foreign language	0	0.56	0	0.48
Lithuanian	0	0.44	2.2	0.41
hybrid	0	0.71	0	0.62

TABLE 2.
THE RECOGNITION ACCURACY AND AVERAGE CONFIDENCE FOR DIFFERENT TYPES OF RECOGNIZERS AND DIFFERENT TYPES OF RECOGNIZERS

Recognizer	Isolated commands		Groups of words	
	Error %	Average confidence	Error, %	Average confidence
adapted foreign language	0.5	0.47	2.4	0.38
Lithuanian	3.1	0.31	6.7	0.37
hybrid	0	0.65	0	0.57

speaker voices will be more specific and more difficult to recognize.

Another group of experiments was performed to evaluate the accuracy of VUI in continuous and fixed input modes in real home environments. We understand that the continuous input when audio is recording is fed to the recognizer continuously and it is the task of the recognizer or the preprocessor to find the boundaries of voice command. As the fixed input mode we understand the audio recording mode when the user is instructed to press a key and to tell the voice command. The experiments were performed in the room were typical usual activities were performed (there were 2-3 persons in the room and everyone was engaged in own area of interest – working, maintenance, reading, watching TV, etc.). The set of commands used in these experiments consisted from 22 often used Lithuanian commands to control home appliances. The commands were composed using 1-3 words pronounced continuously (12 commands were composed from a single word, 8 commands from two words and 2 commands from three words). The utterances of 50 speakers (25 males and 25 females) were used to train or to adapt the recognizer while 10 speakers (5 males and 5 females) were used for testing. Each speaker uttered each command 5 times. The command utterances were fed to the recognizer in isolated command mode (single command fed in one try) or in a group of three commands at once. In the latter case the correct recognition was fixed only if all three commands (independently from the command content) were recognized and performed appropriately.

The main goal of this experiment was to evaluate the possibility to develop VUI for the smart home environment using only adapted Lithuanian voice commands for the foreign language recognizer and the utilizing the importance to implement a hybrid approach to achieve the robust performance of the VUI for the impaired persons.

Table 2 summarizes the results of these experiments.

In these experiments very high recognition accuracy was achieved too. The more complicated testing conditions caused slightly higher number of the recognition errors. It is important to note that using hybrid approach there were no recognition errors during testing, while using adapted or proprietary Lithuanian recognizer few errors were observed. This shows that both recognizers have different discriminant capabilities and may supplement each other. The higher number of recognition errors using proprietary Lithuanian recognizer may be caused by the limited amount of data used to train it.

Summarizing the VUI efficiency we could conclude that the achieved recognition accuracy satisfies the user expectations and may be implemented to control home environment. Particular benefits may be achieved by the people with motoric disabilities.

The last group of experiments was performed with the aim to evaluate the performance time when household devices are controlled using voice commands or more typical key-based controlled approach. In this case user was given the small mobile type device keyboard to control the household appliances. It meant that a user needed to press down the number of appropriate commands or to say this command by voice. Users were instructed to execute 1, 2 or 3 commands in a row. 3 people with motoric disabilities participated in these experiments. It should be noted that these people suffered from easy and medium types of handicap (people with temporal handicaps, e.g. broken legs) and weren't representatives of the very complicated impairments. They were experienced users of the computer-like devices too. For the comparison, 3 ordinary and healthy people took part in the experiments too. Each user was obliged to perform each command fully. This means that if the command wasn't recognized properly or the incorrect number of command was sent to the controller user was forced to repeat the task until completion.

From the functional point of view each user needed to learn the number of the commands or to learn the appropriate voice command. Each user should complete the commands in the random order: the supervisor presented the number of command (not the same as the key number) and then the user needed to complete it (to remember the command and to press the key or to say the name). This method was selected trying to avoid the mechanical repetition of the same key combinations or repetition of the same voice commands.

TABLE 3.
THE AVERAGE COMPLETION TIME (IN SEC) OF 1,2 AND 3 CONTROL COMMANDS USING VOICE AND KEYBOARD MODES

Mode	Tested group	Number of commands in a row		
		1	2	3
voice	impaired	1,6	3,8	5,7
	average	1,5	3,6	5,4
key	impaired	1,8	4,7	7,6
	average	1,8	4,8	7,8

Table 3 shows the average completion times of 1, 2 and 3 commands in a row using voice and keyboard modes.

Despite the fact that these results can't be treated as statistically reliable some observations could be made. It seems that the voice based user interface enables to perform the household devices control operations faster than using keyboard based principles. The increase in performance speed is more visible when more commands needs to be performed. These experiments showed no difference between the performance of healthy people and those with motoric impairments since all impaired people had good skills using computerized devices. The general observation from all experiments is that VUI controlled smart home and disabled people oriented platform is feasible and acceptable by the users if voice interface is carefully designed (appropriate, easy to remember and easier to recognize commands are selected).

V. CONCLUSIONS

The paper presented a universal platform for the control of smart home environment (mainly household appliances and assistive tools) oriented to the people with motoric disabilities. The key properties of the proposed platform are scalability and universality. The platform is composed from easy to get and relatively cheap hardware elements. Hence it is easy to rescale the platform and to involve different number of the controlled appliances and tools. The preferable mode of control is a voice based interface. The VUI is realized using a hybrid recognizer adapted to recognize the Lithuanian voice commands and a proprietary HMM based Lithuanian speaker-independent recognizer to find the optimal decision. This approach enables the minimization of the resources necessary to build the voice based interface: Lithuanian recognizer is used to reinforce the decision of the adapted foreign language recognizer and at the same time we need significantly less resources to build a proprietary recognizer comparing with the case when no adapted foreign language recognizer was used. The achieved performance is high and is acceptable for the vast majority of users, particularly the impaired people.

Our experience with the multimodal interfaces suggests that implementation of the elements of multimodal interfaces in the future could be valuable. Of particular importance in such type of applications could be gesture recognition.

REFERENCES

[1] T. Ernst, "E-wheelchair: a communication system based on IPv6 and NEMO" in *Proc of ICOST'04*, Singapore,, September 2004.

[2] A. Vovos, B. Kladis, N. Fakotakis, "Speech operated smart-home control system for users with special needs" *in Proceedings of Interspeech' 2005*, Lisbon, Portugal, 2005 pp. 193-196.

[3] S. Moncrieff, S. Venkatesh, and G. West, "Dynamic privacy assessment in a smart house environment using multimodal sensing," in A*CM Trans. Multimedia Comput. Commun. Appl*., vol. 5, no. 2, Nov. 2008, pp. 1–29,

[4] C.-L. Wu, L.C. Fu, "Design and Realization of a Framework for Human-System Interaction in Smart Homes", *IEEE Transactions on Systems, Man, and Cybernetics*, Part A 42(1): 2012, p.p. 15-31

[5] X. Ye; J. Huang, "A Framework for Cloud-based Smart Home" in *Proceedings of International Conference on Computer Science and Network Technology (ICCSNT)*, Shengdu, China, December 2011, p.p. 894 – 897

[6] M.R. Alam,; M.B. Reaz,; M.A. Ali, S.A. Samad, F.H. Hashim; M.K. Hamzah,, "Human activity classification for smart home: A multiagent approach," in *Proc of Industrial Electronics & Applications (ISIEA), 2010 IEEE Symposium on* , Penang, Malaysia, October 2010, pp.511-514

[7] P. Manchón, C. del Solar, G. Amores, G. Pérez., "Multimodal Interaction Analysis in a Smart House," in *Proceedings of the 9th international Conference on Multimodal interfaces ICMI '07,* Nagoya, Japan, November , 2007.

[8] V. Rudžionis, R. Maskeliūnas, A. Rudžionis. "Assistive Tools for the Motor-Handicapped People Using Speech Technologies: Lithuanian Case". in: *Lecture Notes in Business Information Processing,* Vol. 97, Berlin : Springer, 2011, pp. 123-135

[9] V. Rudzionis, R. Maskeliunas. "Control of assistive tools using voice interface and fuzzy methods"in *Proc. 15th conference of Bussiness Information Systems*, Vilnius, Lithuania, May 2012, *Lecture Notes in Business Information Processing,* Berlin : Springer, 2011

[10] Hao Yee Chan, R.Rosenfeld. "Discriminative Pronunciation Learning for Speech Recognition for Resource Scarce Languages", *Proc. Of ACM DEV 2012*, Atlanta, March 2012, article 12

[11] R. Maskeliunas."Lithuanian Voice Commands Recognition Based on the Multiple Transcriptions". Summary of Ph. D. Thesis, Kaunas, Technologija, 2009, p. 34

[12] N. Kumar, A. Andreou. "Heteroscedastic discriminant analysis and reduced rank HMMs for improved speech recognition", *Speech Communication*, vol. 26, issue 4, Elsevier, 1998, p.p. 283-297.

[13] D.Vergri, A. Stolcke, G. Tur, "Exploiting User Feedback for Language Model Adaptation in Meeting Recognition," in *Proceedings of IEEE ICASSP*, Taipei, 2009, p.p. 4737-4740