ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 *2012. No. 6(122)*
ELEKTRONIKA IR ELEKTROTECHNIKA

TELECOMMUNICATIONS ENGINEERING
T 180
TELEKOMUNIKACIJŲ INŽINERIJA

Interactive Digital Television and Voice Interaction: Experimental Evaluation and Subjective Perception by Elderly

V. Bures
Faculty of Informatics and Management, University of Hradec Králové,
Rokitanskeho 62, 500 03 Hradec Kralove, Czech Republic, phone: +420493332259, e-mail: vladimir.bures@uhk.cz

crossref http://dx.doi.org/10.5755/j01.eee.122.6.1827

Introduction

Advanced and modern technologies employed in telecommunication devices allow their incorporation into the smart environment concept, which is based on ambient intelligence or pervasive computing principles. Various tools from such diverse areas as user behavior [1], knowledge management [2], or technologies encouraging physical activity in workplaces [3] have already been designed and implemented in this field of study. Contemporary, latest developments of interactive digital television (iDTV) enable to use this technology for new purposes and include it on the list of available components for the smart environment concept. In this realm the interaction with the used devices represent very important issue. Therefore, the goal of this paper is to conduct an experiment and provide with an analysis of verbal and non-verbal human to iDTV-based application voice interaction, and consequently compare it with other available interaction modalities. Moreover, this paper is focused primarily on specific target group. As television (TV) is a device widely spread among households, generally accepted by elderly and voice represents common way of seniors' interaction, the usage of voice to control applications for the given target group seems to offer relatively efficient way of interaction. Hence, a group of elderly users interact using particular methods to control applications in a TV-simulated computer-based environment in the study. The paper is organized as follows. In the following section the brief review of literature is presented. In the next section the research methodology is introduced. The fourth section outlines and discusses acquired results. Paper is concluded in the last section of the paper.

Theoretical background

According to the vision promoted in Ambient Intelligence the user should be able to interact with the devices in an unobtrusive way [4]. This can be achieved by providing natural ways for interaction [5]. It seems inevitable to substitute the windows-iconsmenus-pointers interactions by more inartificial ways of interactions [6]. Using voice to mediate commands in human to human interactions is quite natural and thus obviously there have been many attempts to implement this interaction method in human to computer communication as well. However, there are some inherent weaknesses of that limit the voice interaction [5]. Thus the voice is often used to complement other interaction modalities. The combination of voice with other interaction methods or computational techniques [7] helps to compensate for the above mentioned shortcomings [8]. Despite the effort of speech recognition researchers the accuracy is still lagging behind. The above mentioned limitations and problems draw speech recognition researchers to focus attention to non-verbal, prosodic features in speech to directly control devices or applications [9].

The non-verbal interaction techniques include control by continuous voice, rate-based parameter control by pitch, discrete control by tonguing, or volume based control [10]. Related research in non-verbal vocal interaction includes comparative study of speech recognition and non-speech input, employed to control game like applications such as Tetris [11]. The study points to verbal and non-verbal input as a low cost interaction modality for users with motor impairments or with limitation of mobility to control application in which fast reaction and otherwise hand-eye coordination would be necessary. In the [12] authors proposed voice-based assistive technology termed the "Vocal Joystick" which exploits a large set of continuous acousticphonetic parameters like pitch, loudness, vowel quality to offer an interaction modality to people with motor impairments. The combination of verbal and non-verbal gestures for cursor control was also explored in [11]. This study presents an approach in which, speech commands provided the direction. The authors argue that mapping speech rate to cursor speed is easy to understand

and allows the user to execute slow, high-precision cursor movements thus being more intuitive than using command only in verbally. In [13] the rate based interaction was used to control the speed of scrolling and zooming.

The experiment methodology

The main research question of this study is whether the non-verbal vocal interaction is an option in controlling applications for a group of elderly users. Therefore, the main goal of the research is to test suitability of verbal and non-verbal vocal interaction and non-vocal interaction for a specific user group. It appears that the voice is a preferable way of interacting for inexperienced technically less competent users as was the case of the studied user group. The thorough qualitative research is important since the preliminary testing and general perception showed a negative attitude toward non-verbal interaction using humming instead of words. The use of vocal interaction studied in this paper is only for the one way human to computer interaction in which the given application is controlled using a limited set of commands either verbal or non-verbal. In this way the problems of linearity and locality is not evinced. Using a limited set of one word commands which are normally placed in discrete time interval can be classified as isolated word recognition thus reducing the speech recognition process to three steps [14]. The non-verbal recognition requires only one step in that process.

The organization of the experiment is separated into two stages: design phase and implementation phase. Within the design phase a simulation of the user interface (UI) is tested, comments and discussion of users on ease-of-use, likes and dislikes are collected. Twelve volunteers all aged 60+ with an average age of 68 years are tested in this stage. There are three modalities of UI tested: automatic speech recognition (ASR), non-verbal vocal input (NVVI) and combination of ASR and NVVI. During the testing period the "wizard of OZ" method is used. This method consists in the usage of two computers – one used by observer for controlling of the application and one used by participants. Four tasks are performed by participants during one-hour period: (1) reaction to showing a picture – response task, (2) reproduction of a specific time interval – start/stop task, (3) navigation in a simple cell system – up/down/right/left task, and (4) navigation in a field grid where certain fields have to be visited – grid task. After each task the observers ask each participant to comment on the speed and ease of the application and to rate them on four level Likert scale ranging from I strongly agree (1) to I strongly disagree (4). The particular attention of the research is devoted to test the verbal and non-verbal vocal interaction in respect to the usability and preference of both modalities, the speed of processing the voice signals and the accuracy of recognition the voice signals.

Within the implementation phase the developed UI is tested in a structured way and difficulties or misunderstandings of users are identified. In the implementation phase the research involves 15 users all aged 60+ with an average age of 70 years. In this stage the voice interaction is confronted with more complex and technologically better developed interaction modalities - remote control with up/down/left/right cursor keys and Wiimote absolute pointing device (component of the Nintendo Wii game console). There are two applications used for testing purposes: physical exercise with virtual instructor and family photographs organizer and viewer created by members of the Czech Technical University team. Each application and each user is tested using all three interaction methods at once (in case of physical exercise) and one by one (in case of family photographs). Each volunteer works individually, being watched by two observers. The session duration with one application is 45-60 minutes depending on the volunteer's experience and personality. After finishing the work with application, volunteers are asked to fulfill the questionnaire with 20 questions and predefined four level Likert scale.

All technological limitations of TV as a medium are carefully handled during the experiment (e.g. TV screen action safe area, available image recognition, colors, or fonts). The simulated environment consisted of PC with standard operation system, Bluetooth receiver, USB module, TV card with remote control, browser Opera and Flash player application, widescreen TV Panasonic and additional equipment such as microphones, and testing applications. The voice interaction system consists of following modules: (a) Sound card and low-level library that provides access to microphone input; (b) monophonic pitch detection (MPD) module; (c) Vocal gesture recognition (VGR) module; (d) Automatic speech recognition (ASR) module; and (e) Vocal control layer (VCL) module.

Experiment results and discussion

Results of the design stage reveal that concerning the preference for verbal or non-verbal interaction the testing showed that approximately half of the users prefer verbal and half prefer non-verbal. The users who prefer verbal interaction strongly denounce the non-verbal humming. They regard humming as weird and even stupid. However, once the advantages of non-verbal interaction are explained they admit that it might be useful (e.g. for motorically disabled people). In general, users consider verbal commands as being slower than using humming or combination of both modalities (mean=1,56 (n=34); mean=1,21 (n=34); and mean=1,25 (n=8) respectively). A number of volunteers have problems to give out required different pitch of tone. Problems are also experienced with users wearing hearing aid. Some users attribute the uneasiness about the NVVI to the fact that they are unaccustomed to utter non-verbal sounds intentionally. Combination of two new ways of interaction is not generally accepted. The combination of verbal and non-verbal interaction is commented as „too complex for seniors, it´s better to have one type of control. More words are confusing." Only one person preferred the combination since it made her to think and concentrate. Two users recommended the combination mode for long interactions. Some of the users confuse left and right and some used words when humming is required. Results for easiness-of-use are summarized in Table 1. The results of the design phase also show that in case of elderly the new interaction method should be as simple as possible.

Concerning the speed of processing the verbal or non-verbal signal, none of the users note any significant difference although technically there is a noticeable difference in processing verbal commands especially in task 1. The explanation is that users are convinced that something will happen and they feel no need to get nervous because of the delay in application reaction. It also has to be noted that the user group consists of elderly commonly technically skilled people. Further, the applications are new to the users so they might even be comfortable that there is some time to rest.

Table 1. Subjective perception of easiness of use

	Device	Ease	N	StDev
Task1	ASR	1,56	8	0,527
	NVVI	1,44	9	0,527
Task2	ASR	1,75	8	0,707
	NVVI	1,33	9	0,500
Task3	ASR	1,50	8	0,756
	NVVI	1,78	9	0,833
Task4	ASR	1,88	8	0,991
	NVVI	1,89	9	0,991
	Combination	1,67	8	0,866

Regarding the accuracy of signal processing, the non-verbal interaction poses no particular problem. Only short and long humming is used and the users are able to produce these gestures appropriately. The threshold for short and long gesture is set in approximately one second and the recognizer distinguishes the time feature accurately. For the verbal interaction the accuracy is an issue. To produce comparably accurate interaction the "wizard of oz" method is employed. Thus the once the speech recognition engine wrongly recognizes command this is corrected by the wizard.

In the implementation phase results in a form of answers to twenty questions focused on particular aspects of tested applications are analyzed. Both applications receive the majority of negative answers to the Q15 (I would need guidelines or help while using this application.). Therefore, both applications seem to be quite complicated for participants when using given interaction methods and need to be supported with help or short illustrative guidelines. This conclusion is supported by the fact that in the Family tree application the Q2 (I always know what to do next with this application) receives the majority of negative answers as well. Although both applications receive positive reactions during the evaluation by participants in general, the main attention needs to be paid to the ways of interaction with applications in particular. In the questionnaire following questions are aimed at interaction modalities: 1 - The speed of application is adequate in all situations; 4 - Learning to operate this application initially is without problems; 8 - The use of remote control was engaging; 9 - The use of NVVI control was engaging; 10 - The use of Wii control was engaging; and 13 - The application almost always reacted as I had expected. Speed of Physical exercise and Family three applications is considered as sufficient by the majority of participants (average 1,47 and 1,73; modus 1 a 2 respectively). The most engaging way of interaction is traditional remote control, while NVVI is the least engaging (average 1,27 and 2,27; modus 1 and 2 respectively). There is not any negative experience with expected reaction of the application. In case of the Family three application 25 % of participants experience problems with the application operation. Similarly to the first application, the most engaging way of interaction is traditional remote control, while NVVI was the least engaging (average 1,27 and 2,67; modus 1 and 3 respectively). In addition to the aforementioned description of results focused on particular questions, a more comprehensive view of the responses can acquired by grouping data into the defined subscales. The investigated attributes are Efficiency (Q 1-4, 16-17), Affectivity (Q 5 – 7), Control (Q 8-12), and Learnability (Q13 – 15). This sub-group data is presented in Figure 1. The figure depicts that the affectivity sub-group achieved the most favorable ratings for both applications. It indicates that the participants' general emotional reaction to the applications is positive. Nevertheless, the speed and facility, with which participants feel that they are able to master the system, seems to be quite problematic.

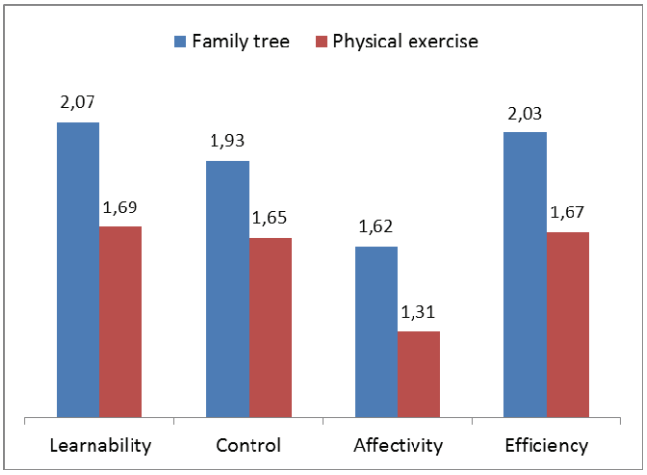

Fig. 1. Averaged ratings of subscales for both applications.

Conclusions

Extended functionalities of iDTV enable to consider this device as a component of smart environments concept together with intelligent refrigerators or wearable devices helping people to live healthier. This paper investigates possibilities voice interaction between elderly and iDTV applications. The tests reveal that voice interaction is positively perceived by elderly only if other traditional control devices are not available. The majority of them got used to use of remote control and therefore, when they can make a choice, they prefer it. Although participants are able to control applications using the NVVI and NVVI offers certain advantages, the use of NVVI needs some training and most probably this type of control will never be preferred by some users. The major disadvantage is that interacting using non-verbal gestures is considered awkward.

Albeit results show that elderly prefer traditional remote control, inclusion of iDTV into the smart

environment concept can be connected with all investigated modalities of interaction. All of them have their own disadvantages and advantages and further research can connect them with useful functionalities. For instance, while voice interaction can be used as a preventive tool for checking correct breathing or for speech problems indicating stroke related troubles, traditional remote control or a pointing device can help with identification of physical impairments, or unusual tremor of hands. Apparently, iDTV based applications in combination with other telecommunication devices such as smart mobile phones [15] have the potential to improve quality of life of elderly.

Acknowledgements

This paper was created with the support of the GAČR project SMEW, project num. 403/10/1310, and the Vital Mind project - project num. 215387. Authors would like to thank to all consortium members, especially to the Czech Technical University team for their technological support.

References

1. **Malý I., Mikovec Z., Vystrcil J., Franc J., Slavik, P.** An evaluation tool for research of user behavior in a realistic mobile environment // Personal and Ubiquitous Computing – Online First™, 2011. – P. 1–12.
2. **Bureš V.** Conceptual Perspective of Knowledge Management // E+M Economics and Management, 2009. – Vol. 12. – No. 2. – P. 84–96.
3. **Kim H., Swarts M., Lee S. T., Do E. Y.** HealthQuest: Technology That Encourages Physical Activity in the Workplace // Lecture Notes in Computer Science, 2010. – Vol. 6159. – P. 263–266.
4. **Mikulecký P.** Remarks on Ubiquitous Intelligent Supportive Spaces // Proceedings 15th American Conference on Applied Mathematics and Information Sciences. – Houston, United States, 2009. – P. 523–528.
5. **Kim H. C.** Weaknesses of Voice Interaction // Proceedings 4th International Conference on Networked Computing and Advanced Information Management. – Washington, United States, 2008. – P. 740–745.
6. **Oviatt S.** Ten myths of multimodal interaction // Communications of the ACM, 1999. – Vol. 42. – No. 11. – P. 74–81.
7. **Panuš J.** Parallel Computing for Modified Local Search // Proceedings 10th WSEAS International Conference on Applied Computer Science. – Iwate, Japan, 2010. – P. 508–513.
8. **Dang N. T., Tavanti M., Rankin I., Cooper M.** A comparison of different input devices for a 3D environment // International Journal of Industrial Ergonomics, 2009. – Vol. 39. – No. 3. – P. 554–563.
9. **Igarashi T., Hughes J. F.** Voice as sound: using non–verbal voice input for interactive control // Proceedings 14th annual ACM symposium on User interface software and technology. – Orlando, United States, 2001. – P. 155–156.
10. **Olwal A., Feiner S.** Interaction techniques using prosodic features of speech and audio localization // Proceedings 10th international conference on Intelligent user interfaces. – San Diego, United States, 2005. – P. 284–286.
11. **Sporka A. J., Kurniawan S. H., Mahmud M., Slavik P.** Non–speech input and speech recognition for real–time control of computer games // Proceedings 8th ACM SIGACCESS conference on Computers and accessibility. – Portland, United States, 2006. – P. 213 – 220
12. **Bilmes J. A., Li X., Malkin J., Kilanski K., Wright R., Kirchhoff K., Subramanya A., Harada S., Landay J. A., Dowden P., Chizeck H.** The vocal joystick: a voice–based human–computer interface for individuals with motor impairments // Proceedings Human Language Technology and Empirical Methods in Natural Language Processing. – Vancouver, Canada, 2005. – P. 995–1002.
13. **Igarashi T., Hinckley K.** Speed–dependent automatic zooming for browsing large documents // Proceedings 13th annual ACM symposium on User interface software and technology. – San Diego, United States, 2000. – P. 139–148.
14. **Buchsbaum A. L., Giancarlo R.** Algorithmic aspects in speech recognition: an introduction // Journal of Experimental Algorithmics, 1997. – Vol. 2. – No. 1 – P. 1–44.
15. **Krejcar O., Jirka J., Janckulik D.** Use of Mobile Phones as Intelligent Sensors for Sound Input Analysis and Sleep State Detection // Sensors, 2011. – Vol. 11. – No. 3. – P. 2334–2346.

Received 2012 03 11
Accepted after revision 2012 05 12

V. Bures. Interactive Digital Television and Voice Interaction: Experimental Evaluation and Subjective Perception by Elderly // Electronics and Electrical Engineering. – Kaunas: Technologija, 2012. – No. 6(122). – P. 87–90.

Advanced functionality of interactive digital television (iDTV) enables its integration into the smart environment concept, in which several modes of interaction with devices is investigated and implemented. Therefore, the aim of his paper is the experiment based analysis of verbal and non-verbal voice interaction in the human-to-iDTV application context with respect to a particular target group - elderly. Moreover, comparison with more complex interaction modalities is also conducted. The study reveals that voice interaction can be used by elderly in general. However, results show that if they can make a choice, they prefer traditional control devices such as a remote control. Elderly are also less tolerable to combination of two voice-based ways of interaction. Therefore, the study proves that voice interaction can be employed in connection with iDTV for purposes of elderly as the most common users of television devices and it can potentially contribute to quality of their lives. Ill. 1, bibl. 15, tabl. 1 (in English; abstracts in English and Lithuanian).

V. Bureš. Skaitmeninės televizijos ir balso interakcija: pagyvenusių žmonių vertinimas ir subjektyvus suvokimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2012. – Nr. 6(122). – P. 87–90.

Interaktyvios skaitmeninės televizijos (iSTV) funkcionalumas leidžia integruoti ją į sumaniosios aplinkos koncepciją, kuria remiantis tyrinėjama ir įdiegta keletas sąveikos su įtaisais būdų. Šio straipsnio tikslas yra eksperimentu pagrįsta verbalinės ir neverbalinės balso sąveikos analizė žmogaus ir iSTV taikymo tikslinei pagyvenusių žmonių grupei kontekste. Be to, atliktas palyginimas su sudėtingesniais sąveikos būdais. Tyrimas rodo, kad pagyvenę žmonės paprastai gali naudoti balsinę sąveiką, tačiau, jei jie galėtų rinktis, teiktų pirmenybę tradiciniams valdymo įtaisams (nuotoliniams pultams). Il. 1, bibl. 15, lent. 1 (anglų kalba; santraukos anglų ir lietuvių k.).