Development of interactive applications with multi-modal interfaces for mobile devices

Sergey Glazkov
St. Petersburg Institute for Informatics and Automations of RAS
Russia, 199178, St. Petersburg, 14 line, 39
s.glazkov@gmail.com

Abstract

There was a lot of research on speeding up the data entry on mobile devices, most aimed at adapting existing desktop input methods to mobile devices. But unfortunately, this approach has not yet yielded any significant results. The speed of data entry on mobile devices is far from the desired and the existing input methods require user's full attention and are almost impossible to use while in motion or while driving. Therefore, the existing interfaces impose strict restrictions on mobility of the user. All this merely confirms that mobile devices require fundamentally different interfaces and input methods like multimodal interfaces, where not a sole but any available in mobile interface modality can be used for data input. This paper describes the main restriction on development of multimodal interfaces for mobile devices and some approaches for their removal.

Index Terms: Multi-modal, Mobile, Interface.

I. Introduction

Multi-modal interfaces provide user interaction through various communication channels, processing two or more user information input methods in conjunction with a multimodal system output. Modalities are the channels that people can use to interact with each other or equipment. These channels examples are vision, hearing, touch, taste, smell, and proprioception (orientation of the body) [1]. Unlike traditional interfaces based on the keyboard and mouse input that are uni-modal, multi-modal systems provide more flexible use of information input streams. This approach grants a person the opportunity to choose the most convenient way to send and receive information.

II. MAIN PART

The multi-modal interfaces are based on modality types used by the person communicating with another person [2]. Table 1 shows the main modalities found in people interaction. The most often used modalities for the construction of interfaces are visual and sound modalities. Moreover, sense of presence and co-presence are marked out in the behavior of people who are directly in a team or participating in activities remotely using different telecommunication means [3]. In medicine and safety areas other biometrics like fingerprint, hand shape, face thermogram, ratinal scan are also analyzed.

Table 1. The main types of modalities and examples of their processing technologies.

Senses	Types of modalities	Examples of modalities	Examples of technologies
Vision	Visual	Hand gestures, body and lips movements, gaze direction, animated objects	Image analysis, virtual reality

Senses	Types of modalities	Examples of modalities	Examples of technologies
Hearing	Sound	Speech, sounds, melodies	Speech synthesis and recognition
Touch	Tactile	Touch, vibration,	Touchpad and touchscreen
Smell	Olfactory	Odor	Methods of chemical analysis and synthesis of
Taste	Taste	Taste	
Vestibular apparatus	Vestibular	Position of the body	Systems based on accelerometers and gyroscopes
Joints, nervous system	Proprioceptive	Relative position of body parts and their motion	Systems with exoskeleton

Depending on the used input and output modalities several main types of multi modal interfaces are defined (speech + gestures, speech + lip-reading, line of sight + indication + speech, etc.).

The most know multimodal concept is Bolt's (1980) original "Put That There" demonstration that combined speech and manual pointing during object manipulation. Since that time considerable efforts have been made in developing more general multimodal systems that process complex gestural input other than just pointing [4]. New systems have been extended to process different mode combination – the most noteworthy being speech and pen input, and speech and lip movements. In contrast to Bolt's initial concept, which was a limited prototype, significant progress also has occurred in building a variety of real applications that range from map-based and virtual reality systems for simulation and training, to field medic systems for mobile use in noisy environments, to web-based transactions and standard text-editing applications. Major progress has occurred in both the hardware and software for component technologies like speech, pen, and vision. In addition, the basic architectural components and framework have become established for designing more general multimodal systems. [5].

A. Mobile devices hardware and software features

Mobile devices applications development should take into account the limitations in size, weight, battery life as well as performance and internal memory. The small size of these devices and their usage conditions restrict the possibility of implementing a familiar interface (keyboard and mouse). Placing a comfortable full-size keyboard on a mobile device is quite difficult itself and making it suitable for touch-typing is not even assumed in situations where there is no opportunity to place a mobile device on a desk or other stable surface. Therefore, the most preferred input modalities for interaction with mobile devices are audio-visual (as the most close to the interpersonal communication).

The performance of mobile devices is much lower than the performance of personal computers, and even more than the performance of specialized servers. The most effective way to solve complex problems is to use client-server models. The implementation of a complex multi-modal interface with sound and vision as the main modalities and the possibility of parallel processing of more than one modality is not possible only by the means of a mobile device but becomes possible in a distributed system where the mobile device serves primarily as the data collector and all main processing is executed on the server. In this case the use of sophisticated algorithms of speech and image recognition such as hidden Markov models, neural networks, algorithms for global reconstruction is possible with no restrictions. But when using such an approach one should take into consideration the following features of mobile devices as well as communication networks and wireless data transmission:

- CPU clock. Processor performance of a mobile device can be extremely small. This slows down the initial processing of information and speed of sending and receiving data, as well as the speed of encryption / decryption of data sent over secure communication channels.
- Half-and full-duplex radio. Some devices do not support full-duplex mode, that is they cannot simultaneously transmit and receive data. This increases the delay and affects the speed of the device response to user actions.
- Lag. Majority of the mobile devices use the energy-saving protocols. In addition, there are legal restrictions on the specific degree of absorption (Specific Absorption Rate, SAR). SAR is a degree of the biological effects of RF energy that characterizes the safety of cell phones and other mobile devices. These restrictions further increase the response time of mobile devices.
- Throughput. Though mobile devices supporting GPRS (General Packet Radio Services) can operate at the theoretical maximum speed 171.2 kilobits per second (Kbps), most phones still use the Circuit Switched Data technology that supports data rates only up to 9.6 Kbps. It means that mobile devices are especially sensitive to the costs associated with the work of the protocols, and to the volume of transmitted data.
- Discretionary participation feature connections. The mobile device can be moved from the area of sustainable outreach to the zone of partial coverage or even to places where there is no connection at all. The quality of communication is affected by tall buildings, nearby bridges and other conditions. It is therefore necessary to monitor the signal strength and in its absence switch device interface to the local mode notifying the user.

Also particular attention should be paid to such aspects as minimizing the interface response time and the use of different encryption techniques to protect user data.

B. The main problems that arise when a user interacts with a mobile device

It should be said that most difficulties with mobile applications development are connected with increasingly complex tasks users are trying to accomplish with their help and at the same time the lack of psychological knowledge usage in designing user interfaces.

At the heart of many theoretical, practical and empirical studies on the modeling of communicative human-computer interaction stands the ultimate question: what does the developer need to know about the user to organize an effective human-machine interaction. User communication with the mobile computer system may be facilitated if the software form and structure will be consistent with the characteristic features of human perception, thought and action.

Unfortunately, among the existing methods of UI quality evaluation informal methods with significant subjective assessment are dominating. Common approach is in an attempt to identify general, standard characteristics that distinguish high-quality interfaces and with these results create a methodology for evaluating graphical user interfaces effectiveness. Typically, the existing methods consider only one aspect of evaluating the graphical user interface effectiveness pointing out the single most important criterion.

All above concerns user interfaces in general. Now problems associated exclusively with mobile devices will be considered.

Users face a growing variety of mobile devices, their operating systems and, consequently, user interfaces. Usability of a mobile device is determined, most often, by user habits. Interfaces With the development of mobile hardware interfaces become more complex, large number of settings, applications and peripherals appear. Device displays are quite small and do not allow to place all the frequently used settings on a single screen. So the interface creates additional work desks, multiple transitions on various menus, all this does not make the interface easier. Often, owners of mobile devices do not even know all the options granted to them by mobile device or are unwilling to use them because of uncomfortable realization of some functions.

Mobile phones ergonomics are mainly aimed at providing the voice interaction between a pair of users. Mobile phone numeric keypad was originally designed for dialing a subscriber number, but with the adoption of additional services provided by telecom operators, allowed to enter text messages, too. The letters on such keyboards are arranged in alphabetical order. By pressing the same button on the numeric keypad, you can select the desired character, one of the appointed to this key. The average rate of text entry when using the above method is 8 words per minute. More experienced users of mobile devices can gain an average of 20 words per minute, thanks to technology of predictive input T9 [6], but such input fully captures human visual attention. With the adoption of touch screens, push-button keypad began to give place to software counterparts, full-size QWERTY keyboard appeared on the display. Though input became more usual, compared to a desktop computer, but still is very slow.

There was a lot of research on speeding up the data entry on mobile devices [7, 8, 9], most aimed at adapting existing desktop input methods to mobile devices [10]. But unfortunately, this approach has not yet yielded any significant results. The speed of data entry on mobile devices is far from the desired and the existing input methods require user's full attention and are almost impossible to use while in motion or while driving. Therefore, the existing interfaces impose strict restrictions on mobility of the user. All this merely confirms that mobile devices require fundamentally different interfaces and input methods. These interfaces must provide user interaction that will remove following restrictions:

- Failure to input data in motion.
- Failure to input data without having to look at the display device.
- Failure to use the device without direct contact

The most promising way to resolve the failures described above is to use speech modality in the base of mobile device multimodal interface. Unfortunately, the efficiency of continuous-speech recognition systems, especially in difficult environment (for instance in a noisy one) has not reached the level, needed for comfortable use of the mobile device. Nevertheless, the use of additional modality, e.g. visual or tactile will make it possible significantly increase the quality of human computer interaction.

III. Conclusion

A lot of research was conducted in the past years on multimodal interface development with the use of different modalities. Unfortunately not so much work was done for the field of the mobile devices, where the multimodal interfaces so comfortable and needed. This situation is caused by the high demand of the multimodal interfaces for the computational, power and storage resources that are much lower in the mobile devices. The use of client-server architecture enables the implementation of complex models where the mobile device serves primarily as the data collector and all main processing is executed on the server.

REFERENCES

[1] Oviatt, S. L., “Mutual disambiguation of recognition errors in a multimodal architecture. Proceedings of the Conference on Human Factors in Computing Systems (CHI’99)” *New York: ACM Press*, pp. 576-583, 1999.

[2] Шиффман Х, “Ощущение и восприятие.” *СПб:Питер*, 928 с., 2003.

[3] Akker, R., Hofs, D., Hondorp, H., Akker, H., Zwiers J., Nijholt, A., “Supporting Engagement and Floor Control in Hybrid Meetings.” *Springer, LNAI 5641*, pp. 276–290, 2009.

[4] Oviatt, S. L., Cohen, P. R., “Perceptual user interfaces: multimodal interfaces that process what comes naturally” *Communications of the ACM*, vol. 43, Issue 3 pp. 45–53, 2000.

[5] Oviatt, S. L., Cohen, P. R., “Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions” *Human-Computer Interaction*, vol. 15, Issue 4 pp. 263–322, 2000.

[6] James, C.L., Reischel, K.M., “A input for mobile devices: comparing model prediction to actual performance. In Proc. of the SIGCHI conference on Human factors in computing systems” *Seattle, WA: ACM Press*, pp. 365–371, 2001.

[7] Harrison, C., Tan, D. Morris, D., “Skinput: Appropriating the Body as an Input Surface.” *In Proc. of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, April 10 - 15, 2010. CHI '10. ACM, New York, NY,* pp.453-462, 2010.

[8] Harrison, C., Hudson, S., “Minput: Enabling Interaction on Small Mobile Devices with High-Precision, Low-Cost, Multipoint Optical Tracking.” *In Proc. of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, (Atlanta, Georgia, April 10 - 15, 2010). CHI '10. ACM, New York, NY,* pp.1661-1664*,* 2010.

[9] Schwarz, J., Harrison, C., Mankoff, J., Hudson, S. E., “Cord Input: An Intuitive, High-Accuracy, Multi-Degree-of-Freedom Input Method for Mobile Devices.” *In Proc. of the 28th Annual SIGCHI Conference on Human Factors in Computing Systems, (Atlanta, Georgia, April 10 - 15, 2010). CHI '10. ACM, New York, NY,* pp.1657-1660*,* 2010.

[10] Sears, A., Zha, Y., “Data Entry for Mobile Devices Using Soft Keyboards: Understanding the Effects of Keyboard Size and User Tasks” *International Journal of Human-Computer Interaction, Vol. 16,* Issue 2, pp. 163 – 184, 2003.