A Prototype of Gesture-based Interface

Zhiyuan Lu, Xiang Chen*, Zhangyan Zhao
Department of Electronic Science and Technology
University of Science and Technology of China
xch@ustc.edu.cn

Kongqiao Wang
Multimedia Technologies Lab
NOKIA(CHINA) Investment CO., LTD
Kongqiao.Wang@nokia.com

ABSTRACT
This paper introduces a novel gesture-based human-machine interface prototype, which consists of a wearable belt embedding with four surface electromyography (SEMG) sensors, a tri-axis accelerometer and an application program running on NOKIA 5800XM. The sensor b elt captures hand gestures by acquiring SEMG and acceleration (ACC) signals from forearm, and sends the m out via Bluetooth. The application program receives the data and transl ates them into control commands of a given interaction application. Experimental results of two test schemes conducted on hand gesture recognition and media player operation demonstrate the v alidity of the proposed gesture-based interface prototype.

Author Keywords
Gesture-based interface, SEMG, accelerometer

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces – input devices and strategies, interaction styles

General Terms
Performance, Design

INTRODUCTION
As mobile devices become more and more powerful and smart, the predominant forms of input interfaces such as keyboards and t ouch pads may not always satisfy users' needs nowadays. They s eem to be inconvenient sometimes when users' hands are occupied. And keyboards might take up too much space f or handheld devices to ensure go od operability.

Multimodal interaction, which aims to provide more natural and convenient interaction experience, has become a research focus recently. As one form of multimodal interaction, gesture-based interaction is now being popularized. Many kinds of gesture-based interface have been developed for computers [1], household appliances and mobile devices. Some cell phones can mute themselves when they are turned over. Several gesture-based games for iPhone are v ery popular in recent years, including a basketball game named iBasketball. In another game named Labyrinth, user can control a virtual steel ball just by tilting his mobile phone. This novel interaction technology attracts more and more attention from users, and several similar applications including Labyrinth have be en downloaded more than 25 million times total from Nokia Ovi Store [7].

This paper introduces a novel gesture-based human-machine interface prototype for mobile devices. The prototype is composed of the gesture capturin g device, the gesture recognition algorithms and the interaction program, which translates recognition results into commands. The gesture capturing device integrates with a tri- axis accelerometer, four dry surface electromyography (surface EMG, SEMG) electrodes and a Bluetooth module. Recognition algorithms based on acce leration (ACC) and SEMG signals run on the mobile device. Either pre-defined or user-defined gestures are sup ported by this prototype. Mobile phones can be remotely controlled by gestures when using this interface. The demonstrated functions include accepting or rejecting a phone call, sending short messages, controlling the media player, and so on.

MOTIVATION
There are many potential use cases f or gesture-based interface. For exa mple, if one's pho ne rings during a meeting, he might turn off the ringer using a subtle gesture immediately rather than finding his p hone hurriedly. When driving, he might reject a phone call and send a short message to the caller by just grasping the steering wheel.

Moreover, despite the fact that smart phones can be used to take photographs, play videos, browse websites, read books and do other interesting things, their small screens are disappointing in some places, especially when one want to share his screen with others. As a solution, micro-projectors have been introduced into the field of mobile interaction. Some phones are integrated with projectors, and so me can be connected to projectors for better interaction experience. When sharing the screen using the projector, interaction using gestures remotely seems more convenient than touching the phone.

How to capture and recogn ize gestures are two of the crucial issues of gesture-based human-machine interface. The accelerometer and SEMG sensor provide two potential technologies for gesture sensing. Accelerometers can measure both dynamic accelerations like vibrations and static accelerations like g ravity, and they are good at

MobileHCI 2011, Aug 30–Sept 2, 2011, Stockholm, Sweden.

capturing noticeable, large-scale gestures. Most of the existing gesture-based functions and games for mobile devices are based on ACC signals measured by the embedded accelerometer [3]. Surface EMG signal, which represents the contract level of related muscles during a gesture execution, has advantages in capturing fine motions such as extension and flexion of wrist and fingers. Recently, the EMG-based hand gesture interaction has also attracted more and more attentions, and lots of researches have been conducted on EMG-based hand gesture recognition and real-time interfaces for mobile and wearable devices [2, 6].

Because accelerometer and SEMG sensor have their own advantages in capturing hand gestures, we believe that combined sensing approach could improve the performance of hand gesture recognition significantly, and make it more feasible to realize natural gesture-based interaction. For this reason, a novel human-machine interaction prototype is proposed in this paper to demonstrate the feasibility of constructing gesture-based interfaces using accelerometer and SEMG sensors.

DESIGN

Gesture capturing Device

As Figure 1 shows, the gesture capturing device is a wearable elastic belt with four electrodes and a main board, on which the MCU (C8051F411), tri-axis accelerometer (MMA3761L), Bluetooth module produced by Ommitek Electronics Co. and a 1000mAh lithium battery are welded as Figure 2 shows. The MCU integrates with a serial port (UART) and a 12-bit A/D, which converts each channel of analog signals into digital signals with the sample rate of 600sps (samples per second). The accelerometer provides three channels of analog ACC signals along the X, Y, Z axes. Each electrode connected to the battery and A/D by wires acquires one channel of SEMG signals, amplifies them by 500, and filters them within 20Hz~300Hz band-pass, because most of the signals' energy is contained in this range[5]. These electrodes with amplifiers and filters in them are fixed to the belt in series. Three channels of ACC signals and four channels of SEMG signals from A/D are stored in MCU temporarily and then sent to Bluetooth module though UART.

Figure 1. Snapshot of the gesture capturing device

The end-users just wear the belt-like device on forearm and connect it to mobile phone via Bluetooth, and then it will acquire gesture information from human body and send the information to cell phone within 10 meters. A charge of the battery costs 1-2 hours and can support nearly 10 hours continuous work.

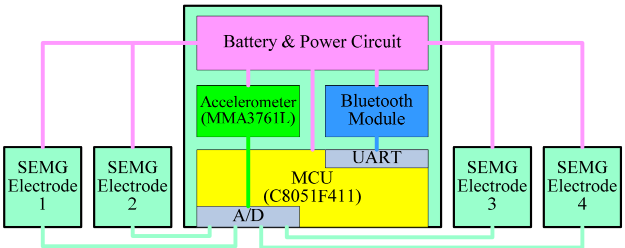


Figure 2. Structure of the gesture capturing device

Data Transmission

Data transmission in this prototype is based on Bluetooth because it is well supported by both MCU and our mobile phone. The baud rate of Bluetooth is set to 115200bps to support the gesture capturing device, which sends out 8400 bytes per second. The Bluetooth module receives data from MCU and finishes all the things automatically. The application program running on the mobile phone reads data from Bluetooth using APIs without caring about the Bluetooth protocols. Though few errors occur in the transmission based on Bluetooth, it is still very difficult for the program on the mobile phone to use these data when facing the endless stream of sensor data without any protocol. So the following encoding scheme is designed.

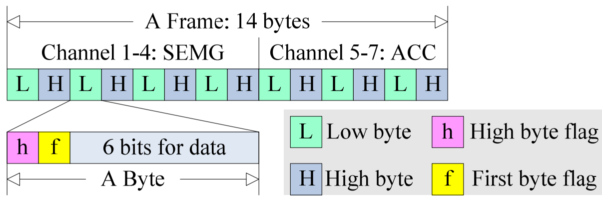


Figure 3. Protocol for transmission

As Figure 3 shows, sensor data are transmitted by frame. A frame contains 14 bytes, 2 bytes for each channel, and the low byte is transmitted firstly. Each data takes up 12 bits in 2 bytes, and the 2 bits left in each byte carry additional information for synchronization and error detection. In our protocol, each byte carries 6 bits of sensor data, which are stored in its lower 6 bits. The "first byte flag" becomes "1" only in the first byte of a frame, and the "high byte flag" becomes "1" when it is a high byte, what means this byte carries bit 6~11 of the sensor data. Application program on the cell phone receives data from Bluetooth, decodes them, and pushes them into a buffer. If an error is detected in one frame, all of the data carried by this frame will be discarded.

Gesture Recognition Algorithms

SEMG signals acquired from forearm have advantages in capturing finger and wrist motions, but are quite insensitive to large-scale movements. ACC signals contain rich information of large-scale movements such as velocities and trajectories. So in our algorithms (Figure 4), segmentation is based on SEMG signals and the first classifier is based on ACC signals. Two thresholds of the mean absolute value (MAV) of SEMG signals are established for segmentation, one for starting point detection and the other for end point detection. The first classifier put gestures into two categories according to their scales: if the variance of ACC signals exceeds a given threshold, it is a large-scale gesture, and vice versa.

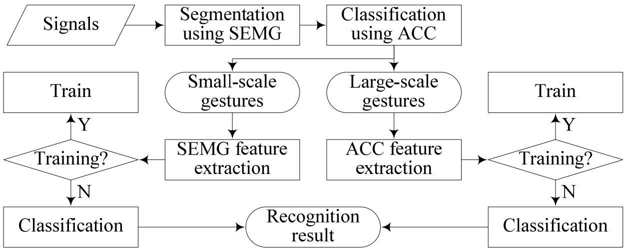


Figure 4. Flow chart of gesture recognition

For small-scale gestures, only SEMG signals are used for further classification. MAV, zero crossing rate, waveform length and Autoregressive (AR) model coefficients etc. have been proved to be effective features [4]. Considering both the classification performance and the computing power of mobile devices, MAV and 3-order AR model coefficients are adopted in this prototype. Bayes classifier based on minimum error probability is employed for classification. The classifier should be trained before being used because of the randomicity of SEMG signals. Based on the experimental results, we found that 32 samples of each pattern are enough for an untrained classifier to reach a stable classification performance.

For large-scale gestures, only ACC signals are used for further recognition. K-nearest neighbor (k-NN) classifier is used to classify the time domain features. More specifically, for each channel of ACC signals, the DC component of the segment is removed firstly, and then the segment is divided into 16 equal frames. Average value of each frame is computed and scaled to [-1, 1]. The result sequences from 3 channels are regarded as one sequence of coordinates of points in a 3D word. The classifier is trained with 8 training samples each pattern and its recognition result is based on three nearest training samples to the unclassified sample. The distance between a training sample and an unclassified sample is defined as the sum of distances between each point in the unclassified sample and the matching point in the training sample. Matching points are found using dynamic time warping (DTW).

Interaction Scheme with Cell Phone

An interaction program is developed on Nokia 5800XM to translate gestures into control commands, so that users can operate their cell phones using gestures. 4 small-scale gestures shown in Table 1, and 5 large-scale gestures shown in Table 2 are defined to accomplish a given interaction task. More gestures can be defined by users. When doing large-scale gestures, the user should wave his arm, and keep his hand tense.

Name	Wrist Extension (EXWR)	Wrist Flexion (FLWR)	Palm Extension (EXPM)	Hand Grasp (HDGP)
Image				

Table 1. Definition of small-scale gestures

Key pressed messages are used by the operating system of our phone to tell programs key press events. All of the key pressed messages are supported by the phone although it has only 3 keys. Each gesture is mapped to a key-press message in this prototype. Our program sends a key-press message when each gesture is done. So that user can do operations using gestures just as he pressed the keyboard.

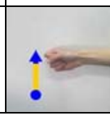
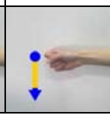
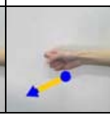
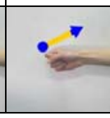
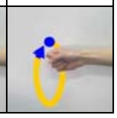
Name	Up	Down	Left	Right	RR
Image					

Table 2. Definition of large-scale gestures

Multiple interaction applications including accepting or rejecting phone calls, sending short messages, controlling the media player and so on can be implemented with these defined 9 kinds of gestures.

TESTING SCHEMES AND RESULTS

Two types of testing schemes have been conducted to testify the performance of this gesture-based human-machine interface prototype. Our subjects are college students aged 21-25 who have used cell phones for more than 2 years, so that they are familiar with cell phone operations and can easily accept novel things.

Testing of Gesture Recognition

6 subjects (2 female) took part in the user-specific testing, in which training and testing data were from the same subject. Each subject repeated each small-scale gesture 32 times and large-scale gesture 8 times in training mode to get his own template before testing. Each gesture was repeated 20 times during the testing to calculate recognition accuracy. All of these 9 classes of gestures were well classified with above 95% accuracies as shown in Figure 5.

5 subjects (2 female) participated in the user-independent testing, in which the classifier was trained with data from 5 well-trained users. Each gesture was repeated 20 times during the testing and about 90% average accuracies were obtained. As shown in Figure 5, most gestures can be well distinguished. Compared with the result of user-specific test, a conclusion can be drawn that further improvement of accuracies will be obtained, in condition that the classifier is trained with the user's own data. Accuracies of EXPM and HDGP drop obviously due to individual differences.

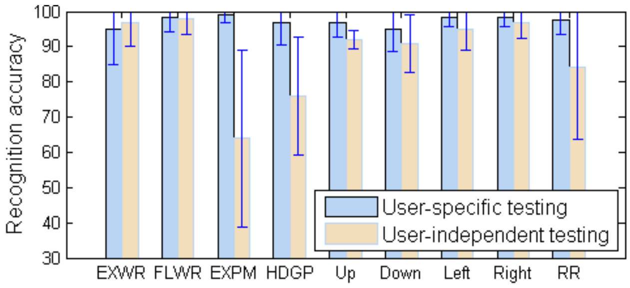


Figure 5. Recognition accuracies of two testing

Testing of Interaction

Two fresh users (FU) and two experienced users (EU, 1 female) have joint the interaction testing, in which each subject should open media player to play the first song in media library and then quit only with gestures. Subjects can practice 3 times at most before the testing to become familiar with this task. The same task needs 8 cli cks and more than 10 seconds using the touch pad on Nokia 5800XM. Figure 6 shows a typical sequence of gestures to finish this task. 11 gestu res are required here because gestures are mapped to key-press messages in this prototype, and this interaction task requires 11 key-press messages.

Figure 6. Sequence of gestures in the interaction task

Every subject was allo wed to correct any mistakes by gestures. During the testing, the number of gestures and time taken were recod ed to evaluate the interaction performance. Each subject repeated this interaction task 5 times. Table 3 gi ves the nu mber of gestures and time for each subject to finish the interaction task.

	EU 1	EU 2	FU 1	FU 2
Number	11.6±0.9	11.8±1.8	13.8±2.3	14.0±1.6
Time (s)	24.5±2.4	29.2±2.9	35.2±6.3	35.7±4.7

Table 3. Interaction performance parameters

The result shows that most users, even fresh users can accept and master gesture-based interaction easily, f or gestures are natural and easy to learn. A lthough doing gestures takes up more time than using keyboards or touch pads, the gesture-based interface provides a new choice and is useful in some use cases. The differences between fresh users and experienced users show that experience is an important factor for improving the performance, and more practice can make this interface more effective.

CONCLUSION

In order to d emonstrate the feasibility of hand gesture recognition and interaction technology based on SEMG and ACC signals, a nov el gesture-based int erface prototype is proposed in this paper. In this prototype, a gesture capturing device is desi gned to capture hand gesture, a SEMG and ACC-based recognition scheme is i mplemented to classify hand gesture, and an i nteraction program is developed to demonstrate the feasibility of gesture-based interface. The experimental results from two testing schemes conducted on hand g esture recognition and media player operation demonstrates that the proposed prototype has satisfying interaction performance. Because it is easy to learn and to use even for fresh users, we expect this gesture-based interface prototype can b e accepted by mobile phone owners, especially by young people.

ACKNOWLEDGMENTS

This work is partly supported by National High Technology of Research and Development Program of China (86 3 Program) under Grant No. 2009AA01Z322, the Fundamental Research Funds for the Central Universities under Grant No. WK2100230002 and WK2100230005.

REFERENCES

1. Benko, H., Saponas, T.S., Morris, D. and Tan, D. Enhancing input on and above the interactive surface with muscle sensing. *In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces 2009*, ACM, 93-100.
2. Costanza, E., Inverso, S.A. and Allen, R. Toward subtle intimate interfaces for mobile devices using an EMG controller, in *Proc. CHI 2005*, ACM Press (2005), 481-489.
3. Liu, J., Zhong, L., Wickramasuriya, J. and Vasudevan, V. User Evaluation of Lightweight User Authentication with a Single Tri-Axis Accelerometer. In *Proc. MobileHCI 2009*.
4. Liu, X., Zhou, R., Yang, L. and Li, G. Performance of various EMG features in identifying arm movements for control of multifunctional prostheses. In *Proceedings of Youth Conference on Information, Computing and Telecommunication 2009*, IEEE 2009, 287-290.
5. Mulavara, A.P., Sastry, K.L.A. and Verstraete, M.C. Frequency characterization of EMG activity during gait. In *Proceedings of the Annual Conference on Engineering in Medicine and Biology 1993*, IEEE (1993), 1215-1216.
6. Saponas, T.S., Tan, D.S., Morris, D., Balakrishnan, R., Turner, J. and Landay, J.A. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 167-176.
7. Ovi Store downloads break new records http://conversations.nokia.com/2010/04/21/ovi-store-downloads-break-new-records/.