LUI: Lip in Multimodal Mobile GUI Interaction

Maryam Azh[1,2], Shengdong Zhao[1,2]
[1] NUS-HCI Lab
National University of Singapore
13 Computing Drive, Singapore 117417
[2] NUS Graduate School for
Integrative Sciences & Engineering
28 Medical Drive, Singapore 117456
{maryam, zhaosd}@comp.nus.edu.sg

ABSTRACT
Gesture based interactions are commonly used in mobile and ubiquitous environments. Multimodal interaction techniques use lip gestures to enhance speech recognition or control mouse movement on the screen. In this paper we extend the previous work to explore *LUI*: lip gestures as an alternative input technique for controlling the user interface elements in a ubiquitous environment. In addition to use lips to control cursor movement, we use lip gestures to control music players and activate menus. A LUI Motion-Action library is also provided to guide future interaction design using lip gestures.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces - *Interaction styles*

Keywords
LUI, gesture input, multimodal mobile interaction

INTRODUCTION
Mobile interactions often happen in ubiquitous environments, where multi-tasking may prevent the human user to use of all available attentional and physical resources for interaction. There are many cases in which both hands are occupied and cannot be used for interaction with the user interface. For instance, while holding on to the bus handrails with one hand and keeping a tablet with the other, or in case of wearing gloves on both hands while holding a mobile phone, it is challenging to use fingers for interaction. Alternative interaction techniques that leverage multiple modalities can be useful in such scenarios (i.e., [11]). One of the under-utilized input channels is the gestures produced by the mouth, which can potentially be used not only to speak, but also to interact with computer interfaces. With the advancements in computer vision, it is possible to utilize webcam or device-integrated camera capabilities to interact with the user interface through mouth gestures. Our goal is to explore such potential as a complementary modality for interacting with computers.

ICMI '12, October 22–26, 2012, Santa Monica, California, USA.

MOUTH GESTURES FOR MULTIMODAL USER INTERFACES
Deploying mouth information as source of information is widely researched in human computer interaction. Deformation of articulatory organs such as *lips*, *tongue* and *teeth* conveys useful information to complement vocal speech data, can be used solitarily as input modality to interact with the user interface. *Lip* movements are most widely researched as opposed to tongue or teeth, as more dynamic and distinctive shapes can be identified through lip motion [7].

Lip gesture in audio visual speech recognition
Lip movement information is used as complementary visual data to enhance speech recognition. Top-down model-based, or bottom-up feature-based approaches are used to analyze lip image sequences, which utilize various lip movement features such as inner and outer lip contours, shape, color and motion [6]. The four main approaches for extracting visual information for speech recognition, include *image-based*, whereby the mouth image is used as a feature vector, *visual-motion-based*, which assumes relevant speech information within the mouth motion data, *geometric-feature-based*, which bases on height or width of the mouth opening as essential features, and finally *model-based*, which relies on a model of visible speech features such as lip contours [3]. It is expected that the bottom-up approaches prevent systematic model errors and the top-down approaches are more noise resistant.

Lip gesture as the input modality
Lip gestures have been studied as a solution to hands-free control of the user interface. A previous work [10] utilizes six distinctive mouth shapes for handling mouse events. The shapes are identified through vertical and horizontal movements of upper and lower lip centers or mouth corners. Another interesting work [2] tracks lip movements for cursor navigation.

LUI: THE LIP USER INTERFACE
Lip User Interface (LUI) is an application for mobile computers, including laptops, phones and tablets with a built-in front camera, which enables user interaction with the GUI through lip gestures. An overview of the proposed system is illustrated in figure 1.

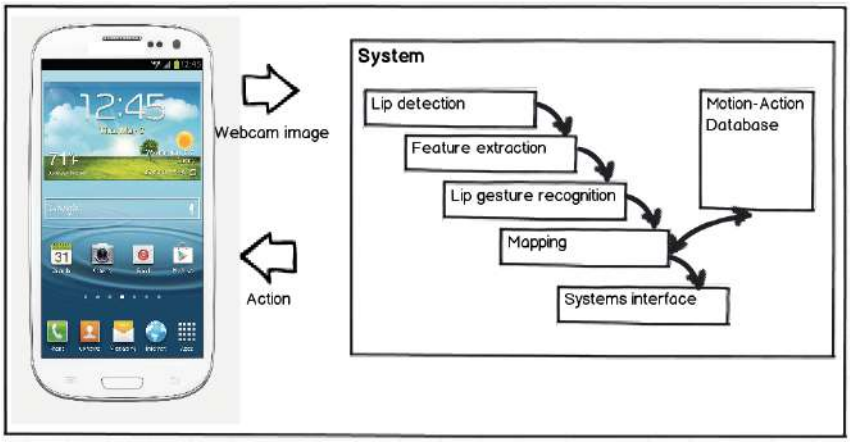


Figure 1. LUI system architecture.

We used an Android-based phone to create a hands-free scenario in a mobile context. The mobile phone built-in camera captures user lip movements in real time. Using the lip detection module implemented in OpenCV library, the frames captured from the video stream are processed and the recognized user lip gestures are mapped to user interface commands, based on the predefined motion-action library.

Table 1 lists some of the predefined motion-action entities. Our system currently supports basic user interface actions such as navigation and selection. A library of distinctive lip gestures can be found in [8] and [4]. The following section provides more detail on the implementation.

Implementation

LUI first detects the lip from the phone camera; processes the lip gesture; and executes action events on the phone user interface. Our system is composed of the following five modules:

Lip detection module: the inner and outer lip contours are identified in OpenCV based on previous work [1, 9].

Feature extraction module is built based on a previous lip feature extraction algorithm [1]. Shape features include lip shape, lip orientation and lip center.

Lip gesture recognition module deploys a Hidden Markov Model for finding the best match between the lip gesture and the image library, which is a commonly used method for lip movement recognition [5].

Mapping module: maps the lip gesture to the *motion* element in the motion-action database. Then the module signals the systems interface module to evoke the corresponding *action* element.

Systems interface module: triggers the action at the Android operating system layer.

User Interface and Interaction

LUI system is developed as an android application and installed on the phone. When the user opens the app, the front camera is activated and lib motions are triggered in real time. The application can be turned off from the Android status bar.

Usage Scenarios

We propose the following navigation and selection scenarios to test our system in the mobile environment.

- *Navigating through the music player interface, including song selection, play, volume change, etc.*
- *Calling a phone number from the contact list.*
- *Opening and scrolling through a PDF document, etc.*

CONCLUSION AND FUTURE WORKS

Our initial implementation of LUI supports basic interactions with lip gestures. In addition to the aforementioned scenarios, LUI can be used for people with hand, vision and speech disabilities as a convenient channel of interaction. User study is yet to be conducted to validate the feasibility of our system for various scenarios in ubiquitous environments.

REFERENCES

1. Çetingül, H. E., Yemez, Y., Erzin, E. and Tekalp, A. M. Discriminative Analysis of Lip Motion Features for Speaker Identification and Speech-Reading. *IEEE Transactions on Image Processing*, *15*(10), 2879-2891 (2006).
2. Dalka, P. and Czyzewski, A. Human-Computer Interface Based on Visual Lip Movement and Gesture Recognition. *IJCSA*, *7*(3), 124-139 (2010)
3. Dupont, S. and Luettin, J. Audio-Visual Speech Modeling for Continuous Speech Recognition. *IEEE Transactions on Multimedia*, *2*(3), 141-151 (2000).
4. Kanade, T., Tian, Y. and Cohn, J. F. Comprehensive Database for Facial Expression Analysis. *FG 2000*, 46-53.

5. Kaynak, M. N., Zhi, Q., Cheok, A.D., Sengupta, K., Zhang, J. and Ko, C.C. Analysis of lip geometric features for audio-visual speech recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, Part A, *34*(4): 564-570 (2004).
6. Matthews, I., Cootes,T. F., Bangham, J. A., Cox, S. and Harvey, R. Extraction of Visual Features for Lipreading. *IEEE Trans. Pattern Anal. Mach. Intell*., *24*(2): 198-213 (2002).
7. Pantic, M. and Rothkrantz, L. J. M. Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Trans. Pattern Anal. Mach. Intell*., *22*(12): 1424-1445 (2000).
8. Tian, Y., Kanade, T. and Cohn, J. F. Recognizing Lower Face Action Units for Facial Expression Analysis. *FG 2000*, 484-490.
9. Tian, Y., Kanade, T., and Cohn, J.F. Robust lip tracking by combining shape, color and motion, In *Proc. of ACCV '00*, 1040-1045.
10. Tu, J., Tao, H. and Huang, T. S. Face as mouse through visual face tracking. *Computer Vision and Image Understanding*, *108*(1-2): 35-40 (2007).
11. Zhao, S., Dragicevic, P., Chignell, M.H., Balakrishnan, R., Baudisch, P. earpod: Eyes-free menu selection using touch input and reactive audio feedback. In *Proc. of CHI 2007*, 1395-1404.

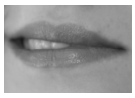
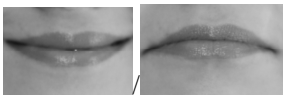
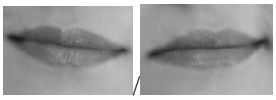
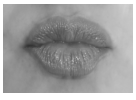
Motion		**Action**	
Lip Gesture	**Description**	**Action Event**	**Action Type**
	Neutral	-	-
	Lips bitten inward	Home	Main Buttons
	Left corner of the lips open and pulled	Back	
	Lips pulled inward	Options	
	Corner of the lips pulled up/down	Navigate up/down	Navigation
	Left/right corner of the lips pulled	Navigate left/right	
	Centre of the upper lip pulled up/ center of the lower lip pulled down	Zoom in/out	
	Lips pushed outward with mouth closed	Click	Selection

Table 1: LUI Motion-Action library