LUI: Lip in Multimodal Mobile GUI Interaction

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

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.

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.
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 lip 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


<table>
<thead>
<tr>
<th>Lip Gesture</th>
<th>Description</th>
<th>Action Event</th>
<th>Action Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Neutral" /></td>
<td>Neutral</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><img src="image" alt="Lips bitten inward" /></td>
<td>Lips bitten inward</td>
<td>Home</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Left corner of the lips open and pulled" /></td>
<td>Left corner of the lips open and pulled</td>
<td>Back</td>
<td>Main Buttons</td>
</tr>
<tr>
<td><img src="image" alt="Lips pulled inward" /></td>
<td>Lips pulled inward</td>
<td>Options</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Corner of the lips pulled up/down" /></td>
<td>Corner of the lips pulled up/down</td>
<td>Navigate up/down</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Left/right corner of the lips pulled" /></td>
<td>Left/right corner of the lips pulled</td>
<td>Navigate left/right</td>
<td>Navigation</td>
</tr>
<tr>
<td><img src="image" alt="Centre of the upper lip pulled up/center of the lower lip pulled down" /></td>
<td>Centre of the upper lip pulled up/center of the lower lip pulled down</td>
<td>Zoom in/out</td>
<td></td>
</tr>
<tr>
<td><img src="image" alt="Lips pushed outward with mouth closed" /></td>
<td>Lips pushed outward with mouth closed</td>
<td>Click</td>
<td>Selection</td>
</tr>
</tbody>
</table>

| Table 1: LUI Motion-Action library |