October 14th, 2010

Dr. Andrew Rawicz
School of Engineering Science, SFU
8888 University Drive
Burnaby, British Columbia
V5A 1S6

Re: ENSC440 Functional Specification for a Facial and Speech Identification System (FASIS) for Nokia Devices

Dear Dr. Rawicz:

The attached document, *Functional Specification for a Facial and Speech Identification System (FASIS) for Nokia Devices*, outlines the functional behavior of our project for ENSC440/ ENSC305. The purpose of this project is to implement a security system to provide secured mobile log-in based on a combination of successful facial match and speech recognition from the user. The project, supported partially by Nokia by providing us with the device, will become a prototype system that can provide one of the teams at Nokia Canada with expertise and insight for any future implementation.

This functional specification document will provide an overview of our system in terms of the necessary functional requirements for its design and implementation. Future work related to this project will use this document as a guide for the actual deployed product.

Ztitch Solutions consists of three motivated fifth-year engineering students: Andrew Au (computer engineering), George Liao (system engineering), and Ching-Hsin Chen (computer engineering). By harvesting our skills and knowledge gained through our undergraduate careers and from industrial co-op experiences, we will be tackling this problem of mobile security and bringing life to this project. If you have any question or concern about our functional specification document, feel free to contact me by phone (778-322-7928) or by e-mail (aau1@sfu.ca).

Sincerely,

Andrew Au

Fifth-year Computer Engineering Student

Enclosure: *Functional Specification for a Facial and Speech Identification System for Nokia Devices*

Functional Specification for a

Facial and Speech Identification System (FASIS) for Nokia Devices

Project Team: Andrew Au
George Liao
Ching-Hsin Chen

Contact Person: Andrew Au
aau1@sfu.ca

Submitted to: Dr. Andrew Rawicz – ENSC 440
Michael Sjoerdsma – ENSC 305
School of Engineering Science
Simon Fraser University

Issued date: October 14th, 2010

Revision: 3.4

Executive Summary

The number of mobile subscribers continues to rise rapidly around the world, but so has the number of stolen or missing mobile devices. The technological capabilities of cell phones are improving, and more personal data are being stored in a cell phone than ever before. For example, mobile users are constantly using their mobile phones to access their internet banking accounts. Clearly, today's mobile devices offer much more than the ability to make a simple telephone call, but at the same time, there has been no dramatic enhancement in mobile security. A stolen or missing mobile device can lead to dire consequences, including identity theft and unauthorized access into personal accounts because personal information, account information/passwords are cached in the phone.

The proposed system will incorporate facial recognition as well as speech recognition to grant or deny access to certain secured features of the phone. Speech recognition was initially not in the proposal, but it shall now be implemented in response to the need of added security (i.e. against intruders using a printed picture of the user). Therefore, if the user desires to log-in to the internet banking account via the cell phone, then a facial match and a speech match must be requested. It utilizes the front facing camera of the Nokia smart phone to capture a picture. The picture is then sent to the network provider who will do the processing, and acts as the middleman/gateway of the service. If access is granted, then the network provider will return a key back to the device stating that the login may proceed; else, the user will continue to be locked out of the service.

Development of FASIS can be divided into four phases that are independent from each other:

1) Development of the microcontroller and transceiver wired to the phone such that it will allow it to be controlled (remotely) by the PC

2) Development of the speech recognition software to recognize a user's spoken password. This stage should be fairly simple as there are many open sources, such as Microsoft's Speech SDK.

3) Development of the facial localization and facial recognition algorithm using C# and MATLAB, respectively.

4) Development of the FASIS application on the phone

The four-month development cycle of this prototype phase is targeted for completion on December 15th, 2010.

Ztitch Solutions

Table of Contents

List of Figures

Glossary

2G Second-Generation Wireless
3G Third-Generation Wireless
HSPA High Speed Packet Access
HTI Harmonized Test Interface
ISO International Organization for Standardization
KB Kilobyte (1024 bytes)
FASIS Facial and Speech Identification System
OEM Original Equipment Manufacturer
PC Personal Computer
RS232 Recommended Serial 232
SDK Software Development Kit
UI User Interface
USB Universal Serial Bus
WEEE Waste Electrical and Electronic Equipment
Wi-Fi Wireless Fidelity

1. Introduction

The Facial and Speech Identification System (FASIS) enables added security to the modern smart phone by recognizing users with matching facial features and voice-based keyword input to access certain security sensitive features of the phone, while locking unrecognized users out. This system comes in response to the growing need of enhanced security for mobile phone users as more personal and confidential information are stored within the device. The requirements for the FASIS are described in this functional specifications document.

1.1. Scope

This document outlines the mandatory functional requirements of the FASIS. The set of requirements fully describes the proof-of-concept prototype and partially describes the production model.

1.2. Intended Audience

This document is intended for use by all members of the Ztitch Solutions team responsible for developing the FASIS. Each member shall refer to this document to ensure progress during development, as well as satisfactory performance and quality of the system during testing. Members shall refer to this document in order to keep product development in the right path toward a common goal. It can also provide an overview of the system for the OEM.

1.3. Classification

Throughout this document, the following convention shall be used to denote functional requirements:

[Rn-p] A functional requirement.

where **n** is the functional requirement number, and **p** is the priority of the functional requirement as denoted by one of three values:

I	The requirement applies to the proof-of-concept system only.
II	The requirement applies to both the proof-of-concept system and the final production system.
III	The requirement applies to the final production system only.

2. System Requirements

The general overview of the FASIS and its functional requirements are provided in this section of the document.

2.1. System Overview

Development of the system can be separated into four stages. The first stage is development of the link way between the device and the PC (server). Since broadcasting a real HSPA signal would not be feasible in the scope of this project, we will be using Bluetooth, USB, or simply a RS232 serial cable to simulate the over-the-air (OTA) connection.

The second stage is the implementation of speech recognition. In the proof-of-concept system, FASIS will attempt to verify that the user speaks the correct password to the microphone (the owner of the device will initially configure this voice-based password). For speech recognition, there are many open source libraries available; including Microsoft's established Speech Recognition SDK5.1 as found in [3], and in the proof-of-concept FASIS, we will not try to uniquely identify the user based on voice patterns, but rather, we will try to verify that the spoken keyword is correct only (since speech pattern recognition is a very in-depth subject on its own and none of the members in the team have expertise in this particular area).

The third stage is the development of the facial detection algorithm on the PC which simulates the service provided by the network server, which acts as the middleman that grants or denies the user access to the service. The PC is constantly on standby for receiving a request, which in this case, is a small bitmap stream. A unique key is returned if access is granted, or an instruction to lock the user out will be returned if access is denied.

The fourth stage is the development of the application on the phone which sends the bitmap stream and receives the key. Since the amount of data being transmitted goes up with cost, we want to minimize the amount of data being transmitted. The bitmap image will be encoded and transformed to grayscale (total size would be 2 to 4-KB in size) and the returned key is 128-bit in size.

The FASIS can be separated into smaller components: device/camera module, the facial recognition module, and the relay of information between the server and the client. Figure 1 provides a high-level overview of the system illustrating its proof-of-concept flowchart and how the components interact with each other.

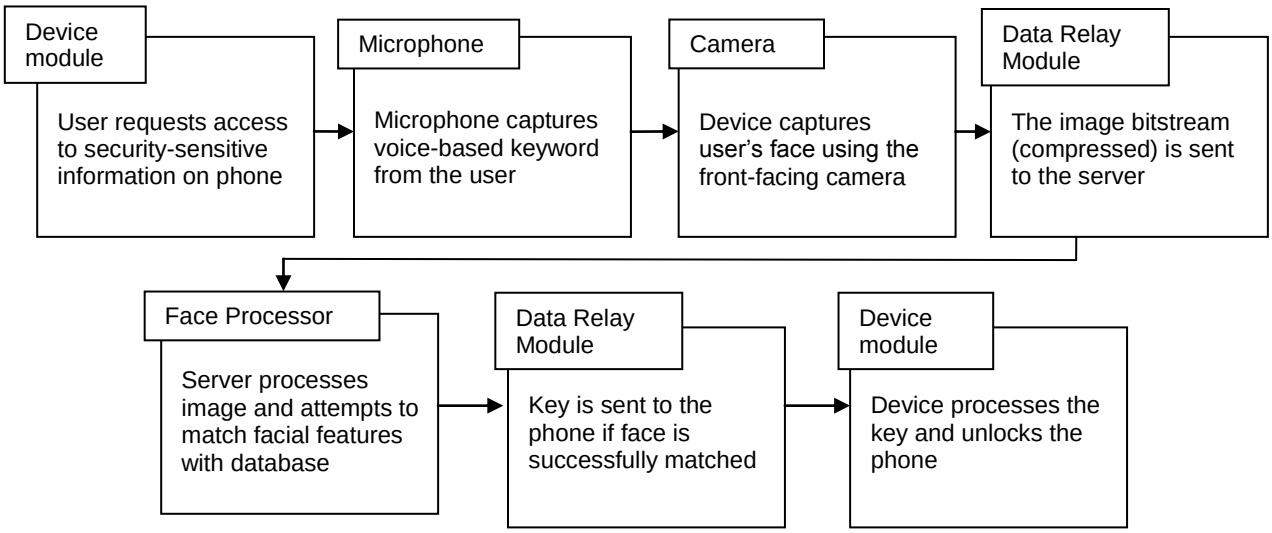

Figure 1. *High-level proof-of-concept system flowchart.*

When the user requests to access a secured item on the phone, such as logging into the internet banking account, the phone should automatically run FASIS. The FASIS application should launch, and when the user is ready, the user can press a button to take the snap shot of his or her face. The application will return an image bitstream, which should undergo image processing (compression and changing to grayscale) in order to minimize the amount of data being transmitted over the network and to the server. The data relay module is the network (i.e. 2G/3G/HSPA) protocol itself, but in the proof-of-concept system, the data relay module is a RS232/USB/Bluetooth connection.

At the core of FASIS is the image processing system. For maximum security and to reduce the amount of processing on the phone, the image processing step is taken care of by the server. This image process will find facial features of the taken picture, and compare them to the face of the authenticate owner(s) in the database. The method of facial detection is based from the EigenFace method for facial identification.

Once the server is able to successfully match the user's face with the database, a standard 128-bit unique key is sent back to the client to unlock it. If the server is unable to match the user's face with the database, an instruction to deny access will be relayed back to the device. Since facial identification is not 100% reliable, the device will not be completely locked out, and there should be a total of three retries to access the secured item. Under all circumstances, normal functions of the phone will be retained (including the ability to dial and receive calls) – only the secured items will be locked out from the user.

Due to limitation of time and resources, the proof-of-concept system (prototype) will exclude several features of the production system. However, the overall operation remains as described above.

2.2. General Requirements

[R1-III] Whenever the user attempts to access secured data, FASIS shall be triggered.

[R2-II] FASIS shall leverage speech recognition to identify a specific voice-based keyword input from the user, before proceeding to the facial recognition stage.

[R3-II] FASIS shall request a picture of the user's face using the front camera, and not the back camera.

[R4-III] User shall be able to install FASIS via software update only, without any need for hardware or physical modifications.

[R5-II] The matching algorithm shall be executed server side.

[R6-II] Under no circumstance shall the basic functionality of the phone be inhibited, even if authorization fails.

[R7-III] The user shall be able to cancel the FASIS process at any time to make emergency calls.

[R8-II] FASIS shall work as a passive system until it is triggered by user actions. As a passive system, it must be running seamlessly in the background and shall not affect any other operations of the phone.

[R9-I] The prototype system shall be priced under $800.

2.3. Physical Requirements

[R10-II] The mobile device must be pre-equipped with a front-facing camera from the manufacturer in order to install FASIS.

[R11-II] The mobile device must be pre-equipped with a microphone that is suitable to capture audio clearly for speech recognition.

[R12-I] The prototype system shall communicate to the server via USB and/or RS232 serial connections. An external communication board will be required and integrated with the device.

[R13-I] The prototype system shall be wired to a microcontroller in order for it to be controllable by the PC.

[R14-III] The commercialized system shall communicate to the server via cellular network protocols.

2.4. Facial Identification System (FASIS) Software Requirements

[R15-II] The captured image of the user shall be encoded and compressed before being sent to the server.

[R16-III] During authorization, the device shall wait for a response from the server. The response will be returned as a unique key which unlocks the secured-data, and the key must be hidden from view of the user.

[R17-III] In the prototype system, FASIS shall match only the voice-based keyword.

[R18-II] Multiple voice-based keywords can be set by the owner. However, security risk is increased as more keywords are set.

[R19-III] In the commercialized system, FASIS shall be able to uniquely recognize the voice of the owner.

[R20-I] For the prototype system, the Eigenface facial recognition algorithm shall be used.

[R21-III] For the commercialized system, a more accurate but more complex algorithm, such as the 'neuronal motivated dynamic link matching' method shall be used.

[R22-II] When FASIS authorization fails, a bitmap image of the infiltrator shall be forwarded to the owner's e-mail address.

[R23-II] Multiple users may be added to the FASIS database for matching, so that there can be more than one owner for any given device.

[R24-III] In the event that there are multiple faces within a single captured image, only the face closest to the screen (which has the largest relative proportion) shall be used for matching.

[R25-II] The speech recognition algorithm shall be based on the established Microsoft Speech SDK 5.1 [3].

[R26-III] FASIS shall be triggered automatically when the user attempts to browse secured data, such as accessing a HTTP Secure (HTTPS) location, as shown in figure 2.

Figure 2. *Screenshot of a HTTPS login page (PayPal). Here, FASIS should request facial and speech recognition from the user. A successful match will allow the user to bypass the manual password punching.*

2.5. Normal Operating Conditions

[R27-II] Sufficient lighting shall be expected in order for FASIS to work accurately.

[R28-II] The lighting condition between the database and the current picture capture shall not vary too greatly.

2.6. Upgrade and Compatibility

[R29-II] Facial recognition is still an active research topic, and newer/better methods may be uncovered in the future, so the facial recognition algorithm shall be upgradeable server-side at anytime without requiring the user to download anything on the client-side.

[R30-III] New firmware or software updates from the OEM (Nokia) shall be tested against FASIS to ensure that the system is not broken.

2.7. Performance Requirements

[R31-II] The facial matching algorithm shall take less than 5 seconds to perform.

[R32-III] It shall be assumed that there exist uncontrollable variations in network performance (due to varying bandwidth and traffic) which can affect the total overall performance of FASIS.

[R33-III] The camera shall be deactivated immediately after authorization to conserve battery.

[R34-II] FASIS shall take less than 2 seconds to launch when there are no other applications running.

[R35-II] The encoded and compressed image shall be less than 10KB total in size.

[R36-III] FASIS shall not have detrimental effects on battery life.

[R37-I] External power sources shall be required for the external hardware attached to the phone.

[R38-II] The localization of the face shall be calculated intelligently and is not confused with other objects in the background.

2.8. Usability Requirements

[R39-II] During the facial image capture stage, the device shall, in real-time, display the live camera feed of the user to facilitate in obtaining a suitable image.

[R40-II] The user interface (UI) shall be written in English.

[R41-II] A clear message along with a distinguishable icon must be displayed after FASIS authentication is complete to clearly indicate whether the operation succeeded or failed.

[R42-III] The algorithm must be able to detect whether the captured image is clear enough for processing. If not, it shall trigger a new image capture request.

[R43-II] An authenticated user shall have the option to remove FASIS from the phone at any time.

[R44-III] Database for registered users must be updatable only by an authenticated user.

[R45-III] Database management UI shall be user-friendly and intuitive.

[R46-II] Intuitive icons or images shall accommodate important messages to aid international users whose primary language is not English.

2.9. Reliability Requirements

[R47-III] FASIS matching shall have a success rate of at least 95% under reasonable variations in lighting conditions.

[R48-III] FASIS matching shall have a success rate of at least 95% for all types of skin color in the world.

[R49-III] The database shall be backed up to multiple servers in different locations.

[R50-III] Installation of FASIS shall not void the original warranty of the phone.

[R51-III] Main server running image processing shall be operational at all times except during malfunctions.

2.10. User Documentation Requirements

[R52-II] All user documentation shall be written in English.

[R53-III] User documentation shall include a website with general and technical information of the system, a user's manual for end-point users.

[R54-III] All user documentation shall be written for an audience assumed to have minimal technical knowledge of the system.

[R55-III] If commercialized internationally, the language of the user documentation shall be localized.

2.11. Standards and Regulations Requirements

[R56-I] The hardware in the proof-of-concept system shall conform to the Waste Electrical and Electronic Equipment (WEEE) directive [1].

[R57-II] FASIS shall not undermine the privacy of the user or make use of any personal data.

[R58-III] FASIS shall conform to the ISO 9000 standard [2] for software quality assurance.

3. System Test Plan

To ensure strong confidence and quality assurance of the system, comprehensive testing will be performed at all stages of development. Tests will be performed on the unit components, but a set of test cases (simulations) will also be developed to cover a wide range of use-case scenarios in the final overall system. These tests are intended to

discover bugs so that they can be alleviated as early as possible. In the proof-of-concept system, some testing of the hardware will also be required to ensure that project demo will go flawlessly. However, the commercialized system will contain no such hardware modification, since data is transferred over the network rather than an external communications board, so the hardware test stage will not be as comprehensive as the software test stage.

3.1. Automated Tests

Testing will be automated whenever possible. The software portion of the project will be updated frequently, and these software components include the algorithm for facial matching, the logic for locking/unlocking the phone, and the transfer of data between the phone and the server/PC. Since they are updated frequently, we want to save time by automating these tests rather than have a member of the group manually test them each time they are updated.

The objective here is to test each component with a variety of input arguments to validate that the returned results are correct. Automated tests can be executed using batch/bash files or with free software test tools such as AutoIt. These automated tests can be prescribed to run a specific number of loops for stress testing.

3.2. Unit Test

The proof-of-concept FASIS system can be divided into three sub-units: an application on the phone, a hardware communication board to relay information between the phone and the PC, and the server/PC providing the services.

The phone software can be tested by triggering FASIS by attempts to access secured data. The expected response is the camera application opening up which also powers up the front-facing camera. A button is then pressed (either the camera capture button or the center push button). The image then awaits transfer to the PC. Fortunately, these two steps can be automated using a USB connection controlled by Nokia's HTI (Harmonized Test Interface). The test will be running in loops that try to access different secured data of the phones, and each time, FASIS is expected to fire up. Then, in the second step, the automation should trigger the camera capture by simulating the push of a button, and the expected response is an image now waiting to be transferred.

Testing of the hardware communication board can be done by sending data from the phone to the PC, and vice-versa. We then verify that the data is sent successfully

without loss. FASIS will only relay data in two forms: an encoded bitmap image, and an encrypted key. This sub-unit of the system will be tested manually only, because this component will not appear in the final commercialized product, since data will actually be relayed over the network rather than an external board.

The facial matching algorithm can be tested independently from the other sub-units, with the only required input being a small image containing the face. The expected response is an encrypted key which contains the instruction for the phone to either lock or unlock the device. Testing of this sub-unit can also be automated, which is important. The facial matching algorithm will most likely undergo many revisions, and if we automate this step, it can save a lot of time by running the automated test after each revision compared to manually testing the algorithm. During each cycle of the test, one image will be chosen either randomly or from a consecutive list from a set of many different images. The database of test images will be derived from many different faces that exhibit a wide range of variations of facial features and skin color; this can ensure accuracy for all types of faces.

3.3 Integrated Test and Simulations

After each component has satisfied its requirement, an integrated test will be performed to ensure final product quality and that the performance meets the requirements. Simulations will cover actual use-case scenario: each member of the team will act as the owner of the device, and have his/her face data stored in the database. We will ask people from around the campus to be part of the tests, and these unrecognized users should not be authenticated. Furthermore, we will try to vary the lighting conditions by switching to different environments during these tests. The success rate should be at least 95%. Of course, the lighting conditions are assumed to be reasonable, both indoors and outdoors. However, areas of complete darkness or extreme brightness will probably not occur in practical situations, and would most likely cause the facial matching algorithm to fail, so we will not test in these areas.

If a test fails during the integrated stage, we will fix the sub-unit causing the failure and repeat unit testing before repeating the integrated tests.

4. Conclusion

This functional specification document clearly outlines capabilities, functions, requirements and standards for the FASIS. By using the specifications as guideline, we will provide more secure and convenient methods for mobile users when access secured data on the phone. The proof-of-concept system is well in progress of development and it is expected to be completed and meet all functional requirements defined by the target date of December 15th, 2010. In addition, the final product will be completed in one year from the prototype production date, should the system is to be commercialized.

5. References

[1] Agile Business Consultants, EU RoHS Directive Compliance Solutions. 2008. http://www.pb-free.info/index.htm

[2] International Standards for Organizations, ISO 9000 Standards. 2008. http://www.iso.org/iso/iso_catalogue/management_standards.htm

[3] Microsoft Speech Technologies. 2010. Microsoft Speech Technologies for Developers. http://www.microsoft.com/speech/default.aspx [Accessed September 29, 2010]