Chapter Summaries

Chapter 1 – Introduction

The project seeks to explore the use of Voice User Interfaces (VUIs) to augment buying and selling in Ashesi University with the aim of building a Natural Language Processing (NLP) system uses voice features and commands to accumulate change amounts electronically.

It was realized that there was the inconvenience of getting change for buyers especially when the change amount is quite small, such as GHC 1.50, 20 pesewas etc. and this was the problem. Unproductive methods such as chits, books for recordkeeping, word of mouth and other forms of paper receipts as a means to tackle the problem, have proven to be unsatisfactory and tend to create auxiliary issues.

The project therefore proposes a more efficient approach to solving the problem by building a system that will allow sellers to give buyers electronic change through a voice interface and thus enhance the buying and selling of items on the Ashesi campus. Some of the project' s contributions include a Voice User Interface (VUI) for change collection, electronic change that can be integrated with other electronic currencies, electronic piggy bank for change and research on the problem of change collection.

Although the problem may not have been formally designed for academic research yet, others have presented research contributing to the knowledge-base of this project. These related works include:

1. Live Speaker Identification in Conversations – In this paper a Speaker Identification System with a microphone is demonstrated where it is able to identify the speaker recorded with high accuracy of 85%. The paper presents an online speaker identification system demonstrated by the live test that can be embedded in a large range of web applications.
2. Pattern Recognition in Speaker Verification – This paper generally explores the concept of Speaker Verification. It uses the concept of pattern recognition using phrases that carry speaker-dependent

information. Charts and other distributions where used to demonstrate the experiments of speaker verification on two real speakers and the results proved that pattern recognition for speaker verification has proven to be successful with better accuracy than other conventional methods such as the Adeline Procedure.

3. Real Time Speaker Recognition from Internet Radio – This project explores the use of speaker recognition with recordings from internet radio as input. The project presents the general overview and architecture of speaker recognition and its requirements when it has to be live. The paper uses a Gaussian Mixture Model algorithm for modelling of speech and MPEG 3-layer compression on mel frequency cepstral coefficients for feature extraction. They experiment the system live with real speakers already added to their database, experiment with background noises and uses of thresholding to avoid invalid speaker detection.
4. "Are you There Margaret? It' s me, Margaret" : Speech Recognition as a Mirror – This paper investigates the advantages of speech recognition as input devices over button-pushing modes for taking input and its pro as more reflective Human Computer Interaction tool. After experimenting prototypes, Flounders (the author) realized a few issues of which include necessity of error-correction and the interference of noisy environments which contribute vitally to the knowledgebase of the OkNsesa project.
5. Findings with the Design of a Command-based Speech Interface for a Voice Mail System – This paper explores the use of a speech interface system embedded in a voice mail system. Similarly, the aim was to point out that speech interfaces for control are much more efficient designs than touch-tone control. The system used the concepts of commands such as help, previous, replay, yes and no to allow users interact it. After three levels iteration based on the idea of more intuitive commands, the authors verified that the speech interface is much more efficient.
6. A Voice Controlled E-Commerce Application – This project is a typical example of how Voice Systems can be applied in commerce. The authors used a couple of Speech Recognition Systems (SRS) including IBM' s Watson speech-to-text to demonstrate purchasing of items from an

ecommerce platform. The architecture of their project was the integration of the SRS with the e-commerce web application and exhibiting how voice commands can be used to select categories and products.

Chapter 2 – Requirement Analysis

Requirements specify how the project should be designed and built to meet user expectations and thus a very essential part of the project. Requirements for the OkNsesa Project were gathered through informal interviews and observation.

At the end of both activities of interactions with potential users (sellers and buyers) of the system, a list of requirements was generated grouped into functional (fundamental functions of the system) and non-functional (quality attributes of the system).

The functional requirements of the system, which could be further categorized into user and system requirements, include: buyers should be able to log onto system with voice features, update electronic account with voice commands, check past transactions and the system should have an electronic change account for each buyer, be able to recognize buyer' s voice and speech, do rigorous security verifications and a few other functionalities.

The Non-functional requirements of the system include Security where system must employ features to ensure that the information that is stored in the system is safeguarded from both internal and external attacks, Availability, where the system would be available to use 24 hours a day, seven days a week, Usability where the system shall be easy to use, and consistency with regards to user actions would be enforced and a few other qualities.

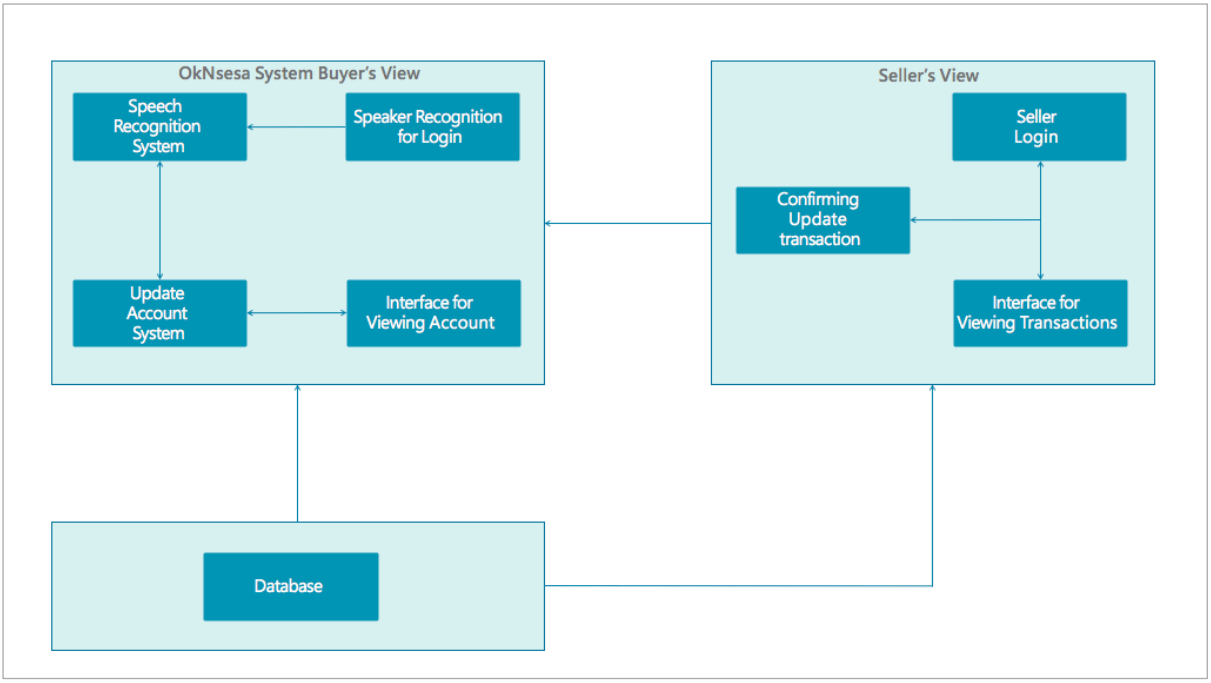
Chapter 3 – Architecture and Design

The system is organized into four main components: Speaker Recognition System, Speech Recognition System, Change Account System and an Interactive Interface. These four components constitute the process of a buyer being automatically recognized by voice his/her features, then updating his/her account with the said amount.

The first component is Speaker Recognition which is responsible for identifying which buyer spoke, so as to map on to the right buyer account. It is made of two consequent parts identification which is identifying the person speaking in an audio file, given a group of prospective speakers and verification where an input voice and phrase are compared against the enrollment's voice signature and phrase –in order to verify whether or not they are from the same person.

The Second Component is Speech Recognition which is responsible for understanding a what exactly a buyer said, to trigger an action, for instance update account or display transactions.

The third component is the Change Account System which is responsible for the managing of buyers' change account with regards to creating, updating and deleting accounts. The last component is an interactive interface where the user can communicate to track his/her transactions and account details. All of these components are designed to interact sequentially in a single architecture as shown below:



Chapter 4 – Implementation

The four fundamental components of the system (Speaker Recognition, Speech Recognition, Change Account System, Interactive Interface) were developed using different technologies and techniques of which are discussed in this chapter, emphasizing on how these components interact to achieve a full implementation of the system.

The Speaker recognition Component was built using Microsoft' s Cognition Services for Speaker Recognition. The two tasks of Identification and Verification that make up Speaker Recognition, are executed sequentially, where the identified ID from the Identification process is passed onto the Verification process to verify that the identified ID.

Speech Recognition comes in when the user has been identified and verified. This component was implemented by making calls to Google' s speech recognition API. The user' s speech is recognized, and the intent of the speech is extracted by repeatedly searching for specific patterns such as "add 20cedis" to create the actions to take on whether to add to the user' s account or remove from the account.

The Change Account System component uses the commands extracted to update the database of the particular user and it is implemented using SQLite. The interactive interface for allowing and viewing this management is implemented in flask, and both of these components work hand in hand to allow the system to successfully manage change amounts.

The technologies used in building the project include Python, PyAudio, Flask, SQLite, Google Speech Recognizer, Microsoft Speaker Recognizer and the choices for these technologies boil down to speed, simplicity, cost, large external community or forum etc.

Chapter 5 – Testing and Results

Testing is a vital aspect of any system development process as it confirms that the developed system meets the needs of its users. Various levels of testing were conducted for the OkNsesa project. These include Unit Testing, End-to-End Testing, Case Testing and Usability Testing.

For Unit Testing, the idea was to test individual units of the components of the system with the aim of reducing defects in newly added features. In undergoing these tests, unit test cases were developed, and they include testing enrollment for Identification, testing formatting of enrollment audio for identification, testing enrollment for verification, testing microphone recording with PyAudio, testing reading and writing of audio as wav files, testing inserting and updating database etc.

Case testing was not much different from unit testing only that it was used to test rare cases, error cases and since the system would interact with humans, certain typical and odd environments were designed as cases for test. Some of these cases include noise cancellation, sound error, no sound/speech, wrong identification and allowed or disallowed verification, wrong command extraction, testing at the Big Ben cafeteria etc.

End-to-end testing was conducted with five (5) users that wholly test the system from end to end six (6) times each with certain parameters such as no noise or with noise. The results for these repeated end-to-end tests are displayed in the table below, and the essentially show that the system works 10

out of 15 times with noise on an average and 3 out of 4 times without noise on an average.

The aim of the Usability test was to understand the interaction between the system and potential user. The five users after undertaking the end-to-end test were asked to answer some usability study including questions such as rate how easy it was to learn to use the system or easy it was to complete tasks. The general feedback presented that the system was quite easy to understand and use.