PAL: Personal Assistant System Using Low-Cost Computer

Melanie Louisa Khong Fui Yee and Sarah Samson Juan
Faculty of Computer Science and Information Technology, University Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, MALAYSIA.
louisa940321@gmail.com

***Abstract*—Automatic Speech Recognition (ASR) describes the ability of a computer to capture, identify and recognize the variety of human speech. It has been applied in many technologies such as personal assistant systems. Unfortunately, many personal assistant systems has been built in a way that may not always be disability-friendly and this causes the affected users whom are blind, disabled, illiterates and those who have physical limitations unable to enjoy the benefits of operating a computer. Hence, PAL is introduced. PAL is a personal assistant system built using low-cost device called the Raspberry Pi and open source voice-controlled software called Jasper. The functionalities of PAL includes searching for information on the internet, check for unread emails, schedule events in the calendar, manage a to-do list and translate texts through voice commands. Apart from that, a friendly graphical user interface (GUI) is also designed to display the output of each of the functional modules. Lastly, a number of tests are conducted to evaluate the performance and accuracy of the functional modules, GUI output display as well as response rate of the system. These tests include GUI output display test, user acceptance testing, the Command Success Rate (CSR) and Word Error Rate (WER) tests as well as response rate test. With the development of this project, it is hoped that PAL will be able to provide users with the benefits of using a computer in a more convenient and cost efficient manner.**

***Index Terms*—Automatic Speech Recognition; Artificial Intelligence; Raspberry Pi; Jasper.**

I. INTRODUCTION

Speech recognition has been applied in many voice command devices throughout the years such as personal assistant systems. Personal assistant systems has been tremendously useful and convenient in many ways in our daily life especially to users who are blind, disabled, illiterate and have physical limitations such as visual impairments. Hence, the proposed personal assistant system, PAL will act as a personal assistant that enables the affected users to search for information on the internet, check for unread emails, schedule events on the calendar, manage a task list and translate texts to a different language through voice commands. The objectives of PAL are to develop the system with lightweight device, the Raspberry Pi [1] and open-source speech recognition engine called Jasper, to design a graphical user interface (GUI) that displays the output results for each of the modules as well as to evaluate the performance and accuracy of the customized modules. The Raspberry Pi is a low-cost, credit card-sized single-board computers developed with a purpose to promote the teaching of basic computer science in schools [2]. On the other hand, Jasper is an open source voice-control software introduced by Shubhro Saha and Charlie Marsh which is basically a Siri clone running on the Raspberry Pi [3].

The following section will be covering the topic on the concept of ASR and background of personal assistant systems. After that, Section III discusses on the installation and configuration of the Raspberry Pi and Jasper, the development of the functional modules and GUI layout as well as testing and evaluations. All the results and analysis obtained through the tests and evaluations conducted on PAL are discussed in Section IV. Finally, Section V concludes our work.

II. BACKGROUND / MOTIVATION

A. Concept of Automatic Speech Recognition

Figure 1 shows the standard concept of Automatic Speech Recognition (ASR) system that explains how speech recognition works. The speech recognition system will first analyse the input speech signal before decoding it to find the best match between the input speech and its corresponding word string.

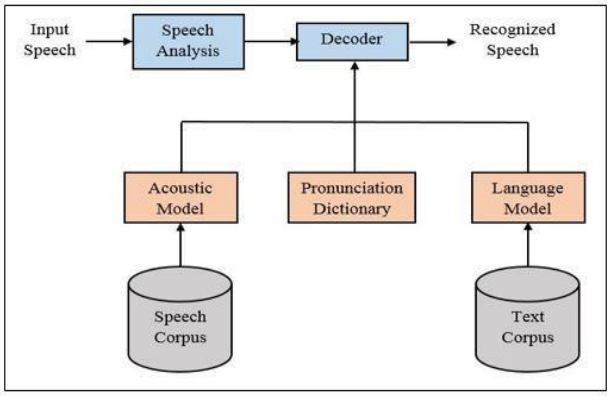

Figure 1: A standard concept of an ASR system

The three basic components, which are the acoustic model, language model and pronunciation dictionary, are responsible for finding the best match in the decoder. Acoustic models are statistical models used to estimate the probability that a certain phoneme has been uttered in a recorded input speech [4]. In other words, it converts sound to phoneme. The Hidden Markov Model (HMM) has been widely used to train acoustic models as it is an efficient algorithm for training and recognition [5]. After that, phonemes are converted into graphemes and to do so, a pronunciation dictionary is used to form valid words by combining various combinations of phonemes together. Several techniques have been used by researchers to create pronunciation dictionary with thousands of words such as rule based or statistical approach. For example, Juan and Besacier [6] built an Iban pronunciation dictionary for ASR through a semi-supervised Grapheme-to-Phoneme (G2P) bootstrapping strategy. Lastly, a language