Speech recognition for Indonesian language and its application to home automation

Zulkarnaen Hatala
Electrical Engineering Department
Politeknik Negeri Ambon
Ambon, Indonesia
dzulqarnaenhatala@gmail.com

***Abstract*— The practical aspects of developing an Automatic Speech Recognition System with Hidden Markov Toolkit are reviewed. Steps are explained concerning hardware, software, libraries, applications and computer programs used. The common procedure to rapidly apply the speech recognition system is summarized. The procedure is illustrated, to implement a speech-based electrical switch in home automation for the Indonesian language. The main key of the procedure is to match the environment for training and testing using the training data recorded from the testing program, HVite. Often the silence detector of HVite is wrongly triggered by noises because the microphone is too sensitive. This problem is mitigated by simply scaling down the volume. In this sub-word phone-based speech recognition, noise is included in the training database and labeled particularly. Illustration of the procedure is applied to a home automation application. Electrical switches are controlled by Indonesian speech recognizer. The results show a 100% command completion rate.**

Keywords—Indonesian speech recognizer; hidden Markov toolkit; home automation

I. INTRODUCTION

An automatic voice recognition system or automatic speech recognition system (ASR) aims to translate the digital signals of the human voice into text or written forms that are meaningful according to certain grammar. The human voice signal is captured by the microphone and processed by the computer to get a regular text. Thus ASR can be used for further purposes such as giving a command to a computer to do a group of tasks by just talking. For example, someone at his home instructs the computer to turn off and turn on the bathroom lights, garden lights without having to press the electrical switch, enough to give a voice command. Another case of applying ASR is when someone is driving a vehicle and wants to instruct his smartphone to notify the current position, or destination route information or even information about the on-off condition of all the electrical switches at his home.

Based on the number of words supported, ASR is categorized into small size vocabulary ASR (SVASR) and large vocabulary ASR (LVASR). SVASR only supports very few words compared to LVASR. An example of SVASR is turning on-off electrical switches at home. Another example would be digit recognition applications or instruction of a phonebook contact to dial automatic calls. In contrast, LVASR is used as a keyboard replacement for typing by naturally speaking naturally on search engines for topics and unlimited words. Even LVASR can be used as an automatic translator machine between two different languages such as between English to Indonesian and vice versa. Here the speaker with an LVASR machine (desktop or smartphone) does not need to understand the language of the interlocutor because it will be translated by the device [1].

But LVASR requires relatively high-quality hardware with a powerful processor and large memory. This makes LVASR difficult to apply directly to limited hardware environments such as embedded devices, smartphones, and home appliance peripherals. This is where the SVASR has the opportunity to be developed. A small number of words certainly do not require high hardware prerequisites. SVASR can also be developed quickly with high accuracy. Some of these things make the SVASR get a significant place to translate simple voice commands that are useful for everyday life. LVASR often is not practical when applied to some daily life cases.

Many reports on ASR research are based on a speech corpus that recorded on the soundproof noise-free environment that results in crystal clear speech databases[2][3]. The acoustic modeled achieved from these speech databases significantly suffer low performance when applied directly to the noisy environment[4]. Unfortunately, most applications are deployed in environments that are contaminated by noise and disturbance. To the best of our knowledge, no noisy ASR is applied for Indonesian especially for small vocabulary let alone implementation of ASR for home automation.

Hidden Markov Toolkit (HTK) [1] is a group of programs and libraries used to develop ASR with the Hidden Markov model. Actually, HTK itself is complete to build the whole ASR starting from speech database, feature extraction, construction and training of acoustic models and conducting tests offline and real-time online. To the best of our knowledge, no Indonesian ASR for home automation has been specifically implemented using HTK.

In the next section II summarizes the proposed steps for implementing noisy speech recognition for Indonesian. Section III adds a detail explanation with a real case illustration. Section IV discusses and summarizes the important result and the drawback of this research. The last section concludes the most important results of this writing.

II. PROPOSED PROCEDURES

A procedure is proposed to build a small vocabulary automatic speech recognition system (SVASR) using HTK. In general, these steps are to determine the grammar of the language to be supported, creating a speech training database,