VOICE TO TEXT SYSTEM FOR DISABLED (V2TS)

SIVATHIBAN A/L KRISHNAMURTTHU

A thesis submitted in fulfillment of the
requirement for the award of the degree of
Bachelor of Computer Science

Faculty of Computer Systems & Software Engineering
Universiti Malaysia Pahang

MAY 2010

ABSTRACT

Voice To Text System for Disabled (V2TS) is a speech recognition system that is aimed for Disabled people that can't type but have the speaking ability. V2TS deals with a number of useful functions where the disabled user can give command to the application for general operations such as reading text and closing the application itself. It has the Text To Speech (TTS) functionality where the text that has been input in the text area will be synthesized and read out by the computer. This report will discuss on the preparation, analysis, development and result that have been collected throughout the development cycle of this system. A complete reference and research details have been inserted in this document. This document will be a quick reference to refer on the specification and requirements of the system.

ABSTRAK

Sistem Suara ke Teks untuk Orang Kurang Upaya (V2TS) merupakan sebuah sistem yang disertakan dengan teknologi pengesanan pertuturan. Sistem ini adalah khas untuk Orang Kurang Upaya (OKU) yang tidak berkebolehan untuk menaip tetapi boleh bertutur. V2TS menyediakan pelbagai kemudahan dimana OKU boleh memberi arahan bersuara kepada aplikasi untuk operasi umum seperti membaca teks dan menutup aplikasi. Sistem ini juga mempunyai kebolehan untuk sintesis teks yang dimasukkan oleh pengguna dan membacanya. Laporan ini akan membincangkan perseediaan, analisis, pembangunan, dan juga keputusan yang telah dikumpul selama proses pembangunan system ini. Kesemua rujukan dan butiran mengenai sumber telah disertakan di dalam dokumen ini. Dokumen ini boleh digunakan sebagai rujukan untuk spesifikasi dan keperluan system.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter presents an outline of the entire project and the introduction into problem statements, objectives, scopes and thesis organization.

1.1 Introduction

Voice to Text System for disabled (V2TS) is a speech recognition embedded system. Speech recognition converts spoken words to machine-readable input. It simply means that this program enables the computer to generate correct output through our voice by utilizing some related algorithm. Speech recognition can be used for isolated words or continuous speech. Isolated word means that the system will only take one word at a time while continuous speech has continuous speech characteristic and the system need to recognize and convert the utterance at once. Speech Application Programming Interface (SAPI) will be a very important part of this system because it needs to recognize wide range of voice and tones. The output for any voice input depends on the matching between available Grammar and the utterance speed.

The term "voice recognition" is sometimes used to refer to speech recognition where the recognition system is trained to a particular speaker as is the case for most desktop recognition software, hence there is an aspect of speaker recognition, which attempts to identify the person speaking, to better recognize what is being said. Speech recognition is a broad term which means it can recognize almost anybody's speech. V2TS deals with a number of useful functions where the disabled user can give command to the application for general operations such as reading text and closing the application itself. It has the Text To Speech (TTS) functionality where the text that has been input in the text area will be synthesized and read out by the computer. Furthermore to facilitate multiuser as the Disabled might need help from secondary user, the Multipoint® Technology have been implemented where more than one mouse can be utilized in the system. For processing purpose, the text can be saved as text file or send to word processor for editing and other related works.

If a user has lost the use of his hands, or for visually impaired users when it is not possible or convenient to use a Braille keyboard, the systems allow personal expression through dictation as well as control of many computer tasks. Some programs save users'

speech data after every session, allowing people with progressive speech deterioration to continue to dictate to their computers. [1]

1.1.1 Research on Speech

1.1.1.1 Speech research in Microsoft Corporation

Microsoft Research has a group in Redmond and another in Beijing working together to improve spoken language technologies. Their main goal is to build applications that make computers available everywhere, and work with its Speech Products Group to make this vision a reality. The research are interested not only in creating state-of-the-art spoken language components, but also in how these disparate components can come together with other modes of human-computer interaction to form a unified, consistent computing environment. Microsoft is pursuing several projects to help reach its vision of a fully speech-enabled computer [2]

1.1.1.2 Speech research in Carnegie Mellon University (CMU)

The Sphinx Group at Carnegie Mellon University is committed to releasing the long-time, DARPA-funded Sphinx projects widely, in order to stimulate the creation of speech-using tools and applications, and to advance the state of the art both directly in speech recognition, as well as in related areas including dialog systems and speech synthesis.

The Sphinx Group has been supported for many years by funding from the Defense Advanced Research Projects Agency, and the recognition engines to be released are those that the group used for the various DARPA projects and their respective evaluations.

The packages that the CMU Sphinx Group is releasing are a set of reasonably mature, world-class speech components that provide a basic level of technology to anyone interested in creating speech-using applications without the once-prohibitive initial investment cost in research and development; the same components are open to peer review by all researchers in the field, and are used for linguistic research as well. [3]

1.2 Problem Statement

This system is aimed at Disabled people therefore the problems are Disabled people that can't write or type encounters difficulties in expressing their spoken words in text or documents form. Disabled people have difficulties to communicate well in this fast paced community where they are left behind in Information Technology. Most of them are not computer literate therefore they need an application that simplifies their task by just using speech command.

1.3 Objective

i. Develop a prototype of Speech Recognition application that reacts to speech command and output words.

ii. Design a simple and easy Speech Recognition application to enable Disabled people to get accessed to computer.

1.4 Scope

The scopes of the project are:

Project:

i. Recognize spoken words in English either command or grammar.

User:

i. Disabled people

ii. Normal user (Tutor/Assistant)

1.5 Thesis Organization

Chapter 1: Introduction

The purpose of this chapter is to introduce to the readers about the project that will be developed later. This chapter contains introduction, problem statement, objective, and scope and thesis organization.

Chapter 2: Literature review

This chapter explains about the reviews for the chosen project. This chapter is divided into two sub reviews that require students to study to get complete information about the project.

Chapter 3: Methodology

This chapter discusses the approach and framework for the project. Method, technique or approach that will be and will be used while designing and implementing the project will be included in the content. Justification and of method on approach used and hardware and software necessary is stated here.

Chapter 4: Implementation

This chapter acts to document all processes that involve in the development of the project. Designed project development is explained here. The content of this project depends on the system. It contains information of database and tools used. Data in database is shown in this chapter.

Chapter 5: Results and Discussion

The purpose of this system is to explain about the results and data analysis that had been acquired. Result analysis, project limitation and suggestion and project enhancement are contents for the chapter.

Chapter 6: Conclusion

This chapter explains briefly and summarizes the developed project.

CHAPTER 2

LITERATURE REVIEW

This chapter explains about the reviews for the chosen project. This chapter is divided into two sub reviews that require students to study to get complete information about the project.

2.1 Introduction to Speech Recognition

Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. The recognized words can be the final results, as for applications such as commands & control, data entry, and document preparation. They can also serve as the input to further linguistic processing in order to achieve speech understanding. Speech recognition systems can be characterized by many parameters. An isolated-word speech recognition system requires that the speaker pause briefly between words, whereas a continuous speech recognition system does not. Spontaneous, or extemporaneously generated, speech contains disfluencies, and is much more difficult to recognize than speech read from script. Some systems require speaker enrollment where a user must provide samples of his or her speech before using them, whereas other systems are said to be speaker-independent, in that no enrollment is necessary. Some of the other parameters depend on the specific task. Recognition is generally more difficult when vocabularies are large or have many similar-sounding words. When speech is produced in a sequence of words, language models or artificial grammars are used to restrict the combination of words. The simplest language model can be specified as a finite-state network, where the permissible words following each word are given explicitly. One popular measure of the difficulty of the task, combining the vocabulary size and the language model, is *perplexity*, loosely defined as the geometric mean of the number of words that can follow a word after the language model has been applied. Finally, there are some external parameters that can affect speech recognition system performance, including the characteristics of the environmental noise and the type and the placement of the microphone. [4]

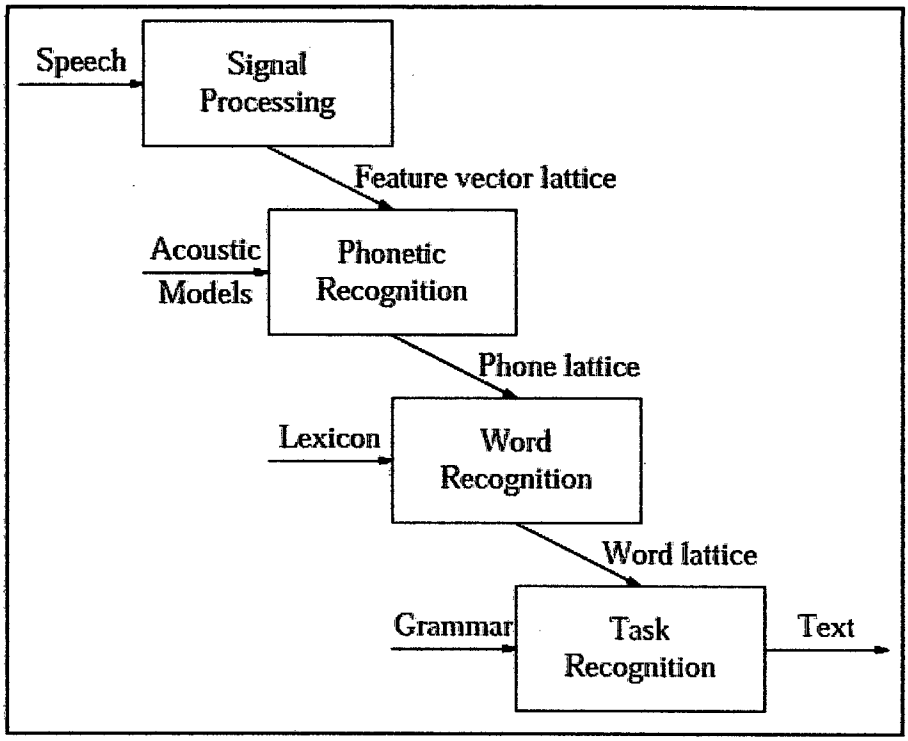

Figure 2.1: Block Diagram of Speech Recognition

Table 2.1: Parameters of speech recognition systems

Parameters	Range
Speaking Mode	Isolated words to continuous speech
Speaking Style	Read speech to spontaneous speech
Enrollment	Speaker dependent to Speaker independent
Vocabulary	Small (less than 20 words) to large (more than 20,000 words)
Language Model	Finite state to context sensitive
Perplexity	Small (<10) to large (>100)
SNR	High (>30dB) to low (<10dB)

2.2 Existing system

2.2.1 Thai Automatic Speech recognition

2.2.1.1 Introduction

This research was performed as part of the DARPA-Babylon program aimed at rapidly developing multilingual speech-to-speech translation capability in several languages. It is built on extensive background in ASR, language portability, and speech translation, the group has built Arabic-English and Thai-English Speech-to-Speech translation systems in less than 9 months per language. This system has been used in an external DARPA evaluation involving medical scenarios between an American Doctor and a naive monolingual Thai patient. [5]

2.2.1.2 Technique

Hidden Markov Model has been used in this system.

2.2.1.3 Objective

To develop a robust and flexible Thai Speech Recognizer that can be integrated to Thai-English speech translation.

2.2.1.4 Feature

I. Automatic pronunciation generation.
II. Rapid bootstrapping.
III. Phone set and pronunciation variation.
IV. Real Time recognizer for medical dialogs

2.2.2 HMM-Based Speech Synthesis System Applied to English

2.2.1.2 Introduction

Although many speech synthesis systems can synthesize high quality speech, they still cannot synthesize speech with various voice characteristics such as speaker individualities, speaking styles, emotions, etc. To obtain various voice characteristics in speech synthesis systems based on the selection and concatenation of acoustical units, a large amount of speech data is necessary. However, it is difficult to collect store such speech data. In order to construct speech synthesis systems which can generate various voice characteristics, the HMM-based speech synthesis system (HTS) was proposed. In the training part, spectrum and excitation parameters are extracted from speech database and modeled by context dependent HMMs. In the synthesis part, context dependent HMMs are concatenated according to the text to be synthesized. Then spectrum and excitation parameters are generated from the HMM by using a speech parameter generation algorithm. Finally, the excitation generation module and synthesis filter module synthesize speech waveform using the generated excitation and spectrum parameters. The attraction of this approach is in that voice characteristics of synthesized speech can easily be changed by transforming HMM parameters. In fact, it is shown that we can change voice characteristics of synthesized speech by applying a speaker adaptation technique, a speaker interpolation technique, or an Eigen voice technique. [6]

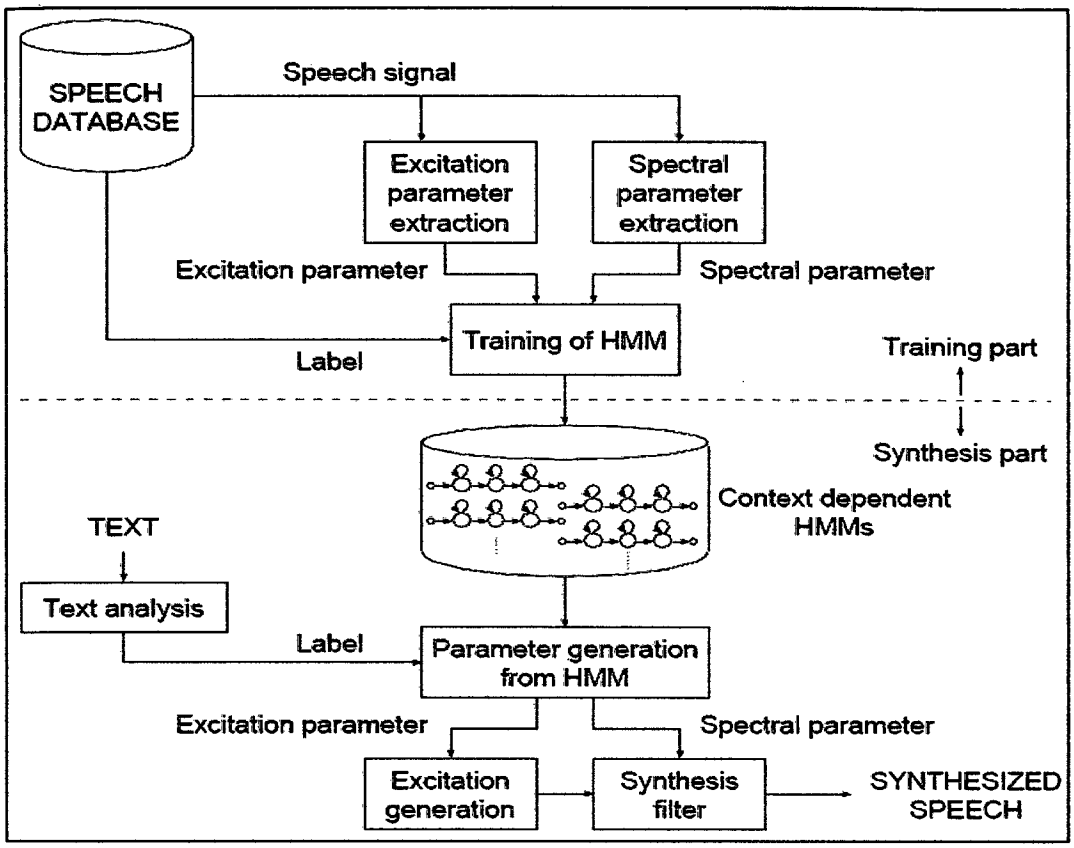

Figure 2.2: HMM-Based Speech Synthesis System

2.2.2.2 Technique

Hidden Markov Model has been used in this system.

2.2.2.3 Objective

To utilize HMM-based speech synthesis system (HTS) to English speech synthesis.

2.2.2.4 Feature

I. Spectrum Modeling.