IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Speech Recognition for Vietnamese: A Literature Review

Le Minh Tri[1], Do Dinh Thanh[2]

Department of Mathematics– Informatics, The People's Police University, Ho Chi Minh City, Vietnam[1]

Faculty of Information Technology, Ho Chi Minh City University of Foreign Languages and Information Technology, Ho Chi Minh City, Vietnam[2]

Abstract: This paper describes a literature review of automatic speech processing for Vietnamese. Our work also presents what research has been done around for dealing with the problem of Vietnamese speech processing. The main objective of this review paper is to provide the information of the progress made for speech processing of Vietnamese language. We intend to describe distinctive and novel features of selected systems and their relative merits and demerits.

Keywords: Vietnamese speech processing; voice applications for Vietnamese; Vietnamese speech recognition; speech synthesis; automatic speech recognition for Vietnamese.

I. INTRODUCTION

Speech recognition is a pattern recognition process where each pattern is a identification unit, it can be a word or phoneme. The fundamental challenge of this problem is the speech (voice) is always variability over time. And there is a big difference between the voices of different speakers, speed, context and different acoustic environments. The study of speech recognition based on three basic principles:

- Voice signals are represented accurately by a short-term amplitude spectrum. So we can extract voice features from short period time windows and use these features for speech recognition propose.
- The voices are expressed as letters which are series of phonetic symbols. Hence the significance of a pronunciation is conserved when we pronounce the phonetic sequence of acoustic symbols.
- Speech recognition is a cognitive process, all languages have meaning. So the semantics and pramatics are important in the recognition task.

Research field of speech recognition is quite broad involving many different sectors: digital signal processing, acoustic, computer science theory, linguistics, physiology… The speech recognition system can be divided into two different types: discrete and continuous word based speech recognition. In the speech recognition system, we again distinguish identification system small-sized dictionaries, medium-sized and large-sized dictionaries. In the field of applying speech processing techniques for Vietnamese, at now, we know some newest remarkable publications of some research groups such as Thien Khai Tran et al. [1,2,3,4,5], Luong Chi Mai et la. [6,7] as well as Quan Vu et al. [8,9,10,11]. In this paper, we make a literature review on Automatic Speech Recognition and Understanding for Vietnamese. The rest of the paper is organized as follows. In section II, the automatic speech recognition is described. In Section III, we present the voice applications for Vietnamese. The conclusion is in Section V.

II. AUTOMATIC SPEECH RECOGNITION

2.1 Building Blocks

Figure 1 shows the building blocks which perform transformations pertinent to speech recognition. In which, the input is natural speech and the output is the recognized text.

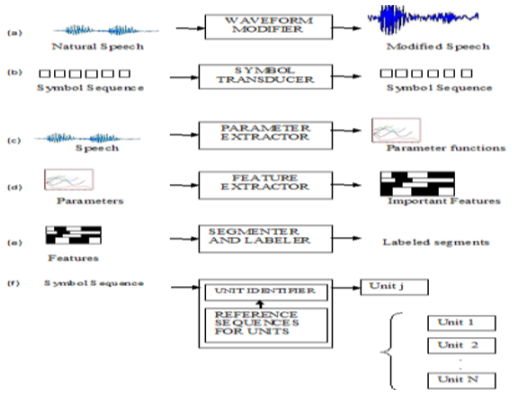

Fig.1. Basic building blocks for speech processing[12]

2.2 Speech recognition based on Hidden Markov Model

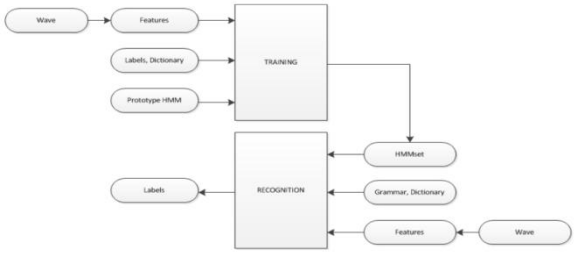

Fig.2. Steps to build the Automatic Speech Recognizer with HTK [3]

DOI 10.17148/IJARCCE.2016.54258

Hidden Markov Mode(HMM) [13]is a statistical model in which the system being modelled assumed to be a Markov process with unknown parameters, and the challenge is to determine the hidden parameters from an observation parameters. By applying HMM, we construct a statistical model on each phone that its states are assigned specific possibilities in comparison with reference value. The possibility of each state depends on itself and the previous one. The goal of speech recognition system is to find out the sequence of states that has the maximum probability.
Nguyen Tien Dung, Vu Tat Thang, Luong Chi Mai, [7] in the publication on 2004, the authors proposed a way to recognize spoken names of Vietnamese language. Since the number of names was big and their resources were limited, they decided to develop an affordable recognizer that had 3 characteristics: continuous speech, context dependent and large vocabulary. They used HMM model to recognize names. In their evaluation they achieved 80%at word accuracy rate (WAR).

Based on HMM, Quan Vu et al. [8] obtained the precision rate of over than 93% and this group successfully built many voice applications on this base. Nevertheless, these works have not been accompanied with a efficient semantic processing mechanism yet, which is the important mechanism helping the system with understanding command.

2.3 Speech recognition based on Dynamic Time Warping (DTW)

Dynamic time warping (DTW) is a time series alignment algorithm developed originally for speech recognition. DTW aims at comparing and aligning two sequences of feature vectors by warping the time axis repetitively until an optimal match between the two sequences is found. This warping between two time series can then be used to find corresponding regions between the two time series.
Minh-Son Nguyen and Tu-Lanh Vo [14] approached speech recognition for home automation in Vietnamese language by using the improvement MFCC and Dynamic Time Warping (DTW). They showed a combination of pattern matching approach to speech recognition for smart home is proposed and developed with the important improvement in MFCC extraction that increase the accuracy up to 20%.

III. VOICE APPLICATIONS

Quan Vu et al. [11]created a voice server for Stock Information Inquiry. This group also developed a Spoken Information Based Approach for the Retrieval of Soccer Video Events [10], Is ago: The Vietnamese Mobile Speech Assistant for Food-court and Restaurant Location [9] and A Robust Vietnamese Voice Server for Automated Directory Assistance Application [8]. The authors crucially concentrated on improving the efficiency of their voice recognition system which obtained the precision rate of over than 93% . Nevertheless, all the applications have not been accompanied with a efficient semantic processing mechanism yet, which is the important mechanism in view of helping the system to be more intelligent. Thien et al. adopted DCG for resolving syntactic parsing and semantic analysis of Vietnamese voice commands and questions in Senti Voice [1,2], Edu Voice [4] and Edu ICR systems [5]. They have provided evidence of the importance of syntax and semantics processing in voice applications, specially Vietnamese speech applications.

IV. CONCLUSION

In this paper, a review on Speech Recognition for Vietnamese is presented. With this research, we showed the importance of text language processing in voice applications. In next steps, our jobs to be accomplished have essential characters in executing an smart application with a combination of the spoken language recognition and the written language processing.

REFERENCES

[1]. Thien Khai Tran (2015), "SentiVoice - a system for querying hotel service reviews via phone", The 11th IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF 2015), Cantho, 2015: 65-70.
[2]. Thien Khai Tran, Tuoi Thi Phan:An upgrading SentiVoice - a system for querying hotel service reviews via phone. IALP 2015: 115-118.
[3]. Thien Khai Tran, Dang Tuan Nguyen (2013). "Semantic Processing Mechanism for Listening and Comprehension in VNS Calendar System". International Journal on Natural Language Computing (IJNLC) Vol. 2, No.2, April 2013.
[4]. Thien Khai Tran, Tien Cat Khai Tran, Tho Anh Mai, Nhat Minh H. Nguyen and Hien Thanh Vu (2014), "EDUVoice - a system for querying academic information via PSTN", The Third Asian Conference on Information Systems (ACIS 2014). Nha Trang, 2014.
[5]. Thien Khai Tran, Dung Minh Pham, Binh Van Huynh, "Towards Building an Intelligent Call Routing System, International Journal of Advanced Computer Science & Applications 1 (7), 458-462.
[6]. Thang Vu and Mai Luong, (2012). "The Development of Vietnamese Corpora Toward Speech Translation System". RIVF-VLSP 2012. Ho Chi Minh City, Viet Nam.
[7]. Nguyen Tien Dung, Vu Tat Thang, Luong Chi Mai, " Vietnamese spoken names recognition", National Informatics Conference, Da Nang, Vietnam, 2004
[8]. Duong Dau, Minh Le, Cuong Le and Quan Vu, (2012). "A Robust Vietnamese Voice Server for Automated Directory Assistance Application". RIVF-VLSP 2012. Ho Chi Minh City, Viet Nam.
[9]. Hue Nguyen, Truong Tran, Nhi Le, Nhut Pham and Quan Vu, (2012). "iSago: The Vietnamese Mobile Speech Assistant for Food-court and Restaurant Location". RIVF-VLSP 2012. Ho Chi Minh City, Viet Nam.
[10]. Quan VU, et al., "A Spoken Information Based Approach for the Retrieval of Soccer Video Events", Speech Technologies, Book 2, Intech Open Access Publisher, 2011.
[11]. Quan VU, et al., A Spoken Dialog System for Stock Information Inquiry, IT@EDU, 2010.
[12]. S.J.Arora and R.Singh, "Automatic Speech Recognition: A Review," International Journal of Computer Applications, Vol.60 - No.9, December 2012.
[13]. Steve Young et al., (2006). The HTK Book (version 3.4). [On-line]. Available: www.htk.eng.cam.ac.uk/docs/docs.shtml [Nov. 1, 2012].
[14]. Minh-Son Nguyen, Tu-Lanh Vo:Vietnamese Voice Recognition for Home Automation using MFCC and DTW Techniques. ACOMP 2015: 150-156.