Chapter 63

Soft–Computational Techniques and Spectro–Temporal Features for Telephonic Speech Recognition:
An Overview and Review of Current State of the Art

Mridusmita Sharma
Gauhati University, India

Kandarpa Kumar Sarma
Gauhati University, India

ABSTRACT

Speech is the natural communication means, however, it is not the typical input means afforded by computers. The interaction between humans and machines would have become easier, if speech were an alternative effective input means to the keyboard and mouse. With advancement in techniques for signal processing and model building and the empowerment of computing devices, significant progress has been made in speech recognition research, and various speech based applications have been developed. With rapid advancement of the speech recognition technology, telephone speech technology are getting more involved in many new applications of spoken language processing. From the literature it has been found that the spectro-temporal features gives a significant performance improvement for telephone speech recognition system in comparison to the robust feature techniques used for the recognition purpose. In this chapter, the authors have reported the use of various spectral and temporal features and the soft computing techniques that have been used for the telephonic speech recognition.

INTRODUCTION

Speech is the vocalized form of communication between one speaker and one or more listeners and is the most effective, reliable and common medium of sending messages in real time systems. Speech is the result of time varying vocal tract system excited by the time varying excitation source signal. According to Information Theory, speech can be represented in terms of its message content, or information. An alternative way of characterizing speech is in terms of signal carrying the message information that is the acoustic waveform. Speech is a natural phenomenon which is easy, fast to communicate and do not require any technical knowledge. The bandwidth of human speech communication is approximately the frequency range up to 7 kHz (O'Shaughnessy, 2000). From the very childhood a human being starts to learn the basic linguistic information without any strict instructions and hence develop a large vocabulary in the brain throughout their lives. The physiology of human speech production is however not an easy process. It requires the co-functioning of biological organs such as the lungs, larynx, pharynx, etc. The human vocal tracts and the articulators are biological organs with non-linear properties whose parameters are largely affected by the factors ranging from gender to upbringing to emotional state. As a result, vocalization is largely influenced by the accent, pronunciation and various other vocal tract parameters (Sarma & Sarma, 2014). With the advent of the speech processing technology the human-machine interaction has become much easier. It is much comfortable on the part of the human being to communicate directly with the machine than to use primitive interfaces such as keyboard, mouse or other pointing devices because of the fact that the primitive interfaces like keyboard and pointing devices require certain amount of skill for their effective usage. In order to use the computer efficiently, apart from a certain level of literacy, the user is also expected to have a sound proficiency in English and a proper typing skill. However, a physically challenged person finds it difficult to use the computer as well the interfacing devices. Further, non-English literate persons also find it difficult to use these interface devices. For a language with dialectal and ethnographic variations, the complexities grow further. These difficulties can be overcome with the speech based interfaces. The interfacing of the human being with the machine with a user friendly interface has always been an important technological issue. This is one of the most widely followed issues in human computer interaction (HCI). Recognition of speech with native content enables the common man to make use of the benefits of information technology and hence facilitates better HCI in a much easier way (Kurian, 2014). Now-a-days, telephone speech technology is getting more acceptance in many new applications of spoken language processing such as Voice Service Centre in hotels and restaurants, voice navigations in traffic and transportation systems, call center support in medical, banking, agriculture etc. sectors and many more. However, there are significant challenges that need solutions while designing system for real time telephone speech recognition with better accuracy. Greater reliability in real time telephonic speech recognition is another constraint often seen while dealing with the speech that comes through a telephone channel. Recording over the telephone lines introduces severe distortions due to the variations in the transmission channels (Zuo, Liu & Ruan, 2003). Speech recognition over the lines has also become an integral part of the various applications of Large Vocabulary Continuous Speech Recognition (LVCSR).

From the literature it has also been found that over the past decades, there has been considerable amount of works done in the field of English Conversational Telephone Speech Recognition (CTSR). Also, India being a country with vast linguistic variations, it provides a sound area of research toward language-specific speech recognition technology. However, not much work has been done in Assamese as well as other north-eastern languages of the country in telephonic speech recognition. This chapter

Related Content

Sociomateriality Implications of Multi-Agent Supported Collaborative Work Systems
Tagelsir Mohamed Gasmelseid (2012). *International Journal of Intelligent Information Technologies (pp. 1-16).*
www.irma-international.org/article/sociomateriality-implications-multi-agent-supported/69387

User Experience in Social Human-Robot Interaction
Beatrice Alenljung, Jessica Lindblom, Rebecca Andreassonand Tom Ziemke (2017). *International Journal of Ambient Computing and Intelligence (pp. 12-31).*
www.irma-international.org/article/user-experience-in-social-human-robot-interaction/179287

Blind Signal Detection Techniques for Spectrum Sensing in Satellite Communication: Blind Signal Detection Techniques for Satellite Communication
Bilal Muhammad Khanand Rabia Bilal (2017). *Handbook of Research on Recent Developments in Intelligent Communication Application (pp. 1-48).*
www.irma-international.org/chapter/blind-signal-detection-techniques-for-spectrum-sensing-in-satellite-communication/173238

Optimising Object Classification: Uncertain Reasoning-Based Analysis Using CaRBS Systematic Research Algorithms
Malcolm J. Beynon (2008). *Artificial Intelligence for Advanced Problem Solving Techniques (pp. 234-253).*
www.irma-international.org/chapter/optimising-object-classification/5325

Influence Maximization for MOOC Learners Using BAT Optimization Algorithm
Kirti Aggarwaland Anuja Arora (2022). *International Journal of Fuzzy System Applications (pp. 1-19).*
www.irma-international.org/article/influence-maximization-for-mooc-learners-using-bat-optimization-algorithm/296588