

Preface

Speech is a naturally occurring nonstationary signal essential not only for person-to-person communication but has become an important aspect of Human Computer Interaction (HCI). Some of the issues related to analysis and design of speech-based applications for HCI have received widespread attention. With continuous upgradation of processing techniques, treatment of speech signals and related analysis from varied angles has become a critical research domain. It is more so with cases where there are regional linguistic orientations with cultural and dialectal elements. It has enabled technologists to visualize innovative applications. This work is an attempt to treat speech recognition with soft computing tools oriented toward a language like Assamese spoken mostly in the northeastern part of India with rich linguistic and phonetic diversity. The regional and phonetic variety observed in Assamese makes it a sound area for research involving ever-changing approaches in speech and speaker recognition. The contents included in this compilation are outcomes of the research carried out over the last few years with emphasis on the design of a soft computing framework for phoneme segmentation used for speech recognition. Though the work uses Assamese as an application language, the concepts outlined, systems formulated, and the results reported are equally relevant to any other language. It makes the proposed framework a universal system suitable for application to soft computing-based speech segmentation algorithm design and implementation.

[Chapter 1](#) provides basic notions related to speech and its generation. This treatment is general in nature and is expected to provide the background necessary for such a work. The contents included in this chapter also provide the necessary motivation, certain aspects of phoneme segmentation, a review of the reported literature, application of Artificial Neural Network (ANN) as a speech processing tool, and certain related issues. This content should help the reader to have a rudimentary familiarization about the subsequent issues highlighted in the work.

Speech recognition research is interdisciplinary in nature, drawing upon work in fields as diverse as biology, computer science, electrical engineering, linguistics, mathematics, physics, and psychology. Some of the basic issues related to speech processing are summarized in [Chap. 2](#). The related theories on speech perception and spoken word recognition model have been covered in this chapter. As ANN is the most critical element of the book, certain essential features necessary for the subsequent portion of the work constitute [Chap. 3](#). The primary topologies covered

include the Multi Layer Perceptron (MLP), Recurrent Neural Network (RNN), Probabilistic Neural Network (PNN), Learning Vector Quantization (LVQ), and Self-Organizing Map (SOM). The descriptions included are designed in such a manner that it serves as a supporting material for the subsequent content.

Chapter 4 primarily discusses about Assamese language and its phonemical characteristics. Assamese is an Indo-Aryan language originated from the Vedic dialects and has strong links to Sanskrit, the ancient language of the Indian sub-continent. However, its vocabulary, phonology, and grammar have substantially been influenced by the original inhabitants of Assam such as the Bodos and Kacharis. Retaining certain features of its parent Indo-European family, it has many unique phonological characteristics. There is a host of phonological uniqueness in Assamese pronunciation which shows variations when spoken by people of different regions of the state. This makes Assamese speech unique and hence requires a study exclusively to directly develop a language-specific speech recognition/speaker identification system.

In **Chap. 5**, a brief overview derived out of a detailed survey of speech recognition works reported from different groups all over the globe in the last two decades is given. Robustness of speech recognition systems toward language variation is a recent trend of research in speech recognition technology. To develop a system that can communicate between human beings in any language like any other human being is the foremost requirement of any speech recognition system. The related efforts in this direction are summarized in this chapter.

Chapter 6 includes certain experimental work carried out. The chapter provides a description of a SOM-based segmentation technique and explains how it can be used to segment the initial phoneme from some Consonant–Vowel–Consonant (CVC) type Assamese word. The work provides a comparison of the proposed SOM-based technique with the conventional Discrete Wavelet Transform (DWT)-based speech segmentation technique. The contents include a description of an ANN approach to speech segmentation by extracting the weight vectors obtained from SOM trained with the LP coefficients of digitized samples of speech to be segmented. The results obtained are better than those reported earlier.

Chapter 7 provides a description of the proposed spoken word recognition model, where a set of word candidates are activated at first on the basis of phoneme family to which its initial phoneme belongs. The phonemical structure of every natural language provides some phonemical groups for both vowel and consonant phonemes each having distinctive features. This work provides an approach to CVC-type Assamese spoken words recognition by taking advantages of such phonemical groups of Assamese language, where all words of the recognition vocabulary are initially classified into six distinct phoneme families and then the constituent vowel and consonant phonemes are classified within the group. A hybrid framework, using four different ANN structures, is constituted for this word recognition model, to recognize phoneme family and phonemes and thus the word at various levels of the algorithm.

A technique to remove the CVC-type word limitation observed in case of spoken word recognition model described in **Chap. 7** is proposed in **Chap. 8**.

This technique is based on a phoneme count determination block based on K-means Clustering (KMC) of speech data. A KMC algorithm-based technique provides prior knowledge about the possible number of phonemes in a word. The KMC-based approach enables proper counting of phonemes which improves the system to include words with multiple number of phonemes.

[Chapter 9](#) presents a neural model for speaker identification using speaker-specific information extracted from vowel sounds. The vowel sound is segmented out from words spoken by the speaker to be identified. Vowel sounds occur in a speech more frequently and with higher energy. Therefore, situations where acoustic information is noise corrupted vowel sounds can be used to extract different amounts of speaker discriminative information. The model explained here uses a neural framework formed with PNN and LVQ where the proposed SOM-based vowel segmentation technique is used. The speaker-specific glottal source information is initially extracted using LP residual. Later, Empirical Mode Decomposition (EMD) of the speech signal is performed to extract the residual. The work shows a comparison of effectiveness between these two residual features.

The key features of the work have been summarized in [Chap. 10](#). It also includes certain future directions that can be considered as part of follow-up research to make the proposed system a fail-proof framework.

The authors are thankful to the acquisition, editorial, and production team of the publishers. The authors are thankful to students, research scholars, faculty members of Gauhati University, and IIT Guwahati for being connected in respective ways to the work. The authors are also thankful to their respective family members for their support and encouragement.

Finally, the authors are thankful to the Almighty.

Guwahati, Assam, India
January 2014

Mousmita Sarma
Kandarpa Kumar Sarma

Phoneme-Based Speech Segmentation using Hybrid
Soft Computing Framework

Sarma, M.; Sarma, K.K.

2014, XXI, 187 p. 62 illus. in color., Hardcover

ISBN: 978-81-322-1861-6