

Chapter 42

Bangla User Adaptive Word Speech Recognition: Approaches and Comparisons

Adnan Firoze

North South University, Bangladesh

Md Shamsul Arifin

North South University, Bangladesh

Rashedur M. Rahman

North South University, Bangladesh

ABSTRACT

The paper presents Bangla word speech recognition using two novel approaches with a comprehensive analysis. The first approach is based on spectral analysis and fuzzy logic and the second one uses Mel-Frequency Cepstral Coefficients (MFCC) analysis and feed-forward back-propagation neural networks. As human speech is imprecise and ambiguous, fuzzy logic – the base of which is indeed linguistic ambiguity, could serve as a precise tool for analyzing and recognizing human speech. The authors' systems revolve around the visual representations of voiced signals – the Fourier energy spectrum and the MFCC. The essences of a Fourier energy spectrum and the MFCC are matrices that include information about properties of a sound by storing energy and frequency in discrete time. The decision making process of their systems is based on fuzzy logic and neural networks. Experimental results demonstrate that their fuzzy logic based system is 86% accurate whereas the Artificial Neural Networks (ANN) based system is 90% accurate compared to a commercial Hidden Markov Model (HMM) based speech recognizer that shows 73% accuracy on an average. Moreover, the authors' research derives that, even though ANN gives a better recognition accuracy than the fuzzy logic based system, the fuzzy logic based system is more accurate when it comes to “more difficult” or “polysyllabic” words. In terms of runtime performance, the fuzzy logic based system outperforms the ANN based Bangla speech recognition system.

DOI: 10.4018/978-1-4666-7456-1.ch042

1. INTRODUCTION

Human speech recognition has a broad elucidation that refers to the technology that can recognize speech. The recognition process is still open because none of the current methods is fast and precise enough compared to human recognition abilities. Research in this area has attracted a great deal of attention over the past five decades. Several technologies have been applied and efforts were made to increase the performance up to marketplace standard so that the users would have the benefit in a number of ways.

During this long research period, several key technologies were applied to recognize isolated words such as Hidden Markov Models (HMM) (Hasnat, Jabir, & Mumit, 2007), Artificial Neural Networks (ANN), Support Vector Classifiers with HMM, Independent Component Analysis, HMM and Neural-Network Hybrid, the stochastic language model and more (Juang & Rabiner, 2005).

In recognizing Bangla speech, most of the research efforts were developed by HMM and ANN techniques but no research work has been reported till the date that uses Fuzzy logic and develops a Fuzzy Inference System (FIS), MEL (a shortened form of the word ‘melody’ named by mathematicians – Stevens, Volkman and Newman) filtering and Short-time Fourier transform (STFT) methods. Therefore, in this research we investigate, propose and implement two distinct models that can recognize Bangla isolated words. The first approach is by using spectral analysis and fuzzy logic and the second one is by using cepstral analysis and feed-forward back-propagation artificial neural networks.

The ambiguity in phonemes in Bangla/Bengali speech is more intense and varied than that of English speech since Bangla stems from the ‘Indo-European language family’ just as Hindi, Urdu, Persian and numerous languages from South Asia having native speakers of over 3 billion (Weiss, 2006). Therefore, the approach for speech recognition considers the ‘word level’ rather than

the ‘phonetic level’. In other words, the base or smallest entity of this system is a ‘word’ (in Bangla) rather than sounds (phonemes) that construct the words. The authors also want to mention that HMM based speech recognizers work from the phonetic level as opposed to a ‘word level’ since most of the HMM based systems are optimized for English speech.

In the authors’ FIS, three inputs have been used, e.g., frequency, energy level of the sample, and the energy level of the target or description. Since human ear is more susceptible to lower frequencies of sounds, FIS rules are made accordingly to put emphasize on the lower frequencies. The output of FIS is the similarity between two ‘segments’ of a word and the overall evaluation of the FIS has been cumulated to reach the verdict of a word recognition.

However, the authors rely on cepstral analysis for our inputs of the ANN based system. We designed our network using the feed-forward back-propagation architecture with one hidden layer (between the input and output layers) consisting of 60 neurons (the justification is elaborated in subsection 3.10.2). The inputs for our ANN based system are somewhat similar to the inputs of our fuzzy logic based system; however, instead of using spectral analysis directly, we take the energy levels and frequency of audio signals to the paradigm of cepstral analysis. Using Mel-frequency cepstral coefficients (MFCC) as features in the input, we recognize the Bangla words from the outputs of the ANN.

The organization of the paper is as follows: Section 2 discusses the related works done till date in relevance to speech recognition emphasizing on Bangla speech (phoneme descriptions, vowels and recognition systems) in particular. Section 3 presents the detailed descriptions of the strategies that we implement to build our systems. In Section 4 we report and analyze the experimental results. Finally, Section 5 concludes and gives directions of our future research.

31 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/bangla-user-adaptive-word-speech-recognition/124535

Related Content

How to Design and Frame a Questionnaire

Rayees Farooq (2018). *Innovations in Measuring and Evaluating Scientific Information* (pp. 50-60).

www.irma-international.org/chapter/how-to-design-and-frame-a-questionnaire/199964

Analyzing Qualitative Data: Words, Words, Words!

Janice E. Jones (2020). *Applied Social Science Approaches to Mixed Methods Research* (pp. 87-105).

www.irma-international.org/chapter/analyzing-qualitative-data/244117

Application of Multi-Methods in Understanding Knowledge Retention in the Cross-Border Mergers of the Telecommunications Industry in Lesotho

Tahleho Emmanuel Tsele (2022). *Handbook of Research on Mixed Methods Research in Information Science* (pp. 440-469).

www.irma-international.org/chapter/application-of-multi-methods-in-understanding-knowledge-retention-in-the-cross-border-mergers-of-the-telecommunications-industry-in-lesotho/291205

Ethical Issues Underlying the Instructor's Manual for Teaching With Cases: Exploring the Dark Side of the Pedagogy

Herbert Sherman and Adva Rachel Dinur (2019). *Scholarly Ethics and Publishing: Breakthroughs in Research and Practice* (pp. 69-92).

www.irma-international.org/chapter/ethical-issues-underlying-the-instructors-manual-for-teaching-with-cases/222302

Hybrid Metaheuristic to Optimize Traceability in the Food Industry

Saima Dhoubi (2021). *International Journal of Strategic Engineering* (pp. 14-27).

www.irma-international.org/article/hybrid-metaheuristic-to-optimize-traceability-in-the-food-industry/279643