

SPEECH RECOGNITION SYSTEM

Abstract:

“Speech Recognition System” also termed as “Speech to Text” is a software technology that lets the user control the computer functions and dictate text by voice hence reducing human efforts. “Speech Recognition System” or the problem of “Automated Speech Recognition” has been studied over past several years. “Speech Recognition Systems” have revolutionized the approach of people with disabilities and have made lives easier by reducing Human Efforts. Early speech recognition systems tried to apply a set of grammatical and syntactical rules to speech. If the words spoken, fit into a certain set of predefined rules, the “Speech Recognition System” could determine what the words were. However, there were certain limitations such as different densities of voice, varied accents, change in dialects and mannerisms in the human speech which eventually had a lot of flaws in the approach.

Today’s “Speech Recognition Systems” use powerful and statistical modeling systems. The current systems use concepts like Probability and Mathematical Functions to determine the likelihood of an event. According to John Garofolo, Speech Group Manager at the Information Technology Laboratory of the National Institute of Standards and Technology, the two models that dominate the field today are the Hidden Markov Model (HMM) and Neural Networks. These methods involve complex mathematical functions. In HMM, speech generation is modeled as a Stochastic Process. Hidden Markov Models are expressed in terms of phonetic states and acoustic emissions and are parameterized by transition and emission probabilities. In this model, each phoneme is like a link in a chain, and the completed chain is a word. However, the chain branches off in different directions as the program attempts to match the digital sound with the phoneme that's most likely to come next. During this process, the program assigns a probability score to each phoneme, based on its built-in dictionary and user training. John Garofolo said, “The user training and statistical systems need a lot of exemplary training to reach out to the optimal performance.”

None of the Speech Recognition Systems can be 100% efficient and accurate. Current Speech Recognition Systems have certain limitations. Some of them are Low Signal-to-Noise ratio, Overlapping speech, Intensive use of Computer Power, Homophones. The above stated limitations have been focused on in our Speech Recognition System.

References:

- [1] Geoffrey Zweigh and Stuart Russell, “Probabilistic Modeling with Bayesian Networks for Automatic Speech Recognition”, IBM T. J. Watson Research Center, University of California, Berkeley.

- [2] Mathias De Wachter, Mike Matton, Kris Demuynck, “IEEE Transactions on Audio, Speech and Language Processing”, Volume: 15, Issue: 4, Date: May 2007.