Decision level data fusion in Speech and Image Recognition Systems

Gaurav Sanjay Newalkar (12EC43), Nithin Rao Koluguri(12EC51),
Nikhil Lunavath(12EC68)

Project Guide: Dr.Ashvini Chaturvedi
Dept of E & C Engg

July - December 2015

Contents

Introduction

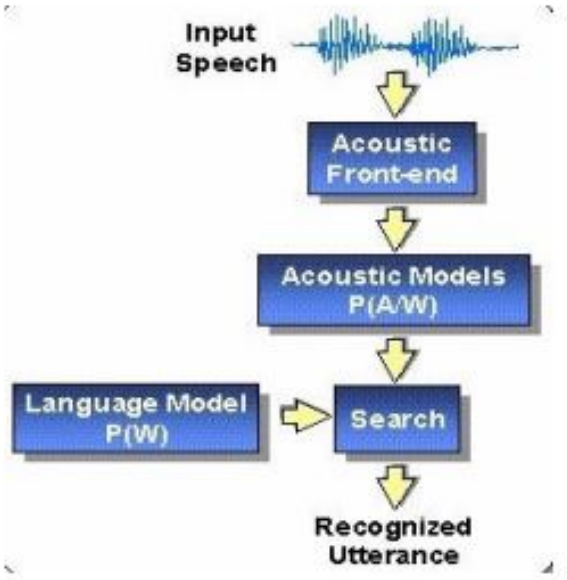
- Automatic Speech Recognition System
- Its applications are numerous,like in our everyday devices etc.

Objectives

- To recognize 4 different words through automatic speech recognition and convert to 3 digit binary word.
- Based on the 3 digit binary word, A Image Recognition system to be developed to recognize the object pertaining to the command given.

Literature Survey

- Peeling and Moore(1987) applied MLPs to digit recognition with excellent results using NN over HMM.[1]
- Kammerer and Kupper (1988) applied a variety of networks to the TI 20-word database, finding that a single-layer perceptron outperformed both multi-layer perceptrons and a DTW template-based recognizer [2]
- Lipmann(1989) and Burr(1988) also shown how NN perform better recognition over HMM in small data sets.[3]



(IJCSIS

Fig.1 Basic model of speech recognition

Figure: Basic model of speech recognition

Procedure

- Speech recognition systems can be separated in several different classes by describing what types of utterances they have the ability to recognize
- ISOLATED WORDS, Connected Words, Continuous Speech, Spontaneous Speech

Approaches to Speech Recognition

- a)Acoustic Phonetic Approach
- b)PATTERN RECOGNITION APPROACH
- c)Artificial Intelligence Approach

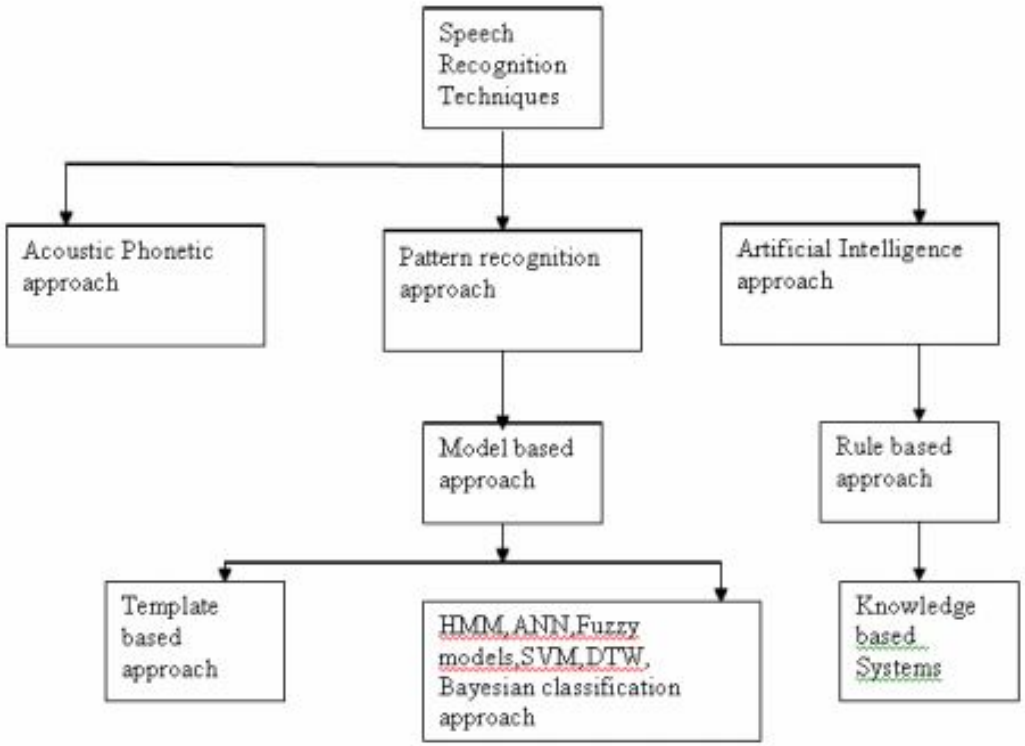


Figure: Approaches

General Process

- The Automatic Speaker Recognition(ASR) loads a list of words to be recognized. This list of words is called a grammar.
- The ASR loads audio from the speaker. This audio is represented as a waveform, essentially the mathematical representation of sound.
- The ASR compares the waveform to its own acoustic models and are what allow the engine to recognize speech[4]

General Process

- The ASR compares the words in the grammar to the results it obtained from searching its acoustic models.
- It then determines which words in the grammar the audio most closely matches and returns a result.

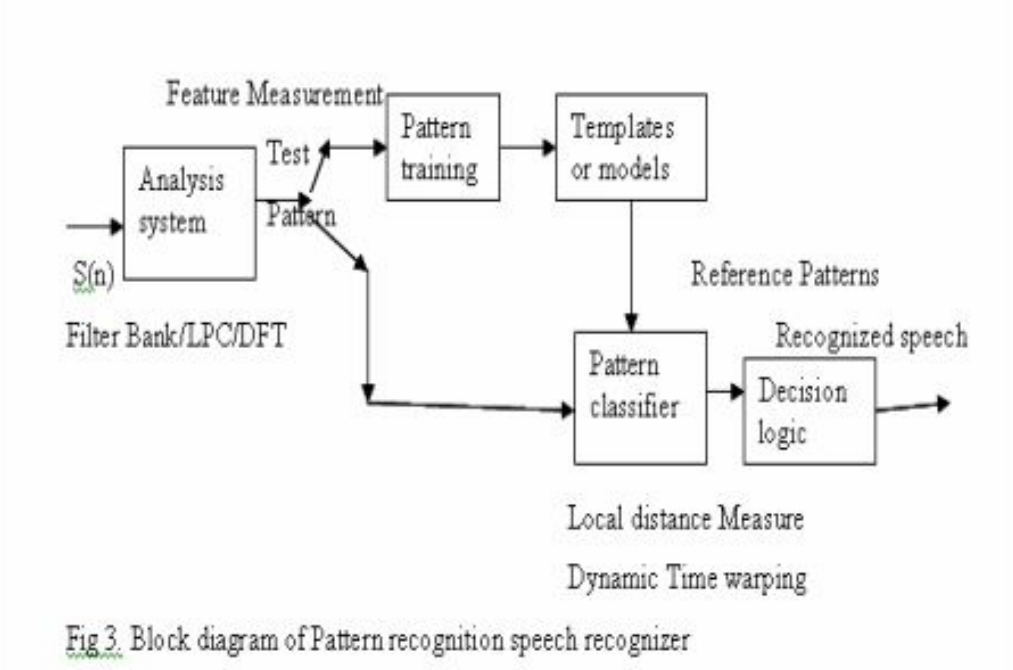


Figure: Process

Feature Extraction and Classifier

Method	Property	Comments
Principal Component Analysis(PCA)	Non linear feature extraction method, Linear map; fast; eigenvector-based	Traditional, eigenvector based method, also known as karhuneu-Loeve expansion; good for Gaussian data.
Linear Discriminant Analysis(LDA)	Non linear feature extraction method, Supervised linear map; fast; eigenvector-based	Better than PCA for classification;
Independent Component Analysis (ICA)	Non linear feature extraction method, Linear map, iterative non-Gaussian	Blind course separation, used for de-mixing non-Gaussian distributed sources(features)

Figure: Feature Extraction

Future Scope

- There are copious future applications for speech recognition.
- In Military and Internet of Things for example.

References

S.M. Peeling, R.K. Moore, *" Isolated digit recognition experiments using the multi-layer perceptron"*,1987

Bernhard R. Kammerer, Wolfgang A. Kuapper *"Experiments for isolated-word recognition with single- and two-layer perceptrons"*, 1988

Burr, D.J., *""Experiments on neural net recognition of spoken and written text"*, IEEE Transactions on , vol.36, no.7, pp.1162-1168, Jul 1988.

. M.A.Anusuya ,S.K.Katti , *"Speech recognition by machine: A review"*, in Proceedings of the IEEE , vol.64, no.4, pp.501-531, April 2009.