International Journal of Advanced Research in Engineering Technology & Science

ISSN: 2394-2819

Email: editor@ijarets.org Volume-3, Issue-1 January- 2016 www.ijarets.org

INTRODUCTION TO TECHNIQUES USED IN VOICE RECOGNITION

[1]Vennu Santosh Kumar, [2]Dr. K.P. Yadav

[1]Research Scholar,[2] Research Supervisor

[1,2]Department of Computer Science & Engineering

[1,2]Sunrise University, Rajasthan, Rajasthan

ABSTRACT:

Voice is a natural mode of communication for people. We learn all the relevant skills during early childhood, without instruction, and we continue to rely on voice communication throughout our lives. It comes so naturally to us that we don't realize how complex a phenomenon voice is. The human vocal tract and articulators are biological organs with nonlinear a property, whose operation is not just under conscious control but also affected by factors ranging from gender to upbringing to emotional state.

KEY WORDS*:* communication, phenomenon, nonlinear

INTRODUCTION:

As a result, vocalizations can vary widely in terms of their accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed; moreover, during transmission, our irregular speech patterns can be further distorted by background noise and echoes, as well as electrical characteristics (if telephones or other electronic equipment are used). All these sources of variability make speech recognition, even more than voice generation, a very complex problem. Yet people are so comfortable with voice that we would also like to interact with our computers via speech, rather than having to resort to primitive interfaces such as keyboards and pointing devices. A voice interface would support many valuable applications — for example, telephone directory assistance, spoken database querying for novice users, "hands busy" applications in medicine or fieldwork, office dictation devices, or even automatic voice translation into foreign languages. Such tantalizing applications have motivated research in automatic speech recognition since the 1950's. Great progress has been made so far, especially since the 1970's, using a series of engineered approaches that include template matching, knowledge engineering, and statistical modeling. Yet computers are still nowhere near the level of human performance at voice recognition, and it appears that further significant advances will require some new insights. What makes people so good at recognizing voice? Intriguingly, the human brain is known to be wired differently than a conventional computer; in fact it operates under a radically different computational paradigm. While conventional computers use a very fast & complex central processor with explicit program instructions and locally addressable memory, by contrast the human brain uses a massively parallel collection of slow & simple processing elements (neurons), densely connected by weights (synapses) whose strengths are modified with experience, directly supporting the integration of multiple constraints, and providing a distributed form of associative memory. The brain's impressive superiority at a wide range of cognitive skills, including voice recognition, has motivated research into its novel computational paradigm since the 1940's, on the assumption that brain like models may ultimately lead to brain like performance on many complex tasks. This fascinating research area is now known as *connectionism*, or the study of *artificial neural networks*. The history of this field has been erratic (and laced with hyperbole), but by the mid-1980's, the field had matured to a point where it became realistic to begin applying connectionist models to difficult tasks like voice recognition. By 1990, many researchers had demonstrated the value of neural networks for important subtasks like phoneme recognition and spoken digit recognition, but it was still unclear whether connectionist techniques would scale up to large speech recognition tasks. This proposed work demonstrates that neural networks can indeed form the basis for a general purpose voice recognition system, and that neural networks offer some clear advantages over conventional techniques.

BASIC CONCEPTS

A Hidden Markov Model is a collection of states connected by transition, as illustrated in Figure 1. It begins in a designated initial state. In each discrete time step, a transition is taken into a new state and then one output symbol generated in that state. The choice of transition and output symbol are both random, governed by probability distributions. The HMM can be thought of as a black box, where the sequence of output symbols generated over time is observable, but the sequence of states visited over time is hidden from view. This is why it's called a Hidden Markov Model.

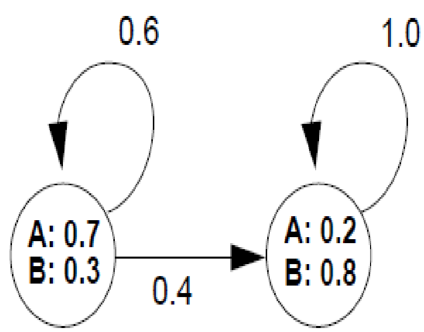

Figure 1: Hidden Markov Model, with two states A and B

HMMs have a variety of application. When an HMM applied to speech recognition, the state are interpreted as acoustics model, indicating what sounds are likely to be heard during their corresponding segment of speech; while the transition provide temporal constraints, indicating how the states may follow each other in sequence. Because speech always goes forward in time, transition in an speech application always go forward (or make a self-loop, allowing a state to have arbitrary duration). Figure 2 illustrates how states and transition in an HMM can be structured hierarchically, in order to present phonemes, words, and sentences.

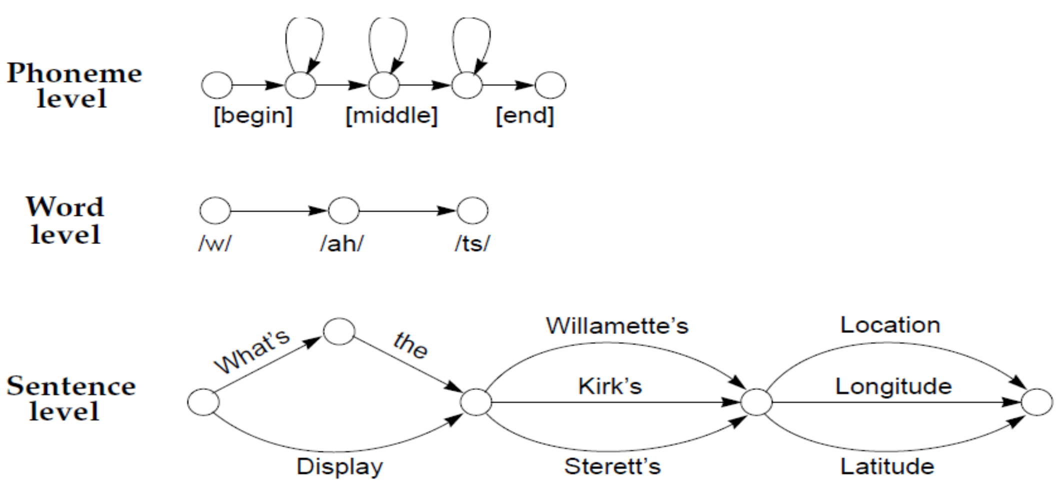

Figure 2: A hierarchically structure of HMM

Formally, an HMM consists of the following elements:

{s} = A set of states

{a_{ij}} = A set of transition probabilities, a_{ij} is the probability of taking the transition from state i to state j.

{$b_i(u)$} = A set of emission probabilities, where bi is the probability distribution over the acoustic space describing the likelihood of emitting each possible sound u while in state i.

Since a and b are both probabilities, they must satisfy the following properties:

$$a_{ij} \geq 0, \; b_i(u) \geq 0, \qquad i,j,u \tag{1}$$

$$\sum a_{ij} = 1, \qquad \forall \; i \tag{2}$$

$$\sum b_i(u) = 1, \qquad \forall \; i \tag{3}$$

In using this notation we implicitly confine our attention to First-Order HMMs, in which **a** and **b** depend only on the current state, independent of the previous history of the state sequence. This assumption, almost universally observed, limits the number of trainable parameters and makes the training and testing algorithms very efficient, rendering HMMs useful for speech recognition.

RESULTS:

It is produced by forcing air through the glottis, proper adjustment of the tension of the vocal cords results in opening and closing of the cords, and a production of almost periodic pulses of air. These pulses excite the vocal tract. Psychoacoustics experiments show that this part holds most of the information of the speech and thus holds the keys for characterizing a speaker.

REFERENCES:

1. Youssef, A. M. and Hassoun, M. H. (1989). "Dynamic Autoassociative Neural Memory Performance vs. Capacity," *Proc. SPIE, Optical Pattern Recognition*, H.-K. Liu, Editor, **1053**, 52-59.
2. Widrow, B., Gupta, N. K., and Maitra, S. (1973). "Punish/Reward: Learning with a Critic in Adaptive Threshold Systems," *IEEE Trans. on System, Man, and Cybernetics*, **SMC-3**, 455-465.
3. Tollenaere, T. (1990). "SuperSAB: Fast Adaptive Back Propagation with Good Scaling Properties," *Neural Networks*, **3**(5), 561-573.
4. Reeves, C. R. (1993). "Using Genetic Algorithms with Small Populations," *Proceedings of the Fifth International Conference on Genetic Algorithms* (Urbana-Champaign 1993), S. Forrest, Editor, 92-99. Morgan Kaufmann, San Mateo.
5. Pandey A., Bansal K.K. “Performance Evaluation of TORA Protocol Using Random Waypoint Mobility Model” International Journal of Education and Science Research Review Vol.1 (2) pp 193-199
6. Yadav R.K., Bansal K.K. “Analysis of Sliding Window Protocol for Connected Node “International Journal of Soft Computing and Engineering (IJSCE) Vol.2 (5) PP. 292-294
7. Tiwari S.P.,Kumar S.,Bansal K.K.(2014) “A Survey of Metaheuristic Algorithms for Travelling Salesman Problem ” International Journal Of Engineering Research & Management Technology Vol.1(5)
8. Kumar N.,Bansal K.K.,(2012) “Different Compression Techniques and Their Execution In Database Systems To Improve Performance” International Journal of Advanced Research in Computer Science and Software Engineering Vol.2(6) pp. 293-296
9. Lay, S.-R. and Hwang, J.-N. (1993). "Robust Construction of Radial Basis Function Networks for Classification," in *Proceedings of the IEEE International Conference on Neural Networks* (San Francisco 1993), vol. III, 1859-1864. IEEE, New York.

10. Hassoun, M. H., Wang, C., and Spitzer, A. R. (1994b). "NNERVE: Neural Network Extraction of Repetitive Vectors for Electromyography, Part II: Performance Analysis," *IEEE Transactions on Biomedical Engineering*, XXX to appear XXX.
11. Barto, A. G. and Anandan, P. (1985). "Pattern Recognizing Stochastic Learning Automata," *IEEE Trans. on Systems, Man, and Cybernetics*, **SMC-15**, 360-375.