International Journal of Advance Research in Science and Engineering
Vol. No.4, Special Issue (01), Spetember 2015
www.ijarse.com
IJARSE
ISSN 2319 - 8354

AUTOMATIC SPEECH RECOGNITION FOR MACHINES IN DIGITAL SIGNAL PROCESSING USING ARTIFICIAL NEURAL NETWORKS: A REVIEW

Sudha Sharma[1], Dr. Manish Mann[2]

[1,2]Department of Computer Science and Engineering, L.R.I.E.T. Solan,
Himachal Pradesh Technical University, (India)

ABSTRACT

Speech Recognition technology is a fast growing engineering technology now days. Speech Recognition is the ability for a device to recognize individual words or phrases from human speech. These words further can be command the operation of a system. This paper describes the basics of speech recognition including its types and the algorithm used in speech recognition. This paper described the artificial neural networks used for automatic speech recognition in detail. There would be number of factors that can affect and does matters in speech recognition, few of these are discussed in this paper with application in numerous fields. The history of speech recognition is described in detail in this paper.

Keywords: Speech Recognition, Artificial Neural Networks, Isolated Speech, Hidden Markow Model, Dynamic Time Warping.

I. INTODUCTION

Digital signal processing refers to various techniques for improving the accuracy and reliability of digital communications. A digital signal processing is able to differentiate between human made signals, which are orderly, and noise. Digital signal processing is used to remove the noise from the speech sample. Signal processing is the process of extracting the relevant information from the speech signal in an efficient and robust manner.

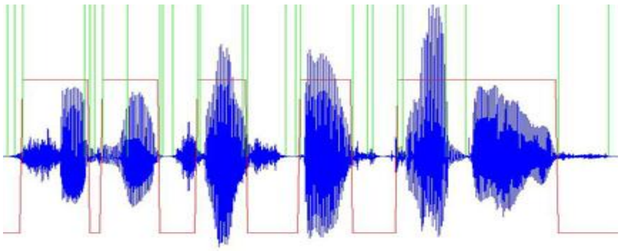

Figure1. Waveform Representation of Speech Signal

There are many applications of digital signal processing. The speech recognition is one of the application of the digital signal processing. A speech recognition system comprises a collection of algorithms drawn from a wide variety of disciplines, including statistical pattern recognition, communication theory, signal processing, combinational mathematics, and linguistics, among others. Although each of these areas is relied on to varying degrees in different recognizers, perhaps the greatest common denominator of all recognition systems is the signal processing front end, which converts the speech waveform to some type of parametric representation for further analysis and processing. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. [8] The smallest unit of spoken language is known as a Phoneme. The English language contains approximately 44 phonemes representing all the vowels and consonants that we use for speech. [5]

Historically the sounds of spoken language have been studied at two different levels: (1) *phonetic* components of spoken words, e.g., vowel and consonant sounds, and (2) *acoustic* wave patterns. A language can be broken down into a very small number of basic sounds, called phonemes (English has approximately forty). An acoustic wave is a sequence of changing vibration patterns (generally in air), however we are more accustom to "seeing" acoustic waves as their electrical analog on an oscilloscope (time presentation) or spectrum analyzer (frequency presentation). Also seen in sound analysis are two-dimensional patterns called spectrograms. [1]

There are seven layers in describing a speech. These are:-

Signal Processing	
Acoustic	Language Independent
Phonetics	
Phonological	
Morphology	
Syntax	Language Dependent
Semantic	
Pragmatics	

Speech is a complex phenomenon. People rarely understand how is it produced and perceived. The naive perception is often that speech is built with words, and each word consists of phones. The reality is unfortunately very different. Speech is a dynamic process without clearly distinguished parts. It's always useful to get a sound editor and look into the recording of the speech and listen to it. Here is for example the speech recording in an audio editor.

Speech is a continuous audio stream where rather stable states mix with dynamically changed states. In this sequence of states, one can define more or less similar classes of sounds, or phones. Words are understood to be built of phones, but this is certainly not true. The acoustic properties of a waveform corresponding to a phone can vary greatly depending on many factors - phone context, speaker, style of speech and so on. Remember nine (N AY N vs. N AY AH N) shown here in one of the "standard" phoneme sets.

1.1 History of Speech Recognition

In 1950s and 1960s, the researches made it possible to control device using only their voice. The first speech recognition systems could understand only digits. Bell Laboratories designed in 1952 the "Audrey" system, which recognized digits spoken by a single voice. Ten years later, IBM demonstrated at the 1962 World's Fair its "Shoebox" machine, which could understand 16 words spoken in English. 1970s, The DoD's DARPA Speech Understanding Research (SUR) program, from 1971 to 1976, was one of the largest of its kind in the history of speech recognition, and among other things it was responsible for Carnegie Mellon's "Harpy" speech-understanding system. Harpy could understand 1011 words, approximately the vocabulary of an average three-year-old.

In 1980s, Speech Recognition Turns Toward Prediction Over the next decade, thanks to new approaches to understanding what people say, speech recognition vocabulary jumped from about a few hundred words to several thousand words, and had the potential to recognize an unlimited number of words. One major reason was a new statistical method known as the *hidden Markov model*. However, whether speech recognition software at the time could recognize 1000 words, as the 1985 Kurzweil text-to-speech program did, or whether it could support a 5000-word vocabulary, as IBM's system did, a significant hurdle remained: These programs took discrete dictation, so you had … to … pause … after … each … and … every … word.[6]

Speech recognition research has been ongoing more than 80 years. By 2001, computer speech recognition had reached 80% accuracy and no further process was reported till 2010. Speech recognition technology development began to edge back into the front with one major event: the arrival of the “Google voice search app for the iPhone “. In 2010 Google added “personalized recognition to “voice search on android phones, so that the software could record users”. Voice search and produce a more accurate speech model. It draws its knowledge about the speaker to generate a conceptual reply and responds to voice input. [2]

II. TYPES OF SPEECH RECOGNITION

Divided into the number of classes based on their ability to recognize that words and list of words they have:

1) Isolated speech:- the isolated speech involve a pause between two utterance ,it doesn’t mean that it only accepts a single word but instead it requires one utterance at a time.[4]
2) Connected speech:- connected words or connected speech is similar to isolated speech but allow separate utterance with minimum pause between them.[4]
3) Continuous speech: - continuous speech allows the user to speak almost naturally; it is also called the computer dictation.[4]
4) Spontaneous speech: - Spontaneous speech is natural sounding and not reheard. An automatic speech recognition system with spontaneous speech ability should be able to handle a variety of natural speech features such as words being together “ums” and “ahs” and even slight stutters.[4]

III. MODELS IN SPEECH RECOGNITION

According to the speech structure, three models are used in speech recognition:

1) An acoustic model contains acoustic properties for each senone. There are context-independent models that contain properties and context-dependent ones (built from senones with context).
2) A phonetic dictionary contains a mapping from words to phones. This mapping is not very effective. For example, only two to three pronunciation variants are noted in it, but it's practical enough most of the time. The dictionary is not the only variant of mapper from words to phones. It could be done with some complex function learned with a machine learning algorithm. [7]
3) A language model is used to restrict word search. It defines which word could follow previously recognized words (remember that matching is a sequential process) and helps to significantly restrict the matching process by stripping words that are not probable. Most common language models used are n-gram language models-these contain statistics of word sequences-and finite state language models-these define speech sequences by finite state automation, sometimes with weights.

IV. APPLICATIONS OF SPEECH RECOGNITION

1) **Medical perspective: -** people with disability can benefit from speech recognition programs. Speech recognition is especially useful for people who have difficulty using their hands.
2) **For military purpose: -** In air force speech recognition has defined potential for reducing pilot workload. Beside the air force such programs can also be trained to be used in helicopters battle management and their applications. They didn’t have to use their hands.
3) **For education perspective: -** Individual with learning disability who have problem with thought to paper communication can benefit from system.
4) Radiology scanning hundreds of X-rays, ultra sonograms, CT scans and simultaneously dictating conclusion to a speech recognition system connected to word processor. The radiologist can focus his attention on the image rather than writing the text.[9]
5) Voice recognition could also be used on computer for making airline and hotel reservation. A user required simply stating his needs to make reservations, cancel a reservation or making enquiry about schedule.[9]

V. ALGORITHMS USED IN SPEECH RECOGNITION

The digital signal processes such as Feature Extraction and Feature Matching are introduced to represent the voice signal. Several methods such as Liner Predictive Coding (LPC), Hidden Markov Model (HMM), Artificial Neural Network (ANN) etc are evaluated with a view to identify a straight forward and effective method for voice signal. There are mainly 3 algorithms that are used for SR. Those are given below:

- Hidden Markov Model(HMM)
- Dynamic Time Warping(DTW)
- Artificial Neural Networks(ANN)

5.1 Artificial Neural Networks (Ann)

A neural network can be defined as a model of reasoning based on the human brain. The brain consists of a densely interconnected set of nerve cells, or basic information-processing units, called neurons. The human brain incorporates nearly 10 billion neurons and 60 trillion connections, *synapses*, between them. By using multiple neurons simultaneously, the brain can perform its functions much faster than the fastest computers in existence today.

Each neuron has a very simple structure, but an army of such elements constitutes a tremendous processing power. A neuron consists of a cell body, soma, a number of fibers called dendrites, and a single long fiber called the axon.

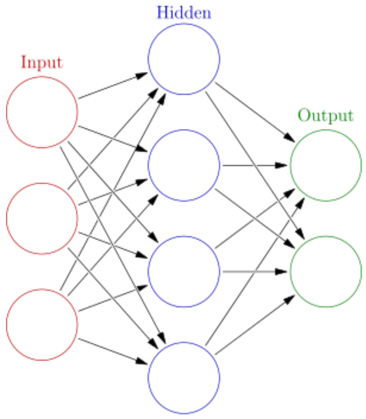

Figure2: Artificial Neural Netork

An artificial neural network consists of a number of very simple processors, also called neurons, which are analogous to the biological neurons in the brain. The neurons are connected by weighted links passing signals from one neuron to another. The output signal is transmitted through the neuron's outgoing connection. The outgoing connection splits into a number of branches that transmit the same signal. The outgoing branches terminate at the incoming connections of other neurons in the network.

5.2 Advantages of Ann

- ANNs are highly non-linear modeling.
- ANN is nonlinear model that is easy to use and understand compared to statistical methods.
- ANN is non-parametric model while most of statistical methods are parametric model that need higher background of statistic.
- ANN with Back propagation (BP) learning algorithm is widely used in solving various classifications and forecasting problems. Even though BP convergence is slow but it is guaranteed.
- Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms.

5.3 Applications of Ann

- Character Recognition
- Image Compression
- Stock Market Prediction
- Travelling Salesman Problem
- Medicine and Security.

VI. CONCLUSION

Speech recognition is the interaction between the human and the machine. Speech Recognition should be helpful for the people who are suffering from various disability such as in case of blindness or if unable to use their hands effectively. Speech recognition can be helpful by operating input through voice input. It does involve various processes, by using number of method and techniques for having the entire process/ operation to be done. Number of different techniques can be studied and applied for checking the better results in future

RFERENCES

[1] Dunn, and Lacey, "The sound spectrograph was first described in Koenig", *Journal of the Acoustical Society of America* (18, 19-49) in 1946.

[2] Suma Swamy & K.V. Ramakrihnan, "An Efficient speech recognition system", *CS & Engineering: An International Journal(CSIJ,)* Vol.3,No. 4,2013.

[3] Pahini A. Trivedi," Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks*", International Journal of Engineering Development and Research*, Volume 2, Issue 4, 2014.

[4] http://en.wikipedia.org/wiki/accoustic_model

[5] http://www.cs.stir.ac.uk/~kjt/research/match/resources/tutorial/Speech_Language/Speech_Recognition/Rec_2.html

[6] http://www.techhive.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html

[7] http://cmusphinx.sourceforge.net/wiki/tutorialconcepts

[8] Urmila Shrawankar, Anjali Mahajan," Speech: A Challenge to Digital Signal Processing Technology For Human-to-Computer Interaction", Conference proceeding National a conference on recent trends in electronics & information technology, pp.206-212, 2013.

[9] Raghvendra Priyam, Rahmi Kumari & Dr. Prof Videh Kihori Thakur, "Artificial Intelligence Applications for speech Recognition ", *Conference on Advances in communications and Control system*,pp.475-476,2013.