International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: 6.614
Available Online at www.journalijcar.org
Volume 8; Issue 01(F); January 2019; Page No. 17102-17107
DOI: http://dx.doi.org/10.24327/ijcar.2019.17107.3189

Review Article

SPEAKER RECOGNITION AND ITS ASSOCIATION WITH COGNITIVE SCIENCE: A BRIEF REVIEW

RishabhDeo Pandey[1]., Sumit Srivastava[2] and Sushma Pandey[3]

[1]CSE (Pursuing) Birla Institute of Technology, Mesra, Ranchi, India-835215
[2]Department of CSE, Birla Institute of Technology, Mesra, Ranchi, India-835215
[3]Department of Psychology, D. D.U. Gorakhpur University, Gorakhpur, U.P. India-273009

A R T I C L E I N F O

Article History:
Received 4th October, 2018
Received in revised form 25th October, 2018
Accepted 18th December, 2018
Published online 28th January, 2019

Key words:
Artificial intelligence, cognitive psychology, cognitive science, linguistics, speech perception, speaker recognition.

A B S T R A C T

Over the last few decades, a remarkable growth of research in the field of cognitive science has been identified. Cognitive science tries to answer about mind. It is a multidisciplinary field which includes within its scope various disciplines i.e., Psychology, Philosophy, Anthropology, Neuroscience, Linguistics and Computer Science. The purpose of this article is to briefly elucidate the nature of cognitive science and its link with computer science (Artificial Intelligence). Moreover the association between artificial intelligence and linguistics is outlined. Specifically, Speaker Recognition as an active process of cognition has been analyzed in term of its background, nature and forms. Finally, the application of speaker recognition and its limitations are discussed.

INTRODUCTION

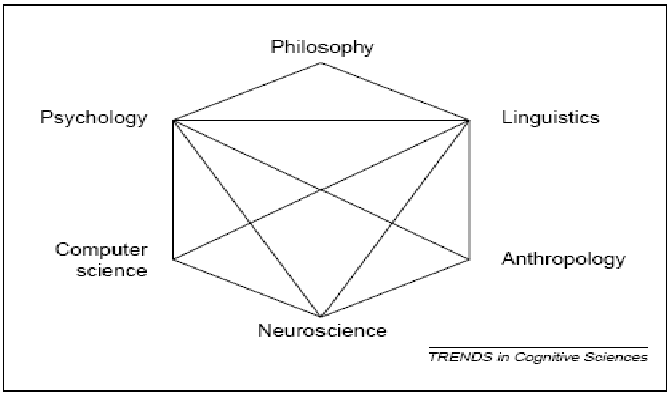
The development of research interest in the field of cognition has been considered as a great revolution of 20th century (Gardner, 1985; Maltin, 1989). Three causative factors worked in the background of this revolution during 1940 to 1960 (Lachman et al., 1979). First, Behaviorism, the dominant paradigm in that era, failed to explain how people understand and acquire language (Chomsky, 1959). Secondly, the development of Information Processing/ Communication Theory (Shannon and Weaver, 1949) provided a method of measuring how the amount of information flowing through a given system and third, the advent of digital computers offered psychologists both computational metaphor and computer simulation for the investigation of mind.

This article endeavours to present an overview of cognitive science, its scope and links with other disciplines. Moreover, the pervasive association between cognitive science, linguistics and artificial intelligence (speech/ speaker recognition) are described. Lastly, the historical background, nature and current forms of speaker recognition, its utility and shortcomings are also elucidated.

Corresponding author:* **RishabhDeo Pandey
CSE (Pursuing) Birla Institute of Technology, Mesra, Ranchi, India-835215

Cognitive Science

Cognitive Science is the multidisciplinary scientific study of mind and its functioning. It investigates what cognition is, what it does and how it works. Cognitive Science consists of multiple research disciplines including psychology, philosophy, anthropology, neuroscience, linguistics and computer science (Fig.1).



The concept behind cognitive science is "that thinking can be understood in terms of representational structures in the mind and computational procedures that operate on those structures". A central tenet of cognitive science is that a complete understanding of mind cannot be done by studying only a single level. Therefore, studying a particular phenomenon from multiple levels creates a better understanding of the processes that occur in the brain to give rise to a particular behaviour.

Marr (1962) gave a famous description of three levels of analysis i.e. (1) The computational theory, specifying the goals of the computation, (2) Representation and algorithm, giving a representation of the input and output and the algorithm which transforms one into the other; and (3) The hardware implementation, how algorithm and representation may be physically realized.

Therefore, Cognitive Science is considered an interdisciplinary field with contributions from various fields. Cognitive Science leads to observe the world outside the mind much as other sciences do. Thus, it too has an objective, observer-independent existence. The field is usually seen as compatible with the physical sciences and uses the scientific method as well as simulation or modeling, often comparing the output of models with aspects of human behaviour (Pandey, 2014). Cognitive scientists prefer interdisciplinary researches however; they have less interaction with other fields. Whereas, Cognitive psychologists collaborate in research with Cognitive Science and its other branches viz; Cognitive Neuroscience and Artificial Intelligence (Al).The association between artificial intelligence and cognitive psychology/ cognitive science has been briefly discussed in the following the section.

Artificial Intelligence and Cognitive Psychology/Cognitive Science

Artificial Intelligence (AI), a branch of computer science aims to explore human cognitive processes by creating models that exhibit “intelligent” behavior (Wagman, 1999). Researchers in artificial intelligence have tackled many cognitive tasks such as medical problem solving, legal reasoning, and spatial-map learning (Thrun, et al., 1998;Wagman, 1999). Some aspects of Artificial Intelligence (AI) are described.

***The computer metaphor**:* Throughout the history of cognitive psychology the computer has been a popular metaphor for the human mind (different kinds of metaphors have intrigued theorists for centuries). As early as 430 B.C., philosophers compared the human mind with a machine (Marshal,1977). The activity of the brain has also been compared to a telephone exchange and to weaving on a loom. So the computer metaphors represented in artificial intelligence is one of the more recent in a long list of machine metaphors.

According to the computational metaphor, "our cognitive processes work like a computer, a complex, multipurpose machine that processes information quickly and accurately”. Of course researchers acknowledge obvious differences in physical structure between the computer and the human brain that manages our cognitive processes. However, both may operate according to similar general principles. Like humans, computers feature a variety of internal mechanisms. For example, both computers and humans can compare symbols and can make choices according to the results of the comparison. Furthermore, computers have a central-processing mechanism with a limited capacity (Luger, 1994) and humans also have a limited attention capacity.

Researchers who favour the computational approach try to design the appropriate ‘Software’ with the right computer program and sufficient mathematical detail, researchers hope to mimic the adaptability and the efficiency of human cognitive processes (Guenther 1995). Scientists working in the area of artificial intelligence (Al) point out the analogy between the human mind and the computer because computer programs must be detailed, precise unambiguous, and logical, Researchers can represent the functions of a computer with a flowchart that shows the sequence of stages in processing. The flowchart also illustrates the relationships equivalent performance on a particular task, and then the researchers can speculate that the program which directed the computer represents an appropriate theory for describing the human's cognitive processes (Carpenter & Just, 1989, Lewandowsky, 1993).

However, computer cannot finely duplicate human cognitive processes. For example, human have more complex and fluid goals, People playing a game of chess may be concerned about how long the game lasts, whether they have other social obligations, and now they will interact socially with their opponent. Contrary to this, computer's goals are simple and rigid; the computer deals only with the outcomes of the chess game (Eysenck, 1984; Neisser, 1963).

***Pure Artificial Intelligence**:* Pure Al is an approach that seeks to accomplish a task as efficiently as possible For example., the most successful computer programs for chess will evaluate as many potential moves as possible. The goal of pure Al is to be efficient, not to be human. Not surprisingly, Deep Blue won most of the matches. Franklin (1995) lists some to the tasks that can be accomplished by pure Al systems, such as playing chess, speaking English, and diagnosing an illness.

Computer Simulation: The goal of computer simulation is to design a system that simulates or resembles human performance on a selected cognitive task (Carpenter and Just 1989).Computer simulation research has been most active in such areas as basic visual processing, language processing, and problem solving. For example, Carpenter and Just (1989) created a computer-simulation model for reading sentences. The model was based on the assumption that humans have limited capacity to process information. A sizeable number of studies have been conducted by cognitive psychologists for other cognitive processes like, pattern recognition and language processes i.e. speech perception or speaker recognition.

Natural languages are those that have evolved in human societies and are used by human beings, such as Hindi, English, Spanish, and French. These are in contrast to formal computer language such as C^{++}, or linguistic expressions of logic. There are two kinds of natural language processing. Understanding a natural language involves an individual’s assimilation of linguistic expression in some from, such as speech or writing, extracting its meaning, and then undertaking some action that constitutes a response to this meaning. Understanding is what a computer would need to do if it were to interpret a spoken human taking a formal symbolic representation of an idea and converting it to an expression in English or some other natural language. A computer would be generating language if it could transform this idea into a spoken utterance that a human being could understand. These two processes are thus the computer equivalent of natural language comprehension and production. Cawsey (1998) describes four stages of natural language understanding; Speech recognition, Syntactic analysis, Semantic analysis, and Pragmatic analysis.

In this section, we will concern ourselves exclusively with speech/speaker recognition as that is the area in which research has been concentrated.

Speech/ Speaker Recognition

Speech recognition by machine is a laudable aspiration. Humans use language as a medium to communicate its ideas or thoughts with others. To be able to communicate in a similar way with computers would usher in a new age of efficiency and productivity. Unfortunately, the task of getting a machine to understand speech is much more difficult than it may seem. Let's review some of the steps that speech recognition by machine would have to include and talk about the problems involved.

The most fundamental technique for the process of speech recognition is the use of speech spectrogram. A speech spectrogram is a visual representation of the speech signal. A computer program then attempts to extract the phonemes from the segment of the speech under analysis. If a phoneme is ambiguous, the segment of speech signal that it occupies can be matched against similar utterances that have been recorded and analyzed to "fill it in." The phonemes are then joined together into their corresponding words. This is accomplished in part by a statistical analysis that factors in the probabilities that specific words will crop up in speech, which specific phonemes will crop up in specific words, and that specific word will be surrounded by other specific words.

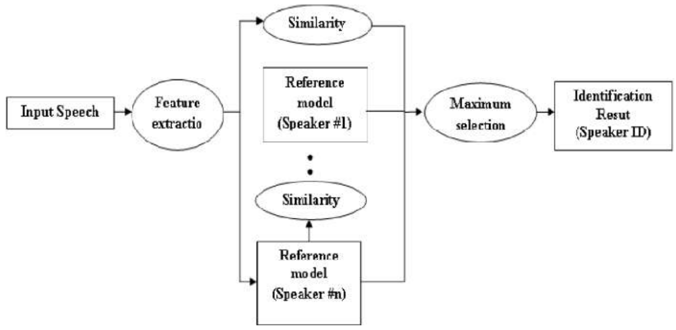
Phoneme-to- words assignment is difficult for two main reasons. The first of these concerns word boundaries. In turns out that there are no pauses in-between words in spoken speech. This makes it hard to tell where one word starts and another ends. To compound the problem, there are often pauses within words. So pauses cannot serve as reliable indicators of word boundaries. The second major issue is phoneme variability. If each phoneme were pronounced clearly and uniformly, speech recognition would be much easier. This is not the case. Speakers vary tremendously with respect to the pitches and durations of phonemes.

Therefore, speech recognition is the ability to identify spoken words, and speaker recognition is the capacity to recognize who is speaking. Speaker recognition refers to the technique of identifying identity of a user based upon his/her voice. The human ear is a relevant organ of human body. Beyond ourunique human ability to receive and decode spoken language, the ear supplies us with the ability to performmany diverse functions. These include, for example, localization of objects, enjoyment of music, and the identification of people by their voices.

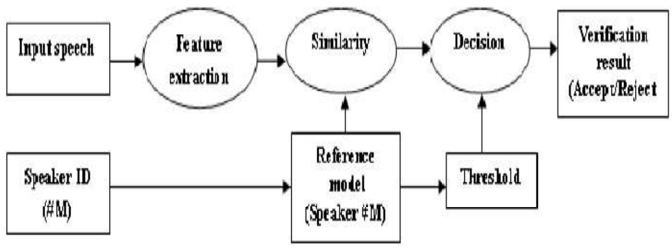
However, the main goal of speaker recognition is to automatically identify a speaker by his/her voice among a population. This technique is being used in voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers etc. Speaker recognition may be classified in two categories speaker identification and speaker verification. Speaker identification is the process of determining the identity of the person based upon his/her voice among a group of persons. Speaker verification is the process of accepting or rejecting the identity claim of the speaker. Mathematically we can say that Speaker verification is 1:N problem while Speaker identification is 1:1 problem. Based on the text to be spoken, speaker recognition methods can also be divided into text-dependent and text-independent speaker recognition. In Text-dependent speaker recognition systems, the speaker is required to produce speech which is same for testing and training, whereas in Text-independent speaker recognition, text in both training and testing may not be same.

Currently, along with efforts to develop computer procedures that understand spoken messages, there is also considerable interest in developing procedures that identify people from their voices (George, 1985). Being able to speak to your personal computer, and have it recognize and understand what you say, would provide a comfortable and natural form of communication. It would reduce the amount of work you have to do leaving your hands free. It would also help in some cases if the computer could tell who was speaking (Richard, Peacocke and Daryl Graf 1990). The schematic presentation of Speaker Identification System and speaker verification system is displayed below (Fig 2-a, b).

Speaker Identification System (Fig 2-a)



Speaker Verification System (Fig 2-b)



The Background of Speaker Recognition System

During the last few decades, the speaker recognition (SR) has made major advances in the area of computer science. In 1962 an article was published in Nature by Bell Laboratories Physicist Lawrence Kerstaentitled, "Voiceprint Identification" (Kersta, 1962). Two years previous, Bell Laboratories had been approached by law enforcement agencies about the possibility of identifying callers who had made verbal bomb threats over the telephone (Lindh, 2004). After two years of research Kersta claimed he had developed a method to identify individuals with high success rates. His method utilized earlier work performed by other Bell Laboratories' scientists, Potter, Kopp and Green who were working on voice identification for military applications during World War II (Potter, Kopp and Green 1947). They had developed a visual representation of speech called a spectrogram. A spectrogram displays the frequency and intensity of a speech signal with respect to time. Kersta's method was an aural-visual method. A spectrogram was inspected visually for pattern matching andscored by an interpreter.

Kersta's research, which produced extremely good results, sparked much research over the next few years. In fact, his article sparked an entire field of research. The first few years following Kersta's publication were intense. There were plenty of researchers with dissenting views. No researcher was able to replicate the incredible results of Kersta's work. To help settle the matter, a research project was undertaken by Tosi, a professor at Michigan State who had doubts about Kersta's "voiceprint". His research was done in conjunction with the Michigan State Police and sponsored by the Federal Department of Justice. When his research was finished, Tosi's work yielded promising results for the emerging field (Tosi, Oyer, Lashbrook, Pedrey, Nicol, and Nash, E. 1972).Tosi's research was not without critics of its own. One year after Tosi's research was published his results were refuted by MIT scientist, Richard Bolt. Bolt's team illustrated holes in Tosi's methodology (Bimbot, et.al., 2004; Bolt, et.al., 1973). The primary criticism was that Tosi's research lacked in practical applications. The FBI, being interested in the forensic application of speaker dentification, requested another study be performed by the National Academy of Sciences. The results from this study showed that the technical uncertainties in forensic applications were substantial enough to claim the use of voiceprints were unreliable in any legal, forensic application.

However, voiceprints are still found useful in certain circumstances. In fact the FBI has utilized a form of Kersta'sspectrogram analysis as late as 2002 (Lindh, 2004). Kersta had not developed 'the solution' to speaker recognition. Today, the success rates with the spectrogram inspection method, given an expert interpreter and proper environmental circumstances, can be very high. But, "the good performance reported in Kersta's paper has not been observed in subsequent evaluations simulating real-life conditions" (Doddington, 1985).

Feature Extraction Techniques

This module converts the speech waveform to various type of parametric illustration. According to the speaker recognition application, feature extraction is the process of retaining necessary information of the speech signal while rejecting redundant and unwanted information this is nothing but analysis of speech signal. Sometimes while removing the unwanted information, we may lose some useful information.
Various techniques used for feature extraction are: Mel-Frequency Cepstrum Coefficients (MFCC), Linear Prediction Coding (LPC), Linear Predictive Cepstral Coefficients (LPCC) and Perceptual Linear Predictive Cepstral Coefficients (PLPC). Here is a review of most commonly used techniques.

Linear Predictive Coding (LPC)

Linear prediction is based on the idea that the current sample is based on the linear combination of past samples. The analysis estimates the values of a discrete-time signal as a linear function of the previous samples.

The spectral envelope is represented in a compressed form, using the information of the linear predictive model. This can be mathematically represented as −

$$S(n) = \sum_{k=1}^{p} \alpha_k S(n-k)$$

Whereas(n) is the current speech sample
k is a particular sample
p is the most recent value
α_k is the predictor co-efficient
s(n - k) is the previous speech sample
For LPC, the predictor co-efficient values are determined by minimizing the sum of squared differences (over a finite interval) between the actual speech samples and the linearly predicted ones.

$$E(n) = S(n) - S(n')$$

Mel Frequency Cepstrum Coefficient (MFCC):

MFCC is a feature extraction technique widely used in automatic speech and speaker recognition. It was introduced by Davis and Mermelstein in the 1980's, and has been state-of-the-art ever since. Prior to the introduction of MFCCs, Linear Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficients (LPCCs) were the main feature types for automatic speech recognition (ASR), especially with HMM classifiers. To implement them following steps should be followed:

1. Framing and Windowing: An audio signal is constantly changing, so to simplify things we assume that on short time scales the audio signal doesn't change much. This is why we frame the signal into 20-40ms frames.
2. Applying FFT (Fast Fourier Transform) to convert speech signal from time domain to frequency domain.
3. Taking its absolute value.
4. Mel scaled filter bank: Incorporating this scale makes our features match more closely what humans hear.
$$mel(f) = 2595 * log_{10}\,(1 + \frac{f}{700})$$
5. Obtaining Discrete cosine transform (DCT). DCT transforms the frequency domain into a time-like domain called quefrency domain. These features are referred to as the mel-scale cepstral coefficients.

Perceptual Linear Predictive Cepstral Coefficients (PLPCC):
It is based on the magnitudespectrum of the speech analysis window. Unlike MFCC and LPC which are cepstral methods, the PLPCC is a temporal method and models the speech auditory spectrum through a low order all pole model (Revathi, Revathi and Venkataraman 2009) details the steps followed to calculate the coefficients of the PLPCC. First, compute the power spectrum of a windowed speech. Second, group the results to 23 critical bands using bark scaling for sampling frequency of 8 kHz. Third, perform loudness equalization and cube root compression to simulate the power law of hearing. Fourth, perform inverse Fast Fourier Transform (IFFT). Fifth, perform LP analysis by Levinson- Durbin algorithm (Delsarte and Genin, 1986). Lastly, convert LPco efficients into cepstral coefficients. The relationship between frequency in Bark and frequency specified as in f(bark) = 6*arcsinh (f(Hz)/600).

Speaker Modelling

During enrolment, speech from a speaker is passed through the feature extraction module and the feature vectors are used to create a speaker model. Desirable attributes of a speaker model are: (1) a theoretical underpinning so one can understand model behavior and mathematically approach extensions and improvements; (2) generalizable to new data so that the model does not over fit the enrolment data and can match new data; (3) parsimonious representation in both size and computation.

There are many modelling techniques that have some or all of these attributes and have been used in speaker verification systems. The selection of modelling is largely dependent on the type of speech to be used, the expected performance, the ease of training and updating, and storage and computation considerations. A few techniques of modelling are briefly discussed in the following segment.

Template Matching

In this technique, the model consists of a template that is a sequence of feature vectors from a fixed phrase. During verification a match score is produced by using dynamic time warping (DTW) to align and measure the similarity between the test phrase and the speaker template. This approach is used almost exclusively for text-dependent applications.

Nearest Neighbour

In this technique, no explicit model is used; instead all features vectors from the enrollment speech are retained to represent the speaker. During verification, the match score is computed as the cumulated distance of each test feature vector to its k nearest neighbours in the speaker's training vectors. To limit storage and computation, feature vector pruning techniques are usually applied.

Neural Networks

The particular model used in this technique can have many forms, such as multi-layer perceptions or radial basis functions. The main difference with the other approaches described is that these models are explicitly trained to discriminate between the speakers being modeled and some alternative speakers. Training can be computationally expensive and models are sometimes not generalizable.

Hidden Markov Models

This technique uses HMMs, which encode the temporal evolution of the features and efficiently model statistical variation of the features, to provide a statistical representation of how a speaker produces sounds. During enrollment HMM parameters are estimated from the speech using established automatic algorithms. During verification, the likelihood of the test feature sequence is computed against the speaker's HMMs. For text-dependent applications, whole phrases or phonemes may be modeled using multi-state left-to right HMMs. For text-independent applications, single state HMMs, also known as Gaussian Mixture Models (GMMs), are used.

Strengths and Weaknesses of Speaker Recognition

Speaker Recognition has its various strengths and weaknesses and the following criteria are used to evaluate the suitability of speaker recognition as biometrics.

Collectability: The major advantage of voice recognition is that it can be implemented using telephone lines or computer microphones, with varying recordings and transmission quality. Pattern matching algorithms must be able to handle different quality of the recordings and noises.

Portability: Speaker verification is easy to use, has low computation requirements (can be ported to cards and handhelds) and, given appropriate constraints, has high accuracy.

Acceptability: Speaking is a natural process so no abnormal actions are required. When speaker recognition is used for surveillance purpose or in general when the subject is not aware of it then the common privacy concerns of identifying unaware subjects apply. Moreover speaker information can be obtained easily from almostany where using the familiar telephone network (orinternet) with no special user equipment or training.

Circumvention: A major concern with speaker recognition is spoofing. The risk of spoofing with voice recordings can be lessen if the system requests a randomly generated phrase to be repeated so that any impostor cannot anticipate the random phrase that will be required and therefore cannot attempt a playback spoofing attack.

Performance: Robustness depends a lot on the setup. When telephone lines or computer microphones are used, the algorithms must compensate for noise and issues with room acoustics. Furthermore speaker recognition is, because the voice is a behaviouralbiometric, impacted by errors of the individual such as misreading and mispronunciations.

Mobility: Mobility of system means that people are using verification systems from more uncontrolled and harsh acoustic environments (cars, crowded airports), which can stress accuracy.

Variability: The varied microphones and channels that people use can cause difficulties since most speaker verification systems rely on low-level spectrum features susceptible to transducer/channel effects.

Universality: Obviously for people who are mute or having problems with their voice due to some illness, this biometric solution is not useable.

Permanence: Speech signal used for recognition is a behavioural signal that may not be consistently reproduced by a speaker because voice is influenced by various factors such as sickness, stress, tiredness, ageing etc.

CONCLUSION

In this article an effort has been made to present interdisciplinary nature and characteristics of cognitive science and its association with speaker recognition. Apart from this, we have presented an overview of speaker recognition system which includes various methods of feature extraction and feature modelling. Focus is also given on the factors affecting the system, as well as its strength and weaknesses are discussed. The speaker recognition system still has various drawbacks which can be further reduced by carrying out research in sub-domains and merging of other biometrics with speaker recognition. The main application for the technology is in the area of access control, where the speakers are required to be authenticated before they can be allowed to access certain facilities or some other restricted services in various domains which is having some secured information.

The future trends in access controlis to integrate speaker verification technology into amulti-level and a hybrid authentication approach, where results from different biometric technology like fingerprint, face, iris and speaker recognition could be fused together to achieve better reliability in authentication. However, the biggest advantage of speech based biometrics is the ability perform authentication where a direct physical or visual contact with the subject is not feasible. Thus, the technology has a clear advantage for authenticating transactions that occur over the voice channel like telebanking. A more controversial application of speaker verification technology is in the area of forensics where the results of the technique could be offered as evidence in judicial trials. Compared to finger-printing and DNA based authentication technology, the existing speaker verification techniques have their drawbacks and limitations due to their sensitivity to corruption by noise and the ability to masquerade the signal using voice recording devices. Despite this, there is an enormous potential for speaker recognition/ verification and recognition technology in multimedia and biometric applications. However, key challenges still remain to be solved and are currently limiting the wide-scale deployment of the technology. The research field is still in its infancy stage and massive efforts are needed for comprehensive understanding and applications of speaker recognition and its link with cognitive science.

References

Bimbot, F., Bonastre, J., Fredouille, C., Gravier, G., Chagnoleau, I., Meignier, S., Merlin, T., Ortega-Garcia, J., Petrovska-Delacretaz, D., Reynolds, D. 2004.A Tutorial on Text-Independent Speaker Verification, EURASIP. *Journal on Applied Signal Processing*, 4, 431-450.

Bolt, R. Cooper, F., David Jr. E., Denes, P., Pickett, J., and Stevens, K. 1973. Speaker Identification by Speech Spectrograms: Some Further Observations. *Journal of the Acoustical Society of America*, 54, 53-534.

Carpenter, P. A., and Just, M. A. 1989.The role of working memory in language comprehension. In D. Klahr& K. Kotovsky (Eds.), Complex information processing: The impact of Herbert A. Simon, 31-68. Hillsdale, NJ: Erlbaum.

Cawsey, A.1993.User Modeling in Interactive Explanations. *User Modeling and User-Adapted Interaction* 3(3), 221-248.

Chakraborty, P., Ahmed, F., MonirulKabir,Md., Shahjahan1, Md., and Kazuyuki Murase. 2008. “An Automatic Speaker Recognition System”, M. Ishikawa et al. (Eds.): ICONIP 2007, Part I, LNCS 4984, 517–526.

Chomsky, N. 1959.A review of B.F. Skinner’s Verbal Behaviour, Language, 25, 26-58.

Delsarte, P. and Genin, Y. 1986. The split Levinsonalgorithm," Acoustics, Speech and Signal Processing. *IEEE Transactions on*, 34, 470- 478.

Doddington, G. 1985. Speaker Recognition – Identifying People By Their Voice. *Proceedings of IEEE*, 73, 1651-1664.

Eysenck, H. J. 1984. Is behavior therapy on course? *Behavioral Psychotherapy*, 12(1), 2-6.

Franklin, S. 1995. Artificial minds. Cambridge, MA : MIT Press.

Gardner, R. C. 1985. Social psychology and second language learning: The role of attitudes and motivation. London: Edward Arnold Publishers.

George, R. Doddington, Member, IEEE. 1985. “Speaker Recognition- Identifying People by their Voices. *Proceedings of the IEEE*, 73(11), 1651-1664.

Guenther, F. H. 1995. Speech sound acquisition, coarticulation, and rate effects ina neural network model of speech production. *Psychological Review* 102, 594–621.

Jose Ramon Calvo de Lara. 2005. A Method of AutomaticSpeaker Recognition Using Cepstral Features and Vectorial Quantization, M. Lazo and A. Sanfeliu (Eds.): CIARP LNCS, 3773, 146 – 153.

Kersta, L. 1962.“Voiceprint Identification”, Nature Magazine, 196, 1253.

Lachman.R., Lachman, J. L, and Butterfield, E. C. 1979. Cognitive Psychology and Information Processing: An Introduction. Hillsdale, NJ: Lawrence Erlbaum.

Lewandowsky, S. 1993. The rewards and hazards of computer simulations. *Psychological Science*, 4, 236-243.

Lindh, J. 2004. “Handling the Voiceprint Issue”, FONETICK Proceedings.

Lugar, G.F. 1994. Cognitive Science: The science of intelligent systems. San Diego: Academic Press.

Marr, J.W.S. 1962. The natural history and geography of the Antarctic krill (*Euphausiasuperba* Dana).*Discovery Rep* 32, 33–464.

Marshall, J. C. 1977. Minds, machines and metaphors. *Social Studies of Science* **7**:475–88.

Matline, M.W. 1989. Cognition. (2nd ed.). New York: Holt, Rinehart and Winston.

Neisser, U. 1963.Decision time without reaction time; Experiments in visual scanning. *American Journal of Psychology*, 76, 376-385.

Pandey, S. 2014. Cognitive science and Psychology: Some Reflections. *Indian Journal of Psychological Issues*, 22, 8-14.

Potter, R., Kopp, G., Green, H. 1947.Technical Aspects of VisualSpeech”, Bell Labs, New York.

Revathi, A. and Venkataramani,Y. 2009. Text Independent Composite Speaker Identification/Verification Using Multiple Features, in Computer Science and Information Engineering, WRI World Congress on, 257-261.

Richard D. Peacocke, and Daryl H. Graf. 1990. AnIntroduction to Speech and Speaker Recognition. *IEEE*, 26-33.

Shannon, C. E and Weaver, W. 1949.The Mathematical Theory of Communication. Urbana, IL: University of IIIinois Press.

Tosi, O. Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., Nash, E. 1972. Experiment on Voice Identification.*Journal of the Acoustical Society of America*, 51, 2030-2043.

Thrun, S., et al. 1998. Map learning and high speed navigation in RHINO. In D. Kortenkamp, R.P. Bonasso and R. Murphy (Eds), Artificial Intelligence and mobile rebots, 21-52. Cambridge, MA: MIT Press.

Wagman, M. 1999. The human mind according to artificial intelligence: Theory, research and implications. Westport, CT: Praeger.

****