

# 3

## SPEECH RECOGNITION AS AN INPUT DEVICE FOR EFFECTIVE COMMUNICATION IN INFORMATION ACCESS: A CONCEPTUAL APPROACH

Goshie Rhoda **WUSA**<sup>1</sup>, Daniel **MUTTEE**<sup>2</sup>, & Caroline **MUTWIRI**<sup>3</sup>  
Federal University of Technology Library, Minna Niger State, Nigeria<sup>1</sup>, Department of Library and Information Science, Kenyatta University, Nairobi, Kenya<sup>2,3</sup>  
goshie wusa@students.ku.ac.ke<sup>1</sup>, muthee.daniel@ku.ac.ke<sup>2</sup>,  
mutwiri.caroline@ku.ac.ke<sup>3</sup>

### Abstract

**Purpose:** The purpose of this paper is to evaluate the speech recognition as an input device for effective communication in information access. Speech Recognition has been one of the most exciting research areas in this computerized world. In order to establish communication between the humans and the machine, the technology 'Speech Recognition' is essential with the help of technology users can easily control devices and create documents by speaking.

**Design/Methodology/Approach:** The paper adopted conceptual approach. It determines the concept of speech recognition, voice input device, usage of speech recognition, usages in education and daily life, people with disabilities, evaluation of speech recognition for effective communication, related works on speech recognitions, key features of effective speech recognition for communication, advantages and benefits of speech recognitions, challenges and recommendations of speech recognitions.

**Implication:** Libraries should find out its place and roles in the usage of speech recognition software for individuals with special needs to help enhance a close captioning of conversations such as discussions in conference rooms and classroom lectures for effective communication and easy access to information

**Originality/Value:** The use of speech recognition software, in conjunction with a digital audio recorder and a personal computer running word-processing software has proven to be positive for restoring damaged short-term memory capacity, in stroke and craniotomy individuals for easy communication and access to information.

**Keywords:** Speech Recognition, Speech Recognition Usage, Mobile Devices, Automatic Speech Recognition, Speech Processing and Voice Recognition.

Paper Type: Conceptual research

### Introduction

Since ages, speech has been an important means of communication between humans. Speech Recognition is the process of converting an acoustic speech into text, and / or identification of the speaker. Over the years with recent advent in technology it has become an essential and integral part of our lifestyle due to the increasing communication between human and computers or automated systems. A system built at Bell Laboratory in 1952 which was the first word recognition system which was trained to recognize digits. Some of the widely used speech recognition systems are speaker dependent systems, speaker

independent system, isolated word recognizer, connected word recognizer and spontaneous recognition system. (Zheng, T.& Li, L. (2017).

### Speech Recognition

Speech recognition can be defined as a process through which a computer system acquires a speech signal produced by a human speaker, possibly distorted by noise in the environment, and tries to re-construct the utterance produced by the speaker. Humans are excellent at understanding other human speakers; for machines, the problem is harder because they do not (usually) have access to or understand clues such as facial expressions, gestures, and discourse context people use to extract

meaning from a voice signal While speech recognition is commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voice. (Paul, R. & Saini, R. 2018)

**Voice input device:** Voice input device is a device in which speech is used to input data or system commands directly into a system. Such equipment involves the use of speech recognition processes, and can replace or supplement other input devices. Some voice input devices can recognize spoken words from a predefined vocabulary, some have to be trained for a particular speaker. When the operator utters a vocabulary item, the matching data input is displayed as characters on a screen and can then be verified by the operator. The speech recognition process depends on the comparison of each utterance with words appearing in a stored vocabulary table. (Deng, L & Dong, M. 2014)

A speech sound input device for providing information to a computer in a system of computer recognition of speech comprising (i) a voiced monitor which provides to the computer a signal indicating whether the sound input to the device has a fundamental repetitive structure, and signals indicating the periodicity of any such structure; (ii) an unvoiced monitor which provides to the computer a signal indicating whether the sound input to the device has no low frequency component and has at the same time high frequency components in a wide range of frequency. (Kumar, M. & Shimi, S. L. 2015).

### **Usage of Speech Recognition**

Speech Recognition has been most commonly used in following sectors: healthcare centre, military organizations, telecommunication, robotics and so on. In order to establish communication between the humans and the machine, the technology 'Speech Recognition' is essential. Many researchers' have more interest on this area from various fields. Some examples where this speech recognition might encounter are: Automated Phone Systems, Google Voice and AppleSiri (Apple Phone). Speech

Recognition has to be carried out in various kinds such as: Supervised Speech Recognition, Unsupervised Speech Recognition, and Isolated Speech Recognition. (Prabhu, U & Liao, H 2018).

### **Usage in Education and Daily Life**

For language learning, speech recognition can be useful for learning a second language. It can teach proper pronunciation, in addition to helping a person develop fluency with their speaking skills. Students who are blind (see Blindness and education) or have very low vision can benefit from using the technology to convey words and then hear the computer recite them, as well as use a computer by commanding with their voice, instead of having to look at the screen and keyboard. The use of voice recognition software, in conjunction with a digital audio recorder and a personal computer running word-processing software has proven to be positive for restoring damaged short-term memory capacity, in stroke and craniotomy individuals. (Saksamudre, S. K & Shrisrimal, 2015).

### **People with Disabilities**

People with disabilities can benefit from speech recognition programs. For individuals that are deaf or hard of hearing, speech recognition software is used to automatically generate a closed-captioning of conversations such as discussions in conference rooms, classroom lectures, and/or religious services. Speech recognition is also very useful for people who have difficulty using their hands, ranging from mild repetitive stress injuries to involve disabilities that preclude using conventional computer input devices. This type of technology can help those with dyslexia but other disabilities are still in question. (Kaur, G. & Vastava, M, 2018)

### **Evaluation of Speech Recognition for Effective Communication**

The performance of speech recognition systems is usually evaluated in terms of accuracy and speed. Accuracy is usually rated with word error rate (WER), whereas speed is measured with the real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR). Speech recognition by machine is a very complex

problem, however vocalizations vary in terms of accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed. (Karpagavalli, S & Chandra, E, 2016)

#### **Related Works on Speech Recognition**

(Suman, et al, 2017) presented the basic idea of speech recognition, proposed types of speech recognition, issues in speech recognition, different useful approaches for feature extraction of the speech signal with its advantage and disadvantage and various pattern matching approaches for recognizing the speech of the different speaker. Many researchers are interested in this area that supports speaker independent operations and continuous speech in different language. (Karpagavalli et al, 2016) studied automatic speech recognition developed by an architecture to support the communication between human and machines. Performance of speech recognition applications deteriorated in the presence of reverberation and even low levels of ambient noise. Robustness to noise, reverberation and characteristics of the transducer is still an unsolved problem that made the research in the area of speech recognition still very active.

(Preeti et al, 2021) presented a study of basic approaches to speech recognition and the results showed better accuracy. The paper also presented what research has been done around for dealing with the problem of ASR. The design of speech recognition system require careful attentions to the challenges or issue such as various types of speech classes, speech representation, feature extraction techniques, database and performance evaluation.

(Santosh et al, 2020) gave an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gave overview technique developed in each stage of speech recognition. The paper helped in choosing the technique along with their relative merits and demerits. It also included with the decision on feature direction for developing technique in human computer interface system using Marathi Language.

#### **Key Features of Effective Speech Recognition for Communication**

Many speech recognition applications and devices are available, but the more advanced solutions use AI and machine learning. They integrate grammar, syntax, structure, and composition of audio and voice signals to understand and process human speech. Ideally, they learn as they go evolving responses with each interaction. Companies, like IBM, are making inroads in several areas, to improve human and machine interaction. (Vyas, M. 2013)

Neural networks learn this mapping function through supervised learning, adjusting based on the loss function through the process of gradient descent. While neural networks tend to be more accurate and can accept more data, this comes at a performance efficiency cost as they tend to be slower to train compared to traditional language models. Speaker Diarization (SD): Speaker diarization algorithms identify and segment speech by speaker identity. This helps programs better distinguish individuals in a conversation and is frequently applied at call centers distinguishing customers and sales agents. (Yuliu, X & Zhen, X, 2021)

#### **Advantages of Speech Recognition**

Speech recognition technology has become an increasingly popular concept in recent years. From organizations to individuals, the technology is widely used for various advantages it provides. One of the most notable advantages of speech recognition technology includes the dictation ability it provides. With the help of technology, users can easily control devices and create documents by speaking. Speech recognition allows documents to be created faster because the software generally produces words as quickly as they are uttered, which is usually much faster than a person can type. Dictation solutions are not only used by individuals but also by organizations that require massive transcription tasks such as healthcare and legal. (IBM, 2020)

With the help of the voice recognition technology, callers can input information such as name, account number, the reason for their call, etc. without interacting with a live agent.

Instead of having callers remain idly on hold while agents are busy, organizations can engage their callers without live customer representatives. Speech recognition technology; delivers a great customer experience while improving the self-service system's containment rate encourages natural, human-like conversations that create more satisfying self-service interactions with customers, automates what touchtone cannot by collecting dynamic data such as names and addresses, enables organizations to save agents for more critical tasks.(Chen,Q& Huang, D, 2016).

The technology; transforms complicated IVR menus to easy-to-use systems, it is compatible with all major operating systems and speech standards, improves security by eliminating agents from data processing, provides an alternative for the customers who prefer to skip touchtone menus and connect to agents, guarantees better customer service experience increased customer satisfaction and decreased operational costs.(Dua, M & Aggarwal, R. K, 2019)

### **Benefits of Speech Recognitions**

Whether a client invests in their own speech recognition software for in-house use or turns to a third party that provides it, there are significant benefits to using this type of service. Those benefits may include: It's faster. Since most people speak faster than they write, speech recognition software provides a simple way to get words into a document without having to be delayed in the process. This speed is what makes many people seek out its use. Typing, on the other hand, can slow down the communication process. It is fairly accurate. Although it should always be proofread, speech recognition software can result in a document more or less free of errors (Zhang,S. &Sun,S, 2021).

In addition, newer programs tend to be well designed and can offer reliable results for some applications, It allows for hands-free work. When working with a client or completing a task, the use of speech recognition tools facilitates easy note taking, use of other materials, and professional eye contact. Each of these activities is limited when someone has to type information into a computer behind a

screen. While these benefits are compelling, there are also some limitations that users have to take into consideration as well (Zorila, C.& Li, M.H.2020).

### **The Challenges Speech Recognition**

As with any automated tool, there are limits to how well a product like this can work. Consider these typical concerns when using speech recognition tools. Accuracy is always imperfect. "More or less accurate" is not perfectly accurate. A client's professional reputation hangs in part on how well they present themselves in writing; the software cannot understand complexities of jargon or phrasing. Every business sector has its own vocabulary and idioms, and if these are not part of standard English, the tool may not be successful in translating them to written format, Some voices do not come across well.

Speech recognition software may not be able to transliterate the words of those who speak quickly, run words together or have an accent. It also drops in accuracy when more than one speaker is present and being recorded, Noise: voice recording machines detect sound waves that are generated through speech. Background noises in rooms make it hard for systems to understand and distinguish between the specific sound waves from the host voice process. (Rybach, D. & Parada, C. 2016).These are the barriers highlighted by respondents to a survey as part of the Speechmatics report on Trends and Predictions for Voice Technology in 2021.( Chan, W. &Jaitly, N. 2016)

### **Conclusion**

Speech is the primary and the most convenient means of communication between people. Speech recognition has now become one of the most emerging technologies in order to enhance human-machine interaction and effective communication. This research work also has some advantages and challenges on this technology. In this research work, we also highlighted the benefit of speech recognition technology to how to handle the devices from our mobile devices from anywhere.

## Recommendations for Successful Speech Recognition for Effective Communication

The followings are the Recommendations for this study.

- On-premises deployment of voice technology should be used to enable users to keep their data secure within their own environments with no need for data to go into the cloud.
- Dark site environments- : Offline licensing should be supported in dark site deployments, meaning all work should be completed within an organization's private environment. This delivers a more robust solution for compliance and data privacy needs.
- **Cloud deployment:** Private cloud deployments should be secure enough to keep data safe for lots of applications. If cloud deployment security is good enough for the business and use case needs, cloud deployment is often the preferred option due to low operational cost and less complexity.

## References

- Chan, W. & Jaitly, N. (2016). *Computer Vision and Pattern Recognition (CVPR)*. pp. 3444–3453. ISBN 978-1-5386-0457-1"
- Dua, M. & Aggarwal, R.K (2019) GFCC BASED DISCRIMINATIVELY TRAINED NOISE ROBUST CONTINUOUS ASR SYSTEM FOR HINDI LANGUAGE *Journal of Ambient Intelligence and Humanized Computing*, 10 (6), 2301-2314
- Chen, Q. & Huang, D. (2016). "DepAudio Net: an efficient deep model for audio based depression classification," in *AVEC '16: Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge* (New York, NY), 35–42.
- Deng, L & Dong, H (2014). "Deep Learning: Methods and Applications" (PDF). *Foundations and Trends in Signal Processing*. 7 (3–4): 197–387.
- IBM cloud Education (2020) <https://www.ibm.com/cloud/learn/speech-recognition>
- Kaur, G., Srivastava, M., & Kumar, A. (2018). Integrated speaker and speech recognition for wheel chair movement using artificial intelligence. *Informatica*, 42, 587–594
- Kumar, M. & Shimi S.L.(2015) Voice Recognition Based Home Automation System for Paralyzed People. *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)* 4 (10),
- Karpagavalli, S. & Chandra, E. (2016) A Review on Automatic Speech Recognition Architecture and Approaches. *International Journal of Signal Processing, Image Processing and Pattern Recognition* 9 (4) 394-398.
- Paul, R. & Saini, R. (2018) "A Review on Speech Recognition Methods", *International Journal on Future Revolution in Computer Science & Communication Engineering*: 4 (2), 292 –298.
- Prabhu, U. & Liao, H. (2018). "Large-Scale Visual Speech Recognition". *arXiv:1807.05162 [cs.CV]*.000000
- Preeti, C. (2021) A deep learning Survey on Content Detection from Images Powered by Computer vision framework. Proceedings of the third International Conference on Innovative Research in Computing Application. ICIRCA. Pp. 1-172, 2(67)
- Rybach, D. & Parada, C. (2016). Personalized speech recognition on mobile devices. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5955-5959.
- Srivastava, N. (2014) "Speech Recognition using Artificial Neural Network", *IJESIT*, 3(3)
- Saini, P. & Kaur, P. (2013) Automatic Speech Recognition: A Review. *International Journal of Engineering Trends and Technology* 4 (2), 133-135.
- Santosh, K.C. (2020) A Lazy Learning- based Language Identification from Speech Using MFCC-Z Features. *International Journal of Machine Learning and Cybernetics*. 11, 1-14
- Speech Analysis Extraction Reorganization Recognize Words (2017) ISSN 1943-023X 98 *Journal of Advanced Research in Dynamical & Control Systems*, Vol. 9, No. 6.
- Suman, M. (2017) Analysis of Emotion Recognition System of Telugu using Prosodic and format features. *Journal of Speech and Language Processing for Human-Machine Communications*. Pp. 137-144, Vol.664
- Saksamudre, S.K., Shrishrimal, P.P. (2015) A Review on Different Approaches for Speech Recognition System. *International Journal of Computer Applications* 115 (22) 23-26.

- Swamy, S. & Ramakrishnan, K.V. (2013) An Efficient Speech Recognition System. *International Journal of Computer Science and Engineering* 3 (4) 23-24.
- Vyas (2013) "A Gaussian Mixture Model Based Speech Recognition System Using Matlab", SIPIJ, Vol.4, No. 4.
- Yu Liu, X. & Zhen, X. (2021) Depression Speech Recognition with a Three-Dimensional Convolutional Network Front. Hum. Neurosis. 30 September 2021 <https://doi.org/10.3389/fnhum.2021.713823>
- Zorilă, C., & Li, M., Hayakawa, (2020) Toshiba's Speech Recognition System for the CHiME 2020 Challenge. Proc. The 6th International Workshop on Speech Processing in Everyday Environments (CHiME 2020), 62-66.
- Zhang, S. & Sun, S. (2021). Depression detection on reddit with an emotion-based attention network: algorithm development and validation. *JMIR Med. Inform.* 9:e28754.
- Zheng, T. & Li, L. (2017). Robustness-Related Issues in Speaker Recognition. Springer Briefs in Electrical and Computer Engineering. Singapore: Springer Singapore.