International Journal of Research in IT and Management (IJRIM)
Available online at: http://euroasiapub.org
Vol. 11 Issue 7, July- 2021
ISSN(o): 2231-4334 | ISSN(p): 2349-6517 | Impact Factor: 8.106

AUTOMATIC SPEECH RECOGNITION FOR THE HEARING IMPAIRED USING MFCC FOR DEAF USERS

D Rajkumar, Research Scholar, Dept. of Computer Science, Himalayan University

Dr. Manish Saxena, Professor, Dept. of Computer Science, Himalayan University

ABSTRACT

The extraction of a unique feature set is crucial for creating a pattern recognition program. In order to distinguish between the various talks, a recognition system also needs to extract a discriminating feature set. As more complicated characteristics are being employed in proposed work on speech recognition, some specific features used in various works may differ greatly. The classification of ASR systems into low, medium, and high dataset speech recognition systems is based on the vocabulary employed. Biometric speech recognition systems, voice to text conversion, audio conferencing, emotion recognition, robotics, synthetic speech recognition, the educational sector, etc. are just a few of the many uses for ASR systems. The majority of ASR systems are trained using clean data; however, due to noise addition in the testing signal, such a system performs poorly in real-time speech recognition. Strong voice recognition is therefore required in noisy environments. Techniques for noise reduction or noise isolation have often been utilized to address such a situation. However, noise reduction may damage the signal content if the noise cannot be identified. Additionally, the pre-processing time required for noise reduction is longer. A tremendous amount of research has been done over the years using deep learning techniques based on artificial intelligence for speech recognition applications. In order to increase the classification accuracy of speech processing, the automated speech recognition (ASR) faces issues primarily in the preprocessing, feature extraction, and classification phases. Mel-frequency Cepstral coefficients (MFCCs) and linear predictive coefficients (LPCs) feature extraction of speech signals have been combined with the Spectral Subtraction (SS) method of denoising to address these challenges. This has resulted in the development of an enhanced speech recognition methodology. The classification of speech signals for ASR is then done using back propagating artificial neural networks (BP-ANN). The simulation results demonstrate that, when compared to current ASR algorithms, the suggested approach provides a higher level of classification accuracy.

KEYWORDS: Emotion Recognition, Speech Signals, Robotics, Simulation

1. INTRODUCTION

Speech is the most common type of communication that people use to express themselves. Information of this kind can be transferred in a variety of slangs and with a variety of persons. One of the most unrestrained emotions features in the sophisticated technology world's research domain was speech identification.

This must be carried out in order to acknowledge and recognize human speech and move from one location to other using controls to run the devices. This program allows users to input spoken expressions and vocabulary as strings of words as input.

The system can be a computer. The package for this gadget should be able to decrypt human sound frequency and output some kind of action.

Implementing this voice recognition technology with an effective approach and algorithm is the goal of my research. In order to create a framework that utilizes the voice recognition technology in the various fields covered in this dissertation, we must first suggest a methodology.

The goal of the job is to suggest a method for creating a framework or algorithm that offers a practical way to integrate speech recognition technology into query processing and in a cloud-based environment. A person with vision impairments has found this technology to be quite helpful in accessing the info.

To provide multiple techniques prepared to implement the technology of speech recognition in different fields, all of which have been empirically proven.

1.1 EFFECT OF HEARING LOSS ON SPEECH PERCEPTION

Speech perception refers to how language sounds are perceived, interpreted, and understood. It is a broad field that has connections to phonetics, phonology, cognitive psychology, and hearing psychology. One of the main things that influence how well people perceive speech is hearing loss. Any difficulty in hearing an approaching sound is considered hearing loss. Among the several types of hearing loss, sensorineural hearing loss damages the auditory system permanently and impairs speech perception.

The 20 Hz to 20,000 Hz frequency range is the range of normal human hearing. People with hearing loss experience problems with this wide range of audibility. High frequencies are more damaged than low frequencies in the majority of people with sensory neural hearing loss because the basal region is more susceptible to damage. As a result, they fail to pick up on crucial key bands that are essential for understanding speech. The perception of speech is negatively impacted by this.

People with hearing loss frequently say that they can hear sounds, but they are unable to distinguish or recognize the sounds. It's crucial to be able to tell the two signals apart in order to recognize speech sounds. Just Noticeable Difference (JND) is the size needed to distinguish between the two signals. According to Turner and Nelson (1982), individuals with sensory neural hearing loss have greater JNDs, meaning their frequency resolution is significantly worse than that of people with normal hearing. This is yet another element that influences how people perceive speech in general.

1.2 NEED OF THE STUDY

Due to its diverse cultural legacy and linguistic ethnicities, India has a great impact on speech and language development, which has a big impact on communication. For the testing and treatment of people with communication impairments, speech and hearing experts have been employing modified versions of western materials. Given the stark disparities in the linguistic and cultural systems, this would be improper. As a result of recent efforts to combat this, test materials, screening diagnostic tools, and therapy materials pertinent to the Indian setting have been developed. India contains more than 30 official languages in addition to a large number of other widely spoken fluent languages, according to the Census of India 2001. A situation of this magnitude demands careful thought when developing language-specific content for our people.

We are currently living in a technological age where users of all ages use software for a variety of communication-based applications. Software-based programs are increasingly more of a requirement than a need in our line of work. Computer-based programs offer advantages such as simplicity of use, portability, data storing, and structure that make them straightforward and practical to employ for both diagnostic and therapeutic purposes. Software for screening, diagnosis, and therapy has been successfully developed by SLPs (Speech Language Pathologists) in India. Comparatively, this is obviously less in the field of audiology and more so in the area of auditory rehabilitation, where the pediatric population has always received more attention. Adult auditory rehabilitation is becoming more and more important in our nation. As a result of recent improvements in medical technology, people are living longer and are more willing to choose hearing aids in order to retain a higher quality of life. This is supported by the technology's quick development in both signal processing and aesthetics. These factors taken together result in a rise in hearing aid usage among this demographic. There is never a situation when older persons with hearing loss are satisfied, even with the best fitting of these hearing aids. Therefore, hearing aid fitting should be followed by training in order to achieve best practices in the field of hearing aids.

A concentrated group discussion (expert survey) was conducted to help determine the need for and approach to the study. This qualitative data collection approach focused on 15 people with more than two years of expertise in the field of auditory rehabilitation. In order to spark a discussion about auditory rehabilitation in older persons with hearing impairment, a series of 10 questions was put out. The study group members concurred that adults needed more training to improve their communication skills rather than just having an amplifying device fitted to them. The panel came to the unanimous conclusion that it was essential to create an auditory rehabilitation module that would take into account the differences in Indian population's culture, language, and dialects.

All of these call for the creation of a systematic program for auditory rehabilitation in order to supplement the advantages of amplification for older people with hearing loss. The focus of the current study is specifically on Kannada-speaking people who live in Karnataka State, India, where 66.9% of the older population has been found to have hearing impairment.

2. REVIEW OF LITERATURE

Regarding the enormous assurance planning of the Social Internet of Things (SIoT), Zhiyuan Li et al. (2015) were cited. They considered the trend equivalency that isn't addressed by the current request frameworks and is put up in intellectual coordinates without proper consideration of the actual planning requirements for the framework.

According to ShwetaDoda et al. (2014), there is a lot of interest in discourse affirmation from many different domains. To understand kept words, techniques are required. They distilled numerous noteworthy practices from various eras of the discourse affirmation system, such as Dynamic Time Warp and Hidden Andrei Markov Model approaches for isolated discourse affirmation.

According to Gheorge et al. (2014), a brief analysis of changed discourse affirmation structures highlighted the existing successes and limitations of present-day plans, as well as their inherent controls and flaws. They established an enhanced point of view associated algorithmic standard for a structure assistant modified talk confirmation framework that suitably alters its affirmation model in an unattended manner by monitoring continuous discourse.

A massive language affirmation system with low turbidity, accuracy, and just enough memory and compute power to run faster than continuous on a Nexus 5 Android phone was developed by Rohit et al. (2016). To accurately visualize phone targets and relate them to any reduction in their weight, they used a live long fundamental scholarly system (LSTM) acoustic model set up with a connectionist transient plan (CTC).

According to Shikha Gupta et al. (2014), as the requirements for embedded handling and interest in increasing introduced phases increase, it is essential that the discourse affirmation structures be open on them as well. PDAs and other hand-held devices are continuing to get uglier and worse. Running transmission on these frameworks has become possible. Discourse affirmation structure was examined in that paper along with several ways. In a similar vein, it displayed the overview of the strategy's features for feature extraction and feature planning. Through this review paper, it was discovered that VQ is superior to DTW for feature extraction and that MFCC is frequently employed.

3. RESEARCH METHODOLOGY

The field of speech recognition has long been recognized as crucial to human computer interaction. A voice recognition system uses methods from a wide range of fields, including statistical sequence identification, communication theory, signal perception, and linguistics. The fundamental or main objective of research in voice processing is the human-machine interface. Interdisciplinarity and the ability to address scientific problems are two of the most important goals of research into speech signal detection using machines. Hindi is our mother tongue, thus its local application is a significant and motivating factor in the choice of language for our appreciation framework, for instance. Real-time communication recognizers can be used extensively in a variety of settings, including simpler man-appliance signal, assistance for physically disabled and hearing-impaired individuals, telephone additional worker, and other man-appliance edge tasks. Due to its relatively straightforward methodology, the insulated word identification systems were among the earliest speech recognition systems produced. In this analysis, the idea and practice of Automatic Speech Recognition as a framework for Isolated Digit Recognition are discussed. The goal of automatic communication identification is to develop methods and systems for early computer systems that receive input from communication signals. Speech signal recognition evolves, such as the speaker recognition, isolated word recognition, and speech signal to text conversion challenge. Users wanted the device to recognize their speech signals. By utilizing some basic elements, such as Wavelets, we have achieved significant progress in (ASR) designed for effectively-well defined presentations such as notation and average terminology operation doing duties in both contexts. In extremely loud environments, fiber optic communication is a cradle for big development. Our objective in this research is to acquire advanced techniques.

4. RESULTS AND INTERPRETATION

In recent years, speech recognition has developed into a useful computer technology with a variety of interactive speech-based applications. Speech serves as the fundamental building block for interpersonal communication. The speech recognition method is developed in the technological field using this method of communication. This method takes into account the input voice in order to extract important information and come to an accurate conclusion regarding the subject text. Here, the research suggests a method for handling online portals via voice for data access (from any location or time) by efficiently incorporating speech recognition techniques. The development of the STT (speech-to-text) conversion paradigm, which converts speech to text for use by the Deaf and in other fields, is a result of the advancement and development of technology. Examining the acoustic properties connected to sound and voice through data mining has been quite successful.

The study presents voice analysis and recognition as a speech process. The document level and the sentence level are the first and second layers, respectively, of the STT (speech to text) model. The suggested technology helps deaf people access data from any location at any time, as well as in other contexts. The technology produces results with excellent precision in the allotted time.

In recent years, speech recognition has developed into a useful computer technology with a variety of interactive speech-based applications.

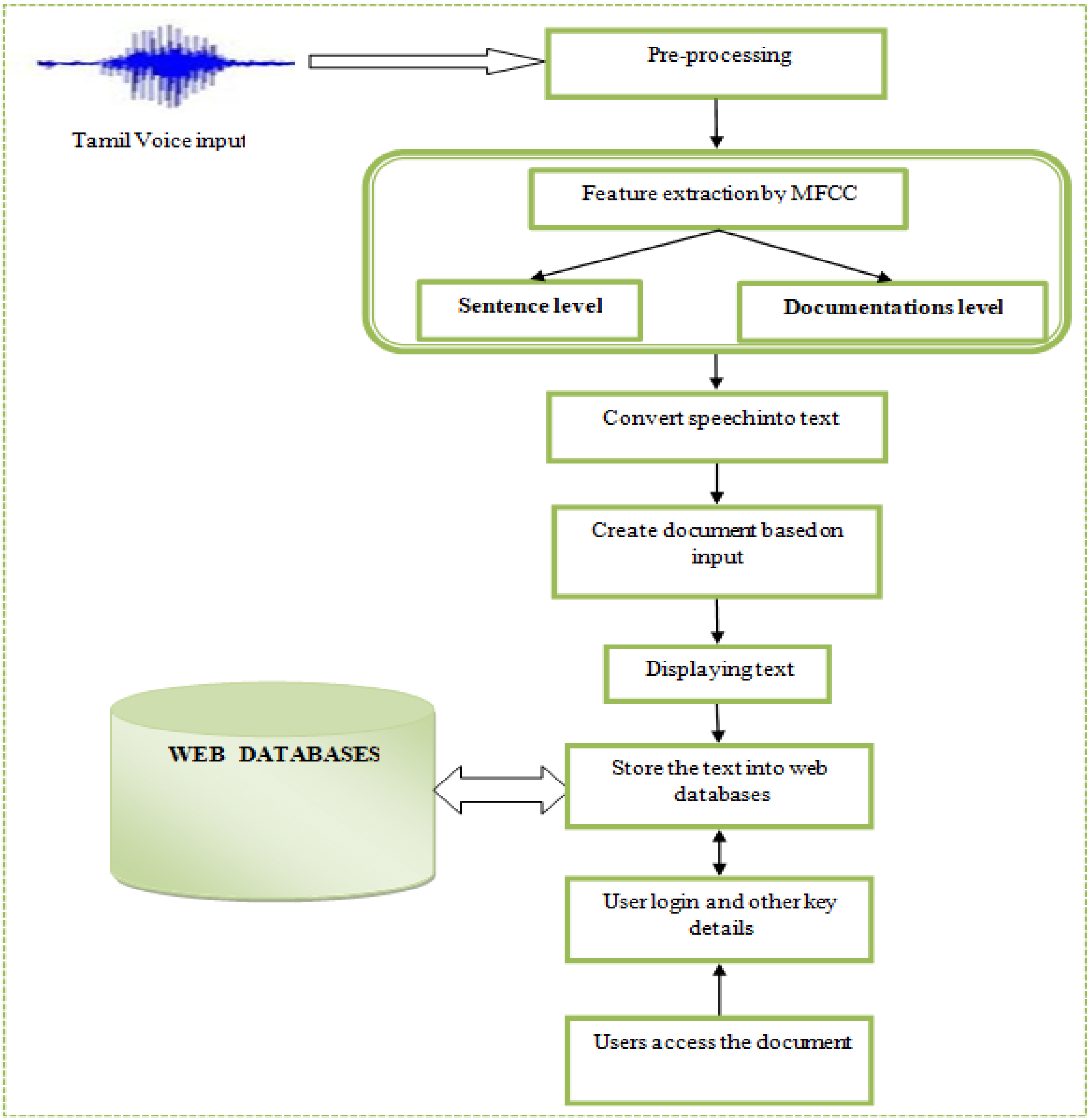
Tamil Voice input
Pre-processing
Feature extraction by MFCC
Sentence level
Documentations level
Convert speech into text
Create document based on input
Displaying text
WEB DATABASES
Store the text into web databases
User login and other key details
Users access the document

FIGURE-1 DESIGN OF ARCHITECTURE

Without a doubt, SR applications work well in an ideal context (i.e., one without distracting sound signals), but in order to use this technique in more real-world settings, the real problem of "sound source separation" needs to be addressed. The study offers a method for improving speech recognition technology's performance while using a web portal's voice interface to access data at any time and from any location. Our method demonstrates that it significantly outperforms the method that uses optimum noise removal, related to several objective measures taking speech recognition into account as well, without evaluating the humming noise.

4.1 COLLECTION OF INPUT VOICE DATA

Speech technologies help by providing common interfaces via which the public can access digital data and by streamlining information transmission between various people, such as speaking Tamil. The STT (Speech to Text) method accepts real-time input from a microphone in the form of audio, which is then converted into text format and displayed on the monitor or desktop. Tamil voice is used as input data in the work, which is then transformed to text format. The system operates by taking user speech or voice as input, which is then put through pre-processing to remove any silence sections from the voice stream. Systems that interact with people are trained and educated using the data.

4.2 FEATURE EXTRACTION OF SPEECH SIGNAL

Since each speech utterance has different unique attributes attached to it, it is an important step in the speech recognition process to distinguish one speech from another. These features are recovered using a variety of proposed feature extraction strategies, successfully utilizing them in the speech recognition process. The initial stage of the element extraction is the discourse assessment. It orders a quick spectral analysis of the sign and provides basic characteristics. Features showing the capacity range's outer limits as shown by brief conversation breaks. An all-inclusive element vector with both fixed and dynamic highlights is being arranged at the second level. The third and last organize change the aforementioned include vectors into incredibly cautious with powerful vectors, which is sent as contribution to the recognizer.

4.3 MELFREQUENCYCEPSTRUMCOEFFICIENTS (MFCC)

It is required and a challenging problem for voice recognition to recognize the signals used by the deaf in order to produce similar textual data to that used for communication by hearing persons. The results of the psychophysical study make it clear that a person's perception of the sound frequency contents of speech signals is not linear. Therefore, theoretical pitch is determined on a scale referred to as the "Mel" scale for every tone having a real repeat f, expressed in Hz. By changing the logarithmically consolidated channel yield energies, which are obtained by using a perceptually separated triangular-sift bank that brings about the handling of a DFT (Discrete Fourier Transformed) speech signal, the MFCC coefficients include the gathering of DCT de-associated factors. In contrast to the conventional MFCC, the approach limits the number of Feature vectors by an identical number of channels. For the purpose of deriving the coefficients, the speech sample is taken into consideration. The voice or speech is transformed to text format by using MFCC.

4.4 MFCC Structure

For the purpose of producing good recognition performance, it is crucial to retrieve the most appropriate parametric portrayal of acoustic signals. Efficiency at the current levels is crucial for the next step since it can affect how it behaves.

Step 1: High frequencies are pre-emphasized by processing the signal through a filter in this step. Here, a large frequency is used to increase the signal's energy.

Step 2: Framing - This step entails dividing the speech tests that were recovered from ADC (easy to modernized change) into segments that are between 20 and 40 msec in length. The voice sign recovered is divided into housings using N tests. M divides edges that are adjacent (where M>N). The usual values used are M = 100 and N = 256.

Hamming windowing at stage three. In order to solidify all of the nearby repetitive lines, the Hamming window is used while taking into account the resulting square in the feature extraction process. The Hamming window condition is up next.

Consider the window communicated as

W (n), $0 \leq n \leq N-1$ where N = number of tests in each edge

Y[n] = Output signal X (n) = data signal

W (n) = Hamming window, yield of the windowing sign is addressed as:

Stage 4: Fast Fourier Transform—Each edge of the N time zone tests is converted into a repeat space. The Fourier Transform deals with converting the convolution of the vocal tract drive response H[n] and the glottal heartbeat U[n] into temporal space. This assertion is supported by the following condition:

Y (w) = FFT [h (t)* X (t)] = H (w)* X (w)

Step 5: Processing the Mel Filter Bank-

The voice sign is non-straight and there are significant frequencies that fall inside the FFT range. The size recurrent reaction of the channel is triangular in shape, similar to solidarity at the inner rehash, and decreases abruptly to zero at the focal rehash of the two involved connected channels. Every single channel result displays the sum of its filtered absurd components. From that moment on, the following criteria are utilized to calculate the Mel for current recurrence (f)as far as

Hz. F (Mel) = [2595 * log 10 (1 + f)]700]

Algorithm:

Input:*i^{th}frame,$x_{p,i}(n)$*

Output: *MFCC cepstralco-efficients for i^{th} framefori=1, 2… Numbersof frames do*

$X_{p,i}(f) = DFT\{x_{p,i}(n)\}$

```
forj=1,2….,$N_F$do
$N_F$- number of sub band filters used in the Mel filter bank
$Y_{i,j}(f)=A_j(f)X_{p,i}(f)$ {output of $j^{th}$ filter of Mel filter bank A}
$Y_{i,j}(n)=IDFT\{Y_{i,j}(f)\}$ {sentencelevel output}
Where
$Y_{i,j}(n)$=Mel sub band filtered signal$Z_{i,j}(n)=$
$\{y_{i,j}(n),DI\}$
Where
$Z_{i,j}(n)$-Mel filter bank
DI-Dependency index
$E_{i,j}=Mean\{Z_{i,j}(n)\}$
$S_{i,j}=log\{E_{i,j}\}$ {document level output}
end for
end for
```

4.5 TRANSLATE SPEECH TO TEXT

This method helps the systems respond to live user input with accuracy and dependability, providing valuable services. People use such a quick approach since live voice communication using a computer system is more prompt than using a microphone. Given that Tamil is the primary language of communication among people, it is clear that the public wants speech computer interfaces. The speech recognition system was developed using STT (speech-to-text), which enables the conversion of voice/audio requests and dictation into a text format, to address this issue. When it comes to STT, speech recognition includes converting an auditory signal to text format. The information obtained may also be used for documentation purposes.

5. CONCLUSION

Talk confirmation has attracted observers as a fundamental request, has had a creative impact on society, and is predicted to succeed in any area of human-machine collaboration. The fundamentals and forward movement of this discourse affirmation advancement are examined in this work. The main objective of our project is to improve a structure that can be developed and realized with a few calculations and methods, allowing the PC to translate voice commands into content and carry out actions according to those commands. These days, most businesses and household appliances employ speech recognition technology to make tasks easier. Physically challenged people and those with vision impairments can both greatly benefit from this technology.

REFERENCES

1. Dillon, H., James, A., &Ginis, J. (1997). Client Oriented Scale of Improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids. American Journal of Audiology, 8, 27-43.
2. Divenyi, P. L., &Haupt, K. M. (1997). Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects. Ear Hearing, 18(1), 42-61.
3. Erber, N. P. (1996). Communication therapy: for adults with sensory loss: Clavis Publishing.
4. Fairbanks, G., &KodmanJr, F. (1957). Word intelligibility as a function of time compression. The Journal of the Acoustical Society of America, 29(5), 636-641.

5. Ferguson, M. A., &Henshaw, H. (2015). Auditory training can improve working memory, attention, and communication in adverse conditions for adults with hearing loss. Frontiers in psychology, 6, 556. doi: 10.3389/fpsyg.2015.00556
6. Ferguson, M. A., Henshaw, H., Clark, D. P., & Moore, D. R. (2014). Benefits of phoneme discrimination training in a randomized controlled trial of 50- to 74- year-olds with mild hearing loss. Ear Hearing, 35(4), e110-121.
7. Fu, Q. J., & Galvin, J. J., 3rd. (2007). Computer-Assisted Speech Training for Cochlear Implant Patients: Feasibility, Outcomes, and Future Directions. Seminars in hearing, 28(2). doi: 10.1055/s-2007-973440
8. Füllgrabe, C., & Rosen, S. (2016). On the (un) importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. Frontiers in psychology, 7, 1268.
9. Gagné, J.-P., Stelmacovich, P., &Yovetich, W. (1991). Reactions to requests for clarification used by hearing-impaired individuals. The Volta Review. 93, 129-143
10. Gagne, J. P. (2000). What is treatment evaluation research? What is its relationship to the goals of audiological rehabilitation? Who are the stakeholders of this type of research?. Ear and Hearing, 21(4), 60S-73S.
11. Gatehouse, S. (1999). Glasgow hearing aid benefit profile: derivation and validation of. Journal of the American Academy of Audiology, 10(80), 103.
12. Gatehouse, S., Naylor, G., &Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. International Journal of Audiology, 42(sup1), 77-85.
13. Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). International Journal Audiology, 43(2), 85-99.
14. Gil, D., &Iorio, M. C. M. (2010). Formal auditory training in adult hearing aid users. Clinics, 65(2), 165-174. doi: 10.1590/s1807-59322010000200008

15. Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. Journal of Speech, Language, and Hearing Research, 36(6), 1276-1285.
16. Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. Journal of Speech, Language, and Hearing Research, 42(2), 300-311.
17. Green, T., Rosen, S., Faulkner, A., & Paterson, R. (2013). Adaptation to spectrally-rotated speech. The Journal of the Acoustical Society of America,134(2), 1369- 1377. doi: 10.1121/1.4812759
18. Hande, N., Archana, Krishna. (2015) Analysis of commuincation strategies used by hearing impaired individuals (Unpublished dissertation), University of Manipal, India
19. Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. Journal of Speech, Language, and Hearing Research, 33(1), 149-155.
20. Henshaw, H., & Ferguson, M. A. (2013). Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. PLoS One, 8(5), e62836. doi: 10.1371/journal.pone.0062836
21. Hickson, L., Worrall, L., &Scarinci, N. (2007). A randomized controlled trial evaluating the active communication education program for older people with hearing impairment. Ear Hearing, 28(2), 212-230.