HIGH SCHOOL EDITION
Journal of Student Research
Volume 14 Issue 1 (2025)

A Review of Automatic Speech Recognition Technology and its Applications in the Medical Field

Abhinav Sood[1] and Manisha Srivastava[#]

[1]Suncity School Gurgaon, India
[#]Advisor

ABSTRACT

The last few years have brought significant advances in Automatic Speech Recognition (ASR) technology due to the rise of deep learning technology. The paper discusses various challenges relating to memory usage and computing power required for ASR are discussed as well as novel methods for combating them. It also highlights the implementation of ASR in embedded devices in the medical field. It discusses its diagnostics capabilities for various neurodegenerative disorders and its role in improving life for those afflicted with deafness or dyslexia through cochlear implants and reading therapy. Despite significant improvements in such medical technology, there are still challenges such as lack of availability of data or noisy environments with reverberant conditions.

Introduction

Speech is one of the most important aspects of human communication. Due to the importance of communication, there is a pressing need to develop novel technologies that can improve communication between individuals and enhance the lives of those with communication impairments due to diseases. ASR brings technology closer to humans. In the last few years, there has been a significant increase in technology in this field due to the developments in artificial intelligence (AI). Virtual assistants such as Siri and integrated smart-home devices such as Amazon Echo or Alexa have been largely successful, and they are capable of performing complex tasks relating to speech-processing; they are able to understand speech even in environments with significant noise.

However, ASR technology is highly computationally demanding and requires a large amount of resources, bringing significant challenges to its application in embedded systems rather than utilising the cloud, a network of servers accessed over the internet where software and databases can run. Still, there is a growing need for the use of ASR in such systems since they're compatible with wearable devices fit for medical use. These devices have constraints relating to memory and computing power, requiring sophisticated techniques to maintain a high degree of accuracy and a real-time component, while minimising the energy consumption and memory usage.

There are several papers discussing the fundamentals of ASR technology and solutions to some common issues relating to it [1-4], the applications of ASR technology for treating people with deafness through cochlear implants [5-10], training people with dyslexia to read more effectively [11-12], and diagnosing people with various neurodegenerative disorders [16-19].

As opposed to the papers mentioned above, this paper analyses hardware constraints and potential improvements associated with specific applications of ASR technology in medicine. The working behind the various applications and the motivation for suggested improvements is presented in detail, along with graphical representations of the systems as needed. The primary purposes of this paper are to: (i) discuss recent literature related to ASR technology and their medical applications and (ii) provide a comprehensive analysis of the future directions and shortcomings of medical devices using ASR. This paper describes the fundamentals of ASR systems, including features and feature extraction methods. Differences between traditional pipeline and end-to-end systems are discussed. The paper delves

into broad challenges faced across all ASR systems and how those challenges may be addressed. Finally, specific medical applications are discussed.

Basics of ASR Systems

Feature Extraction Methods

To extract information from a raw speech signal in the time-domain, the most basic method is a simple Fourier Transform, which returns the same signal in the frequency-domain. Since statistics of speech signals are constantly shifting across the temporal domain, Fourier Transforms are inadequate to perform any sort of speech-recognition. To address the shifting of the signal with time, methods such as Short-Time Fourier Transforms (STFTs), Mel-Frequency Spectral Analysis, and Mel-Frequency Cepstral Analysis are used, with the last of these being the most common.

Short-Time Fourier Transform

STFTs are used to extract spectrograms, a basic feature of speech-recognition. The process involves splitting speech signals into short, overlapping time-frames, windowing the signal (multiplying it with a window function, the amplitude of which smoothly decreases to zero rather than sharply cutting off) to smooth the frame border discontinuities, then applying the Fast Fourier Transform algorithm. A logarithm is applied on the data received. This is then plotted to provide a spectrogram. A spectrogram represents the power of the frequency bins at various points in time.

STFTs effectively capture the variation of a signal with time, being able to perform ASR, but they still require a large amount of data to be processed by the neural network (NN) and do not yield very accurate results. Further, humans do not perceive all frequencies the same way, which STFTs fail to account for.

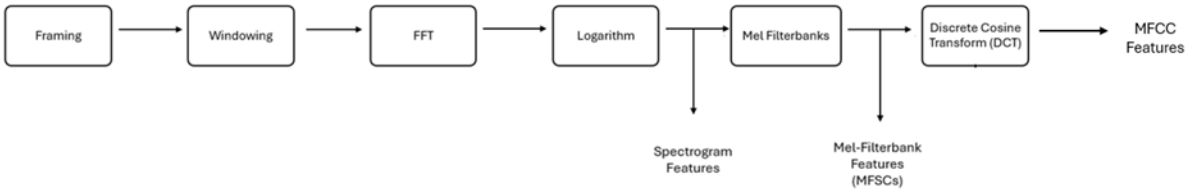

Figure 1. Different speech features obtained at different stages.

Mel-Frequency Coefficients

Mel-Frequency Coefficients were designed to address the shortcomings in STFTs. The data obtained from the STFTs is put through a set of triangular Mel-filterbanks, which translate the data from a wide to a narrow range. It is adjusted to better represent human ear perception of the data. Critical bandwidth, or the frequency resolution of the human ear, is roughly constant on a linear scale below 1000 Hz [20]. Moreover, human perception of changes in frequency is much more sensitive for lower than higher frequencies. So, the Mel-filterbanks treat frequencies between 0 and 1000 Hz approximately linearly and those above 1000 Hz logarithmically. The features obtained during this process are known as Mel-Frequency Spectral Coefficients (MFSCs).

Upon being further put through a Discrete Cosine Transform (DCT), this data is compressed into discrete blocks. DCTs have strong energy compaction properties as they retain high quality at high data compression ratios. The features obtained are Mel-Frequency Cepstral Coefficients (MFCCs). This is the feature most widely used by ASR technology.

Pipeline vs End-to-End Systems

Pipeline and End-to-End systems refer to the arrangement of segments in the ASR technology. Pipeline systems refer to the traditional system where multiple components as discussed in 2.2.1 work together in a sequential manner. End-to-End systems refer to a single NN, which takes a raw signal as an input, and a fully processed speech transcript as an output.

Pipeline ASR

The main task of a pipeline system is to identify the most probable sequence of words given the probability of the speech signal being generated by that sequence of words. Pipeline ASR is divided into various phases. First, feature extraction via methods discussed earlier takes place. Then, the data goes through an Acoustic Model (AM), which is trained on a large amount of speech data from speakers with varying enunciations and tones. The acoustic model provides the probability of each signal sequence represented by those corresponding speech features, given the corresponding sequence of words. The data then goes through a Phonetic Model (PD) and Language Model (LM), which are both implemented as probabilistic components, which deal with the word sequences to match them in phrases with a logical meaning.

The AM uses a mathematical model trained to calculate the probability of each word. This results in the base unit for the pipeline system being words. Base units of a system should fulfil two major requirements: 1) they must be trainable; there must be enough occurrences of them in acoustic contexts and enough data to be able to correctly estimate the unit, and 2) they must be generalisable; it must be possible to create any other word from a string of units. Words are neither generalisable nor trainable as opposed to a sub-lexical unit like phonemes. This is a major drawback of this type of system.

End-to-End ASR

In end-to-end systems, the acoustic, phonetic, and language models are integrated into one neural network as opposed to pipeline systems. In some systems, feature extraction beforehand is an optional step as it is integrated in the NN. The base unit for the acoustic model in an end-to-end system are characters (graphemes) instead of words. It is possible to have an approach independent of the lexicon (a well-defined set of words broken down into phonemes or pronunciation units) using end-to-end networks, which allows the network to handle out-of-vocabulary words, which are words not seen in the training data. This, however, has to be balanced with the risk of getting words with no meaning. End-to-end systems are also fully discriminative, which means they learn everything from data, and there are no required initial forced-alignments to start training [1].

Broad Limitations in ASR Technology

Major Challenges Faced

The main limitations faced in the implementation of ASR technology in medicine come from the need for embedded devices. Since these devices do not utilise the cloud, they require immense computing power and memory storage to be able to perform ASR. Most ASR technology works using Convolutional NNs, which are computationally intensive. The features obtained during the feature extraction stage consist of precise floating point values, which increases complexity of operations. Medical devices using ASR must also perform operations fast, so that user-perceived latency is very low. This also requires compression of the NN and data.

Potential Improvements

Depending on the requirements of the device, any number of the methods described below may be used. Drawbacks of each method have to be carefully balanced with the required optimisation, and specific needs and constraints can be catered to using combinations of these methods.

Architecture Optimisation

Architecture optimisation is a method that involves compression of NNs into those with a smaller number of layers. There are various accelerated algorithms to optimise network architecture and various different types of architecture optimisation. Neural Architecture Search (NAS) is a technique that is used to automate the design of a NN. Data has shown that NAS-created NNs are on par with or outperform hand-crafted NNs [2], while requiring less computational power.

Data Quantisation

The data NNs are normally trained on consists of 32-bit floating point values. The resulting model then has a very large number of parameters, and it is slow during inference time. To reduce the model footprint, it is thus necessary to compress the NN. In the post-training stage, quantisation is often applied to NNs that compresses them down to 16-bit, 8-bit, or even 4-bit values. By operating with a much smaller number of bits, the model becomes faster. It then uses less memory and is able to perform with real-time user interaction. This post-training compression, however, can cause significant performance downgrades and accuracy issues.

One way around this is Quantisation-Aware Training (QAT) schemes [3]. During model-training, the effect of quantisation is taken into account, essentially nullifying the performance issues associated with it. Experimental data measuring Word-Error Rate (WER) without using data quantisation and using QAT has shown that using QAT results in little to no degradation in accuracy after quantisation.

Most existing work using QAT has focused on replicating the effect of compression during the forward propagation under which the model loss is evaluated. Such loss already includes the quantisation effect, leading to no quantisation-related terms in the backpropagation phase. Thus, quantisation is only loosely incorporated in the gradient evaluation and weight update. Existing QAT schemes also require intensive computation and high memory usage.

Much work is being done to improve QAT schemes to address these drawbacks. One novel method involves imposing on the model weights a distribution with a similar shape as that of a quantised model, using Absolute Cosine Regularisation (ACosR) [4]. The total loss thus consists of both accuracy loss and ACosR loss, in which the ACosR loss represents how similar the model is to a quantised model. Data shows that ACosR quantisation reliably forces convergence of the weights to quantisation levels with very minimal performance loss while compressing data to 8-bits. Compressing data to 6-bits remains a challenge as it shows some, albeit little, degradation compared to the floating-point baseline. Using only one regularisation function requires much smaller additional memory and computation resources compared to already existing QAT schemes.

Applications of ASR

Through novel optimisation methods and technical advancements in the field as discussed above, ASR technology was able to be implemented in embedded devices with a high level of accuracy. This opened up many new applications of ASR in various different fields from music to medicine with real-time capabilities and computational efficiency. This section discusses three modern medical applications of ASR technology designed for patients of specific diseases along with some limitations and potential improvements for the same.

Deafness and Hearing Loss

A cochlear implant (CI) is a small electronic device that can help profoundly deaf or hard-of-hearing individuals perceive sound. As of December 2019, more than 700,000 people worldwide use CIs [9]. It consists of an external portion placed behind the ear, and a second portion placed surgically under the skin. It consists of a microphone that picks up sound signals, a processor that selects and arranges those sound signals, a transmitter and stimulator that receives these signals and converts them into electrical impulses, and an array of electrodes that collects these signals and stimulates different portions of the auditory nerve.

These devices do not restore normal hearing, but rather provide a representation of speech to individuals that allows them to understand speech. Unlike hearing aids, which simply amplify incoming audio, CIs actually bypass the damaged portions of the ear by directly sending signals to the auditory nerve. As such, proper use of CIs often requires extensive therapy and training. As of 2020, CIs can be administered in infants as young as nine months. Research and studies have shown that if CIs are implanted early in life, speech processing abilities of individuals can match those without hearing impairment.

However, there are many more individuals facing severe, profound, or moderate bilateral sensorineural hearing loss who do not benefit from hearing aids.

Major Challenges

Although cochlear implants can restore hearing in profoundly deaf individuals, they face significant performance issues in reverberant environmental conditions as opposed to anechoic quiet conditions. Speech perception issues arise when there is a background masker present. There is also a significant increase in WER in ASR systems trained on anechoic conditions when signals with reverberant backgrounds are passed as an input.

In environments where there are multiple signals to be received and processed simultaneously, CIs are very poor at distinguishing the direction of origin of each signal.

Potential Solutions/Augments

To improve the working of CIs, they are often augmented with other hardware or specialised ASR. A specialised ASR system trained on both anechoic quiet conditions and reverberant conditions can be used to improve speech recognition in conditions with background maskers present. The ASR system output text transcription is run through a text-to-speech synthesiser, which is finally sent to the CI as an input. The block diagram for the same is attached below in Fig. 2. Training data for these systems consists of anechoic clean train and test sentences from various individuals' voices convolved with four room-impulse responses (RIR) with various reverberation times. In research conducted by Neuman et al. [6], the RIR was recorded in a room of size 10.06m x 6.65m x 3.4m (length x breadth x height), with reverberation times being gradually varied between 1.0s to 0.8s, 0.6s, and 0.3s. For intelligibility tests, both naturally spoken and synthesised speech were considered. Experimental results show that the WER using ASR trained on non-reverberant conditions reached up to 64%, but could be reduced to as low as 3% with ASR trained as described above.

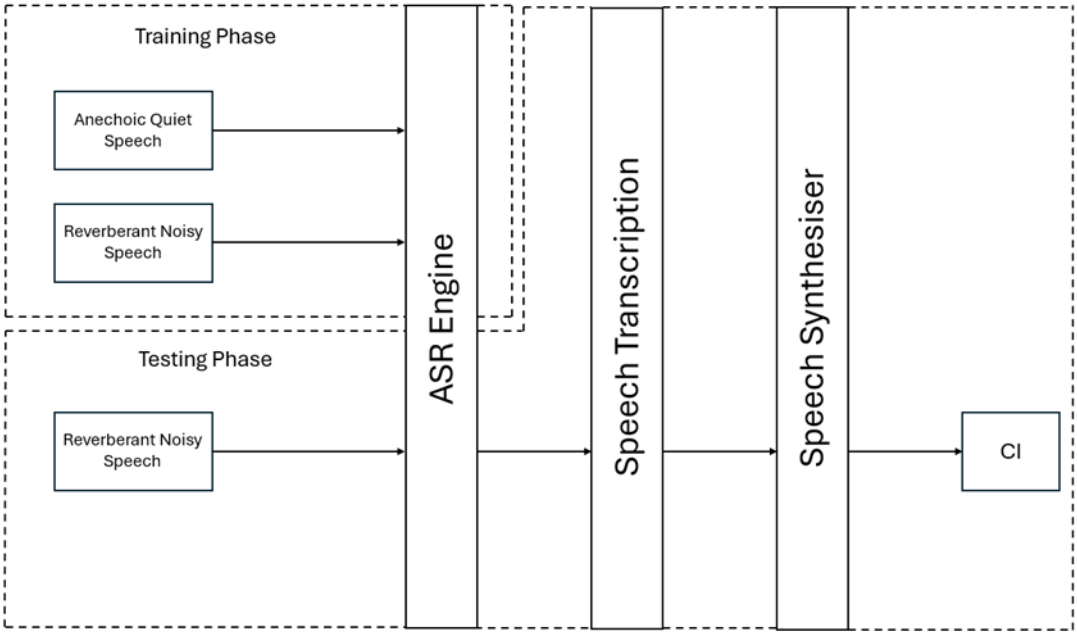

Figure 2. Block diagram of a hybrid reverberant and anechoic speech handling CI system.

In human ears, spatial cues for audio include arrival time and sound intensity differences between the two ears (interaural time and level differences). Since most CI users only have the device implanted in one ear, there is limited access to these cues. Thus, in cases where sound and noise or multiple sound signals are spatially separated, CIs show performance issues. Some researchers combat this with Electro-Haptic Stimulation (EHS) [8, 10], which involves a small non-invasive body-wearable device that sends haptic impulses to the wrist. Speech amplitude envelope cues were extracted from signals in the environment and delivered to each wrist via haptic stimulation. Thus, the intensity difference between the wrists corresponds to the level difference between the ears. Fletcher et Al. [8] recently showed that EHS greatly improves sound localisation and spatial recognition amongst CI users.

Dyslexia

Dyslexia, or word blindness, is a common learning disability that affects reading and writing abilities of individuals. Dyslexia patients often face trouble with spelling, sounding out words in their mind, reading or writing quickly and accurately, and comprehending what they read. It usually develops due to environmental and genetic factors, but can sometimes develop due to traumatic brain injuries and strokes, in which case it is referred to as acquired dyslexia or alexia. Millions of people worldwide suffer from dyslexia. Further, training individuals to improve reading abilities is a task requiring specialised personnel and time. As such, a robust method to transfer some of the load of assisting afflicted individuals towards technology is greatly needed. ASR-driven simulated reading systems using multi-modal presentation techniques are used for the same.

Training Process

The method traditionally used with speech therapists to train dyslexia patients to be able to read better is known as the ‘Book-and-Tape’ method. The patient is given a book, a segment of which is played out loud by an audio recording while the patient reads that segment. Then, the patient reads that same segment aloud and progresses to a new segment of text. Then, the speech therapist provides feedback, corrects any errors, and evaluates performance.

ASR-driven dyslexia training largely aims to mimic this method without the presence of a speech therapist, using similar multiple sensory channel stimulation to enhance reading and text decoding capabilities of individuals. The explicit way of implementing this automatic speech therapy is through applying a user-centred design process and an elaborate usability testing phase for different scenarios to gauge the general preferences of the target group [11]. The process can be condensed into three major systems. The feedback system, the assistance system, and the evaluation system. Each system has to address two issues: format and timing i.e. how and when to provide feedback, assistance, and evaluation.

The feedback system acts as a progress tracker for the patient, allowing them to see what has been read so far and whether or not it was read correctly. The format of providing this feedback can be either through highlighting text in different colours, striking out incorrectly read text, etc. The timing can either be at a word-level, i.e. providing feedback as each word is read, or at a sentence-level, i.e. providing feedback for a whole sentence after it is read.

The assistance system acts as a method to help the user when they are incapable of reading out a particular word being prompted to them. The format of providing this assistance depends heavily on the user's preferences; it can be through pre-recorded speech, visual cueing , contextual switching, morphological re-phrasing (i.e. into the morphological root word), or sub-division of words into syllables. The timing can be either user-requested, initiated by the system itself, or both.

The evaluation system provides a performance review of the user. The format can be through providing numerical or statistical data, word mark-up in green and red colours, or word mark-up in shades of red and green depending on the degree or rate of misreading. The timing can be successive or continuous, i.e. providing the evaluation at the end of a text segment, or simultaneously as the user reads.

Since each aspect of improving reading proficiency in dyslexic patients is so customisable depending on user preferences, it allows for automated, personalised treatment.

Future Directions

One of the current work-in-progress improvements is reduction of WER for dyslexic reading-oriented ASR engines. One method being worked on is phoneme refinement and alternative pronunciation [12]. ASR-based dyslexic reading training technology has so far just been tested through experiments and not put into effect on a larger scale. While the data collected so far suggests users are capable and willing to use this technology for training their reading abilities, the long-term effects of the same remain undisclosed and require further research and data collection to determine.

Neurodegenerative Disorders

Neurodegenerative disorders (NDs) involve a progressive loss of neurons due to neuronal damage and death.There is currently no known cure to neuronal damage, but it may be treated and slowed down. Thus, early detection of these diseases is crucial. NDs can lead to episodic or total memory loss, loss of motor function, impairments in speech or speech processing, problem-solving related issues, etc. Speech impairments are one of the earliest noticeable symptoms, and can therefore serve as early detectors of these diseases. ASR technology can thus be put to use in diagnosis of several NDs that exhibit speech-related symptoms. This form of diagnosis is cost-effective, non-invasive, and time-efficient.

The disorders discussed within the scope of this paper are — Alzheimer's Disease (AD), which is a disease that leads to extensive neuronal loss, faulty synaptic connections, and damage to essential neurotransmitters required for proper brain function [13]. Executive function and speech impairment are also clinical manifestations that usually appear early. Friedreich's Ataxia (FA), which is the most common inherited ataxia, may lead to degradation of peripheral and central nervous systems, the heart, skeleton, and more [14]. Multiple Sclerosis (MS), which is a disease that can cause optic neuritis, brain and spinal cord syndromes, cognitive impairment, sensory ataxia, and more [15]. All these NDs cause patients to show noticeable effects on patients' speech features. They tend to speak slower and more variable speech rates, exhibit lower and more variable pitch, and show lower spectral clarity.

Potential Solutions

Acoustic features of speech can be used as objective clinical markers for diseases affecting the central nervous system. ASR has been used to distinguish between healthy individuals' speech and pathological speech. Further, NDs like Parkinson's disease and spasmodic dysphonia, which have single and well-defined patient populations, have had significant success with using ASR for diagnosis [16]. However, efforts are being made to be able to use ASR to specifically identify the disease by analysing details between different NDs with similar speech phenotypes such as the ones discussed in this paper.

Speech features such as speech rate, duration, and number of pauses between words, and the pitch and timbre of the voice, are the most commonly used for diagnosing various pathologies. The diadochokinetic (DDK) task is widely used for this, in which a patient is tasked with repeating a string of syllables as quickly and clearly as possible within one breath for up to 10 seconds. The information obtained using DDK is highly consistent amongst speakers of various nationalities and ethnicities, having different accents or intonation. One study using this method and a specialised NN concluded that the successful diagnosis rate is significantly more than chance - almost 36% for MS [17].

To improve diagnosis rate of AD, technology to more successfully identify high-quality segments of speech from recordings is being worked on, since a large amount of speech data for the same is low-quality.

Major Challenges

Although ASR shows immense potential in ND diagnosis, there is still a lack of accuracy. Most NDs have a small amount of readily available speech data of afflicted patients, which makes training NNs successfully more difficult. AD has a significant amount of data in databases such as DementiaBank; however, most of this data is very low-quality and limits accuracy of diagnosis [18]. Development of features that can help distinguish between similar NDs is still underway as well; features such as spectral kurtosis and skew, which are not generally used for ASR, are being explored with various new studies [19].

Conclusion and Future Directions

In this work, the fundamentals of ASR systems have been discussed, with a particular focus on their applications in medicine. Various ASR-utilising embedded devices, the challenges they face, and the techniques used to optimise performance were talked about. Specific applications such as cochlear implants, reading training in dyslexia patients, and ND diagnosis were examined. ASR has immense potential in the field of medicine; technology using it encompasses various sectors of the field - personalised medicine, treatment of patients, and disease diagnostics.

However, there are various challenges that must be addressed: the robustness and accuracy of ASR systems in noisy and reverberant environments needs to be worked on. Further advancements in data quantisation techniques and architecture optimisation would play an important role in improving feasibility of using ASR in wearable medical devices, which are constrained to low-power consumption while requiring high efficiency. Speech data for NDs must be made more accessible and high-quality to improve the accuracy of ASR-based diagnostic tools. As these technologies continue to evolve, they have the potential to greatly improve the quality of life for individuals with communication impairments and contribute to early diagnosis and treatment of various medical conditions.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

Abbreviations

ACosR: Absolute Cosine Regulation
AD: Alzheimer's Disease
AI: Artificial Intelligence
AM: Acoustic Model
ASR: Automatic Speech Recognition
CI: Cochlear Implant
DCT: Discrete Cosine Transform
DDK: Diadochokinetic
EHS: Electro-Haptic Stimulation
FA: Friedreich's Ataxia
LM: Language Model
MFCC: Mel-Frequency Cepstral Coefficients
MFSC: Mel-Frequency Spectral Coefficients
MS: Multiple Sclerosis
NAS: Neural Architecture Search
ND: Neurodegenerative Disorder
NN: Neural Network
PD: Phonetic Model
QAT: Quantisation-Aware Training
RIR: Room Impulse Response
STFT: Short-Time Fourier Transform
WER: Word-Error Rate

References

1. Georgescu, Alexandru-Lucian, Alessandro Pappalardo, Horia Cucu, and Michaela Blott. 2021. "Performance vs. Hardware Requirements in State-of-The-Art Automatic Speech Recognition." *EURASIP Journal on Audio, Speech, and Music Processing* 2021. https://doi.org/10.1186/s13636-021-00217-4

2. Cong, Shuang, and Yang Zhou. 2022. "A Review of Convolutional Neural Network Architectures and Their Optimizations." *Artificial Intelligence Review*, June. https://doi.org/10.1007/s10462-022-10213-5

3. Zhao, Qiuming, Guangzhi Sun, Chao Zhang, Mingxing Xu, and Thomas Fang Zheng. 2024. Review of *Enhancing Quantised End-To-End ASR Models via Personalisation*. IEEE. March 18, 2024. https://doi.org/10.1109/ICASSP48485.2024.10448012

4. Nguyen, Hieu Duy, Anastasios Alexandridis, and Thanasis Mouchtaris. 2020. "Quantization Aware Training with Absolute-Cosine Regularization for Automatic Speech Recognition." Amazon Science. 2020. https://doi.org/10.21437/Interspeech.2020-1991

5. Hazrati, Oldooz, Shabnam Ghaffarzadegan, and John H.L. Hansen. 2015. Review of *Leveraging Automatic Speech Recognition in Cochlear Implants for Improved Speech Intelligibility under Reverberation*. IEEE. August 6, 2015. https://doi.org/10.1109/ICASSP.2015.7178941

6. Neuman, Arlene C., Marcin Wroblewski, Joshua Hajicek, and Adrienne Rubinstein. 2010. “Combined Effects of Noise and Reverberation on Speech Recognition Performance of Normal-Hearing Children and Adults.” *Ear and Hearing*. https://doi.org/10.1097/aud.0b013e3181d3d514

7. Huang, Juan, Benjamin Sheffield, Payton Lin, and Fan-Gang Zeng. 2017. “Electro-Tactile Stimulation Enhances Cochlear Implant Speech Recognition in Noise.” *Scientific Reports* 7. https://doi.org/10.1038/s41598-017-02429-1

8. Fletcher, Mark D., Haoheng Song, and Samuel W. Perry. 2020. “Electro-Haptic Stimulation Enhances Speech Recognition in Spatially Separated Noise for Cochlear Implant Users.” *Scientific Reports* 10. https://doi.org/10.1038/s41598-020-69697-2

9. National Institute on Deafness and other Communication Disorders. 2021. “Cochlear Implants.” NIDCD. U.S. Department of Health and Human Services. March 24, 2021.

10. Fletcher, Mark D., Robyn O. Cunningham, and Sean R. Mills. 2020. “Electro-Haptic Enhancement of Spatial Hearing in Cochlear Implant Users.” *Scientific Reports*. https://doi.org/10.1038/s41598-020-58503-8

11. Pedersen, Jakob Schou, Lars Bo Larsen, and Børge Lindberg. 2008. “Usability of ASR-Based Reading Training for Dyslexics.” *Interspeech 2008*, September. https://doi.org/10.21437/Interspeech.2008-474

12. Husni, H., and Z. Jamaludin. 2010. Review of *Minimizing Word Error Rate in a Dyslexic Reading-Oriented ASR Engine Using Phoneme Refinement and Alternative Pronunciation.* EDULEARN10 Proceedings. 2010.

13. Lamptey, Richard, Bivek Chaulagain, Riddhi Trivedi, Avinash Gothwal, Buddhadev Layek, and Jagdish Singh. 2022. “A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics.” *ProQuest* 23. https://doi.org/10.3390/ijms23031851

14. Koeppen, Arnulf H. 2011. “Friedreich’s Ataxia: Pathology, Pathogenesis, and Molecular Genetics.” *Journal of the Neurological Sciences*. https://doi.org/10.1016/j.jns.2011.01.010

15. Dobson, R., and G. Giovannoni. 2019. “Multiple Sclerosis - a Review.” *European Journal of Neurology* 26. https://doi.org/10.1111/ene.13819

16. Schultz, Benjamin G., Venkata S. Aditya Tarigoppula, Gustavo Noffs, Sandra Rojas, Anneke van der Walt, David B. Grayden, and Adam P. Vogel. 2021. Review of *Automatic Speech Recognition in Neurodegenerative Disease*. International Journal of Speech Technology. May 4, 2021. https://doi.org/10.1007/s10772-021-09836-w

17. Schultz, Benjamin G., Zaher Joukhadar, Usha Nattala, Maria del Mar Quiroga, Gustavo Noffs, and Sandra Rojas. 2023. “Disease Delineation for Multiple Sclerosis, Friedreich Ataxia, and Healthy Controls Using Supervised Machine Learning on Speech Acoustics”. IEEE. October 4, 2023. https://doi.org/10.1109/tnsre.2023.3321874

18. Pan, Y, B Mirheidari, M Reuber, A Venneri, D Blackburn, and H Christensen. 2020. “Improving Detection of Alzheimer’s Disease Using Automatic Speech Recognition to Identify High-Quality Segments for More Robust Feature Extraction - White Rose Research Online.” http://dx.doi.org/10.21437/Interspeech.2020-2698

19. Bertini, Flavio, Allevi Davide, Lutero Gianluca, Danilo Montesi, and Calzà Laura. 2021. “Automatic Speech Classifier for Mild Cognitive Impairment and Early Dementia.” *ACM Transactions on Computing for Healthcare*. https://doi.org/10.1145/3469089

20. Oxenham, Andrew J. 2018. “How We Hear: The Perception and Neural Coding of Sound.” *Annual Review of Psychology* 69. https://doi.org/10.1146/annurev-psych-122216-011635