e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

A VOICE RECOGNITION APPROACH FOR IDENTIFYING CYBER ATTACKS

Swapnil Mahajan*1, Dr. Harsh Lohiya*2, Mr. Gaurav Saxena*3

*1Research Scholar, Department Of CSE, School Of Engineering, SSSUTMS, Sehore, M.P., India.

*2,3Associate Professor, Department Of CSE, School Of Engineering, SSSUTMS, Sehore, M.P., India.

ABSTRACT

Basic speech knowledge is designed to recognize the principles of sound. Using MATLAB software to encrypt speech recognition, the director's voice can be recognized. For further analysis and processing it is necessary to convert the noise waveform into a parametric model. When it comes to biometrics, biometrics has become a security measure that reduces cases of fraud and theft because biometrics uses physical features and characteristics to identify people. Fingerprints and handwriting are the oldest types of biometric authentication; however, newer types of biometric authentication include iris/eye scans, facial scans, voice prints, and handprints. Biometric speech recognition technology focuses on training techniques to recognize specific speech patterns (e.g., voice notes). The technology is suitable for many applications, including mobile security management. The voice recognition system has been successfully implemented as a voice command door opening system that can only be used by authorized users. The device has been proven capable of providing moderate security control with adjustable security levels that can take the person's voice into account when performing any speech recognition.

I. INTRODUCTION

More and more devices are now using command-performing voice recognition technology that enables access to stored information and transmits audio to a written document. In automated phones, car entertainment systems or operating systems, for example, "Windows" and "iOS" voice recognition technologies are introduced. However, as voice recognition technologies are in infancy, hackers can experience a wide range of safety vulnerabilities to access sensitive information unauthorized. One of the main weaknesses of voice authentication devices is that voice information, such as fingerprints, can be more accessible than other biometric information. For example, in order for you to collect information about an individual's fingerprints, it might not be sufficient to have access to vocal information, since one has to be physically near items touched by a person. Today many people openly use many online platforms, such as YouTube, Snap Chat and Facebook to make their voice accessible to all. The technology used for voice recognition includes hardware and accompanying software that can decode human voices for different functions (e.g., transcribing voice to text, executing software applications, and verifying the identity of an individual).

In 1952, the first automatic voice recognition system. Although the system was not computerized, the human voice could identify single digits. Xuedong Huang, one of the founders of Microsoft's speech recognition community, was the first developer of a modern voice recognition device, Sphinx-II. The technology was able to perform voice recognition in real time and was ideal for use in modern software applications.

In addition to the advancement of computer technology, other devices have been widely used in different areas, including health care, customer service, avionics, military, automotive security and telecoms. For example, voice recognition technologies enable the immediate processing of voice in medical records in the field of health care. In the area of avionics, pilot training and air traffic, controller technology is used in voice recognition. In the area of car safety, technology used for voice recognition improves safety for passengers in cars, since it allows the driver to make telephone calls without having to use hands. Many businesses have automated menu systems in their service hotlines in the area of customer care. When a customer phones a service hotline, a registered question from a device he or she can hear. After the voice recognition programme processes the customer's response, the device will transfer the customer to an appropriate department.

Dynamic Time Warping Based Voice Recognition

The HMM-based technique was once widely used for speech recognition, but has fallen out of favor due to the widespread success of dynamic time warping. Variation in time or speed of dynamic time warping is an

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

algorithm for measuring similarity between two sequences Similarities in walking style can be found, even if the person is going slowly and if they are accelerating or decelerating during a single step A component of DTW has been used in film, audio, graphical data, and even non-linear applications. Automatic speech recognition is commonly used for various speaking speeds Nonlinear smoothing is a technique that relaxes the given constraints (e.g., restrictions) to find an optimal fit to the given sequences (i.e., for example, time series). This can be best understood as the early sequences being "warped" to fit each other. Hidden Markov models are commonly benefit from this form of positional weighting.

In contrast to hand-made or label-based Hidden Markov models, neural network models make less explicit assumptions about feature properties and are therefore easier to be used for speech recognition. Neural networks are excellent for training discriminative speech classifiers, but in a non-introutrage way. However, in spite of their relative effectiveness for tasks such as individual phonemes and isolated phrases, neural networks had difficulty modeling the time dependencies of continuous recognition, and as a result were seldom applied to longer problems.

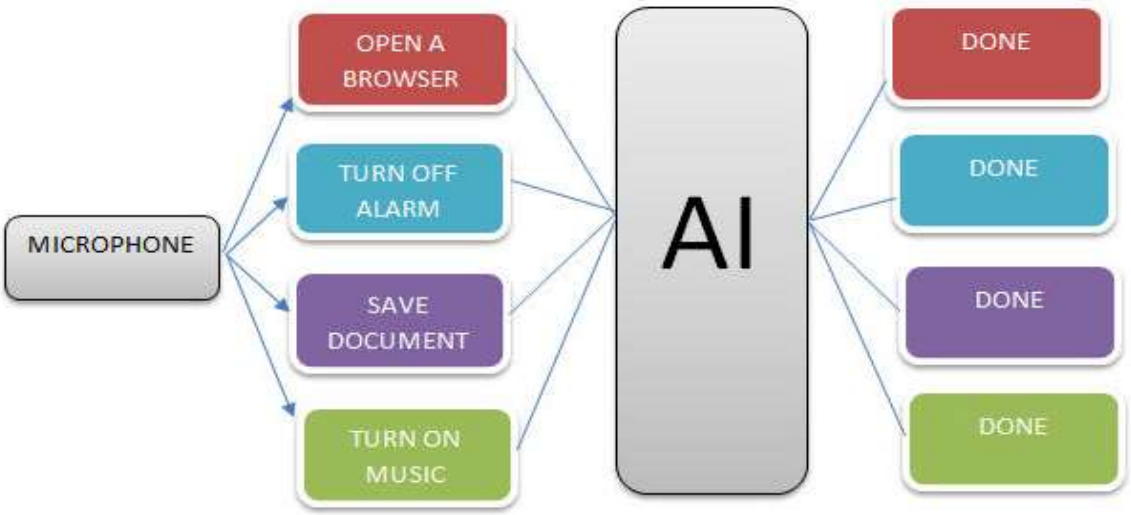

Figure 1: Voice Recognition

End-to-End Automatic Speech Recognition

Since 2014, ASR "ends to end" has received much attention for the earlier (e.g., all earlier) models needed different components and preparation for the pronunciation, each was based on a separate sound-based system (e.g., HMMs). The end-to-end models learn more about the speech recognition part of the system. It simplifies the training and deployment process, since it shortens both. Due to this, an HMM-based framework necessitates an n-gram model and sometimes consumes many gigabytes of memory rendering them unfeasible for mobile devices; an n-gram language model is sometimes employed instead. ASR systems made by Google and Apple are on the cloud, so a network connection is essential.

The first ASR attempt to apply temporal and connectionist techniques was made by Graves and Navas in 2014 with the introduction of CT-based systems by Google DeepMind and the University of Toronto. The model employed recurrent neural networks and a CTC (convolutional transcendental complex network) layers. The RNN-C model can both encode the language, but it cannot do so because of the conditional independence assumptions just like an HMM. However, as they learn speech acoustics directly, CSL models can make several spelling errors but the clean transcript approach would use a separate language model.

We favor a system-focused model over CTC (Closed-loop total care). Smart mirror therapy, developed by Carnegie Mellon University and Google Brain in 2016, was pioneered by Bahdan et al. Listen, track, and transcribe signal characters one character at a time. CTCST-based models cannot distinguish dependent voices and learn all of the components of a speech recognizer, including the pronunciation, acoustic model, and language model, instead, rather than being able to learn solely on attention. This means that during deployment, a reduced memory model is needed.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

II. LITERATURE SURVEY

Human voice is an amazing tool. Every person has a distinct sound, rhythm, pace, and pitch to identify, which differs depending on where they are in the sentence Since the voice of each person is special, it should be obvious that the average male's voice is lower than the average female's Human beings do not only have distinct personalities, but also notable speech patterns. There are several different ways to express even the simplest idea. At its peak frequency, the human voice is about 10 kilohertz; at its lowest, it is about the note C below middle C.

Chakrabartic, et al. use many features to present common acoustic characteristics of speech (2020). Although the MFCCs have performed admirably in the past, this feature extraction has also improved the speech recognition system in environments with background noise. Power-normalized Cepstral Coefficients (the most frequently cited method) and Cepstral normalization (a method popular in many works) are most common, because they deal with noise resulting from the products of convolution. Voice recognition is the method by which a machine determines what was said by analyzing spoken words. In principle, they have the capacity to be further subdivided into text-based and non-text dependent. Text-dependent is about the keywords or phrases, whereas non-specific text is more versatile for voice recognition.

The majority of the published studies found that the wavelet feature extraction demonstrated improvement over conventional Cepstral (Cepstral et al., 2020). The multivariable PWPs already depend on various frequency sub- specific extracted features, which have been combined to form a unique feature vector for each component the speaker recognition framework has been designed and is being implemented computationally Modeling of the unwanted noises and the Gaussian distribution composes the HMM. The spectrum has been calibrated to the way it is in the set. It will have Caprices.

Wavelets provide a tool for visualizing time-frequency content. Generally, it has been applied to the process of decomposing complex messages into high and low frequency components. Its wavelet coefficients represent a given signal's frequency content compared to that of a selected wave function. These signal and wavelet coefficients are determined by a convolution of the wave function that is using an extended band-pass filter; it can be considered as a scaled convolution describe how universal leadership model enables global organizing (2020). As a result, subsequent warped waves are then separated and used as a filter bank called Perceptual Wavelet (PWP). "Undantical" filtering provides better spectral and spatial localization than "Non-red" filtering".

III. RESEARCH METHODOLOGY

This paper is intended for use in a home or office setting where a group of people has been granted access. This project can also be extended. Using VBS, we can protect any software document/file. VBS can be accessed solely and exclusively from the user's conscious state. In addition, if the user is in an unusual state (frightened, concerned, etc.), the voice features that do not match the stored voice will be sampled automatically. As a result, no permission to access the system will be given. The proposed project establishes a mechanism for notifying the primary owner of the system of any actions taken with it. The plan is to send a message to the owner through the GSM network. This can occur in two instances: a) when an approved user is granted entry to the system; and b) when an intruder attempts to crack the door lock.

The architecture for the framework consists of following parts.

- Microphone circuit
- Fingerprint/Voiceprint analysis
- Filters (digital filter)
- Function written in code

Microphone Circuit

This circuit can accept an input and output a few millivolts. Thus, in order to facilitate further processing, the output should be amplified. This amplification is accomplished by the use of three op-amp levels. Additionally, low pass and high pass filters are introduced to aid in the op-amp configuration.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

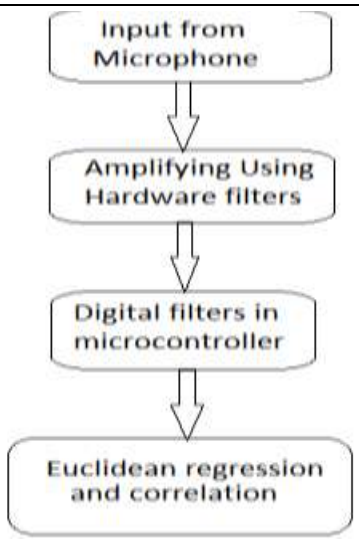

Figure 2: Voice Recognition Security System

Digital Filters

The digital filters are designed using MATLAB 6.5. The bandwidth of the filter is used to calculate the coefficient of the respective filters. In MATLAB, a Chebyshev fourth order high pass filter is used in place of two-second order high pass filters.

Fingerprint/Voiceprint Analysis

The voiceprint and fingerprint would be a combination of the results of the digital filters It is a well-known fact that the voiceprint of even two identical words results in two distinct speech spectrums. In addition, if a single person says a word twice as accurately and similarly as possible, the resulting range would be different. Thus, the difference is fed to the microcontroller in this project to determine if it is capable of detecting a shift in both voice samples. Euclidean distance is known as the summation of the squares of the difference between two phrases/words.

Function Written in Code

Timer 0 is set to use a feature in the microcontroller is programmed. After processing the samples via the analogue to digital converter, the resulting digital data will be filtered using a digital filter. Now the model would use a voiceprint or fingerprint for each word to determine an analysis. The comparator will equate the current voiceprint with the previously stored one.

In order to biometrically authenticate users, this is used. Accepting or denying a speaker's argument is the opposite of speaker verification. It is used to seal an argument's claim of credibility. The Speaker's protocol is commonly referred to as the open mode. Identity verification system, which uses voice recognition on service-specific speakers, is the most important feature of any voice recognition system. The voice recognition is affected by separation of structure and independent voice recognition. This is what the speaker refers to when making the point in the text.

If the words used by the speaker are identical to those that are inputted into the Text Retrieval Software, they are known as the speech data. Instead, it can be argued that a random set of words is called the Voice-independent set of words. There are three basic ways to tell if a voice recognition system is successful: You will find an answer in the text. The Text Collection and the Metaphor are each on their own, while Open Voice and Closed Voice and Closed Statement each seek acknowledgement. The sound of speaking was digitally transformed into an electrical signal using a microphone. Function of a sound card serves Take analogue signal and make it digital with this speech signal, the sound card is able to store and play.

Biometric authentication is now available as an API, so developers can use it as a means of ensuring protection. In this case, biometric authentication means an additional security measure to ensure that only the users' smartphones can access their data. To monitor the smart home, you must register fingerprints on each screen.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

IV. RESULTS & DISCUSSION

The two voice recognition tests are set up to see how well it works. One of the experiments will measure my speech, and another will use other people's as a guide. At any given moment, his/her voice will create a pattern and referred to as a voice waveform. Every voice has a distinct sound. The verification experiments therefore use the two different experiment procedures, the first to test the validity of the analysis, and the second to improve it.

In audio recording, the microphone picks up the user's voice. The voice sample rate is 10000 Hertz and the time is one second. See how the word "HELLO" is said from the microphone in Figure 3.

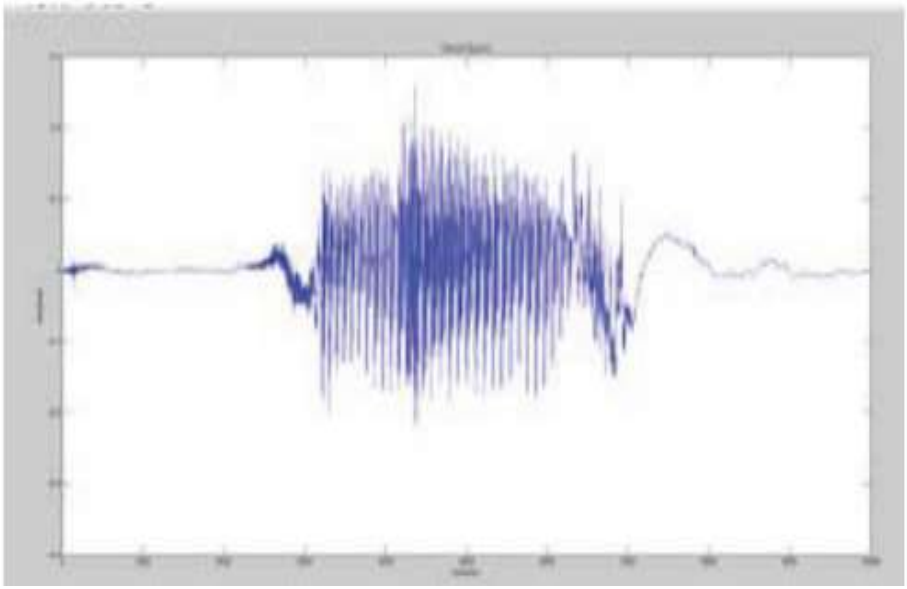

Figure 3: The word "HELLO" as voice input signal.

The input signal is one second of silence, in which the output signal will extract only the actual voice. If the noise is detected in Figure 3, ignore it and move on to Figure 4

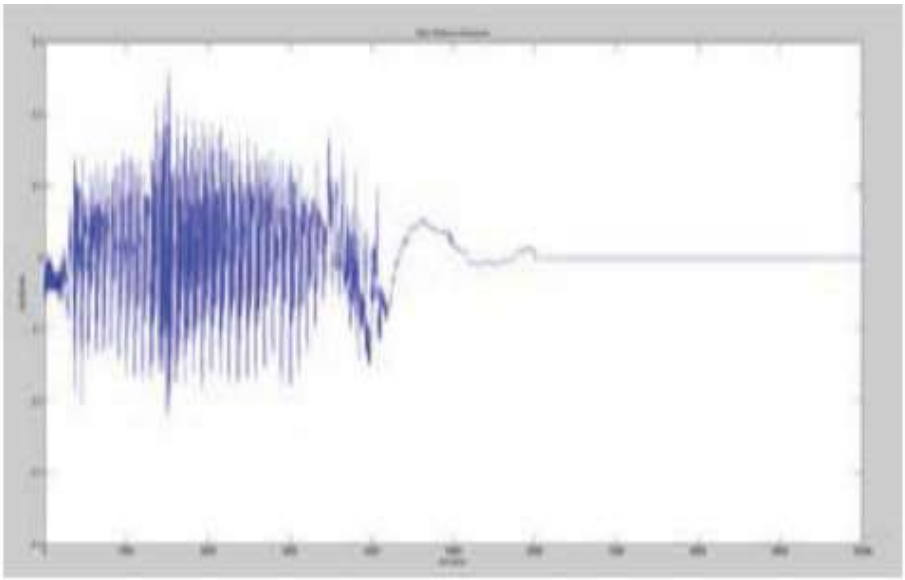

Figure 4: The word "HELLO" after silence detection.

The Hamming window is used to smooth out the input voice signal after extracting the real signal by silence detection. The term "HELLO" after the Hamming window is shown in Figure 4. The signal is on the time domain after it smoothes the input voice signal. The Fast Fourier Transform (FFT) changes the voice input from the temporary field to the frequency field. The term HELLO in the frequency domain is shown in Figure 5.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

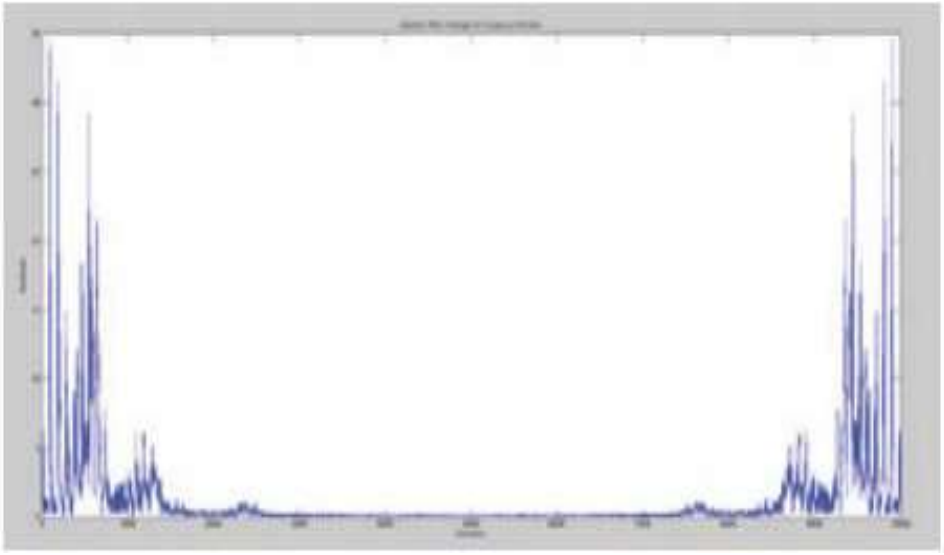

Figure 5: The word "HELLO" after Fast Fourier Transform

This experiment is done with ten separate users, the authenticated user is one person and the other persons are other persons. The voice recognition system recognizes the admin's voice correctly among ten people of different sex or age with the authenticated user. The difference between sex and age is whether the exactness of speech recognition can be affected.

V. CONCLUSION

This article discusses these principles and reviews recent developments. This article compares different methods for creating grammar based on the translation process and grammar. Speech recognition is a type of human speech designed primarily to interpret concepts, sentences, and continuously identify speakers based on the information contained in that speech. This approach allows the use of the speaker's voice, facilitating character identification. Voice provides access to management services such as e-commerce, voice recognition, mobile marketing, automation, home automation, and protection technology. The aim of this work is to recognize speech symbols using three different extraction methods: hardware, digital filters, and Euclidean regression. The speech signal is cleared of noise to eliminate hardware and digital filters, two traditional speech enhancement techniques. Once noise is removed from the speech signal, it is enhanced using digital filters and speech is extracted using feature extraction technology and information speech to best see the accuracy of the stored patterns and actual speech counts. Experimental results show that this strategy can achieve global performance while maintaining accuracy.

VI. REFERENCES

[1] R. Chakroun and M. Frikha, "Efficient text-independent speaker recognition with short utterances in both clean and uncontrolled environments," Multimedia Tools and Applications, vol. 79, no. 29, pp. 21279–21298, Aug. 2020, doi: 10.1007/s11042-020-08824-7.

[2] O. Mamyrbayev, A. Toleu, G. Tolegen, and N. Mekebayev, "Neural architectures for gender detection and speaker identification", Cogent Engineering, vol. 7, no. 1, p. 1727168, Jan. 2020, doi: 10. 1080/23311916. 2020. 1727168.

[3] A. Rinoshika and H. Rinoshika, "Application of multi-dimensional wavelet transform to fluid mechanics", Theoretical and Applied Mechanics Letters, vol. 10, no. 2, pp. 98–115, Jan. 2020, doi: 10. 1016/j.taml.2020.01.017.

[4] N. Holighaus, G. Koliander, Z. Průša, and L. D. Abreu, "Characterization of Analytic Wavelet Transforms and a New Phaseless Reconstruction Algorithm," IEEE Transactions on Signal Processing, vol. 67, no. 15, pp. 3894–3908, Aug. 2019, doi: 10.1109/TSP.2019.2920611.

[5] W. Helali, Z. Hajaiej, and A. Cherif, "Automatic Speech Recognition System Based on Hybrid Feature Extraction Techniques Using TEOPWP for in Real Noisy Environment," IJCSNS - International Journal of Computer Science and Network Security, vol. 19, no. 10, pp. 118–124, Oct. 2019.

[6] A.Mnassri, M. Bennasr, and C. Adnane, "A Robust Feature Extraction Method for Real-Time Speech Recognition System on a Raspberry Pi 3 Board," Engineering, Technology & Applied Science Research, vol. 9, no. 2, pp. 4066–4070, Apr. 2019.

[7] S. N. Truong, "A Low-cost Artificial Neural Network Model for Raspberry Pi," Engineering, Technology & Applied Science Research, vol. 10, no. 2, pp. 5466–5469, Apr. 2020.

[8] A.Koduru, H. B. Valiveti, and A. K. Budati, "Feature extraction algorithms to improve the speech emotion recognition rate," International Journal of Speech Technology, vol. 23, no. 1, pp. 45–55, Mar. 2020, doi: 10. 1007/s10772-020-09672-4.

[9] S. Zhu, C. Xu, J. Wang, Y. Xiao, and F. Ma, "Research and application of combined kernel SVM in dynamic voiceprint password authentication system," in 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), May 2017, pp. 1052–1055, doi: 10. 1109/ICCSN. 2017. 8230271.

[10] Q. Li et al., "MSP-MFCC: Energy-Efficient MFCC Feature Extraction Method with Mixed-Signal Processing Architecture for Wearable Speech Recognition Applications," IEEE Access, vol. 8, pp. 48720– 48730, 2020, doi: 10.1109/ACCESS.2020.2979799.