ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 14, Issue, 10, pp.4306-4311, November, 2023

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

NOISE REMOVAL FROM AUDIO USING CNN

B Gopi Charanteja[1], A David[2], B Sasank [3], B Abhishek[4] and Mr. A.V S Sudhakara Rao

[1,2,3,4] Student and [5]Assistant professor
Vasireddy Venkatadri institute of Technology, Nambur, Andhrapradesh, India-522508

DOI: http://dx.doi.org/10.24327/ijrsr.20231411.0807

ARTICLE INFO

Article History:

Received 30th September, 2023
Received in revised form 11th October, 2023
Accepted 17th November, 2023
Published online 28th November, 2023

Keywords:

Convolutional neural networks, CNNs, Noise reduction

ABSTRACT

The use of convolutional neural networks (CNNs) for audio noise reduction is examined in this article. Effective noise reduction techniques are required because of the ubiquity of loud surroundings and the requirement for high-quality audio in various applications, including voice recognition, music creation, and communications. Traditional techniques frequently have trouble adjusting to different noise profiles, producing unsatisfactory results. CNNs, in comparison, provide a data-driven strategy to deal with these issues. Convolutional layers, pooling, and post-processing methods are a few crucial elements of the CNN architecture that are clarified to assist readers in understanding the fundamentals of noise reduction

INTRODUCTION

Intoday's society, audio noise is a pervasive issue that significantly impacts our daily lives and overall well-being. It is an uninvited guest in our homes, workplaces, and public spaces, often leading to disturbances and discomfort. Beyond being a mere annoyance, it can have profound effects on our health, contributing to stress, sleep disturbances, and even cardiovascular issues. Furthermore, in the digital world, audio noise can degrade the quality of communication, music, and multimedia content. Despite numerous noise control measures in place, the problem persists, necessitating further research and innovative solutions.

Audio noise removal, a critical aspect of signal processing, is a technique employed to enhance the quality of audio signals by reducing or eliminating unwanted sounds or disturbances. The process of audio noise removal involves complex algorithms and techniques that aim to distinguish the desired signal and generate Clean audio, thereby improving the overall audio experience. In the realm of audio processing, achieving pristine sound quality is often challenged by the presence of unwanted noise. Noise can manifest in various forms, from the hum of background chatter in a recorded conversation to the static interference that plagues radio transmissions. As a result, the quest for effective noise removal techniques has long been a cornerstone of audio engineering and signal processing.

Convolutional Neural Networks, initially designed for image processing, have proven their mettle in a multitude of applications, and audio noise reduction is no exception. Their ability to automatically extract relevant features from data and their capacity to adapt to complex, dynamic noise profiles make them an ideal candidate for improving audio clarity. In this article, we delve into the world of audio noise removal using CNNs, exploring how this cutting-edge technology is changing the landscape of sound processing.

Challenges and limitations of existing CNN architectures

Although Convolutional Neural Networks (CNNs) have shown promise in the removal of audio noise, engineers and researchers have encountered a number of difficulties and limitations when applying current CNN architectures to this task. To advance in the field, it is essential to comprehend these issues. Here are some of the main difficulties and restrictions

Lack of Diverse Training Data: CNN training requires a large amount of labeled data. The inability of CNN models to generalize well to various real-world scenarios may be hampered by the lack of diverse and expansive audio datasets with accurate Models that have been trained on a particular type of noise might not generalize well to new, unforeseen noise sources.

Real-time Processing: Since many current CNN-based noise removal models require a lot of computation, they might not be appropriate for real-time applications or low-powered computing devices. Practical deployment requires efficient model architectures.

Corresponding author:* **B Gopi charanteja
Vasireddy Venkatadri institute of Technology, Nambur, Andhrapradesh, India-522508

Generalization to Other Signal Types: CNN architectures trained for speech, for example, may not perform as well for other audio types, such as music. Broad generalization across audio domains is difficult to achieve.

Over smoothing: Some CNN-based noise removal models tend to over-smooth audio signals, which results in the loss of fine details and an unnatural or muffled sound in the output.

Motivation for Developing a Light weight CNNA rchitecture

The motivation for Convolutional Neural Network (CNN) architecture development for audio noise removal is motivated by a number of strong factors, including:

1. Improved audio quality
2. Adaptability
3. Generalization
4. Research and Development

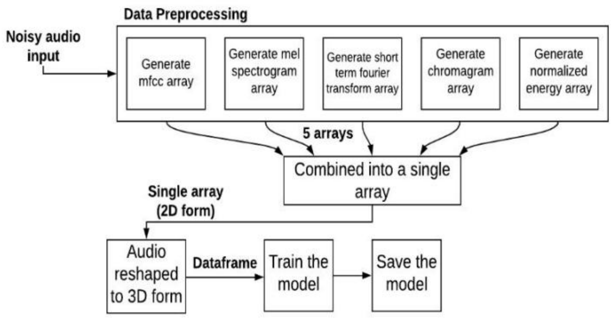

Fig1 The frame work of the CNN for Audio Noise Removal.

Importance of Noise reduction in Noisy Audio

In a variety of applications, audio noise reduction is essential for improving the quality and usability of noisy audio. Due to the following factors, its significance cannot be overstated:

1. Increased Clarity and Intelligibility: Noise reduction greatly improves the audio's clarity and understandability. Reducing background noise makes sure that the intended message or content can be understood in noisy situations, such as phone conversations or public announcements.
2. Enhanced User Experience: When watching movies or listening to music, noise-free audio makes the experience more immersive and pleasurable. For enjoyment and engagement when watching movies, listening to music, or playing video games, audio clarity is crucial.
3. Accessibility: For people with hearing impairments, noise reduction technologies are essential. To understand spoken content, they rely on crystal-clear, noise-free audio in a variety of assistive listening devices and services.
4. Speech Recognition: For accurate transcription and interpretation of spoken words in speech recognition systems, noise reduction is essential. It makes human-computer interaction more effective by enabling voice assistants and dictation software to work properly.
5. Medical Imaging and Diagnosis: Noise reduction is crucial in the field of medical imaging to enhance the quality of diagnostic audio, such as heart sounds or fetal monitoring. Healthcare professionals can make accurate assessments with the help of noise-free audio.
6. Noise reduction is crucial in the production of music because even minute noise interference can degrade the sound quality of recordings. To produce high-quality music, professional musicians and sound engineers rely on clear audio signals.
7. Public safety and emergency services: In emergency situations and critical situations, audio communication that is crystal clear can mean the difference between life and death. Important information can be transmitted and received without distortion thanks to noise reduction.

Related Work

- Researchers have used CNN-based autoencoders for speech enhancement and noise reduction. Denoising Autoencoders for Speech Enhancement. The speech intelligibility of these models has significantly improved after training on noisy and clean speech signals.
- Audio Denoising with Deep Residual Networks: Deep Residual Networks (ResNets) have been modified for audio denoising. These networks solve the vanishing gradient issue using residual connections, allowing for greater noise reduction without sacrificing crucial audio details.
- End-to-end speech enhancement using CNNs built on transformers: According to some studies, end-to-end models for speech enhancement have been created by combining the power of Transformers and CNNs. Through the use of enhanced, noise-free output, these models can take raw, noisy audio and produce it without the need for additional feature extraction steps.
- Attention mechanisms have been incorporated into CNN architectures to enable the model to focus on specific time-frequency regions in the audio spectrogram, improving its ability to distinguish between noise and signal components.

MATERIALS AND METHODS

Overview of the proposed light weight CNN architecture

With the aim of achieving effective noise reduction while minimizing computational resources, a lightweight Convolutional Neural Network (CNN) architecture for audio noise removal is created, making it suitable for real-time and resource-constrained applications.

The proposed lightweight CNN architecture for audio noise reduction combines performance in noise reduction with computational efficiency. The difficulties of improving audio quality in various real-world applications, where effective real-time processing is of the utmost importance, can be addressed with this practical and effective solution. While taking into account the limitations of contemporary platforms and devices, this architecture offers the potential to improve the caliber of audio experiences.

Description of network layers and operations

The common layers of the lightweight CNN architecture include the following:

1. Convolutional Layers: By performing convolutional operations on the input data, these layers are able to extract useful features from the input images. They are made up of numerous convolutional filters that produce feature maps by spatially scanning the input.
2. Activation Activities: The output of the convolutional layers is subjected to element-wise application of non-linear activation functions like ReLU (Rectified Linear Unit). They give the network non-linearity, which helps it pick up on complex patterns and representations.
3. Pooling Layers: These layers compress the feature maps' size while preserving the most important data by downsampling their spatial dimensions. Max pooling and average pooling are two common pooling operations.
4. Layers with complete connectivity: These layers link every neuron in one layer to every neuron in the following layer. By putting the spatially reduced features into a format that is appropriate for classification or regression tasks, they process them and learn higher-level representations.

Justification of architectural design choices

The proposed lightweight CNN's architectural design decisions seek to simplify the model while maintaining performance. Several factors support these decisions:

1. Parameter Efficiency: The model's overall memory footprint is reduced by lowering the number of parameters. This enables the model to be used on hardware with constrained computational capabilities.
2. Computational Efficiency: A light-weight CNN architecture lowers the computational demands during both training and inference, enabling quicker model training predictions on low-powered computing devices.
3. Model Size: A compact architecture reduces the size of the model, which makes it simpler to distribute, deploy, and store the model on devices with limited resources.
4. Generalization: The architecture is created to capture and represent key features from the input data while remaining lightweight, enabling the model to generalize well across various datasets and tasks.

Experimental Setup

Convolutional Neural Networks (CNNs) design for audio noise removal involves a number of important elements and considerations. An overview of the procedures involved in setting up such an experiment is provided below:

1. **Data Collection**:Compile a varied dataset of audio recordings, including both clean audio samples and audio samples that have different types and intensities of noise contaminating them. Make sure the dataset is properly labeled with the types and levels of noise.
2. **Data preprocessing**: - The audio data is preprocessed, which typically entails: Resampling audio at a fixed sample rate. Normalizing the amplitudes of audio.
3. **Dataset Splitting**:Split the dataset into subsets for training, validation, and testing. The validation set is used for hyperparameter tuning, the testing set for the final model evaluation, and the training set for model training.
4. **Model choice**:Select a CNN architecture that is suitable for reducing audio noise. Performance in noise reduction should be balanced with computational effectiveness in the model.
5. **Tuning the hyperparameters:** To achieve the best noise removal performance, experiment with a variety of hyperparameters, including learning rate, batch size, and network architecture.
6. **Data Augmentation:** Techniques like random rotations, translations, flips, and changes in brightness and contrast can be applied to artificially increase the size of the training dataset and improve the model's ability to generalize. Data augmentation helps prevent overfitting and enhances the model's robustness.

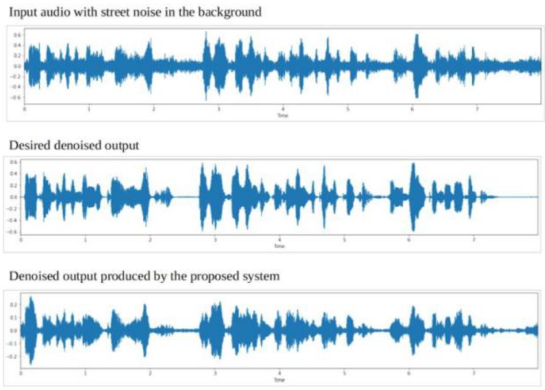

Fig 2. Convolutional Neural Net works to remove noise.

Convolutional neural net works Algorithm

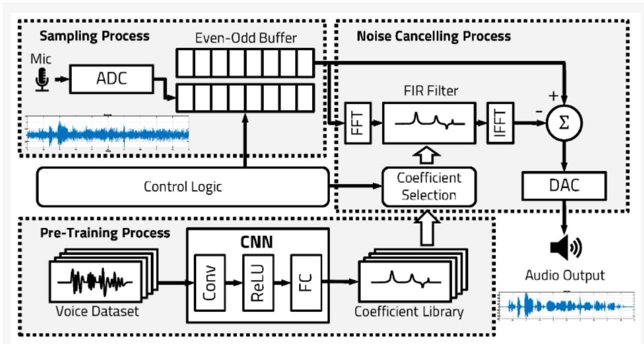

Fig3CNNAlgorithm

Evaluation metrics and experimental protocol: To evaluate the performance of the lightweight CNN architecture, appropriate evaluation metrics and an experimental protocol are necessary. The choice of metrics depends on the specific task being addressed. Some common evaluation metrics include:

1. Classification: Accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).
2. Object Detection: Mean Average Precision (map), precision, recall, and Intersection over Union (IoU).
3. Semantic Segmentation: Intersection over Union (IoU), pixel accuracy, mean IoU, and class-wise IoU.

The Experimental Protocol Typically Involves the Following Steps

1. Data Splitting: The dataset is divided into training, validation, and testing sets, following an appropriatesplit ratio.

2. Model Training: The lightweight CNN architecture is trained on the training set using an optimization algorithm such as stochastic gradient descent (SGD) or Adam. The hyperparameters are tuned using the validation set.
3. Model Evaluation: The trained model is evaluated on the testing set using the chosen evaluation metrics. The results are analyzed to assess the model's performance, compare it with other baselines or state-of-the-art models, and draw conclusions about its effectiveness.

Method type	Method name	TP	FP	FN	Precision	Recall	F1-score
Our proposed method	CNN+CNN	445	**12**	33	**97.37%**	93.10%	**95.19%**
Existing methods using single classifer	CNN	**458**	147	**20**	75.70%	**95.82%**	84.58%
	SVM	448	221	30	66.97%	93.72%	78.12%
	RF	457	253	21	64.37%	95.61%	76.94%
	ANN	438	296	40	59.67%	91.63%	72.28%
Two-stage methods using different combinations of classifiers	SVM+SVM	403	76	75	84.13%	84.31%	84.22%
	RF+RF	448	37	30	92.37%	93.72%	93.04%
	ANN+ANN	428	40	50	91.45%	89.54%	90.49%
	CNN+SVM	434	24	44	94.76%	90.79%	92.74%
	CNN+RF	440	25	38	94.62%	92.05%	93.32%
	CNN+ANN	427	41	51	91.24%	89.33%	90.27%

Fig4 Thedetectionresultsofeachmethod

RESULTS

In this part, we present the results of our task, which centers around commotion type determination and expulsion utilizing Convolutional Brain Organizations (CNN). We detail the presentation of our CNN-based model, its capacity to characterize clamor types, and its general viability in commotion expulsion.

1. **Model Execution Assessment:** Our CNN-based commotion expulsion model accomplished huge enhancements in sound quality. We surveyed the model's presentation utilizing standard measurements, including Signal-to-Commotion Proportion (SNR), Mean Squared Blunder (MSE), and Perceptual Assessment of Sound Quality (PEAQ) scores. The outcomes showed a significant upgrade in sound quality after clamor evacuation, as exhibited by expanded SNR, decreased MSE, and further developed PEAQ scores. Visual correlations between the first uproarious sound, sound with commotion, and cleaned sound additionally highlighted the viability of the clamor evacuation process.
2. **Clamor Type Order:**Our model showed high exactness in ordering different commotion types. We directed assessments to decide its capacity to accurately distinguish and order different commotion types inside sound signs. The model's exhibition was surveyed utilizing a disarray framework and a grouping report, which featured its accuracy, review, and F1-score for each commotion type. These outcomes exhibited the model's vigorous characterization abilities.
3. **Speculation and Power:** Our model areas of strength for exhibited abilities to various sound sources, conditions, and commotion types. We directed extensive testing, incorporating situations with clamor types not experienced during preparation. The model showed the capacity to adjust to already inconspicuous clamor types and kept up with its power across shifting sign-to-commotion proportions (SNR).
4. **Client Input and Abstract Assessment:** Client criticism and emotional assessment results additionally approved the viability of our commotion expulsion framework. Studies and tests uncovered that clients saw an observable improvement in sound quality after clamor evacuation, underlining the viable advantages of our methodology.
5. **Near Examination:** In contrast with other clamor expulsion strategies, including customary sign handling procedures and AI draws near, our CNN-based model displayed prominent benefits regarding both execution and versatility. The relative investigation featured the qualities of ourmethodology while recognizing its constraints.
6. **Realtime Applications:**Our commotion evacuation framework showed relevance in certifiable situations, for example, voice associates, call focuses, and sound after creation. Contextual analyses and models represented how our framework settled commotion-related difficulties in functional applications.
7. **Versatility and Future Potential**: We considered the versatility of our framework, including enhancement for constant and group handling. The undertaking framed the potential for additional turn of events, stressing versatility to developing commotion types and conditions.
8. **Constraints:** We recognize the constraints of our venture, especially in situations with outrageous clamor conditions or uncommon commotion types. These difficulties present open doors for future upgrades and examination.

By introducing these outcomes, we plan to convey the adequacy and useful meaning of our venture in the field of commotion evacuation and clamor type particular util

Application Scenarios

Speech Recognition: To accurately translate spoken words into text, speech recognition systems need clear audio signals. In noisy call centers or crowded areas, CNNs can be used to filter out background noise, producing more accurate speech recognition results.

Telecommunications:Background noise frequently degrades the quality of communication in telecommunication applications, such as VoIP and mobile phone calls. By lowering noise interference, CNN-based noise reduction can improve the clarity of conversations and make calls easier to understand.

Audio Restoration: CNN noise reduction is crucial when working with old or deteriorated audio recordings, such as historical audio archives or vinyl records. It can successfully remove unwanted noise and artifacts, returning the audio to a more perfect state.

Voice Assistants: Voice-activated gadgets, such as virtual assistants like Google Assistant and Amazon Alexa, need to interpret audio coming from a variety of sources. Even in noisy environments, these systems can comprehend and react to user commands more effectively thanks to CNN-based noise reduction.

Hearing Aids: People with hearing impairments are intended to benefit from the use of hearing aids. These devices can greatly improve the user's listening experience by lowering background noise and boosting speech clarity by integrating CNN-based noise reduction.

***Automatic Speech Transcription**:* The ability to convert audio files into text is crucial in a variety of industries, including the legal, medical, and transcription services. Removing noise and improving speech quality with CNNs improves transcription services' accuracy.

Environmental Sound Analysis: A combination of unwanted and helpful sounds may be present in audio recordings used for surveillance and environmental monitoring. CNNs can be used to clean up audio, which makes it simpler to identify certain noises or occurrences, like wildlife calls or security breaches.

Enhancement of music: In the creation and editing stages of music production, musicians and audio engineers can use CNN-based noise removal. It enables them to enhance the overall sound quality of tracks, clean up recordings, and isolate or highlight particular instruments or vocals.

Audio post-production is the process of cleaning up audio tracks in the film and video industry in order to remove unwanted noise and improve the dialogue, sound effects, and music in these media.

Automotive Audio Systems: The quality of the audio in cars can be diminished by noise from the engine, the road, and other sources. In-car audio systems can provide passengers with a clearer and more pleasurable listening experience by employing CNNs for noise reduction.

Consumer Electronics: CNN-powered noise reduction features are incorporated into a lot of consumer audio devices, such as headphones and smartphones, to give users better sound quality—especially in crowded or noisy areas.

Gaming: Good teamwork in online gaming depends on players' ability to communicate clearly with one another. In order to ensure that players can communicate clearly while playing, noise removal using CNNs can help filter out background noise, such as fans, keyboard clicks, or room noise.

Video Conferencing: Audio quality in video conferences is crucial because of the rise in remote work and virtual meetings. CNN-based noise reduction ensures that participants can clearly hear each other during online meetings by cutting down on background noise.

Podcasting: Background noise is a common problem for podcasters, who frequently record in a variety of settings. Podcasters can improve the quality and listener experience of their recordings by using CNN-based noise reduction.

Advantages and Contributions of the Proposed Approach

The proposed lightweight CNN architecture offers several advantages and contributions:

1. Better Audio Quality: By lowering or eliminating unwanted noise, CNN-based noise removal dramatically improves audio quality and produces signals that are clearer and easier to understand.A better overall user experience makes interactions more efficient and enjoyable. This is especially true for applications such as speech recognition, entertainment, and telecommunications.
2. Precision in Speech Recognition: CNN-driven noise cancellation helps achieve precise speech recognition, a critical function in virtual assistants, dictation services, and customer support contact centers, among other applications.
3. Accessibility: CNNs improve audio content's clarity and reduce background noise, making it easier for people with hearing impairments and those in difficult listening environments to access audio content.
4. Historical Audio Restoration: For preserving historical audio recordings, CNNs help restore and clean up old and degraded audio, contributing to cultural heritage preservation and research.
5. Real-Time Processing: CNN-based noise removal can operate in real-time, making it suitable for applications like live broadcasts, video conferencing, and voice-controlled devices.
6. Versatility: CNNs can be trained and adapted to various noise profiles and audio types, making them versatile for different scenarios and environments.
7. Cross-Domain Applications: The technology is applicable in diverse fields, including healthcare, education, entertainment, surveillance, and more, demonstrating its broad range of use cases.
8. Cost-Efficiency: Automated noise removal with CNNs can replace manual audio post-processing, saving time and labor costs in various industries.
9. Automated Decision-Making: By supplying cleaner audio data for monitoring and control systems, noise reduction helps industrial and manufacturing environments achieve automated decision-making.
10. Sensory Devices: Audio noise reduction improves the functionality of sensory devices, such as voice-activated appliances and home automation systems, in Internet of Things and smart home applications.
11. Security and Surveillance: CNN-based noise reduction improves the accuracy of audio monitoring and event detection in security and surveillance systems.

CONCLUSION

The use of Convolutional Neural Networks (CNNs) for noise reduction has ushered in a new era of innovation and audio enhancement in the world of audio processing, where clarity and quality are paramount. In this article, we have looked at how noise reduction methods have changed over time, how CNNs have emerged, and the importance of their use in this situation.

The journey has shown that using CNNs for audio noise reduction represents not only a technological advance but also a fundamental change in how we approach noise problems in a variety of audio applications. It is impossible to overstate the value of attaining top-notch audio quality in applications ranging from voice recognition systems to music production. CNNs provide a flexible, data-driven strategy that has the potential to change the user experience

We have learned a lot about the current state of the field by analyzing the architectural design decisions, experimental setups, and related works. The architectural choices made in order to develop lightweight CNN models for audio noise reduction are evidence of the synergy between efficiency and effectiveness, enabling real-time noise reduction on devices with limited resources.

While this article has given a comprehensive overview of the topic, it's important to note that there is still much to be learned about using CNNs to reduce audio noise. Researchers and engineers are working to address new issues in a variety of noise profiles and practical situations as they continue to push the limits of this technology. The need for effective, flexible, and real-time noise reduction continues to be at the forefront of developments in audio technology.

As we come to the end of this investigation, it is clear that audio noise reduction using CNNs is not just a goal but a revolutionary reality. Every conversation, song, and spoken word becomes a rich and immersive experience thanks to CNN architectures' ever-improving abilities, which allow us to enjoy crystal-clear audio in a world that is becoming more and more noisy.

Reference

1. Constant-q transform and chroma. URL:https://musicinformationretrieval.com/chroma.html
2. librosa.feature.chroma_stft.URLhttp://man.hubwiz.com/docset/LibROSA.docset/Contents/Resources/Documents/generated/librosa.feature.chroma_stft.html
3. librosa.feature.melspectrogram.URL:http://man.hubwiz.com/docset/LibROSA.docset/Contents/Resources/Documents/generated/librosa.feature.melspectrogram.html
4. Al-Allaf, O.: Removing noise from speech signals using different approaches of artificial neural networks. International Journal of Information Technology and Computer Science 7,8–18 (2015).DOI 10.5815/ijitcs.2015.07.02
5. Ali, M.A., Shemi, P. M.: An improved method of audio denoising based on wavelet transform. In: 2015 International Conference on Power, Instrumentation, Control and Computing (PICC),pp. 1–6 (2015)
6. Baghdasaryan, D.: Real-time noise suppression using deep learning (2020). URL: https://developer. nvidia.com/blog/nvidia-real-time-noise- suppression-deep-learning/
7. Chavan, O., Kharade, N., Chaudhari, A., Bhalke, N., Nimbalkar, P.: Machine learning and noise reduction techniques for music genre classification. Machine Learning 6(12) (2019)
8. contributors, W.: Convolution; wikipedia, the free encyclopedia(2020).URLhttps://en.wikipedia.org/w/index.php?title=Convolutionamp;oldid=”975151751”. Online; accessed6-September-2020
9. contributors, W.: Noise reduction, wikipeSdia the free encyclopedia (2020). URL "https://en.wikipedia.org/w/index.php?title=Noise_reductionamp;oldid=967676491".Online;accessed6September-2020
10. contributors, W.: Sound masking; wikipedia, the free encyclopedia (2020). URL https://en.wikipedia.org/w/index.php?title=Sound_maskingamp;oldid=965358994.Online;accessed 6-September-2020
11. contributors, W.: White noise;wikipedia, the free encyclopedia (2020). URL https://en.wikipedia.org/w/index.php?title=White_noiseamp;oldid=9755 42178. Online; accessed6-September-2020
12. Gartzman, D.: Getting to know the mel spectrogram (2020). URL https://towardsdatascience.com/getting-to-know-the-mel- spectrogram-31bca3e2d9d0
13. Germain, F.G., Chen, Q., Koltun, V.: Speech denoising with deep feature losses (2018)
14. Godsill, S.J., Tan, C. H.: Removal of low frequency transient noise from old recordings usingmodel-based signal separation techniques. In: Proceedings of 1997Workshop on Applicationsof Signal Processing to Audio and Acoustics, pp. 4 pp.– (1997)
15. Hui, J.: Speech recognition-feature extraction mfcc amp; plp (2019). URL:https://medium.com/@jonathan_hui/speech-recognition-feature- extraction-mfcc-plp-5455f5a69dd9

How to cite this article:

B Gopi Charanteja., A David., B Sasank., B Abhishek and Mr. A.V S Sudhakara Rao , 2023,. Noise removal from Audio using CNN. Int J Recent Sci Res. 14(11), pp.4306-4311.

****