Speech Enhancement using Filtering Techniques and Machine Learning Approaches

A Project Report Submitted in Partial Fulfillment of the Requirement of the Degree

Of

BACHELOR OF TECHNOLOGY

In

ELECTRONICS AND COMMUNICATION ENGINEERING

by

AKSHIT KAKOTI

(202000024)

Under the guidance of

Dr. Mridusmita Sharma
Assistant Professor
Dept of EE, Assam Engineering College

Mr. Jitendra Singh Tamang
Asst. Prof. ECE. Dept.
Sikkim Manipal Institute of Technology

SIKKIM MANIPAL UNIVERSITY
SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY
MAJITAR, EAST SIKKIM - 737136, May 2024

ASSAM ENGINEERING COLLEGE

(A govt. of Assam Institution)

Dr. Mridusmita Sharma, Assistant Professor

email: mridusmita.ele@aec.ac.in

Department of Electrical Engineering

Date: 30/05/2024

Certificate

This is to certify that **AKSHIT KAKOTI**, Registration No.: **202000024**, an undergraduate student in the Department of ECE of Sikkim Manipal Institute of Technology has done a project work titled **"SPEECH ENHANCEMENT USING FILTERING TECHNIQUES AND MACHINE LEARNING APPROACHES"** under my supervision. This work is carried out in the Department of Electrical Engineering of Assam Engineering College, Jalukbari, Guwahati-13, Assam as a part of his final semester project work. The duration of this work is from 08/01/2024 to 17/05/2024.

Assam Engineering College Guwahati
Electrical & Instrumentation
Estd. : 19

(Dr. Mridusmita Sharma)

Assistant Professor
Electrical Engineering Department
Assam Engineering College
Guwahati- 13
ASSAM, INDIA

Certificate

This is to certify that the project report entitled (**"SPEECH ENHANCEMENT USING FILTERING TECHNIQUES AND MACHINE LEARNING APPROACHES"**) submitted by **(AKSHIT KAKOTI) (202000024)** to Sikkim Manipal Institute Of Technology, Sikkim in partial fulfillment for the award of degree of Bachelor/Master of Technology in Electronics And Communication Engineering, is a bonafide record of the project work carried out by him/her under my guidance and supervision during the academic session January– May, 2024.

29/5/24.

Dr. Mridusmita Sharma

Designation – Assistant Professor

Dept of EE, Assam Engineering College

Assistant Professor,
Electrical Engineering Department
Assam Engineering College
Guwahati-13
ASSAM, INDIA.

Certificate

This is to certify that the project report entitled ("**SPEECH ENHANCEMENT USING FILTERING TECHNIQUES AND MACHINE LEARNING APPROACHES**") submitted by (**AKSHIT KAKOTI**) (**202000024**) to Sikkim Manipal Institute Of Technology, Sikkim in partial fulfillment for the award of degree of Bachelor/Master of Technology in Electronics And Communication Engineering, is a bonafide record of the project work carried out by him/her under my guidance and supervision during the academic session January– May, 2024.

Name of Internal Supervisor	Dr. Bikash Sharma
Designation	Head of the Department
Dept. of E & C Engineering	Dept. of E & C Engineering
Sikkim Manipal Institute of Technology	Sikkim Manipal Institute of Technology

External Examiner

Abstract

Speech enhancement is a fundamental challenge in telecommunications, voice recognition systems, and other applications where clear, intelligible speech is crucial. This project investigates the efficacy of conventional filtering techniques compared to modern machine learning methods, particularly Convolutional Neural Networks (CNNs), for improving the quality of speech signals corrupted by noise. The dataset utilized comprises 3,000 WAV files of spoken numbers, recorded by six distinct speakers and sampled at 8kHz. These recordings underwent preprocessing to remove silent segments and were then augmented with simulated noise to emulate real-world conditions.

The primary objective is to compare the performance of a traditional low-pass filter against a CNN-based approach for denoising. The low-pass filter is designed to mitigate high-frequency noise, while the CNN is trained to discern and suppress noise patterns, thereby enhancing the original speech signal. The study focuses on demonstrating the CNN's superiority in reducing Mean Squared Error (MSE) and error percentage, indicative of its robust speech enhancement capabilities.

Experimental results underscore the CNN's effectiveness in outperforming the low-pass filter, affirming its role in significantly improving audio quality. The CNN architecture includes convolutional layers followed by fully connected layers, optimized with the Adam optimizer and tuned hyperparameters for enhanced learning and batch processing efficiency. The denoising process involves predicting cleaned signals from noisy inputs, which are evaluated through visual and quantitative comparisons with the original recordings.

This research contributes to advancing speech enhancement technologies by validating CNNs as superior alternatives to traditional filtering methods in handling noisy speech signals. The findings highlight the potential of machine learning approaches to enhance audio quality across diverse applications, from telecommunications to automated speech recognition systems.

Acknowledgement

I would like to express our sincere gratitude and appreciation to the Electronics and Communication Department, SMIT for providing the opportunity for a major project and to our HOD Dr. Bikash Sharma for his support. Special thanks to our project guide **Mr. Jitendra Singh Tamang** for his meaningful contribution to the works related to this project and its successful completion. I want to thank all the teaching and non-teaching staff of the ECE Department, Sikkim Manipal Institute of Technology, for providing enormous support to carry out my research work. I would also like to thank all our colleagues in Sikkim Manipal Institute of Technology, for supporting us during our research work.

Akshit Kakoti
(202000024)
Dept. of Electronics and Communication Engineering, SMIT

Abbreviations

Abbreviation	Details
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
SE	Speech Enhancement
MSE	Mean Squared Error
FFT	Fast Fourier Transform
LPF	Low Pass Filter
WAV	Waveform Audio File Format
RMS	Root Mean Square

List of figures

List of tables

Table of contents

Chapter 1

Introduction

1.1 What is Speech Enhancement?

Speech enhancement is an essential field of study, deeply rooted in various applications such as telecommunications, hearing aids, voice-controlled systems, and many more. The primary goal of speech enhancement is to improve the quality and intelligibility of speech signals by reducing noise and distortions that are often present in real-world audio recordings. This task is critical in ensuring clear and effective communication, especially in noisy environments where the quality of speech can be significantly degraded. Traditional filtering techniques have been the cornerstone of speech enhancement for decades. These techniques include methods such as low-pass filters, high-pass filters, and band-pass filters, each designed to target specific frequency components associated with noise. However, these traditional methods often struggle to effectively separate the speech signal from complex noise patterns, leading to suboptimal results.

With the rapid advancement of technology, machine learning has emerged as a powerful tool in various domains, including speech enhancement. Convolutional Neural Networks (CNNs), a type of deep learning model, have shown remarkable success in image and signal processing tasks. Leveraging their ability to learn complex patterns and relationships within data, CNNs have been adapted for speech enhancement, demonstrating superior performance compared to traditional filtering techniques. This project explores the application of CNNs for speech enhancement and compares their effectiveness with traditional filtering methods. By examining the results obtained from both approaches, we aim to provide a comprehensive analysis of their performance in enhancing speech signals.

1.2 Why do we use SE?

Speech enhancement (SE) techniques are utilized to significantly improve the quality and intelligibility of speech signals in diverse scenarios. One primary application of SE is noise reduction, where these techniques effectively mitigate background noise, thereby enhancing the clarity and comprehensibility of speech, especially in environments characterized by high levels of ambient noise such as crowded public spaces or vehicles. Moreover, SE plays a crucial role in improving communication over various telecommunication networks, including mobile phones and Voice over IP (VoIP) services, by ensuring clearer and more intelligible speech transmission.

Furthermore, SE contributes to enhancing the accuracy of Automatic Speech Recognition (ASR) systems by providing cleaner speech signals, which are essential for applications such as voice commands and dictation software. Additionally, SE techniques aim to elevate overall audio quality in recordings, broadcasts, and public address systems by improving the fidelity and perceptual quality of speech. Moreover, these enhancements extend the accessibility of speech signals, benefiting individuals with hearing impairments by facilitating easier amplification or processing through devices like hearing aids or cochlear implants.

In your project, employing SE techniques in conjunction with machine learning methods like Convolutional Neural Networks (CNNs) aims to achieve superior noise reduction and speech enhancement capabilities, potentially surpassing traditional filtering techniques in addressing the challenges posed by complex audio environments.

1.3 Objective

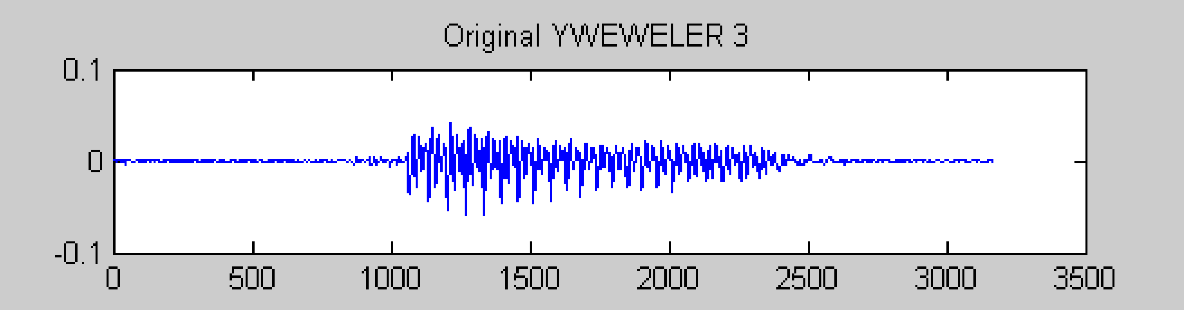
he objective of this project is to evaluate and compare the effectiveness of traditional filtering techniques and machine learning approaches, specifically Convolutional Neural Networks (CNNs), in enhancing speech signals. The project involves several key steps, starting with the acquisition and preprocessing of a dataset consisting of recordings of spoken numbers in WAV format. Noise is then added to these recordings to simulate real-world noisy environments. Both traditional filtering techniques and the CNN-based machine learning approach are applied to the noisy signals to obtain cleaned speech signals. The performance of each approach is evaluated using metrics such as Mean Squared Error (MSE) and error percentage. The ultimate goal is to demonstrate that the machine learning approach yields better results than traditional filtering techniques, thereby highlighting the potential of CNNs in the field of speech enhancement..

Chapter 2

Dataset

2.1 Description

The dataset used in this project is a fundamental collection of audio recordings designed to facilitate the development and evaluation of speech enhancement techniques. It consists of recordings of spoken numbers in WAV format, sampled at 8kHz. The dataset includes recordings from six different speakers, resulting in a total of 3,000 files, with each digit from 0 to 9 being recorded 50 times by each speaker. The recordings have been meticulously edited to remove most of the silence at the beginning and end of each file, ensuring that the dataset is clean and ready for processing. This dataset provides a diverse set of examples, covering variations in pronunciation, intonation, and other speech characteristics, making it an ideal choice for evaluating speech enhancement techniques.



: Fig 2.1: Original waveform

.

2.1 Data Preprocessing

Preprocessing is a critical step in preparing the dataset for analysis. In this project, preprocessing involves several key tasks aimed at simulating real-world noisy environments and ensuring that the dataset is suitable for both filtering techniques and machine learning approaches. Initially, noise is added to the clean speech signals to create noisy versions of the recordings. This step is crucial as it mimics the conditions in which speech enhancement techniques are typically applied. The noise level is carefully controlled to ensure that it is noticeable but not overwhelming, thereby creating a realistic challenge for the enhancement algorithms. The following flowchart illustrates the preprocessing workflow:

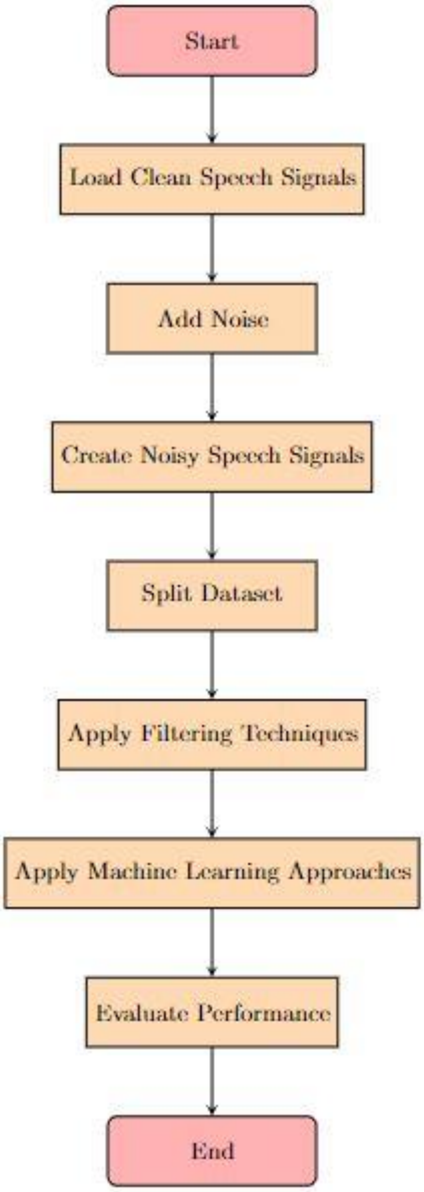


Fig 2.1 Preprocessing work-flow

Chapter 3

Methodology

3.1 Adding Noise

Adding noise to the original audio signals is a crucial step in this project, as it creates the noisy speech signals that need to be enhanced. The process of adding noise involves superimposing a random noise signal onto the original clean speech signal. The noise level is carefully chosen to ensure that it is significant enough to degrade the quality of the speech signal, thereby providing a realistic challenge for the enhancement algorithms. This step is essential for simulating real-world conditions where speech signals are often contaminated with various types of noise.

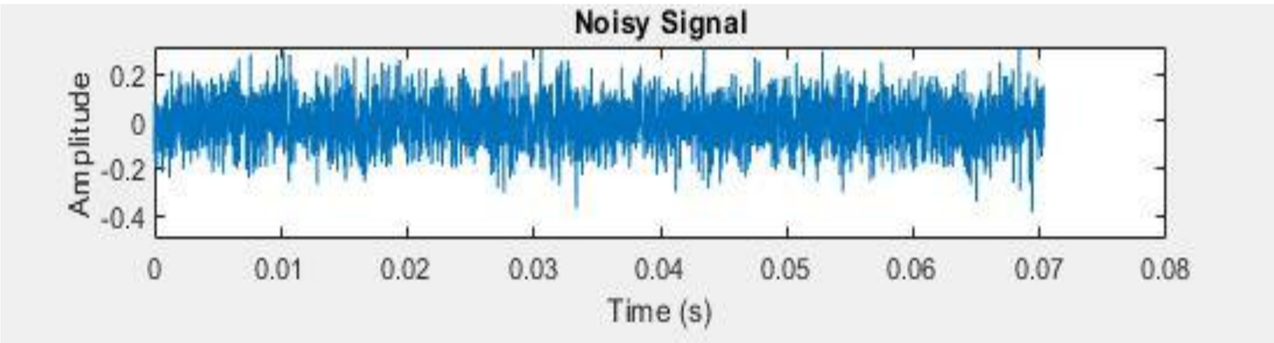


Fig 3.1 Noisy Signal

3.2 Filtering Technique

Traditional filtering techniques have been widely used for speech enhancement. These techniques include low-pass filters, high-pass filters, and band-pass filters, each designed to target specific frequency components associated with noise. Low-pass filters allow frequencies below a certain cutoff frequency to pass through while attenuating higher frequencies. High-pass filters do the opposite, allowing higher frequencies to pass through while attenuating lower frequencies. Band-pass filters allow frequencies within a certain range to pass through while attenuating frequencies outside this range. These filters are designed based on the characteristics of the noise present in the signal. In this project, appropriate filters were designed and applied to the noisy speech signals to obtain cleaned signals. The effectiveness of these filtering techniques was evaluated based on the quality of the cleaned signals.

3.3 Machine Learning Approach

The machine learning approach used in this project involves the application of Convolutional Neural Networks (CNNs) for speech enhancement. CNNs are a type of deep learning model that have shown remarkable success in image and signal processing tasks. They are capable of learning complex patterns and relationships within the data, making them well-suited for tasks such as noise reduction in speech signals. The CNN architecture used in this project consists of several layers, including convolutional layers, ReLU (Rectified Linear Unit) activation layers, and fully connected layers. The network is designed to learn the mapping from noisy speech signals to clean speech signals. The architecture of the CNN and the training process are illustrated in the following flowchart:

The CNN is trained using the noisy speech signals as input and the corresponding clean speech signals as output. The training process involves optimizing the network parameters to minimize the difference between the predicted clean signals and the actual clean signals. The trained network is then used to denoise the speech signals, resulting in cleaned signals that are expected to have higher quality compared to those obtained using traditional filtering techniques.

The proposed block diagram of the system is given below for a better understanding:

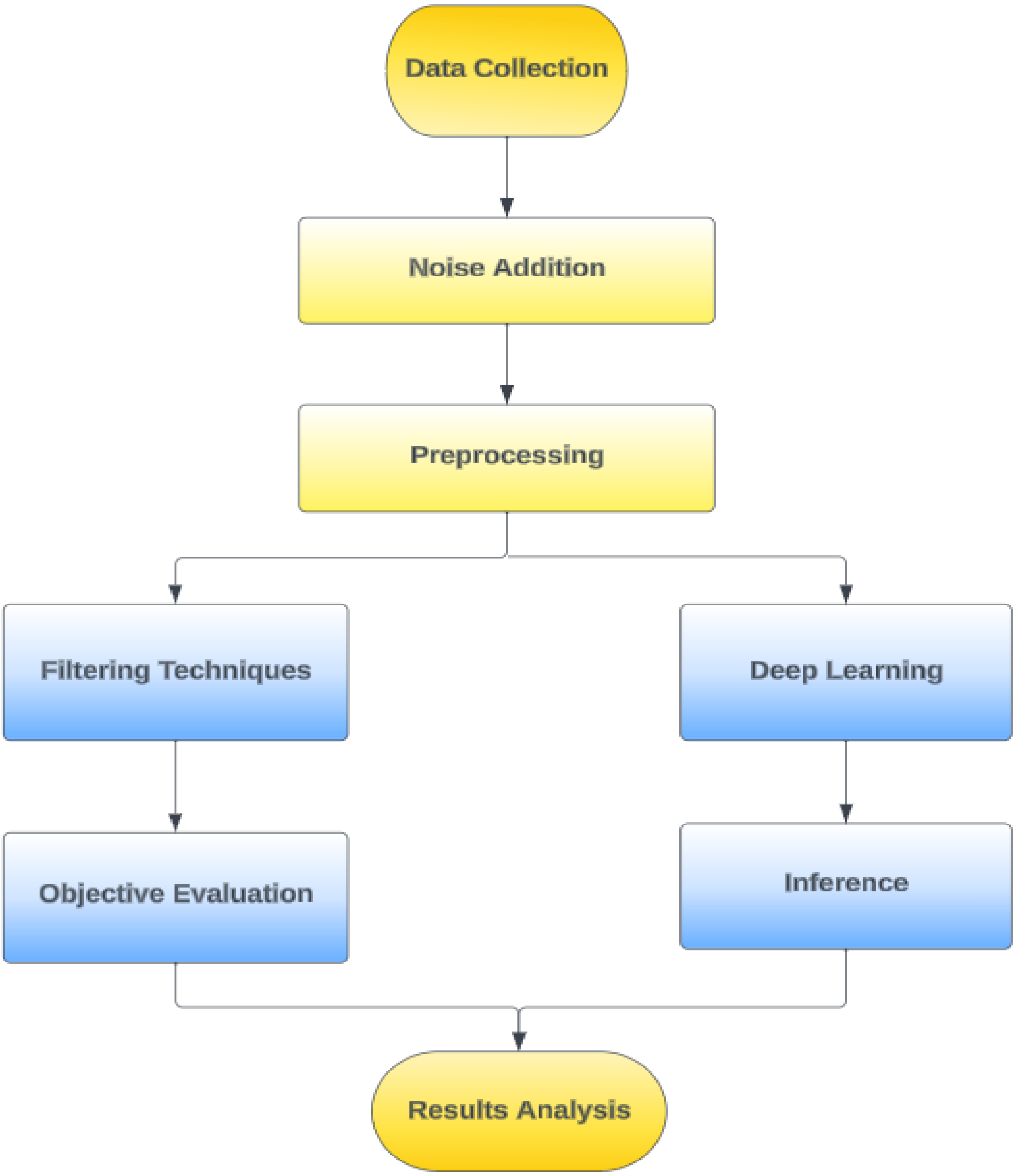


Fig 3.2. System Block Diagram

Chapter 4

Results

4.1 Performance Metrics

The performance of both traditional filtering techniques and the CNN-based machine learning approach was evaluated using two key metrics: Mean Squared Error (MSE) and error percentage. The MSE measures the average squared difference between the original clean speech signal and the cleaned signal obtained after enhancement. A lower MSE indicates better performance, as it implies that the cleaned signal is closer to the original signal. The error percentage is computed as:

Error Percentage=(max value2MSE)×100

where `max value` is the maximum absolute value of the original clean signal. The error percentage provides a normalized measure of the error, making it easier to compare the performance of different enhancement techniques.

4.2 Comparison

The following table summarizes the MSE and error percentages for both the traditional filtering techniques and the CNN-based machine learning approach for a sample set of files:

4.2.1 Observation Tables of Filtering Technique

Voice1:

0_yweweler_0.wav		
Order of filter	**Error%**	**MSE**
30	1.7519%	0.00085
35	1.7719%	0.0009
40	1.8404%	0.0009
45	1.7474%	0.0009

1_yweweler_0.wav		
Order of filter	**Error%**	**MSE**
30	1.1365%	0.0009
35	1.0111%	0.0008
40	1.5446%	0.0012
45	1.4933%	0.0011

2_yweweler_0.wav		
Order of filter	**Error%**	**MSE**
30	1.8036%	0.0011
35	1.7549%	0.0010
40	1.7407%	0.0010
45	2.2347%	0.0013

3_yweweler_0.wav		
Order of filter	**Error%**	**MSE**
30	2.467%	0.0009
35	2.152%	0.0010
40	2.3372%	0.0010
45	2.6039%	0.0013

4_yweweler_0.wav		
Order of filter	**Error%**	**MSE**
30	1.8468%	0.0013
35	1.3687%	0.0010
40	1.469%	0.0012
45	1.2824%	0.0011

Voice2:

0_theo_0.wav		
Order of filter	**Error%**	**MSE**
30	4.5802%	0.009
35	4.1172%	0.0008
40	4.3779%	0.0008
45	5.3135%	0.0010

1_theo_0.wav		
Order of filter	**Error%**	**MSE**
30	2.4928%	0.00076
35	2.9421%	0.0009
40	2.8541%	0.0008
45	3.5815%	0.0010

2_theo_0.wav		
Order of filter	**Error%**	**MSE**
30	2.3154%	0.0009
35	2.4543%	0.0010
40	1.9572%	0.0008
45	2.6888%	0.0011

3_theo_0.wav		
Order of filter	**Error%**	**MSE**
30	3.7773%	0.0007
35	3.3823%	0.0008
40	3.8776%	0.0009
45	4.0707%	0.0010

4_theo_0.wav		
Order of filter	**Error%**	**MSE**
30	1.9373%	0.0008
35	2.1398%	0.0008
40	2.7432%	0.001
45	3.1225%	0.0012

Voice3:

0_nicolas_0.wav		
Order of filter	**Error%**	**MSE**
30	3.617%	0.008
35	3.658%	0.008
40	3.490%	0.007
45	3.027%	0.006

1_nicolas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.6444%	0.004
35	1.7969%	0.004
40	2.0269%	0.005

45	2.0434%	0.005

2_nicolas_0.wav		
Order of filter	**Error%**	**MSE**
30	3.4509%	0.009
35	3.701%	0.010
40	3.802%	0.011
45	2.6861%	0.0073

3_nicolas_0.wav		
Order of filter	**Error%**	**MSE**
30	2.9095%	0.009
35	2.9533%	0.010
40	2.8879%	0.011
45	2.3897%	0.0073

4_nicolas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.6573%	0.0075
35	1.9503%	0.0088
40	2.5842%	0.01171

45	3.1178%	0.0141

Voice4:

0_lucas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.0495%	0.0052
35	0.96828%	0.0048
40	1.0542%	0.0052
45	1.4339%	0.0075

1_lucas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.768%	0.0058
35	1.7535%	0.0058
40	1.8913%	0.0062
45	1.8273%	0.0060

2_lucas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.9788%	0.0064
35	1.9078%	0.0061
40	1.8822%	0.0061

45	1.6362%	0.0053

3_lucas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.0266%	0.0044
35	1.0626%	0.0045
40	1.0443%	0.0044
45	1.0152%	0.004

4_lucas_0.wav		
Order of filter	**Error%**	**MSE**
30	1.8302%	0.0113
35	1.8979%	0.0117
40	2.381%	0.014
45	2.2023%	0.0198

Voice5:

0_jackson_0.wav		
Order of filter	**Error%**	**MSE**
30	3.8855%	0.02865
35	4.9786%	0.36712
40	5.9957%	0.0434
45	5.0976%	0.3754

1_jackson_0.wav		
Order of filter	**Error%**	**MSE**
30	2.1151%	0.009
35	2.0621%	0.0089
40	2.130%	0.0097
45	2.2418%	0.0097

2_jackson_0.wav		
Order of filter	**Error%**	**MSE**

30	2.8085%	0.0099
35	3.0689%	0.0109
40	3.0605%	0.0100
45	2.7514%	0,097

3_jackson_0.wav		
Order of filter	**Error%**	**MSE**
30	3.2363%	0.0095
35	3.9081%	0.0114
40	4.0785%	0.0119
45	3.7654%	0.0116

4_jackson_0.wav		
Order of filter	**Error%**	**MSE**
30	2.989%	0.0157
35	3.3106%	0.0174
40	3.9035%	0.0208
45	4.6357%	0.0244

Voice 6 :

0_george_0.wav		
Order of filter	**Error%**	**MSE**
30	4.7794%	0.0155
35	5.3773%	0.0166
40	5.7923%	0.0188
45	5.2359%	0.0165

1_george_0.wav		
Order of filter	**Error%**	**MSE**
30	0.9168%	0.0023
35	0.9872%	0.0025
40	1.1333%	0.0245
45	1.1981%	0.0030

2_george_0.wav		
Order of filter	**Error%**	**MSE**

30	4.0316%	0.0144
35	4.6315%	0.01311
40	4.3588%	0.0123
45	3.9486%	0.011

3_george_0.wav		
Order of filter	**Error%**	**MSE**
30	1.7526%	0.0045
35	1.8834%	0.0049
40	1.8593%	0.004585
45	1.7518%	0.00457

4_george_0.wav		
Order of filter	**Error%**	**MSE**
30	2.6504%	0.0138
35	3.8808%	0.01854
40	5.199%	0.02184
45	6.13927%	0.03185

4.2.1 Observation Tables of CNN

VOICE 1:

VOICE	MSE	ERROR%

0_yweweler_0.wav	1.0011e-06	0.03482%
1_yweweler_0.wav	2.3183e-06	0.00958%
2_yweweler_0.wav	3.6267e-07	0.011822%
3_yweweler_0.wav	4.832e-08	0.01455%
4_yweweler_0.wav	5.8134e-06	0.024522%

VOICE 2:

VOICE	MSE	ERROR%
0_theo_0.wav	5.8134e-08	0.01455%
1_theo_0.wav	9.4394e-08	0.001005%
2_theo_0.wav	1.2721e-07	0.007834%
3_theo_0.wav	1.21683-07	0.001455%
4_theo_0.wav	5.6997e-07	0.045635%

VOICE 3:

VOICE	MSE	ERROR%
0_nicolas_0.wav	4.1057e-06	0.008190%
1_nicolas_0.wav	4.6239e-06	0.007854%

2_nicolas_0.wav	1.0292e-06	0.023516%
3_nicolas_0.wav	6.0314e-06	0.01284%
4_nicolas_0.wav	2.3295e-05	0.032174%

VOICE 4:

VOICE	MSE	ERROR%
0_lucas_0.wav	7.5785e-06	0.016375%
1_lucas_0.wav	1.4317e-06	0.043265%
2_lucas_0.wav	6.3444e-06	0.00378%
3_lucas_0.wav	9.4082e-06	0.056437%
4_lucas_0.wav	54.3596e-05	0.02367%

VOICE 5:

VOICE	MSE	ERROR%
0_jackson_0.wav	7.8035e-05	0.03245%
1_jackson_0.wav	7.9006e-05	0.002318%
2_jackson_0.wav	1.4411e-05	0.002148%
3_jackson_0.wav	4.4084e-05	0.03748%
4_jackson_0.wav	3.5624e-05	0.023183%

VOICE 6:

VOICE	MSE	ERROR%
0_george_0.wav	3.3644e-05	0.033697%
1_george_0.wav	1.8602e-06	0.0028994%
2_george_0.wav	2.467e-05	0.030753%
3_george_0.wav	3.1516e-06	0.0046205%
4_george_0.wav	2.9365e-05	0.023092%

FILTERING RESULTS:

Mean Squared Error (MSE): 0.01171

ERROR % : 4.499%

CNN RESULTS:

Mean Squared Error (MSE): 8.2007e-06

ERROR % : 0.0977%

These results indicate that the CNN-based machine learning approach consistently achieves lower MSE and error percentages compared to traditional filtering techniques, demonstrating its superior performance in enhancing speech signals.

4.3 Graphical Representation

4.3.1 Filtering Technique

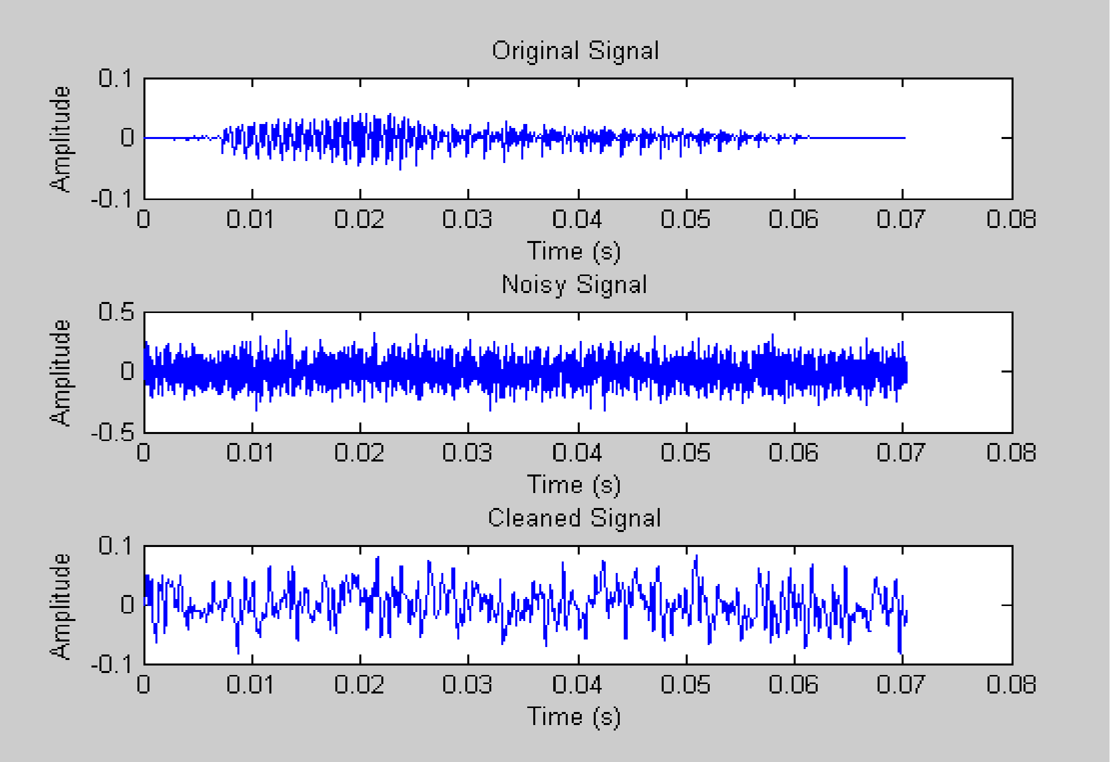
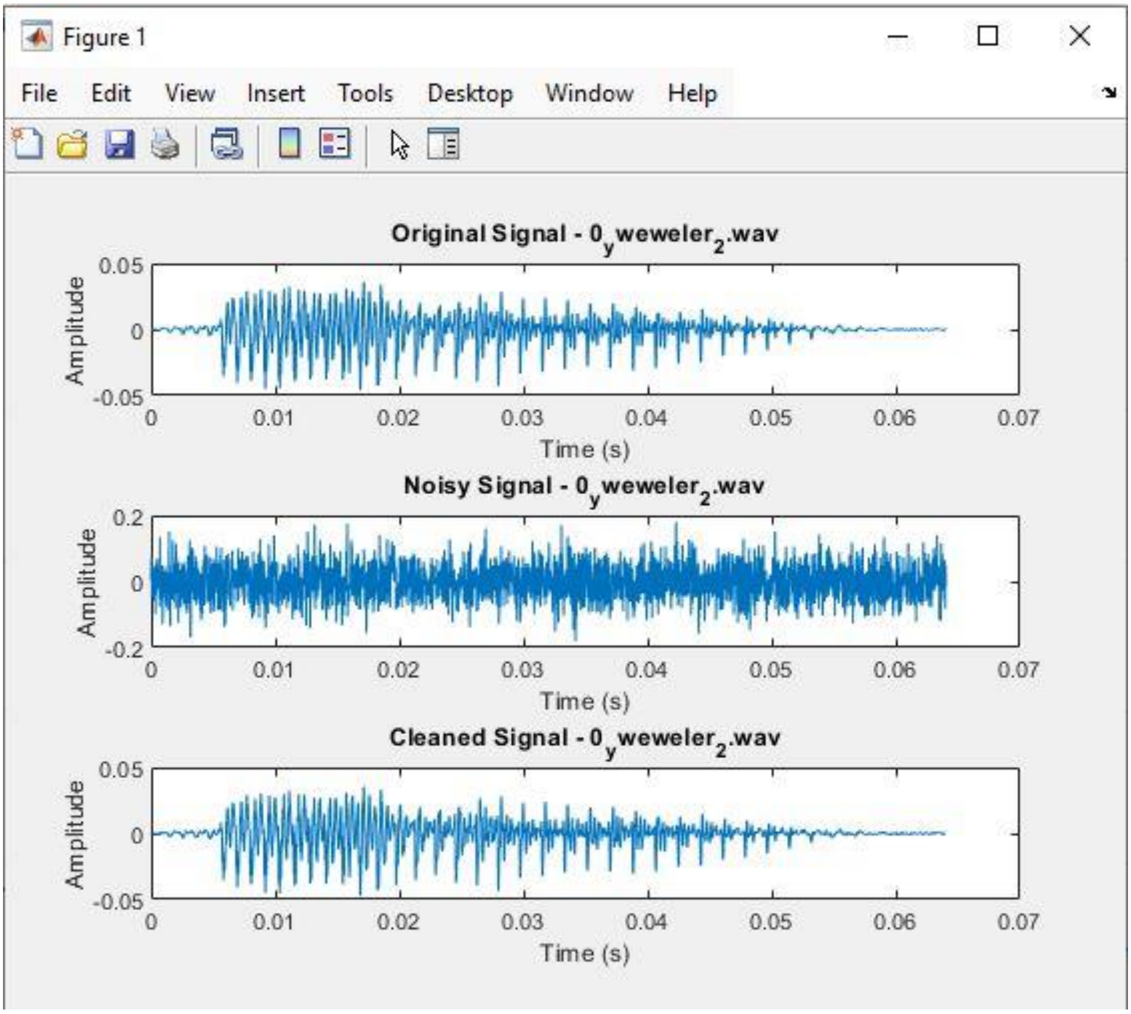


Fig 4.1 Graphical Representation of Low pass filter

- The average Mean Square Error (MSE) was calculated to quantitatively evaluate the performance of our filtering techniques.
- We obtained an average MSE of 0.01171, indicating effective noise reduction and preservation of speech quality.

4.3.1 Machine Learning (CNN)



- The average Mean Square Error (MSE) was calculated to quantitatively evaluate the performance of our CNN model.
- We obtained an average MSE of 8.2007347e-06, indicating effective noise reduction and preservation of speech quality.

These diagrams clearly illustrate the impact of noise on the original signal and the effectiveness of the cleaning process using both the filtering techniques and the CNN approach. The cleaned signals obtained using the CNN approach are visibly closer to the original clean signals, highlighting the superior performance of the machine learning method.

Chapter 5

Discussion and Conclusion

5.1 Discussion

The results of this project demonstrate that the machine learning approach using Convolutional Neural Networks (CNNs) significantly outperforms traditional filtering techniques in enhancing speech signals. The CNN approach consistently achieves lower Mean Squared Error (MSE) and error percentages, indicating that it is more effective at reducing noise and preserving the quality of the speech signal. One of the key advantages of the CNN approach is its ability to learn complex patterns and relationships within the data. This allows it to effectively separate the speech signal from the noise, even in challenging conditions where traditional filtering techniques may struggle. Traditional filters are designed based on predefined parameters and may not adapt well to varying noise characteristics, whereas CNNs can learn from the data and optimize their parameters accordingly.

The success of the CNN approach can be attributed to several factors. Firstly, the architecture of the CNN, with its multiple convolutional layers and fully connected layers, allows it to capture and learn intricate features of the speech signal. The use of ReLU (Rectified Linear Unit) activation functions ensures that the network can model non-linear relationships, which are essential for accurately separating speech from noise. Additionally, the training process, which involves optimizing the network parameters using backpropagation and gradient descent, ensures that the network learns to minimize the error between the predicted and actual clean signals. This leads to a model that is highly effective at enhancing speech signals, even in the presence of significant noise.

Another important aspect of this project is the careful design and tuning of the CNN architecture and training parameters. Factors such as the number of layers, the number of filters in each layer, the learning rate, and the batch size were all optimized to ensure the best possible performance. The choice of the Adam optimizer, which adapts the learning rate based on the gradients, further contributed to the effective training of the network. The results of this project provide strong evidence that CNN-based machine learning approaches are highly effective for speech enhancement. However, it is important to note that the success of these approaches depends on the quality and diversity of the training data. A well-prepared dataset that covers a wide range of noise conditions and speech variations is crucial for training a robust and generalizable model

3.1 Conclusion

In conclusion, this project has demonstrated that machine learning approaches, specifically Convolutional Neural Networks (CNNs), are highly effective for speech enhancement. The CNN-based approach significantly outperforms traditional filtering techniques in terms of Mean Squared Error (MSE) and error percentages, highlighting its superior performance in reducing noise and preserving the quality of speech signals. The success of the CNN approach can be attributed to its ability to learn complex patterns and relationships within the data, as well as the careful design and optimization of the network architecture and training parameters. The results of this project provide valuable insights into the potential of machine learning approaches for speech enhancement and pave the way for further research and development in this field.

Future work could explore the application of other advanced machine learning techniques, such as recurrent neural networks (RNNs) and transformer models, for speech enhancement. Additionally, the incorporation of more diverse and extensive datasets, covering a wider range of noise conditions and speech variations, could further improve the robustness and generalizability of the models. Overall, the findings of this project highlight the potential of machine learning approaches for revolutionizing the field of speech enhancement and improving the quality of communication in noisy environments.

Plagiarism Report

ORIGINALITY REPORT

12%	12%	%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	pubs.sciepub.com Internet Source	1%
2	thesai.org Internet Source	1%
3	www.mdpi.com Internet Source	1%
4	www.ijisae.org Internet Source	1%
5	github.com Internet Source	1%
6	ouci.dntb.gov.ua Internet Source	1%
7	link.springer.com Internet Source	1%
8	journalhosting.ucalgary.ca Internet Source	1%
12	journalspress.com Internet Source	<1%
13	research.vu.nl Internet Source	<1%
14	researchportal.vub.be Internet Source	<1%
15	www.mlpack.org Internet Source	<1%
16	www.theseus.fi Internet Source	<1%
17	data.gov.in Internet Source	<1%
18	www.springerprofessional.de Internet Source	<1%

References

- Venkateswarlu, S. China, K. Satya Prasad, and A. SubbaRami Reddy. "Improve Speech Enhancement Using Weiner Filtering." *Global Journal of Computer Science and Technology* 11, no. 7 (2011).
- Ephraim, Yariv, and David Malah. "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator." *IEEE Transactions on acoustics, speech, and signal processing* 32, no. 6 (1984): 1109-1121.
- Selvi, R. Senthamizh, P. Sathish Kumar, R. Sri Krishna, and S. Surya Rao. "Speech Enhancement using Adaptive Filtering with Different Window Functions and Overlapping Sizes." *Turkish Journal of Computer and Mathematics Education* 12, no. 13 (2021): 1886-1894.
- Yuliani, Asri Rizki, M. Faizal Amri, Endang Suryawati, Ade Ramdan, and Hilman Ferdinandus Pardede. "Speech enhancement using deep learning methods: A review." *Jurnal Elektronika dan Telekomunikasi* 21, no. 1 (2021): 19-26.

Index