Speech Restoration & Enhancement

A Comparative Analysis of Spectral Subtraction vs. Wiener Filtering for Stationary & Non-Stationary Noise.

Team members: Sunny , Aman and Deepanshu

Problem Statement

Real-World Context:

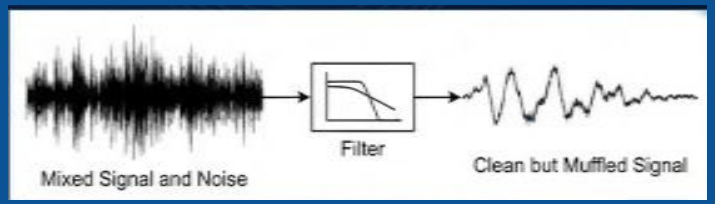
In telecommunications and hearing aids, background noise drastically reduces speech intelligibility.

Common noise types: Stationary (Fan, AC hum) and Non-Stationary/Broadband (Heavy Rain, Traffic).

The Challenge:

Simple filters (Low-Pass/High-Pass) remove noise but also muffle the human voice.

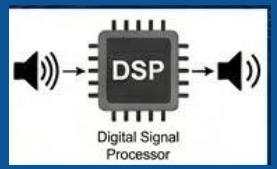
We need an intelligent system that separates "Noise" from "Signal" without damaging the speech harmonics.



Project Objectives & Scope

Primary Objective

To implement and evaluate Digital Signal Processing (DSP) algorithms for speech enhancement.



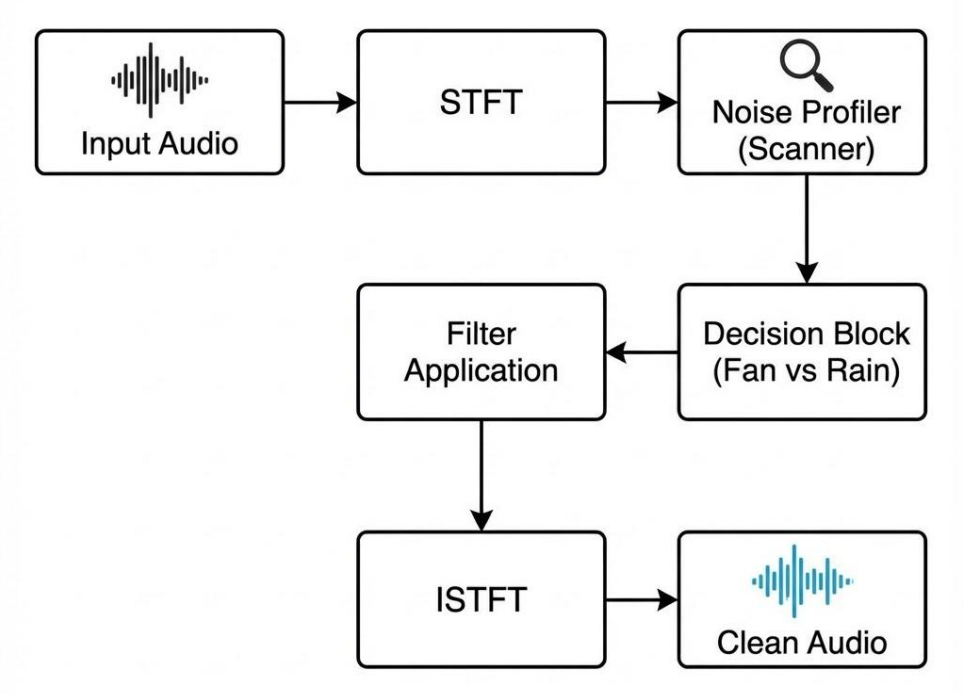
Key Concepts Applied

1. Sampling & Audio Digitization.
2. Time-Frequency Analysis (STFT & Spectrograms).
3. Adaptive Filtering & Noise Gating.

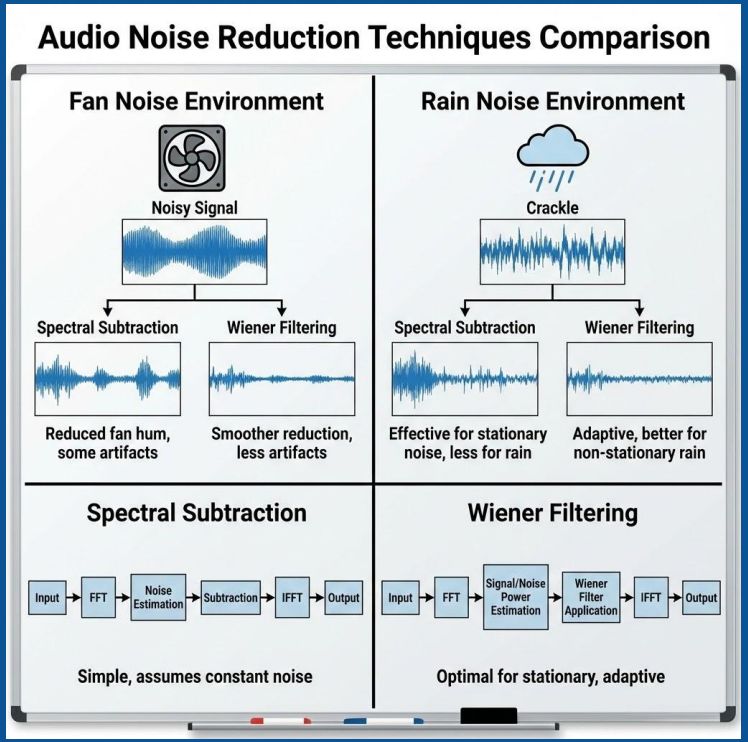
Scope

1. Handling different noise environments (Fan vs. Rain).
2. Comparing Spectral Subtraction vs. Wiener Filtering.

System Architecture



Two Approaches



Methodology 1: Spectral subtraction (Fan noise)

Target Noise: Stationary (Constant Low-Frequency Noise like Fans).

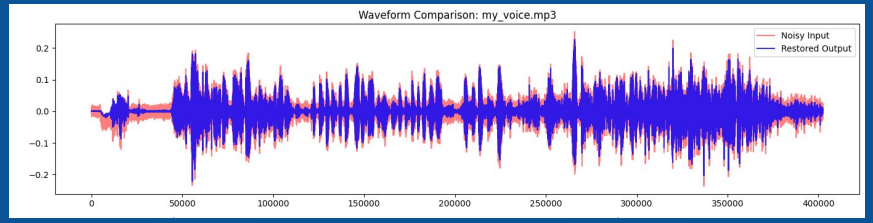
Algorithm Steps:

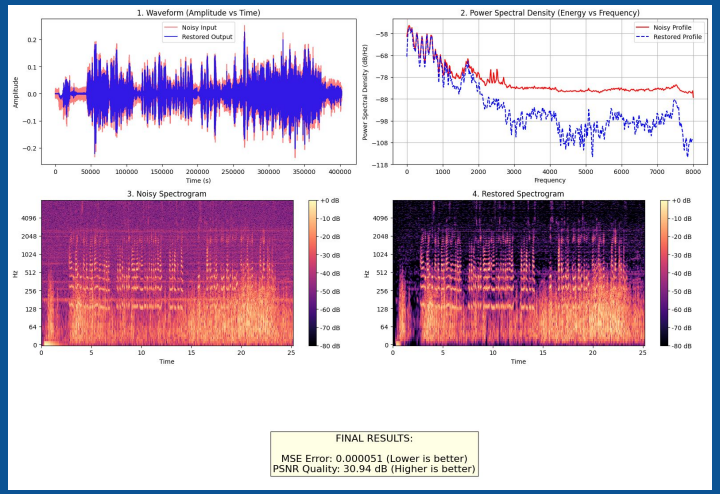
1. Scan: Code analyzes the first few frames to create a "Noise Fingerprint".
2. Transform: Convert Audio to Frequency Domain using STFT.
3. Subtract: Mathematically subtract the Noise Profile from the Signal Magnitude.
4. Reconstruct: Use Inverse-STFT to get clean audio.

Recorded voice	
Unclear voice	
Enhanced voice	

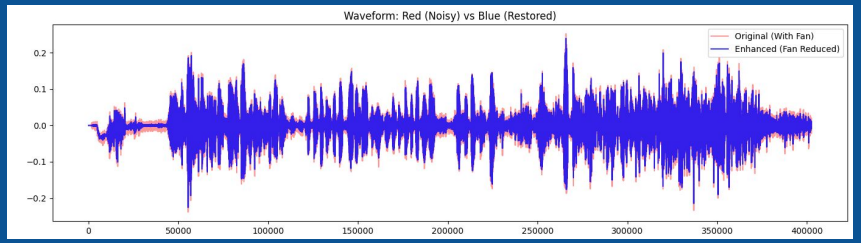
Methodology 1: Spectral subtraction (Fan noise)

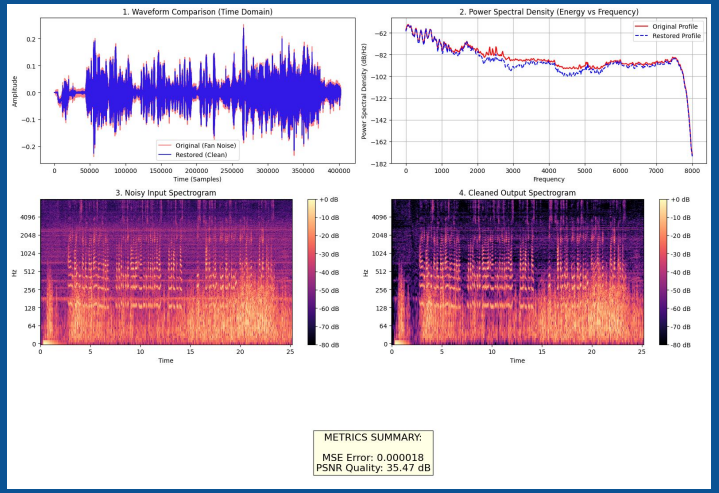
Unclear voice





Enhanced voice





Challenge Encountered - The Rain problem

The Failure Case:

1. When we applied Spectral Subtraction to Rain Noise, it failed.
2. Result: The output had "Musical Noise" (Robotic/Alien artifacts).

Why it Failed:

Rain is Broadband Noise (High Frequency Hiss).

It overlaps perfectly with human speech (Consonants like 's', 't').

Aggressive subtraction deleted the speech details, leaving robotic residues.

Recorded voice
Unclear voice

Link

Methodology 2: Wiener Filtering (The Rain Solution)

The Pivot: We switched to Statistical Filtering (Wiener Filter).

How it Works:

Instead of "subtracting" noise, it calculates the Probability of speech being present at each frequency.

$$H(f) = \frac{\text{Signal Power}}{\text{Signal Power} + \text{Noise Power}}$$

If Signal is strong → Keep it.
If Signal is weak → Suppress it.

Refinement:

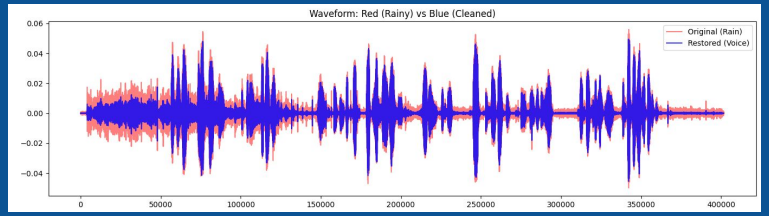
Added a Bandpass Filter (300Hz - 3400Hz) to mimic Telephony standards.

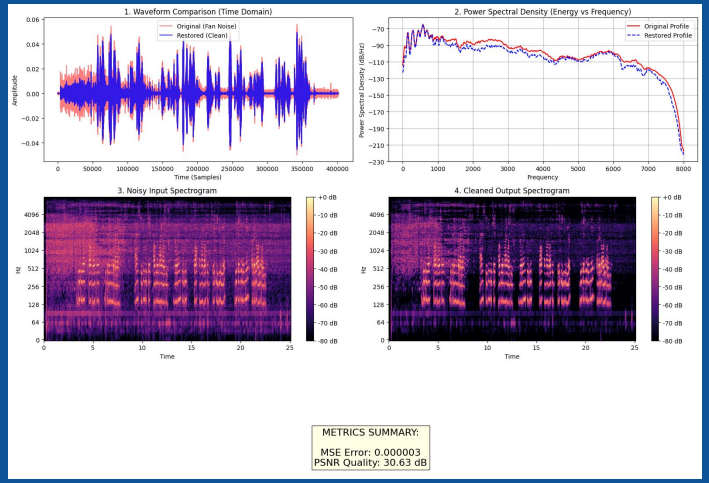
Applied Noise Gating to silence gaps between words.

Enhanced voice

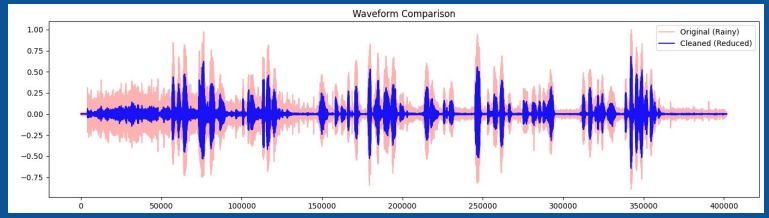
Methodology 2: Wiener Filtering (The Rain Solution)

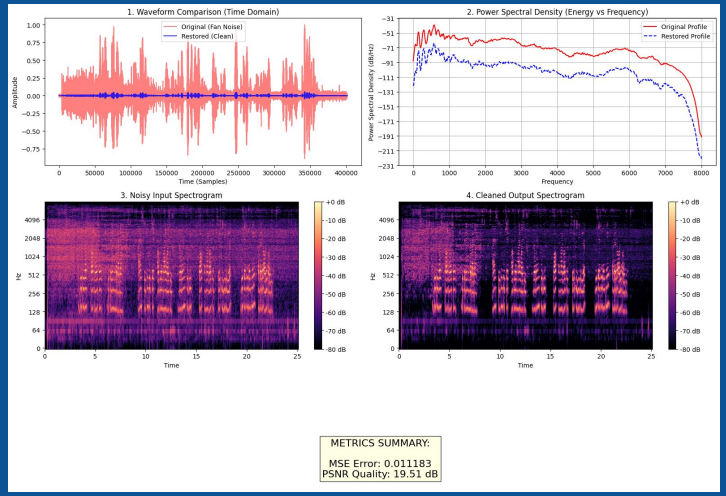
Unclear voice





Enhanced voice





Model Architecture: The U-Net

Why U-Net?

Originally designed for medical image segmentation, U-Net is the industry standard for detailed reconstruction tasks.

- Industry standard for detailed reconstruction

Structure

Encoder (Downsampling)

Compresses the audio image, keeping structured patterns while discarding noise.

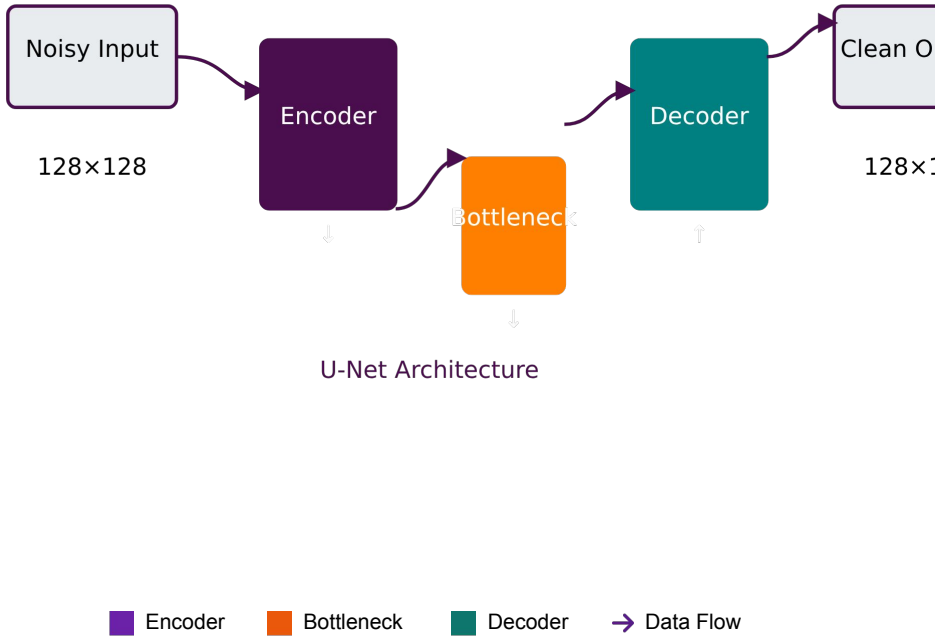
Bottleneck

Deepest layer where audio is represented in its most abstract form.

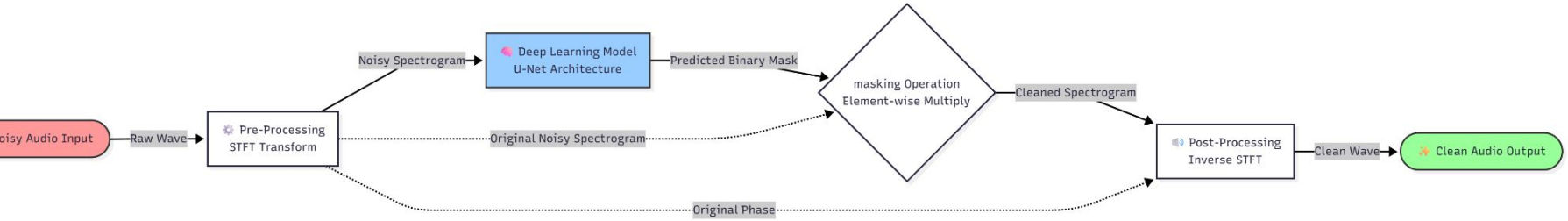
Decoder (Upsampling)

Reconstructs the clean audio image back to its original size.

U-Net Architecture Visualization



U-Net Architecture



The "Secret Sauce": Skip Connections

The Problem with Autoencoders

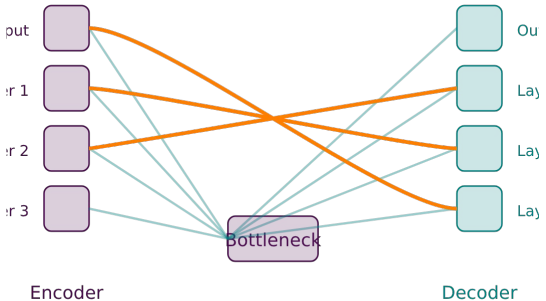
When you compress audio data:

- Losing high-frequency details (clarity)
- Results in "muffled" output
- Audio becomes unrecognizable

The Challenge:

How do we preserve voice clarity while removing noise?

Skip Connections



Direct "bridges" from Encoder to Decoder

The U-Net Solution

Skip Connections:

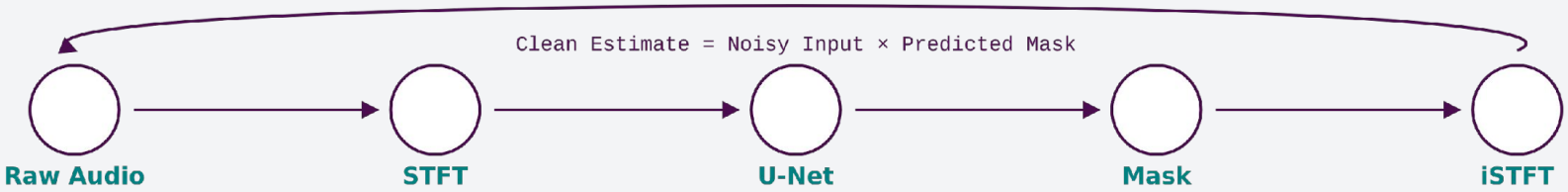
- Direct "bridges" from Encoder to Decoder layers
- Pass high-resolution spatial information
- Bypass the bottleneck entirely

The Result:

Model focuses only on noise detection, while voice remains crisp and clear.

Technical Workflow

Complete Processing Pipeline



1. Input

Raw Noisy Audio (.wav)

2. Preprocessing

- STFT to generate Magnitude Spectrogram
- Slice into 128×128 image patches

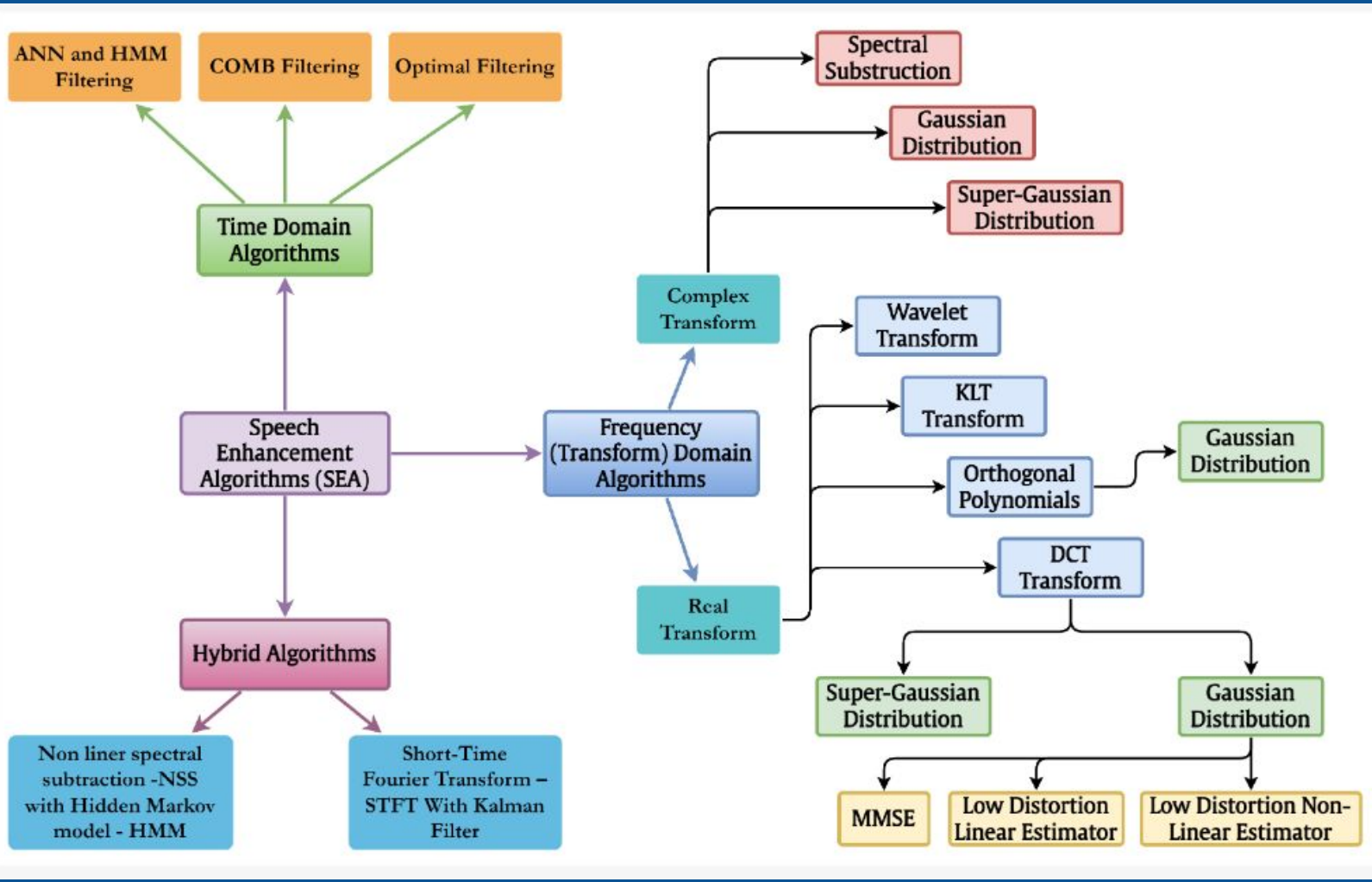
3. Inference

U-Net predicts a Soft Mask

```
Clean Estimate = Noisy Input
× Predicted Mask
```

4. Reconstruction

- Combine cleaned magnitude with original Phase
- Apply iSTFT to convert back to audio waves


ANN and HMM Filtering
COMB Filtering
Optimal Filtering
Time Domain Algorithms
Speech Enhancement Algorithms (SEA)
Hybrid Algorithms
Non liner spectral subtraction -NSS with Hidden Markov model - HMM
Short-Time Fourier Transform – STFT With Kalman Filter
Frequency (Transform) Domain Algorithms
Complex Transform
Real Transform
Spectral Substruction
Gaussian Distribution
Super-Gaussian Distribution
Wavelet Transform
KLT Transform
Orthogonal Polynomials
Gaussian Distribution
DCT Transform
Super-Gaussian Distribution
Gaussian Distribution
MMSE
Low Distortion Linear Estimator
Low Distortion Non-Linear Estimator

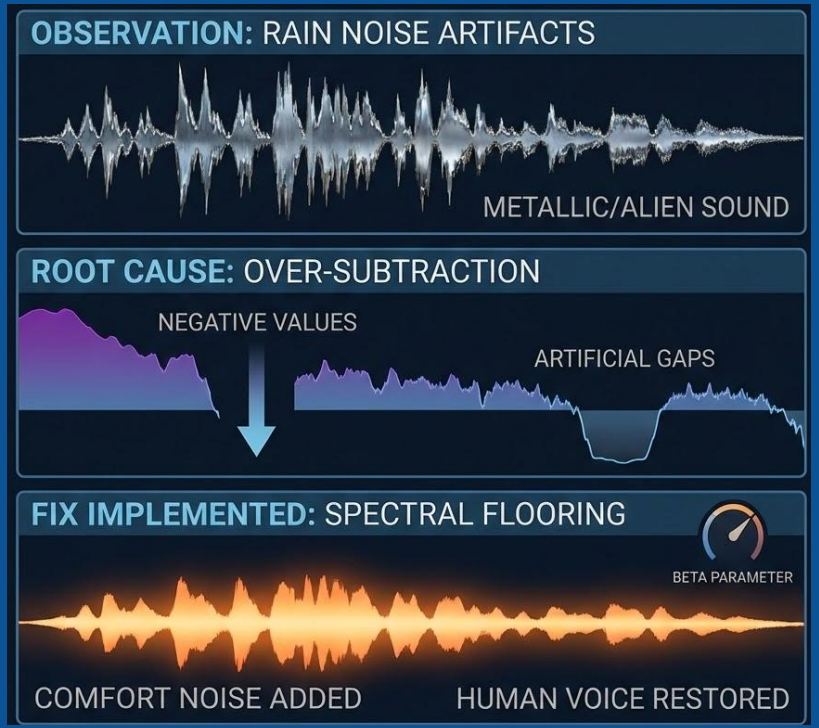
Summary

Observation: Initial attempts on rain noise resulted in metallic/alien artifacts.

Root Cause: "Over-subtraction" of spectral magnitude resulted in negative values, which were floored to zero, creating artificial gaps in sound.

Fix Implemented:

Spectral Flooring: We added a beta parameter to keep a small amount of natural background noise (Comfort Noise), making the voice sound human again.



Research Paper Links

Speech Enhancement Algorithms: A Systematic Literature Review

Deep neural networks for speech enhancement and speech recognition: A systematic review - ScienceDirect

Research Progress in Speech Enhancement Technology | IEEE Conference Publication

(PDF) Speech Enhancement Using Machine Learning

End-to-end feature fusion for jointly optimized speech enhancement and automatic speech recognition | Scientific Reports

Research on Speech Enhancement Based on Deep Neural Network - IOPscience

speech enhancement Latest Research Papers | ScienceGate

Real Time Speech Enhancement in the Waveform Domain - Meta Research

Thank you