Technical Report: Speech Enhancement via Spectral Subtraction

Name : COTTEL Thibault

Field: Audio Signal Processing / High-Level Algorithm Development

Technologies: Python, NumPy, SciPy

1. Introduction

This project implements a stationary noise reduction algorithm based on spectral subtraction. The primary objective is to extract a clear voice signal from a noisy environment while minimizing audible artifacts.

2. Mathematical and Technical Approach

The algorithm operates by transforming the audio signal from the time domain to the frequency domain to manipulate its spectral structure.

A. Short-Time Fourier Transform (STFT)

The audio signal is decomposed into temporal frames of 1024 samples (~23ms at 44.1kHz) with a 50% overlap. A Hann window is applied to each frame to reduce spectral leakage and edge discontinuities.

B. Noise Estimation

The noise profile is estimated during the initial frames of the signal, assuming a period of relative silence (speech-absent frames). The average spectral magnitude is calculated to create a model of the stationary noise:

$$\hat{N}(f) = \frac{1}{K}\sum_{i=1}^{K} |Y_i(f)|$$

C. Optimized Spectral Subtraction

To achieve a natural sound quality, I implemented an enhanced version of spectral subtraction using two critical parameters:

- **Over-subtraction (alpha = 2.0):** This factor allows for more aggressive suppression of noise peaks.
- **Spectral Floor (beta = 0.02):** This prevents the magnitude from dropping to absolute zero, which effectively eliminates "musical noise" (metallic artifacts that can be heard before) and results in a more pleasant auditory experience for the end-user.

3. Results and Analysis

The script generates a visual comparison via spectrograms to validate the algorithm's performance.

- **Noisy Spectrogram:** Shows constant energy density across all frequencies.

- **Denoised Spectrogram:** Background areas have turned dark (near-zero energy), while the speech formants (vertical harmonic structures) are clearly preserved.

The implementation of the Spectral Floor significantly improved the output, moving from a robotic, distorted sound to a much more faithful reconstruction of the original voice.

---

4. Appendix: Sources and Resources

To ensure test reproducibility and algorithmic relevance, the following professional resources were utilized:

Datasets :

- **Microsoft SNSD (Speech Normalization and Simulation Dataset):** Used for stationary noise profiles (Air Conditioner).
  - *Source:* https://github.com/microsoft/MS-SNSD
- **LibriSpeech ASR Corpus:** Used for clean speech samples as the base for SNR mixing.
  - *Source:* https://www.openslr.org/12