Group 32

Sankeerthana Venugopal (2020102008)

Srujana Vanka (2020102005)

Suppression of Acoustic Noise in Speech Using Spectral Subtraction

Objectives:

1) Developing a noise suppression algorithm using spectral subtraction

2) Reducing the spectral error by

- Magnitude averaging
- Bias removal
- Rectification
- Residual noise removal
- Non speech signal suppression

Project summary:

Many times, speech must be processed in the face of unwanted background noise, which degrades its performance, lowers speech quality and intelligibility. In the literature, a range of noise suppressing strategies capable of reducing background noise have been studied. These include- using noise-cancelling microphones, internal modification of the voice processor algorithms to explicitly compensate for signal contamination, or preprocessor noise reduction. Noise-cancelling microphones offer almost null reduction above 1 KHz. However, because of the time, effort, and money invested in the design and implementation of these voice processors, there is a reluctance to modify these systems internally. Coming to preprocessor noise reduction, the advantage is that noise is stripped from the waveform itself, with the output being either digital or analogue speech.

The goal of this project is to create a noise suppression technique, implement a computationally efficient algorithm, and test it in real-world noise situations. Spectral subtraction is a processor-independent, computationally efficient way to arrive at effective digital speech analysis. The assumption is that the noise is a stationary or a slowly varying process, and that the noise spectrum does not change significantly in-between the update periods. Spectral subtraction needs only noisy speech as input. For this, an estimator is obtained by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The signal collected during nonspeech activity provides the spectral information needed to define the noise spectrum.

Modeling the Noisy Speech Signal:

n(k) - noise signal

s(k) - windowed speech signal (Speech, suitably low-pass filtered and digitized, is analyzed by windowing data)

$$x(k) = s(k) + n(k).$$

Taking the Fourier transform gives

$$X(e^{j\omega}) = S(e^{j\omega}) + N(e^{j\omega})$$

where, **DTFT:**

$$X(e^{j\omega}) = \sum_{k=0}^{L-1} x(k)e^{-j\omega k}$$

Inverse DTFT:

$$x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega k}\, d\omega$$

Spectral Subtraction Estimator:

The spectral subtraction estimator is the difference of the average value of the noise spectrum during non-speech activity from the noisy speech spectrum.

$$\hat{S}(e^{j\omega}) = [|X(e^{j\omega})| - \mu(e^{j\omega})]\, e^{j\theta_x(e^{j\omega})}$$

or

$$\hat{S}(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})$$

with

$$H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|}$$

$$\mu(e^{j\omega}) = E\{|N(e^{j\omega})|\}.$$

After the spectral estimator is developed, the spectral error is given by

$$\epsilon(e^{j\omega}) = \hat{S}(e^{j\omega}) - S(e^{j\omega}) = N(e^{j\omega}) - \mu(e^{j\omega})\, e^{j\theta_x}.$$

The modifications for reducing the auditory effects of this spectral error are:

1) magnitude averaging:

- By replacing:

$|X(e^{j\omega})|$ with $\overline{|X(e^{j\omega})|}$ where,

$$\overline{|X(e^{j\omega})|} = \frac{1}{M} \sum_{i=0}^{M-1} |X_i(e^{j\omega})|$$

- On averaging, the spectral error becomes equal to:

$$\epsilon(e^{j\omega}) = S_A(e^{j\omega}) - S(e^{j\omega}) \cong \overline{|N|} - \mu$$

- On taking a longer average, the mean will almost converge with μ.
- The obvious flaw with this modification is that the speech is nonstationary, averaging over a longer duration can decrease intelligibility of the speech signal.

2) half-wave rectification:

- For each frequency w where the spectral subtraction estimate takes a negative value (noisy signal spectrum magnitude is less than average noise spectrum magnitude), the output is set to zero
- This modification can be simply implemented by half-wave rectifying the spectral subtraction filter.

$$\hat{S}(e^{j\omega}) = H_R(e^{j\omega})X(e^{j\omega})$$

where

$$H_R(e^{j\omega}) = \frac{H(e^{j\omega}) + |H(e^{j\omega})|}{2}.$$

- As a result, half-wave rectification causes the magnitude spectrum at each frequency w to be biased down by the noise bias established at that frequency.
- The advantage of half rectification is that the noise floor is reduced by μ(w).

3) residual noise reduction:

- $N_R = N - \mu e^{j\theta_n}$, where $\boldsymbol{N_R}$ = noise residual.
- The noise residual will randomly fluctuate from frame to frame. So it can be suppressed by replacing the current value with the minimum magnitude value from the three consecutive analysis frames in each frequency bin . This is done only when the current amplitude is less than the maximum noise residual observed during nonspeech activity.

- Implementation:

$$|\hat{S}_i(e^{j\omega})| = |\hat{S}_i(e^{j\omega})|, \quad \text{for} \quad |\hat{S}_i(e^{j\omega})| \geqslant \max |N_R(e^{j\omega})|$$

$$|\hat{S}_i(e^{j\omega})| = \min \{|\hat{S}_j(e^{j\omega})| \, j = i-1, i, i+1\},$$
$$\text{for} \quad |\hat{S}_i(e^{j\omega})| < \max |N_R(e^{j\omega})|$$

- The drawback is that it requires extra storage to record the maximum noise residuals and magnitude values for three adjacent frames.

4) additional signal attenuation during nonspeech activity:

- An effective speech activity detector is made by the spectral subtraction approach. This detector necessitates the establishment of a threshold that indicated the absence of speech activity.
- This threshold (T = - 12 dB) is chosen empirically to guarantee that only signals that were clearly background noise would be attenuated.

References

- Paper assigned - “suppression of acoustic noise in speech using spectral subtraction” (https://ieeexplore.ieee.org/document/1163209)

- Cole, C., Karam, M., & Aglan, H. (2008), “*Spectral Subtraction of Noise in Speech Processing Applications” IEEE (https://ieeexplore.ieee.org/document/4480188)*

- S. F. Boll, “Suppression of noise in speech using the SABER method,” in Proc. ZEEE Znt. Conf. on Acoust., Speech, Signal Processing, Tulsa, OK, Apr. 1978, pp. 606-609. (https://ieeexplore.ieee.org/document/1170572)