NON-COHERENT MULTI-CHANNEL SPEECH ENHANCEMENT IN NON-STATIONARY NOISE ENVIRONMENTS

Liav Levi & Nir Kahana

Supervisor: Ronen Talmon

In association with: Israel Police Dept.
Introduction

- **Scenario**
  - A restaurant with a few tables, in each there is a microphone
  - Enhancing the speech in one table, using the other recordings

- ** Modeling the problem**
  - Speech enhancement
  - Non-stationary noise environment
  - Multi-channel recordings
  - Distant, non-coherent, signals
Assume only two speakers

\[ y_p[n] = x_1[n] + h_{p2} \ast x_2[n] + h_{pu} \ast u[n] + d_p[n] \]
\[ y_r[n] = x_2[n] + h_{r1} \ast x_1[n] + h_{ru} \ast u[n] + d_r[n] \]

- \( x_1[n], x_2[n] \) - Independent speech signals
- \( u[n] \) - Stationary environmental noise
- \( d_p[n], d_r[n] \) - Weak stationary noise at each mic.
- \( h_{ij}[n] \) - LTI FIR transfer functions

Our goal is to estimate \( x_1[n] \) using \( y_p[n], y_r[n] \)
Optional Solutions

- **Single-source**
  - Using only one microphone, and consider the noise as stationary
  - Spectral subtraction
  - STSA [Ephraim & Malah, 1984], LSA [Ephraim & Malah, 1985]
  - OM-LSA [Cohen, 2002] on each microphone

- **Multi-sources**
  - Beamforming
    - Not suitable, because of non-coherent signals

- **Make our own solution**
  - Based on OM-LSA
  - Modification for multi-channel
    - Exploit measurements from other microphones
Optimally Modified Log-Spectral Amplitude Estimator (OM-LSA)

- **Single source**
  \[ y[n] = x[n] + d[n] \]

- **Minimizing the MSE of the log-spectral amplitude**
  \[ \min \hat{X}(\ell,k) E\left[ \log|X(\ell,k)| - \log|\hat{X}(\ell,k)| \right]^2 \]

- **Assumptions:**
  - (Quasi) stationary noise and Gaussian distribution
  - Unknown signal to noise ratio (SNR)
  - Speech presence uncertainty

- **General solution:**
  \[ \hat{X}(\ell,k) = G(\ell,k)Y(\ell,k) \]
Developed modules

“Decision directed” a-priori SNR estimator
- Using the estimation of $\hat{X}(\ell, k)$ to estimate the a-priori SNR

Speech absence probability (SAP) estimator
- Detects changes in the estimated a-priori SNR
- Used to control the gain function
- Used to decide whether or not to estimate the noise variance

Noise spectrum estimator
- Estimating while speech is absent
- Applying a temporal recursive smoothing parameter
OM-LSA – Better Performance

- Multi-channel vs. single-channel
- Better noise estimation
  - Studying the transfer function when speech is absent
  - Estimating non-stationary noise, using the other mic.
- Better SAP estimation
- Better noise suppression
Our Algorithm

- In time domain:

\[ y_p[n] = x_1[n] + h_{p2} * x_2[n] + h_{pu} * u[n] \]
\[ y_r[n] = x_2[n] + h_{r1} * x_2[n] + h_{ru} * u[n] \]

- In STFT domain:

\[ Y_p(\ell,k) = X_1(\ell,k) + H_{p2}(k)X_2(\ell,k) + H_{pu}(k)U(\ell,k) \]
\[ = X_1(\ell,k) + N_p(\ell,k) \]
\[ Y_r(\ell,k) = X_2(\ell,k) + H_{r1}(k)X_1(\ell,k) + H_{ru}(k)U(\ell,k) \]
\[ = X_2(\ell,k) + N_r(\ell,k) \]
Hypotheses

- **When no speech is present:**
  
  \[ Y_p (\ell, k) = H_{pu} (k) U (\ell, k) = \overline{H} (k) \overline{U} (\ell, k) \]
  
  \[ Y_r (\ell, k) = H_{ru} (k) U (\ell, k) = \overline{U} (\ell, k) \]

- **Estimating the noises’ variances**
  
  \[ \lambda_u (\ell, k) = E |\overline{U} (\ell, k)|^2 \]

- **Estimating the relative transfer function (RTF)**
  
  \[ |\overline{H} (k)|^2 = \frac{|H_{pu} (k)|^2}{|H_{ru} (k)|^2} \]
Hypotheses

- When only $X_2$ is present:

$$Y_p(\ell,k) = H_{p2}(k)X_2(\ell,k) + H_{pu}(k)U(\ell,k)$$

$$Y_r(\ell,k) = X_2(\ell,k) + H_{ru}(k)U(\ell,k)$$

- Estimating $|H_{p2}(k)|^2$

- Similarly, when only $X_1$ is present:

- Estimating $|H_{r1}(k)|^2$

- When both $X_1, X_2$ are present, using the above estimations to reduce the noise
Future Work

- Further developing
  - NSE – Noise spectrum estimator
    \[ \hat{\lambda}_{N_p}(\ell,k), \hat{\lambda}_{N_r}(\ell,k) \]
  - SAP – Signal absence probability for multi-channel
    \[ P[H_1(\ell,k)], \ldots, P[H_4(\ell,k)] \]

- Coding in Matlab

- Testing
  - Simulated signals & Recordings (from the police)
  - Comparison to single-source OM-LSA