Joint Filtering and Factorization for Recovering Latent Structure from Noisy Speech Data

Colin Vaz, Vikram Ramanarayanan, and Shrikanth Narayanan
Signal Analysis and Interpretation Lab, University of Southern California,
Los Angeles, CA 90089

Motivation

- Observing latent structure in a signal of interest allows for data analysis and interpretation.
- NMF is a widely-used method for observing latent structure, but is sensitive to noise.
- Noise in data prevents proper recovery of latent structure.
- **Goal:** Improve recovery of latent structure by jointly filtering the data to remove noise and factoring it to expose it's latent structure.

Joint Filtering and Factorization

- Non-negative matrix factorization (NMF) provides a parts-based representation of a signal.
- Minimum variance distortionless response (MVDR) filter computes a power spectrum that estimates the spectral envelope of a signal without distorting the spectrum.
- Combine the cost functions of NMF and MVDR to perform joint filtering and factorization.
- $J = \underbrace{\|G \otimes V - WH\|_F^2}_{\text{Factorization}} + \lambda_1 \underbrace{\|G \otimes (WH)\|_F^2}_{\text{Filtering}} + \lambda_2 \underbrace{\|G \otimes (WH) - A \otimes (WH)\|_F^2}_{\text{Distortionless constraint}}$
- G filters the input signal V, which is factored into W and H.
- A is a frequency response that gives prior information about which frequency bands contain the signal of interest.
- Update equations:
- $G = \frac{(WH) \otimes V + \lambda_2 A \otimes (WH)^2}{V^2 + (\lambda_1 + \lambda_2)(WH)^2}$

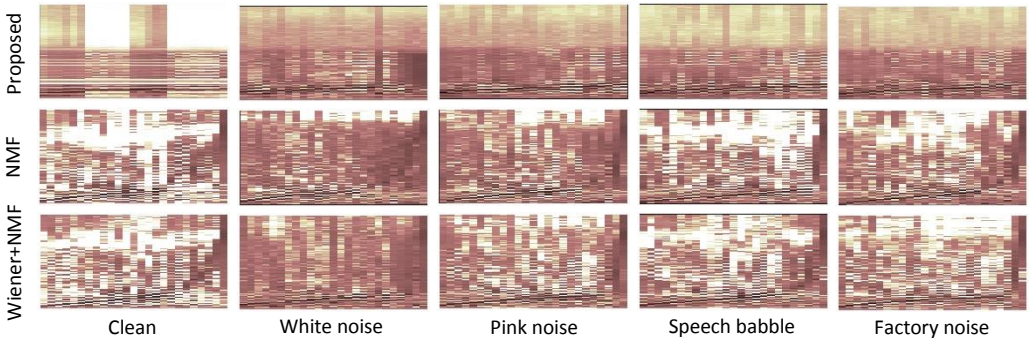
$$W \leftarrow W \otimes \frac{(G \otimes V)H^T}{WHH^T + \lambda_1 W \otimes C + \lambda_2 W \otimes D}$$

$$H \leftarrow H \otimes \frac{W^T(G \otimes V)}{W^T WH + \lambda_1 E \otimes H + \lambda_2 F \otimes H}$$

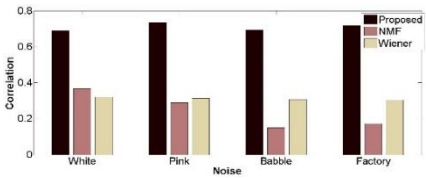
- Filtering operation can be viewed as computing the optimal filter $g_m[n]$ for each frame in the STFT of the noisy signal.
- $X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j\omega n}$
- $Y(m,\omega) = \sum_{n=-\infty}^{\infty} (x[n]w[n-m] * g_m[n])e^{-j\omega n}$
- $= X(m,\omega)G(m,\omega)$ by Parseval's Theorem.

Basis Recovery Experiment

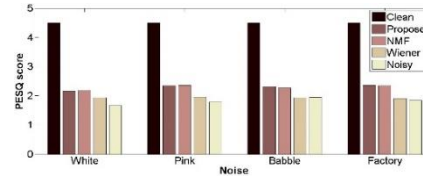
- Correlation metric: $\rho = \frac{1}{K}\sum_{k=1}^{K} \frac{|\boldsymbol{W}_{\text{clean}}^T(:,k)\boldsymbol{W}_{\text{noisy}}(:,k)|}{\|\boldsymbol{W}_{\text{clean}}(:,k)\|_2 \|\boldsymbol{W}_{\text{noisy}}(:,k)\|_2}$
- Added white, pink, speech babble, and factory noises to 100 TIMIT sentences at 5 dB and 10 dB SNR.



- Inverted WH (estimated clean spectrogram of the noisy speech) to obtain denoised speech.



Correlation scores of the recovered basis matrices.



PESQ scores of the denoised speech.

Conclusion & Future Work

- Proposed joint filtering and factorization approach by combining NMF and MVDR into a single cost function.
- Correlation scores show better basis recovery of noisy signals.
- PESQ scores show denoising performance is comparable to other denoising methods.
- Use more generalized divergence metrics in cost function.
- Extend approach to other analysis methods, eg. PCA.
- Evaluate proposed algorithm's performance on other tasks, eg. phoneme recognition.

Work supported by NIH Grant DC007124