May 22–27, 2011
Prague
Czech Republic
2011 International Conference on Acoustics, Speech and Signal Processing

Two-Channel Post-filtering Based on Adaptive Smoothing and Noise Properties

Chengshi Zheng, Yi Zhou, Xiaohu Hu, and Xiaodong Li

Key Laboratory of Noise and Vibration Research
(Communication Acoustics Lab)
Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

Email:cszheng@mail.ioa.ac.cn

中国科学院声学研究所
Institute of Acoustics, Chinese Academy of Sciences

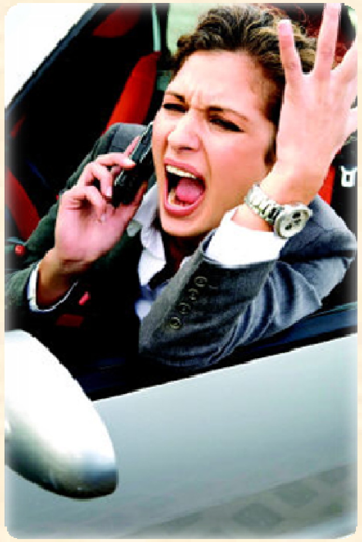
IEEE
IEEE Signal Processing Society

Outline

- Background and motivation
- Theoretical study of two-channel Post-filtering Algorithms
  - Model and derivation
  - Statistical Properties of the post-filter estimator
  - Some remarks
- Two schemes
  - Adaptive time-frequency smoothing scheme
  - Noise properties driven adaptive residual noise floor scheme
- Experiments and conclusions

Background

- Noisy environments
  - Stationary noise: fan noise, white Gaussian noise……
  - Non-stationary noise: car noise, factory noise, airplane noise……
- Effects of noise
  - Reducing both speech quality and speech intelligibility
  - Increasing the listener’s fatigue
- Speech Enhancement
  - Single-Channel Speech Enhancement
  - Multi-Channel Speech Enhancement



Background

- Comparisons of Speech Enhancement through Spatial Filtering and Post-

Algorithm	Spatial Filtering	Post-Filtering	Combination
Directional Noise (Correlated Noise)	☺	☹	☺
Diffuse Noise (Non-Correlated Noise)	☹	☺	☺

- Post-Filtering Algorithms
  - Single-Channel Speech Enhancement
  - Multi-channel post-filtering algorithms: Coherence-based/Phase-based/Power-based
- Com

Algorithm	SC-SE	MC-PF
Stationary Noise	☺	☺
Non-Stationary Noise	☹	☺
Improve Speech Intelligibility	☹	☺

Motivation

- Gain Functions: $W_3(k,l) = [W_1(k,l)]^2 = \dfrac{4|\Phi_{12}(k,l)|^2}{(\Phi_{11}(k,l)+\Phi_{22}(k,l))^2}$

$$W_2(k,l) = \frac{1}{2}\left(\frac{|\Phi_{12}(k,l)|}{\Phi_{11}(k,l)} + \frac{|\Phi_{12}(k,l)|}{\Phi_{22}(k,l)}\right)$$

- Theoretical amounts of noise reduction: (e.g.) $NR_3 = \dfrac{1}{|W_3|^2} = \left(\dfrac{2+\dfrac{1}{SNR_1}+\dfrac{1}{SNR_2}}{4}\right)^2$

- For noise-only segments, in conventional theory, we have

$$NR \to +\infty \quad \text{for } SNR_1, SNR_2 \to 0$$

- **Question:** In practice, $NR << +\infty$? for $SNR_1, SNR_2 \to 0$, why?
- **Question :** How to improve noise reduction without increasing speech distortion in practical consideration.

Model

- Two Assumptions:
  - The observed signals at the two microphones are Gaussian distributed.
  - The noise is uncorrelated between sensors.
- Distribution of the periodogram

$$f_{I_{X_i X_i (k,l)}}(y) = \frac{U(y)}{\sigma^2_{X_i X_i}(k,l)} \exp\left(-\frac{y}{\sigma^2_{X_i X_i}(k,l)}\right)$$

- Distribution of the cross periodogram
  - In theory, the real and the imaginary parts of the cross periodogram are both Laplace distributed.
  - Gaussian approximation

$$f_{\Re\{I_{X_i X_j (k,l)}\}}(y) \propto N\left(0, \mathrm{var}\left(\Re\{I_{X_i X_j (k,l)}\}\right)\right)$$

$$f_{\Im\{I_{X_i X_j (k,l)}\}}(y) \propto N\left(0, \mathrm{var}\left(\Im\{I_{X_i X_j (k,l)}\}\right)\right)$$

Derivation

- Distribution of the auto-spectrum (Using L independent frames)

$$f_{\Phi_{X_iX_i(k,l)}}(y)=\frac{U(y)}{\sigma^2_{X_iX_i}(k,l)\Gamma(L)}\left(\frac{yL}{\sigma^2_{X_iX_i}(k,l)}\right)\exp\left(-\frac{yL}{\sigma^2_{X_iX_i}(k,l)}\right)$$

- Distribution of the cross spectrum (Using L independent frames)

$$f_{\left|\Phi_{X_iX_j(k,l)}\right|}(y)=\frac{yU(y)}{\sigma_1^2(k,l)}\exp\left(-\frac{y^2}{2\sigma_1^2(k,l)}\right)$$

with

$$\sigma_1^2(k,l)=\frac{1}{L}\operatorname{var}\left(\Re\left\{I_{X_iX_j}(k,l)\right\}\right)=\frac{1}{L}\operatorname{var}\left(\Im\left\{I_{X_iX_j}(k,l)\right\}\right)$$

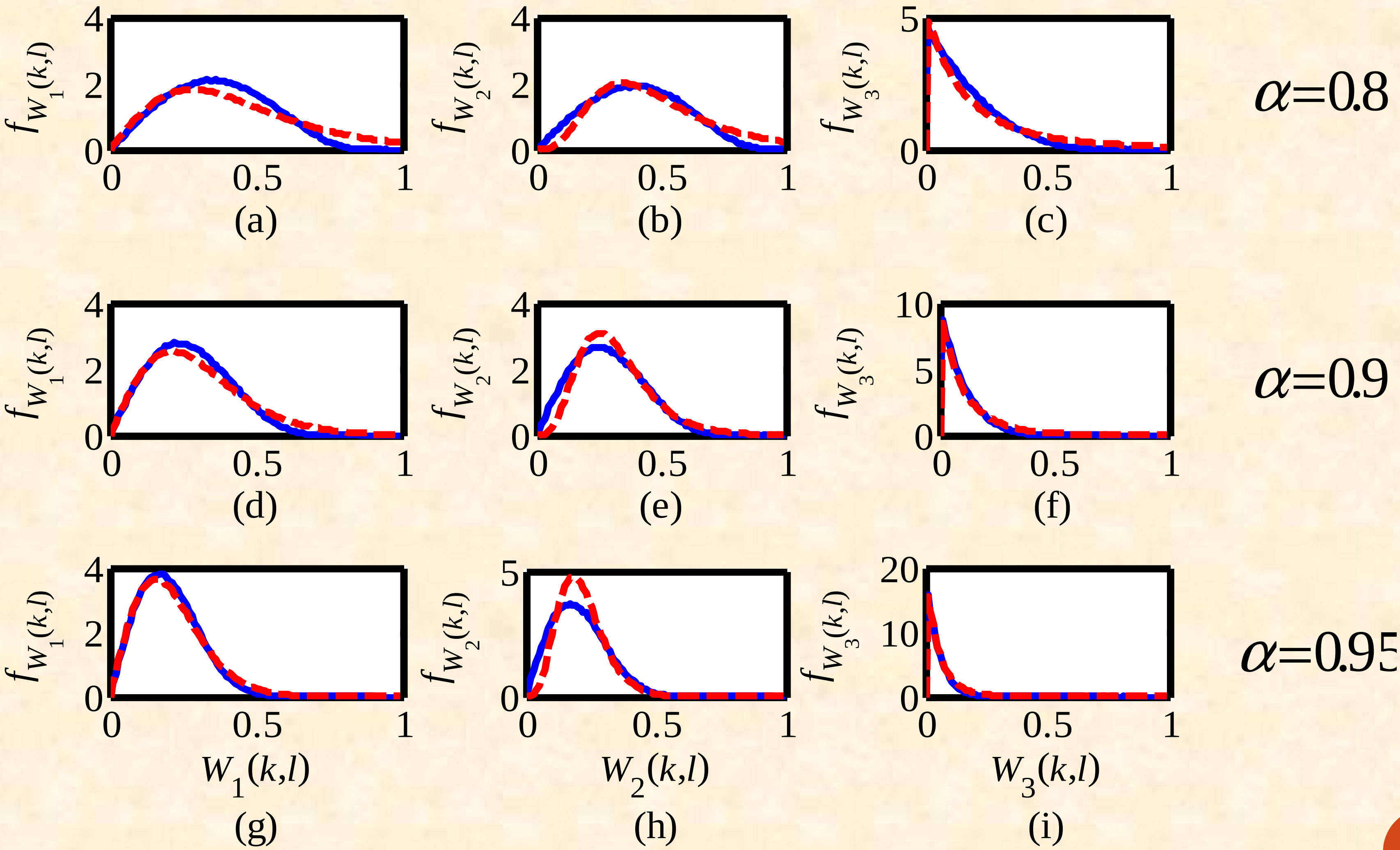
- The relationship between the smoothing factor alpha and the number of frame L

$$L=\frac{1+\alpha}{1-\alpha}\frac{1}{\left(1+2\sum_{l=1}^{\infty}\alpha^l\rho(l)\right)}$$

- With the PDFs of the auto-spectra and the cross-spectrum, the PDF of the gain function can be obtained in theory.

Statistical Properties of the Gain Functions

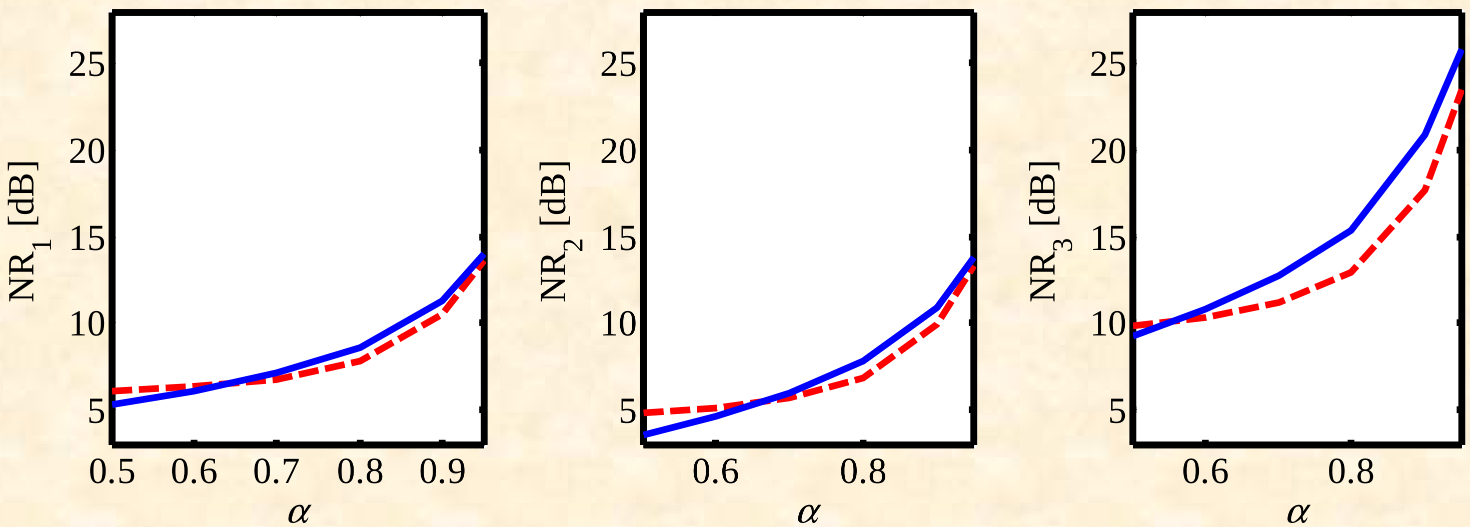
- The PDFs of the gain functions



Statistical Properties of the Gain Functions

- Theoretical limits of noise reduction

$$\mathrm{NR}_i = -20\lg\left(E\{W_i(k,l)\}\right) = -20\lg\left(\int_0^{\infty} y f_{W_i(k,l)}(y)\,dy\right) dB$$



(a) (b) (c)

- By this study, we find that $NR << +\infty$? for $SNR_1, SNR_2 \to 0$
- NR becomes infinity if and only if $\alpha \to 1$

Some remarks

- If the smoothing factor is not large enough, musical noise is inevitable due to a large part of the gain having large values.
- It is better to use a large value of the smoothing factor to increase both noise reduction and reduce musical noise.
- We have to make a trade-off between the noise reduction and the estimation biases by properly selecting the smoothing factor.

Adaptive Time-Frequency Smoothing Scheme

- For the TC-PF algorithms, the sudden change of the system only occurs at the desired speech onsets/offsets.
- For the desired speech onsets:

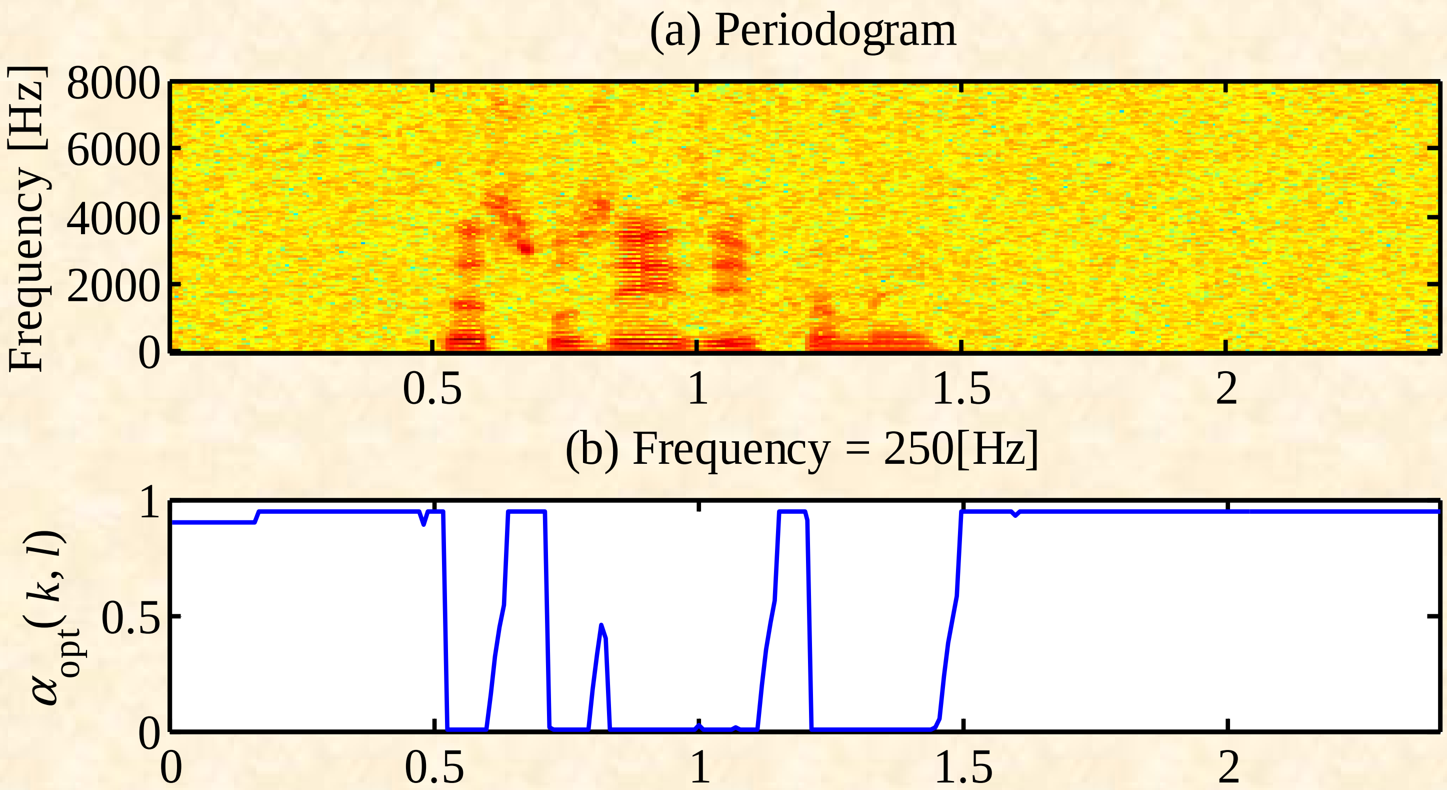
$$\alpha(k,l) = \max\left\{\min\left\{\frac{1}{1+\xi^2(k,l)}, \alpha_{\max}\right\}, \alpha_{\min}\right\}$$

- For the desired speech offsets:

$$\alpha_{\text{opt}}(k,l) = \begin{cases} \beta\alpha_{\text{opt}}(k,l-1) + (1-\beta)\alpha(k,l) & \\ \qquad\qquad \text{if } \alpha_{\text{opt}}(k,l-1) < \alpha(k,l) \;\&\; \alpha_{\text{opt}}(k,l-1) < \alpha_{\text{mid}} \\ \alpha(k,l) \quad \text{otherwise} \end{cases}$$

Adaptive Time-Frequency Smoothing Scheme

- Results of the proposed adaptive time-frequency smoothing scheme.
  - At the desired speech onsets/offsets, the smoothing factor has a small value, which is close to zero.
  - At the noise-only segments, the smoothing factor is close to one.



Adaptive residual noise floor selection

- To mask musical noise, conventional speech enhancement algorithms often use a constant residual noise floor, given by:

$$W_i(k,l) = \max\{W_i(k,l), G_{\min}\}$$

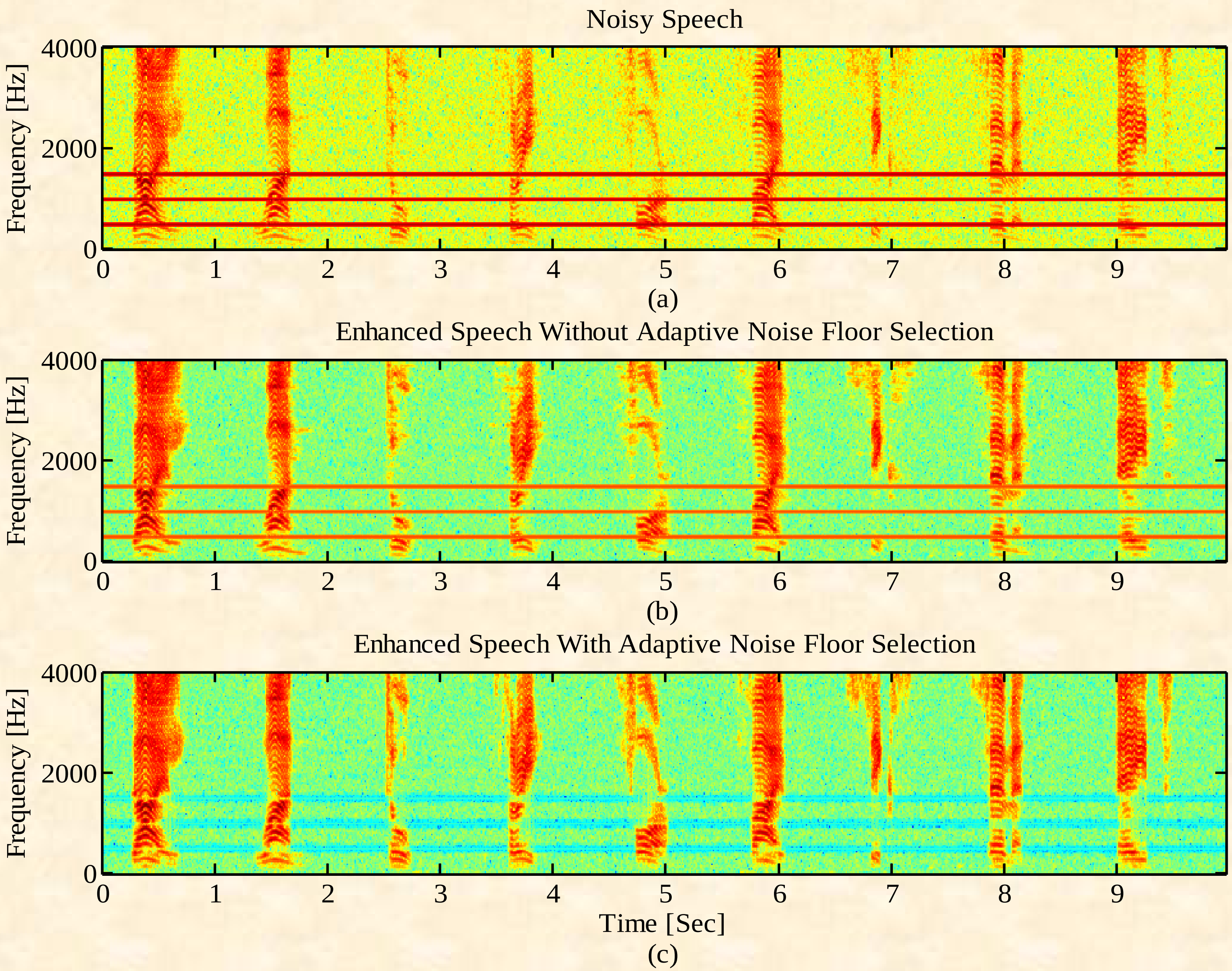
- Based on the psychoacoustic fact, it is difficult for a tone to mask a band of noise. Thus, it is better to further suppress the tonal components.

$$G_i(k,l) = \max\{(1-U(\lambda(k,l)-1))W_i(k,l)U(\lambda(k,l)-1)G_H(k,l), G_{\min}(k,l)\}$$

- $\lambda(k, l)$ is used to indicate the tonal components.
- By using $G_i(k, l)$, the noise reduction could be improved without introducing audible speech distortion, since we only increase the amount of noise reduction at the peaks of the estimated NPSD.

Adaptive residual noise floor selection

- Experimental Results:

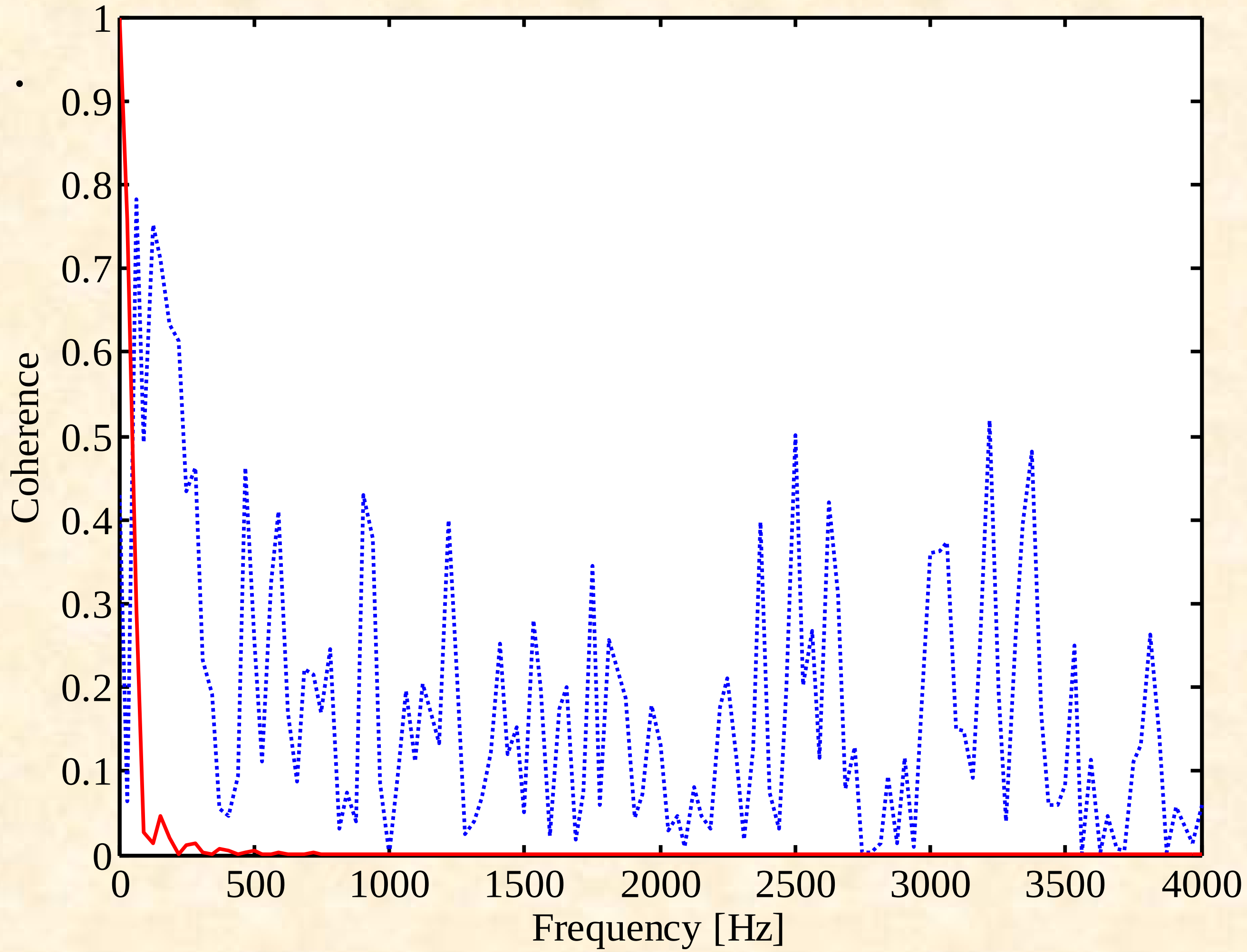


Experiments

- Experiment Setup
  - We use two Omni-directional microphones produced by Goertek
  - The distance between the two microphones is 0.5m
  - The reverberation distance of the room is about 1m
  - The noise speaker is located about 4m away from the center of the two microphones
  - The desired speech is located in front of the center microphone at a distance of 0.5m.

Experiments

- Coherence of the noise between the two microphones



Experiments

- Comparison results of the SegSNR improvement

Algorithm	ARMA(2,2) Non-stationary Noise				AR(24) Non-stationary Noise			
Input SegSNR	0	5	10	15	0	5	10	15
TTC-PF	6.03	4.4	2.12	-0.40	6.21	4.80	2.90	0.44
Proposed	6.51	4.81	2.40	0.16	7.35	5.64	3.63	0.88

- Comparison results of the PESQ improvement

Algorithm	ARMA(2,2) Non-stationary Noise				AR(24) Non-stationary Noise			
Input SegSNR	0	5	10	15	0	5	10	15
TTC-PF	0.44	0.5	0.52	0.41	0.43	0.54	0.54	0.41
Proposed	0.54	0.70	0.68	0.59	0.63	0.75	0.72	0.51

Conclusions

- ***Answer:*** In practice, $NR < +\infty$? for $SNR_1, SNR_2 \rightarrow 0$, why?
  - From statistical properties of the post-filter estimators.
- ***Answer :*** How to improve noise reduction without introducing audible speech distortion in practical consideration.
  - Adaptive time-frequency smoothing scheme
  - Noise properties driven adaptive residual noise floor selection scheme
  - The two schemes could be applied to any speech enhancement that need to estimate the auto/cross spectrum and the gain function.

Thanks for your Attention