Institut für Kommunikationsakustik

RUB

An Evaluation of Noise Power Spectral Density Estimation Algorithms in Adverse Acoustic Environments

Jalal Taghia, Jalil Taghia, Nasser Mohammadiha, Jinqiu Sang, Vaclav Bouse, Rainer Martin

May 27, 2011

AUDIS
Digital Signal Processing in Audiology

Outline

- Motivation
- Overview of algorithms
- Evaluation measures
  - Mean estimation error
  - Estimation error variance
- Experimental results
- Conclusions

Motivation

- Noise power estimation, a crucial part of speech enhancement
- Effective on quality and intelligibility of enhanced speech
- Many new noise power estimators available
- Framework aims:
  - Presenting performance of some recent and some well-known noise estimators
  - New measure for more comprehensive evaluation of performance

Overview of algorithms

- Minimum statistics (**MS**) [Martin, 2001]
- Minima-controlled recursive averaging (**MCRA**) [Cohen, 2002]
- 3 other algorithms belonging to MCRA category
  - Improved minima-controlled recursive averaging (**IMCRA**) [Cohen, 2003]
  - **EMCRA** [Fan et al., 2007]
  - **MCRA-MAP** [Kum et al., 2009]
- Subspace noise tracking (**SNT**) [Hendriks et al., 2008]
- 2 algorithms based on minimum mean-squared error (MMSE) estimation
  - **MMSE-Yu** [Yu, 2009]
  - **MMSE-Hendriks** [Hendriks et al., 2010]

Evaluation measures

- Two issues taken into account:
  - Evaluation shall be independent of speech enhancement system
    - To separate effects of any specific speech enhancement system
  - Having a suitable reference noise is necessary, since
    - During speech activity instantaneous noise power is not available
    - Most noise reduction approaches require a smoothed noise estimate
    - To reduce impact of random fluctuations in original noise periodogram

Evaluation measures ...

- Mean estimation error: averaged log distance $LogErr_{mean}$ between the estimated noise PSD $\hat{\delta}_D^2$ and reference noise PSD δ_D^2

$$LogErr_{mean} = \frac{1}{IK}\sum_{i=1}^{I}\sum_{k=1}^{K}\left|10\log_{10}\left[\frac{\delta_D^2(k,i)}{\hat{\delta}_D^2(k,i)}\right]\right|$$

$I \rightarrow$ **Number of frames**

$K \rightarrow$ **Number of frequency bins**

- Estimation error variance

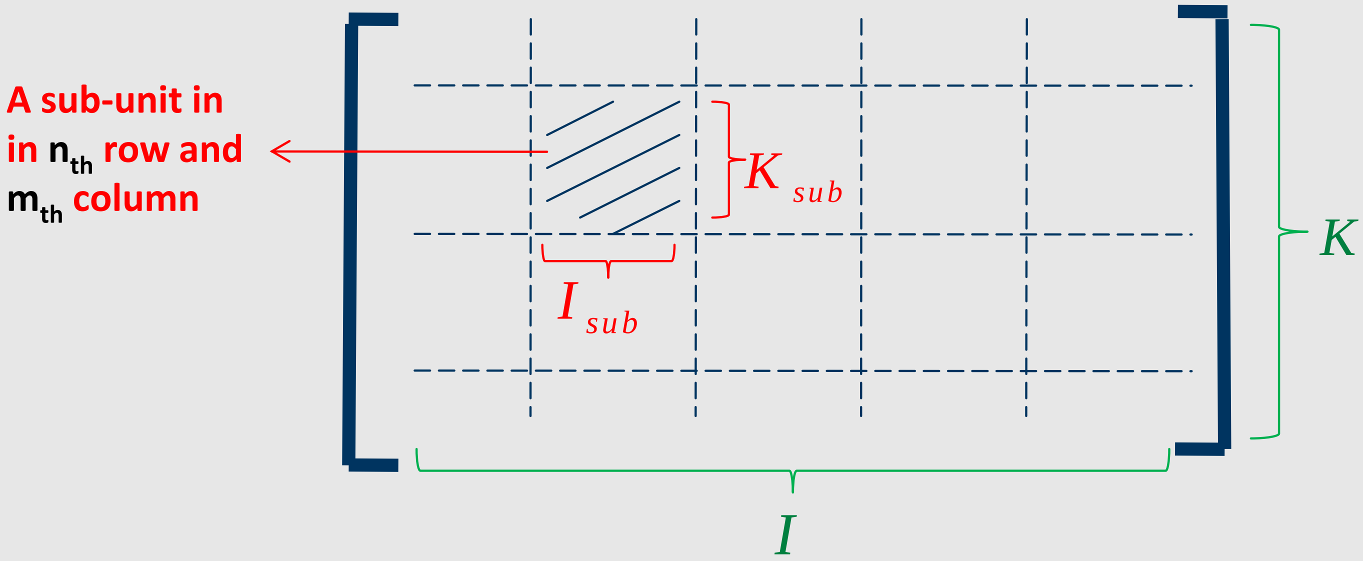
$$LogErr_{var} = \frac{1}{N\,M}\sum_{n=1}^{N}\sum_{m=1}^{M} Variance\left(\left|10\log_{10}\left[\frac{\delta_D^2}{\hat{\delta}_D^2}\right]\right|\right)^{n,m}$$

$M \rightarrow$ **Number of sub-units in the column**

$Variance(.)^{n,m} \rightarrow$ **The variance computed for the sub-unit in the n_{th} row and m_{th} column**

$N \rightarrow$ **Number of sub-units in the row**

Variance computation



$$Variance\left(\left|10\log_{10}\left[\frac{\delta_D^2}{\hat{\delta}_D^2}\right]\right|\right)^{n,m} = \frac{1}{I_{sub}\,K_{sub}} \sum_{i=(m-1)I_{sub}+1}^{m\,I_{sub}} \left[\sum_{k=(n-1)K_{sub}+1}^{n\,K_{sub}} \left(\left|10\log_{10}\left[\frac{\delta_D^2(k,i)}{\hat{\delta}_D^2(k,i)}\right]\right| - \mu_i^n \right)^2 \right],$$

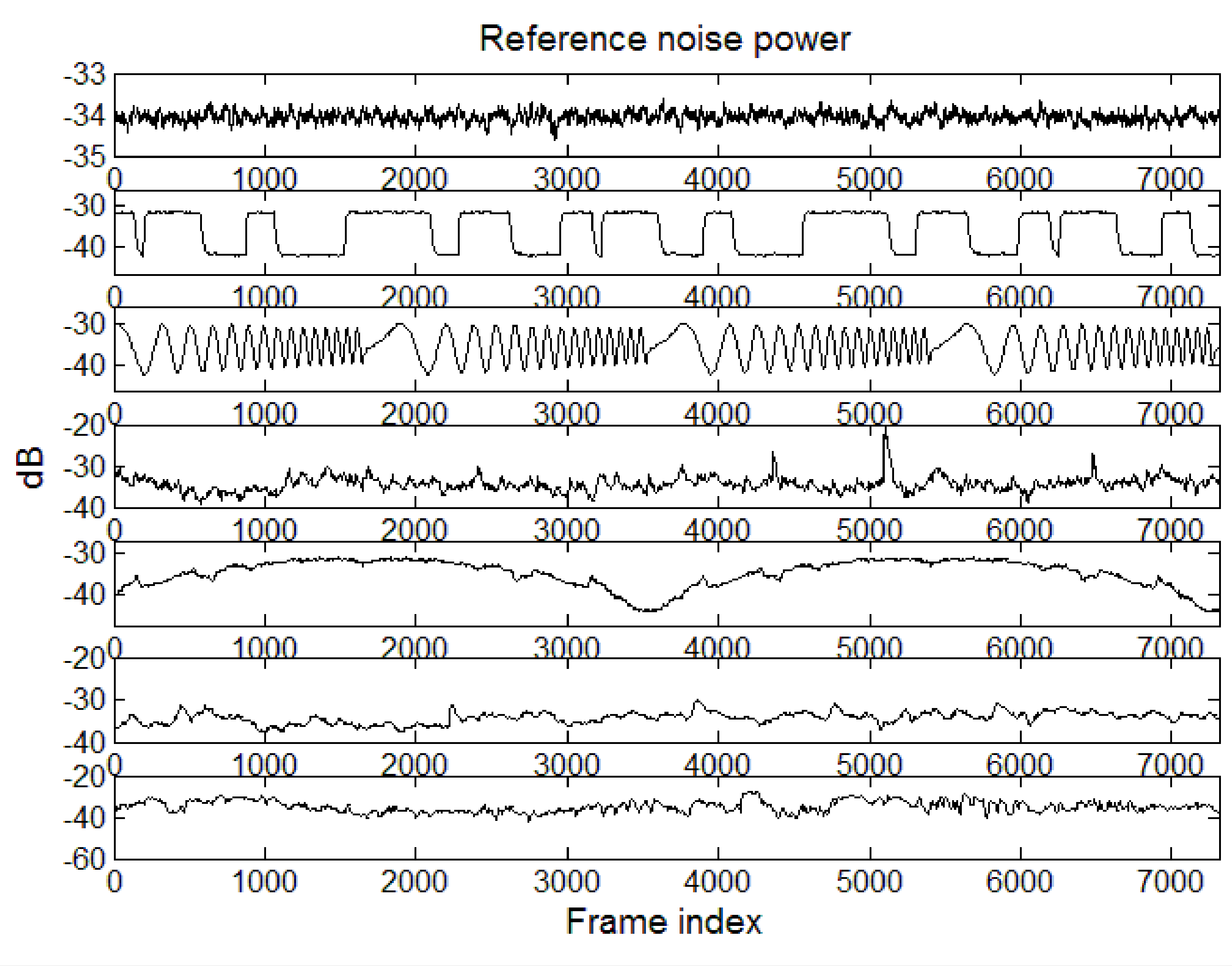
$$\mu_i^n = \frac{1}{K_{sub}} \sum_{k=(n-1)K_{sub}+1}^{n\,K_{sub}} \left(\left|10\log_{10}\left[\frac{\delta_D^2(k,i)}{\hat{\delta}_D^2(k,i)}\right]\right| \right)$$

In our experiments: $I_{sub} = 15$ and $K_{sub} = 10$

Experimental results

- 8 algorithms are considered
- Sampling frequency of all signals is 8 KHz
- Window length as well as the DFT length 256 samples
- Clean speech signals from TIMIT database
  - One female speech and one male speech; each one with 2 minutes length
- 7 different types of noise signals (taken from SOUND-IDEAS)
  - WGN, VaryingStep WGN, Sinusoidally modulated WGN, Babble, Car, Traffic1, Traffic2
  - The range of input SNR is from -5 dB to 20 dB
- Reference noise:
  - Several methods for smoothing were tested
  - Finally, a recursive temporal smoothing of noise periodograms was found to be more appropriate (smoothing factor 0.9)

Experimental results …



WGN

Varying-Step WGN

Sinusoidal WGN

Babble noise

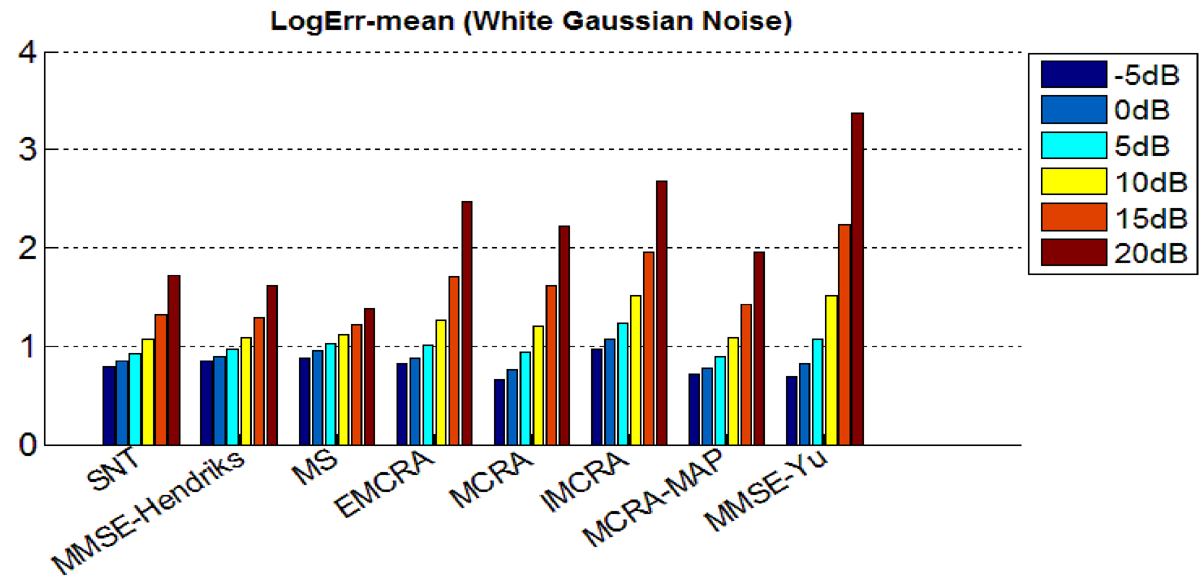
Car noise

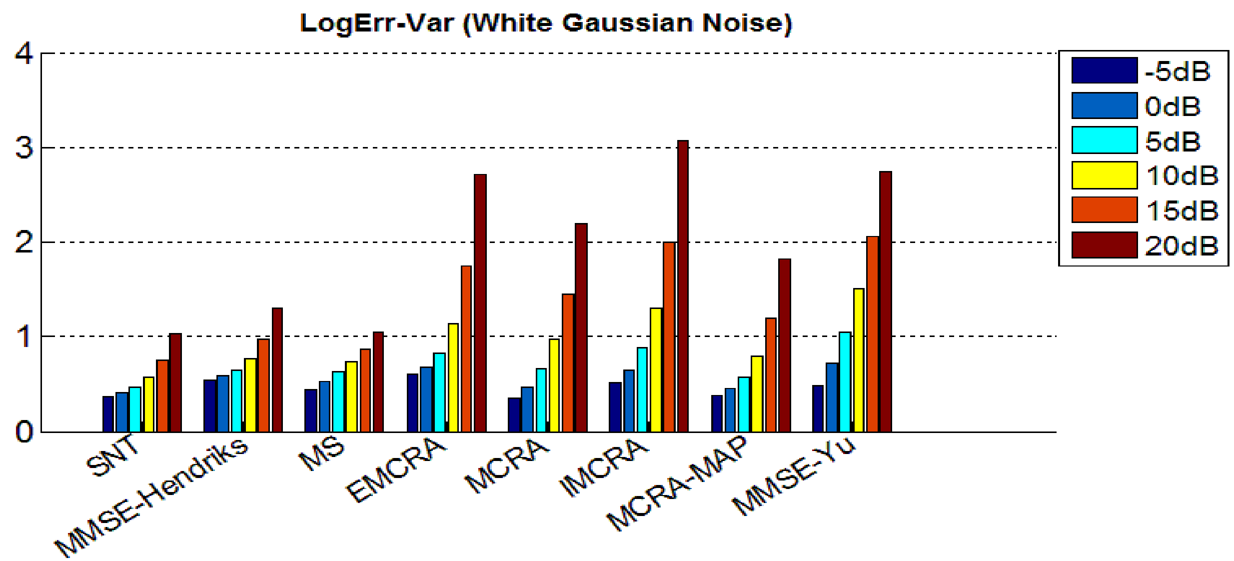
Traffic1 noise

Traffic2 noise

Experimental results …

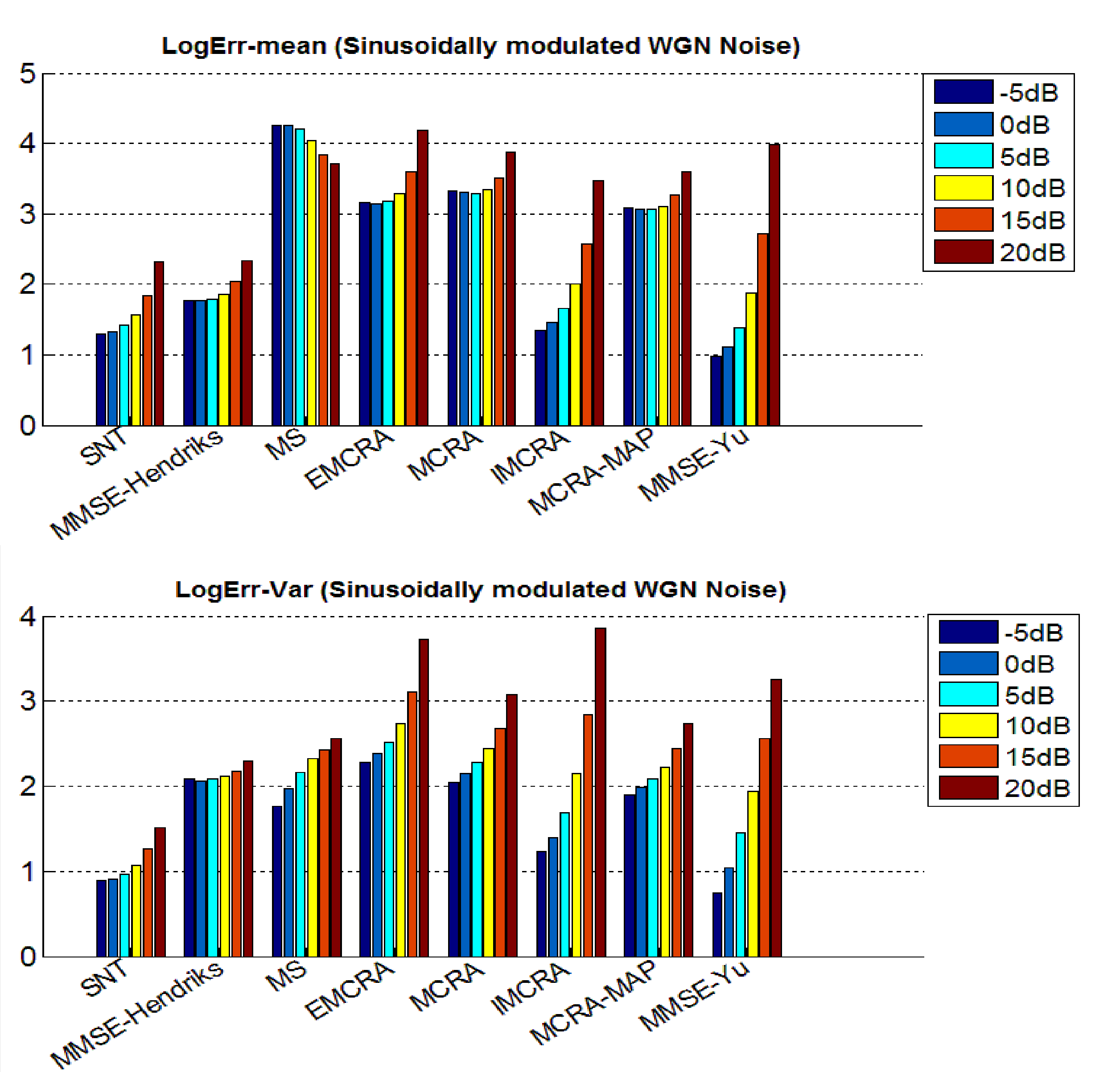
Performance results of 8 algorithms in the case of WGN in terms of LogErr-mean and LogErr-Var





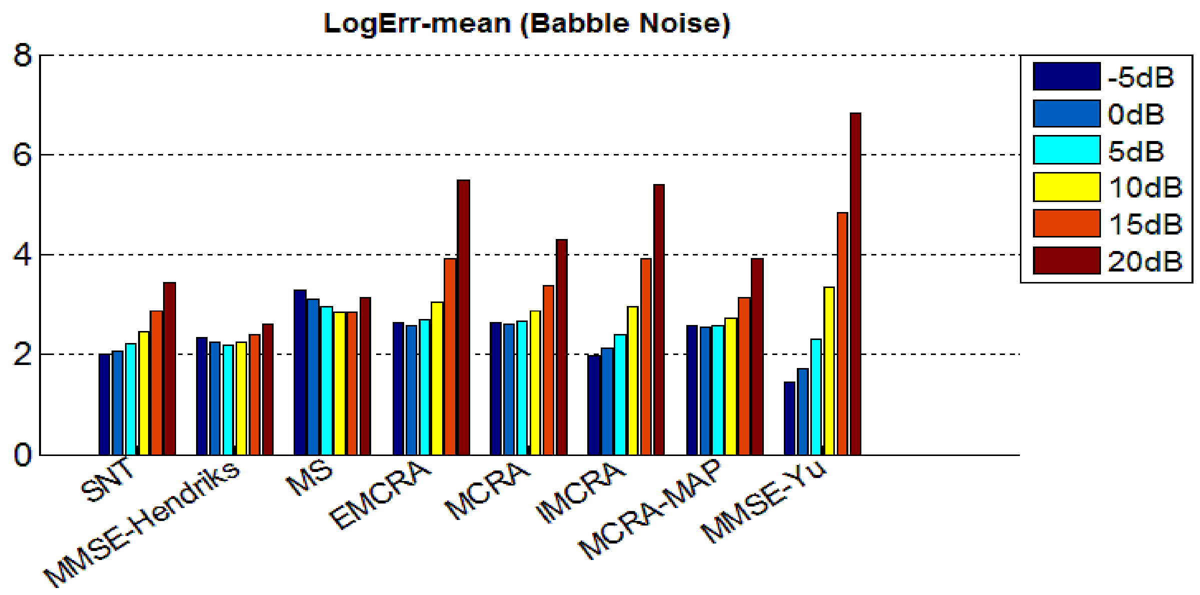
Experimental results …

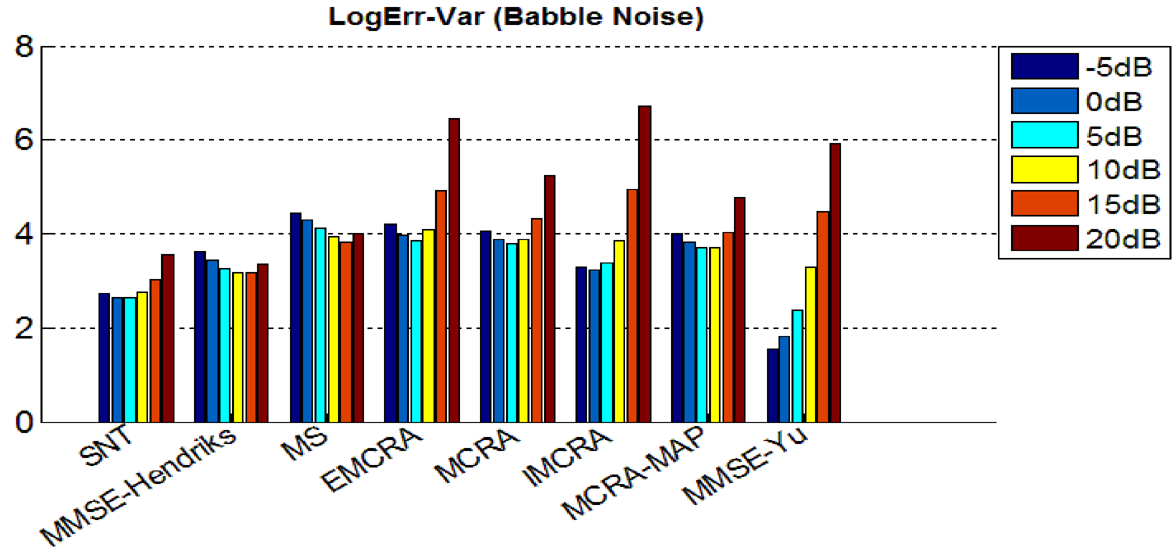
Performance results of 8 algorithms in the case of Sinusoidally modulated noise in terms of LogErr-mean and LogErr-Var



Experimental results …

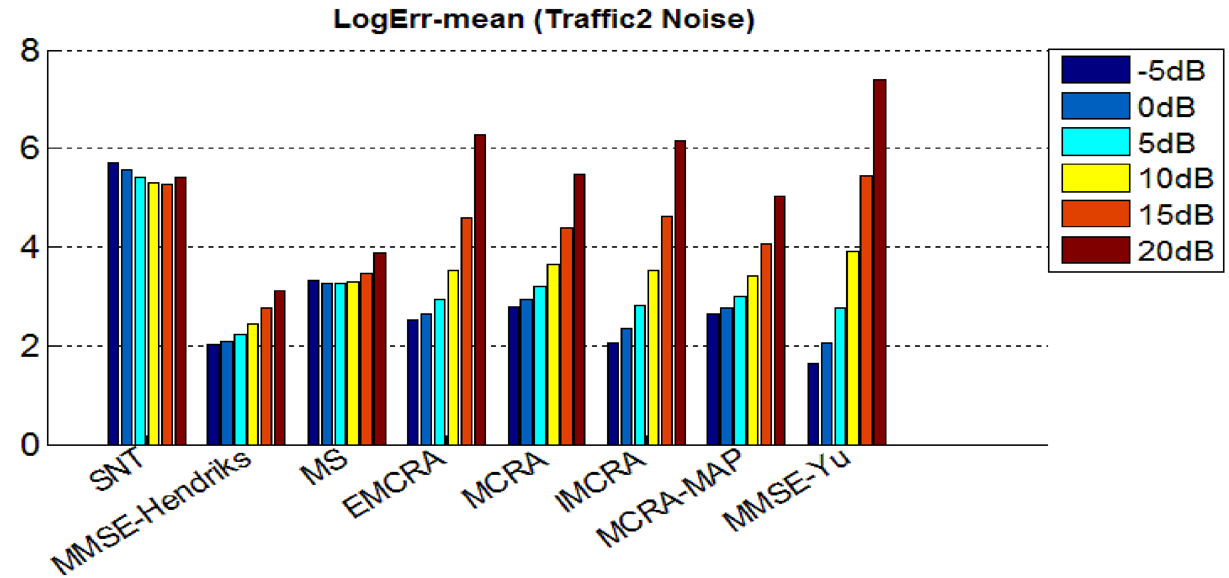
Performance results of 8 algorithms in the case of Babble noise in terms of LogErr-mean and LogErr-Var

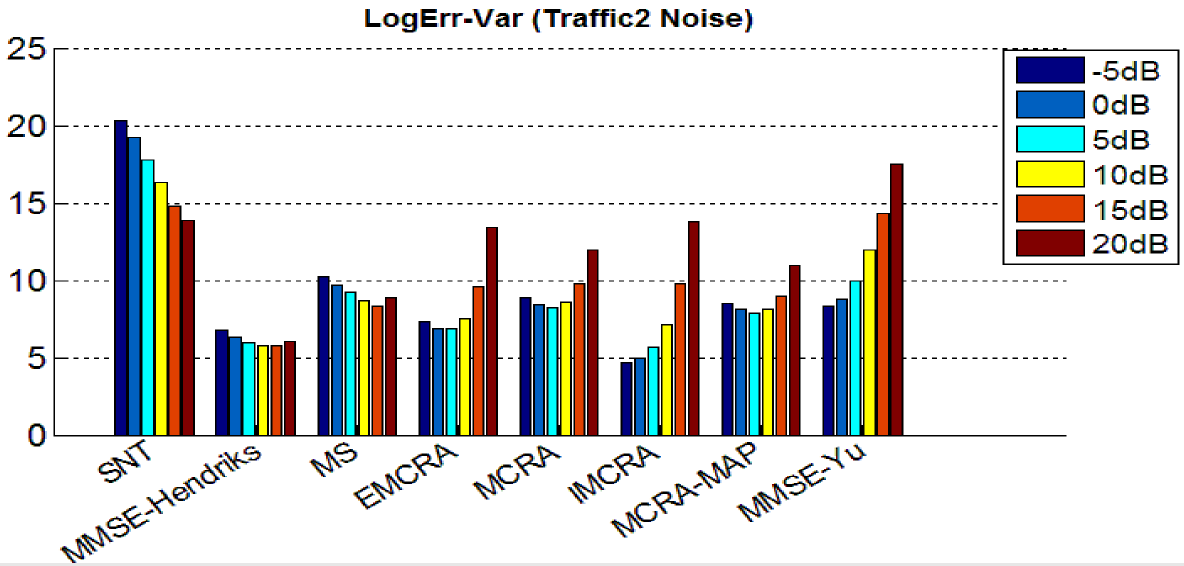




Experimental results …

Performance results of 8 algorithms in the case of Traffic2 noise in terms of LogErr-mean and LogErr-Var

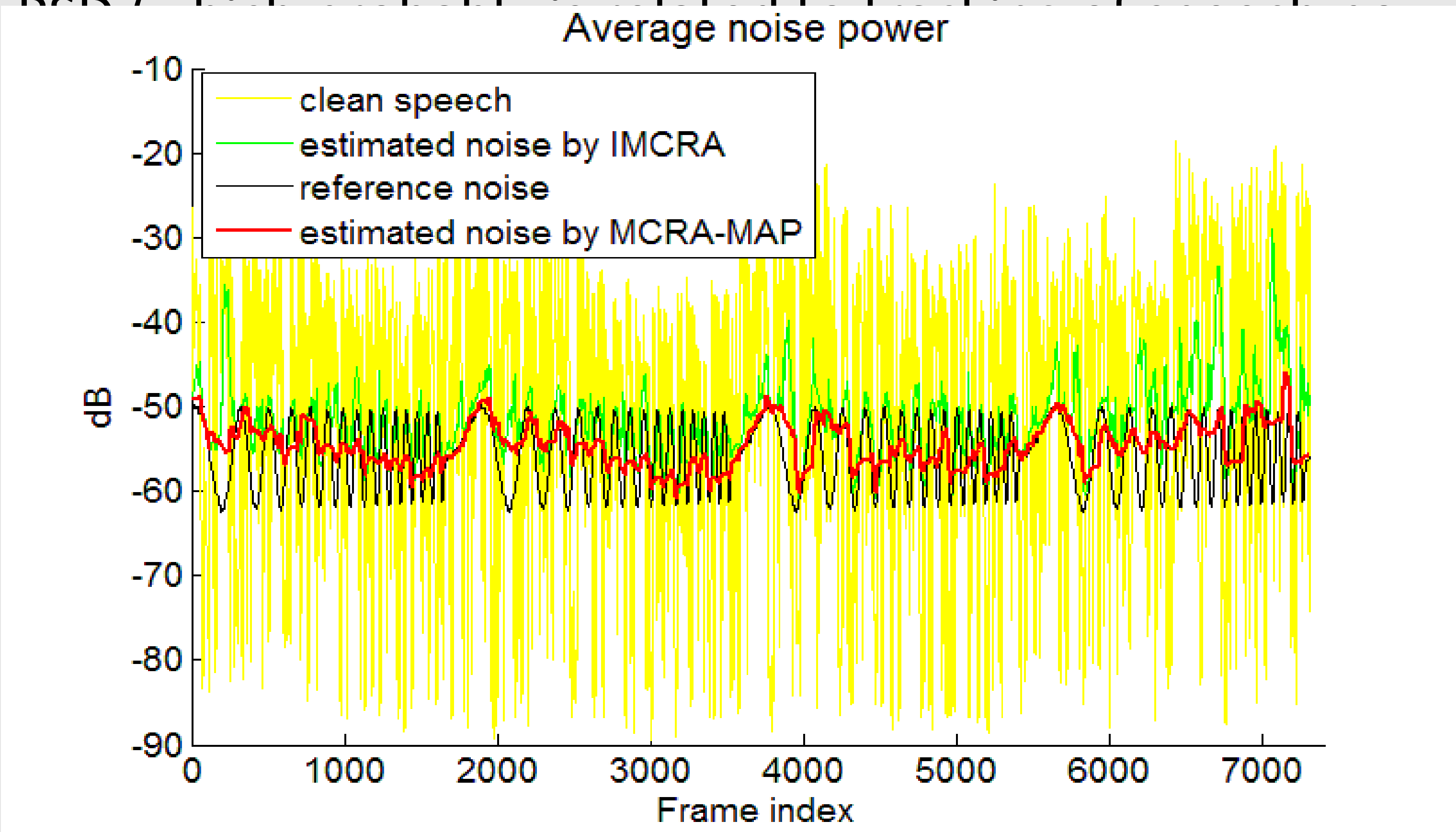




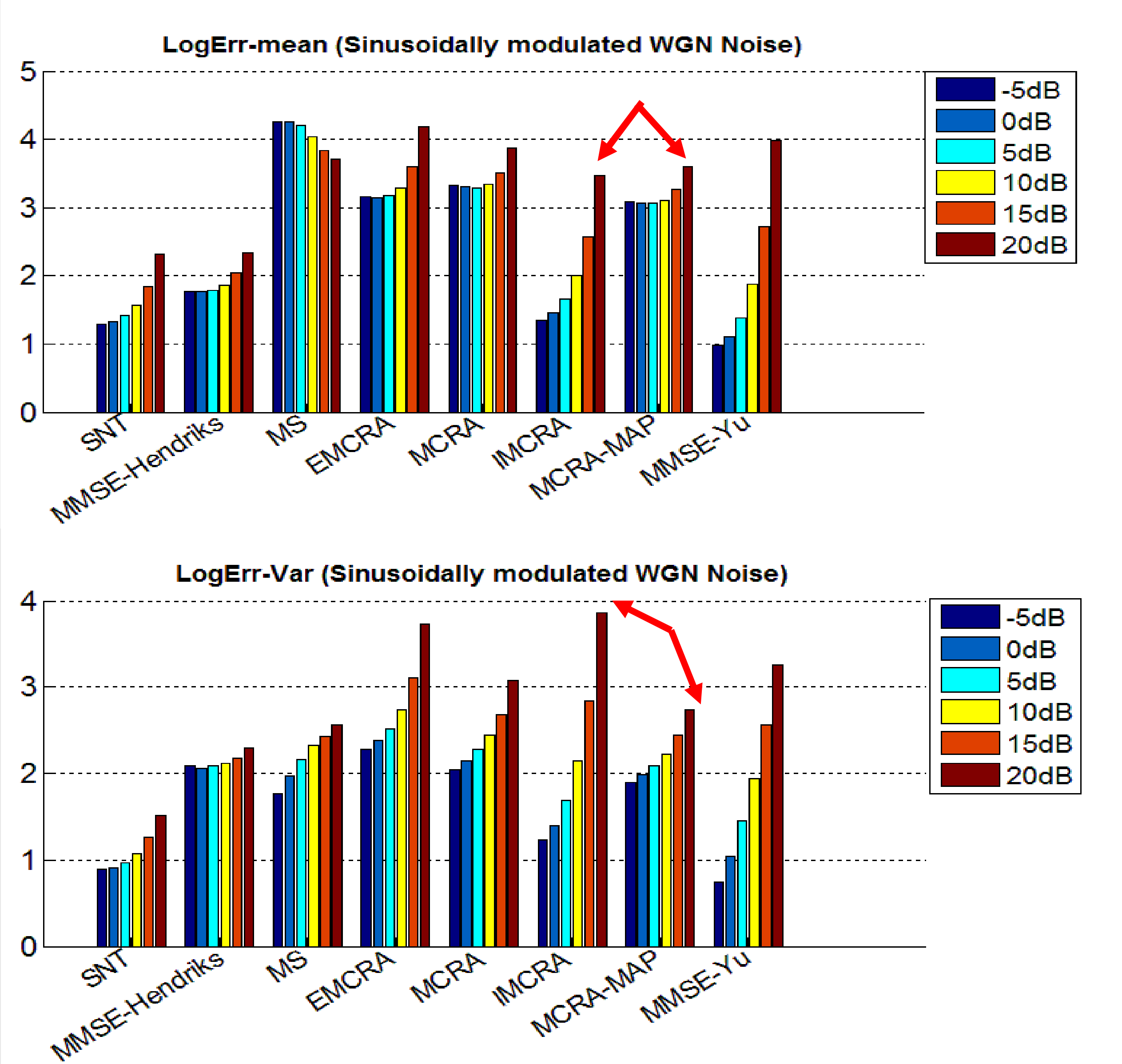
Experimental results …

- Estimation error variance gives additional insight
  - It measures the amount of fluctuations in the estimated noise PSD (… er)
  - … ance in

Female speech degraded by Sinusoidally modulated WGN with 20dB input SNR



Experimental results …



Conclusions

- Some of noise power estimators are very susceptible to the level of input SNR
- **Estimation error variance** allows us to measure *amount of fluctuations* in tracking noise power, and perhaps producing *musical noise*
- For non-stationary noise a few methods show to be robust
  - MMSE-Hendriks → the most robust noise power estimator according to our experiments

Reference

- **Martin** R., “Noise power spectral density estimation based on optimal smoothing and minimum statistics,” IEEE Trans. Speech Audio Process., vol. 9, no. 5, pp. 504–512, 2001.
- **Cohen** I., and B. Berdugo, “Noise estimation by minima controlled recursive averaging for robust speech enhancement,” IEEE Signal Processing Letters, vol. 9, no. 1, pp. 12–15, 2002.
- **Cohen** I., “Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging,” IEEE Trans. Speech and Audio Processing, vol. 11, no. 5, pp. 466–475, September 2003.
- **Kum** J. M., Y. S. Park, and J. H. Chang, “Speech enhancement based on minima controlled recursive averaging incorporating conditional maximum a posteriori criterion,” IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4417–4420, 2009.
- **Fan** N., J. Rosca, and R. Balan, “Speech noise estimation using enhanced minima controlled recursive averaging,” IEEE Int. Conf. On Acoustics, Speech and Signal Processing, pp. IV–581–IV–584, June 2007.
- **Hendriks** R. C., J. Jensen, and R. Heusdens, “Noise tracking using DFT domain subspace decompositions,” IEEE Trans. Audio Speech and Language Processing, vol. 16, no. 3, pp. 541–553, March 2008.
- **Hendriks** R. C., R. Heusdens, and J. Jensen, “MMSE based noise PSD tracking with low complexity,” IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4266–4269, March 2010.
- **Yu** R., “A low-complexity noise estimation algorithm based on smoothing of noise power estimation and estimation bias correction,” IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 4421–4424, April 2009.