Quality assessment of speech enhanced using particle filters

School of Information Technology and Engineering

Frederic Mustiere, Martin Bouchard, Miodrag Bolic

{mustiere,bouchard,mbolic}@site.uottawa.ca

Université d'Ottawa • University of Ottawa

Particle Filtering and Speech enhancement

▶ Particle Filters (PFs): a family of approximate solutions to the generic state estimation problem:

$$x_k = f(x_{k-1}, w_k)$$
$$z_k = h(x_k, v_k)$$

▶ Based on Monte Carlo simulation: draw candidates for the state, assign scores to these candidates, and form a global estimate.

▶ Application to speech denoising via the use of an autoregressive (AR) speech model → part of the model is linear, which allows for the use of Rao-Blackwellisation. The resulting algorithm will hereafter be referred to as RBPF.

▶ Among the numerous existing speech enhancement methods, where do the PF and RBPF approaches stand? What are their strengths and weaknesses?

Challenges of Speech enhancement

▶ Tradeoff between the amount of noise reduction and the resulting speech intelligibility.

▶ How to measure the quality of the enhanced speech?

► Important factors to consider: speech naturalness, and the degree of noise intrusiveness. (E.g., importance in hearing aids devices).

► *Interspeech residual noise*: the noise appearing during pauses in the enhanced speech.

► *Intraspeech residual noise*: the noise still present during an utterance.

Use of speech quality measures:

▶ The overall signal-to-noise ratio (OSNR)

Merely indicates that some enhancement is taking place

▶ The average segmental SNR (ASSNR)

Is good for measuring the amount of interspeech residual noise

▶ The perceptual evaluation of speech quality score (PESQ)

Is a good indicator of the overall speech quality

▶ The log-area ratio (LAR)

Represents the speech naturalness better

Algorithms used for comparison, and experimental conditions

▶ AR model of order 6, white Gaussian observation noise with known variance → simpler for comparisons.

Algorithms used:

▶ PF – Particle Filter: a "plain" PF using the importance density described in [1], with 2000 particles, and a fixed-lag smoothing method with lag equal to 8.

▶ RBPF – Rao-Blackwellised Particle Filter: similar to that described in [2], with 600 particles and a lag of 8.

▶ SSUB – Spectral subtraction: the classical scheme, using the first few frames to estimate the noise spectrum

▶ KF+EM – Kalman Filter + EM algorithm: a Kalman filter-based algorithm using an EM algorithm, with 20 iterations, to update the speech parameters. This algorithm is described in [3]

▶ MMSE-STSA – Minimum mean square error log-spectral amplitude estimation: a method presented in [4], also using the first few frames to estimate the noise spectrum

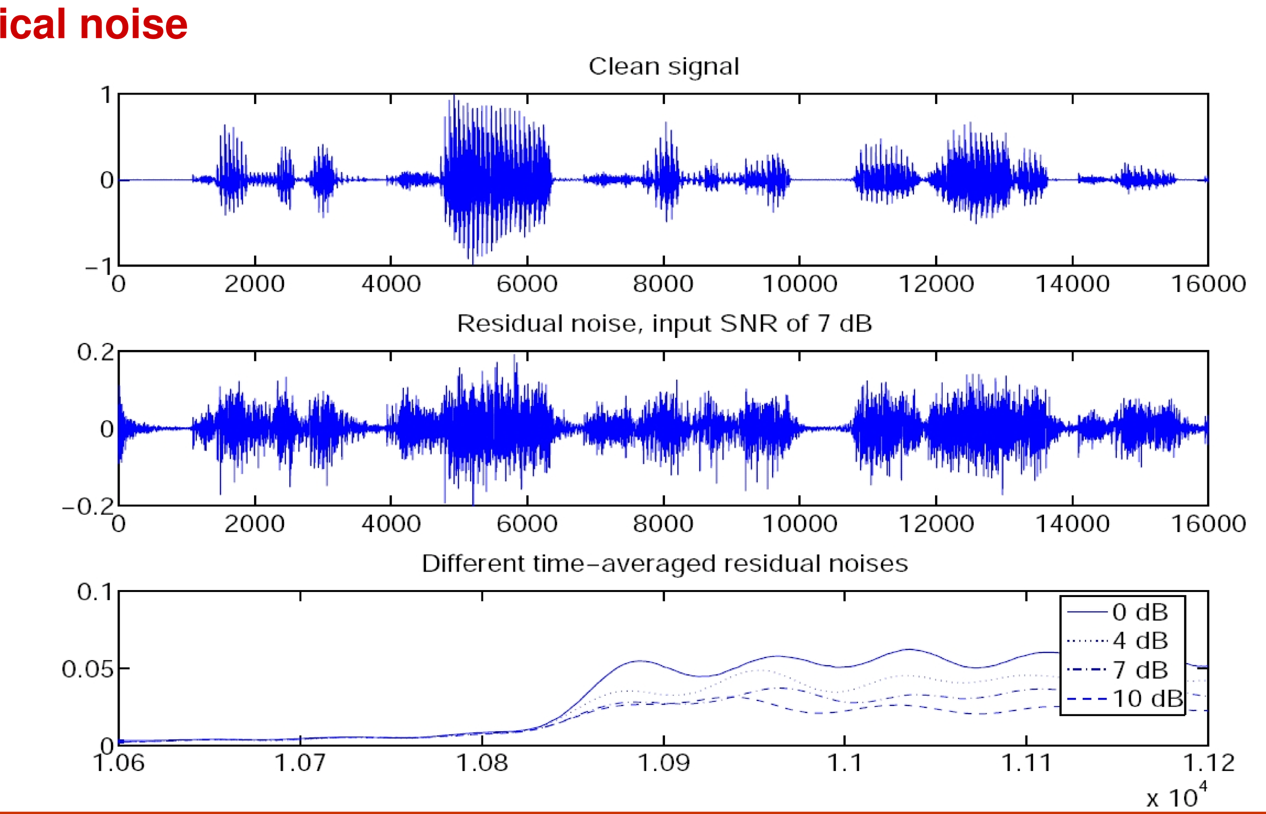
▶ WF+ASNRE – Wiener filter using a priori SNR estimation: an algorithm presented in [5], again using the first few frames to characterize the noise

Quality Assessment of PF-enhanced speech

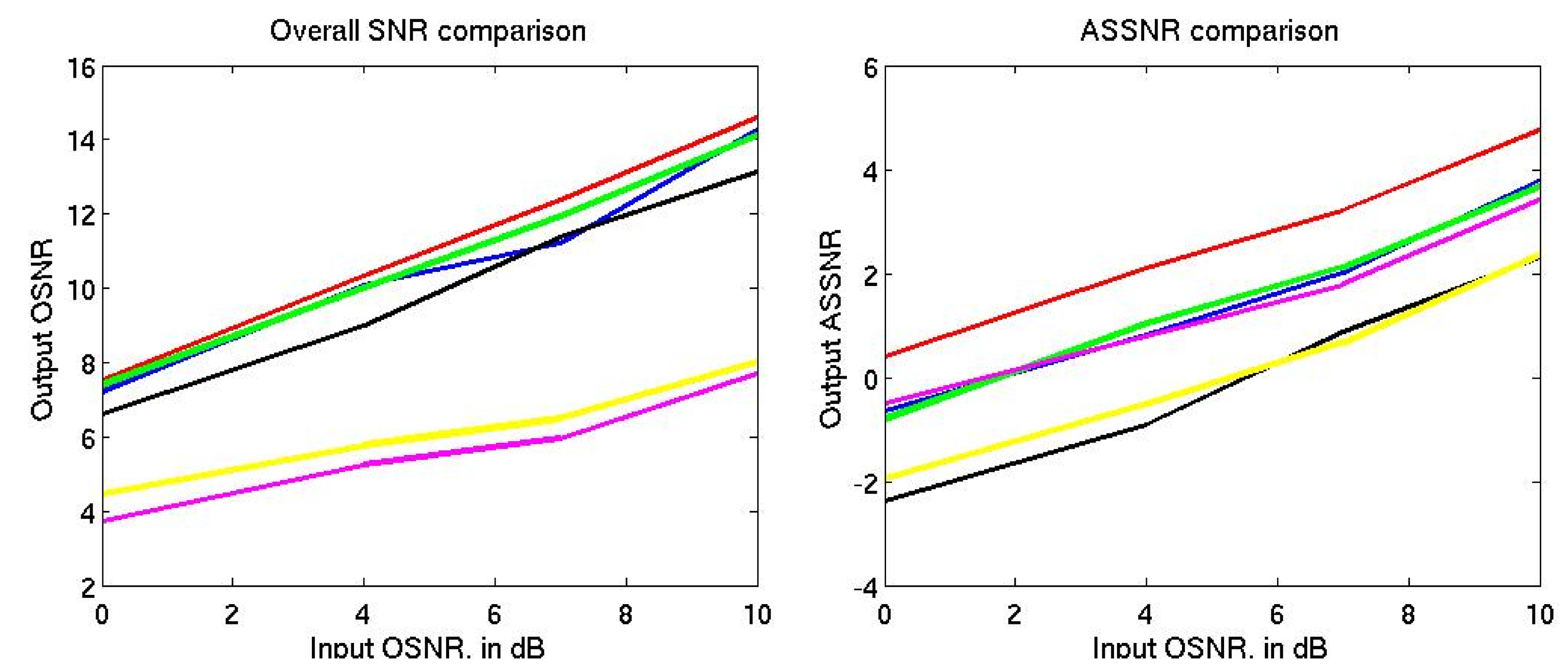
▶ Residual noise: can be identified as a white noise with instantaneous power modulated by the output speech.

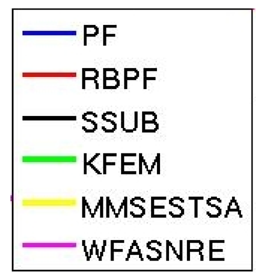
▶ Implications: almost no interspeech noise, but some noise during speech

▶ However, the residual noise does not sound *artificial* — there is no musical noise



Comparison with other algorithms



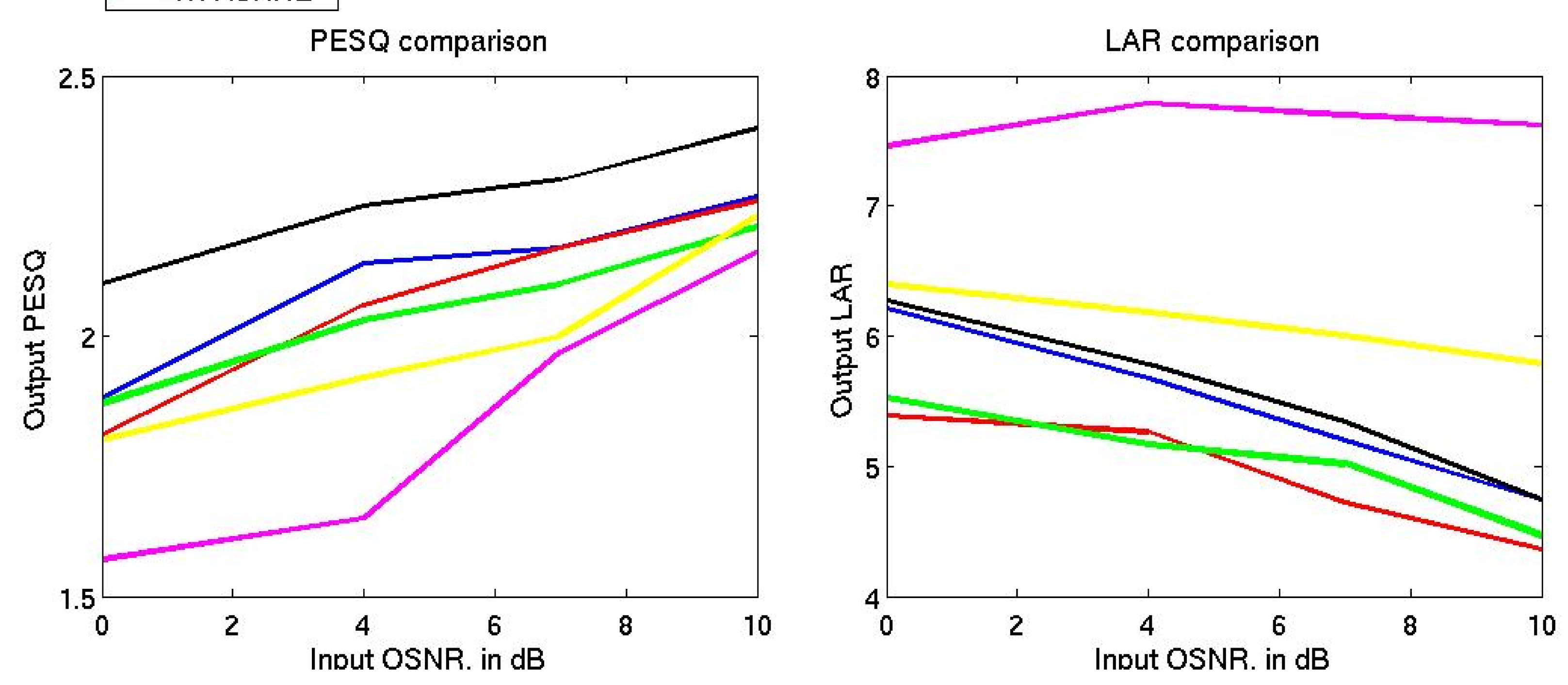


▶ SSUB = spectral subtraction

▶ KFEM = Kalman filter + EM algorithm

▶ MMSESTSA = Minimum mean square error log-spectral amplitude estimator

▶ WFASNRE = A priori SNR estimation



▶ RBPF: best OSNR, ASSNR, and LAR – especially the ASSNR: very low interspeech residual noise. Consistently better than PF (except low SNR PESQ)

▶ For PESQ, although SSUB has the best score, all other scores are lower: good intelligibility, but excessive residual noise and penalized naturalness

Audio Demonstration

Feel free to listen to our demonstrations in order to forge your own opinion on the quality of PF and RBPF-enhanced speech.

The source code is available at www.site.uottawa.ca/~mustiere/

References

[1] S. J. Godsill, A. Doucet, and M.West, "Monte carlo smoothing for nonlinear time series," Journal of the American Statistical Association, vol. 99, no. 465, pp. 156–168, March 2004.

[2] J. Vermaak, C. Andrieu, A. Doucet, and S.J. Godsill, "Particle methods for bayesian modeling and enhancement of speech signals," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 3, pp. 173–85, March 2002.

[3] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and sequential kalman filter-based speech enhancement algorithms," IEEE Transactions on Speech and Audio Processing, vol. 6, no. 4, pp. 373–385, July 1998.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 33, no. 2, pp. 443–5, April 1985.

[5] P. Scalart and J.V. Filho, "Speech enhancement based on a priori signal to noise estimation," in 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, Atlanta, GA, USA, 1996, pp. 629–32.