



Particle Filtering for Speech Enhancement

Particle Filters (PFs): sequential Monte Carlo methods, **approximate solutions to state estimation problem**, they can handle **non-Gaussian non-linear models**.

They operate by:

- Drawing candidates for the state to estimate
- Assigning scores to these candidates (based on model & measurement)
- Forming a global estimate.

Speech enhancement or de-noising can be formulated in **state-space form**, thus **PFs** have been used for this purpose, with **very good reported results in AWGN**.

However they have been **so far**:

- **More computationally expensive** than traditional solutions
- **Not able to endure "real-world" situations** (complex noise, low SNR, etc.)

Main Objective and Constraints

The main objective is to **obtain a PF-based algorithm** with the following **constraints**:

- (1) Capable of operating in **complex noise conditions**, not restricted to stationary noise, and/or using ideal voice activity detectors (VADs)
- (2) As "**light**" as possible computationally
- (3) Able to work for wideband speech (**0-10 kHz bandwidth** in this work)
- (4) Resulting method **must perform favourably** in comparison with well-reviewed and well-established algorithms.

Proposed Solution Overview

In the proposed solution:

- **Subband domain processing** (reduces required processing rate, smaller/lighter models can be used in each subband)
- **Elementary subband PF units:** very light, can operate reliably with less than a dozen particles each
- **Flexibility and robustness** in terms of noise handling, with **internal and/or external noise statistics estimation possible**
- **Psychoacoustic/perceptual constraining** can be incorporated with ease.

Subband Domain Processing

- Use of **maximally decimated filterbanks**, to process minimal amount of data
- "Medium" amount of bands (32) for time-frequency resolution compromise.
- **Three tested filterbank configurations**
 - Multi-stage Wavelet-Packet decomposition of depth 5 (24 coeffs. each)
 - Pseudo-QMF filterbank with a Kaiser prototype window (468 coeffs.)
 - Modified Discrete Cosine Transform (MDC7) with a Kaiser-Bessel Derived window (64 coeffs.).

Elementary Subband Particle Filter Units

Proposed **subband speech and noise model**:

- Speech: 1st order time-varying autoregression (TVAR),
- Noise: 0th order TVAR, i.e. a Gaussian process with time-varying standard deviation.
- Elementary subband **PF units** are thus **very simple**
- In addition, **Rao-Blackwellization*** can be used, the amount of particles required is then small i.e. less than **a dozen particles (good robustness)**.

With known noise variances in each band (see below for details), a **particle update is simple** and consists of:

- Drawing 2 Gaussian random numbers (maintained set of means, fixed variances)
- 1-dimensional Kalman Filter update (~ 5 multiplications, 5 additions of real numbers)
- Computing the particle's weight (exponential and square root functions).

*Rao-Blackwellization: Rao-Blackwellized Particle Filters (RBPFS) are applicable when a subset of a state vector is linear-Gaussian, conditional upon other states. Optimal Kalman filtering can be used for the linear part, and PF filtering for the non-linear part.

Noise Handling

Two solutions are proposed for noise handling:

1. **Draw candidates for noise level (gain) internally** at each iteration
 - Advantageous computationally (very low additional cost, i.e. one Gaussian number to draw and another variable to resample)
 - Allows for very sudden changes in noise level (sample-by-sample treatment)
2. Alternatively, **external background noise Power Spectral Density (PSD) estimation** (e.g. minimum statistics based estimator)
 - Estimated PSD can be "discretized" to a single point in each band
 - More robust than Voice Activity Detection

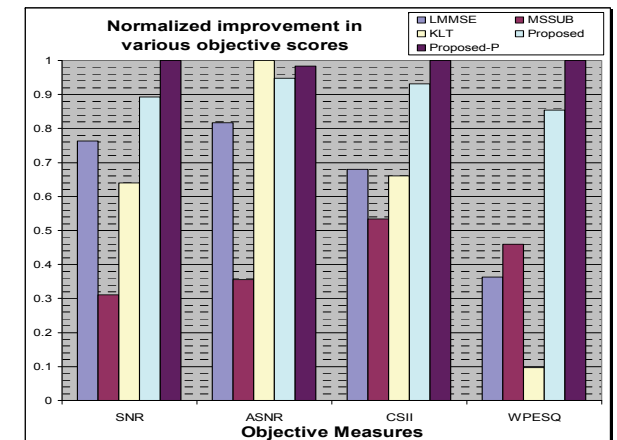
Perceptual Constraining

If external background noise PSD estimation is available:

- Use of **perceptual hearing features**, in particular simultaneous (i.e. frequency) **masking curves**
- Simple **under/overestimation of the returned noise level** can then be applied, depending on some conservative rules
- For example, an overestimation would be removing more noise, but would also introduce more speech distortion. The opposite also holds.

Performance and Simulations

- Comparisons with other schemes: averages over **12 different conditions** – low, mid, high input SNR, and **cafeteria, busy street, stadium and rain noise types**
- Several **objective measures** used for evaluation: **SNR, average (i.e. segmental) SNR, intelligibility index (CSII), wideband perceptual speech quality (w-PESQ)**
- **M-band (parallel) filterbanks** were found to provide **better results** (better frequency selectivity, less inter-band leakage effects)
- **Performance of internal noise PSD estimation** approach **close** to performance with external dedicated noise estimation, but **faster** (less complex)



LMMSE: Log-Spectral Amplitude Estimator **MSSUB:** Multi-band Spectral Subtraction
KLT: Subspace method **Proposed:** proposed method (-P is with perceptual constraints)

Audio Demonstrations

Corresponding audio demonstrations are **available on-line**, for anyone to subjectively assess the quality of enhanced speech signals:

<http://www.site.uottawa.ca/bouchard/papers/cip2010.zip>