A Simple and Effective Framework for A Priori SNR Estimation

Johannes Stahl[1], Pejman Mowlaee[2]

johannes.stahl@tugraz.at, pejman.mowlaee@tugraz.at

[1]Signal Processing and Speech Communication Laboratory, Graz University of Technology
[2]Widex A/S, Nymøllevej 6, 3540 Lynge, Denmark

Abstract

- The **a priori SNR** is key-parameter in DFT-based speech enhancement schemes
- **Decision-directed** (DD) *a priori* SNR estimation: linear combination of estimates along fixed DFT bin k.
- Can speech enhancement performance be improved by combining estimates along **harmonic trajectories** instead of fixed DFT bins?

Speech Enhancement

- DFT based speech enhancement: multiplicative gain function $G(\cdot)$
- Speech Estimate is obtained by

$$\hat{X}(k,\ell) = G(k,\ell,\xi(k,\ell),\zeta(k,\ell)) \cdot Y(k,\ell)$$

- $\zeta(k,\ell) = \frac{|Y(k,\ell)|^2}{\sigma_d^2(k,\ell)} \ldots$ a posteriori SNR
- $\xi(k,\ell) = \frac{\sigma_x^2(k,\ell)}{\sigma_d^2(k,\ell)} \ldots$ a priori SNR

The Decision-Directed A Priori SNR Estimator

DD a priori SNR estimate:

$$\hat{\xi}_{\mathsf{DD}}(k,\ell) = (1-\alpha_{\mathsf{DD}}) \max[\hat{\xi}_{\mathsf{ML}}(k,\ell),0] + \alpha_{\mathsf{DD}}\hat{\xi}_{\ell-1}(k,\ell)$$

with

$$\hat{\xi}_{\mathsf{ML}}(k,\ell) = \hat{\zeta}(k,\ell) - 1$$

$$\hat{\xi}_{\ell-1}(k,\ell) = \frac{|\hat{X}(k,\ell-1)|^2}{\hat{\sigma}_d^2(k,\ell-1)}$$

- Is there a better choice for $\hat{\xi}_{\ell-1}(k,\ell)$?

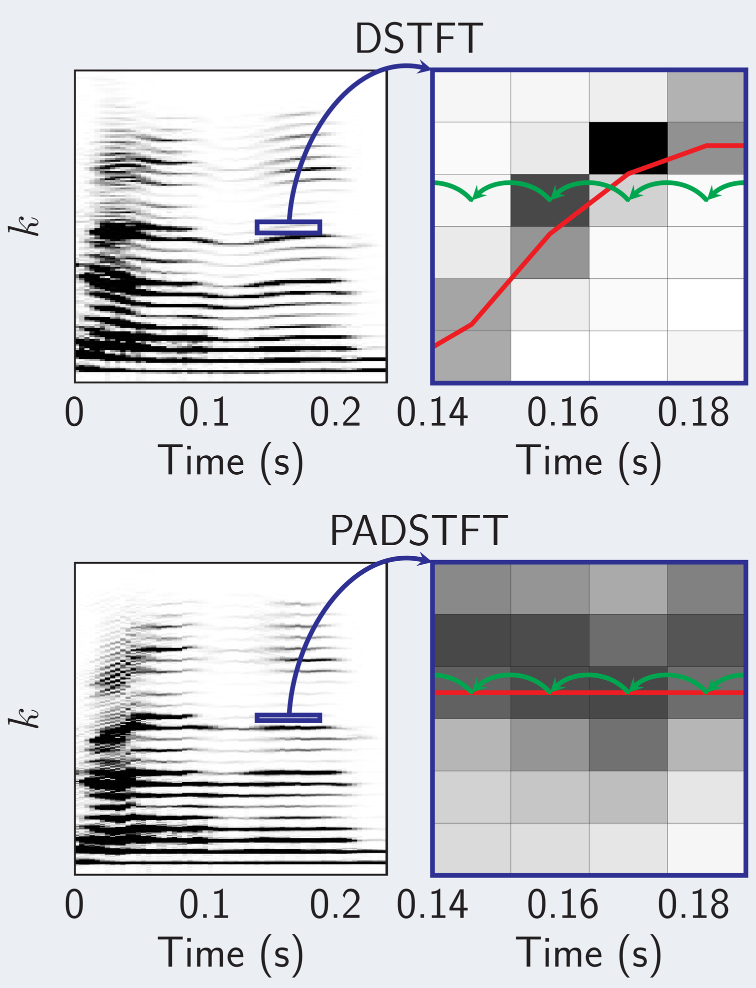
PADDi - The Proposed Method

- Speech exhibits **harmonic structure**
- Fundamental frequency is **time-varying**
- Main idea of this work: ensure that k is dominated similarly by the same harmonic at frames ℓ' and $\ell'-1$
- Pitch-adaptive discrete STFT (**PADSTFT**):

$$N_{\mathsf{DFT}}(\ell) = \text{round}\left[K\frac{f_s}{f_0(\ell)}\right]$$

$$k_h(\ell) = \underset{k}{\text{argmin}}\left|k - N_{\mathsf{DFT}}(\ell)\frac{hf_0(\ell)}{f_s}\right| = Kh$$

independent of ℓ!

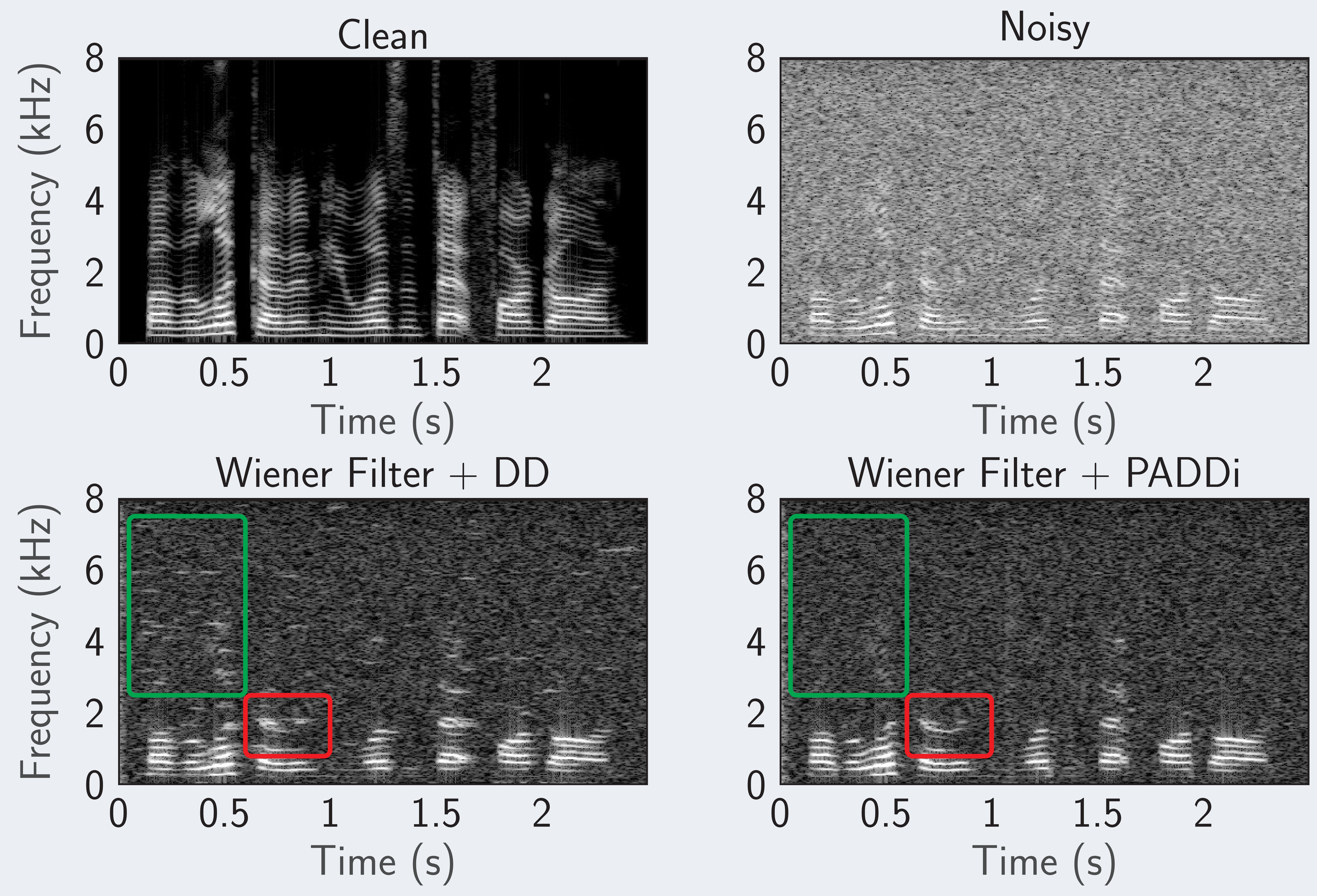


- Red: harmonic trajectory
- Green: smoothing path of *a priori* SNR estimator

- PADSTFT:
  - Fixed mapping from h to k
  - Harmonic trajectory and smoothing path coincide

Proof-of-Concept

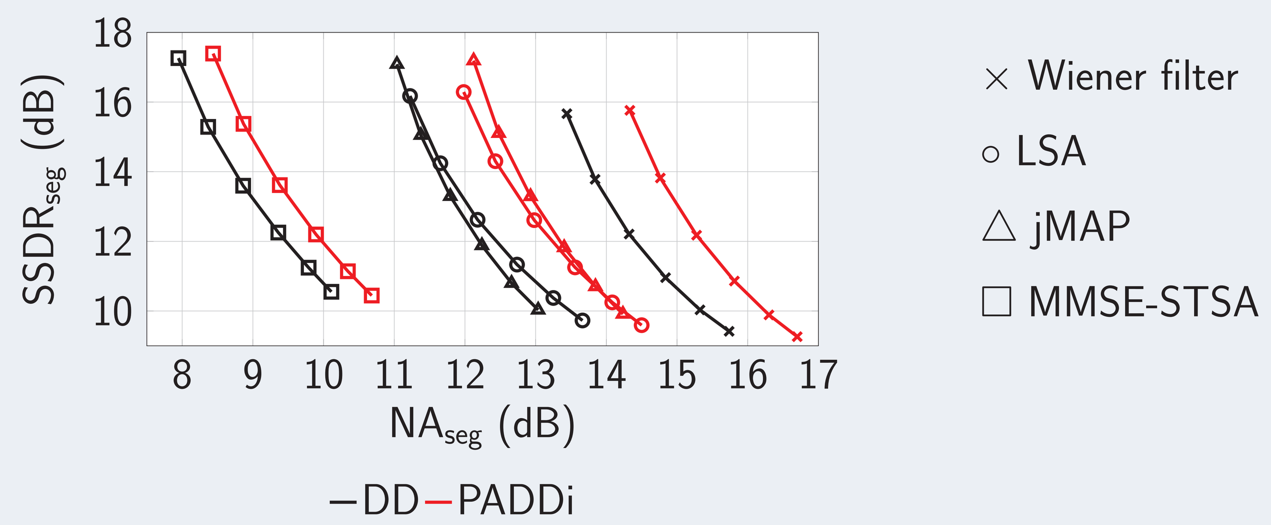
Noisy signal: Speech and white noise mixed at 0 dB SNR.



- **DD**:
  - Spurious spectral peaks → musical noise
  - Harmonics are smeared along time
- **PADDi**:
  - Less isolated spectral peaks
  - Harmonic fine structure is preserved

Results (1)

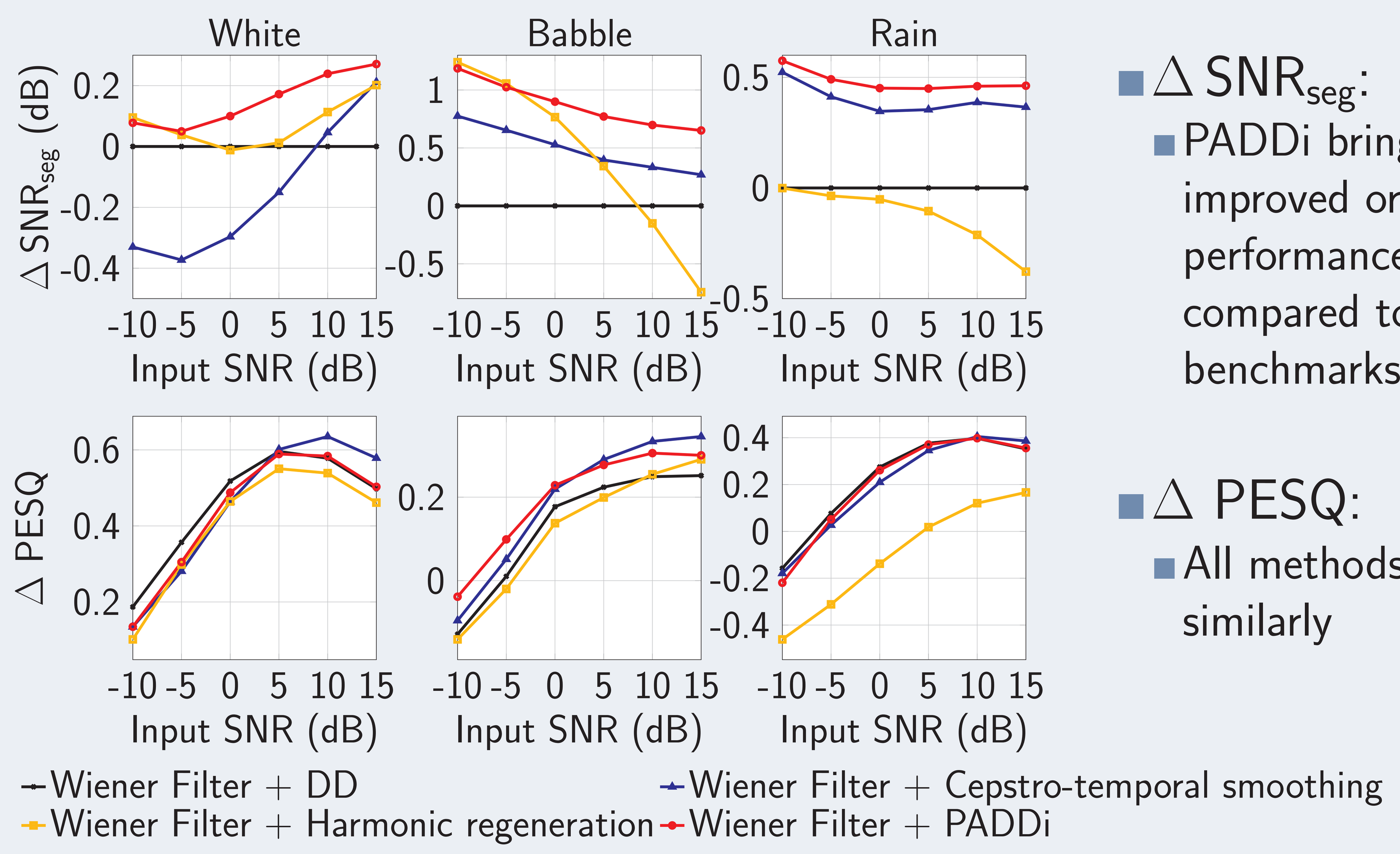
- **Characteristics** of speech estimator **strongly depend on** $G(\cdot)$
- We compared **DD** and **PADDi** for various $G(\cdot)$s
- Evaluation: **Segmental Speech to Speech Distortion Ratio** (SSDR_{seg}) vs. **Segmental Noise Attenuation** (NA_{seg})



- PADDi increases NA_{seg} while preserving SSDR_{seg} compared to DD

Results (2)

Δ-improvement in terms of **PESQ** and $\textbf{SNR}_{\textbf{seg}}$ over noisy speech



- $\Delta\,\text{SNR}_{\text{seg}}$:
  - PADDi brings improved or similar performance compared to benchmarks
- Δ PESQ:
  - All methods perform similarly

Compared to the classical DD approach, PADDi enables
- **more noise suppression** while
- **preserving the level of speech distortions**.

This work was supported by the Austrian Science Fund (FWF): P28070-N33.