Institute of Communication Systems
Prof. Dr.-Ing. Peter Jax

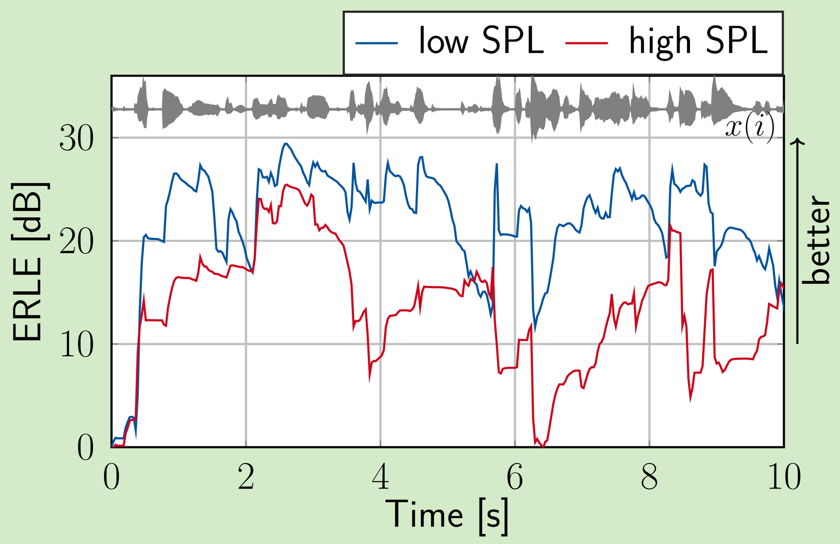
RWTH AACHEN UNIVERSITY

Efficient Nonlinear Acoustic Echo Cancellation by Dual-stage Multi-channel Kalman Filtering

Matthias Schrammen, Stefan Kühl, Shmulik Markovich-Golan, Peter Jax

1 Introduction

- Application: Single-channel Acoustic Echo Cancellation
  - ▶ Example: hands-free call with AEC in mobile phone
  - ▶ Combination of high SPL playback with small and cheap speakers leads to nonlinear distortion
- Problem: Linear AEC cannot compensate nonlinear part of echo



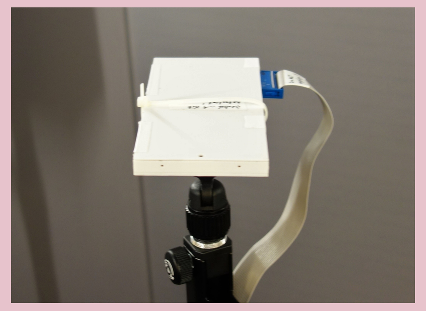
- Goal: Use Nonlinear Acoustic Echo Cancellation

SPL: Sound pressure level, AEC: Acoustic echo cancellation, ERLE: Echo return loss enhancement

2 Acquisition of Realistic Echo Signals

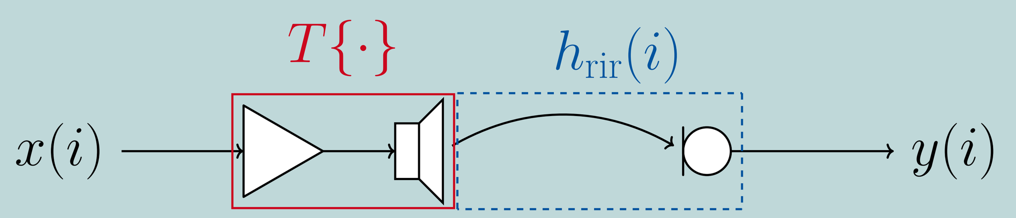
- Evaluation based on simulated nonlinear echo signal not meaningful
- Construction of realistic smartphone mockup comprising Class D amplifier, smartphone loudspeaker and digital MEMS microphone
- Simultaneous playback of far-end signal and recording of microphone signal in studio booth ($T_{60} = 0.12\,\mathrm{s}$)
- Scenarios: Mockup on *desk* (a) and mockup on *microphone stand* (b)

(a) *desk*



(b) *microphone stand*

3 Underlying Digital System Model



i: Discrete time index
$x(i)$: Far-end signal
$y(i)$: Microphone signal

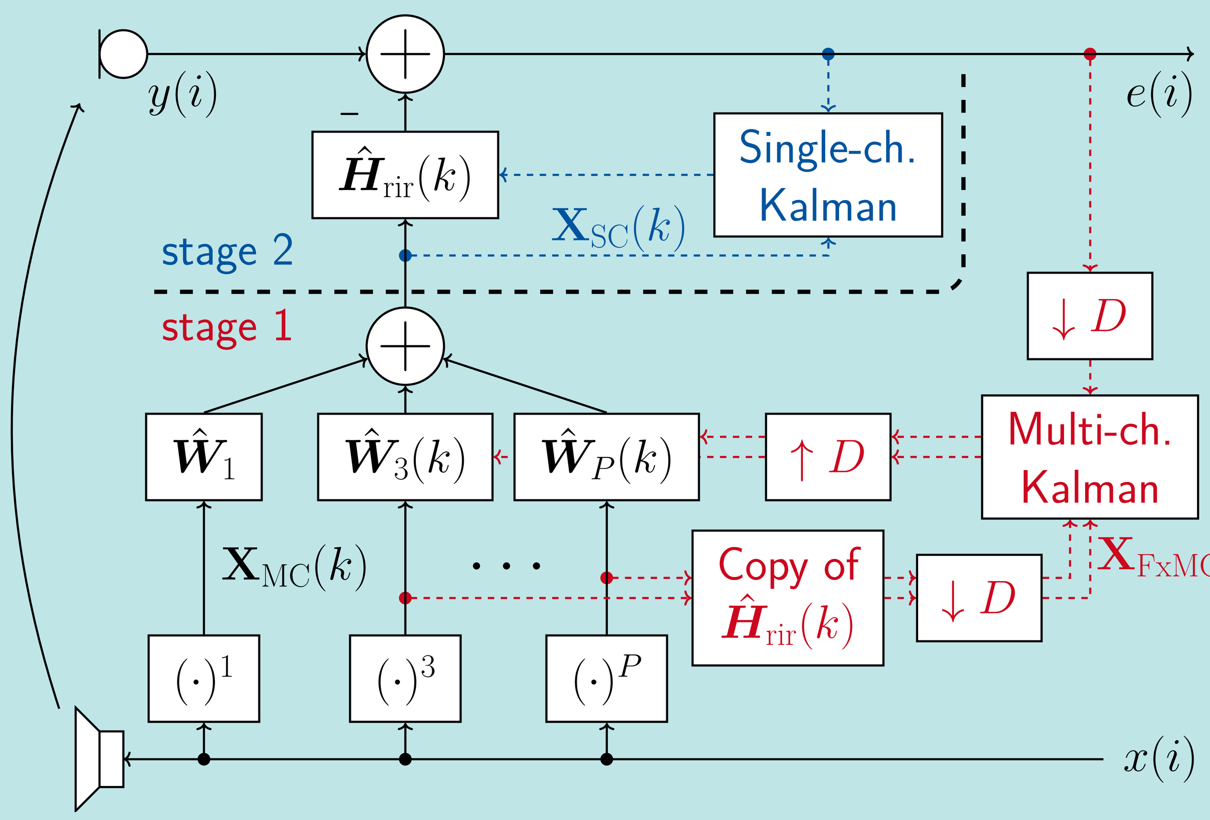
- Odd order power series $T\{\cdot\}$ of order P with weights $w_p(i)$ models nonlinear amplifier and loudspeaker with memory of length N_{nl}:

$$T\{x(i)\} = \sum_{l=0}^{\lfloor P/2 \rfloor} w_{2l+1}(i) * x^{2l+1}(i)$$

- Linear filter $h_{\mathrm{rir}}(i)$ of length $N_{\mathrm{lin}} \gg N_{\mathrm{nl}}$ models transmission from loudspeaker to microphone:

$$y(i) = h_{\mathrm{rir}}(i) * T\{x(i)\}$$

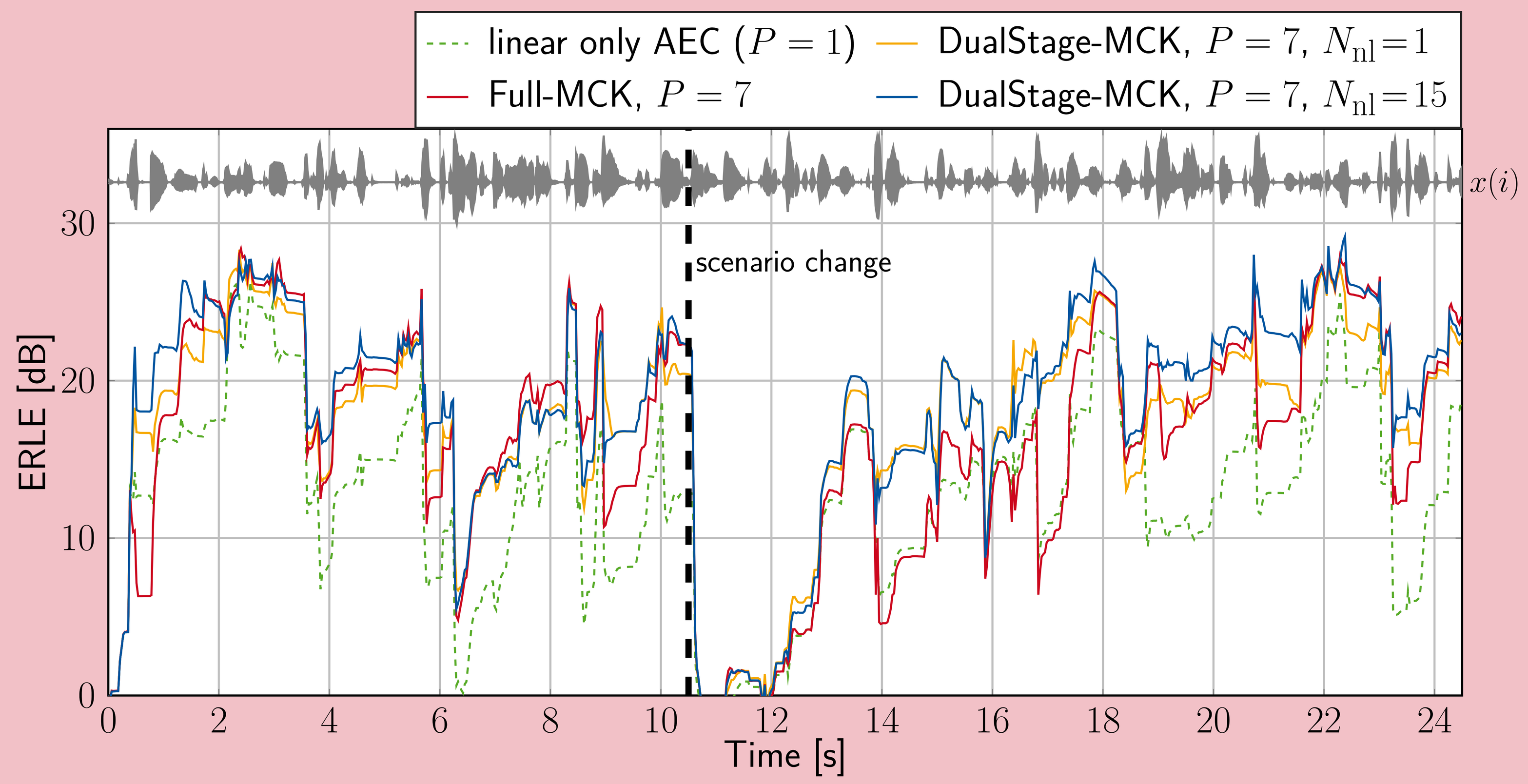
4 Dual-stage Multi-channel Kalman Filter



frame-index k
frame-size $M = 256$
frame-shift $R = 64$
sampling rate $f_s = 8\,\mathrm{kHz}$
$\hat{w}_p(k) \rightarrow \hat{\mathbf{W}}_p(k)$
$\hat{h}_{\mathrm{rir}}(k) \rightarrow \hat{\mathbf{H}}_{\mathrm{rir}}(k)$

- Cascaded structure mimics underlying system model
- Filtering and adaptation is done in short-term Fourier domain
  - Segmentation and reconstruction with overlap-save (not shown)
- Stage 1: Multi-channel Kalman filter (MCK) [1, 2]
  - Filtered-x multi-channel nonlinear reference $\mathbf{X}_{\mathrm{FxMC}}(k)$
  - Complexity reduction by reduced frequency resolution ($\downarrow D$), see 6
- Stage 2: Single-channel Kalman filter

5 Evaluation on Measured Echo Signals



- Full-MCK [1]: Reference system with one multi-channel stage
- Abrupt change of scenario simulated by switching from *desk* to *microphone stand* at $t = 10.5\,\mathrm{s}$
- Proposed DualStage-MCK ($D = 1$) with memory ($N_{\mathrm{nl}} = 15$) outperforms DualStage-MCK without memory ($N_{\mathrm{nl}} = 1$) and Full-MCK

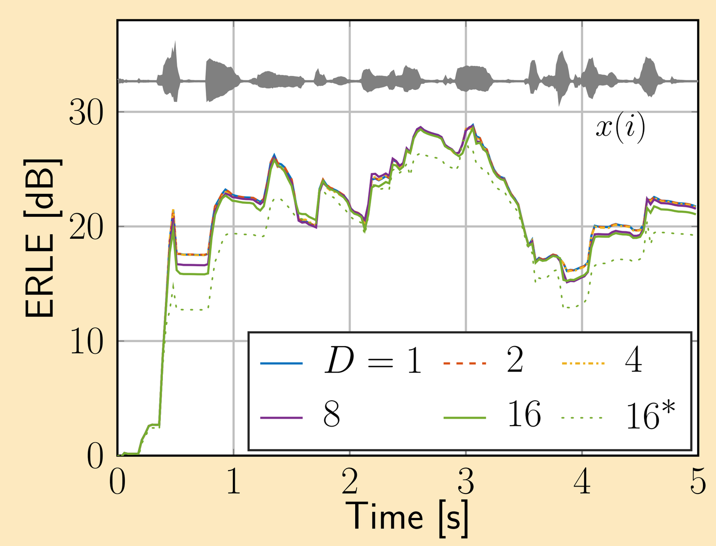
6 Complexity Reduction

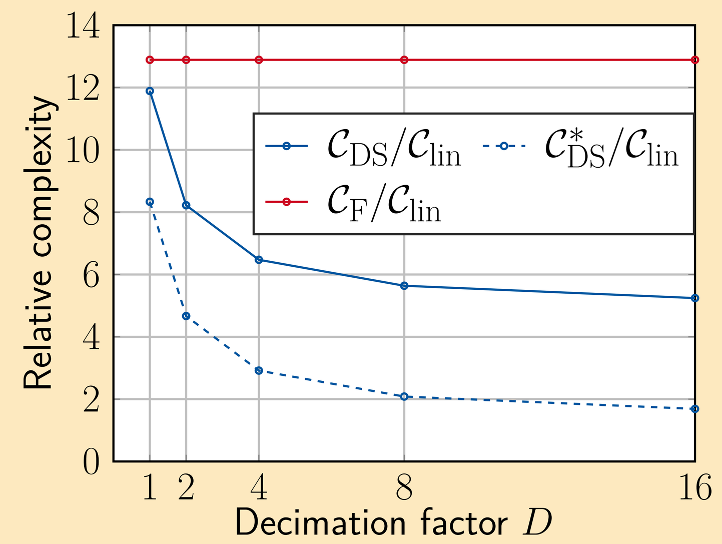
D: Decimation factor

- Nonlinear memory is typically short, see 3
  - $\hat{w}_p(k)$ are short $\rightarrow \hat{\mathbf{W}}_p(k)$ are smooth
  - DFT of size M can be reduced to $M' = M/D$

$$\tilde{X}'_{\mathrm{FxMC}}(\tilde{\mu}) = X'_{\mathrm{FxMC}}(\tilde{\mu}D) \text{ for } \tilde{\mu} = 1, 2, \ldots, M/(2D) - 1$$

- Cyclic convolution without constraining typically introduces only weak artifacts in reference signals $\mathbf{X}_{\mathrm{FxMC}}(k)$ and $\mathbf{X}_{\mathrm{SC}}(k)$
  - Constraining after application of complex weights can be omitted [3]





- ERLE for the DualStage-MCK ($N_{\mathrm{nl}} = 15$) is not impaired by decimation and only slightly decreases without constraining
- Complexity of DualStage-MCK with $D = 16$ and without constraining is only 13% of the complexity of Full-MCK.

DFT: Discrete Fourier transform, $(\cdot)^*$: Constraining omitted

$\mathcal{C}$: Complexity, $\mathcal{C}_{\mathrm{DS}}$: DualStage-MCK, $\mathcal{C}_{\mathrm{lin}}$: linear AEC, $\mathcal{C}_{\mathrm{F}}$: Full-MCK

7 Conclusion

- Novel nonlinear echo canceller with dual-stage structure
  - ▶ Speeds up convergence due to short filters with respect to Full-MCK
  - ▶ Improves ERLE by modelling a nonlinearity with memory
  - ▶ Allows for complexity reduction by reduced frequency resolution
- Significant improvement of ERLE at only 69% higher complexity than linear only AEC
  - ▶ Attractive for real-time speech communication with mobile devices

References

[1] Sarmad Malik and Gerald Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2065–2079, Sept. 2012.

[2] S. Kühl, C. Antweiler, T. Hübschen, and P. Jax, "Kalman filter based stereo system identification with auto- and cross-decorrelation," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, Mar. 2017, pp. 181–185.

[3] Jacob Benesty, Tomas Gänsler, Dennis R. Morgan, M. Mohan Sondhi, and Steven L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Digital Signal Processing. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.