Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP Signal Processing

Timo Gerkmann and Martin Krawczyk-Becker

Phase-Aware Speech Processing

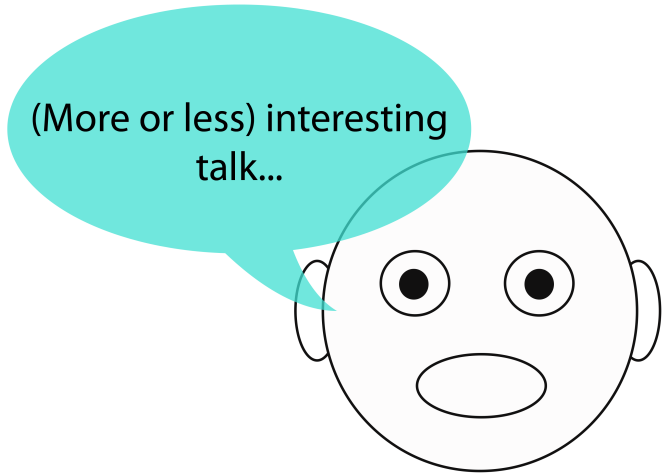
Universität Hamburg
Department of Informatics

Feb, 2019

Outline

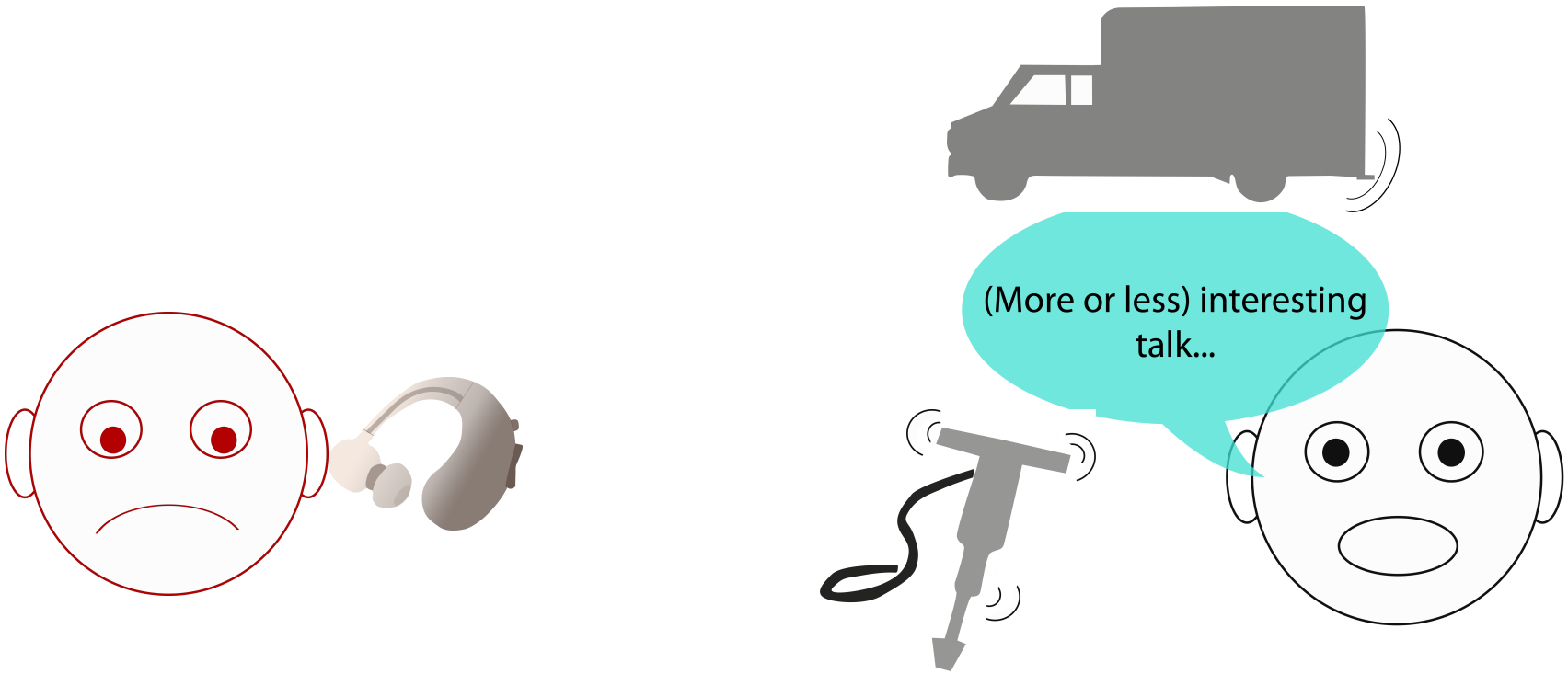
1. Introduction
   - The Role of Phase in Speech Enhancement
   - Different Phase Representation
2. Phase Estimation
   - Iterative Phase Reconstruction
   - Harmonic Model
3. Phase-Aware Speech Estimation
   - Phase-Aware Amplitude Estimation
   - Phase-Aware Complex Estimation
4. Evaluation
5. Outlook and Conclusions

Motivation



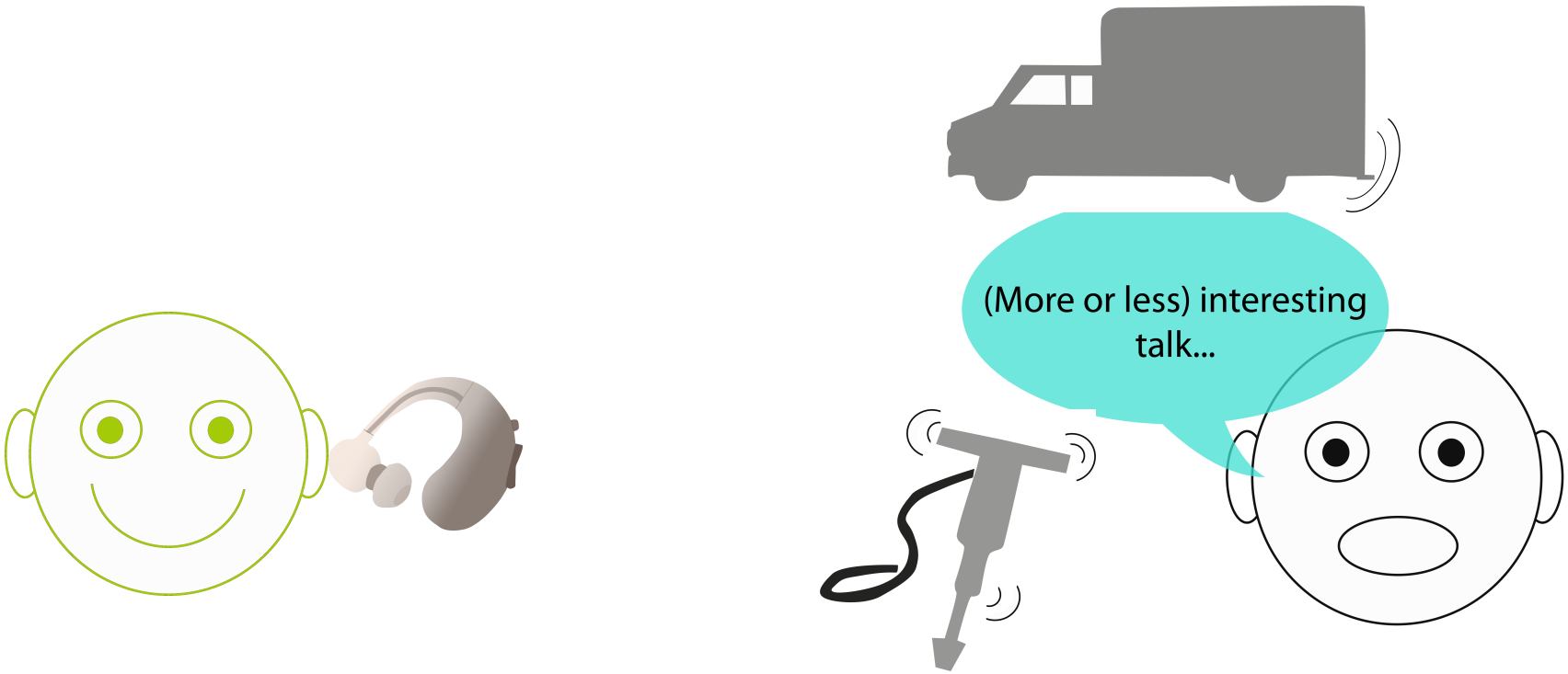
- Hearing aids work well without noise

Motivation



- Hearing aids work well without noise
- In presence of severe noise the performance drops

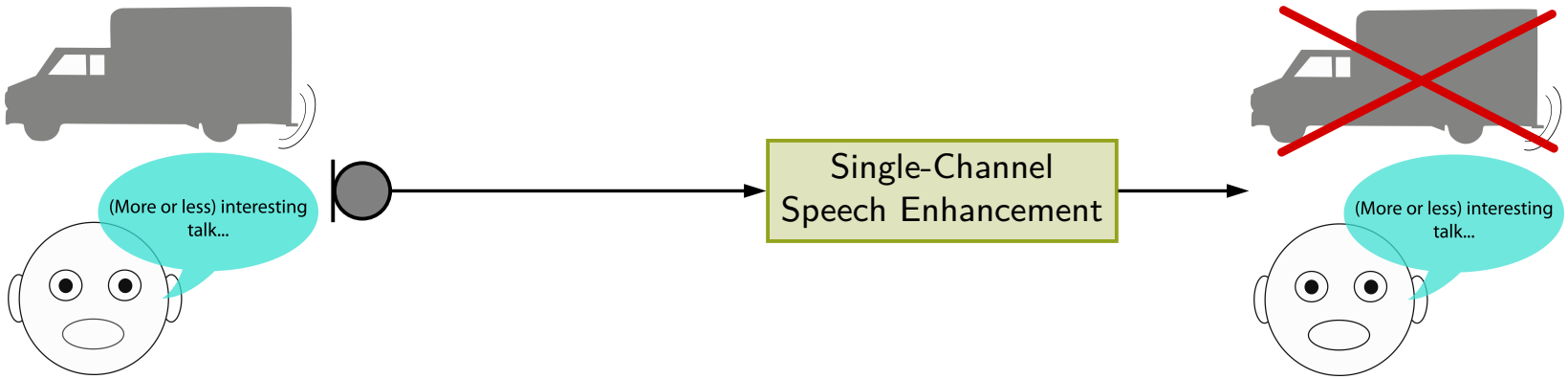
Motivation



- Hearing aids work well without noise
- In presence of severe noise the performance drops
- Apply speech enhancement to reduce the detrimental effect of the noise!

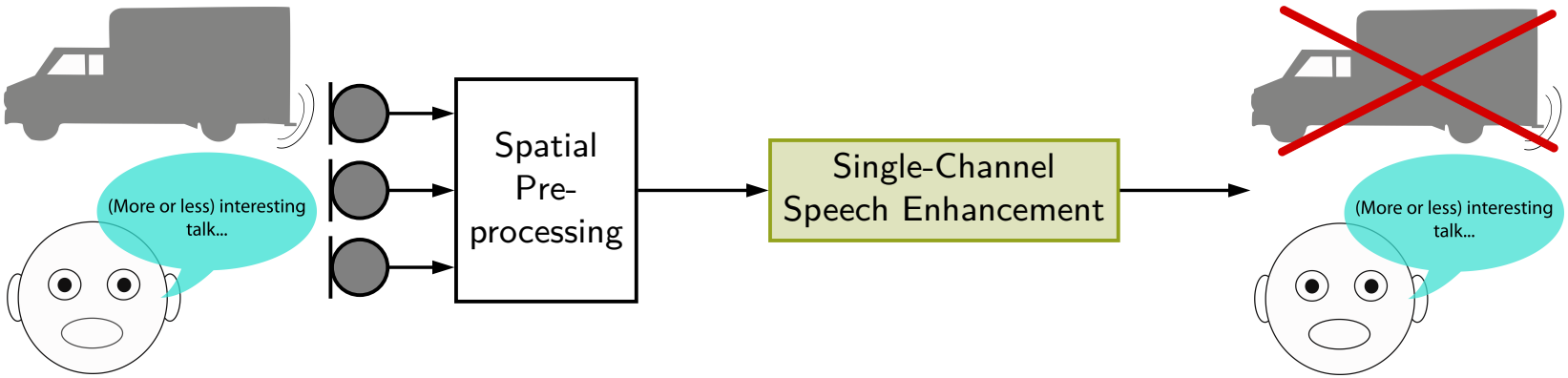
Basic Concept

- The noise is reduced using only a single channel, e.g.
  - a single microphone signal (e.g. in-the-canal hearing aids)

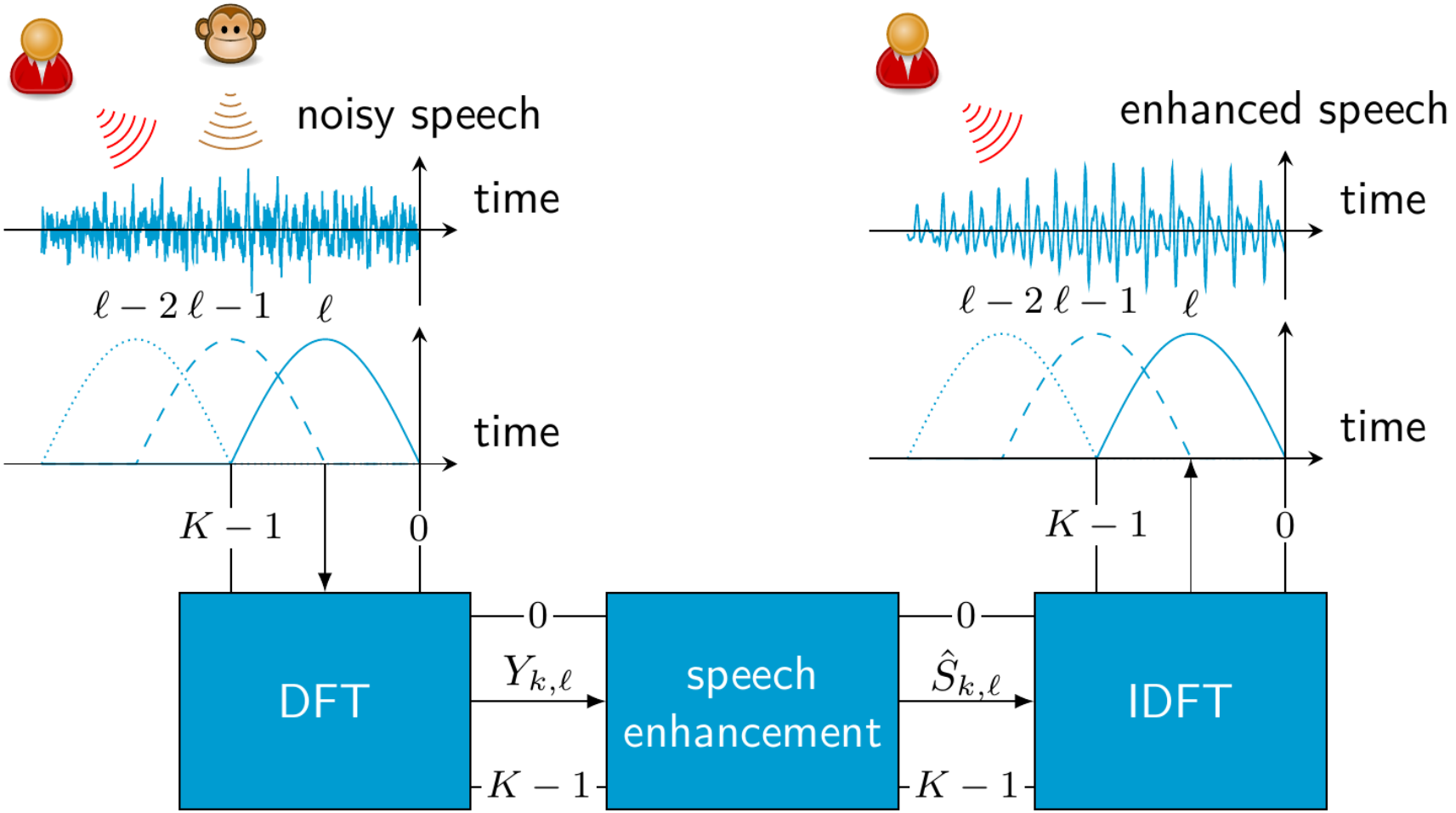


Basic Concept

- The noise is reduced using only a single channel, e.g.
  - a single microphone signal (e.g. in-the-canal hearing aids)
  - the output of a spatial preprocessing stage

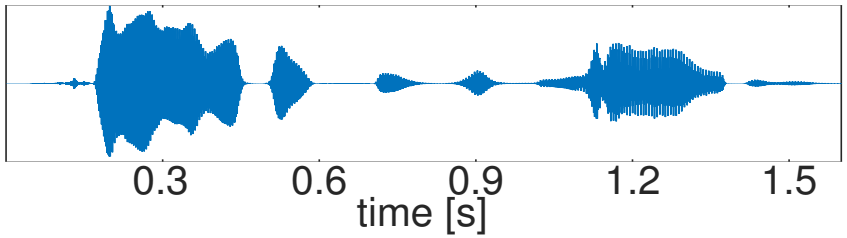


Short-time Fourier Transform (STFT) setup



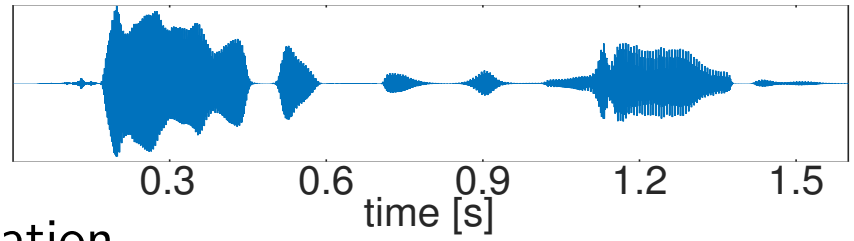
Spectral Representation

- **Real-valued** time domain signal



Spectral Representation

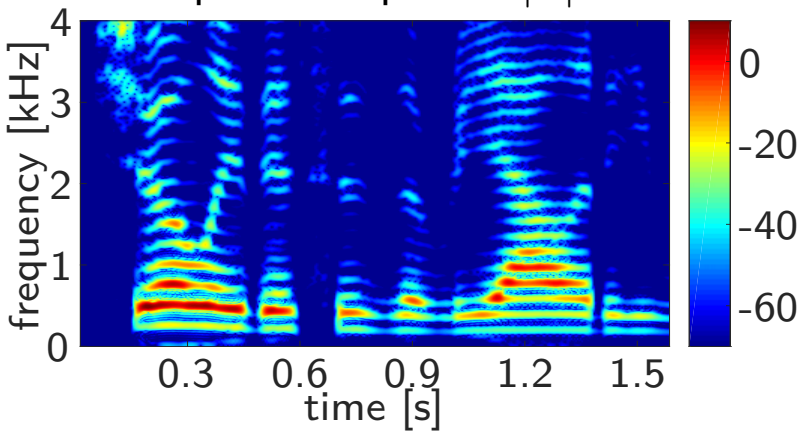
- **Real-valued** time domain signal



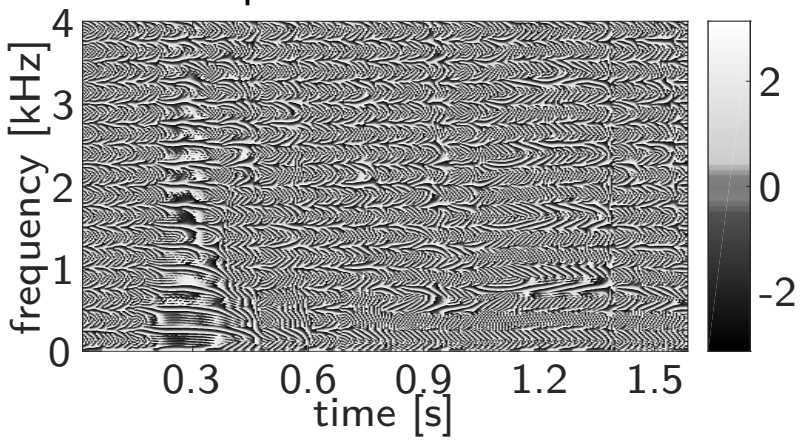
- **Complex-valued** spectral representation

$$\underbrace{Y}_{\text{noisy}} = |Y| e^{j\Phi_Y} = \underbrace{S}_{\text{speech}} + \underbrace{V}_{\text{noise}}$$

Speech Amplitude $|S|$

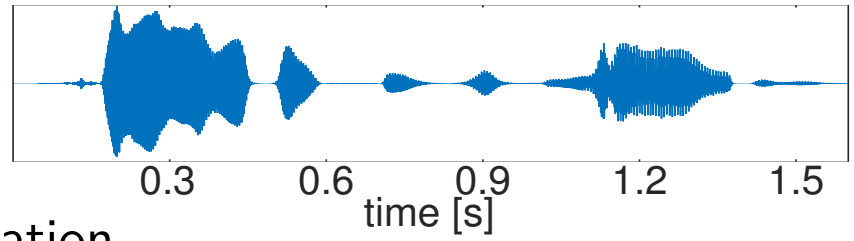


Speech Phase Φ^S



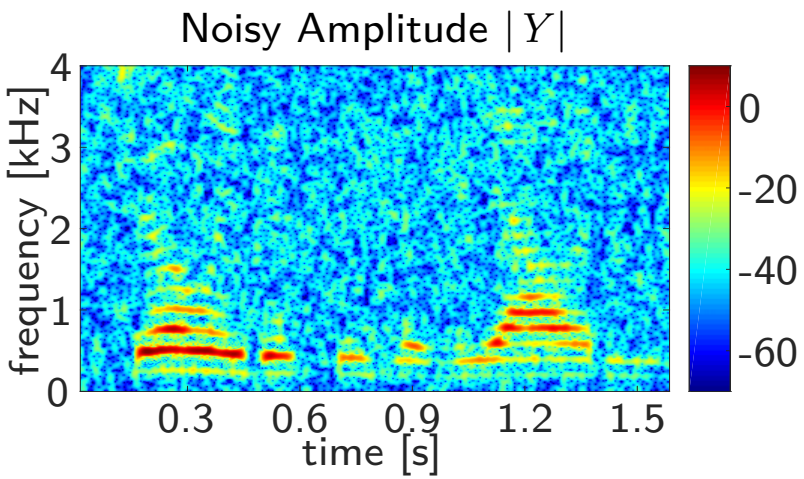
Spectral Representation

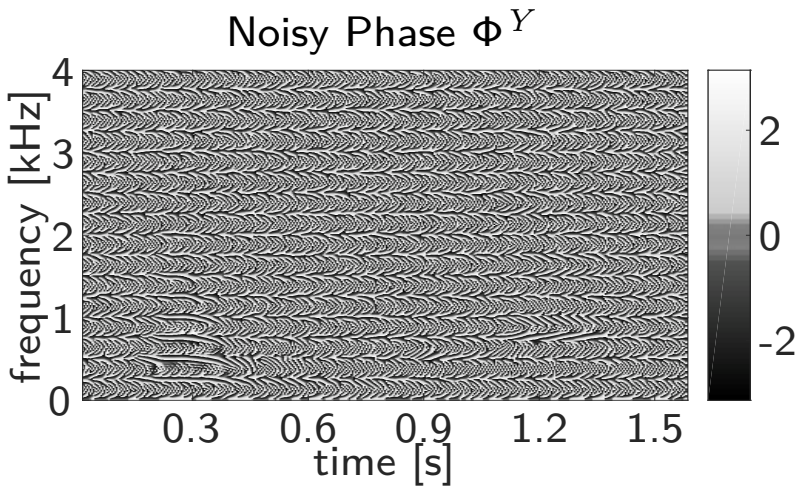
- **Real-valued** time domain signal



- **Complex-valued** spectral representation

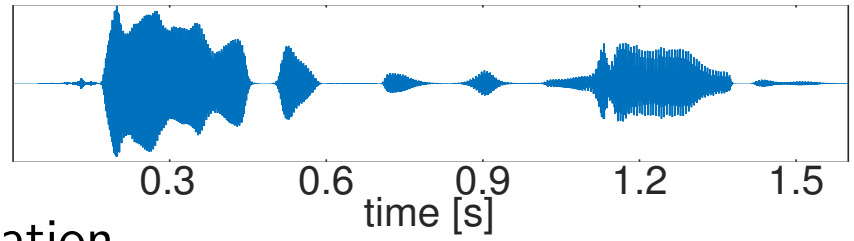
$$\underbrace{Y}_{\text{noisy}} = |Y| e^{\mathrm{j}\Phi_Y} = \underbrace{S}_{\text{speech}} + \underbrace{V}_{\text{noise}}$$





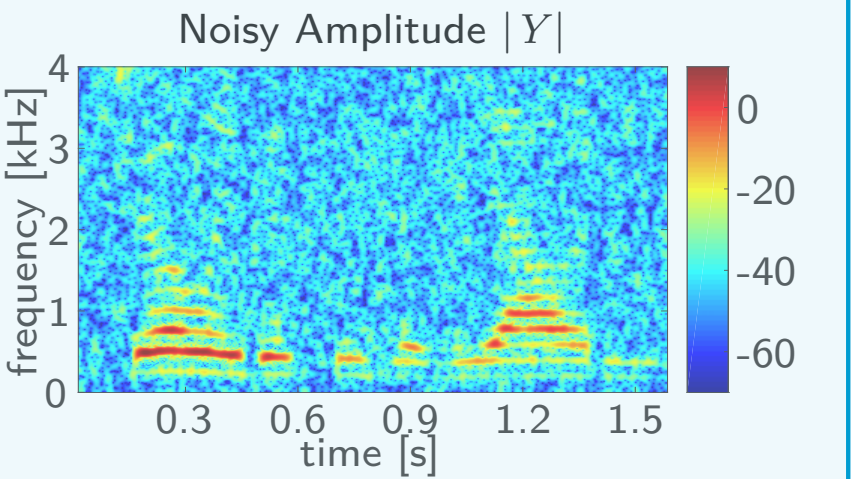
Spectral Representation

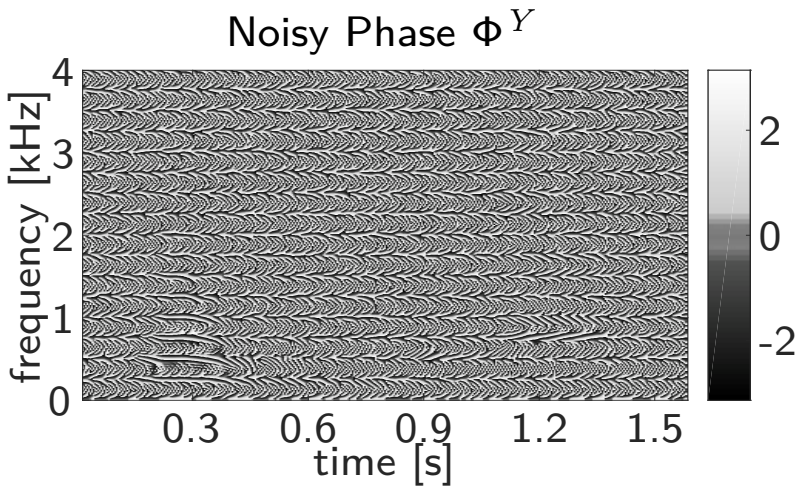
- **Real-valued** time domain signal



- **Complex-valued** spectral representation

$$\underbrace{Y}_{\text{noisy}} = |Y|e^{j\Phi_Y} = \underbrace{S}_{\text{speech}} + \underbrace{V}_{\text{noise}}$$

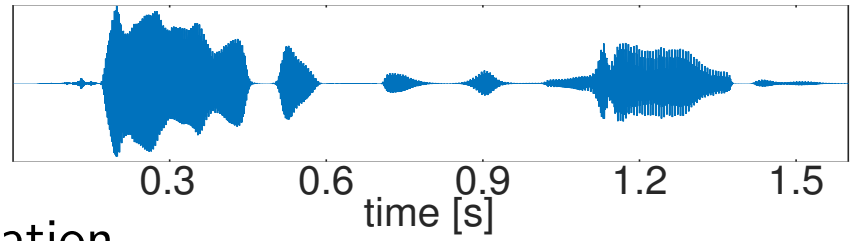




Conventional approaches only enhance the amplitude

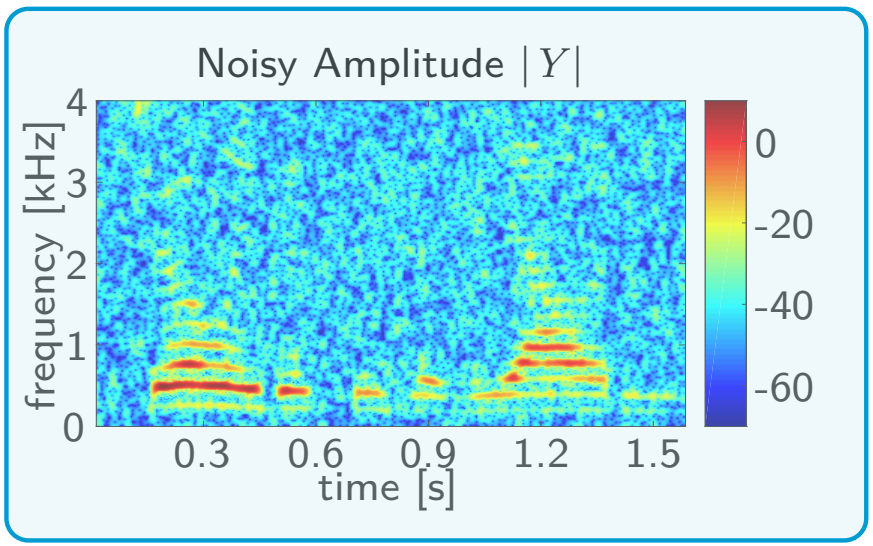
Spectral Representation

- **Real-valued** time domain signal

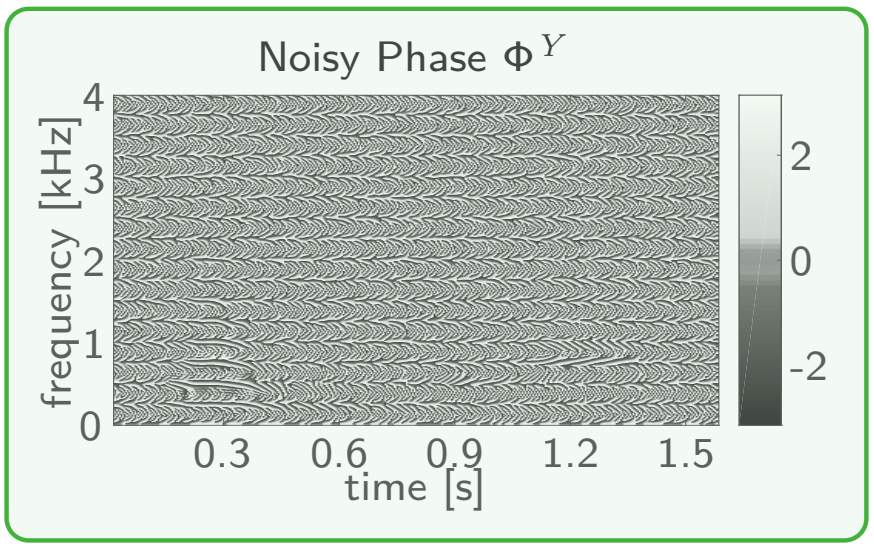


- **Complex-valued** spectral representation

$$\underbrace{Y}_{\text{noisy}} = |Y|e^{j\Phi_Y} = \underbrace{S}_{\text{speech}} + \underbrace{V}_{\text{noise}}$$



Conventional approaches only enhance the amplitude



Phase is neither enhanced nor utilized as a source of information

Why Has Phase Been Neglected?

1982

The Unimportance of Phase in Speech Enhancement

DAVID L. WANG AND JAE S. LIM

Abstract—The importance of Fourier transform phase in speech enhancement is considered. Results indicate that a more accurate estimation of phase is unwarranted in speech enhancement at the S/N ratios where the intelligibility scores of unprocessed speech range from 5 to 95 percent, if the phase estimate is used to reconstruct speech by combining it with an independently estimated magnitude or to reconstruct speech using the phase-only signal reconstruction algorithm.

Why Has Phase Been Neglected?

1982

The Unimportance of Phase in Speech Enhancement

DAVID L. WANG AND JAE S. LIM

Abstract—The importance of Fourier transform phase in speech enhancement is considered. Results indicate that a more accurate estimation of phase is unwarranted in speech enhancement at the S/N ratios where the intelligibility scores of unprocessed speech range from 5 to 95 percent, if the phase estimate is used to reconstruct speech by combining it with an independently estimated magnitude or to reconstruct speech using the phase-only signal reconstruction algorithm.

2010

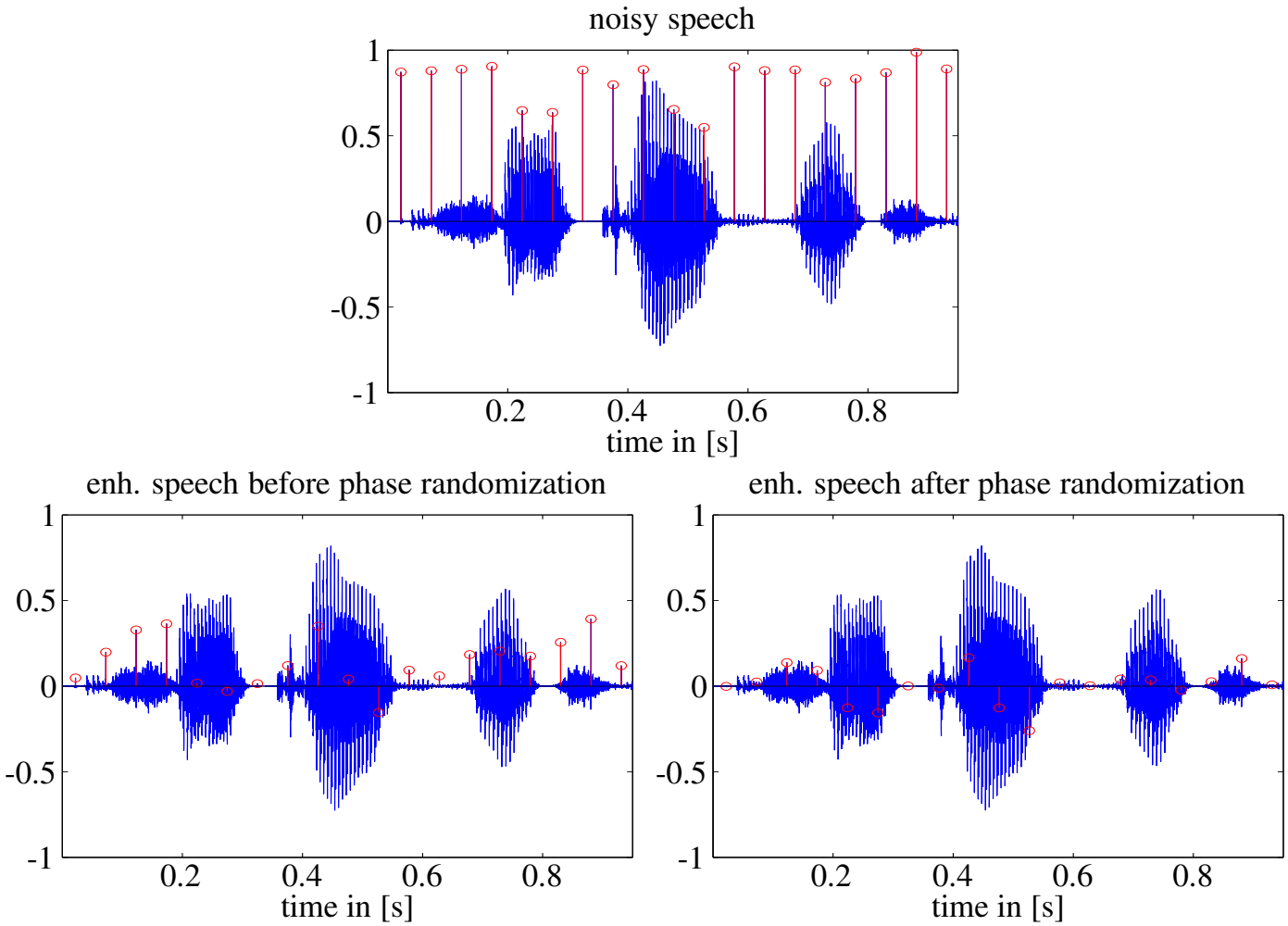
The importance of phase in speech enhancement

Kuldip Paliwal, Kamil Wójcicki *, Benjamin Shannon [1]

Signal Processing Laboratory, Griffith School of Engineering, Griffith University, Nathan, QLD 4111, Australia

Received 24 March 2010; received in revised form 30 November 2010; accepted 6 December 2010
Available online 24 December 2010

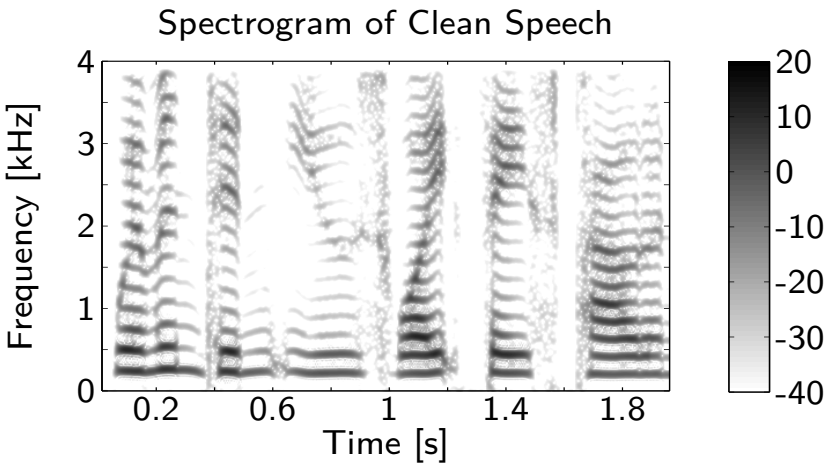
Phase Randomization for Click-Noise Removal[1]

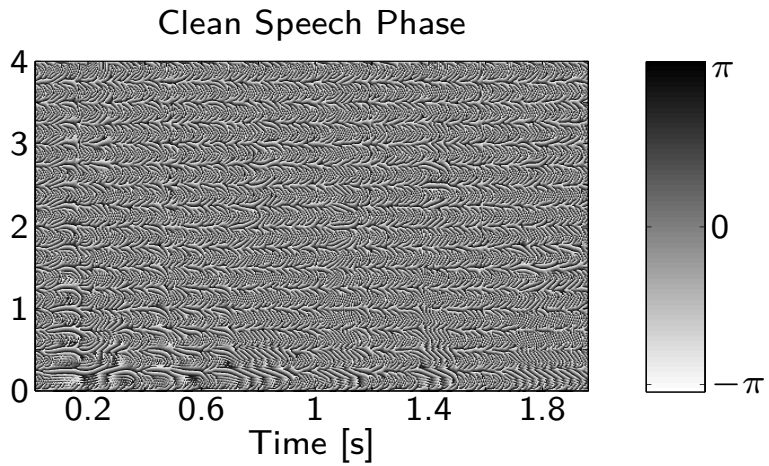


➡ Magnitude modification is not sufficient for click removal!

[1] A. Sugiyama and R. Miyahara. "Phase randomization – A new paradigm for single-channel signal enhancement". In: *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*. Vancouver, Canada, 2013, pp. 7487–7491.

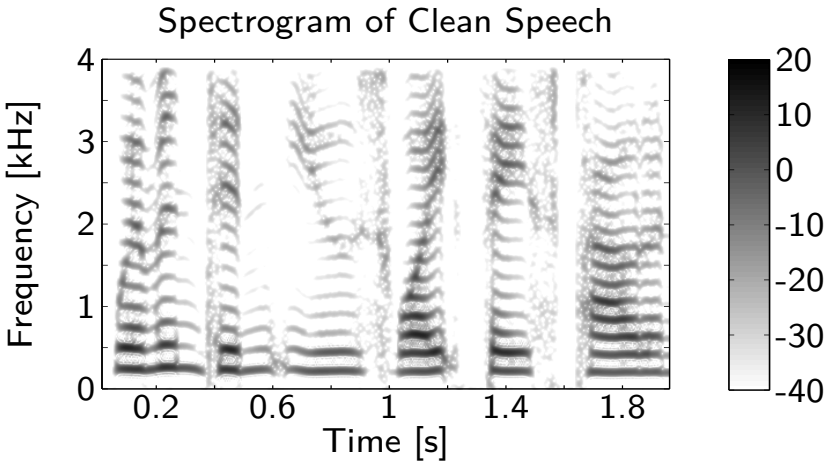
Is Phase Estimation Possible?

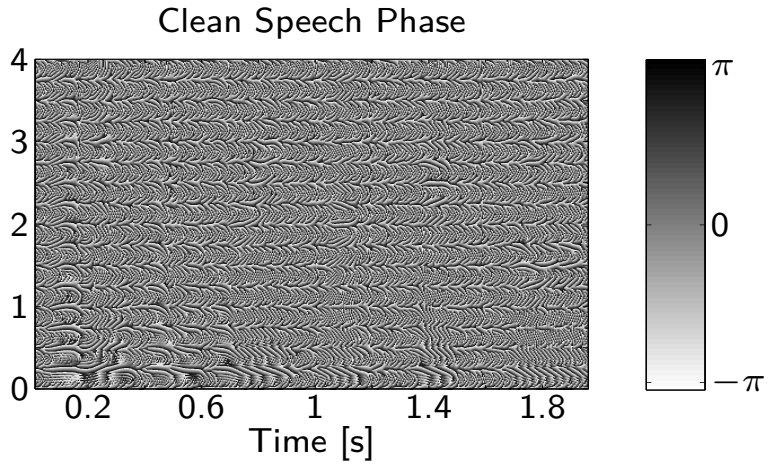




[2] Y. Ephraim and D. Malah. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.6 (1984), pp. 1109–1121.

Is Phase Estimation Possible?

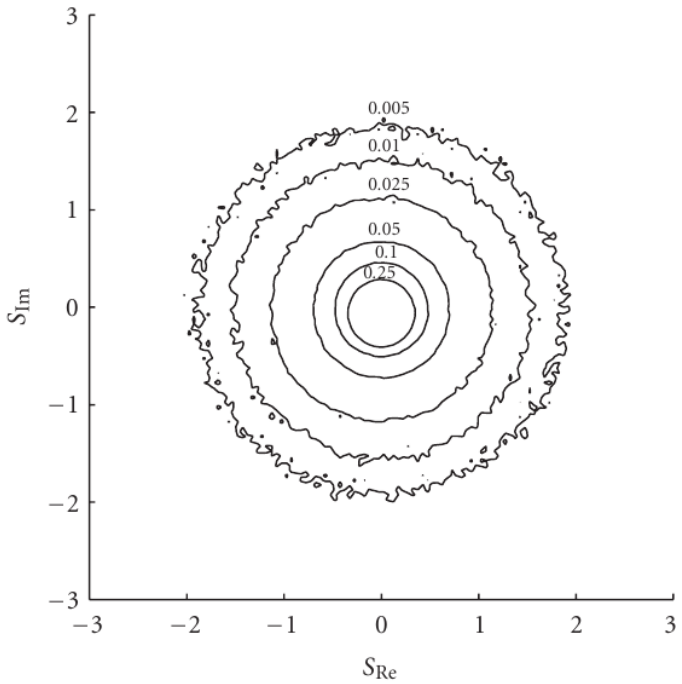


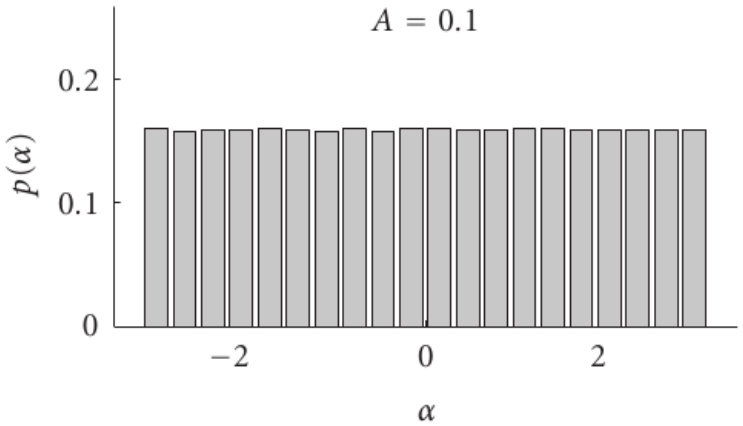


- Is the phase Φ_S just random and independent of A?
- If so, then the noisy phase would be the MMSE-optimal estimate[2].

[2] Y. Ephraim and D. Malah. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.6 (1984), pp. 1109–1121.

Histogram of Complex Speech Data[3]



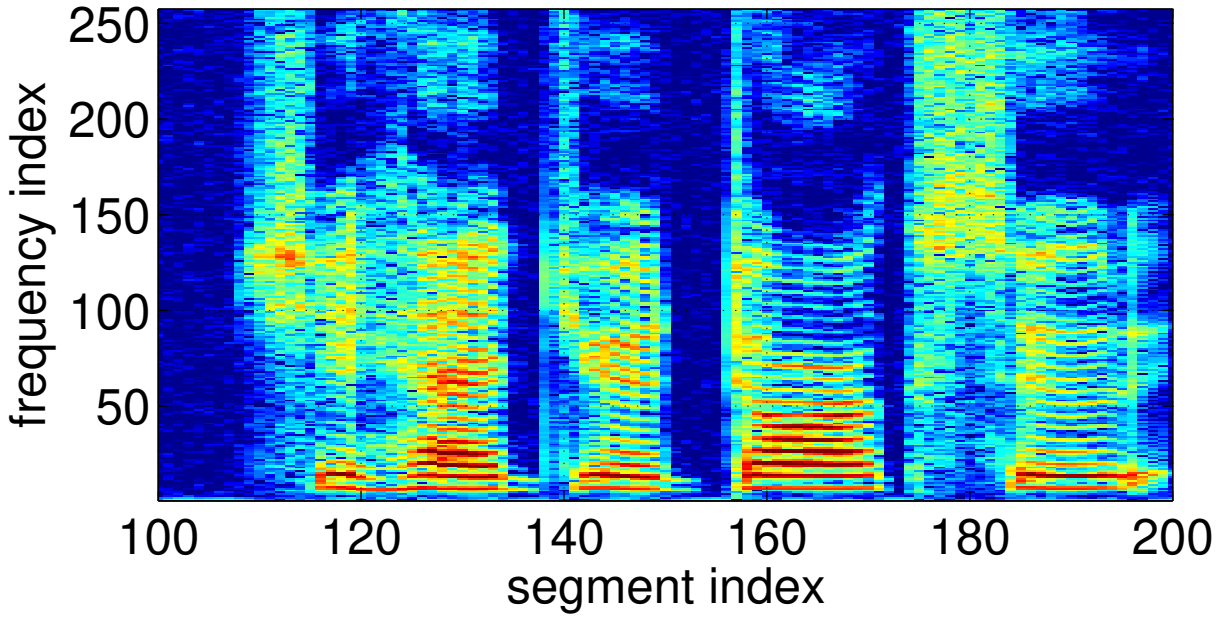


- Circular symmetric joint PDF
  - ➡ Phase is uniformly distributed and independent of the amplitude
  - ➡ The MMSE optimal estimate of the clean speech phase is the noisy phase[2]

[2] Y. Ephraim and D. Malah. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.6 (1984), pp. 1109–1121.

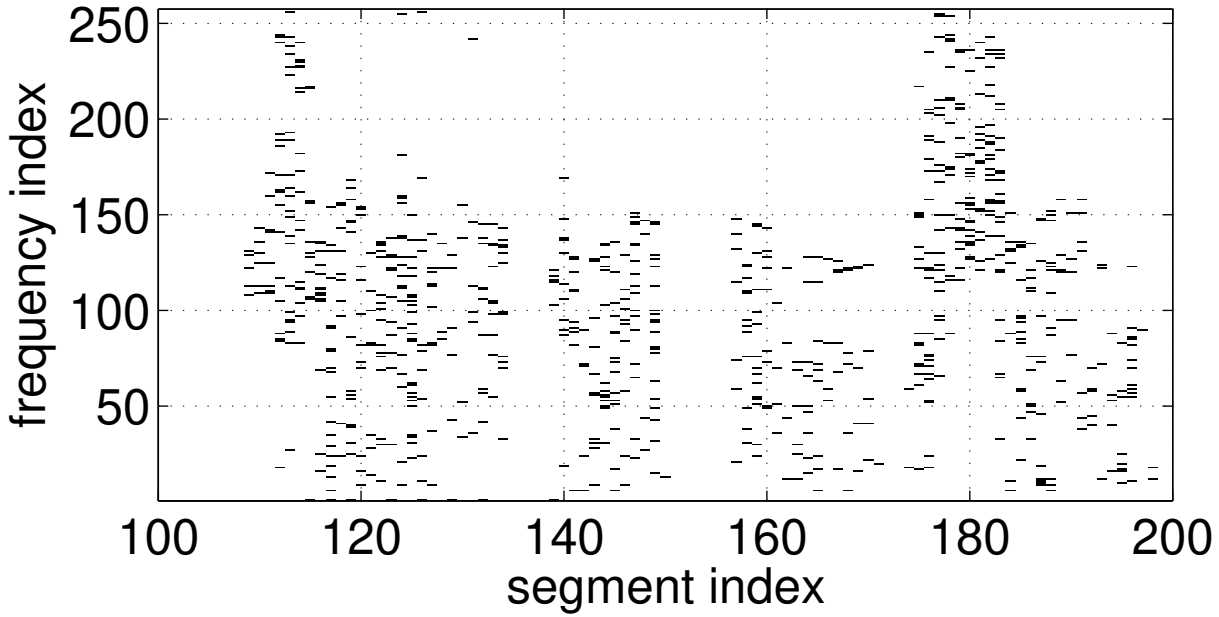
[3] T. Lotter and P. Vary. "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model". In: *EURASIP J. Applied Signal Process.* 2005.7 (2005), pp. 1110–1126.

How are the Histograms Computed?



- Histogram were computed from coefficients with similar SNR (19-21 dB).

How are the Histograms Computed?



- Histogram were computed from coefficients with similar SNR (19-21 dB).

➡ Temporal and spectral correlations are not reflected.

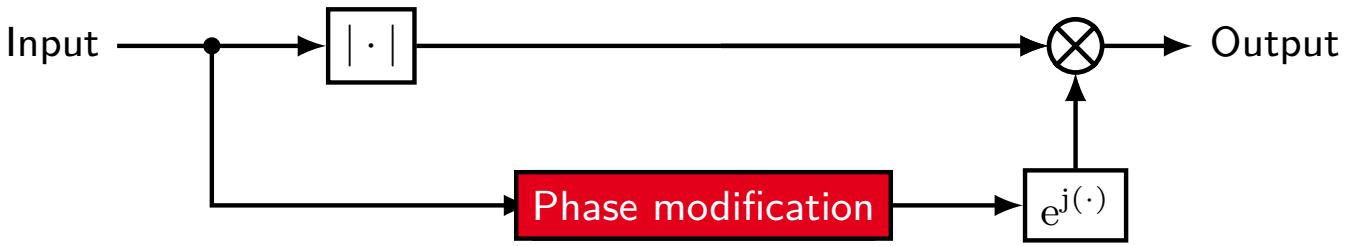
➡ In practice, we have past data available!

How Does a Random Phase Sound?

- Random phase
  -

[4] P. Vary. "Noise Suppression By Spectral Magnitude Estimation – Mechanism and Theoretical Limits". In: *ELSEVIER Signal Process.* 8 (1985), 387–400.

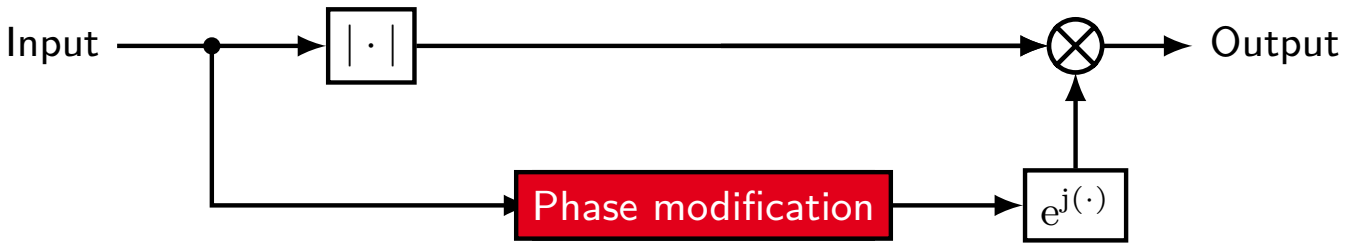
How Does a Random Phase Sound?



- Random phase
  -
  - sounds rough
- Clean phase + random phase
  - $\pi/2$:
  - $\pi/8$:
  - clean:

[4] P. Vary. "Noise Suppression By Spectral Magnitude Estimation – Mechanism and Theoretical Limits". In: *ELSEVIER Signal Process.* 8 (1985), 387–400.

How Does a Random Phase Sound?

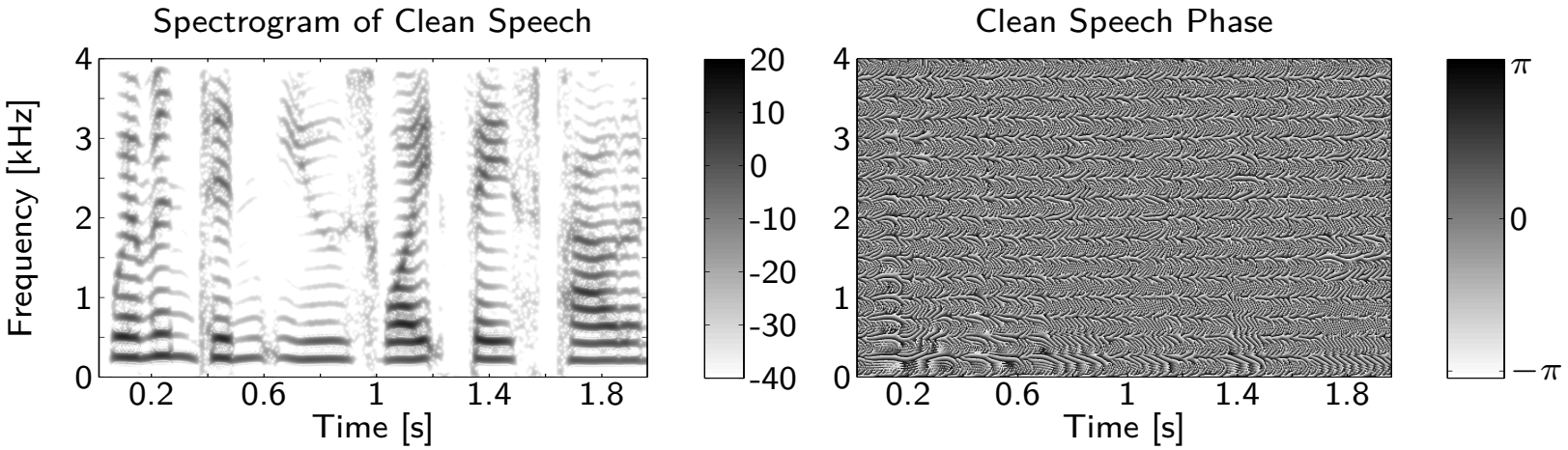


- Random phase
  -
  - sounds rough
- Clean phase + random phase
  - $\pi/2$:
  - $\pi/8$:
  - clean:
  - differences only perceivable when phase error $\approx \pi/4$

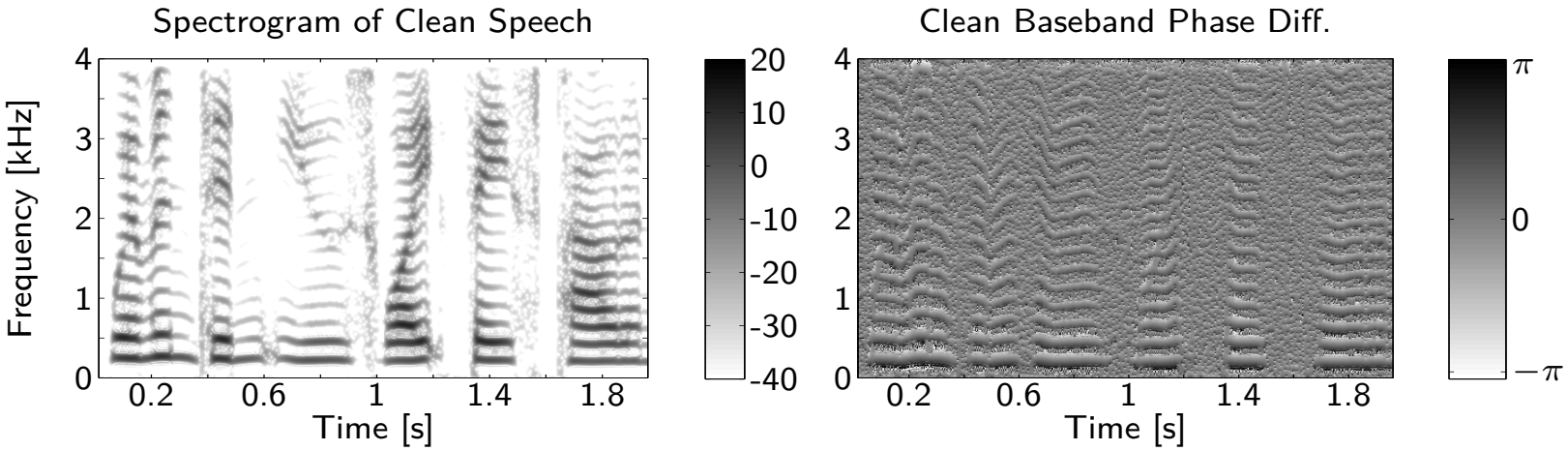
➡ Phase errors perceivable when local SNR $< 6\,\mathrm{dB}$[4]

[4] P. Vary. "Noise Suppression By Spectral Magnitude Estimation – Mechanism and Theoretical Limits". In: *ELSEVIER Signal Process.* 8 (1985), 387–400.

Improved Visualization of the Phase Structure



Improved Visualization of the Phase Structure

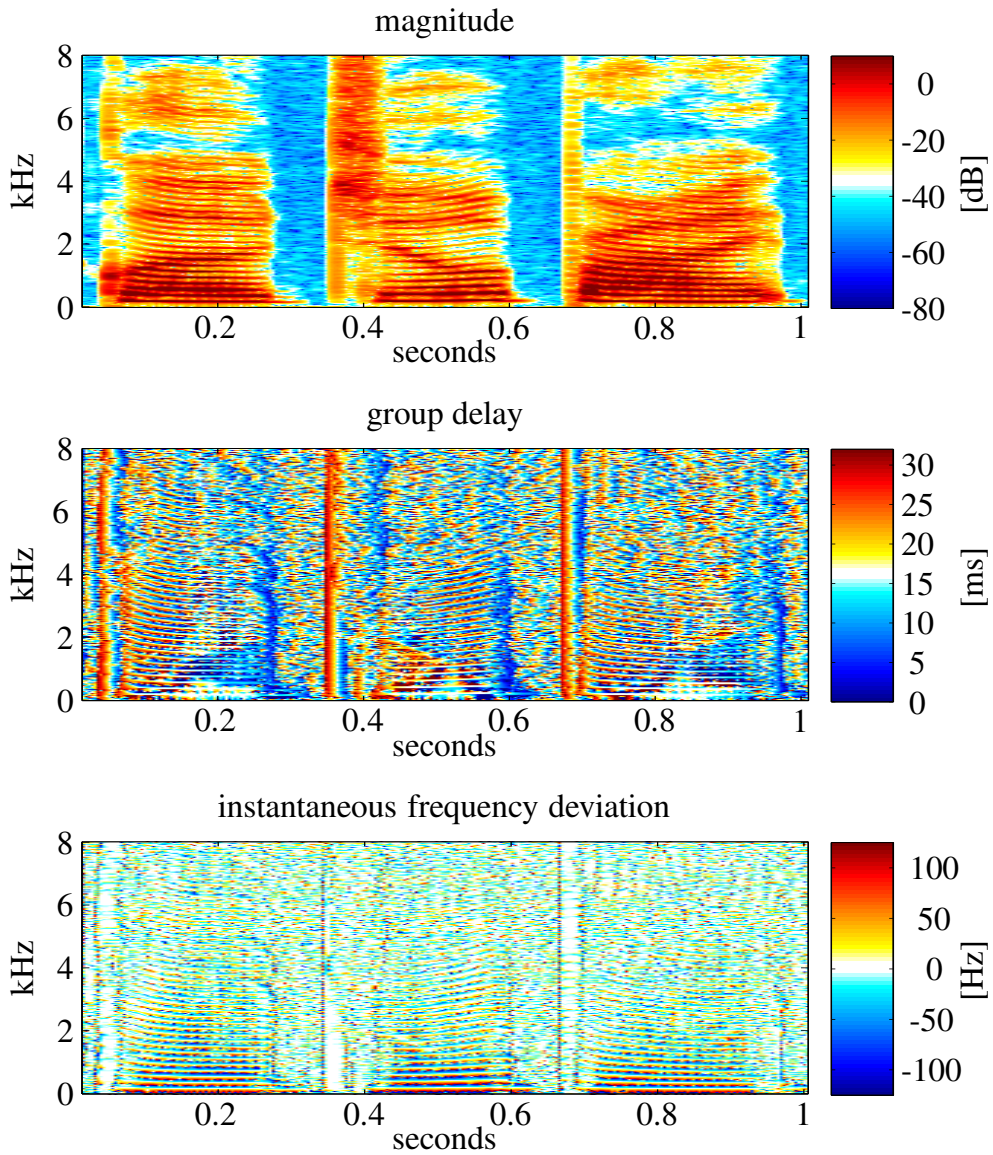


- Transform to baseband by multiplying $\mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}k\ell L}$
- Plot phase difference between neighboring frames

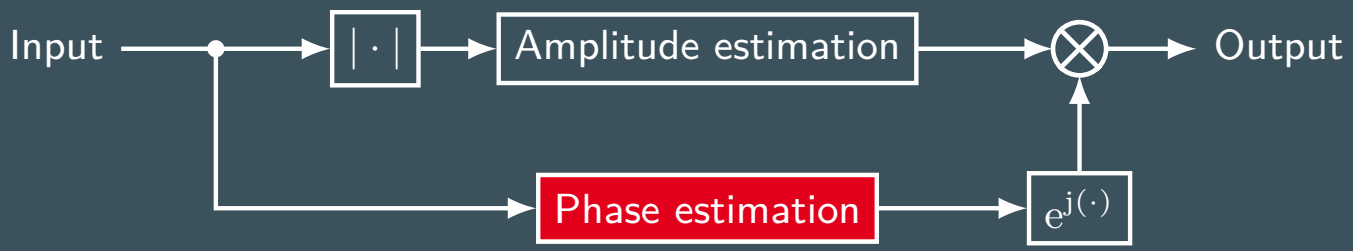
➡ Less phase wrapping!

➡ In voiced segments, the phase-change and amplitude trajectories are clearly related!

Other Phase Representations[5]



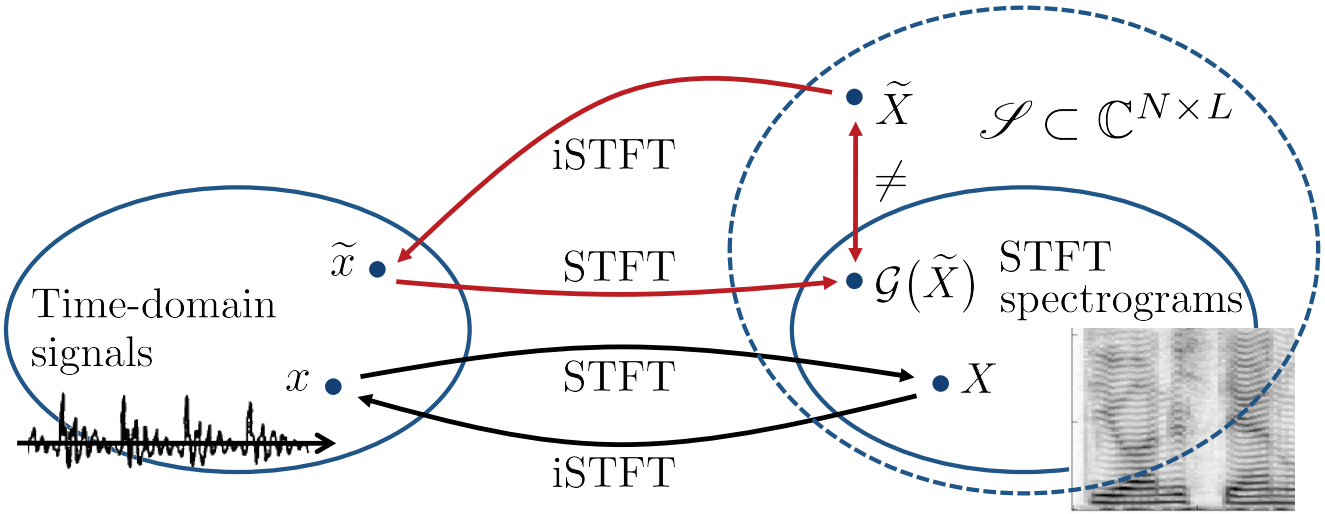
[5] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux. "Phase Processing for Single Channel Speech Enhancement: History and Recent Advances". In: *IEEE Signal Process. Mag.* 32.2 (2015), pp. 55–66.



Phase Estimation

Consistent STFT

- When magnitude and phase are modified independently, their STFT is not *consistent*[5]



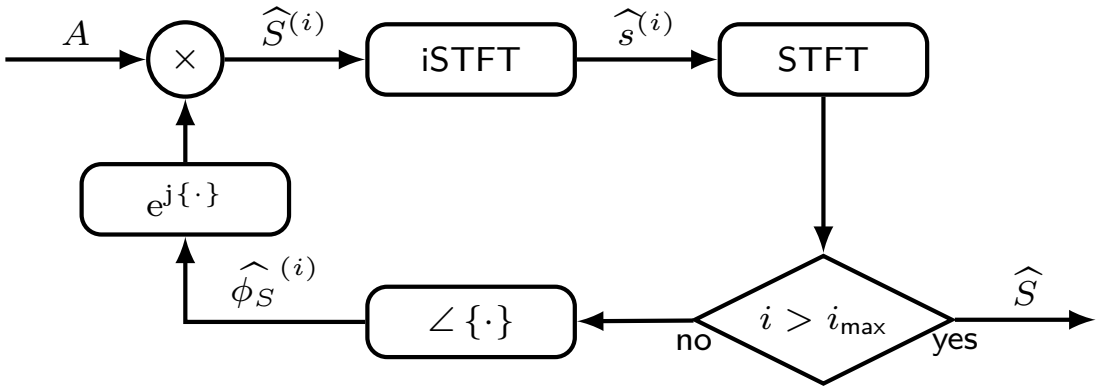
[5] T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux. "Phase Processing for Single Channel Speech Enhancement: History and Recent Advances". In: *IEEE Signal Process. Mag.* 32.2 (2015), pp. 55–66.

Iterative Phase Reconstruction[6]

- Only the spectral amplitudes of a speech signal are known

➡ **Goal:** Find the spectral phase that produces a consistent STFT

➡ Iteratively compute the iSTFT and STFT, refining the phase estimate in each iteration



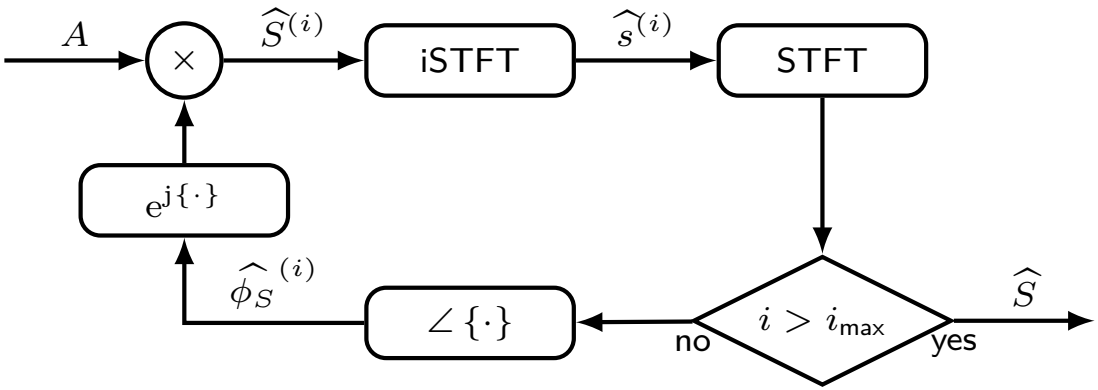
[6] D. W. Griffin and J. S. Lim. "Signal estimation from modified short-time Fourier transform". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.2 (1984), pp. 236–243.

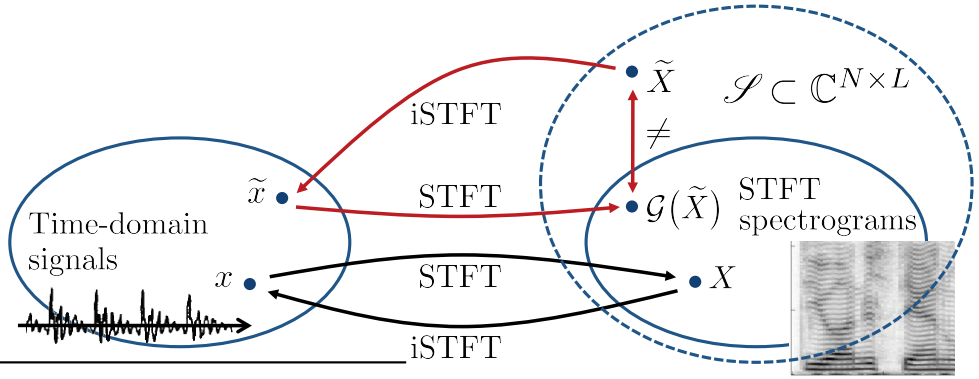
Iterative Phase Reconstruction[6]

- Only the spectral amplitudes of a speech signal are known

➡ **Goal:** Find the spectral phase that produces a consistent STFT

  ➡ Iteratively compute the iSTFT and STFT, refining the phase estimate in each iteration

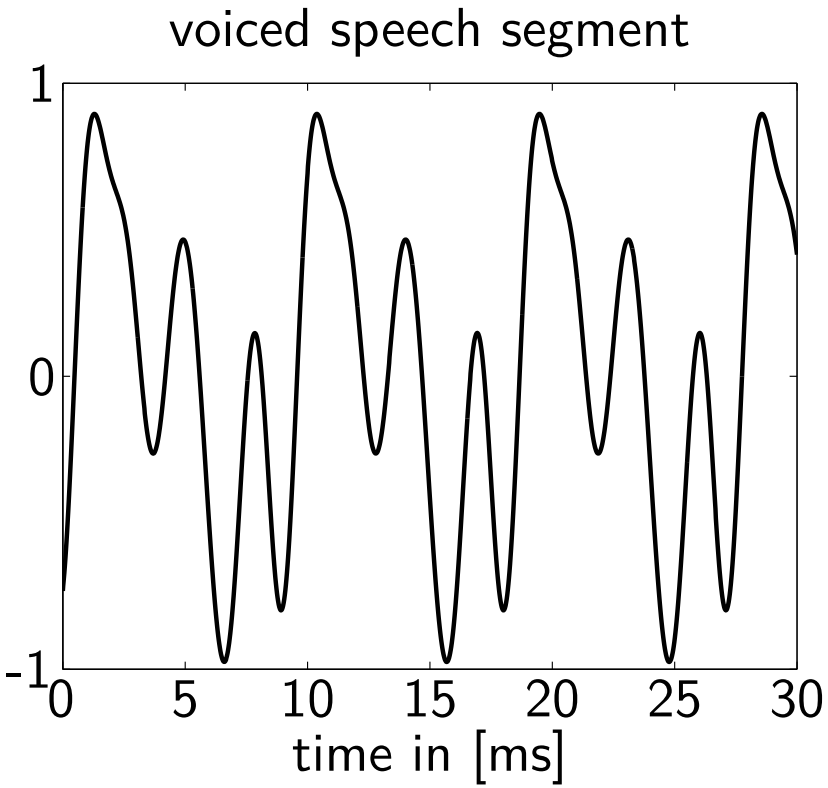




[6] D. W. Griffin and J. S. Lim. "Signal estimation from modified short-time Fourier transform". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.2 (1984), pp. 236–243.

Time Domain

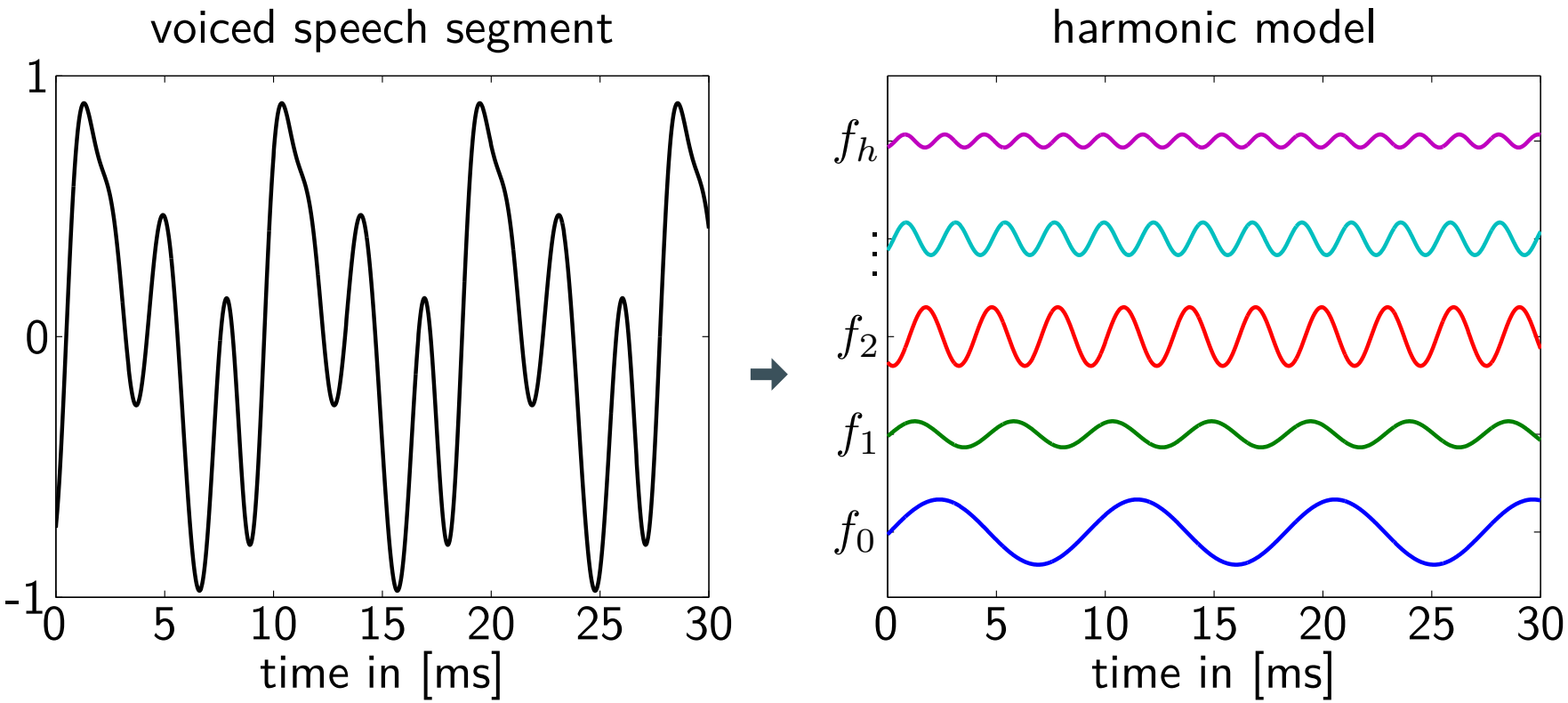
- For now we consider only voiced speech



Harmonic Model

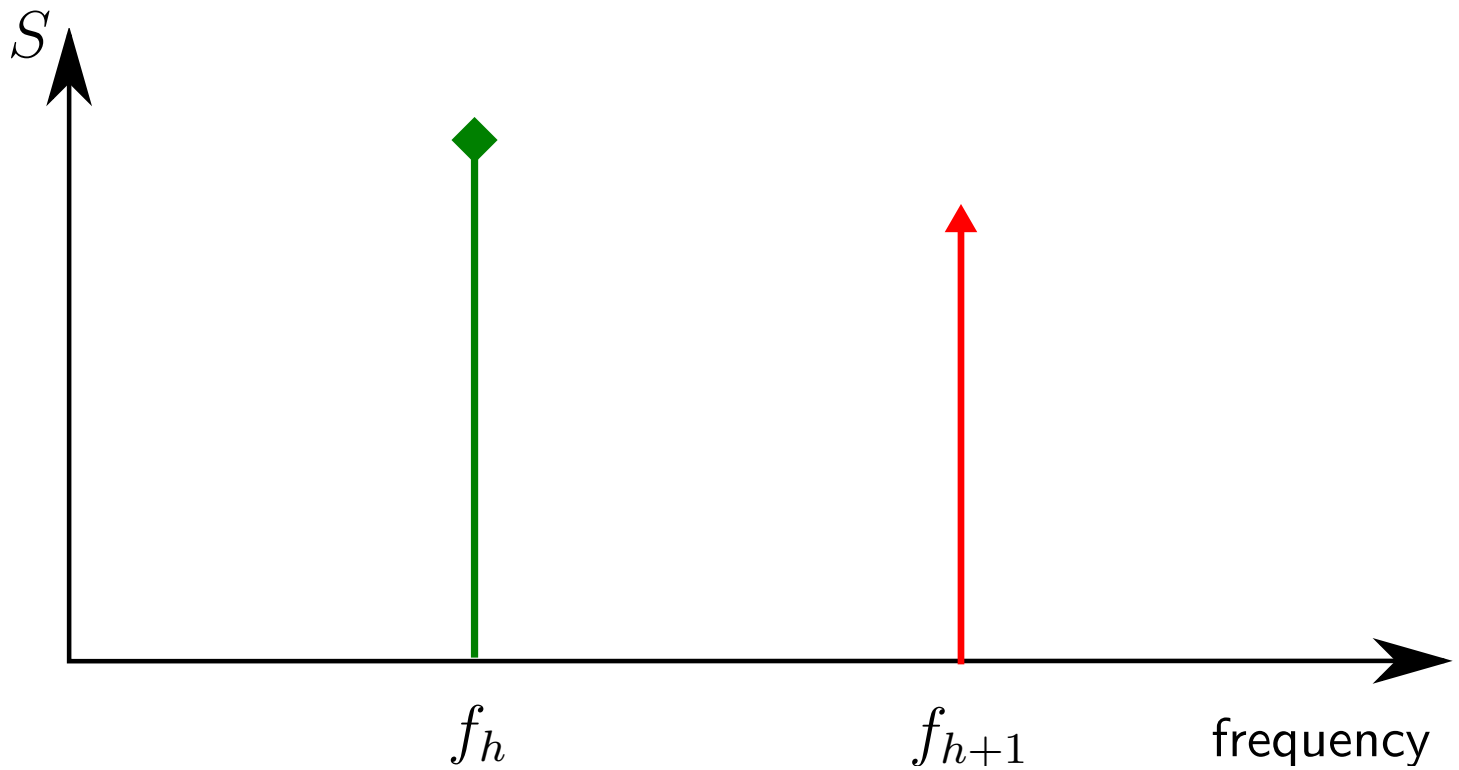
Time Domain

- For now we consider only voiced speech
- We model voiced speech with a harmonic model
  - ➡ Sum of *weighted* and *shifted* sinusoids at the fundamental frequency f_0 and its harmonics f_h



Frequency Domain

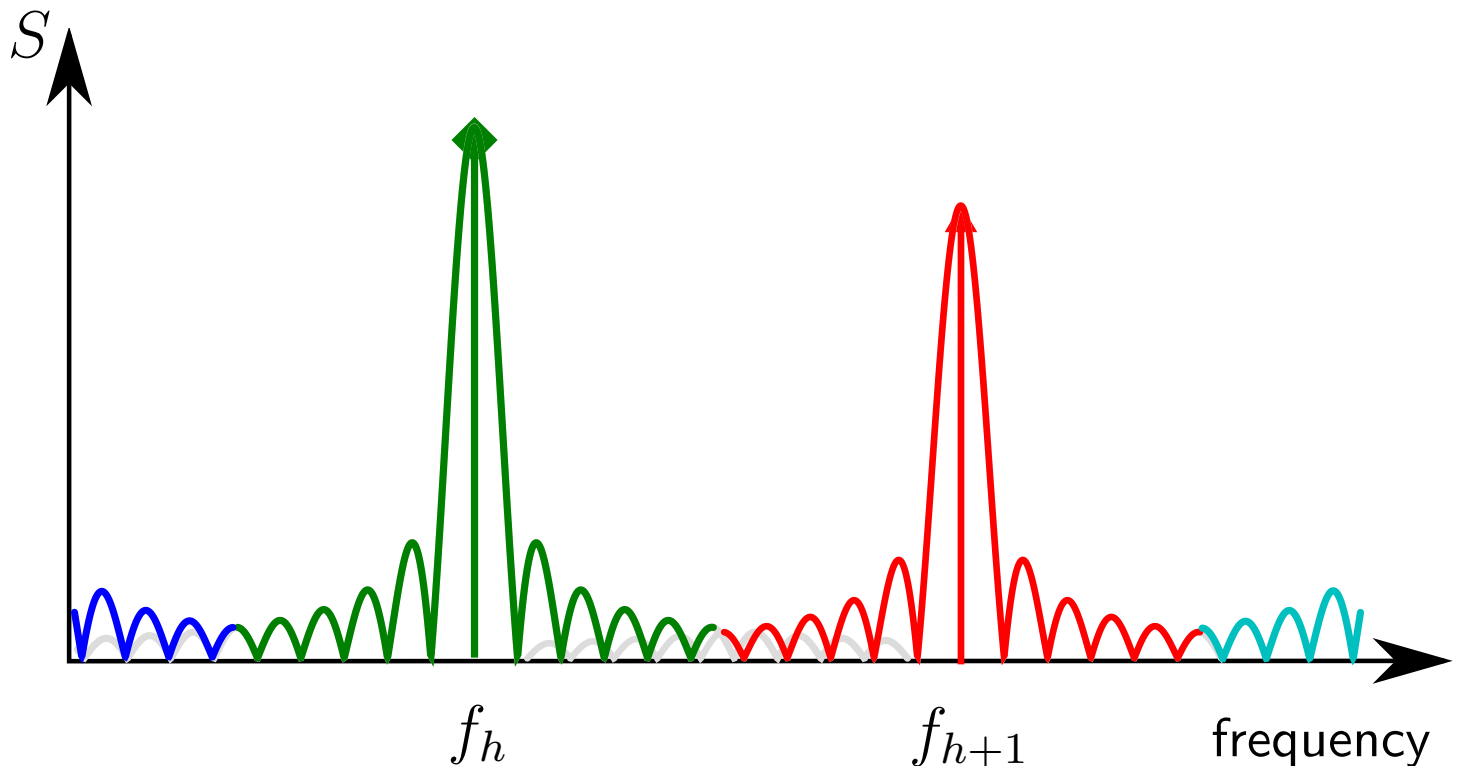
- Infinitely long signal $\rightarrow$ Impulse train at harmonic frequencies f_h



Harmonic Model

Frequency Domain

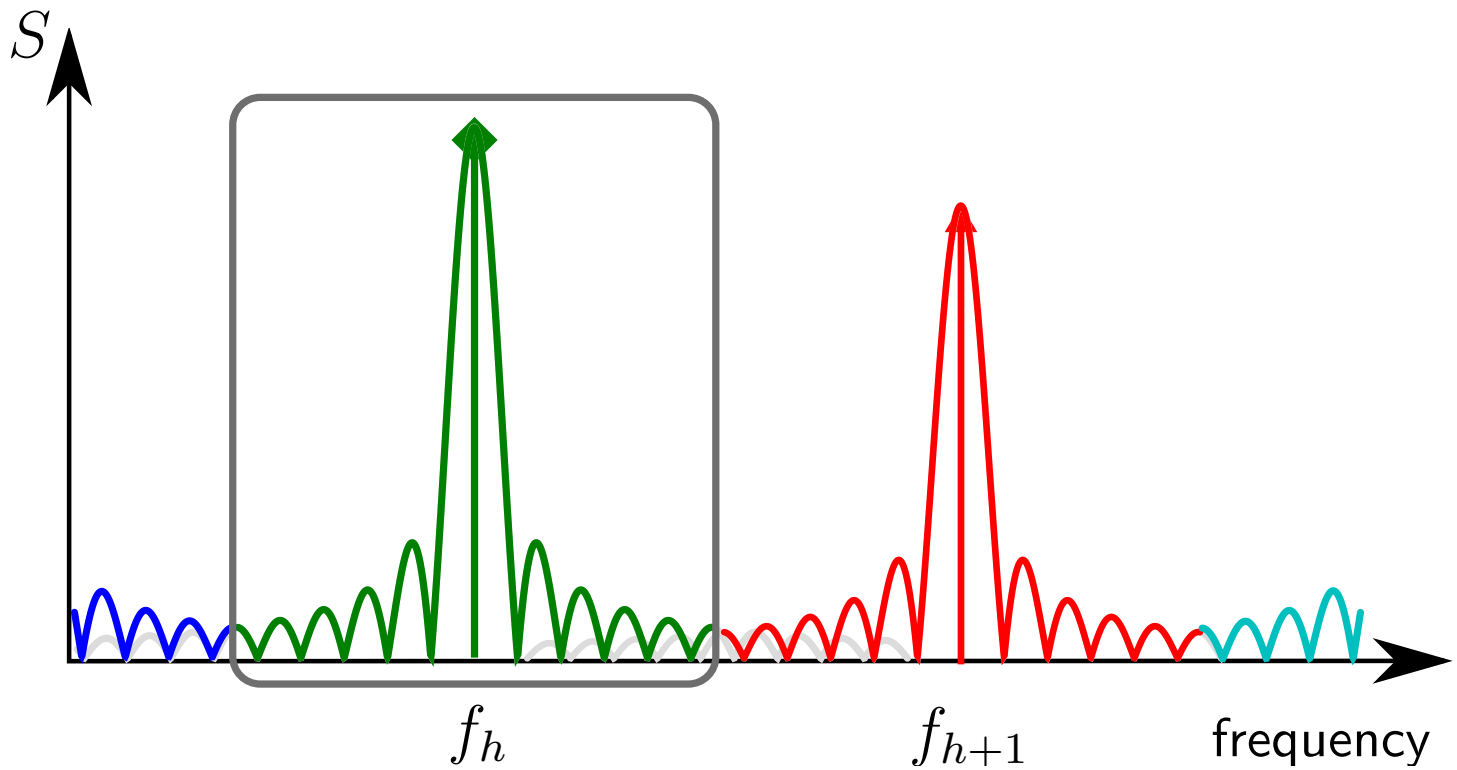
- Infinitely long signal $\rightarrow$ Impulse train at harmonic frequencies f_h
- Finite-length analysis window $w(t) \circ\!\!-\!\!\bullet W(f)$
- Multiplication $\circ\!\!-\!\!\bullet$ cyclic convolution



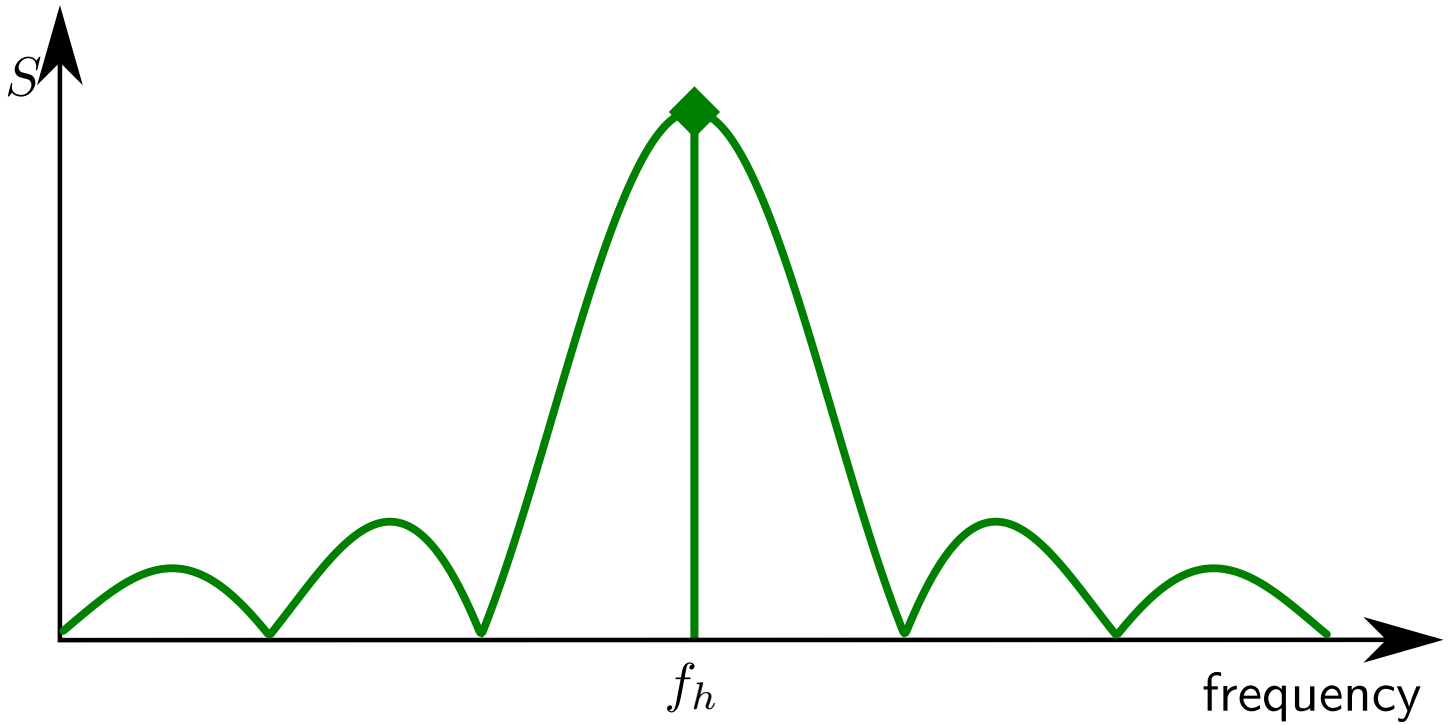
Harmonic Model

Frequency Domain

- Infinitely long signal $\rightarrow$ Impulse train at harmonic frequencies f_h
- Finite-length analysis window $w(t) \circ\!\!-\!\!\bullet\, W(f)$
- Multiplication $\circ\!\!-\!\!\bullet$ cyclic convolution
- Process vicinity of each harmonic separately

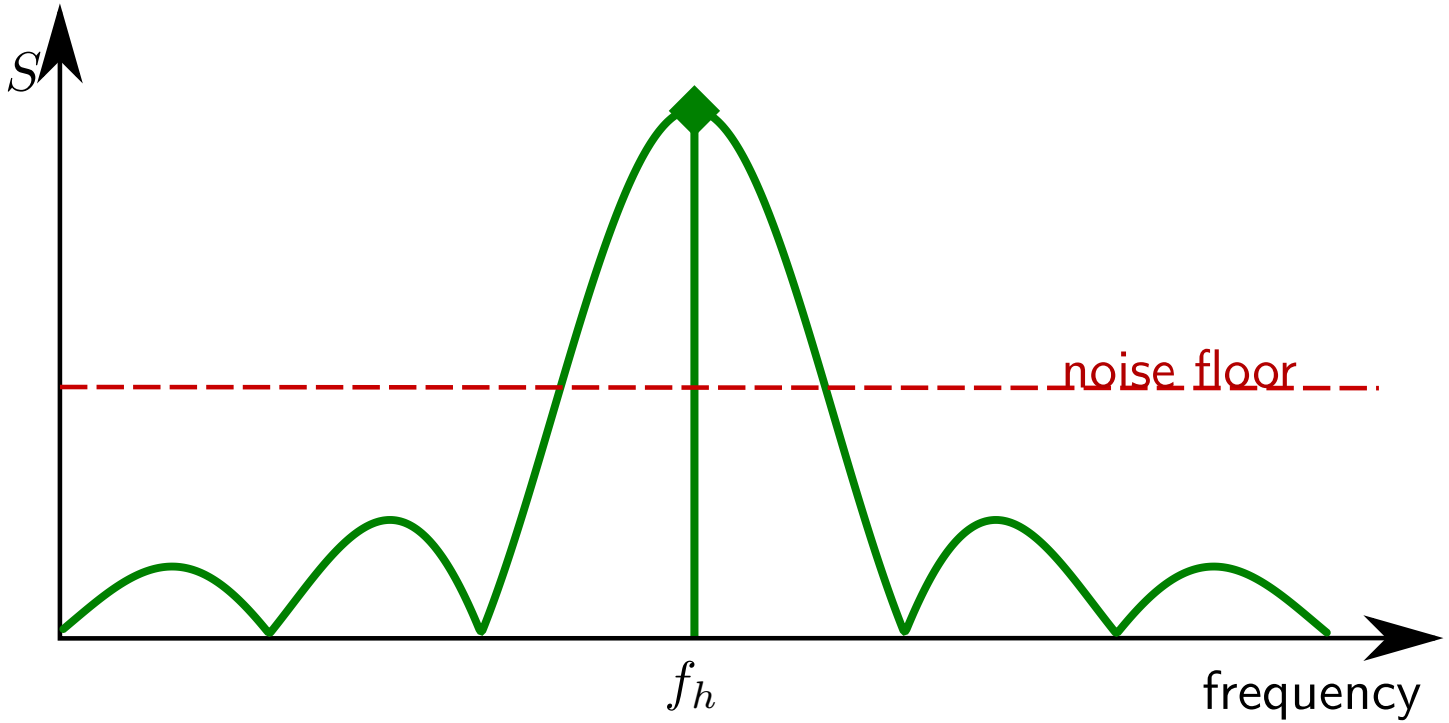


Across Frequency



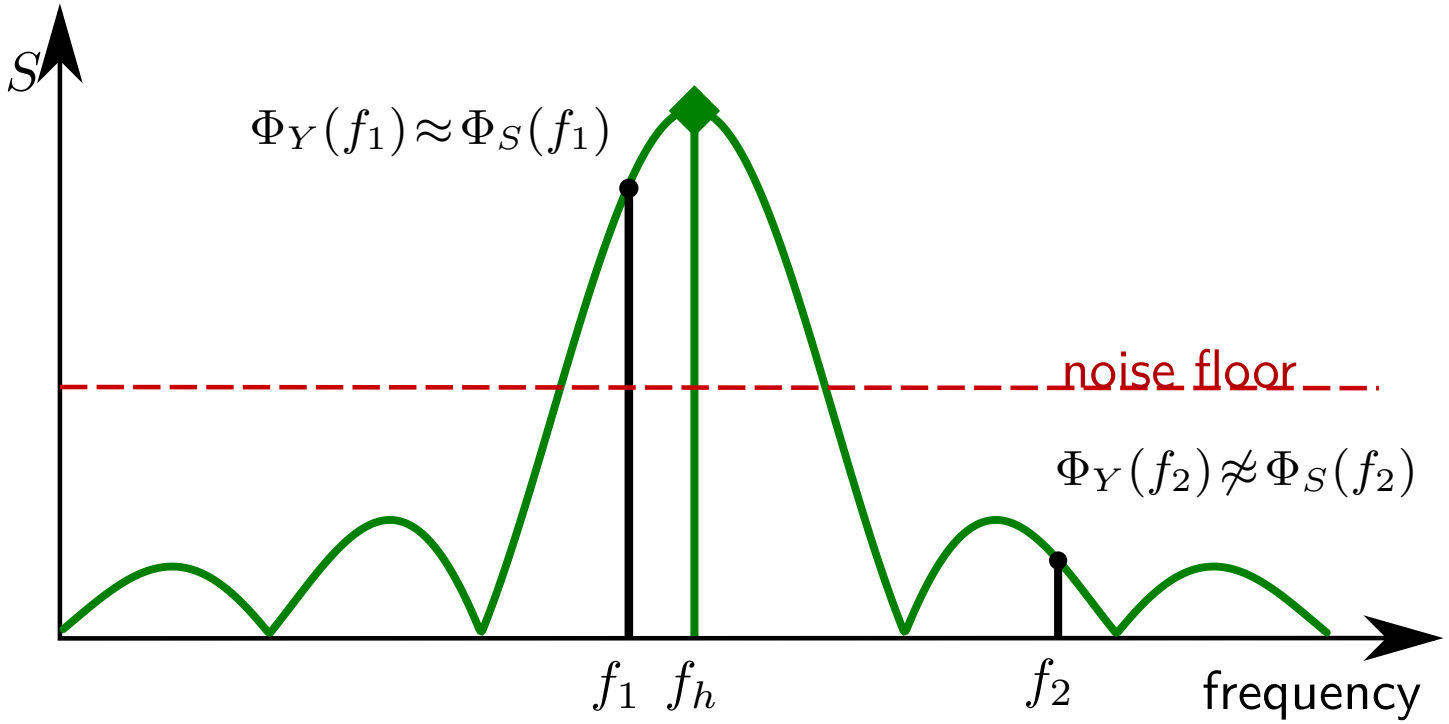
Proposed Phase Reconstruction

Across Frequency



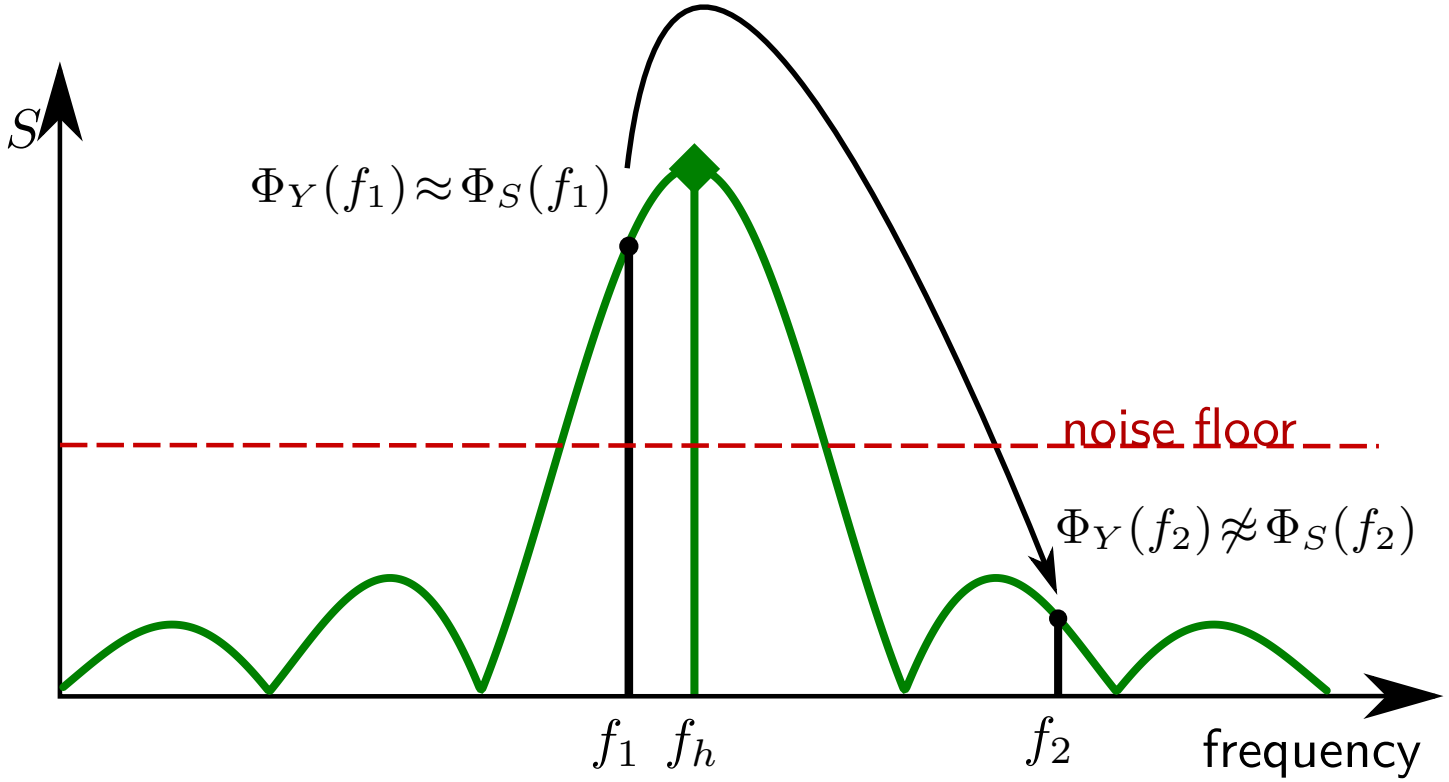
Proposed Phase Reconstruction

Across Frequency



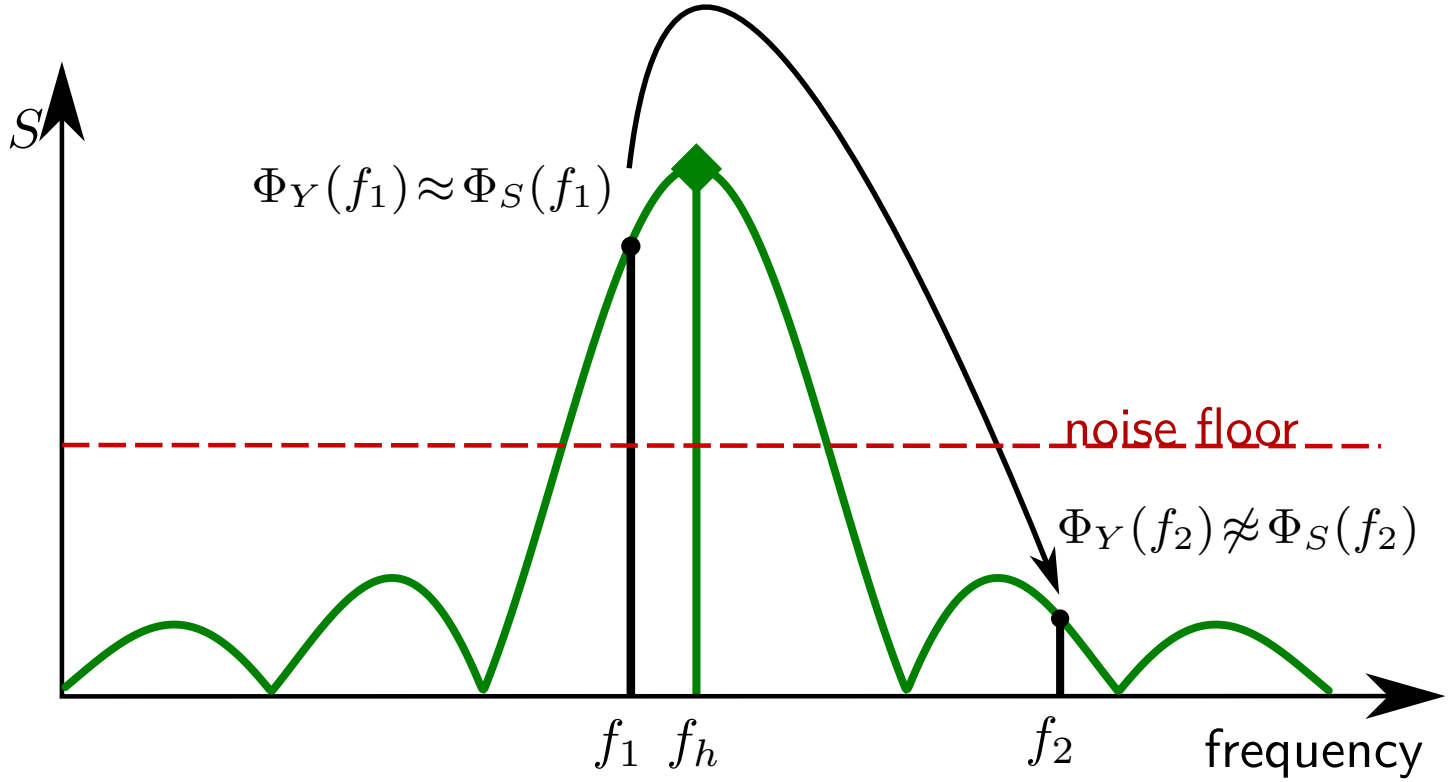
Proposed Phase Reconstruction

Across Frequency



Proposed Phase Reconstruction

Across Frequency



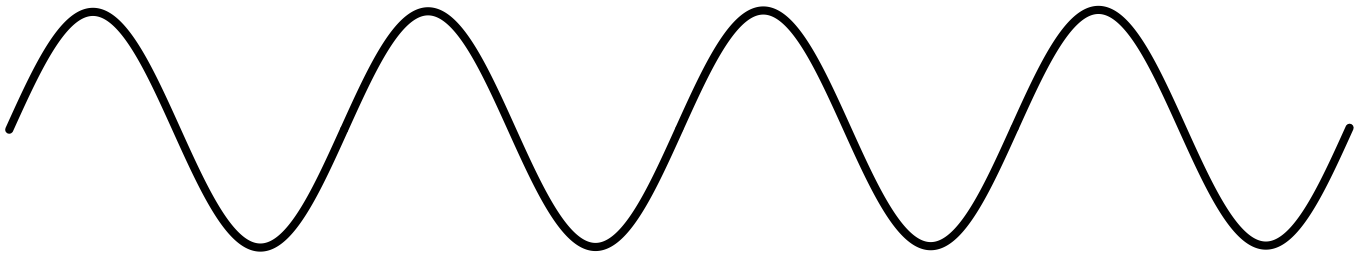
Thanks to the convolution with $W(f)$, the phases are directly related:

$$\begin{aligned}\Phi_S(f_2) &= \Phi_S(f_1) - \Phi_{\mathrm{W}}(f_1 - f_h) + \Phi_{\mathrm{W}}(f_2 - f_h) \\ &\approx \Phi_Y(f_1) - \Phi_{\mathrm{W}}(f_1 - f_h) + \Phi_{\mathrm{W}}(f_2 - f_h)\end{aligned}$$

Proposed Phase Reconstruction

Across Time

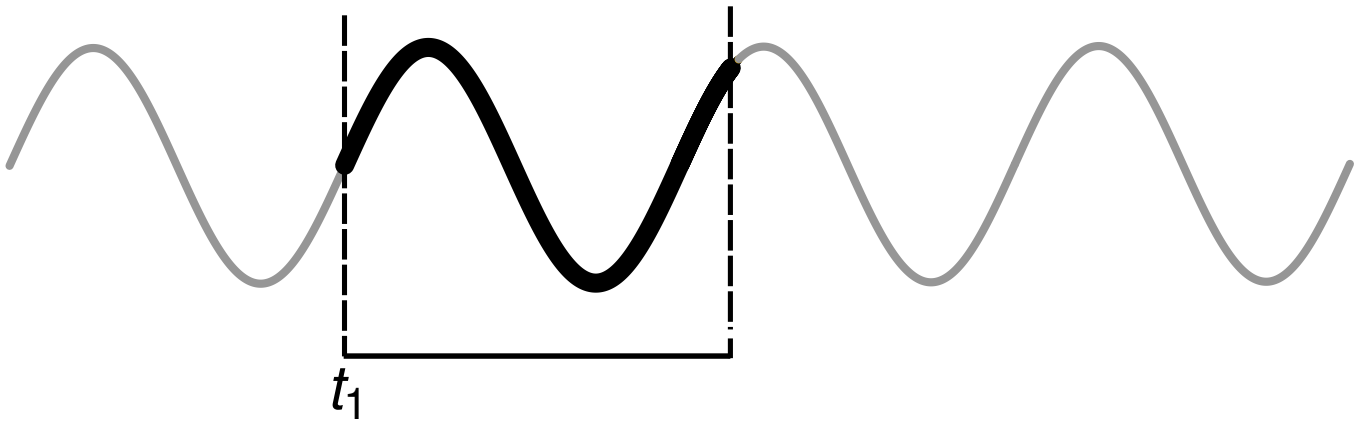
- Assumption: In each frequency band, only one harmonic is active
- Signal with frequency f_0



Proposed Phase Reconstruction

Across Time

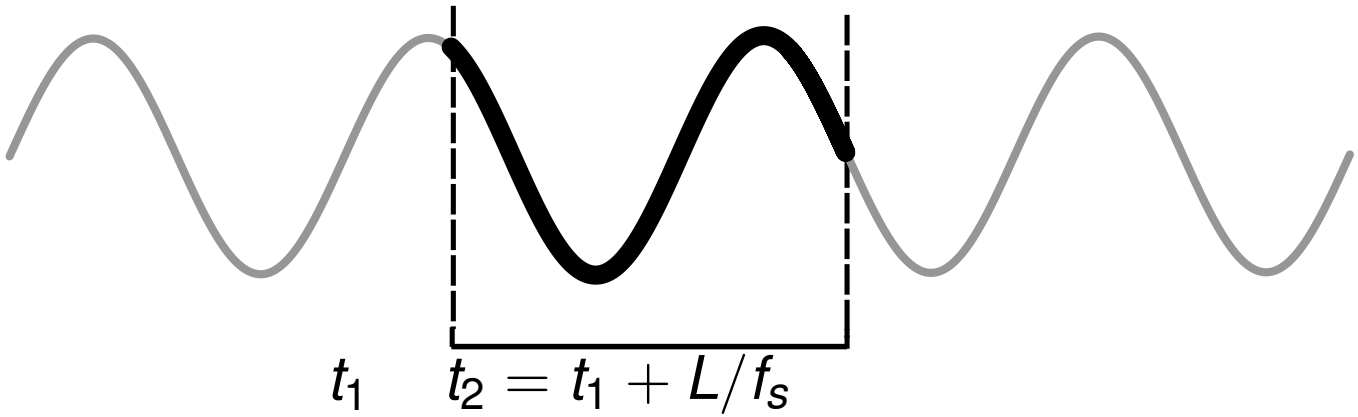
- Assumption: In each frequency band, only one harmonic is active
- Signal with frequency f_0
- Phase at t_1:$\phi(t_1) = 2\pi f_0 t_1 + \phi_0$, with initial phase ϕ_0



Proposed Phase Reconstruction

Across Time

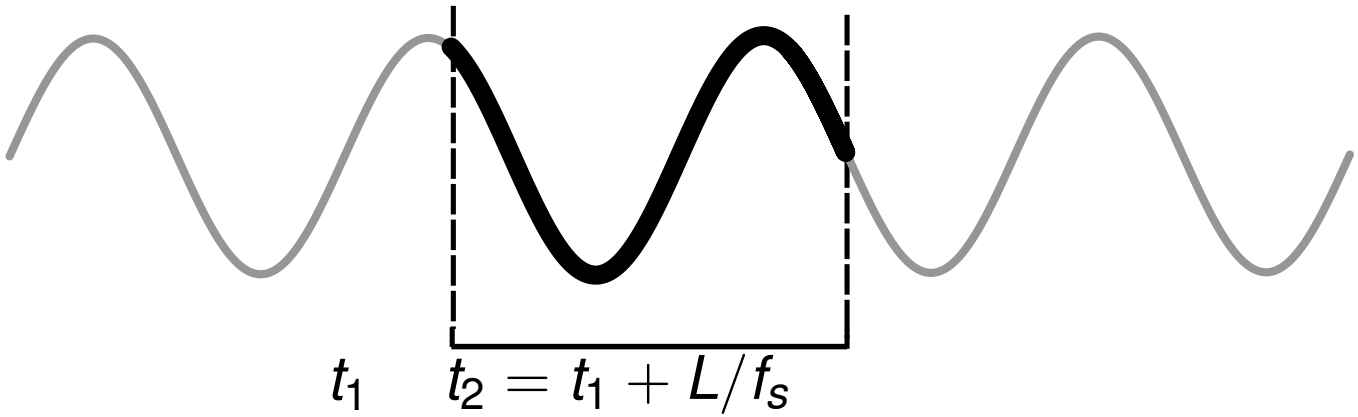
- Assumption: In each frequency band, only one harmonic is active
- Signal with frequency f_0
- Phase at t_1:$\phi(t_1) = 2\pi f_0 t_1 + \phi_0$, with initial phase ϕ_0
- Phase at t_2: $\phi(t_2) = 2\pi f_0 t_2 + \phi_0$



Proposed Phase Reconstruction

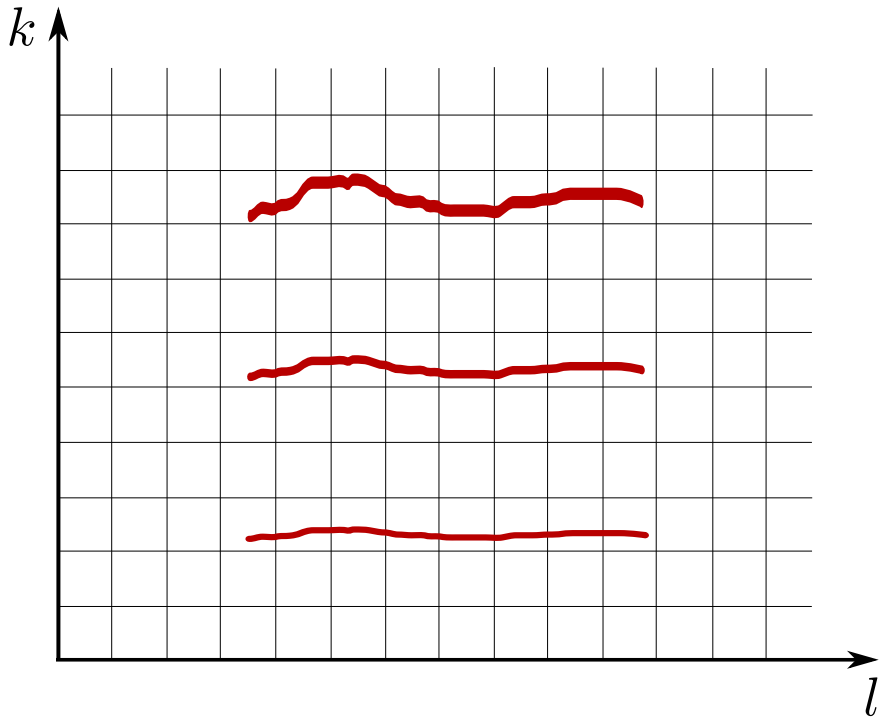
Across Time

- Assumption: In each frequency band, only one harmonic is active
- Signal with frequency f_0
- Phase at t_1:$\phi(t_1) = 2\pi f_0 t_1 + \phi_0$, with initial phase ϕ_0
- Phase at t_2: $\phi(t_2) = 2\pi f_0 t_2 + \phi_0$
- STFT segment shift of L samples: $\Delta\phi = \phi(t_2) - \phi(t_1) = 2\pi f_0 \frac{L}{f_s}$



Proposed Phase Reconstruction[7]

Overview

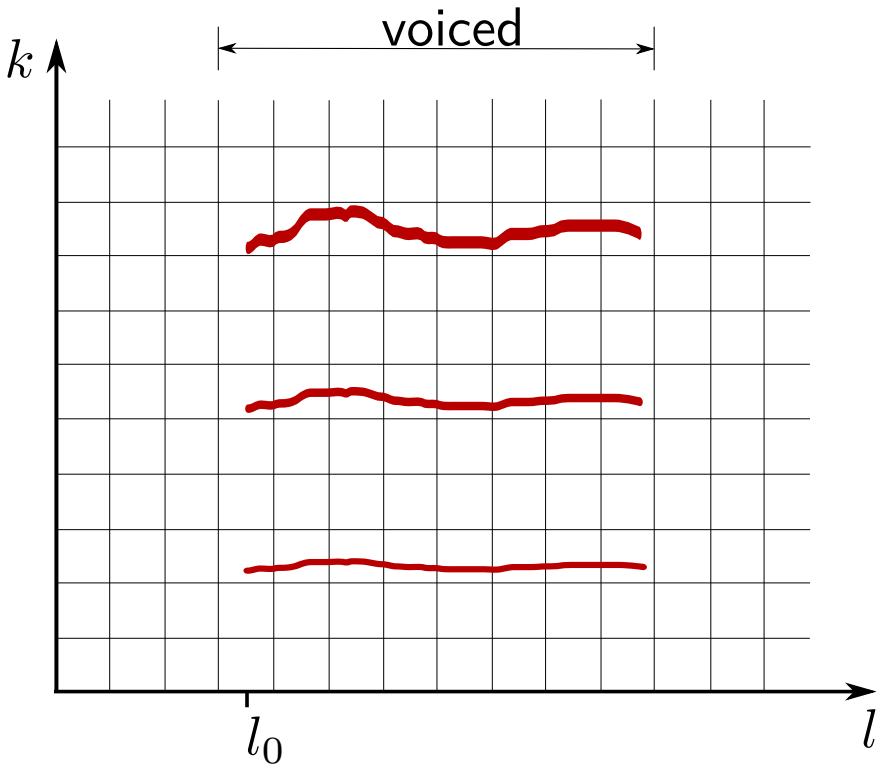


1. Compute STFT

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

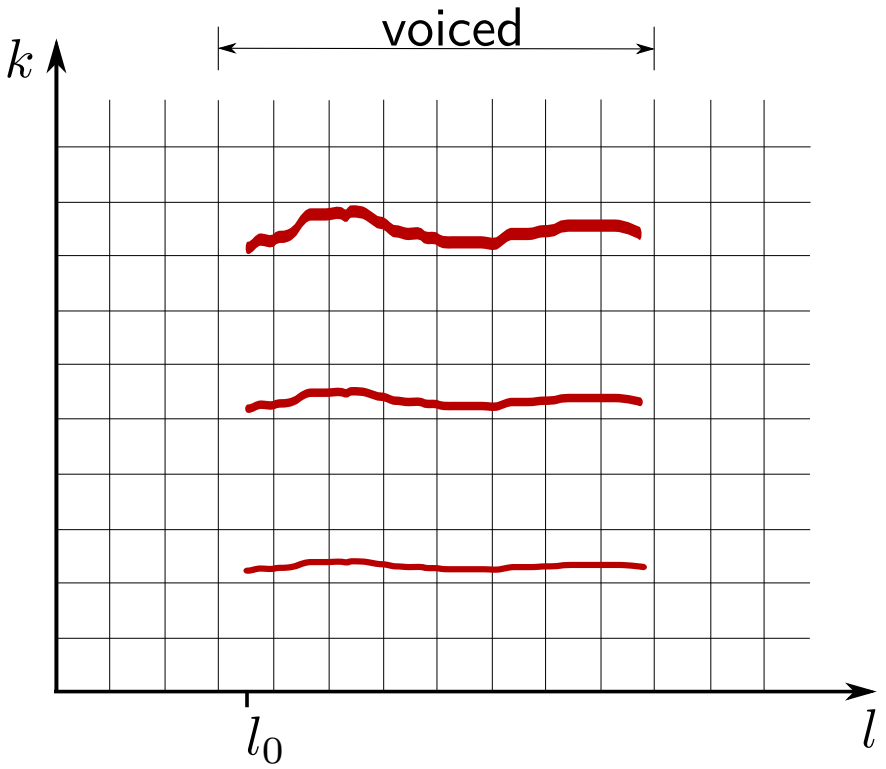


1. Compute STFT
2. Estimate fundamental freq. f_0

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

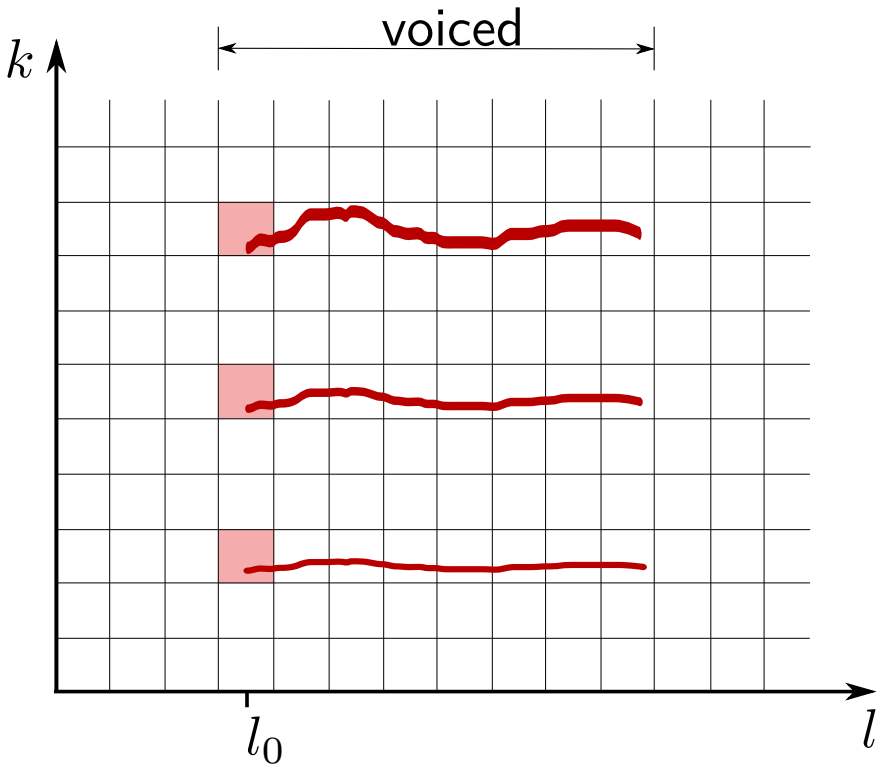


1. Compute STFT
2. Estimate fundamental freq. f_0
3. Reconstruct the clean speech phase Φ_S along time in frequency bands that contain harmonics (red)

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

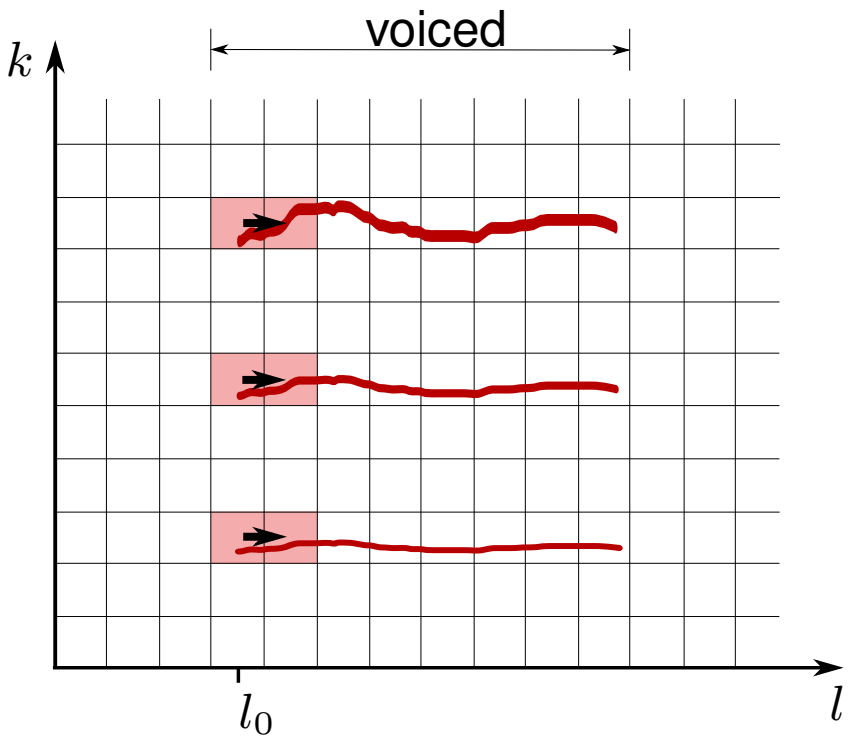


1. Compute STFT
2. Estimate fundamental freq. f_0
3. Reconstruct the clean speech phase Φ_S along time in frequency bands that contain harmonics (red)

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

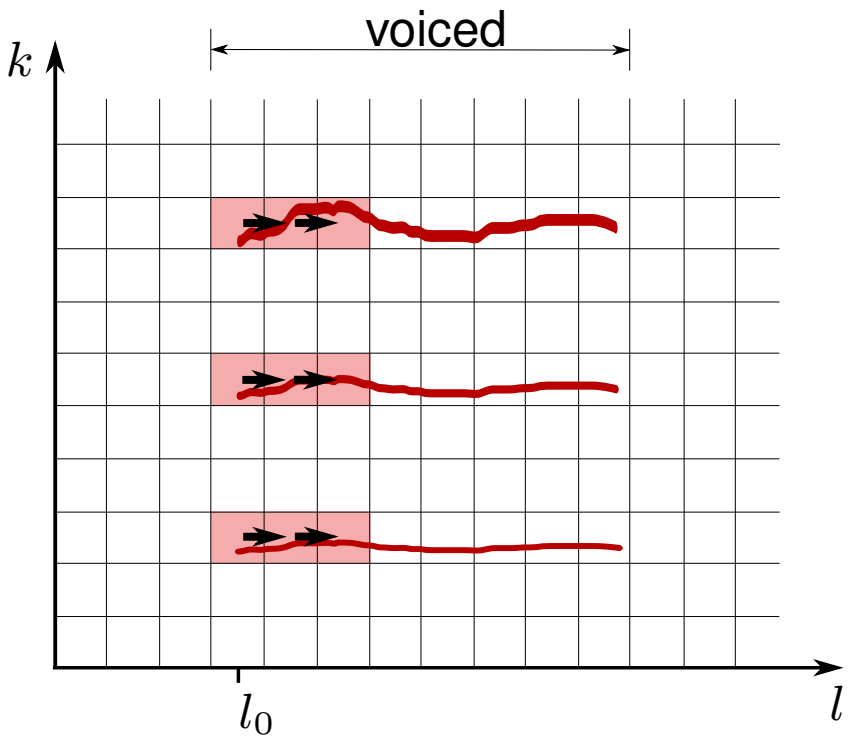


1. Compute STFT
2. Estimate fundamental freq. f_0
3. Reconstruct the clean speech phase Φ_S along time in frequency bands that contain harmonics (red)

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

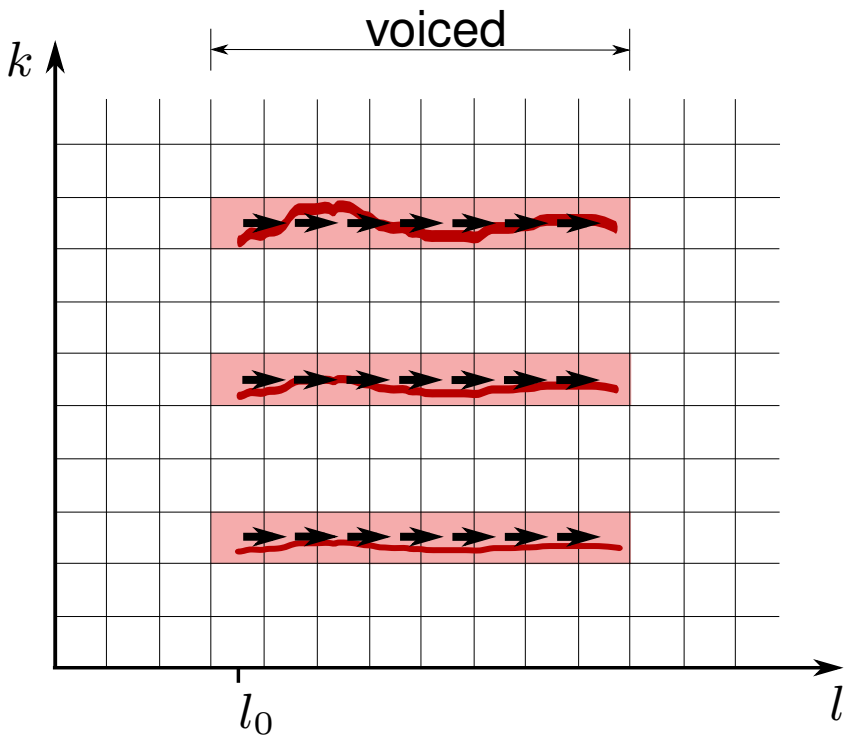


1. Compute STFT
2. Estimate fundamental freq. f_0
3. Reconstruct the clean speech phase Φ_S along time in frequency bands that contain harmonics (red)

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

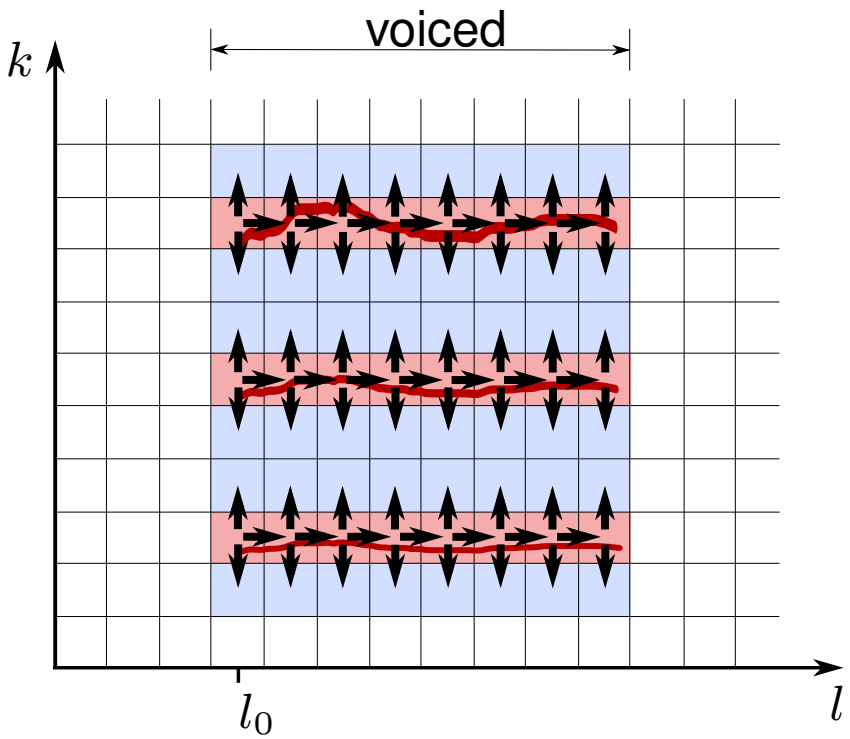


1. Compute STFT
2. Estimate fundamental freq. f_0
3. Reconstruct the clean speech phase Φ_S along time in frequency bands that contain harmonics (red)

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Proposed Phase Reconstruction[7]

Overview

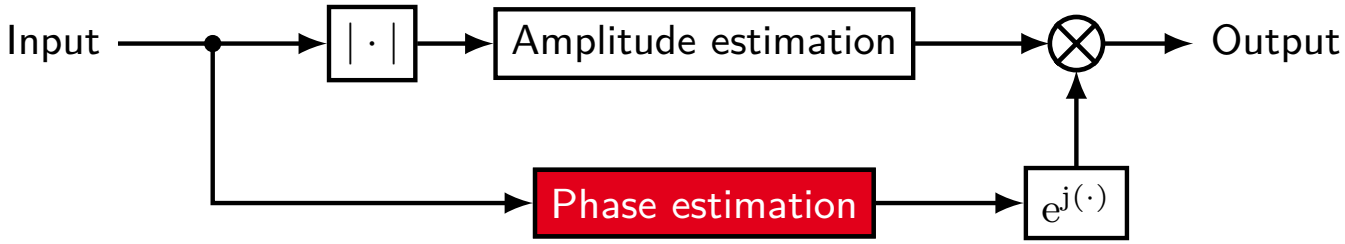


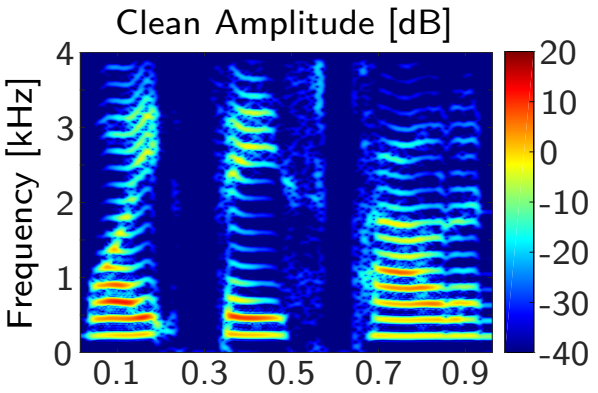
1. Compute STFT
2. Estimate fundamental freq. f_0
3. Reconstruct the clean speech phase Φ_S along time in frequency bands that contain harmonics (red)
4. Starting from these bands, the remaining phases are reconstructed along frequency (blue)

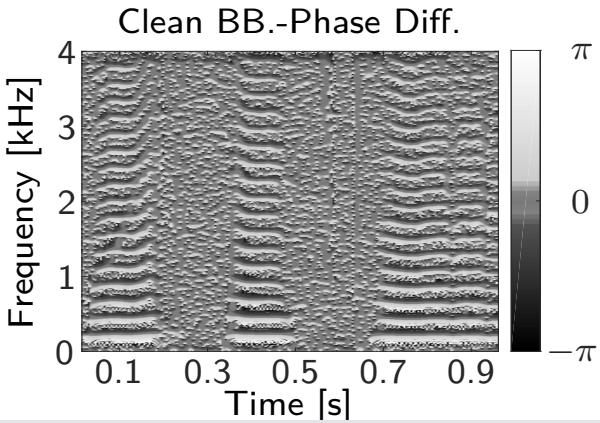
[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

Example

Phase-Only Speech Enhancement

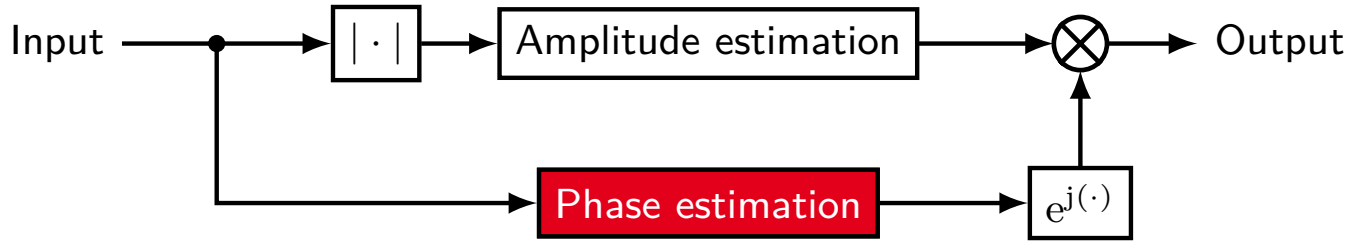


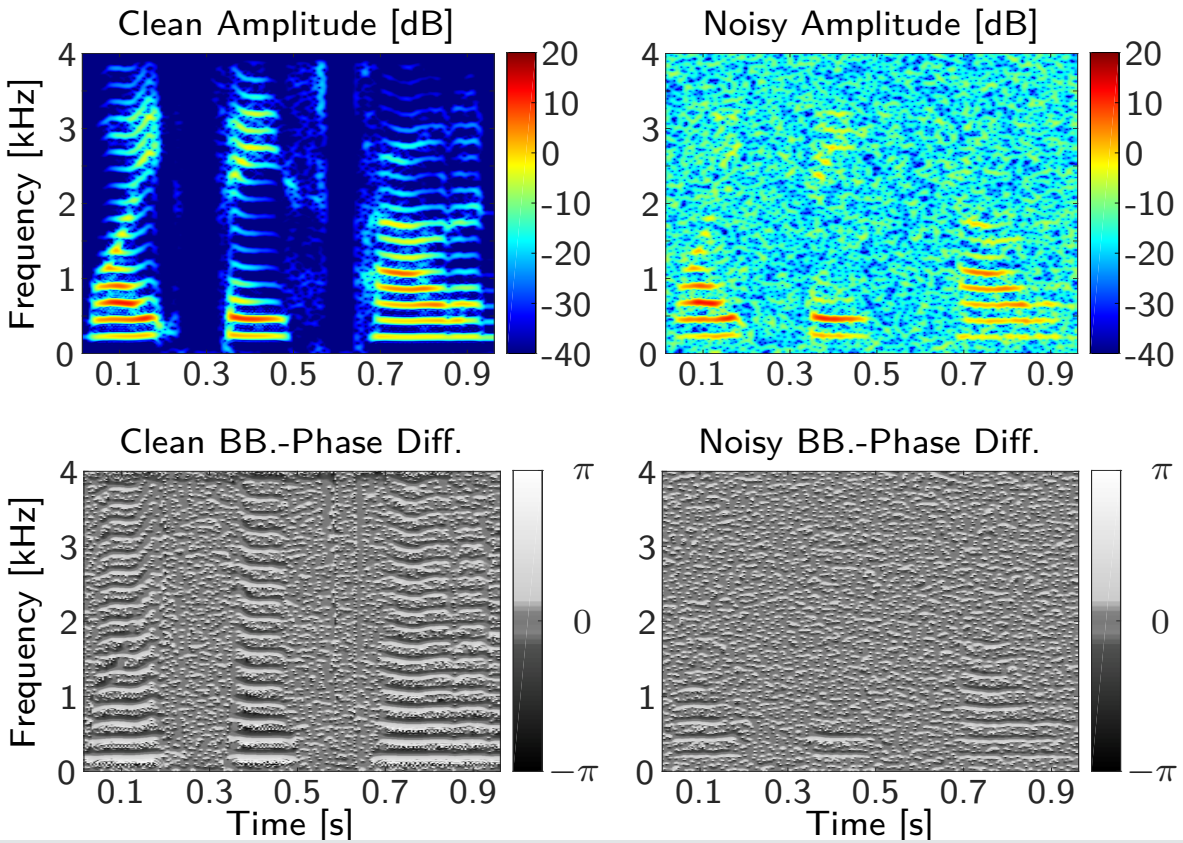




Example

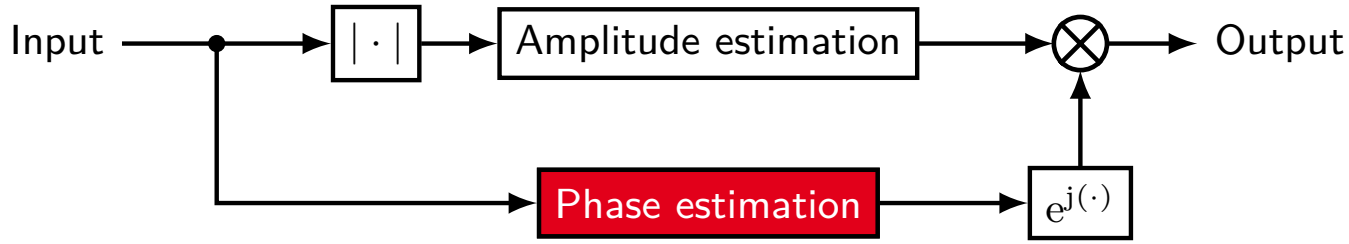
Phase-Only Speech Enhancement

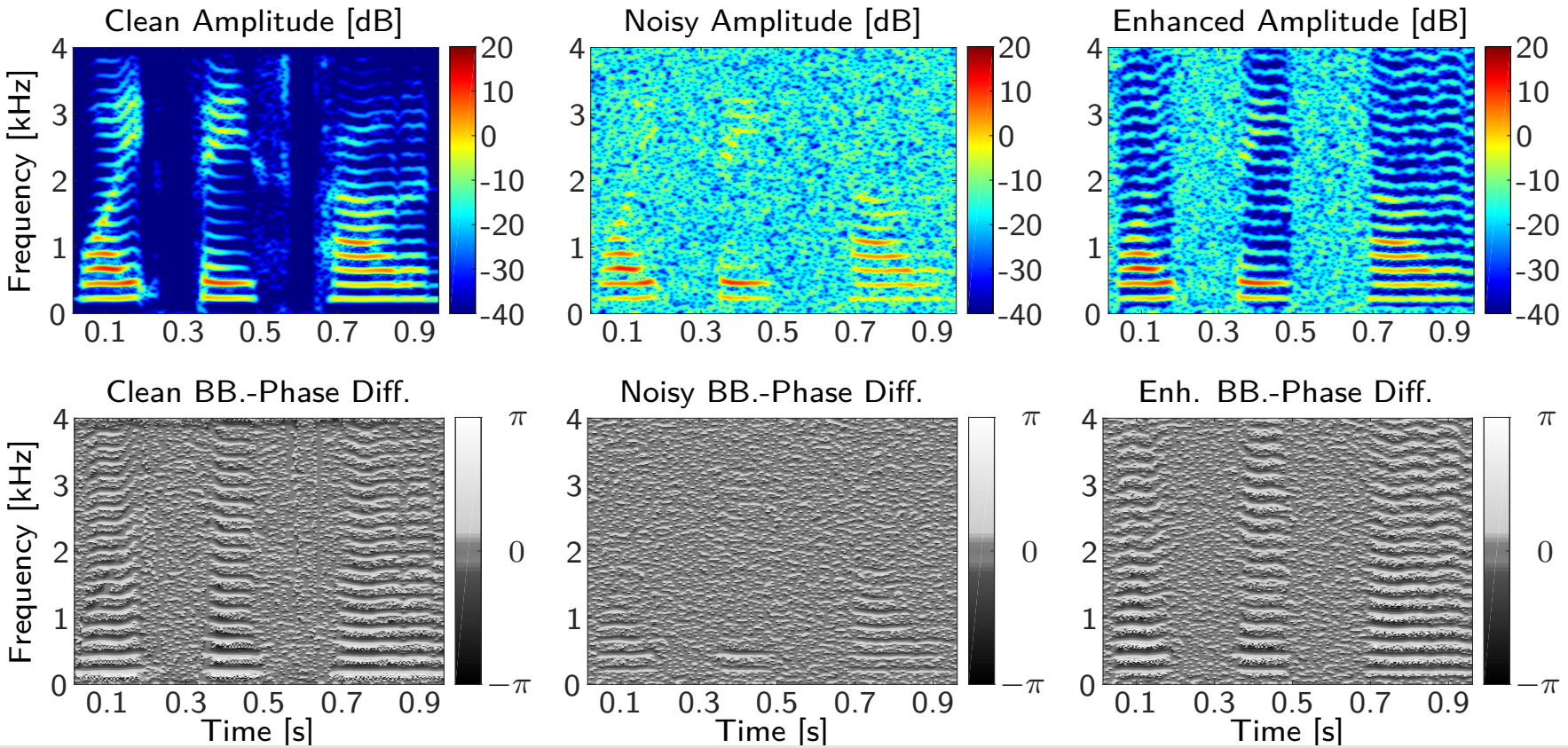


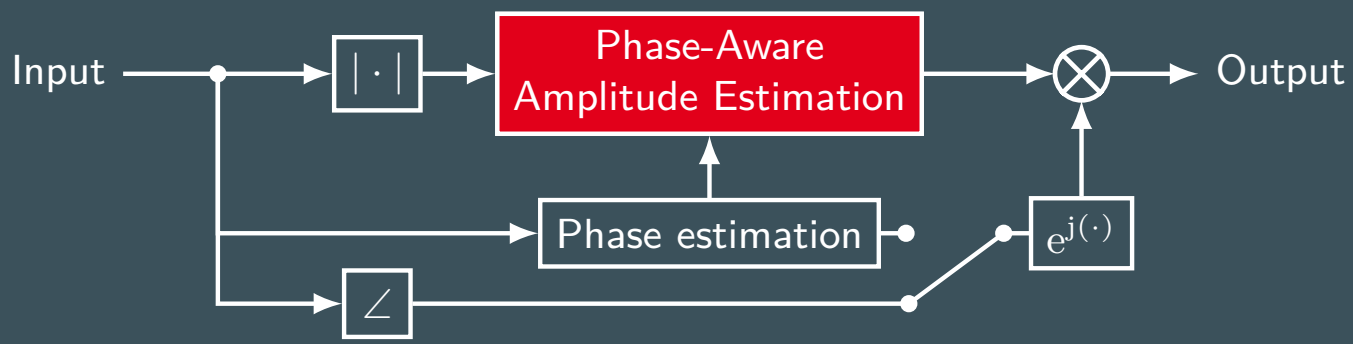


Example

Phase-Only Speech Enhancement

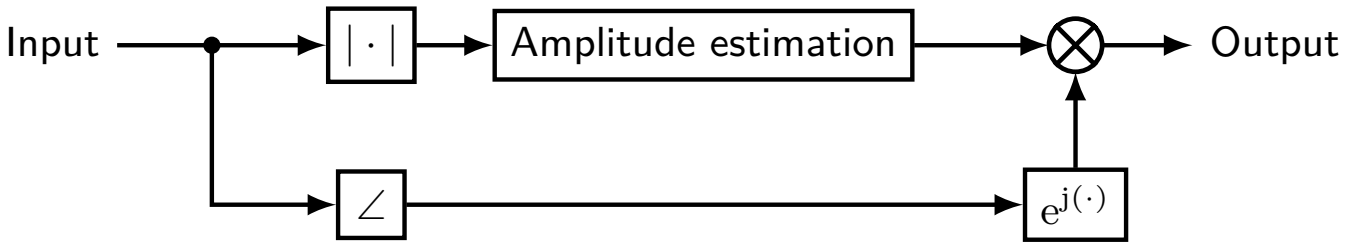






Phase-Aware Speech Estimation

Phase-Aware Speech Enhancement



- Conventional speech enhancement schemes only modify the spectral amplitude

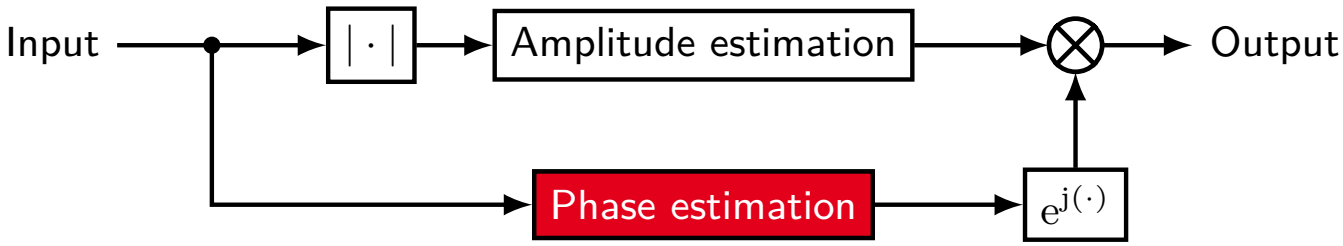
[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

[8] T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

[9] T. Gerkmann. "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase". In: *IEEE Trans. Signal Process.* 62.16 (2014), pp. 4199–4208.

[10] M. Krawczyk-Becker and T. Gerkmann. "On MMSE-Based Estimation of Spectral Speech Coefficients Under Phase-Uncertainty". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.12 (2016), pp. 2251–2262.

Phase-Aware Speech Enhancement



- Conventional speech enhancement schemes only modify the spectral amplitude
- **But:** The clean phase can be estimated (e.g. via[7]) and used
  1. to directly replace the noisy phase

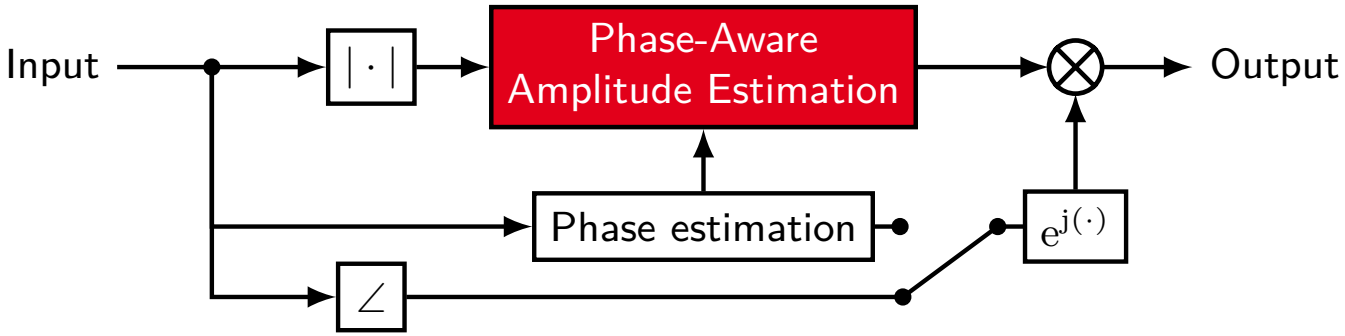
[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

[8] T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

[9] T. Gerkmann. "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase". In: *IEEE Trans. Signal Process.* 62.16 (2014), pp. 4199–4208.

[10] M. Krawczyk-Becker and T. Gerkmann. "On MMSE-Based Estimation of Spectral Speech Coefficients Under Phase-Uncertainty". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.12 (2016), pp. 2251–2262.

Phase-Aware Speech Enhancement



- Conventional speech enhancement schemes only modify the spectral amplitude
- **But:** The clean phase can be estimated (e.g. via[7]) and used
  1. to directly replace the noisy phase
  2. and/or as extra information to facilitate speech estimation (e.g.[8][9][10])

[7] M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

[8] T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

[9] T. Gerkmann. "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase". In: *IEEE Trans. Signal Process.* 62.16 (2014), pp. 4199–4208.

[10] M. Krawczyk-Becker and T. Gerkmann. "On MMSE-Based Estimation of Spectral Speech Coefficients Under Phase-Uncertainty". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.12 (2016), pp. 2251–2262.

Phase-Aware Amplitude Estimation

- Notation:
  - Noisy speech: $Y = |Y|\mathrm{e}^{\mathrm{j}\Phi_Y} = \underbrace{|S|\mathrm{e}^{\mathrm{j}\Phi_S}}_{\text{clean speech}} + \underbrace{|V|\mathrm{e}^{\mathrm{j}\Phi^V}}_{\text{noise}}$
- Classic, phase-blind Bayesian amplitude estimation
  - $\widehat{|S|}^{\beta} = \mathrm{E}\big(|S|^{\beta} \mid Y\big)$

  ➡ The clean speech phase is considered unknown and random.

Phase-Aware Amplitude Estimation

- Notation:
  - Noisy speech: $Y = |Y|\mathrm{e}^{\mathrm{j}\Phi_Y} = \underbrace{|S|\mathrm{e}^{\mathrm{j}\Phi_S}}_{\text{clean speech}} + \underbrace{|V|\mathrm{e}^{\mathrm{j}\Phi^V}}_{\text{noise}}$
- Classic, phase-blind Bayesian amplitude estimation
  - $\widehat{|S|}^{\beta} = \mathrm{E}\big(|S|^{\beta} \mid Y\big)$
  - ➡ The clean speech phase is considered unknown and random.

➡ **Idea:** Use estimated clean speech phase as prior knowledge!

- $\widehat{|S|}^{\beta} = \mathrm{E}\big(|S|^{\beta} \mid Y, \Phi_S\big)$

Phase-Aware Amplitude Estimation[8]

Proposed phase-aware amplitude estimator

$$\widehat{|S|} = \left(\mathrm{E}\left(|S|^{\beta} \mid Y, \Phi_S\right)\right)^{\frac{1}{\beta}}$$
$$= \sqrt{\frac{1}{2}\frac{\xi}{\mu+\xi}\sigma_{\mathrm{V}}^2}\left(\frac{\Gamma(2\mu+\beta)}{\Gamma(2\mu)}\frac{\mathrm{D}_{-(2\mu+\beta)}(\nu)}{\mathrm{D}_{-(2\mu)}(\nu)}\right)^{\frac{1}{\beta}}$$

β, μ: compression, degree of super-Gaussianity

$\mathrm{D}_{\cdot}(\nu)$, $\Gamma(\cdot)$: parabolic cylinder function, Gamma function,

$\xi = \sigma_{\mathrm{S}}^2/\sigma_{\mathrm{V}}^2$: *a priori* SNR

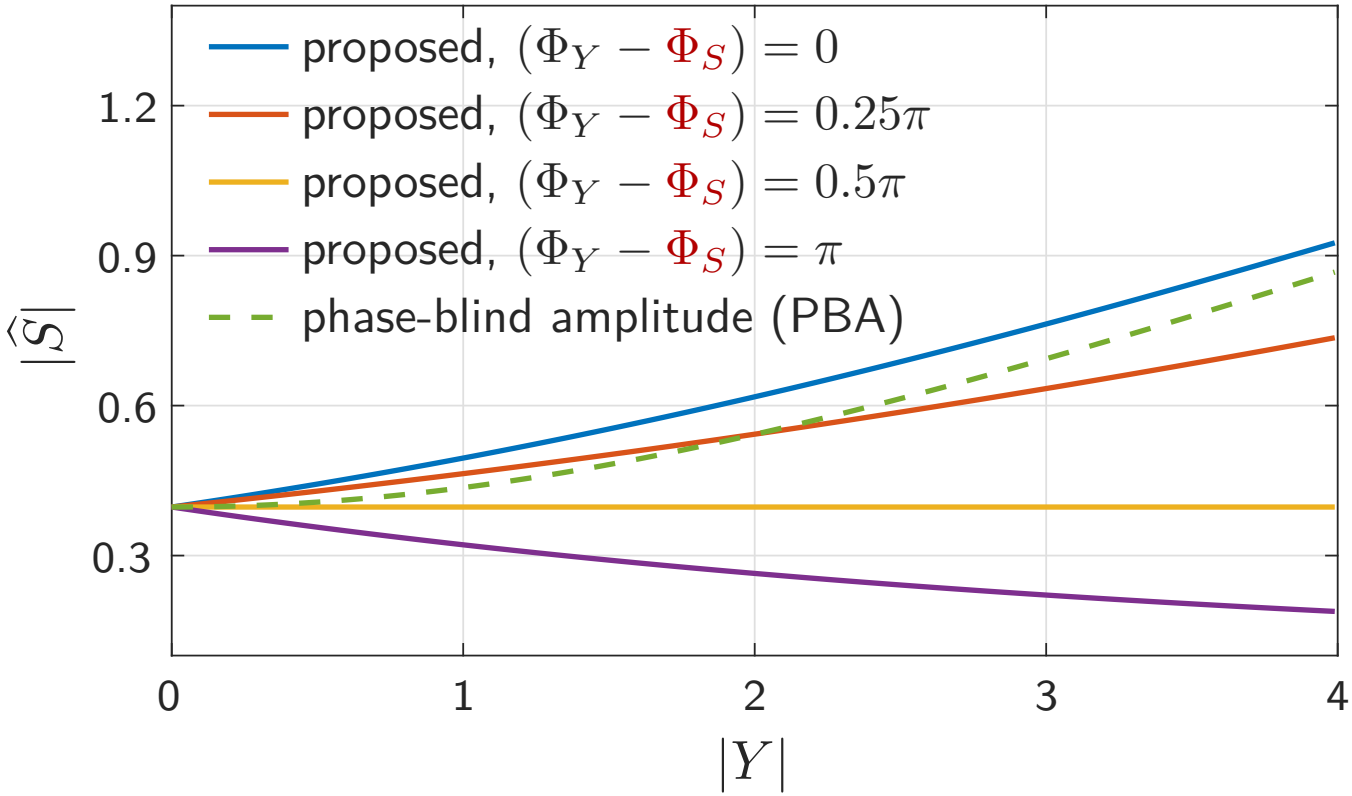
ν: contains the phase difference $\Delta\phi$

$$\nu = -\frac{|Y|}{\sigma_{\mathrm{V}}}\sqrt{2\frac{\xi}{\mu+\xi}}\cos(\underbrace{\Phi_Y - \Phi_S}_{\Delta\phi})$$

[8] T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

Phase-Aware Amplitude Estimation[8]

Input-Output Characteristic at low SNRs

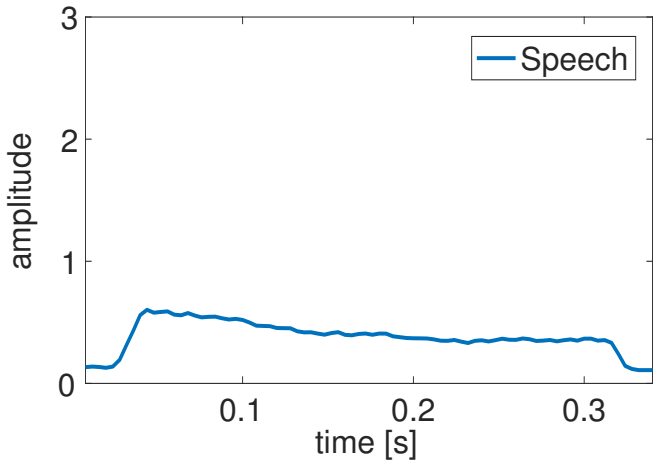


- If $(\Phi_Y - \Phi_S) \to 0$, *less* attenuation is applied,
- If $(\Phi_Y - \Phi_S)$ large, *more* attenuation is applied.

➡ New way to distinguish noise outliers from speech!

[8] T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

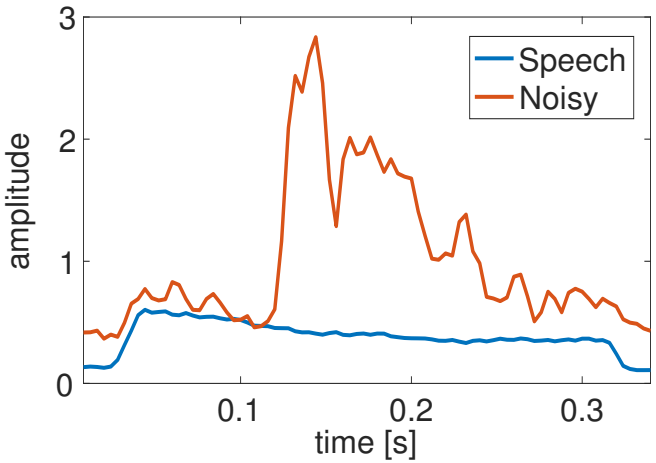
Illustrative Example



- Speech amplitude averaged over 1 kHz

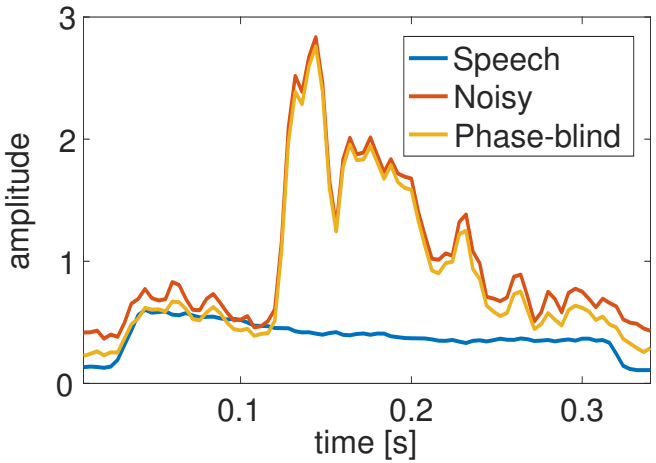
Phase-Aware Amplitude Estimation

Illustrative Example



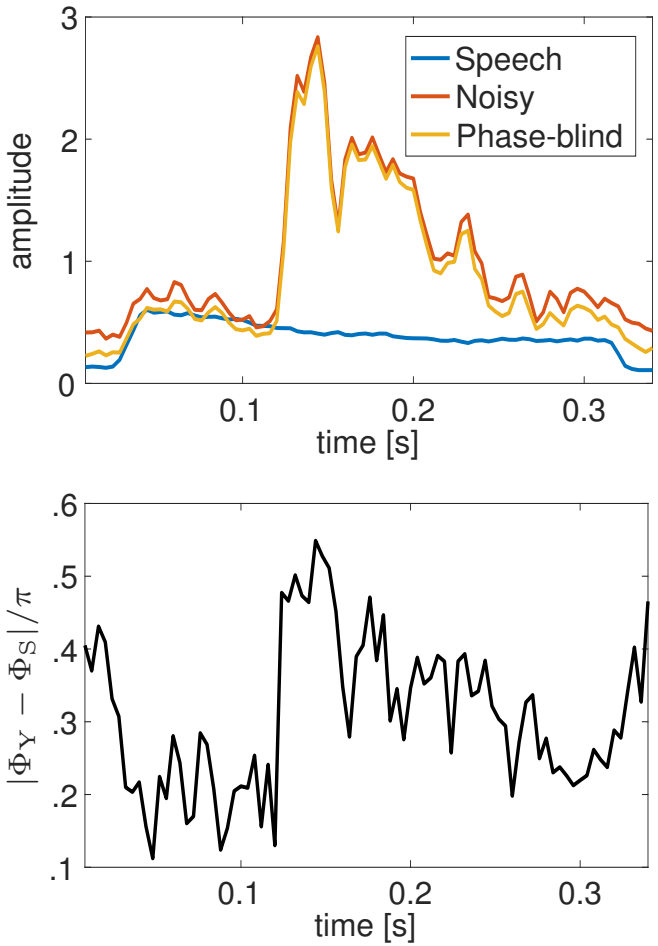
- Speech amplitude averaged over 1 kHz
- Distorted by a noise burst

Illustrative Example



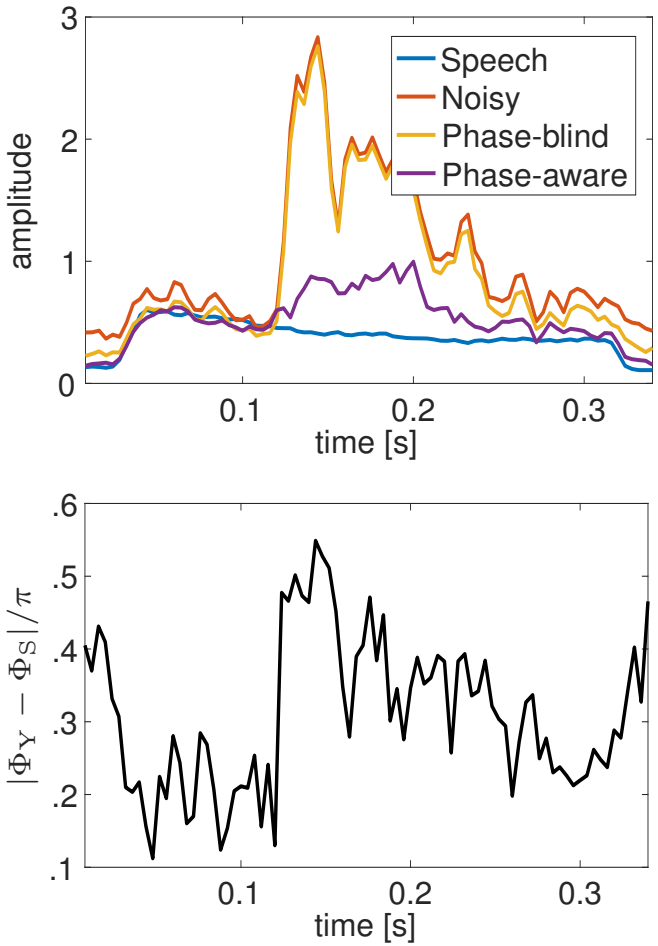
- Speech amplitude averaged over 1 kHz
- Distorted by a noise burst
- Noise is hardly suppressed by conventional phase-blind estimators

Illustrative Example



- Speech amplitude averaged over 1 kHz
- Distorted by a noise burst
- Noise is hardly suppressed by conventional phase-blind estimators
- Phase difference yields additional info

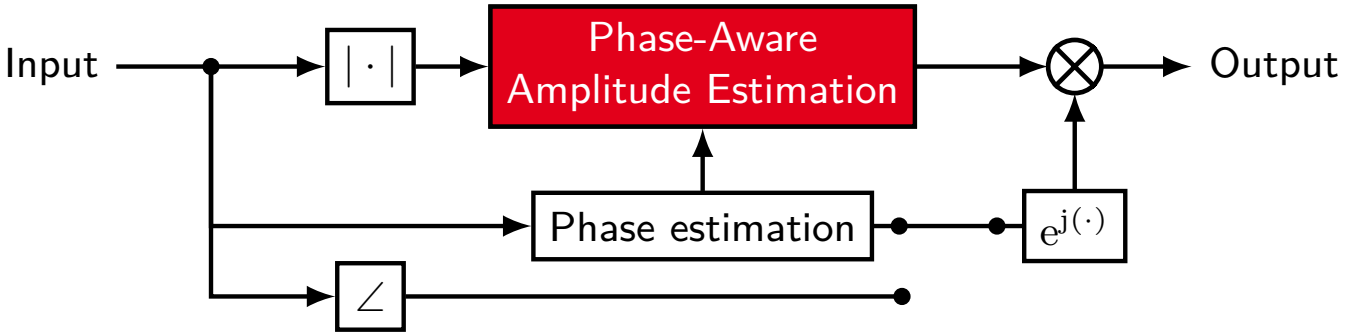
Illustrative Example



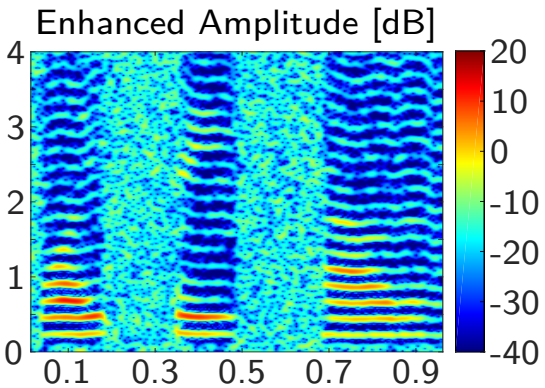
- Speech amplitude averaged over 1 kHz
- Distorted by a noise burst
- Noise is hardly suppressed by conventional phase-blind estimators
- Phase difference yields additional info

➡ New means to distinguish noise outliers from speech

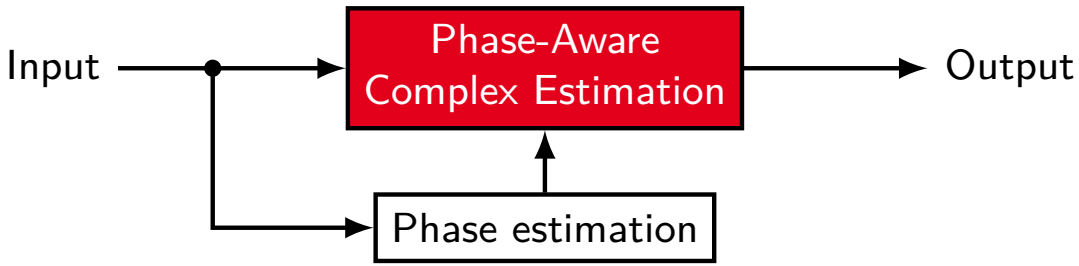
Phase Estimate for Signal Reconstruction?



- Recall: Using the phase estimate may result in harmonic artifacts!
- While combining with enhanced magnitudes reduces the effect, in stationary broadband noise, these artifacts remain audible



Phase-Aware Complex Estimation



Challenge: Use phase estimate also for signal reconstruction.

Idea: Joint estimation of amplitude and phase with **uncertain** phase prior

$$\widehat{S}^{(\beta)} = \mathrm{E}\left(A^{\beta} \mathrm{e}^{\mathrm{j}\Phi_S} \mid y, \widetilde{\Phi_S} \right)$$

- Main difference to existing phase-blind estimators:
  - ➜ Choice of the **phase prior** $p\left(\Phi_S | \widetilde{\Phi_S}\right)$

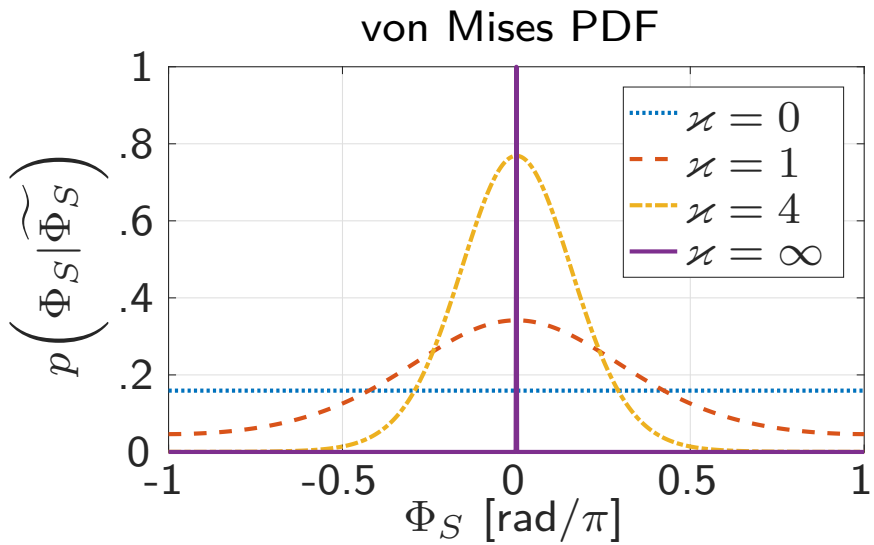
Choice of Phase Prior

- We model the phase prior with a flexible **von Mises distribution**:

$$p\left(\Phi_S|\widetilde{\Phi_S}\right) = \mathcal{M}\left(\widetilde{\Phi_S}, \varkappa\right)$$

$\widetilde{\Phi_S}$: mean direction
$\varkappa$: concentration around $\widetilde{\Phi_S}$

- $\varkappa$ is used to **model the uncertainty** in the initial phase estimate $\widetilde{\Phi_S}$



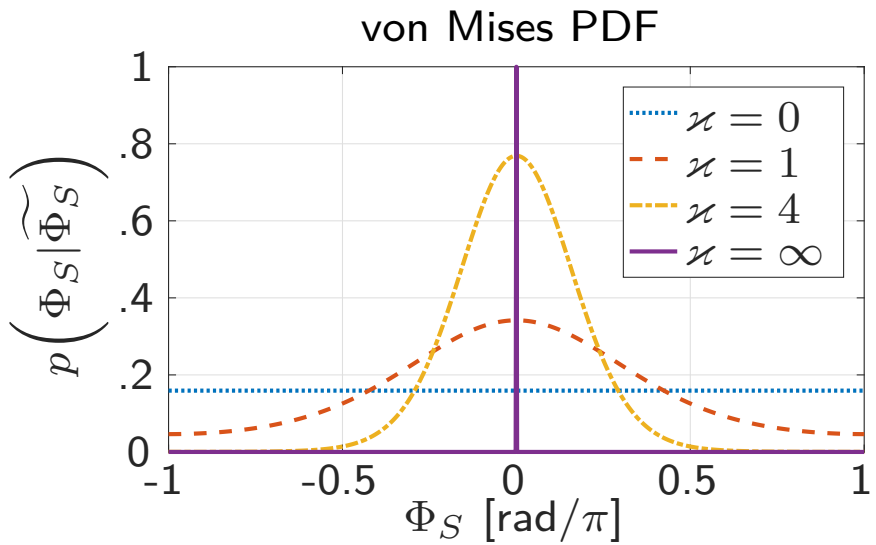
Choice of Phase Prior

- We model the phase prior with a flexible **von Mises distribution**:

$$p\left(\Phi_S|\widetilde{\Phi_S}\right) = \mathcal{M}\left(\widetilde{\Phi_S}, \varkappa\right)$$

$\widetilde{\Phi_S}$: mean direction
$\varkappa$: concentration around $\widetilde{\Phi_S}$

- $\varkappa$ is used to **model the uncertainty** in the initial phase estimate $\widetilde{\Phi_S}$

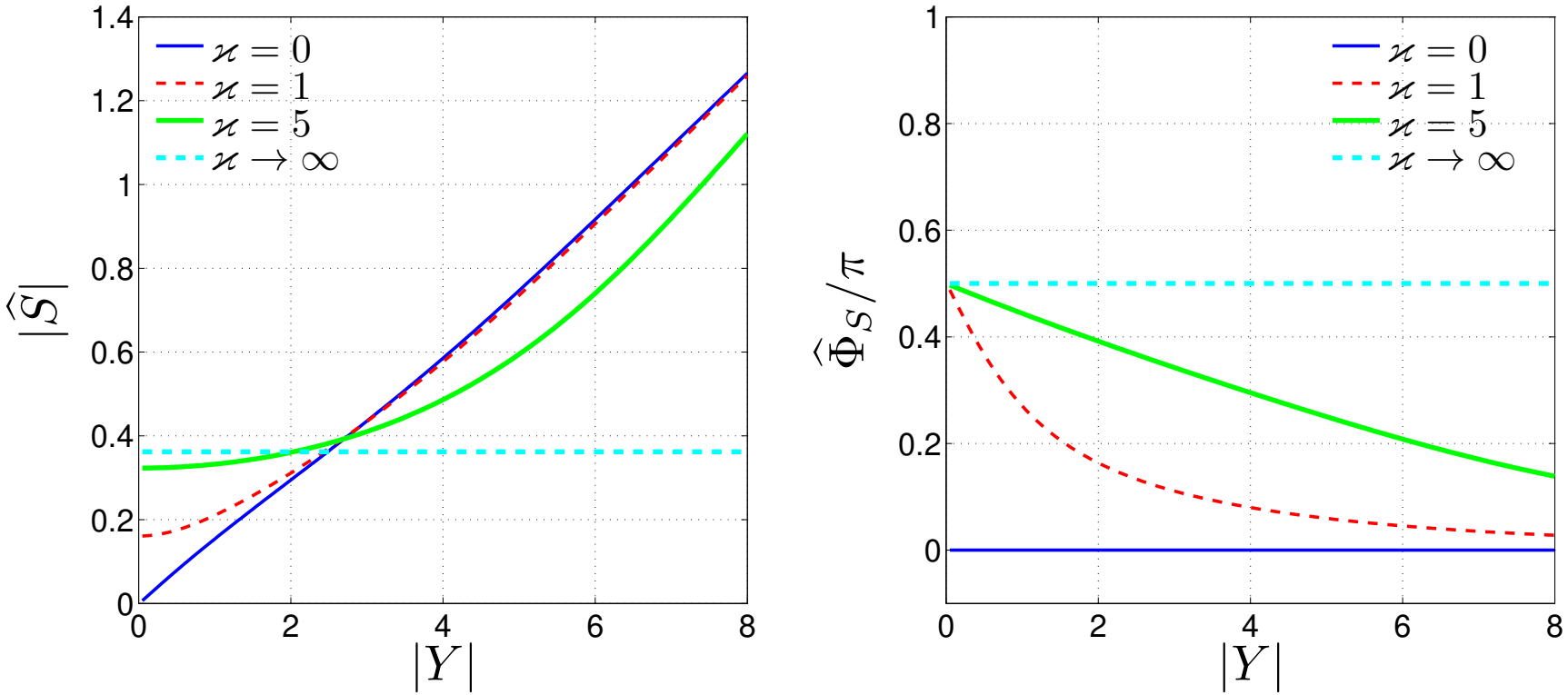


- $\varkappa = 0$: Uniform phase prior
  $\Rightarrow$ Phase-blind estimators
- $\varkappa > 0$: Uncertainty in $\widetilde{\Phi_S}$
- $\varkappa \to \infty$: Assumes that the prior phase is exactly the true clean speech phase $\widetilde{\Phi_S} = \Phi_S$

Phase-Aware Complex Estimation

Input-Output Characteristic at low SNRs[9]

Amplitude and phase responses for $\Phi_Y = 0$ and $\widetilde{\phi}_S = \pi/2$



[9] T. Gerkmann. "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase". In: *IEEE Trans. Signal Process.* 62.16 (2014), pp. 4199–4208.

Estimators for Different Choices of Parameters

	Complex estimators	*Amplitude estimators*
phase-blind $\varkappa = 0$	[10]	PBA[11]
uncertain phase $0 < \varkappa < \infty$	PAC[9]	PAA[10]
certain phase $\varkappa \to \infty$	PAC[8]	PAA[8]

- PBA: Phase-blind amplitude
- PAC: Phase-aware complex
- PAA: Phase-aware amplitude

[10] M. Krawczyk-Becker and T. Gerkmann. "On MMSE-Based Estimation of Spectral Speech Coefficients Under Phase-Uncertainty". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.12 (2016), pp. 2251–2262.

[11] C. Breithaupt, M. Krawczyk, and R. Martin. "Parameterized MMSE Spectral Magnitude Estimation for the Enhancement of Noisy Speech". In: *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. Las Vegas, NV, USA, 2008, pp. 4037–4040.

[9] T. Gerkmann. "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase". In: *IEEE Trans. Signal Process.* 62.16 (2014), pp. 4199–4208.

[10] M. Krawczyk-Becker and T. Gerkmann. "On MMSE-Based Estimation of Spectral Speech Coefficients Under Phase-Uncertainty". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.12 (2016), pp. 2251–2262.

[8] T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

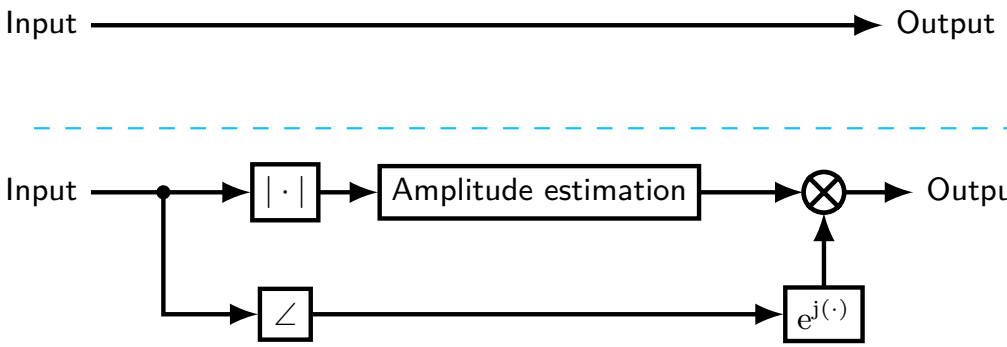
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

SP
Signal Processing

Evaluation

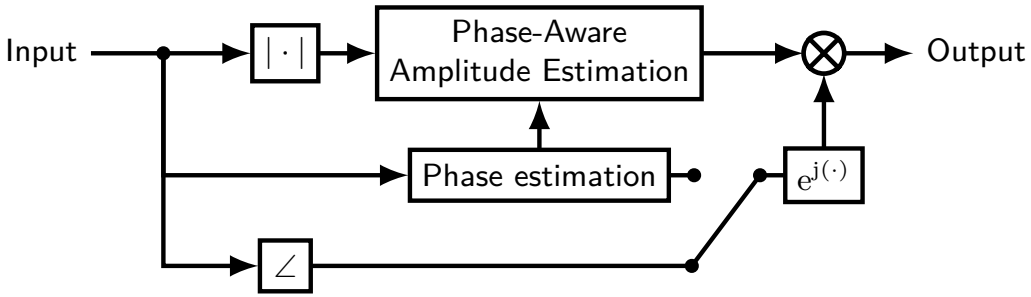
Evaluation

Algorithm Overview

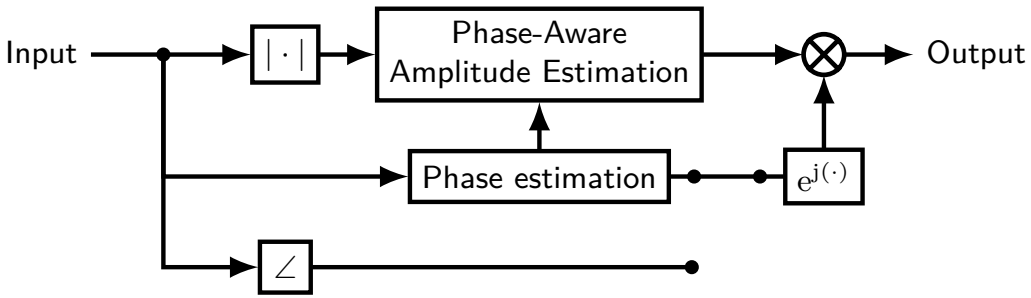


Noisy Clean Reference

PBA
Phase-**B**lind **A**mpl. est.



PAA oracle / estimated
Phase-**A**ware **A**mpl. est.



PAC oracle / estimated
Phase-**A**ware **C**omplex est.

Instrumental Evaluation

Different Degrees of Non-Stationarity

- Speech: Gender-balanced TIMIT core-set at 16 kHz
- Speech PSD σ_{S}^2 estimation via the decision-directed approach[2]
- Noise PSD σ_{V}^2 estimation based on speech presence probability[12]
- Two noise types with increasing amount of non-stationarity
  1. Pink noise modulated with an increasing modulation frequency
  2. Multi-talker babble with decreasing number of talkers

➡ Increasingly challenging for phase-blind speech enhancement

---

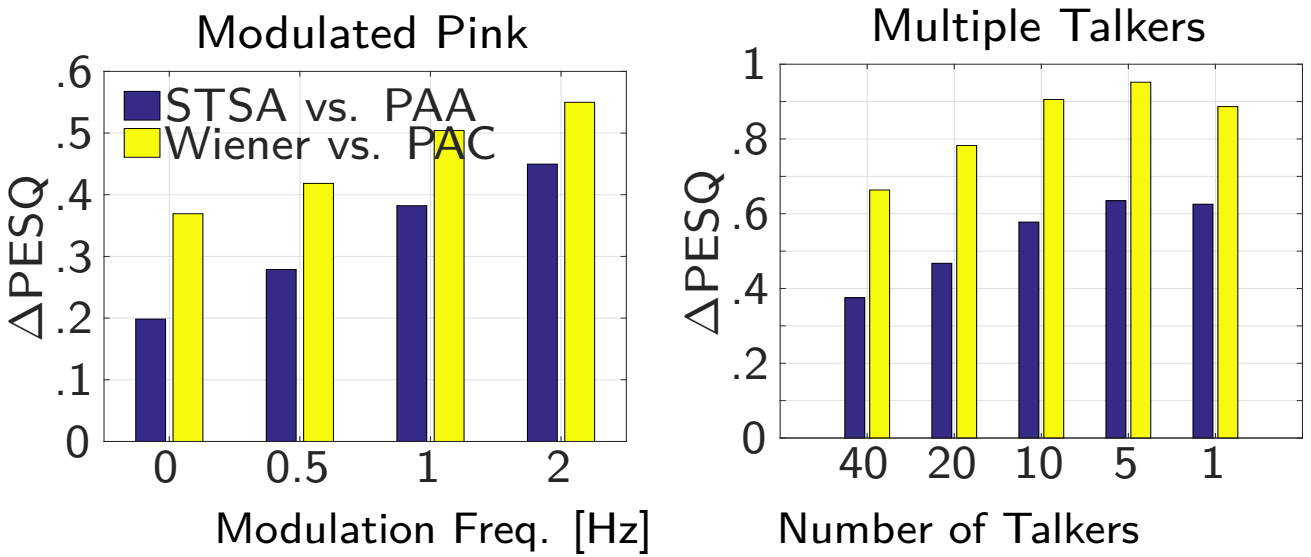
[2] Y. Ephraim and D. Malah. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.6 (1984), pp. 1109–1121.

[12] T. Gerkmann and R. C. Hendriks. "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay". In: *IEEE Trans. Audio, Speech, Language Process.* 20.4 (2012), pp. 1383–1393.

Instrumental Evaluation

Different Degrees of Non-Stationarity, 0 dB SNR, $\varkappa \to \infty$

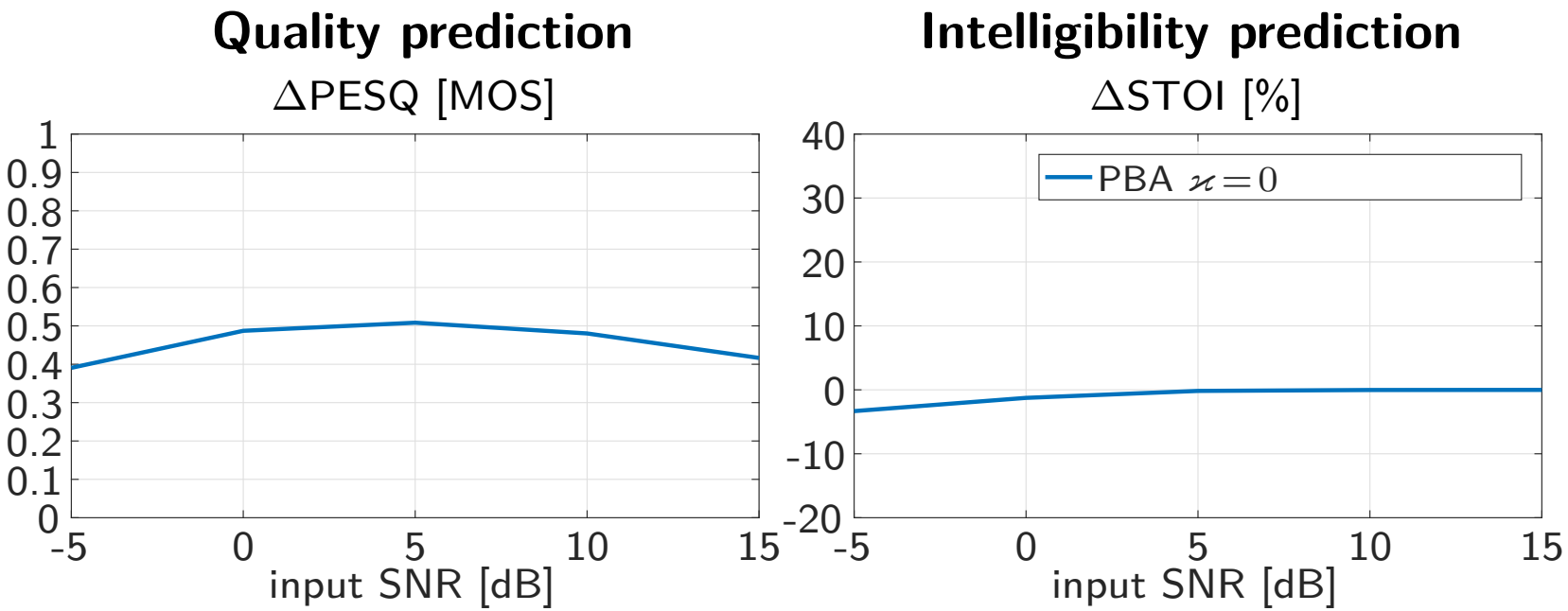
Clean speech phase Φ_S is known



- Increasing non-stationarity ➡ increasing benefit of phase-aware processing

Instrumental Evaluation

Known Clean Speech Phase

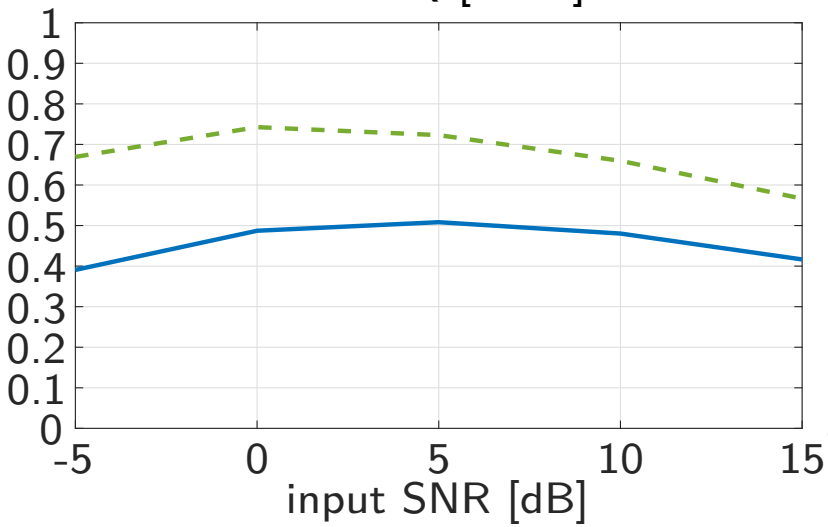


- 128 TIMIT sentences (gender balanced)
- Averaged over 4 noise types (pink, modulated pink, factory, babble)

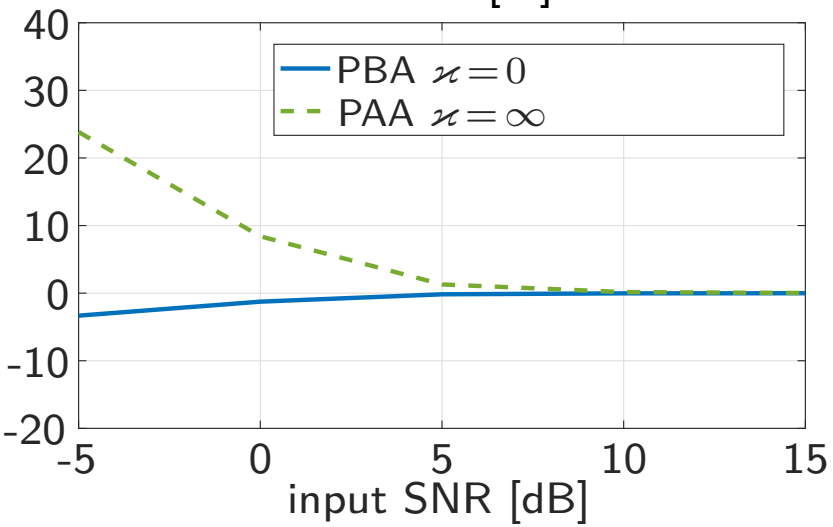
Instrumental Evaluation

Known Clean Speech Phase

Quality prediction



Intelligibility prediction

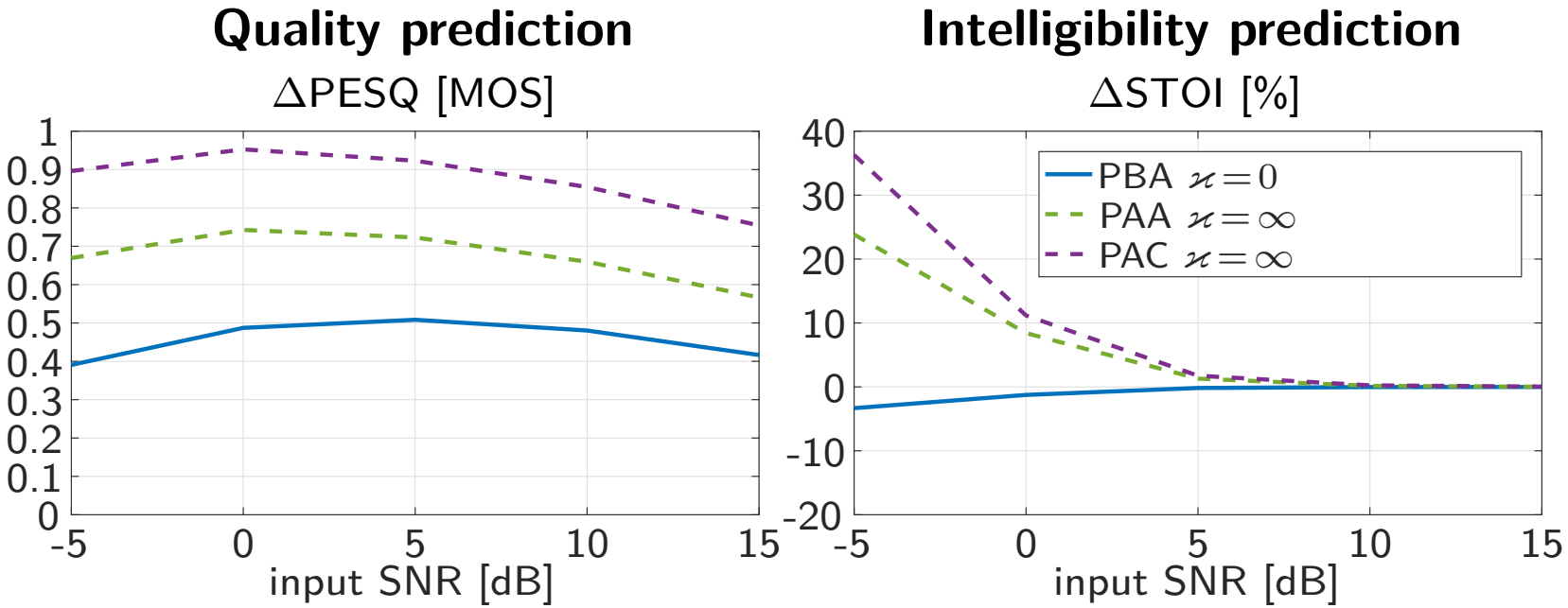


- 128 TIMIT sentences (gender balanced)
- Averaged over 4 noise types (pink, modulated pink, factory, babble)

➡ Phase-aware amplitude estimator outperforms phase-blind counterparts

Instrumental Evaluation

Known Clean Speech Phase



- 128 TIMIT sentences (gender balanced)
- Averaged over 4 noise types (pink, modulated pink, factory, babble)

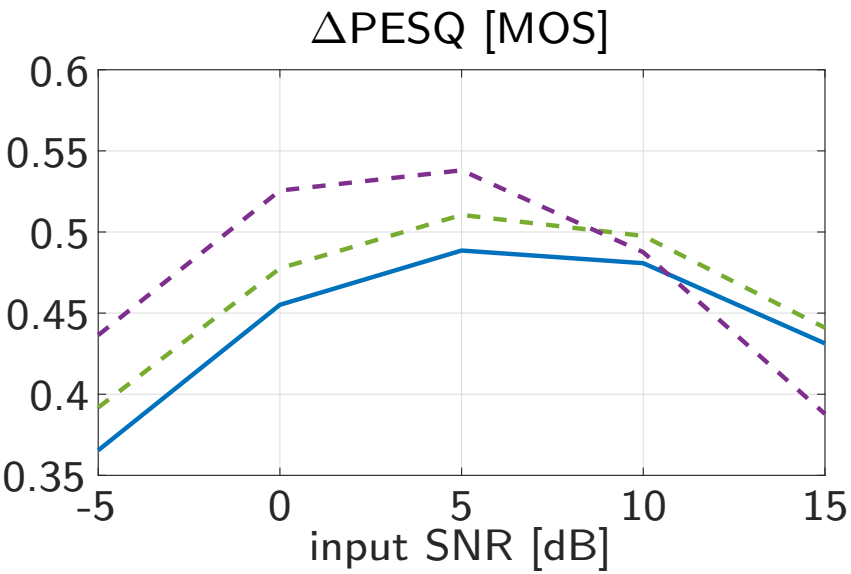
➡ Phase-aware amplitude estimator outperforms phase-blind counterparts

➡ The phase-aware complex estimator yields further improvements

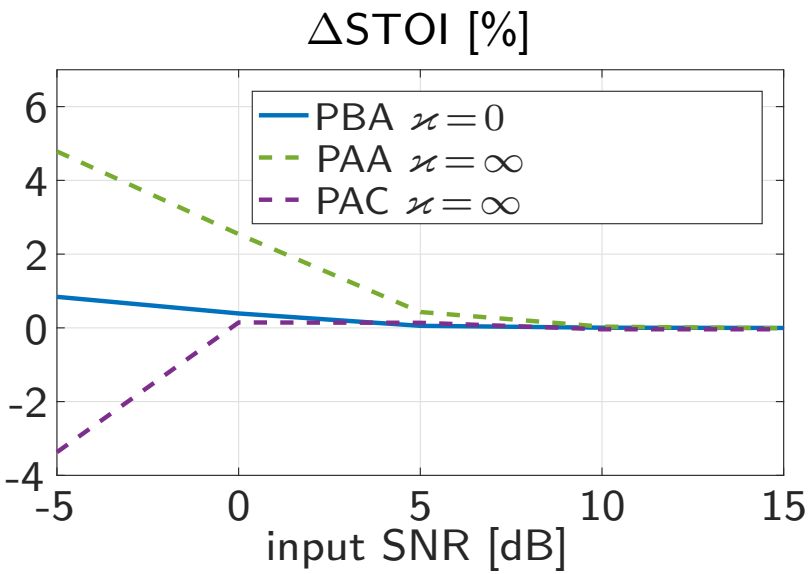
Instrumental Evaluation

Estimated Clean Speech Phase

Quality prediction



Intelligibility prediction



- The initial phase $\widetilde{\Phi_S}$ is estimated from the noisy signal

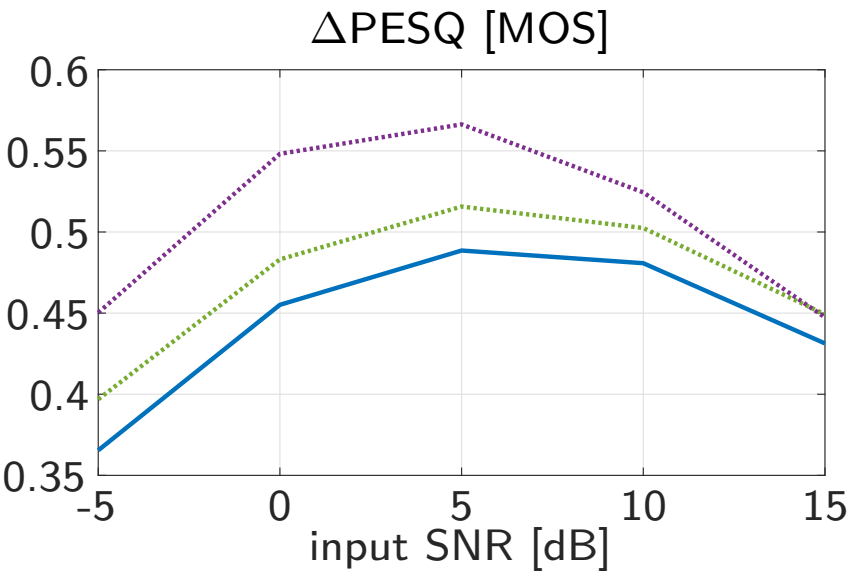
➡ Highest PESQ scores for the phase-aware complex estimator

➡ Highest STOI scores for the phase-aware amplitude estimator

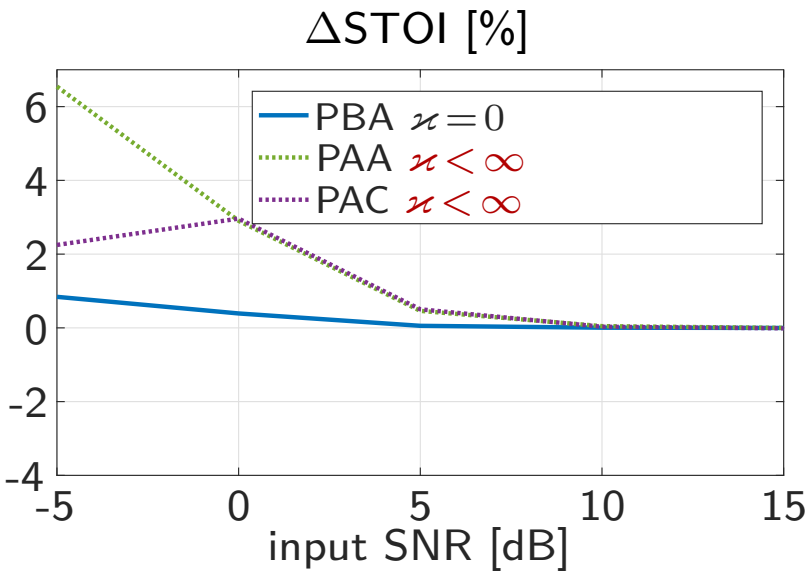
Instrumental Evaluation

Estimated Clean Speech Phase

Quality prediction



Intelligibility prediction



- The initial phase $\widetilde{\Phi_S}$ is estimated from the noisy signal

➡ Highest PESQ scores for the phase-aware complex estimator

➡ Highest STOI scores for the phase-aware amplitude estimator

➡ Considering the uncertainty $\varkappa$ yields further improvements

Perceptual Evaluation

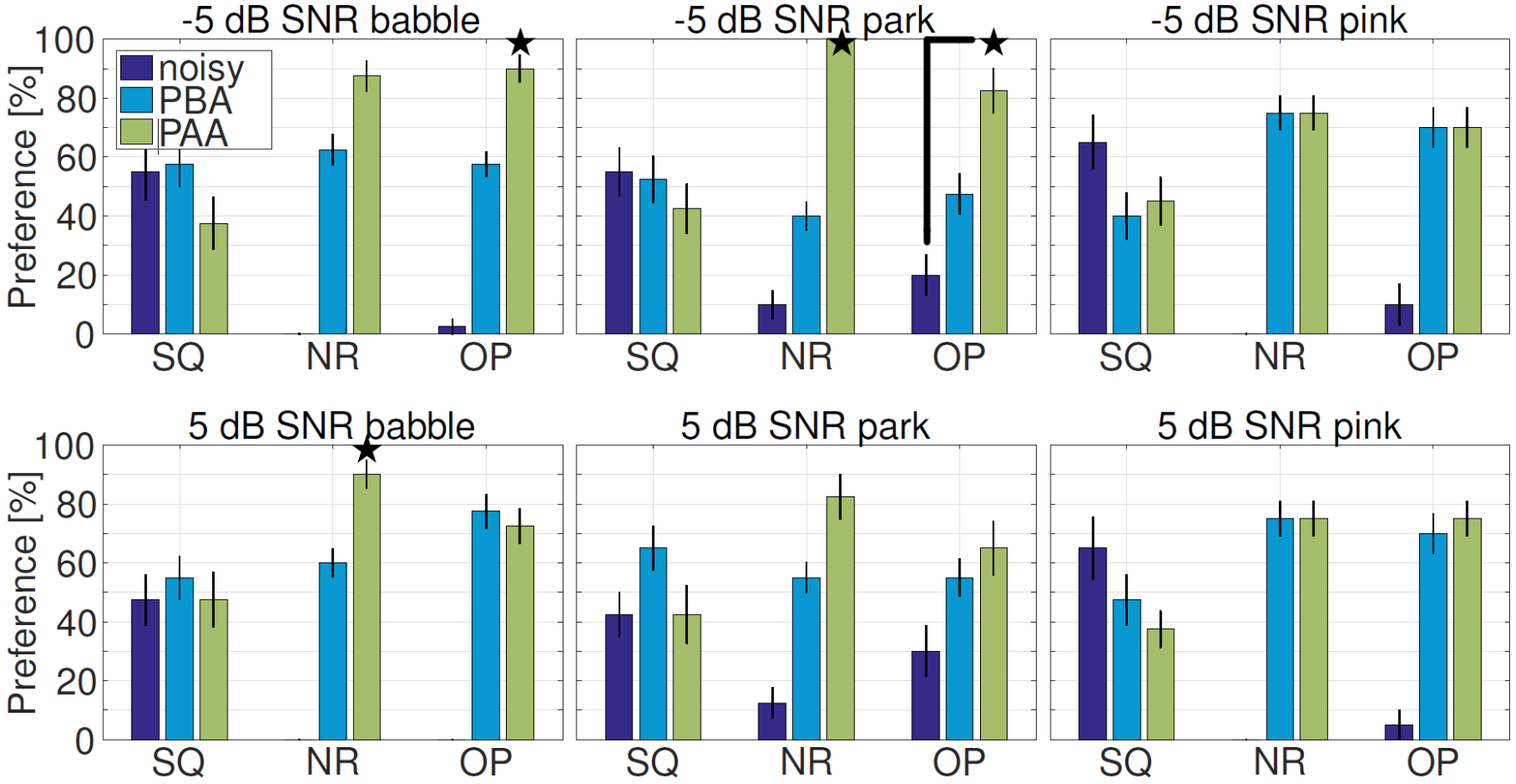
Experimental Setup

- Pairwise comparison test separately for
  - Speech quality (SQ)
  - Noise reduction (NR)
  - Overall preference (OP)
- Clean speech sentences (TIMIT database) mixed with
  - babble noise with hammering (−/+5 dB)
  - walk in the park (−/+5 dB)
  - stationary pink noise (−/+5 dB)
- Clean speech was always provided as a reference
- Diotic headphone presentation at 65 dB SPL
- 20 self-reported normal hearing listeners

Preference Scores

Estimated Speech Phase

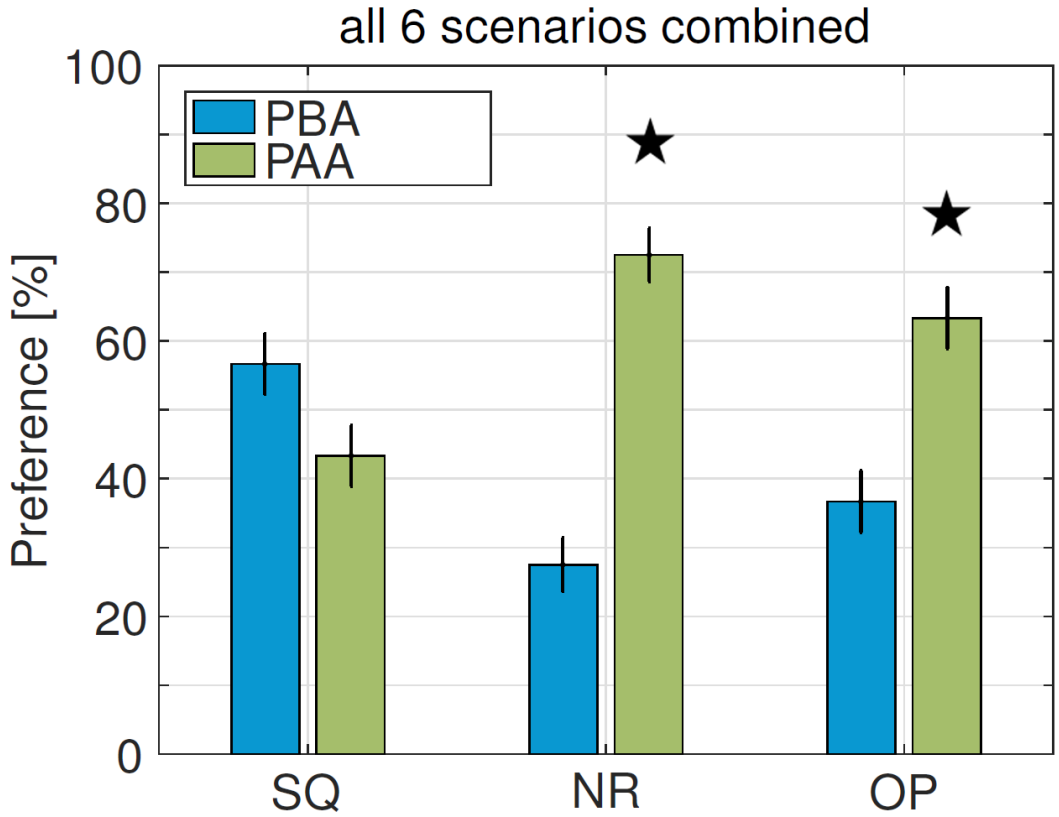
(SQ) speech quality – (NR) noise reduction – (OP) overall preference



- Most gain in challenging acoustic scenarios

Preference Scores[13]

Estimated Speech Phase



- Only head-to-head comparisons of PBA and PAA
- Statistically significant preference in noise reduction and overall

[13] M. Krawczyk-Becker and T. Gerkmann. "An evaluation of the perceptual quality of phase-aware single-channel speech enhancement". In: *J. Acoust. Soc. Amer.* 140.4 (2016), EL364–EL369.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

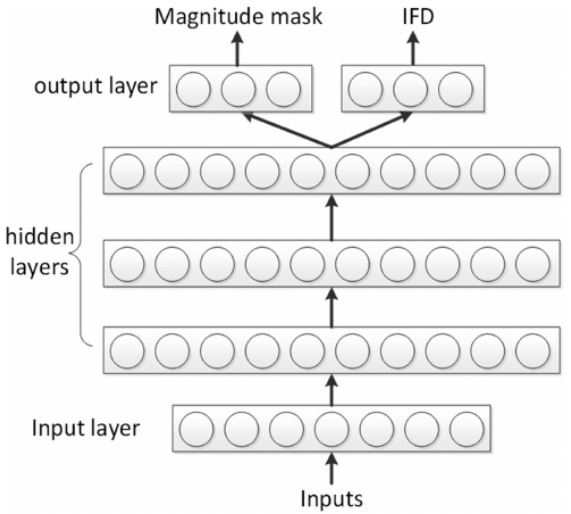
SP
Signal Processing

Outlook and Conclusions

Machine Learning (ML) for Phase-Aware Proc.

- In the last 1-2 years, we observe a drastic increase on ML-based phase-aware processing
- Again, avoiding phase wrapping is crucial
- Phase and amplitude can be jointly estimated in a multi-objective DNN

➡ Exciting and challenging ML-problem



N. Zheng and X.-L. Zhang. "Phase-Aware Speech Enhancement Based on Deep Neural Networks". In: *IEEE Trans. Audio, Speech, Language Process.* 27.1 (2019), pp. 63–76

D. S. Williamson. "Monaural Speech Separation Using a Phase-Aware Deep Denoising Auto Encoder". In: *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*. Aalborg, Denmark, 2018, pp. 1–6

Z.-Q. Wang et al. "End-to-End Speech Separation with Unfolded Iterative Phase Reconstruction". In: *Proc. Interspeech 2018*. Hyderabad, India, 2018, pp. 2708–2712

K. Oyamada et al. "Generative adversarial network-based approach to signal reconstruction from magnitude spectrogram". In: *2018 26th European Signal Processing Conference (EUSIPCO)*. Rome, Italy, 2018, pp. 2514–2518

Conclusions

Is phase important?

Conclusions

Is phase important?

- It yields valuable information that can be used to
  - ➜ Distinguish speech from noise
  - ➜ Benefit speech enhancement in challenging acoustic scenarios

Conclusions

Is phase important?

- It yields valuable information that can be used to
  - ➜ Distinguish speech from noise
  - ➜ Benefit speech enhancement in challenging acoustic scenarios

And it remains an interesting research topic:

- Phase estimation in unvoiced sounds
- Research on phase-aware machine learning based algorithms

References

A. Sugiyama and R. Miyahara. "Phase randomization – A new paradigm for single-channel signal enhancement". In: *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*. Vancouver, Canada, 2013, pp. 7487–7491.

Y. Ephraim and D. Malah. "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.6 (1984), pp. 1109–1121.

T. Lotter and P. Vary. "Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model". In: *EURASIP J. Applied Signal Process.* 2005.7 (2005), pp. 1110–1126.

P. Vary. "Noise Suppression By Spectral Magnitude Estimation – Mechanism and Theoretical Limits". In: *ELSEVIER Signal Process.* 8 (1985), 387–400.

T. Gerkmann, M. Krawczyk-Becker, and J. Le Roux. "Phase Processing for Single Channel Speech Enhancement: History and Recent Advances". In: *IEEE Signal Process. Mag.* 32.2 (2015), pp. 55–66.

D. W. Griffin and J. S. Lim. "Signal estimation from modified short-time Fourier transform". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.2 (1984), pp. 236–243.

M. Krawczyk and T. Gerkmann. "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 22.12 (Dec. 2014), pp. 1931–1940.

T. Gerkmann and M. Krawczyk. "MMSE-Optimal Spectral Amplitude Estimation Given the STFT-Phase". In: *IEEE Signal Process. Lett.* 20.2 (2013), pp. 129–132.

T. Gerkmann. "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase". In: *IEEE Trans. Signal Process.* 62.16 (2014), pp. 4199–4208.

M. Krawczyk-Becker and T. Gerkmann. "On MMSE-Based Estimation of Spectral Speech Coefficients Under Phase-Uncertainty". In: *IEEE/ACM Trans. Audio, Speech, Language Process.* 24.12 (2016), pp. 2251–2262.

C. Breithaupt, M. Krawczyk, and R. Martin. "Parameterized MMSE Spectral Magnitude Estimation for the Enhancement of Noisy Speech". In: *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. Las Vegas, NV, USA, 2008, pp. 4037–4040.

T. Gerkmann and R. C. Hendriks. "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay". In: *IEEE Trans. Audio, Speech, Language Process.* 20.4 (2012), pp. 1383–1393.

M. Krawczyk-Becker and T. Gerkmann. "An evaluation of the perceptual quality of phase-aware single-channel speech enhancement". In: *J. Acoust. Soc. Amer.* 140.4 (2016), EL364–EL369.