Advanced Speech-Audio Processing in Mobile Phones and Hearing Aids

– Synergies and Distinctions –

Peter Vary

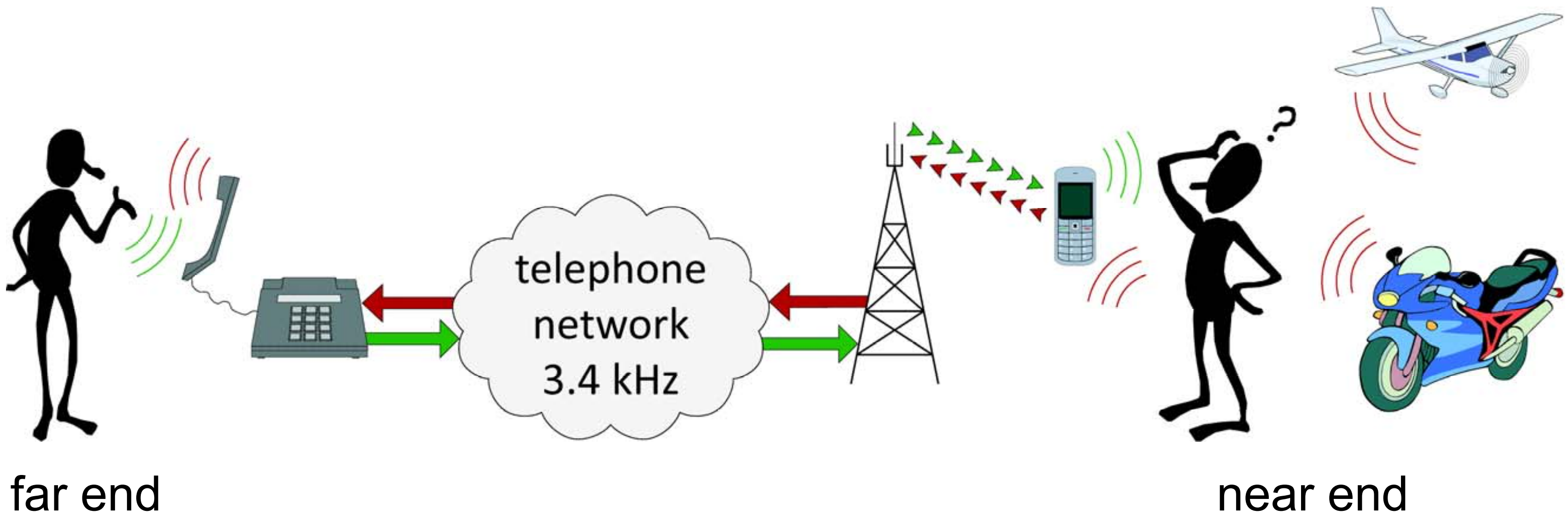
RWTH Aachen University

Institute of Communication Systems

WASPAA, October 23, 2013

Mohonk Mountain House

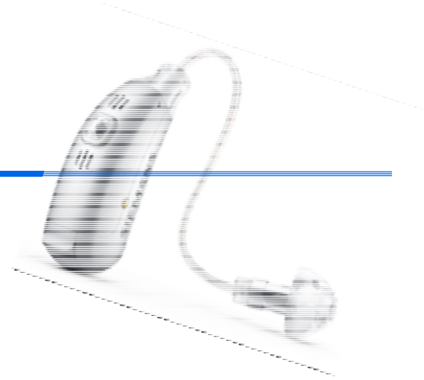
Mobile Phone in Noisy Environment



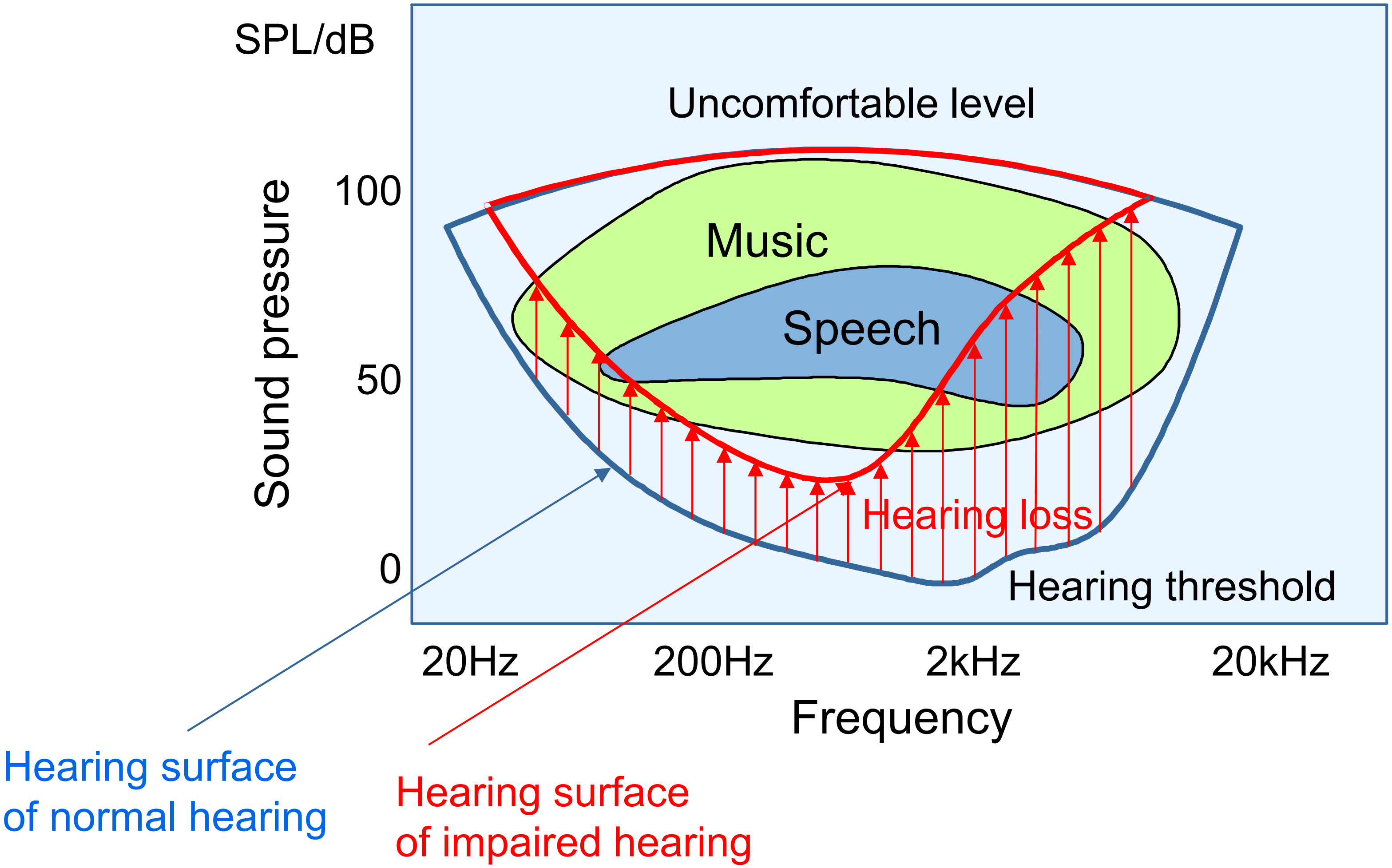
- <u>At both ends</u> of the communication link
  - increased listening effort
  - decreased intelligibility

due to 3.4 kHz frequency limitation and acoustic background noise

Hearing Aid in Noisy Environment



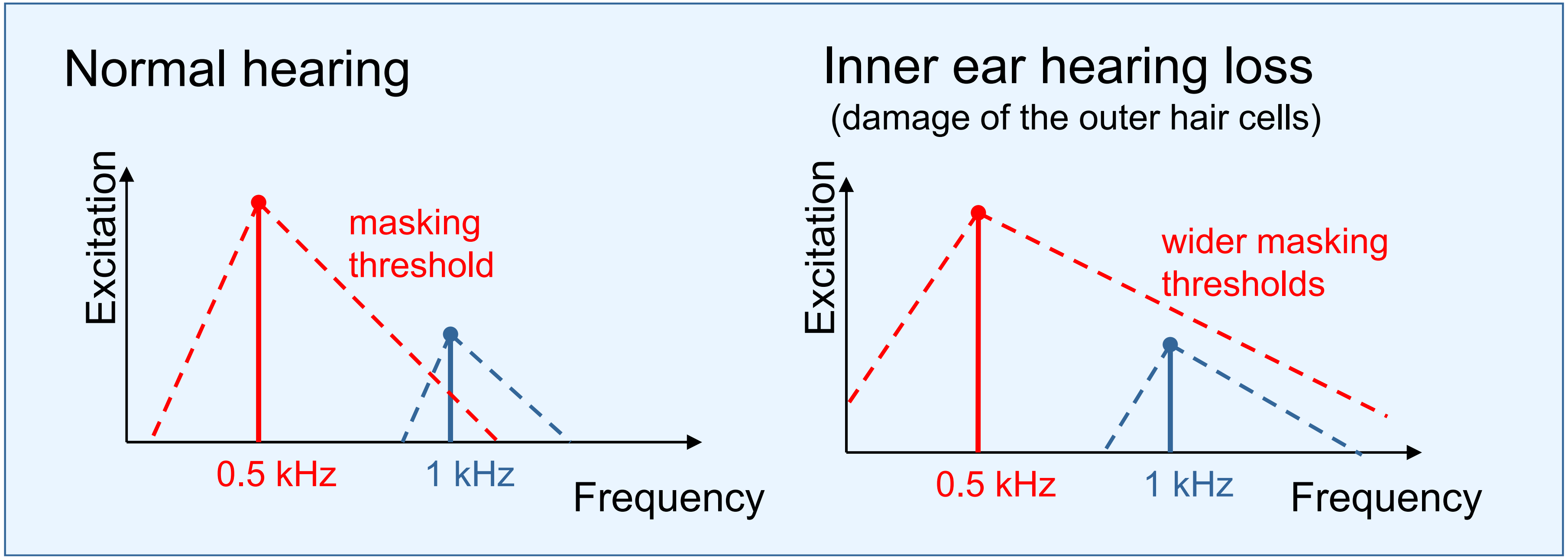
- Hearing area of normal and impaired hearing



Source: H. Puder, Siemens]

Hearing Aid in Noisy Environment

- Normal and impaired frequency resolution and masking



- Limited dynamic range (raised hearing threshold) and stronger masking
  - increased listening effort
  - decreased intelligibility

Source: H. Puder, Siemens]

Advanced Speech-Audio Processing in Mobile Phones and Hearing Aids

1. Introduction
2. Acoustical Distinctions
3. Signal Processing & Coding
4. Selected Algorithms
5. Conclusions

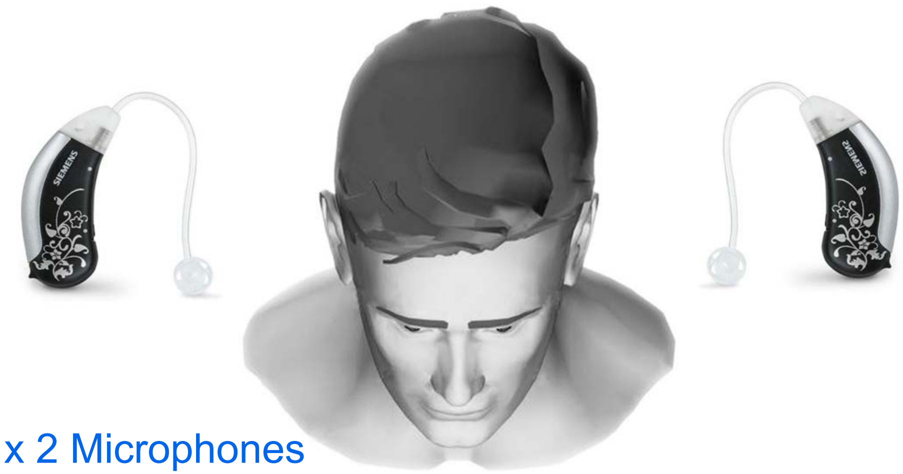
2. Acoustical Distinctions

- Dual- and multi-microphones signal processing capabilities

- Hearing aids
  - monaural

2 Microphones

  - bilateral



2 x 2 Microphones

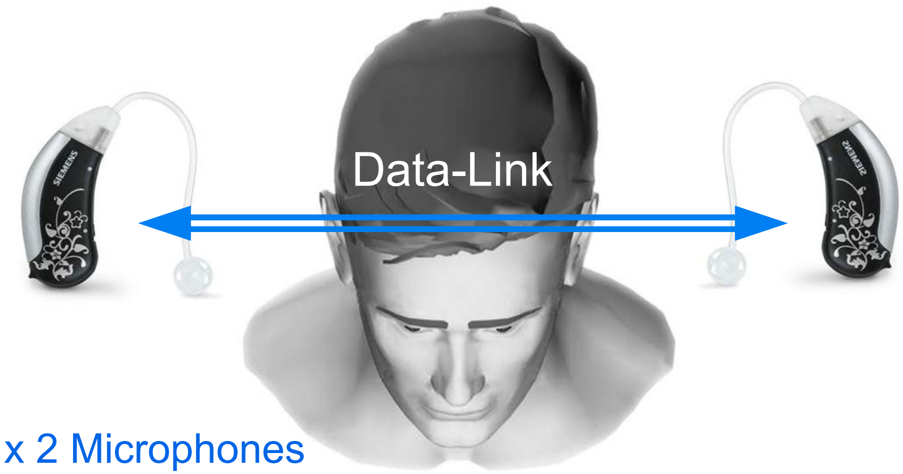
2. Acoustical Distinctions

- Dual- and multi-microphones signal processing capabilities

- Hearing aids
  - monaural

2 Microphones

  - binaural



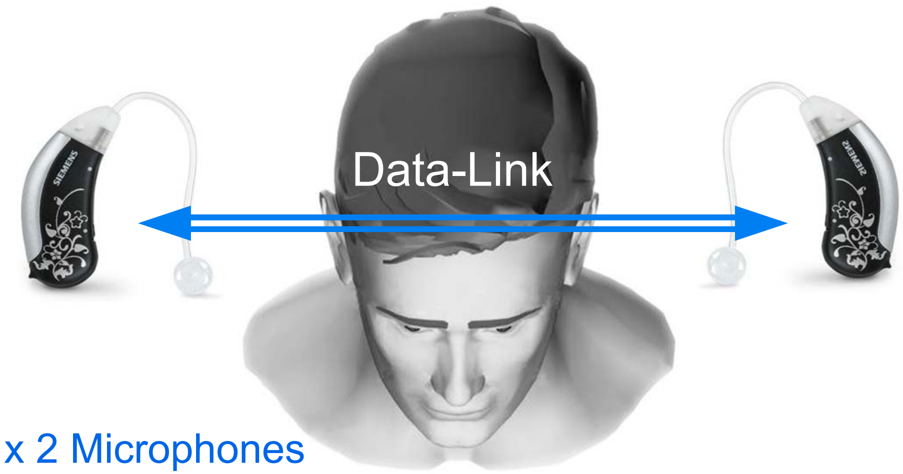
2 x 2 Microphones

2. Acoustical Distinctions

- Dual- and multi-microphones signal processing capabilities

- Hearing aids
  - monaural / bilateral

  - binaural



- Mobile phones
  - monaural

Coherence: Theory

d = 1 cm

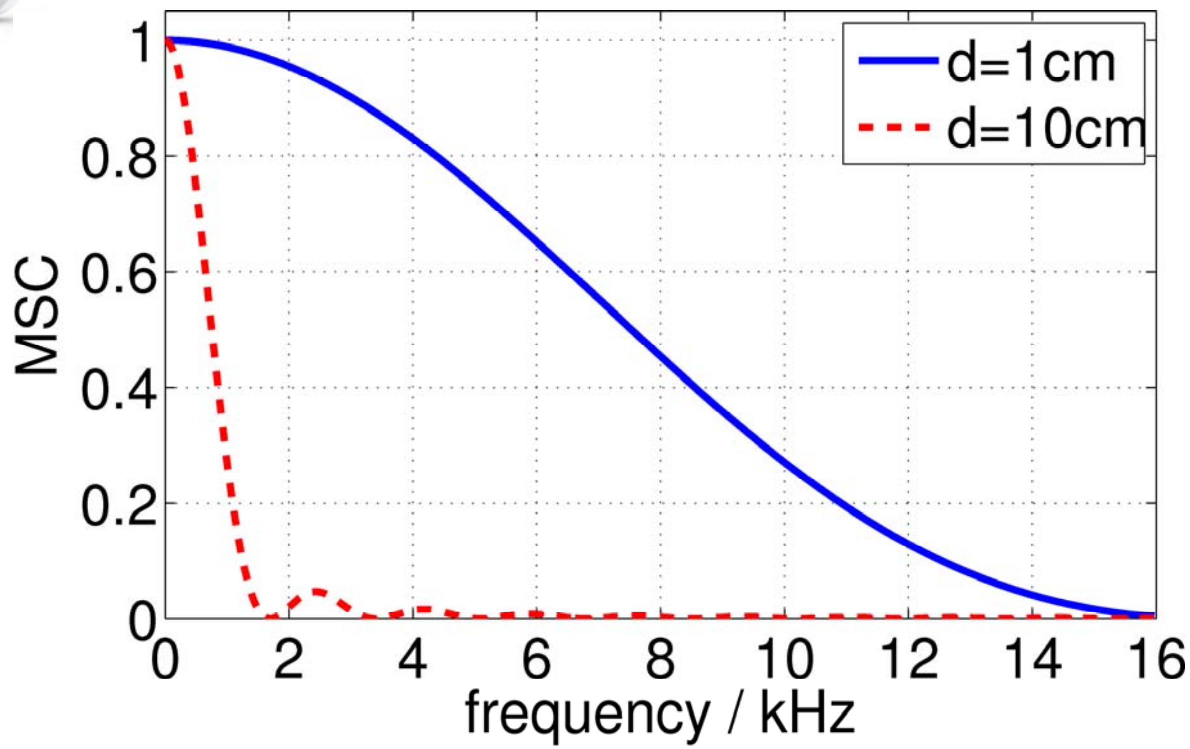
Microphones

MSC: (diffuse noise)
Magnitude Squared Coherence Function

d = 10 cm

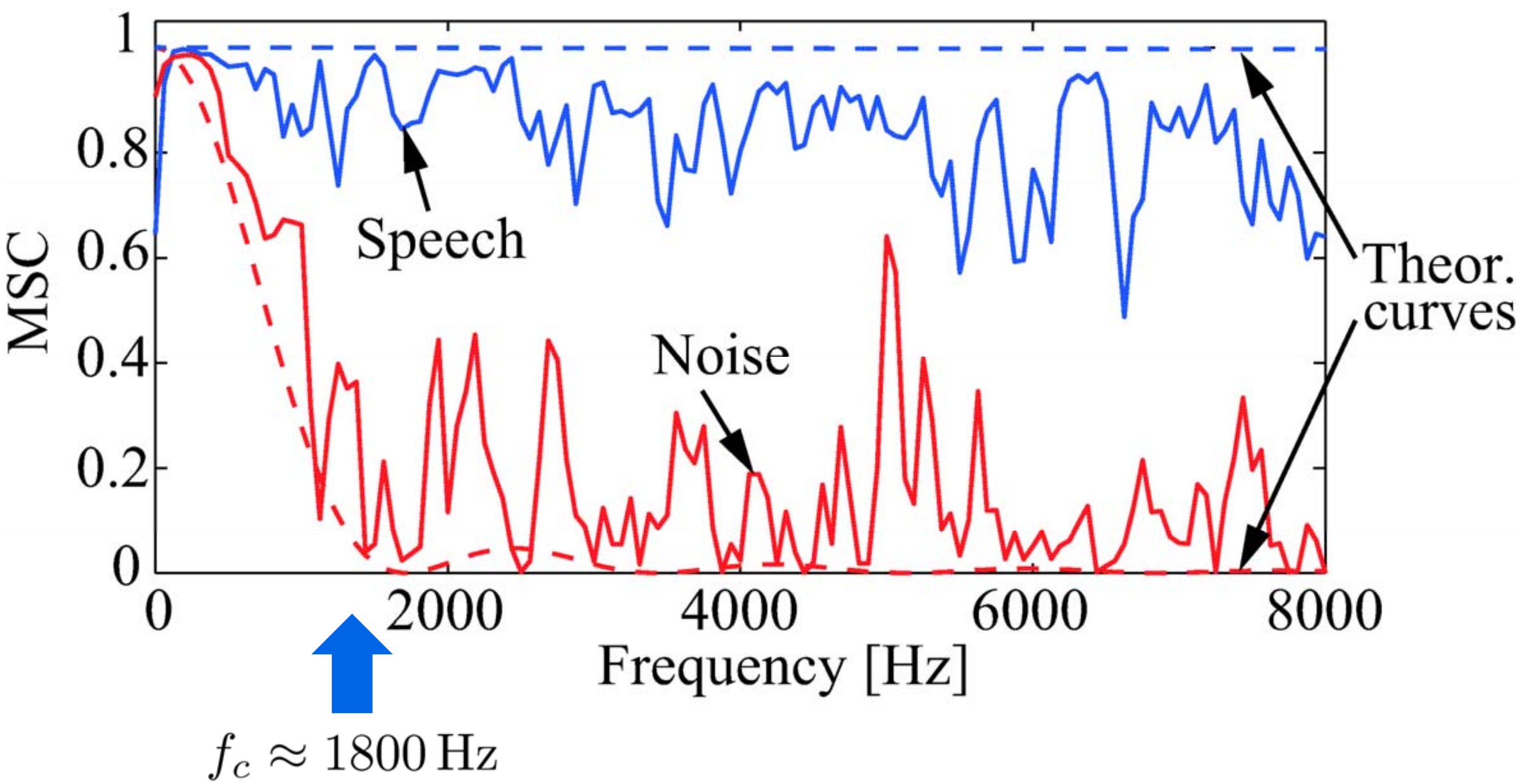
Auxiliary microphone

Primary microphone

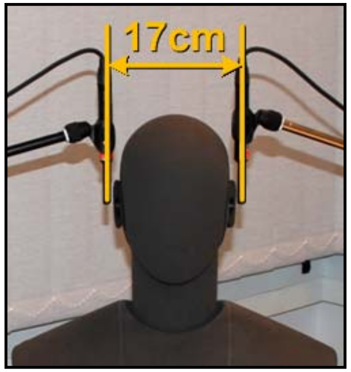


Coherence: Theory & Measurement (1)

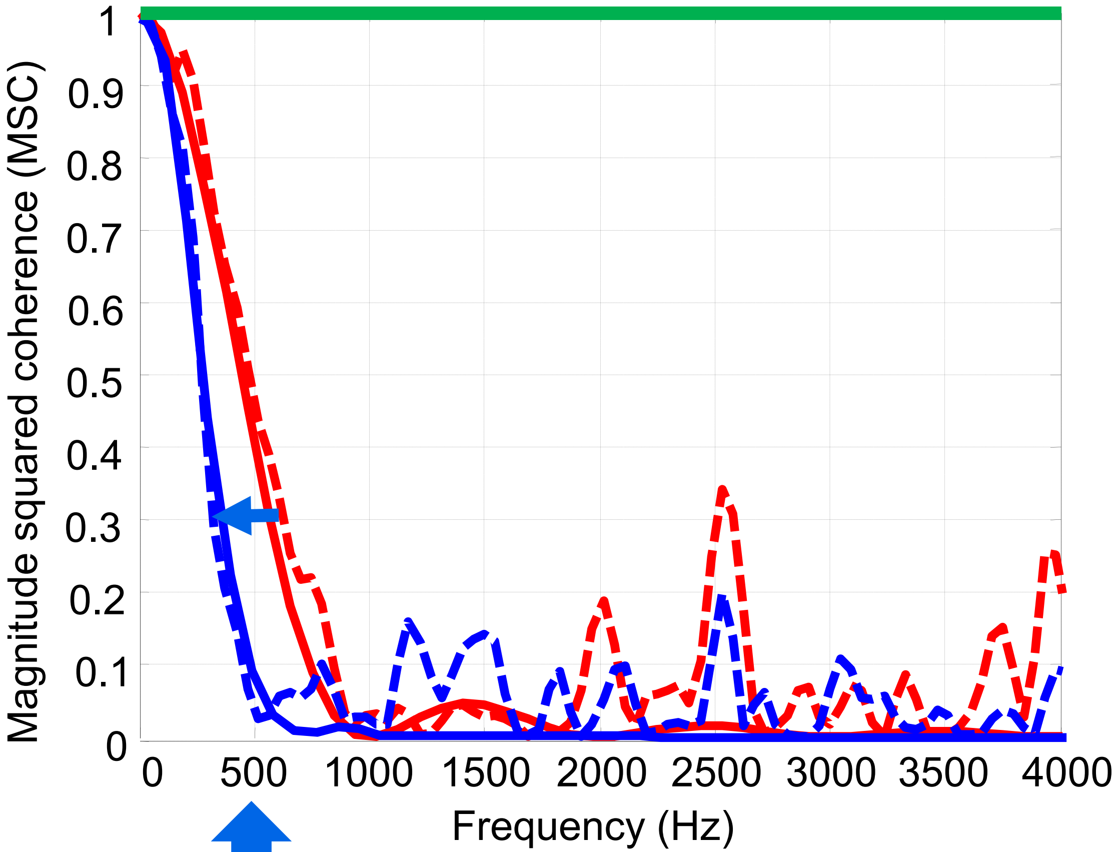
- Mobile phone in hands-free / loudspeaking mode



Coherence: Theory & Measurement (2)



- MSC for d_{mic} = 17 cm
- Head-related (binaural) noise field coherence

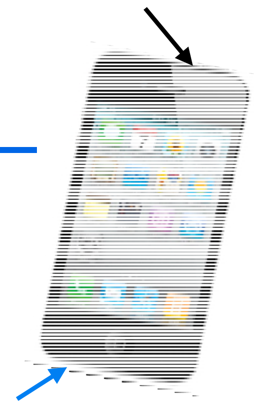


$$\Gamma^{(\text{diff})}_{x_l x_r}(f) = \text{sinc}\left(\frac{2\pi f d_{\text{mic}}}{c}\right)$$

[Dörbecker 1998], [Jeub, Dörbecker 2011]

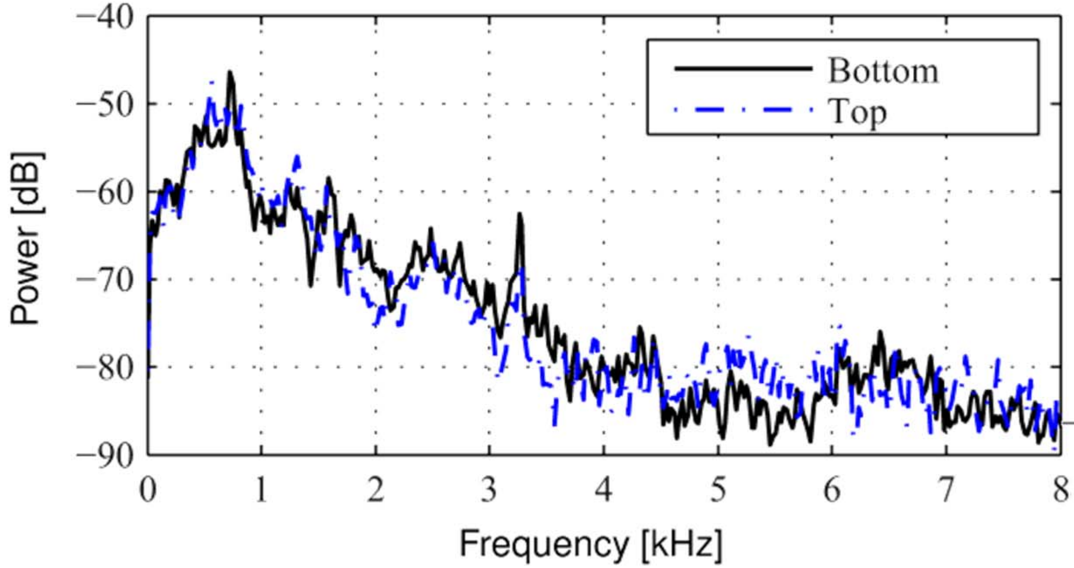
$f_c \approx 500\,\text{Hz}$

Power Level Differences (1)

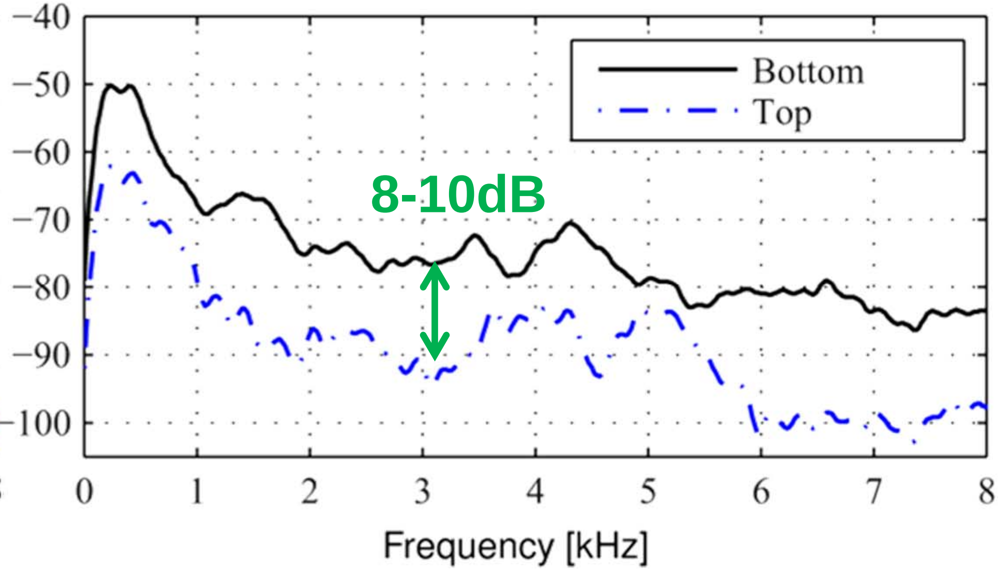


- Mobile phone in handset mode

PSD of **noise**



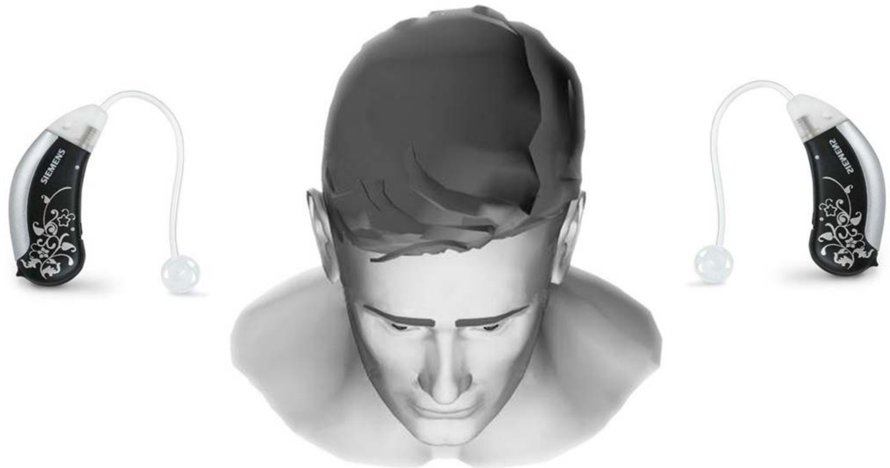
PSD of **speech**



- **Noise:** Diffuse field coherence and small power differences
- **Speech:** Large power differences between the microphones

Power Level Differences (2)

- Hearing aids (bilateral & binaural)



- **Two cases**
  - bilateral: 2 microphones with distance of 1cm at each ear
  - binaural: 1 differential microphone at each ear
- **In both cases**

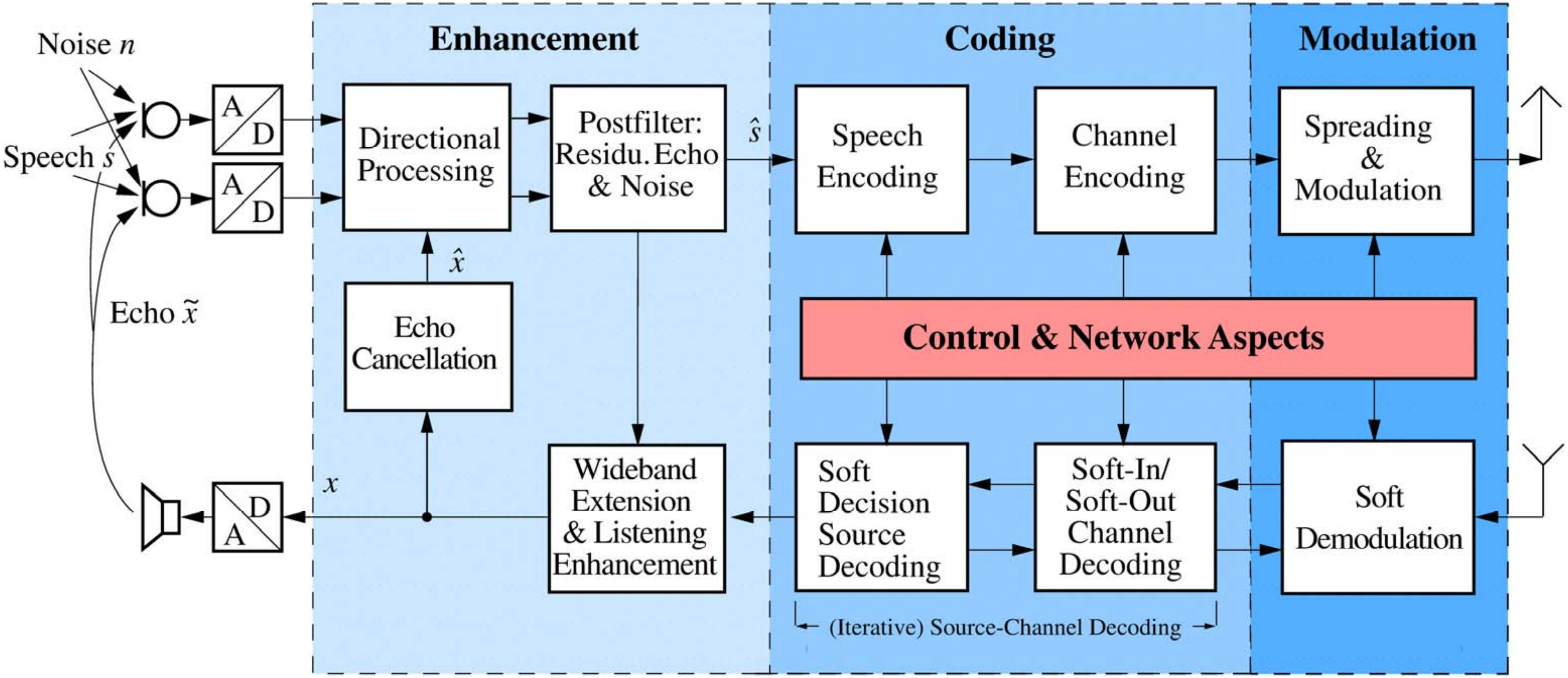
  small power level differences for frontal speech <u>and</u> diffuse noise

Advanced Speech-Audio Processing in Mobile Phones and Hearing Aids

1. Introduction
2. Acoustical Distinctions
3. Signal Processing & Coding
4. Selected Algorithms
5. Conclusions

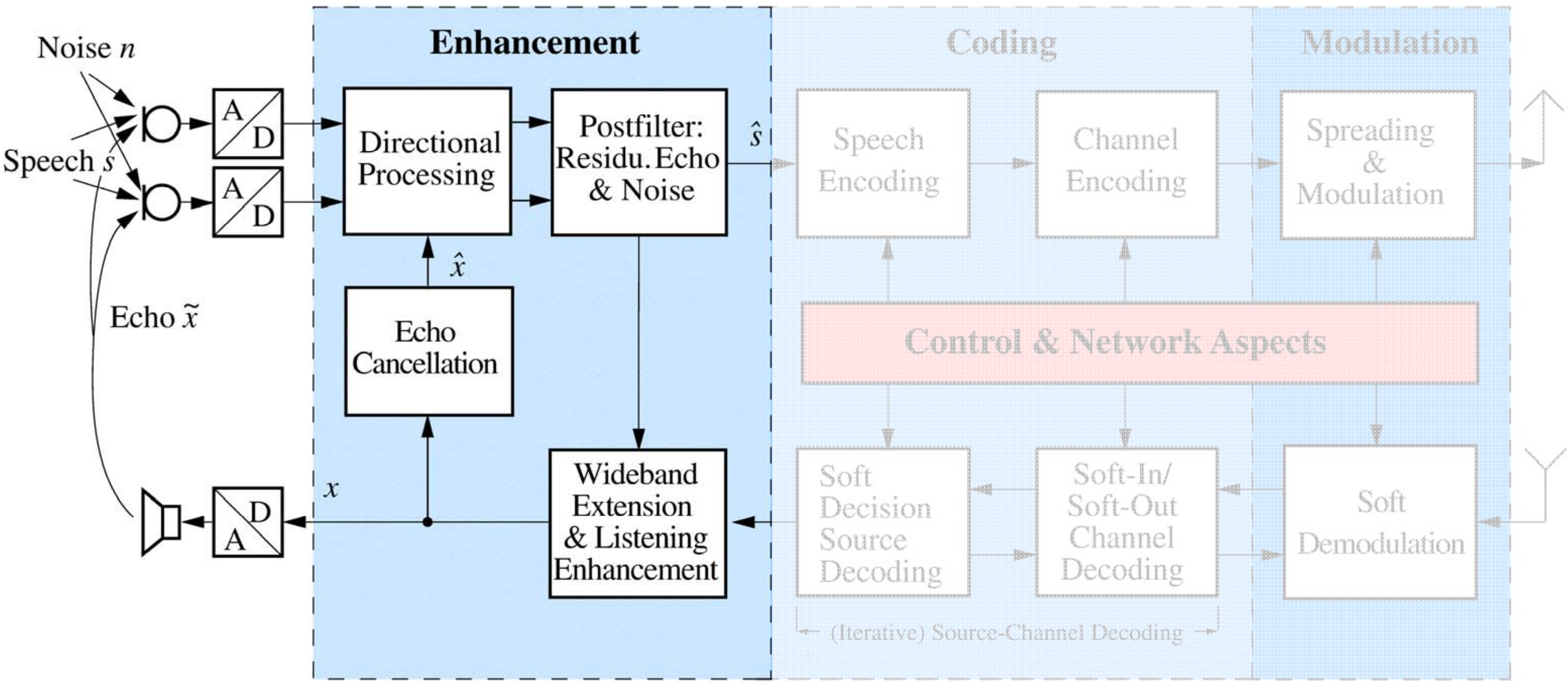
Mobile Phone

- Enhancement, coding & modulation

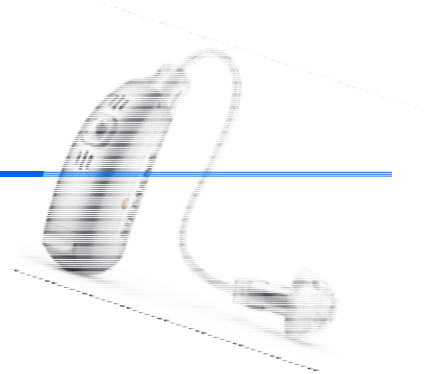


Mobile Phone

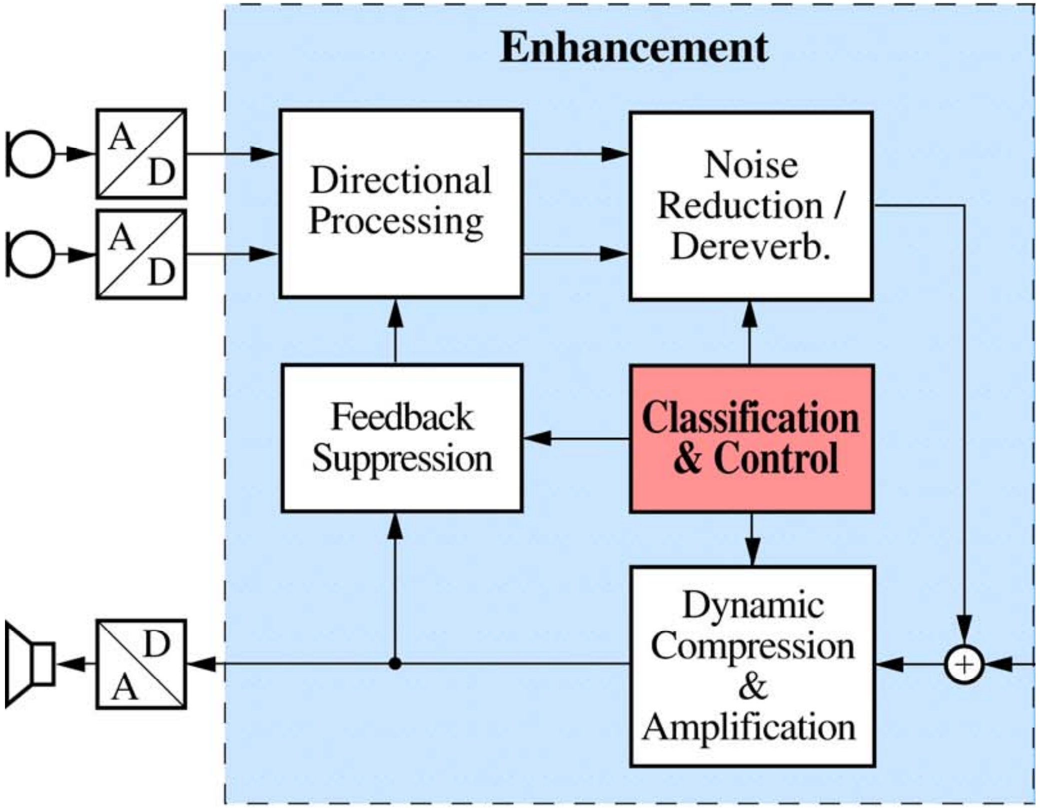
- Enhancement



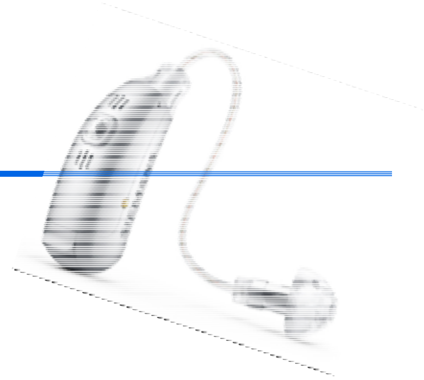
Hearing Aid



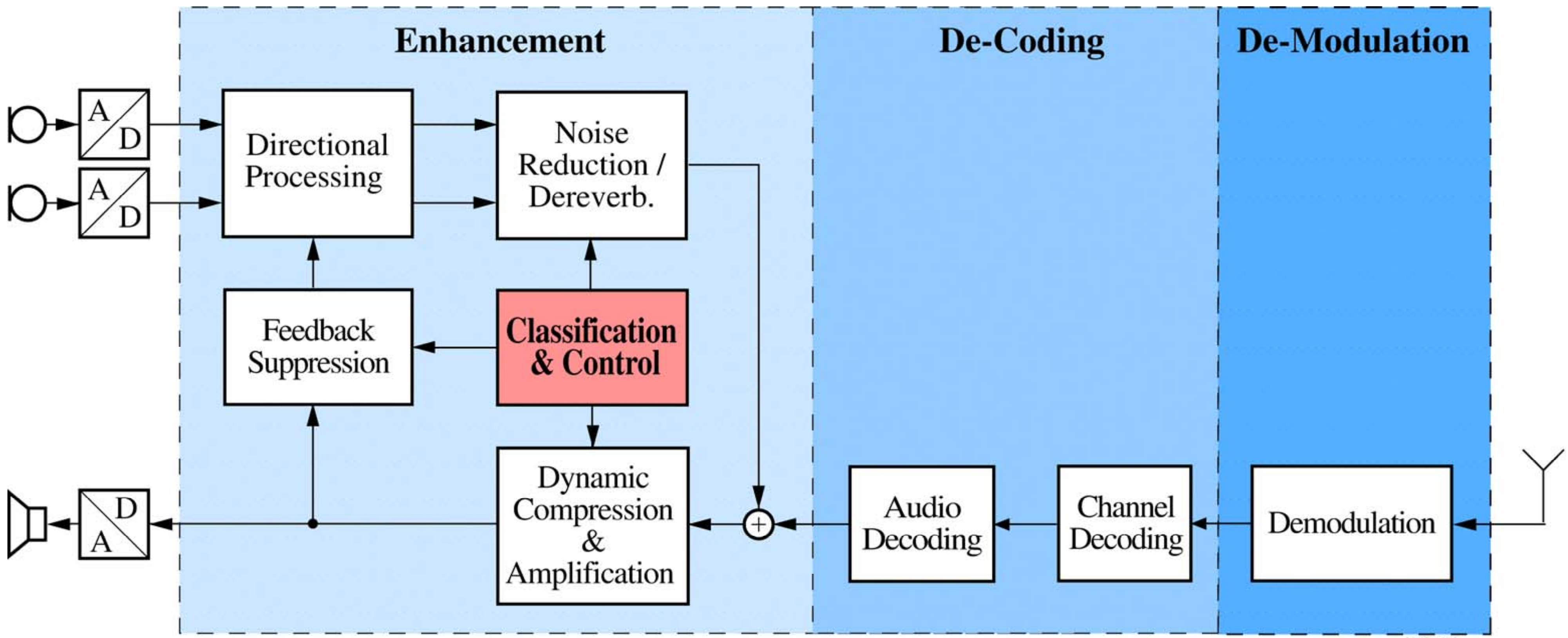
- Enhancement



Hearing Aid



- Enhancement and external digital wireless audio input

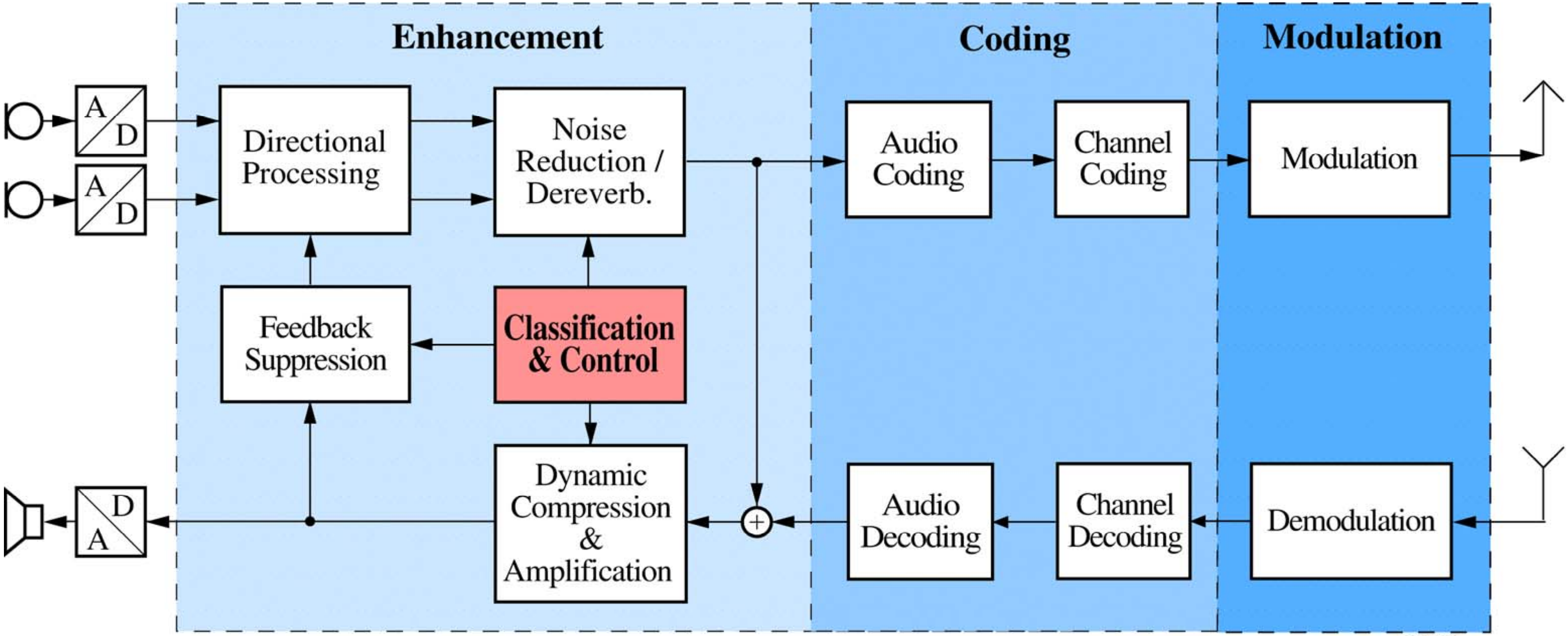


Hearing Area Network – Wireless Connectivity

[Based on PHONAK Hearing Systems]

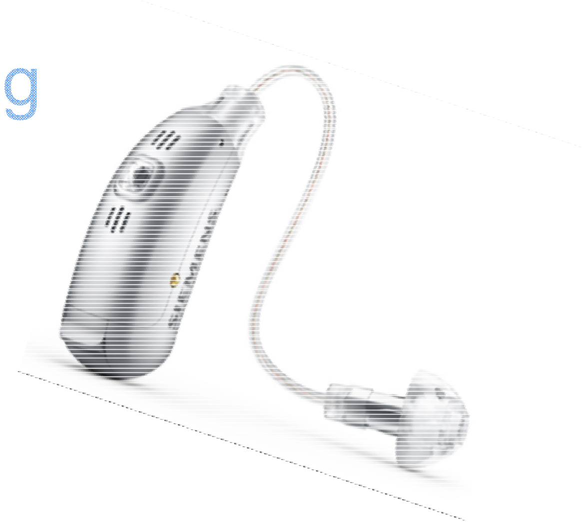
Hearing Aid with Binaural Audio Processing

- Enhancement, coding & modulation



Advanced Speech-Audio Processing in Mobile Phones and Hearing Aids

1. Introduction
2. Acoustical Distinctions
3. Signal Processing & Coding
4. Selected Algorithms
5. Conclusions



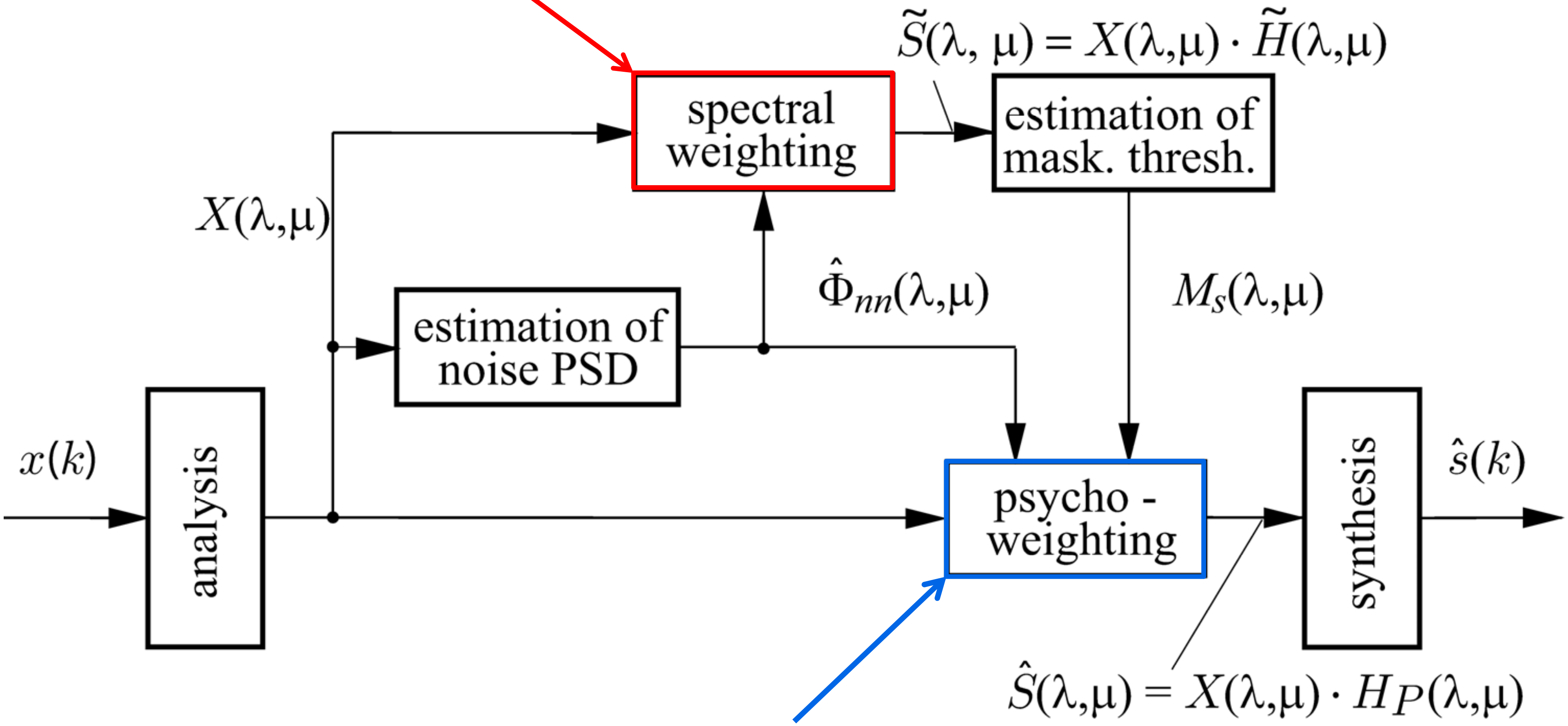
4. Selected Signal Processing Algorithms

1) Single Microphone Noise Reduction
2) Dual Microphone Noise Reduction
3) Speech-Audio Coding
4) Intelligibility / Listening Enhancement
5) Artificial Bandwidth Extension
6) Wind Noise Reduction
7) Spatial HD-Telephony

To which extent may algorithms be re-used in mobile phones and digital hearing aids?

4.1 Single Microphone Noise Reduction

- Background filter $\tilde{H}(\lambda, \mu)$ for masking threshold estimation

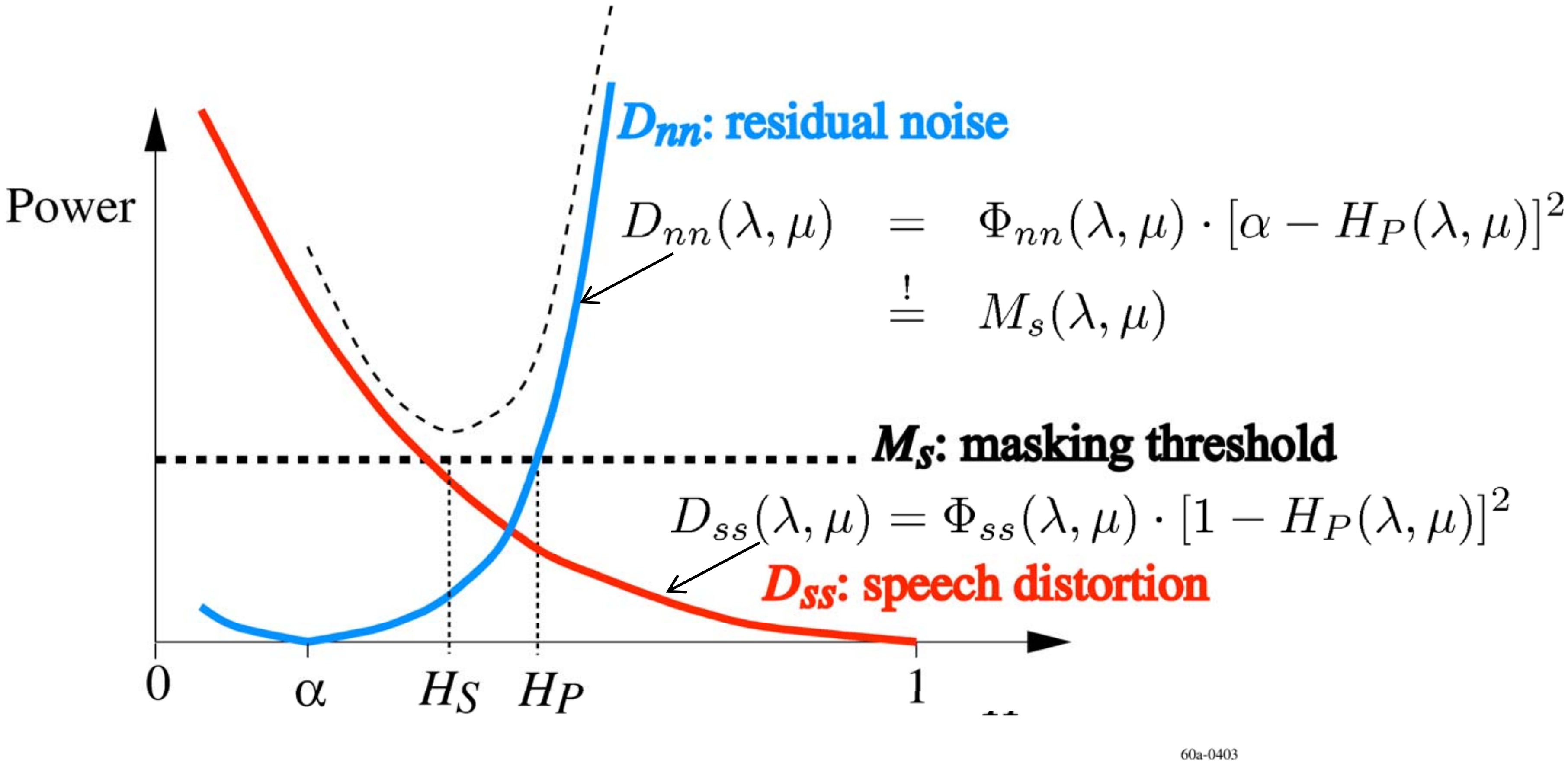


- Psychoacoustic foreground filter $H_P(\lambda, \mu)$

$k = \text{time}, \lambda = \text{frame}, \mu = \text{frequency}$

Psychoacoustic Weighting Rule H_P

- Choose H_P such that residual noise is masked by speech
- Target for frequency bins with noise only: $N_{\text{out}}(\lambda, \mu) \stackrel{!}{=} \alpha \cdot N_{\text{in}}(\lambda, \mu)$

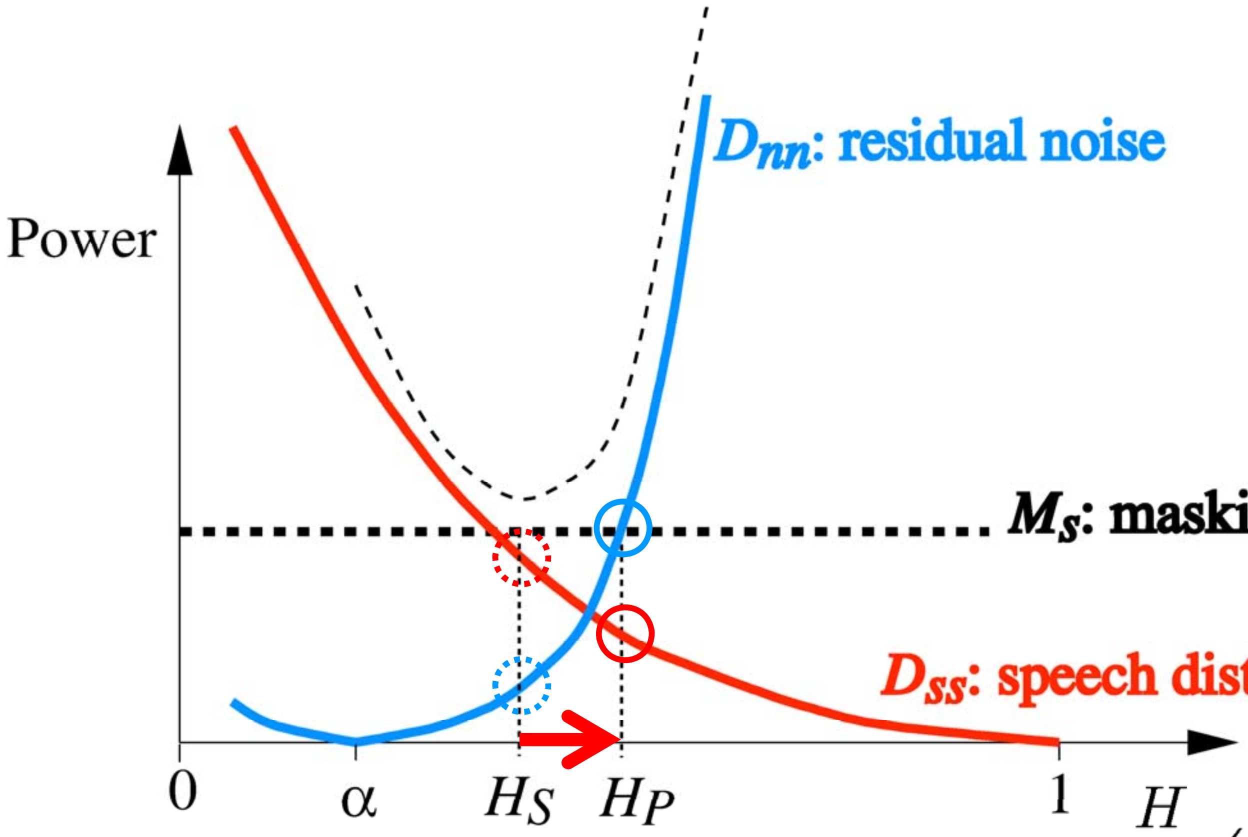


$\lambda = \text{frame}, \mu = \text{frequency}$

[Stefan Gustafsson IEEE Tr. SAP, 2002]

Psychoacoustic Weighting Rule H_P

- Choose H_P such that residual noise is masked by speech
- Target for frequency bins with noise only: $N_{\text{out}}(\lambda,\mu) \overset{!}{=} \alpha \cdot N_{\text{in}}(\lambda,\mu)$



$$H_S = \text{Spectral Subtraction} \;\Rightarrow H_P(\lambda,\mu) = \min\left(\sqrt{\frac{M_s(\lambda,\mu)}{\Phi_{nn}(\lambda,\mu)}} + \alpha,\ 1\right)$$

$\lambda = \text{frame},\ \mu = \text{frequency}$

[Stefan Gustafsson IEEE Tr. SAP, 2002]

Audio Example:

SNR = 10 dB

- Magnitude spectral subtraction

$$H_S = \max\left(\frac{|X| - \sigma_N}{|X|},\, 0\right)$$

input: $s + n$

output: $\hat{s} + \Delta n$

output noise: Δn

- Psychoacoustic weighting rule

$$H_P = \min\left(\sqrt{\frac{M_s}{|\sigma_N^2|}} + \alpha,\, 1\right)$$

$$20\,\mathrm{lg}(\alpha) = -15\,\mathrm{dB}$$

output: $\hat{s} + \Delta n$

output noise: Δn

4.2 Dual Microphone Noise Reduction

- Mobile phone in hands-free / loudspeaking mode

- With coherence functions Γ_s, Γ_n of speech s and noise n as a function of frequency

$$\Phi_{x1,x2} = \Gamma_s \cdot \Phi_{ss} + \Gamma_n \cdot \Phi_{nn}$$

- Noise PSD estimate

$$\hat{\Phi}_{nn} = \frac{\sqrt{\Phi_{x1x1} \cdot \Phi_{x2x2}} - \frac{\Phi_{x1x2}}{\Gamma_s}}{1 - \frac{\Gamma_n}{\Gamma_s}}$$

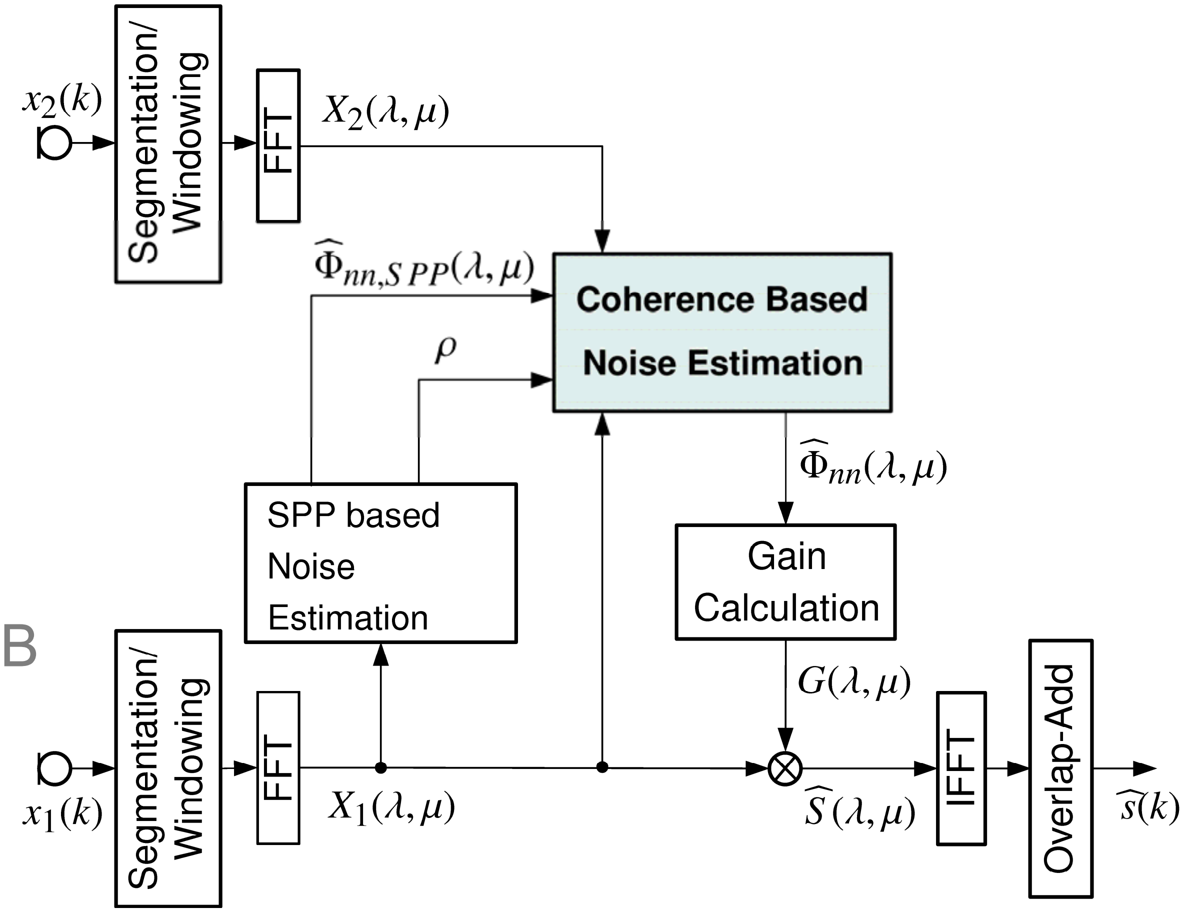
- Adaptive learning of Γ_s, Γ_n using *Speech Presence Probability*

  (SPP, soft decision voice activity detection, T. Gerkmann, R.C. Hendriks, WASPAA 2011)

[Christoph Nelke, ICASSP 2013]

Audio Example: Dual Microphone Noise Reduction

SNR = 5 dB



$$k = \text{time},\ \lambda = \text{frame},\ \mu = \text{frequency}$$

[Christoph Nelke, ICASSP 2013]

4.3 Speech-Audio Coding

Mobile phones:

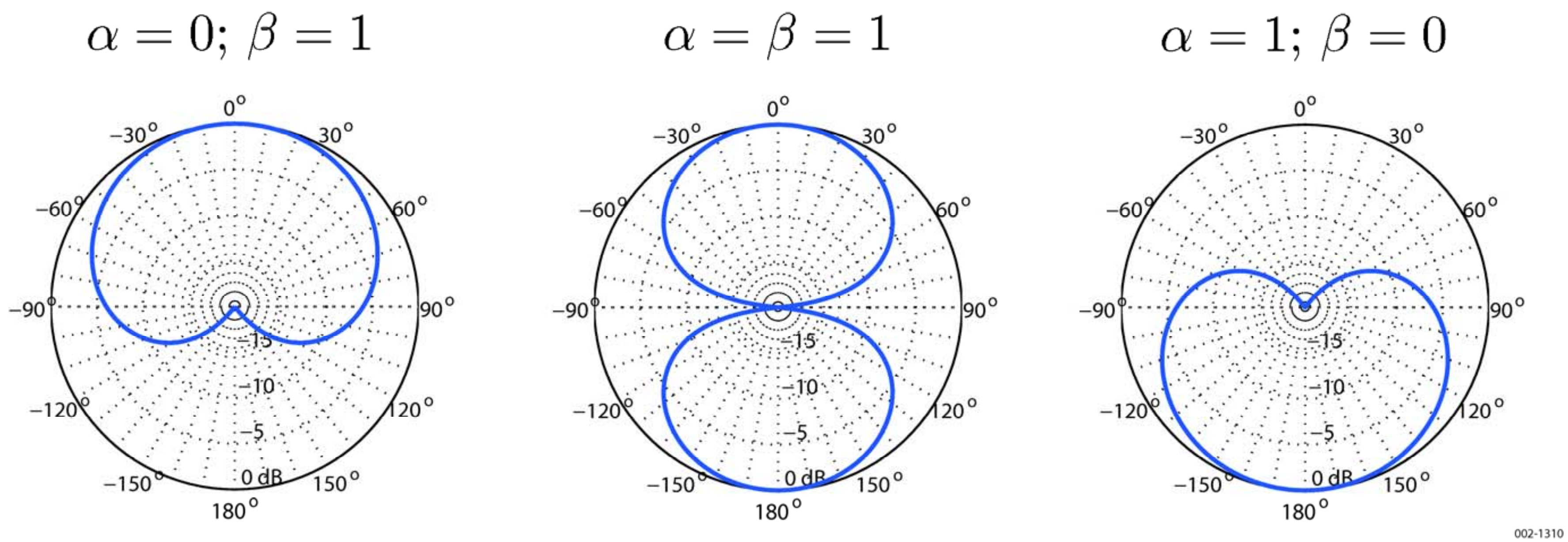
- *Model based* monaural coding with 1–2 bits per sample
- Latency: 20 ms
- Audio bandwidth: 3.4 or 7.0 kHz
- Shaping of quantisation-noise shaping to exploit masking

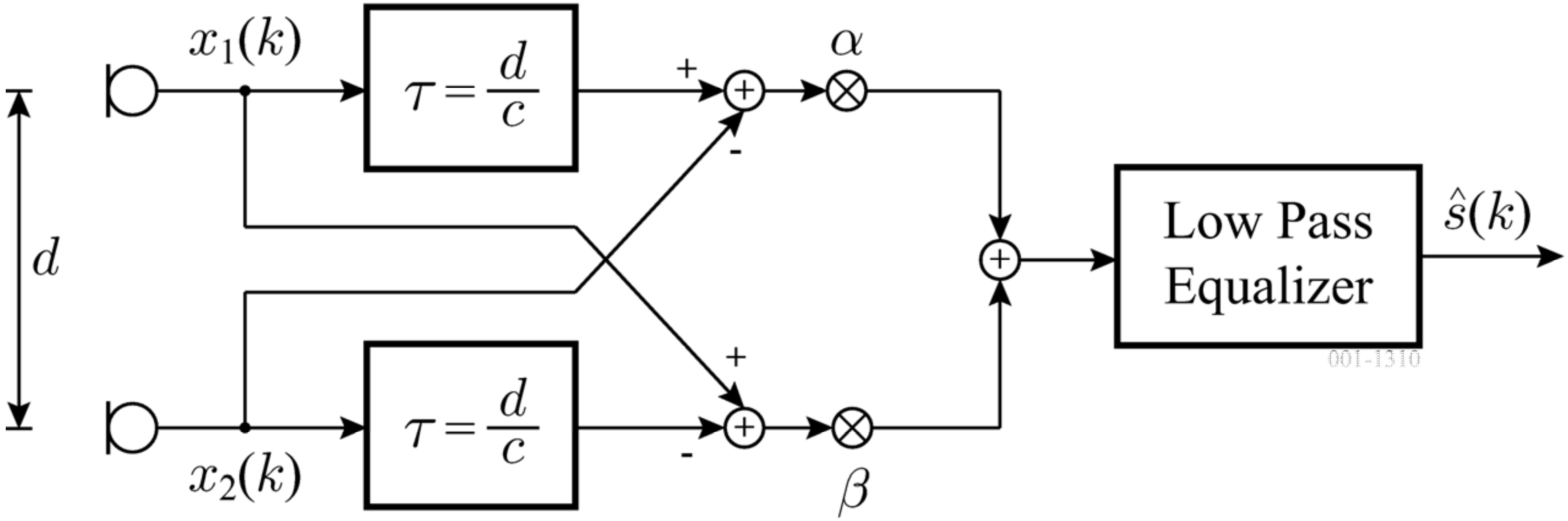
Hearing aids:

- *Waveform coding* (mono or stereo) with more than 2 bits per sample
- Audio bandwidth more than 7 kHz
- External audio link
  - Latency: not critical
  - Noise shaping
- Internal binaural link
  - Latency: 5 ms
  - Noise shaping?

[Bastian Sauert, EUSIPCO 2010]

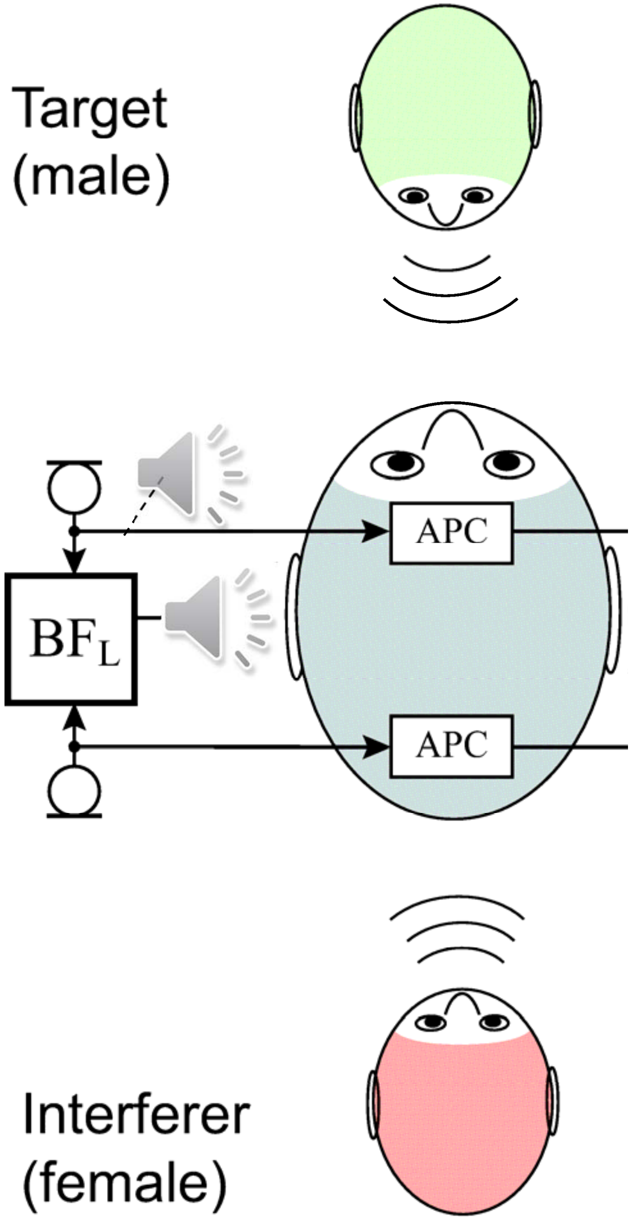
Differential Beamforming (DBF)

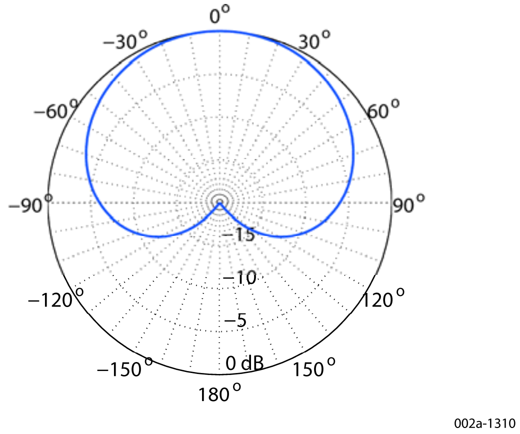




[Based on G. Elko, A.N. Pong, WASPAA 1995]

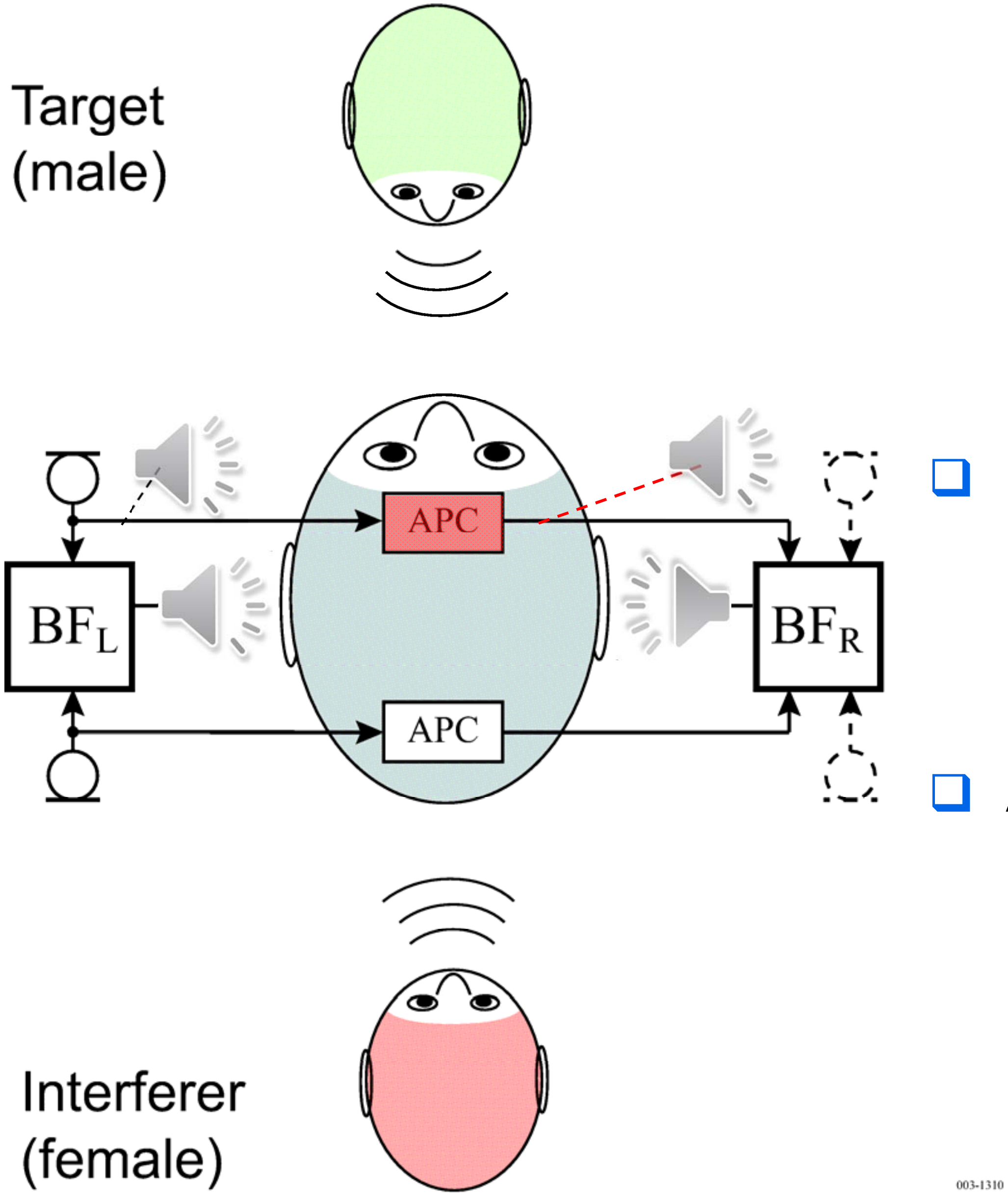
Differential Beamforming and Coding

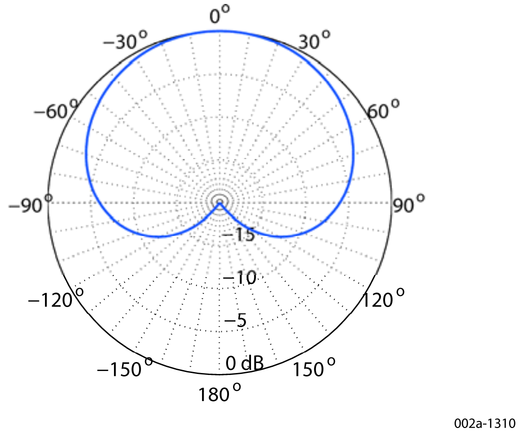




- Differential beamformer (BF):
  - d = 10 mm
  - cardioid

Differential Beamforming and Coding

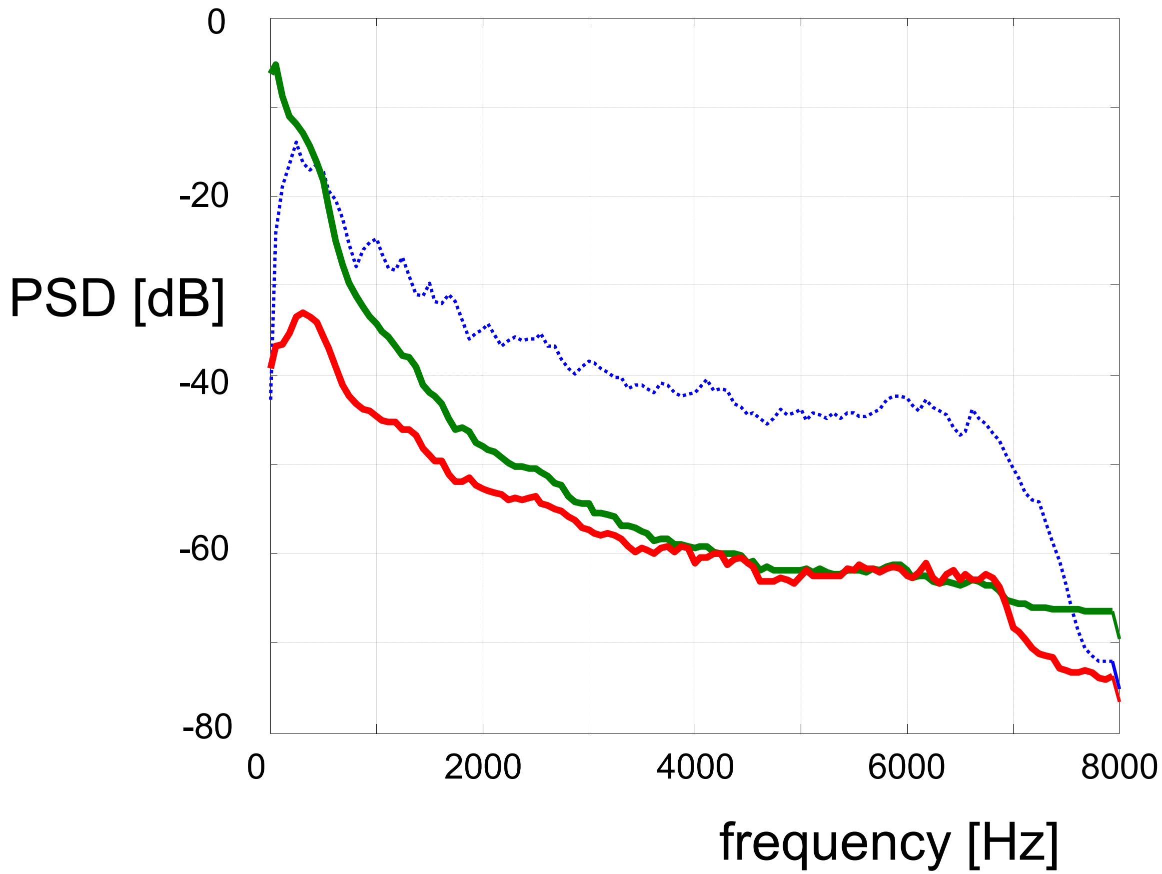




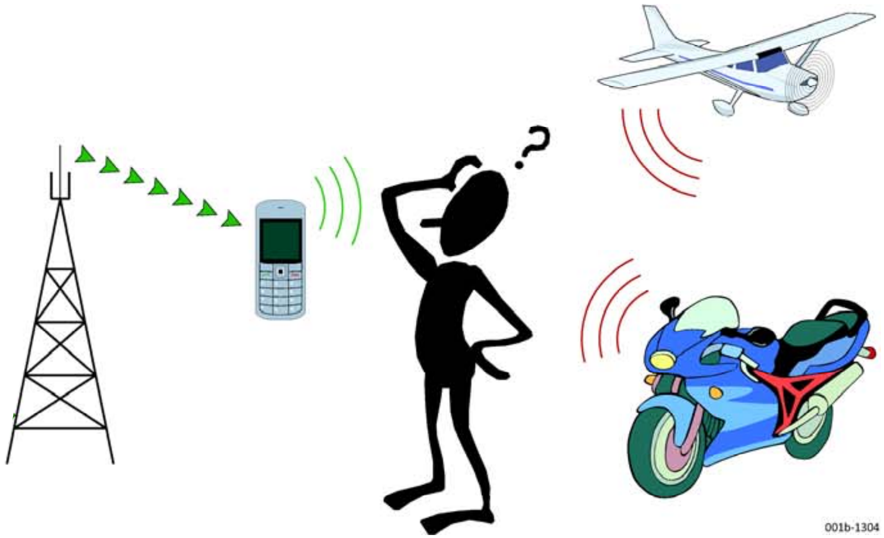
- Differential beamformer (BF):
  - d = 10 mm
  - cardioid
- Adaptive Predictive Coding (APC):
  - f_s = 16 kHz
  - 4 bits/sample
  - noise shaping

Differential Beamforming with / without Coding

- Power Spectral Densities (PSDs):
  - front microphone signal
  - quantization noise of encoded microphone signal
  - beamformer error using encoded signals



4.4 Intelligibility / Listening Enhancement



- Near-end listener experience:
  - Higher listening effort
  - Reduced speech intelligibility

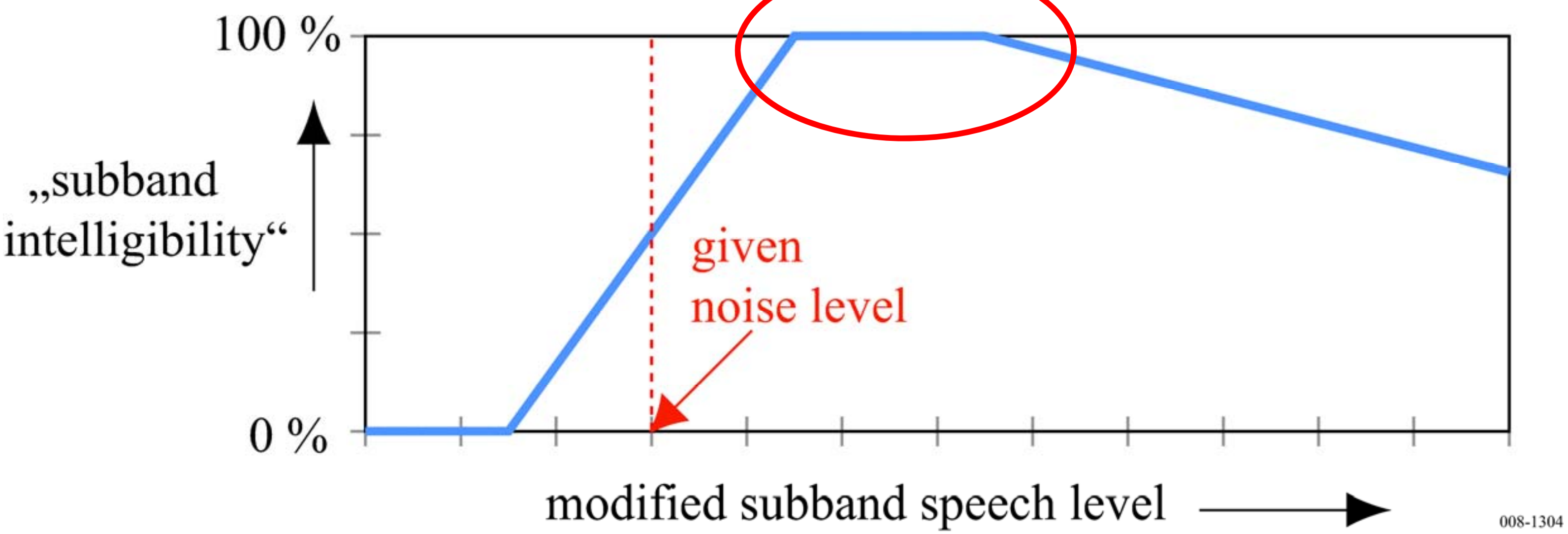
- Approach:
  - *Preprocess* clean far-end speech
  - Enhance *intelligibility* in near-end noise
  - *Re-distribute* signal frame energy over frequency

- Constraints:
  - Ear damage
  - Loudspeaker protection
  - Low delay

Optimization: Speech Intelligibility Index (SII)

- Intelligibility maximization by dynamical spectral weighting in Bark bands

$$\tilde{S}(\lambda,\mu) = W(\lambda,\mu)\cdot S(\lambda,\mu) \qquad W(\lambda,\mu) = \mathrm{f}(\hat{\Phi}_{ss}(\lambda,\mu),\ \hat{\Phi}_{nn}(\lambda,\mu))$$

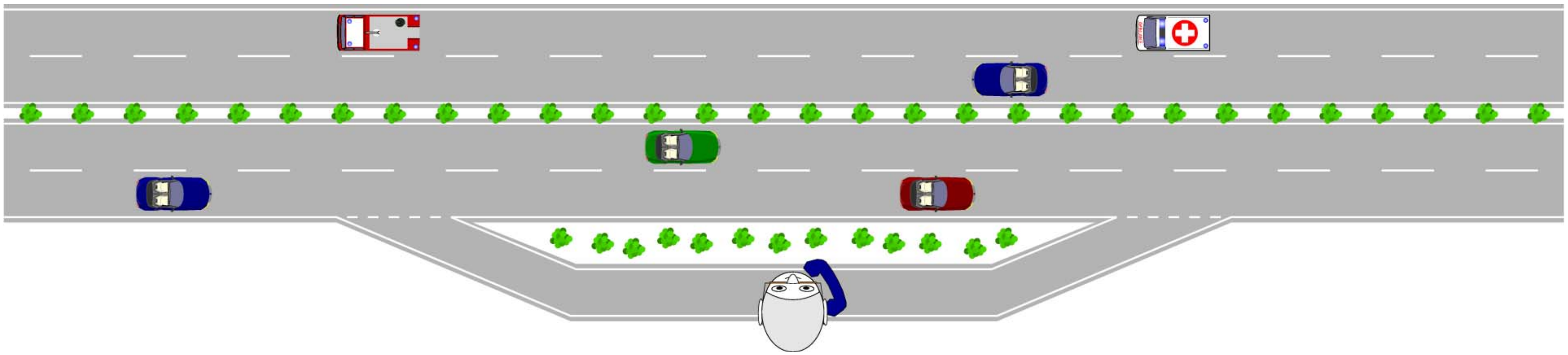


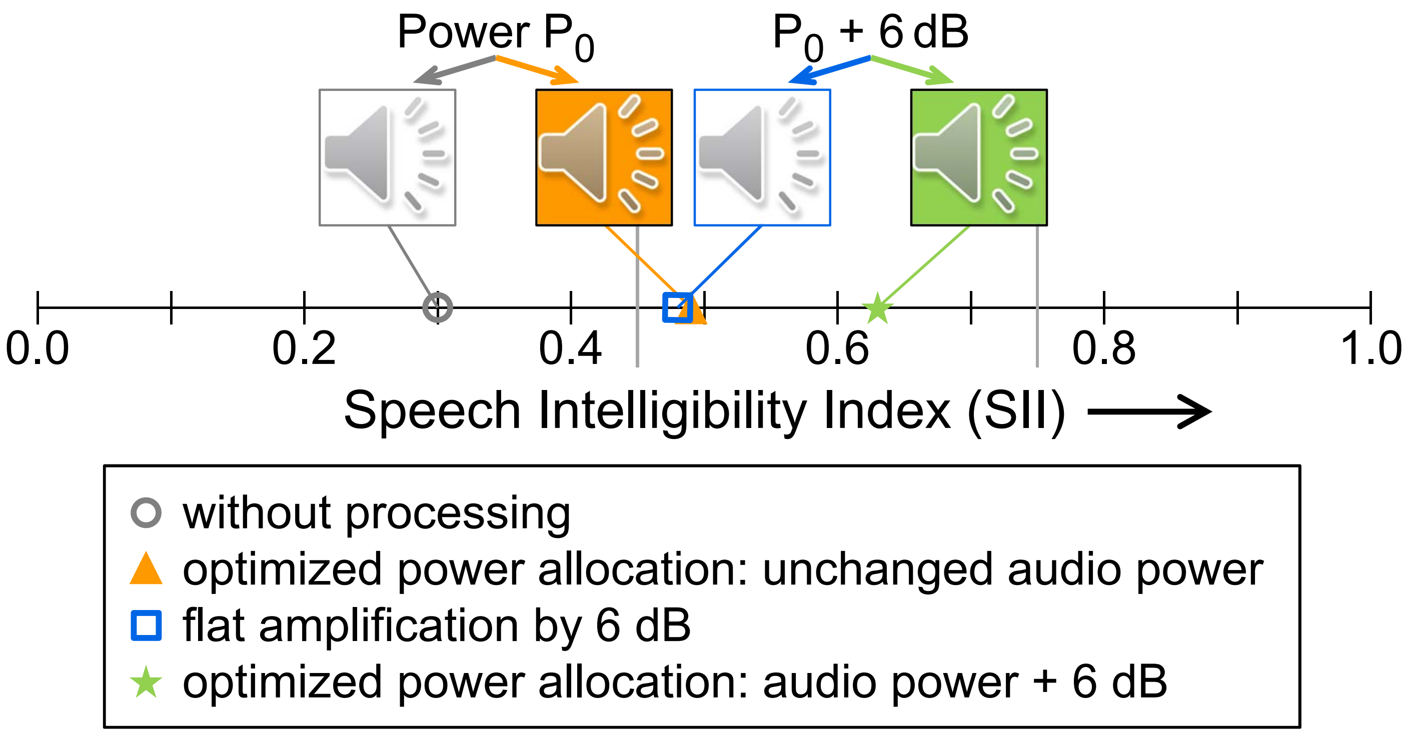
- Constrained 17-dimensional non-linear optimization
- Up to 22 percentage points increase of word recognition rate without increasing total audio power

[Bastian Sauert, EUSIPCO 2010]

Performance in Real Environment

- Near-end listener at a motorway station



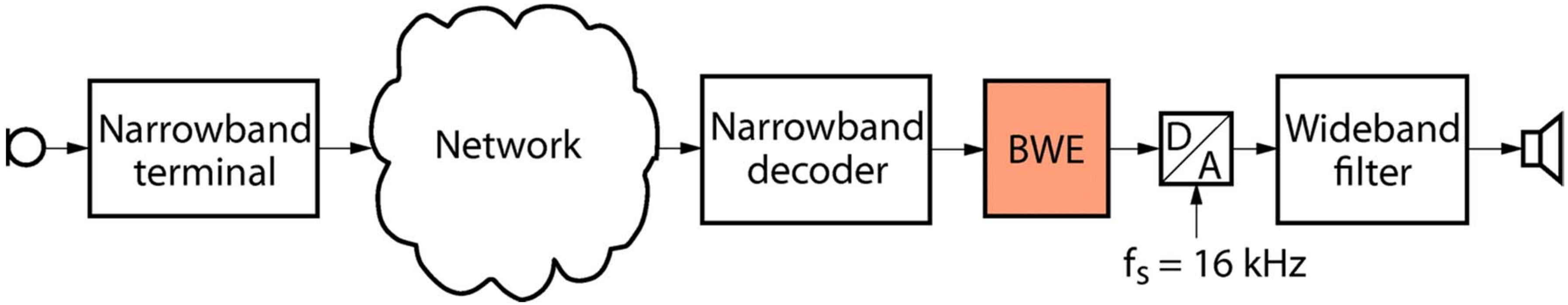


4.5 Artificial Bandwidth Extension (BWE)

- Today's narrowband telephony: 3.4 kHz
- Tomorrow's telephony with wideband speech coding: 7 kHz
- Long transition period

far end with
narrowband terminal

receiving end with artificial
bandwidth extension

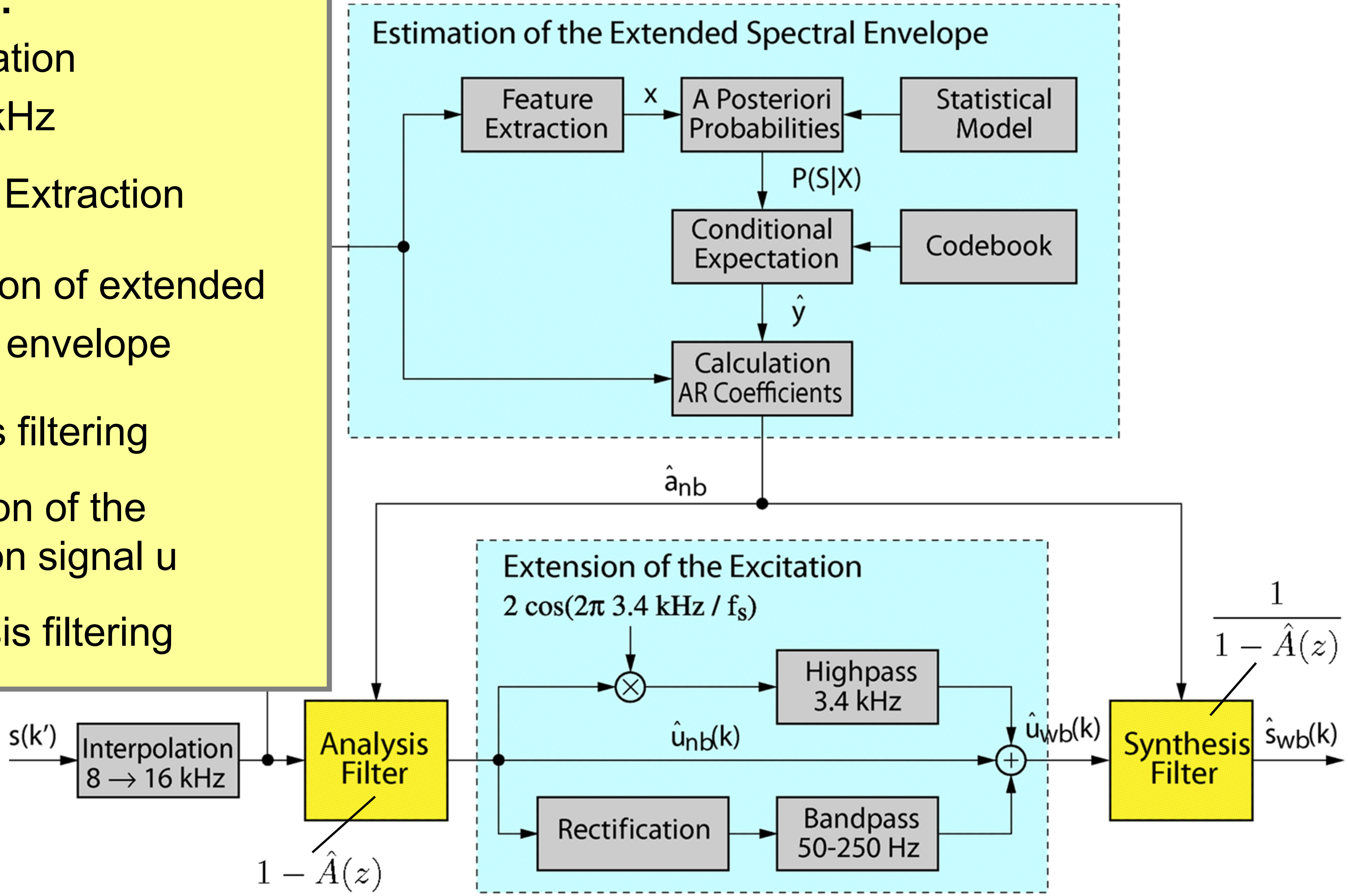


- Bandwidth extension at the receiving end

BWE: Recognition & Estimation & Processing

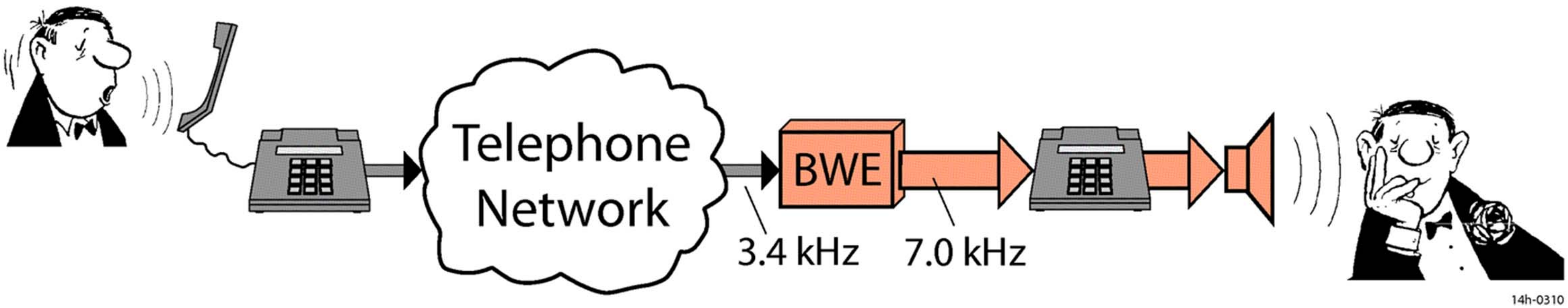
Algorithm:

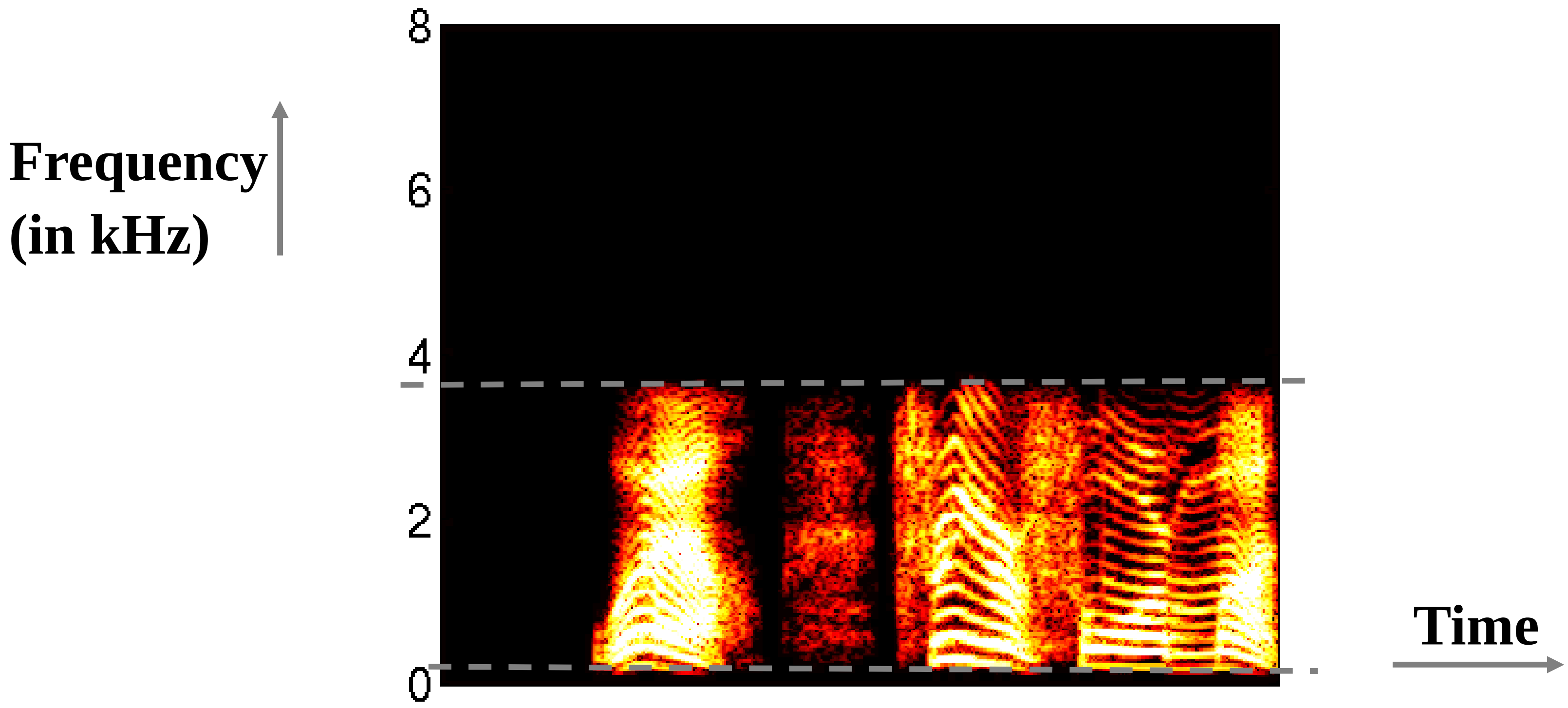
1. Interpolation
   8 → 16 kHz
2. Feature Extraction
3. Estimation of extended spectral envelope
4. Analysis filtering
5. Extension of the excitation signal u
6. Synthesis filtering



[Peter Jax, ICASSP 2004]

Audio Example: Telephone Speech without & with BWE

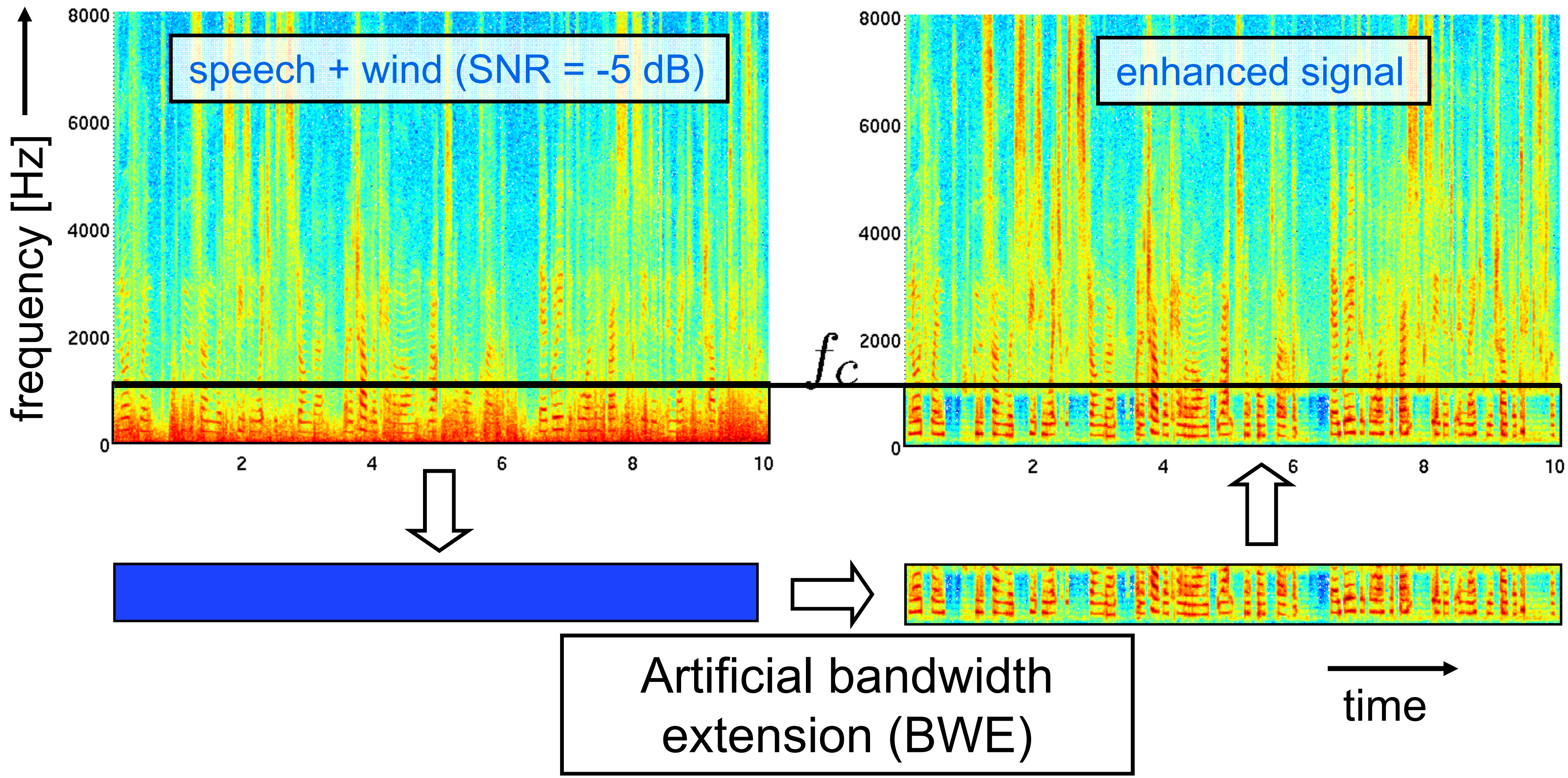




GO

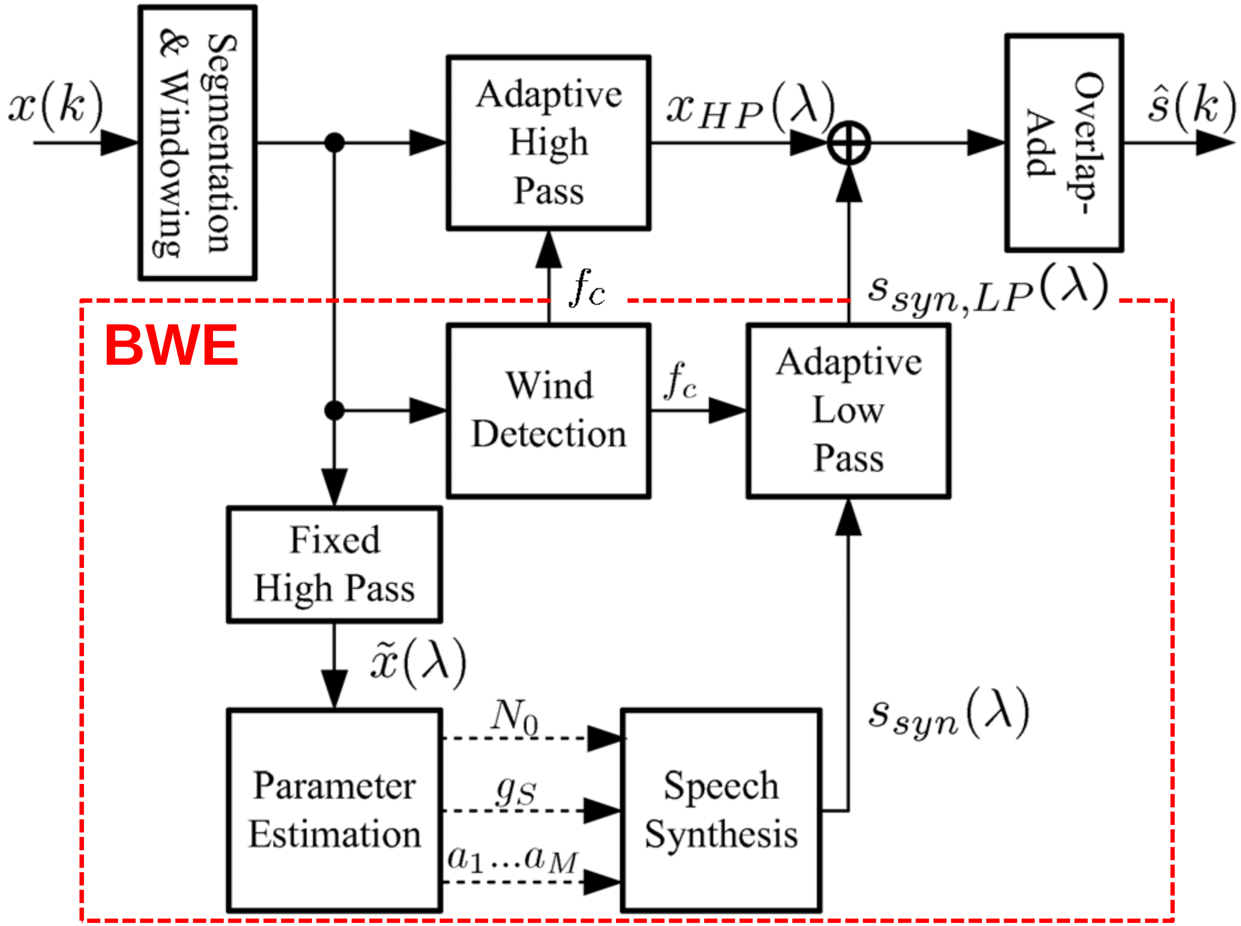
4.6 Wind Noise Reduction

- Wind noise = low frequency noise with $f < f_c$ (time varying)
- Substitution of disturbed frequency band using BWE



Wind Noise Reduction: Algorithm

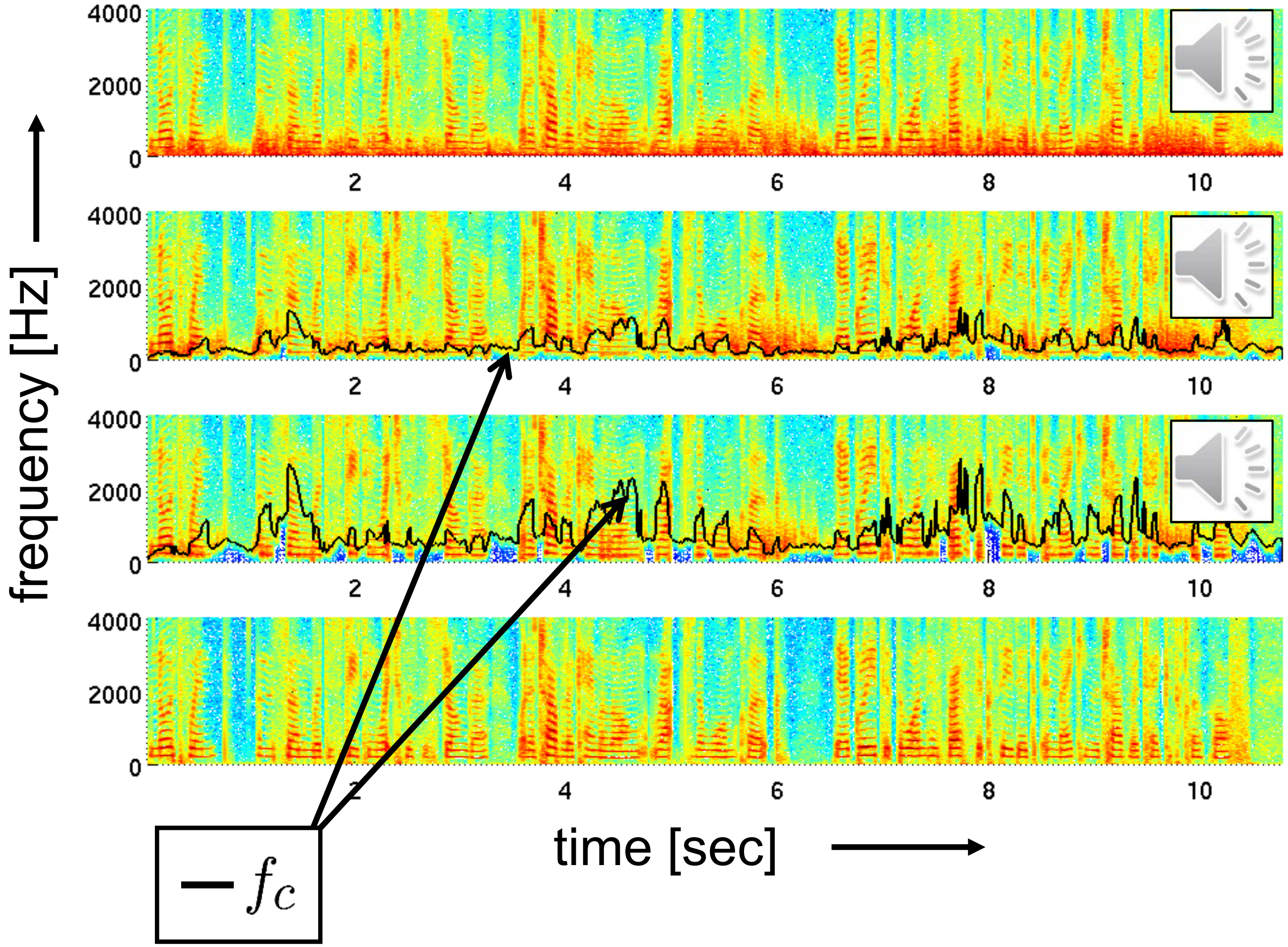
- Substitution of disturbed frequency band by bandwidth extension (BWE)



N_0	: Pitch period	f_c	: Range of synthesized speech
$a_1 \ldots a_M$	: LPC coefficients	s_{syn}	: Synthetic speech
g_S	: Gain factor	λ	: Frame index

[Christoph Nelke, EUSIPCO 2012]

Wind Noise Reduction: Audio Example



Unprocessed speech

Enhanced speech
moderate, $f_{c,\max} = 1.5\,\mathrm{kHz}$

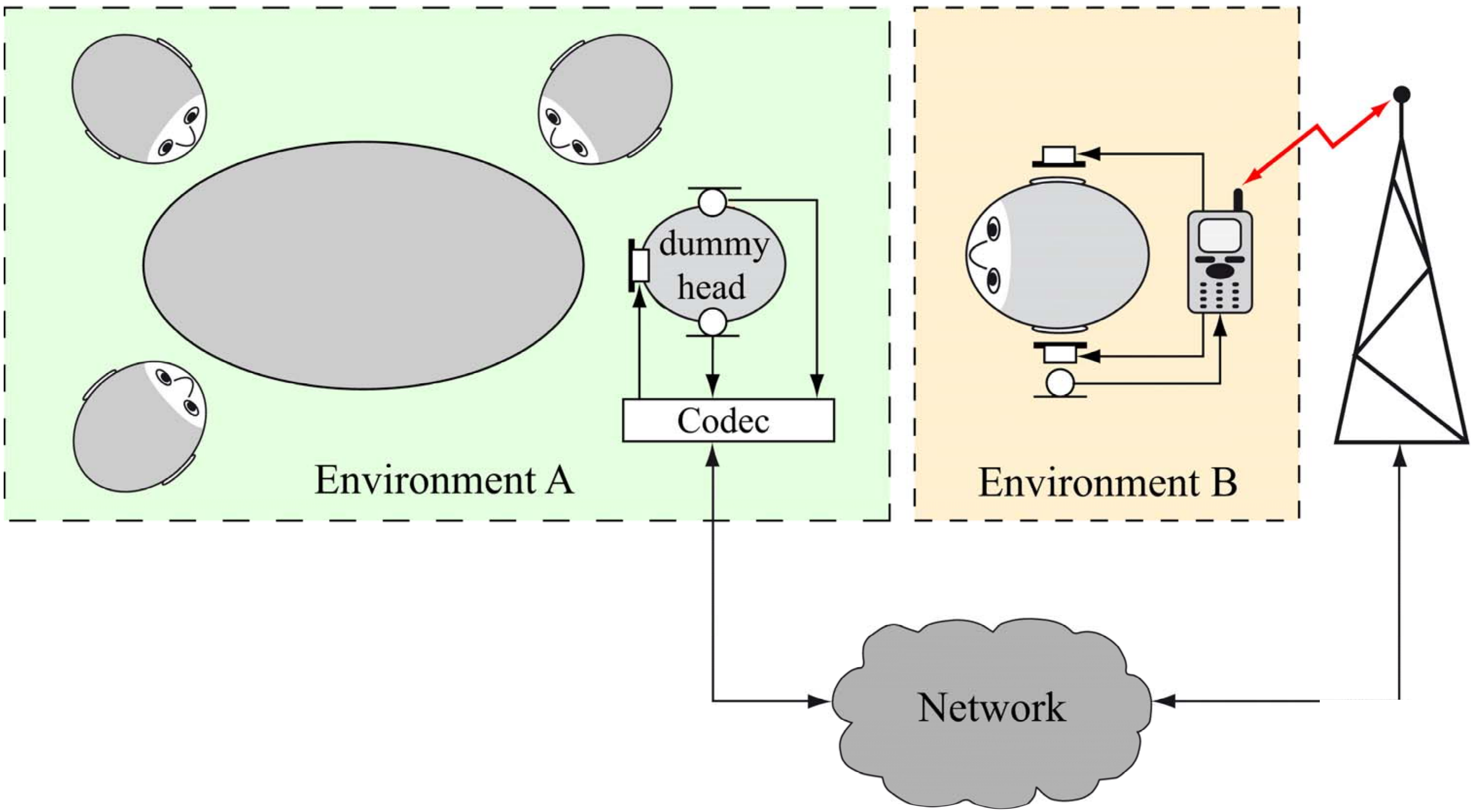
Enhanced speech
aggressive, $f_{c,\max} = 4\,\mathrm{kHz}$

Clean speech

$\mathrm{SNR} = \text{-}5\,\mathrm{dB} \quad f_s = 16\,\mathrm{kHz}$

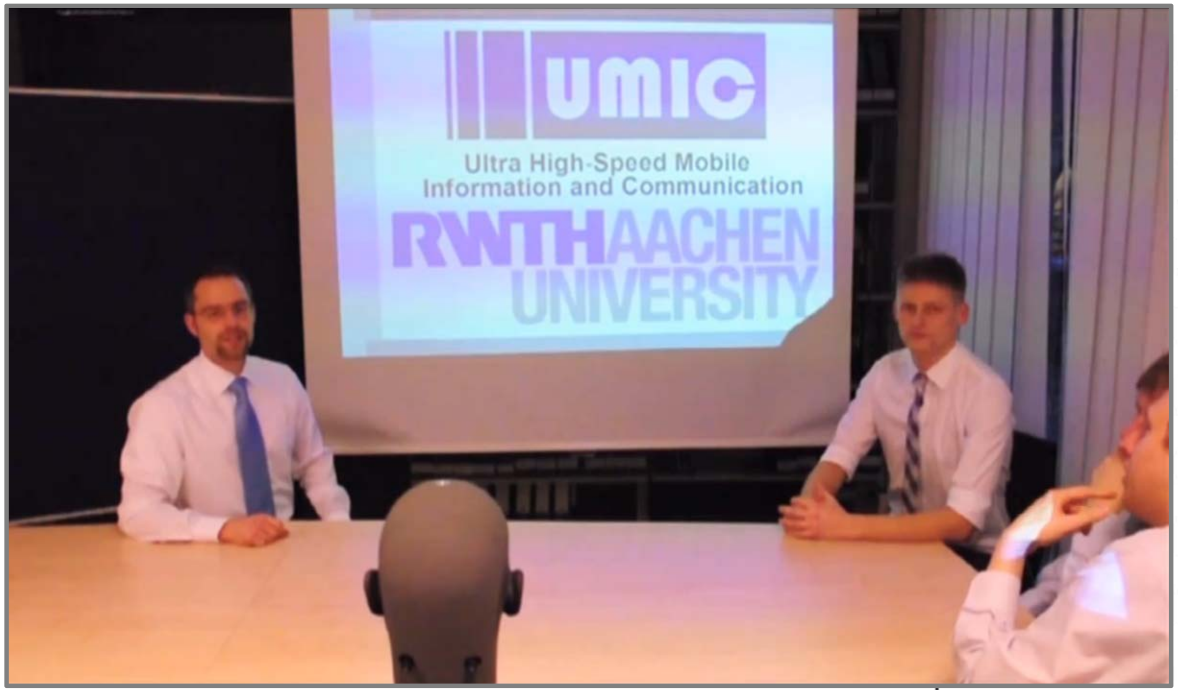
4.7 Spatial HD-Telephony & Audio Conferencing

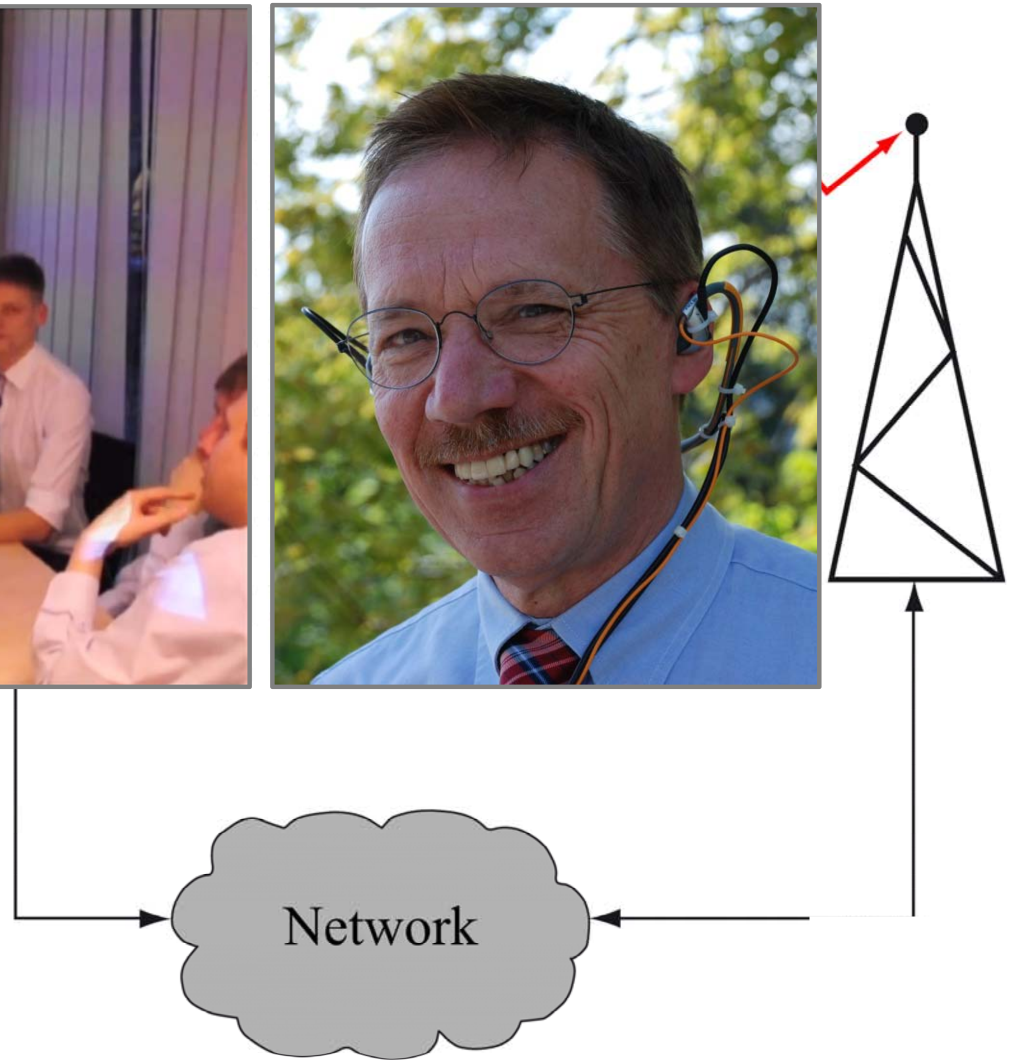
- Binaural headset or dummy head



4.7 Spatial HD Telephony & Audio Conferencing

- Binaural headset or dummy head

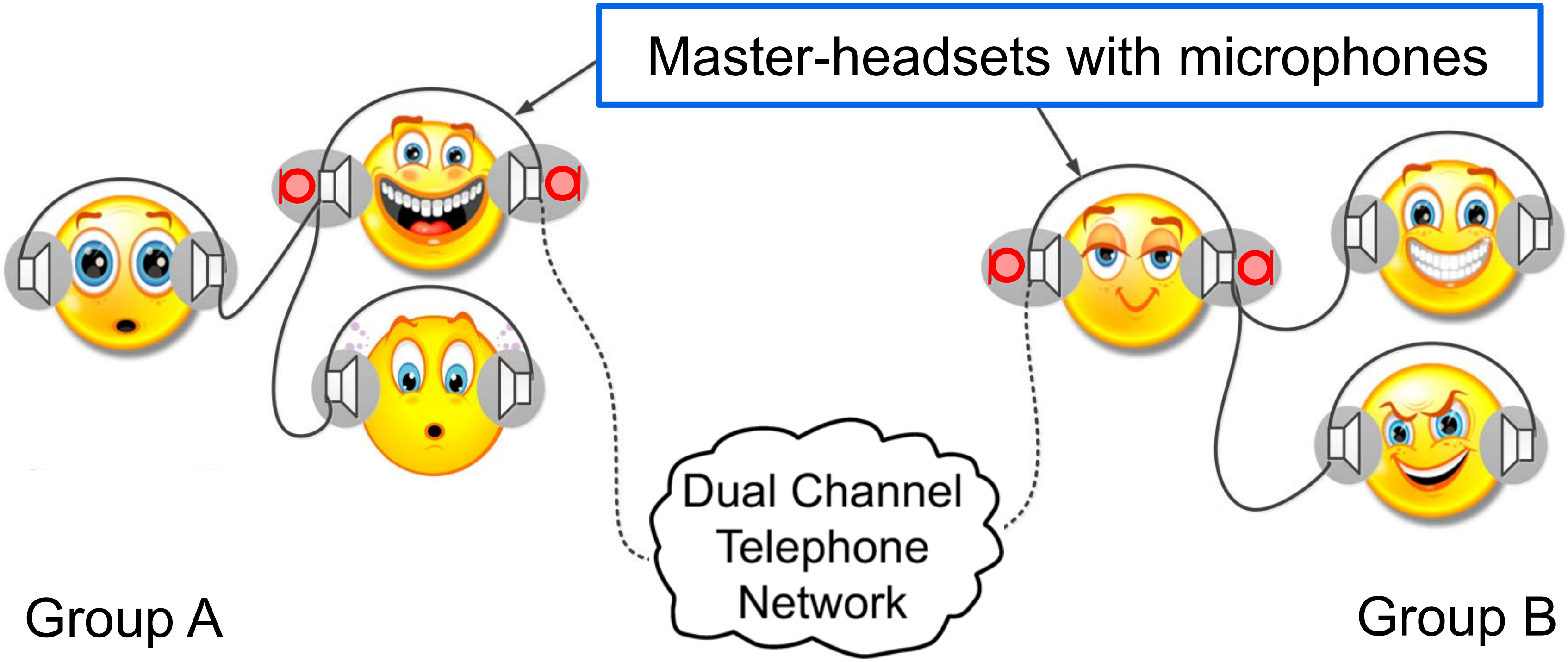




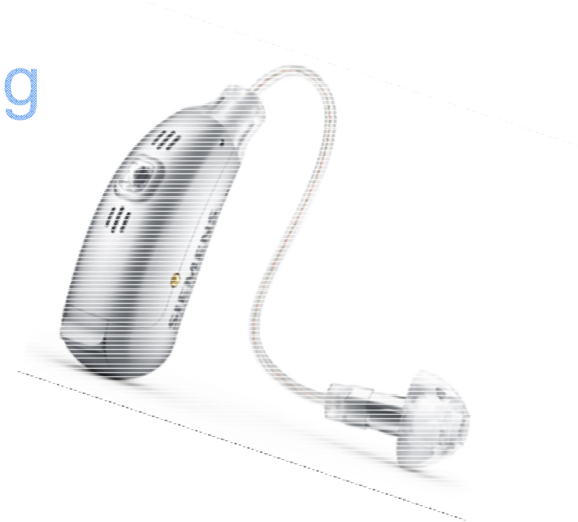
http://www.binaural-telephony.com/
(Listen with head phones)

[Matthias Rüngeler, et. al. , IWAENC 2012]

Binaural Group-Communication



Advanced Speech-Audio Processing in Mobile Phones and Hearing Aids



Re-Usability of Algorithms

		MP	HA
1.	Single Microphone Noise Reduction	✓	✓
2.	Dual Microphone Noise Reduction	✓	✓
3.	Speech-Audio Coding	✓	✓
4.	Near-End Listening Enhancement	✓	(✓)
5.	Artificial Bandwidth Extension	✓	✓
6.	Wind Noise Reduction	✓	✓
7.	Spatial (Binaural) HD Telephony	✓	(✓)

Re-Usability of Algorithms

	MP	HA
1. Single Microphone Noise Reduction	✓	✓
2. Dual Microphone Noise Reduction	✓	✓
3. Speech-Audio Coding	✓	✓
4. Near-End Listening Enhancement	✓	(✓)
5. Artificial Bandwidth Extension	✓	✓
6. Wind Noise Reduction	✓	✓
7. Spatial (Binaural) HD Telephony	✓	(✓)
8. Echo & Feedback Cancellation	✓	✓
9. Dereverberation	?	✓
10. Modulation & Error Protection	✓	✓

Conclusions

Synergy effects:
similar algorithms

single microphone NR

dual microphone NR

wind noise reduction

listening enhancement

spatial HD telephony

BWE, coding

…

Mobile phones and hearing aids are approaching each other

normal/impaired hearing

coherence

level differences

audio bandwidth

power consumption

latency

…

Distinctions:
partly different
design constraints

Advanced Speech-Audio Processing in Mobile Phones and Hearing Aids

Peter Vary
RWTH Aachen University
Institute of Communication Systems

Thanks for contributions:

Christiane Antweiler
Bastian Sauert
Christoph Nelke
Bernd Geiser
Magnus Schäfer
Marco Jeub
Matthias Dörbecker
Stefan Gustafsson
Peter Jax

RWTH AACHEN UNIVERSITY

- Nelke C. M., Beaugeant, C., and Vary P.:
  Dual Microphone Noise PSD Estimation for Mobile Phones in Hands-Free Position Exploiting the Coherence and Speech Presence Probability,
  in: *Proceedings of ICASSP*, Vancouver, Canada, May 26–31, 2013.

- Nelke, C. M., Nawroth, N., Jeub, M., Beaugeant, C., and Vary, P.:
  Single Microphone Wind Noise Reduction Using Techniques of Artificial Bandwidth Extension,
  in: *Proceedings of EUSIPCO*, Bucharest, Romania, Aug. 27–31, 2012, p. 2328–2332.

- Sauert, B. and Vary, P.:
  Listening Enhancement for Mobile Phones – How to Improve the Intelligibility in a Noisy Environment,
  in: *The Listening Talker Workshop (LISTA)*, Edinburgh, United Kingdom, May 2–3, 2012, invited talk.

- Jeub, M., Herglotz, C., Nelke, C. M., Beaugeant, C., and Vary, P.:
  Noise Reduction for Dual-Microphone Mobile Phones Exploiting Power Level Differences,
  in: *Proceedings of ICASSP*, Kyoto, Japan, March 25–30, 2012.

- Sauert, B. and Vary, P.:
  Near End Listening Enhancement: Speech Intelligibility Improvement in Noisy Environments,
  in: *Proceedings of ICASSP*, Toulouse, France, May 14–19, 2006, p. 493–496.

- Jax, P. and Vary, P.:
  On Artificial Bandwidth Extension of Telephone Speech,
  in: *Signal Processing*, vol. 83, no. 8, Aug. 2003, p. 1707–1719.

- Gustafsson, S., Martin, R., Jax, P., and Vary, P.:
  A Psychoacoustic Approach to Combined Acoustic Echo Cancellation and Noise Reduction,
  in: *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, July 2002, p. 245–256.