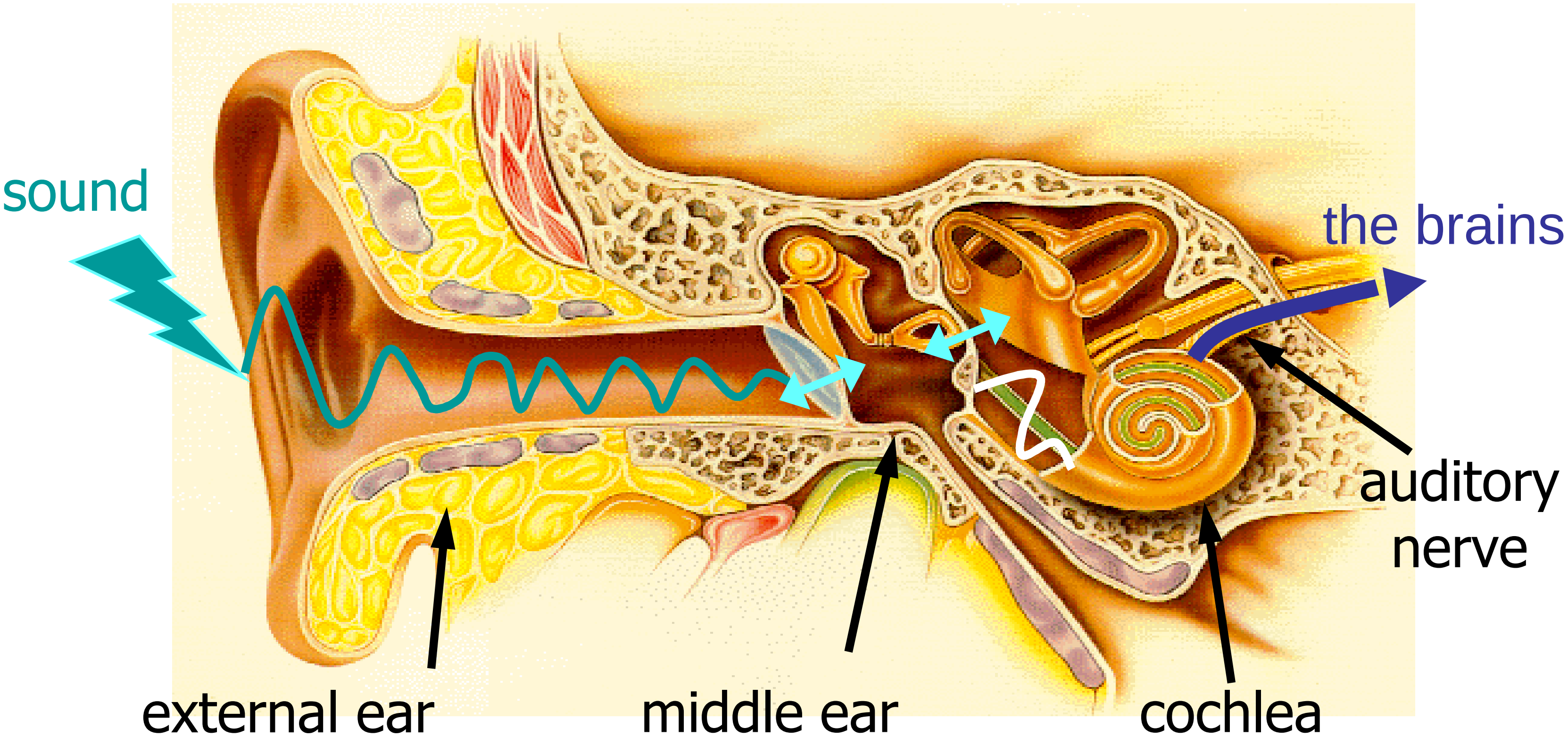
Lecture 2:

Digital signal processing in hearing aids

Prof. Dr. Simon Doclo

University of Oldenburg - Institute of Physics
Signal Processing Group
simon.doclo@uni-oldenburg.de

The auditory system



conductive hearing loss

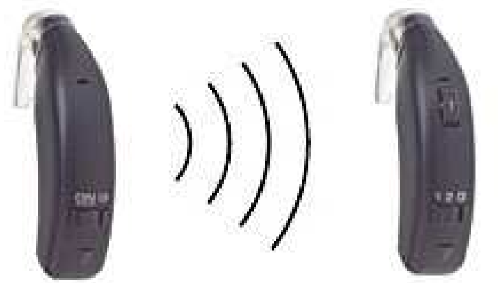
sensorineural hearing loss

↓

hearing aids and cochlear implants

Signal processing in hearing aids

- Possibilities with analog hearing aids = limited !
- **Developments** in HW and micro-electronics:
  - Digital signal processor (DSP)
  - Multiple microphones (2-3)
  - Binaural wireless link between hearing aids

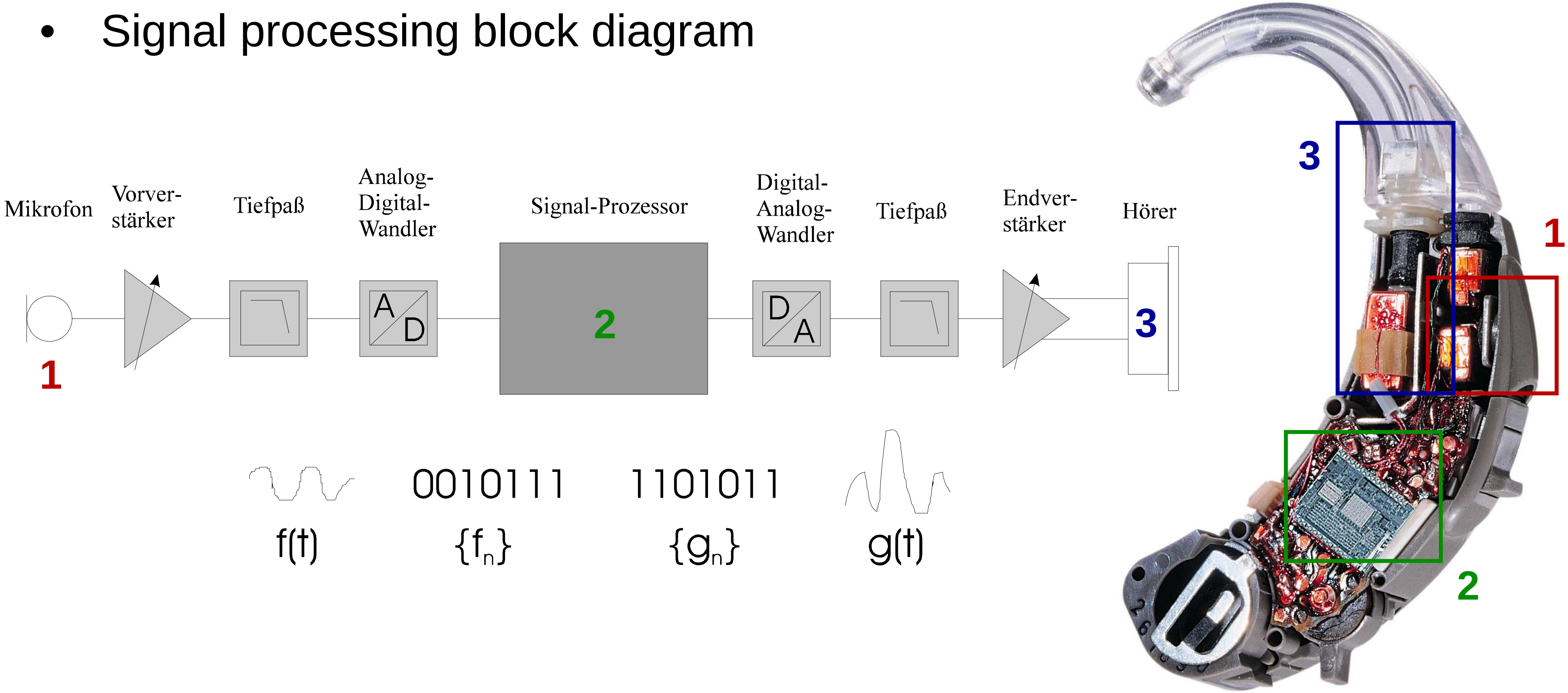


- Digital hearing instruments and cochlear implants allow for **advanced acoustical signal (pre-)processing**

- Important algorithmic **constraints:**
  - Input-output latency (< 10…15 ms)
  - Power constraints from small battery

Signal processing in hearing aids

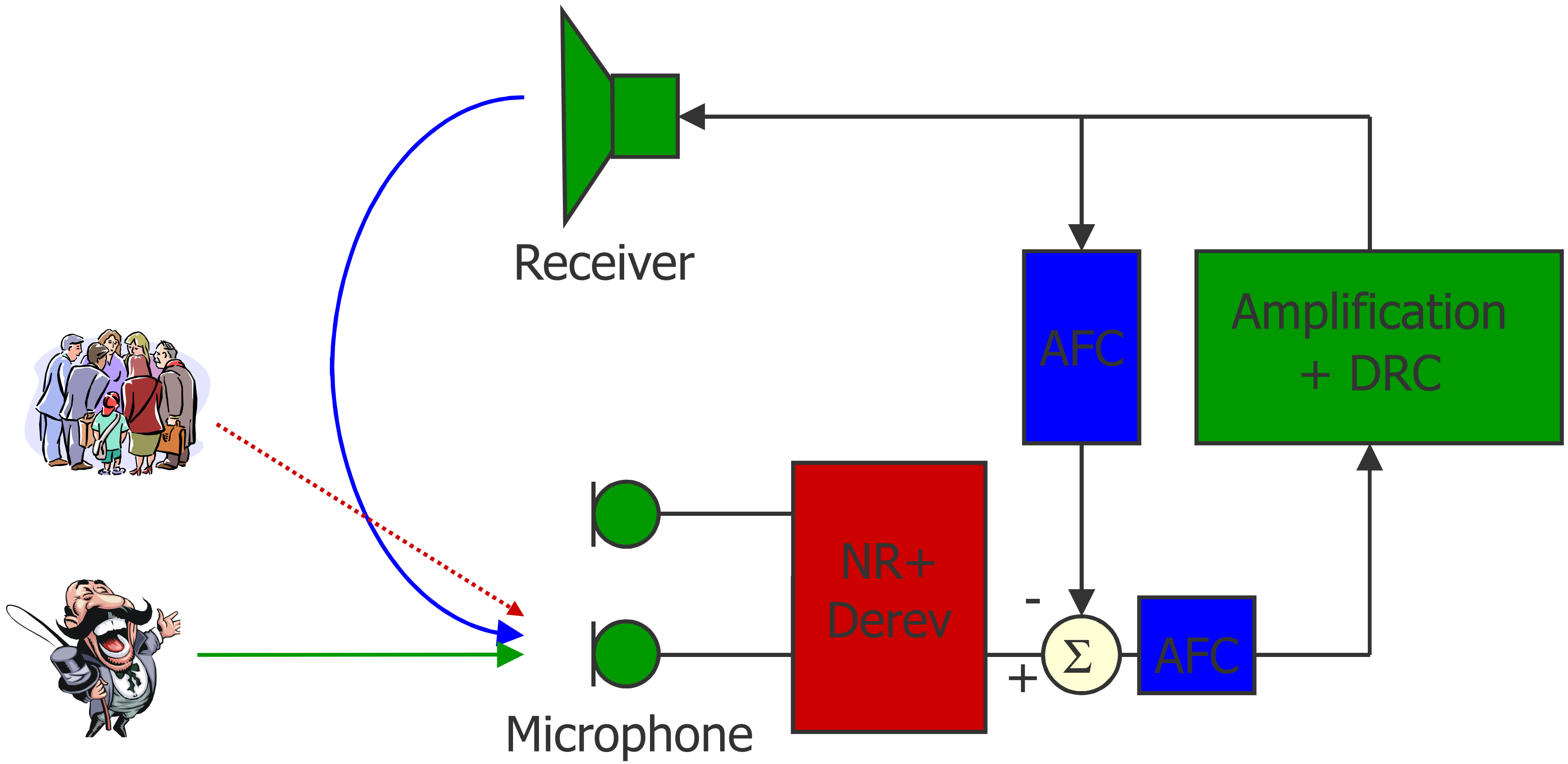
- Signal processing block diagram



Signal processing algorithms

- Cochlear loss:
  – Frequency-specific amplification
  – Dynamic range compression
- Binaural and central loss:
  – Noise reduction
  – Binaural Algorithms
- "Technical" requirements
  – Feedback control (40-60 dB acoustic gain!)
  – Occlusion effect / 'own voice' detection
  – Classification of acoustic environment
  – (fully digital, 1V supply from very small battery, 5-6d battery time, wireless binaural link (new!))

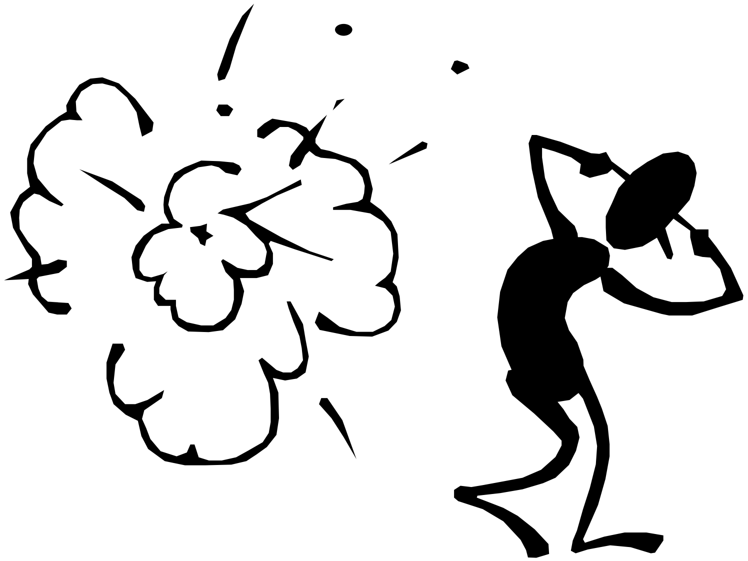
- Basic processing: acoustic amplification and dynamic range compression (frequency-selective)
- Due to acoustic coupling between receiver and microphone (large amplification): acoustic feedback control
- Increase speech intelligibility in background noise: single- or multi-microphone noise reduction and dereverberation

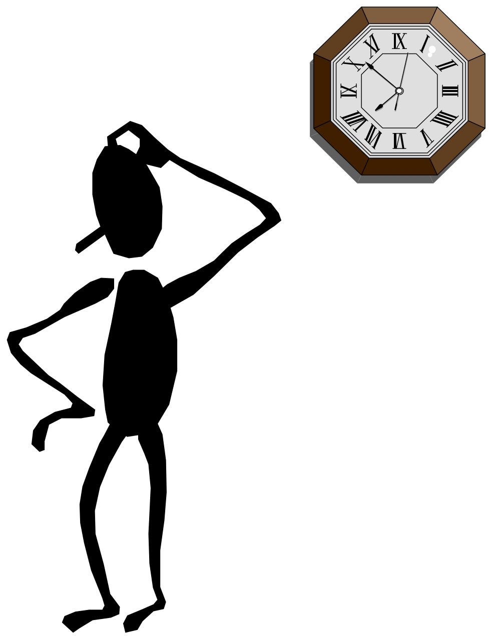


Dynamic range compression

Recruitment phenomenon

Empirical finding:
Reduced dynamic range between threshold of hearing and uncomfortable level

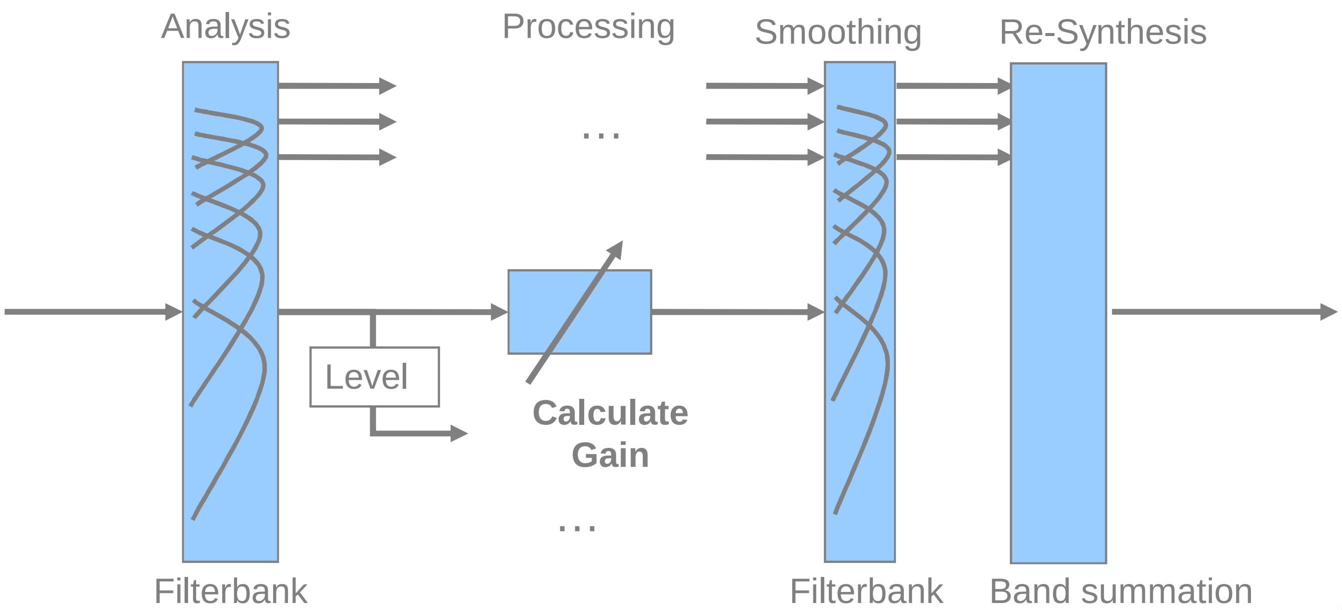




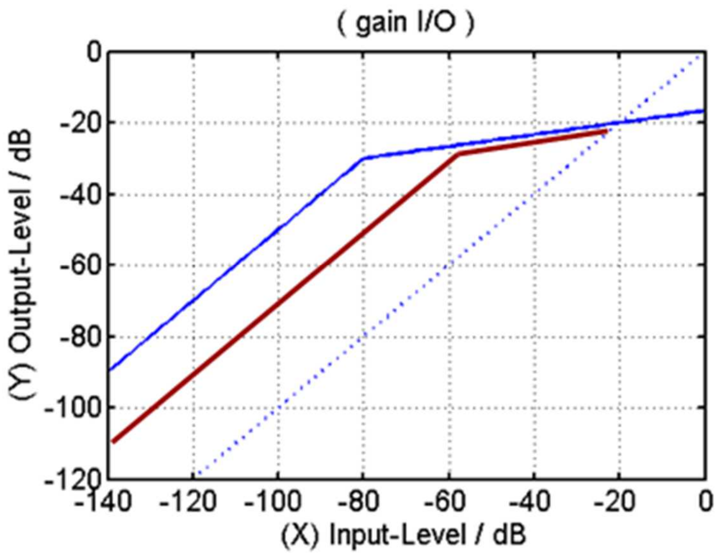
Loud signals are too loud ...

... Soft signals are too soft

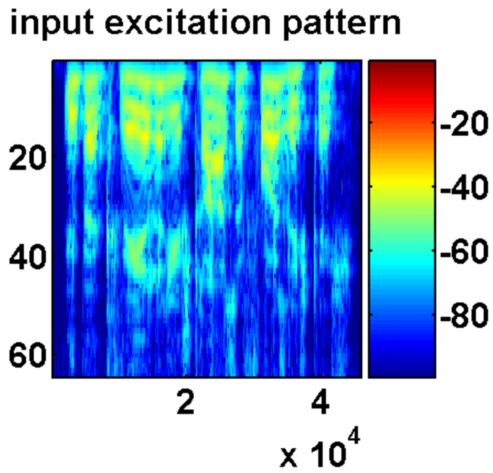
Multichannel dynamic range compression

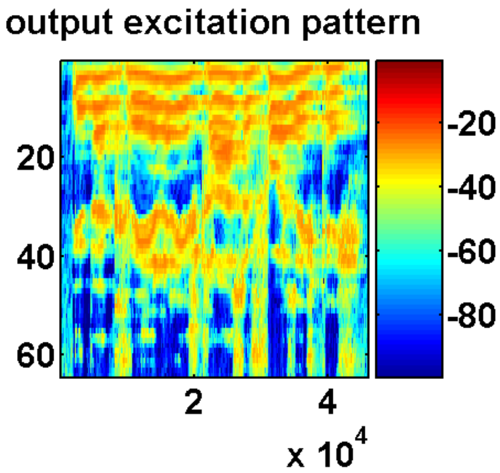


- Limited success with many bands and very short time constants

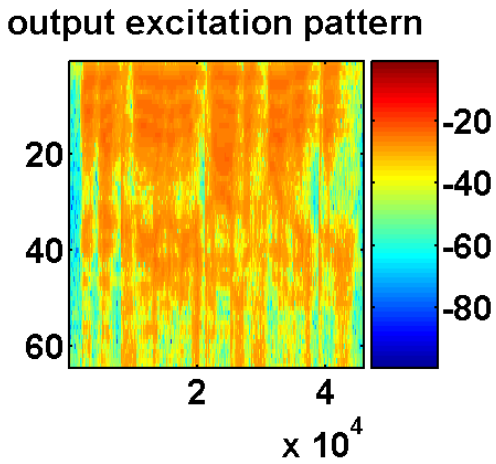


Multichannel dynamic range compression





without IF-control:



- Instantaneous compression including suppression model (instantaneous-frequency (IF) control)
- Gain and compression applied independently in frequency channels flattens spectro-temporal pattern
- Non-linear processing sharpens spectro-temporal pattern

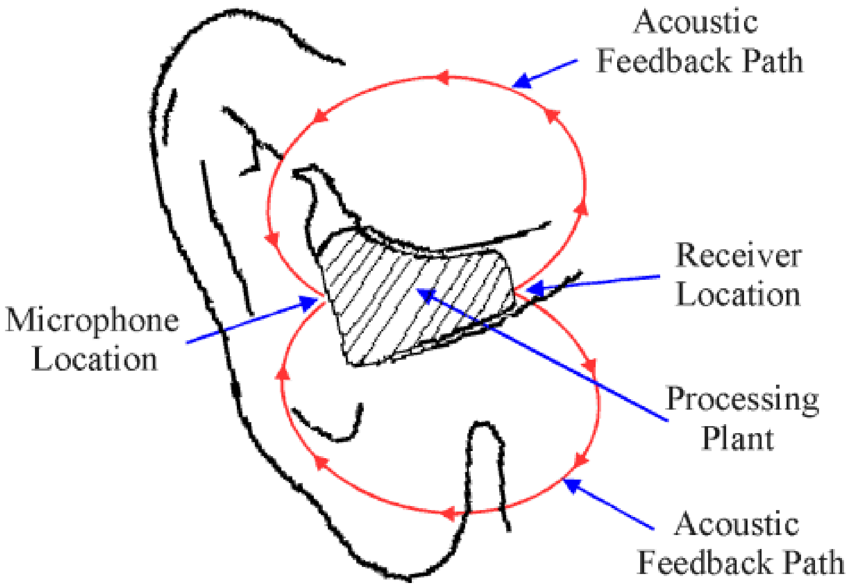
Feedback cancellation

Acoustic feedback

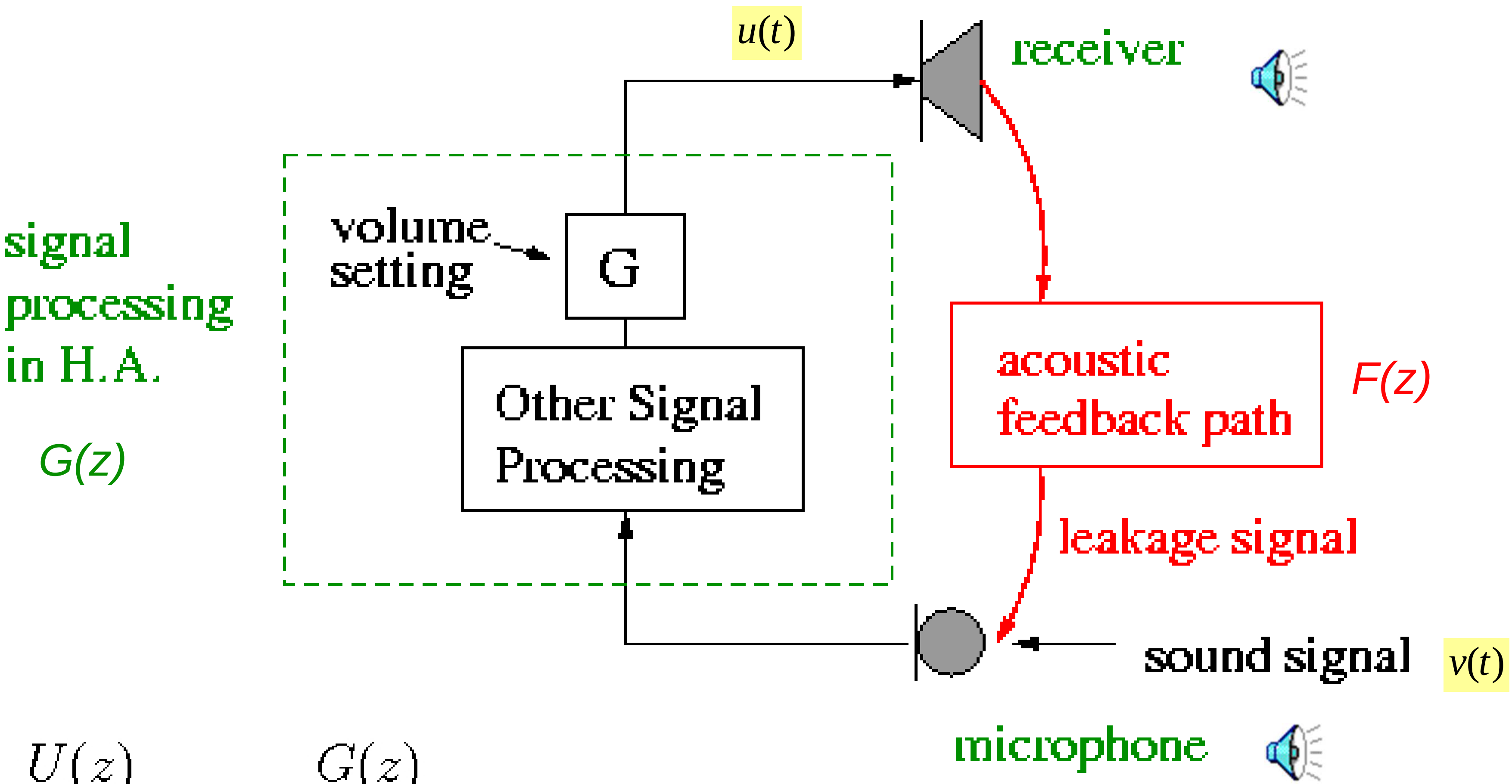
- Amplification of recorded signal needed
- BUT: ringing/howling when amplification is increased above certain limit
- REASON: acoustic coupling between receiver and microphone

  ➡ Acoustic Feedback

- Acoustic feedback limits maximum amplification in hearing aids
  (even more problematic in open-fitting hearing aids)



Acoustic Feedback: illustration



$$\frac{U(z)}{V(z)} = \frac{G(z)}{1-G(z)F(z)}$$

Acoustic feedback cancellation: approaches

- **Notch Filters: traditional solution**

  Suppress the narrow-band oscillations that originate from system instability (when such instability occurs)

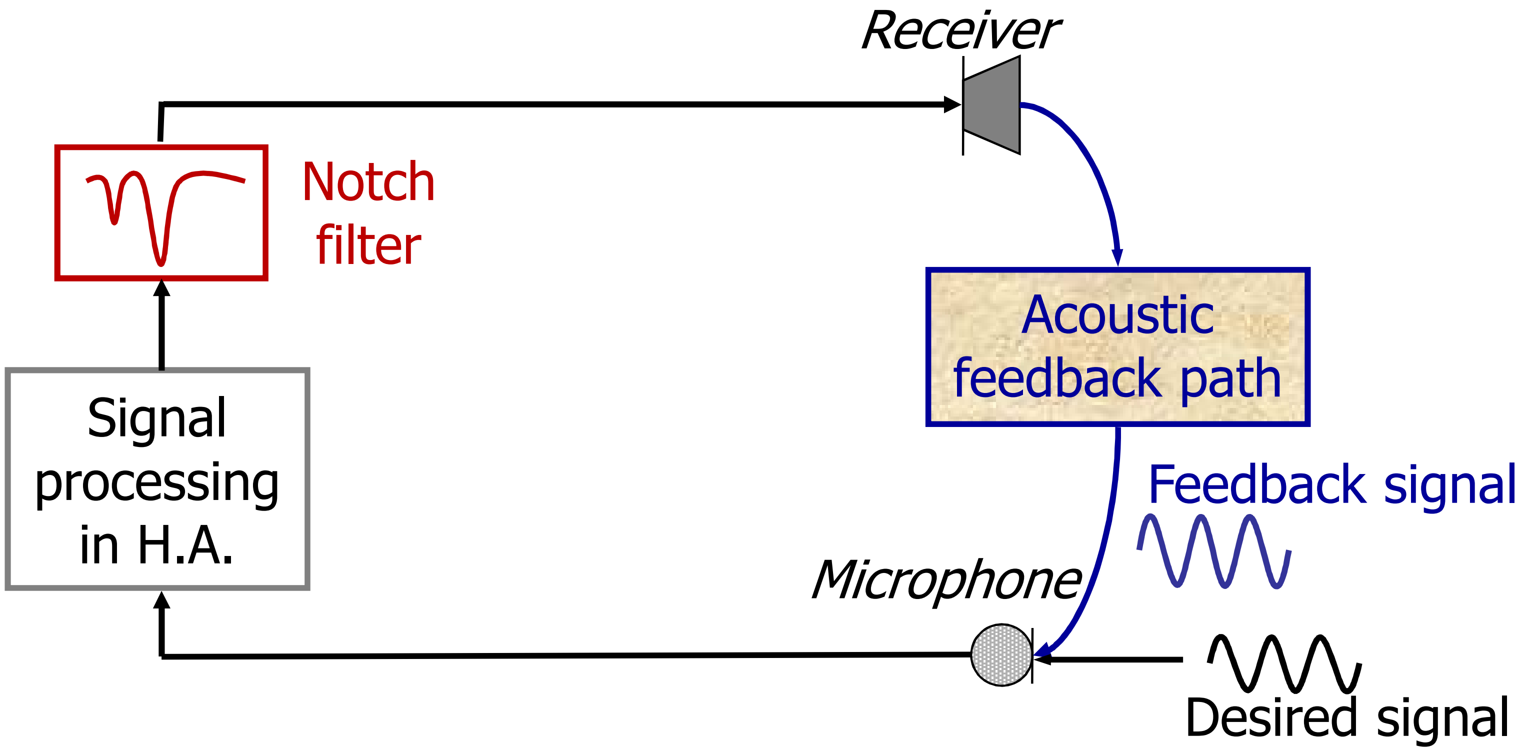
  - Self-adjusting notch filters
  - Adaptive notch filters

- **Adaptive Feedback Cancellation:**

  Estimate and cancel feedback signal by recursively identifying and tracking the unknown feedback path transfer function *F(z)*

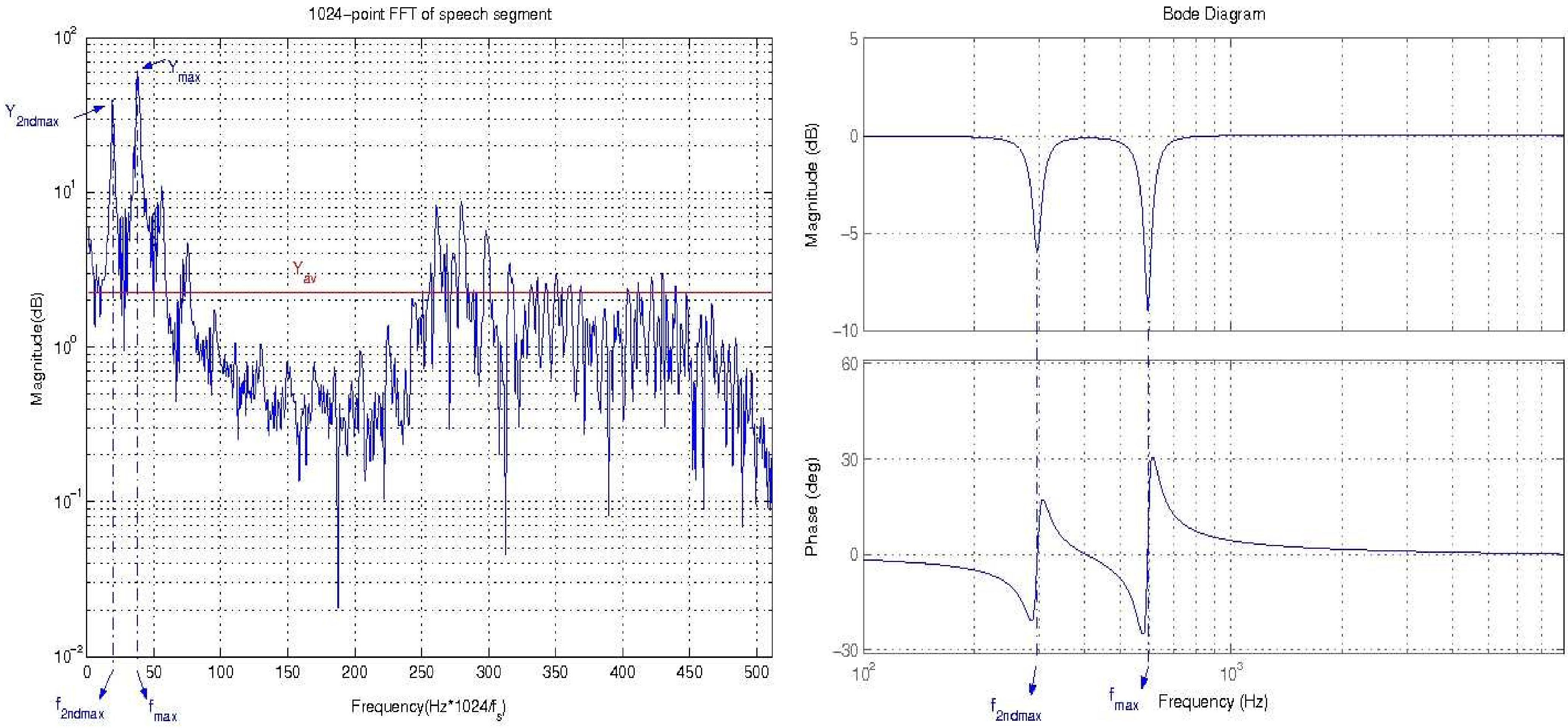
Notch filtering

Notch filtering: detect and attenuate frequencies where instability occurs



- Reactive approach → always too late!
- Amplification is still limited
- Hearing aid response is compromised

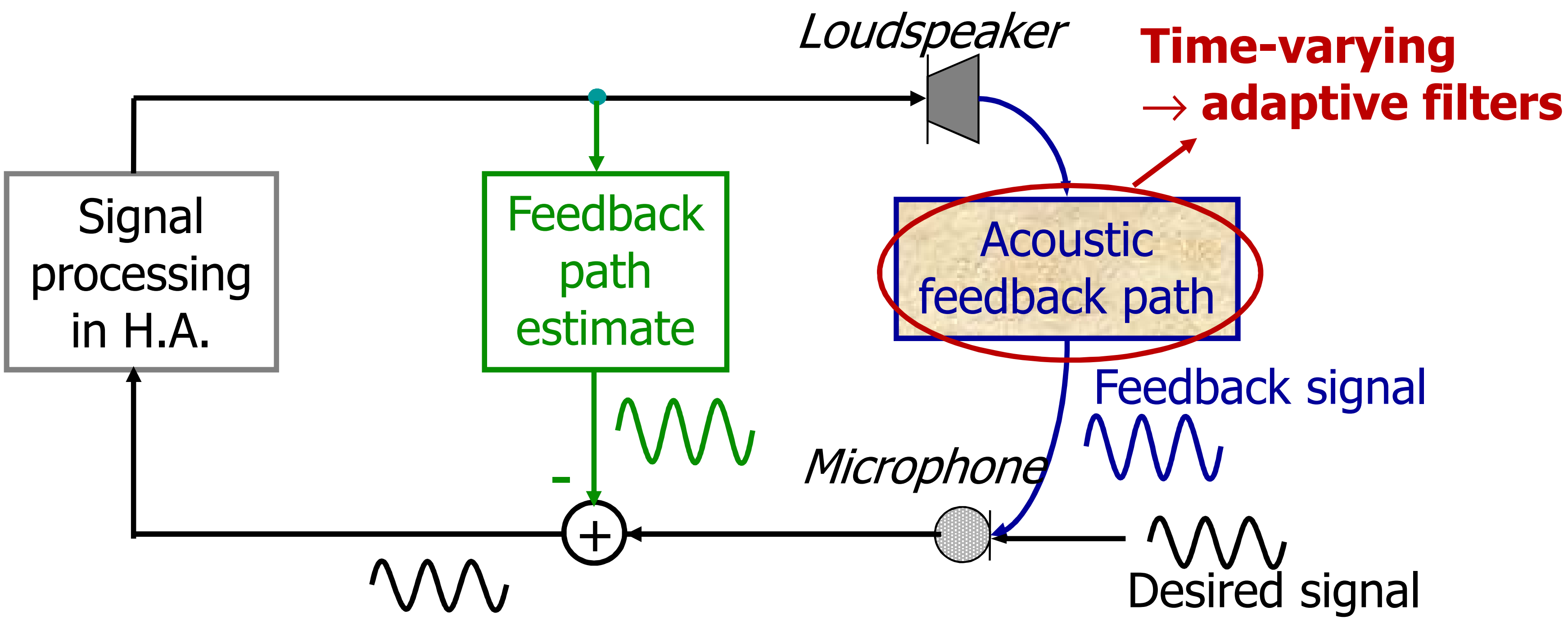
Notch filtering



Microphone signal analysis → Feedback detection → Notch Filter design

Adaptive Feedback cancellation

More promising solution? **Adaptive Feedback cancellation**



Model the leakage signal and subtract it from the microphone signal ➡ increases maximum amplification

Adaptive Feedback cancellation

Due to signal correlation, decorrelation is required, e,g, by

✓ injecting noise signal *r(t)*, possibly psycho-acoustically masked

✓ adding a delay *d* to the forward path:

$$\tilde{e}(t) = e(t-d)$$

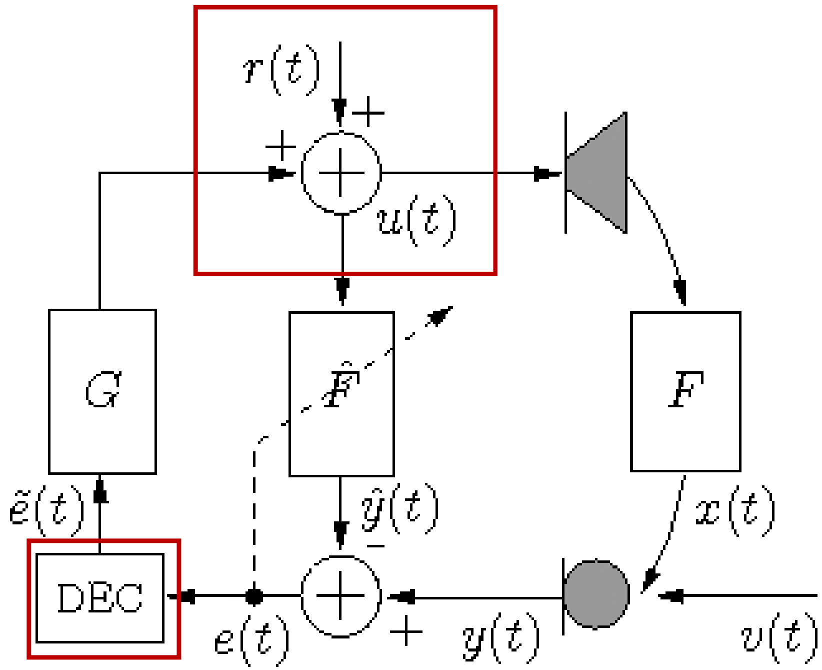
Note: if v(t)=white noise, then d=1 is sufficient !

✓adding a nonlinear operation to the forward path:

– frequency shift

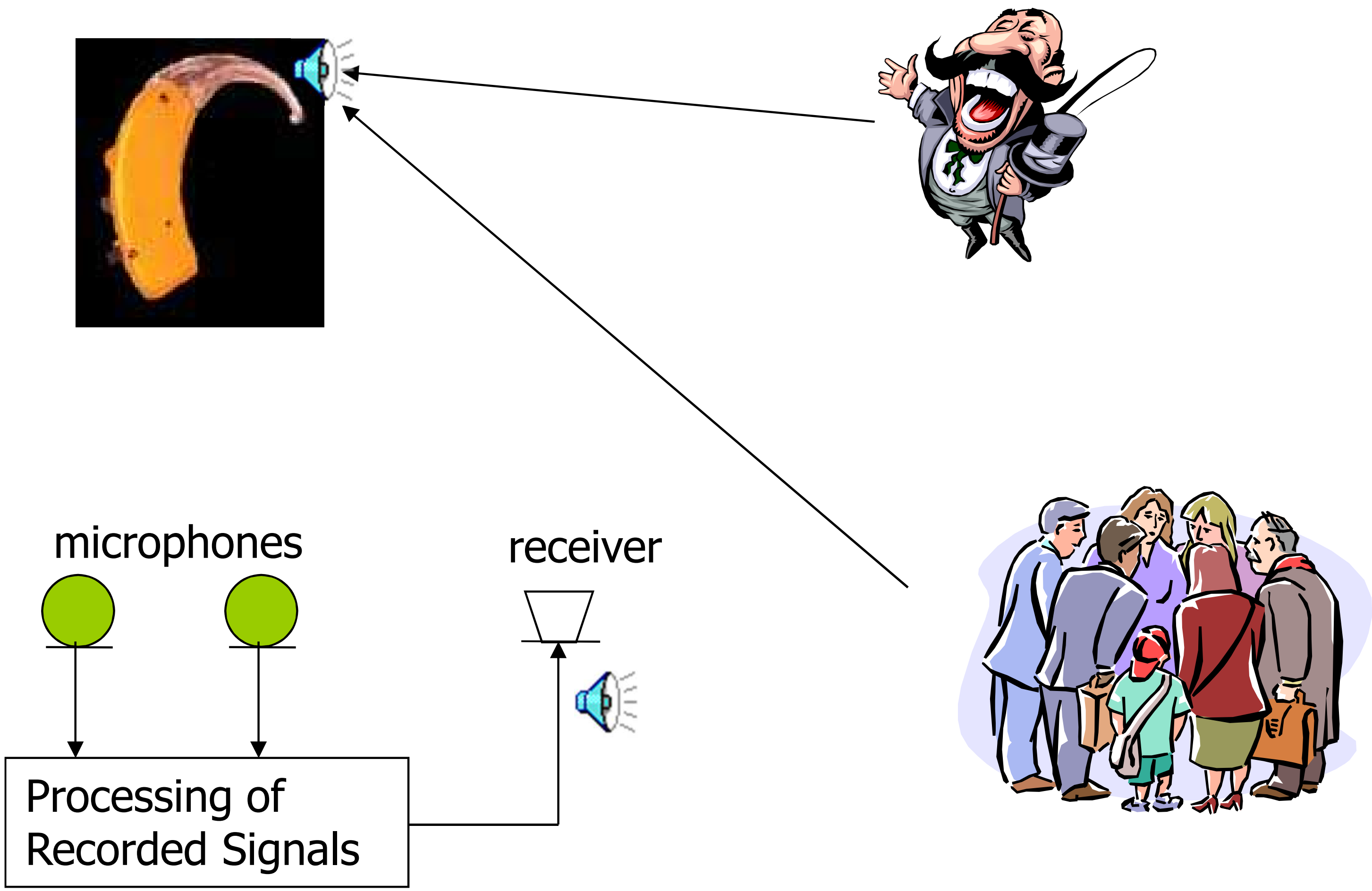
– phase modulation

– half wave rectifier: $\tilde{e}(t) = e(t) + \alpha(e(t) + |e(t)|)$



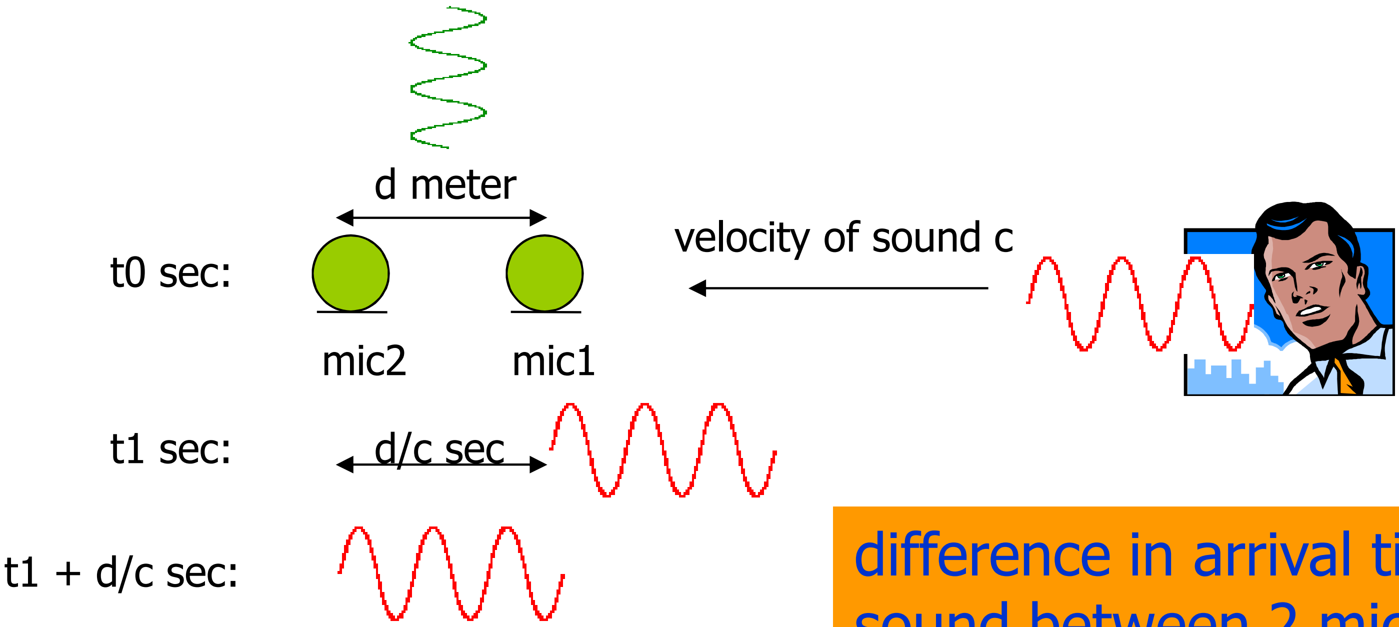
Noise reduction

Speech intelligibility in background noise



Background noise reduction

- **Goal: increase signal-to-noise ratio (SNR)**
- one microphone:
  can only exploit temporal or spectral differences in speech and noise signal
- more than one microphone:
  can also distinguish between signals coming from different positions in space (spatial processing)

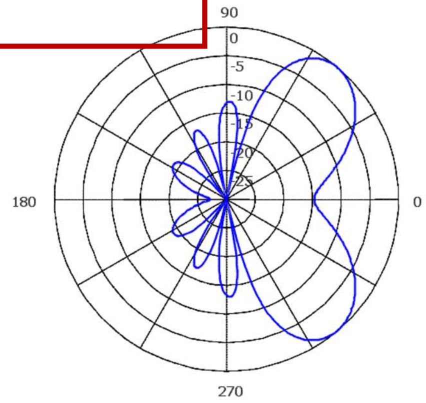


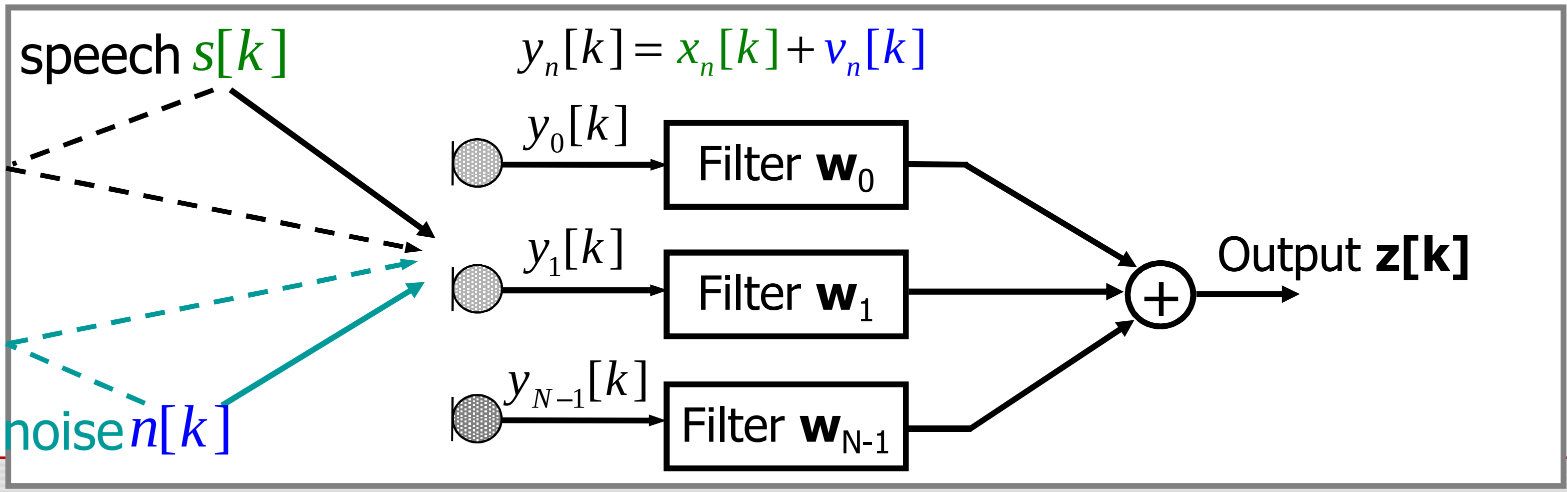
difference in arrival time of sound between 2 mics depends on position of sound source

Background noise reduction

- **Single-microphone techniques:**
  - only temporal and spectral information → limited performance
  - spectral subtraction, Kalman filter, subspace-based

- **Multi-microphone techniques:**
  - exploit spatial information
  - *Fixed beamforming*: fixed directivity pattern
  - *Adaptive beamforming*: adapt to different acoustic environments → improved performance

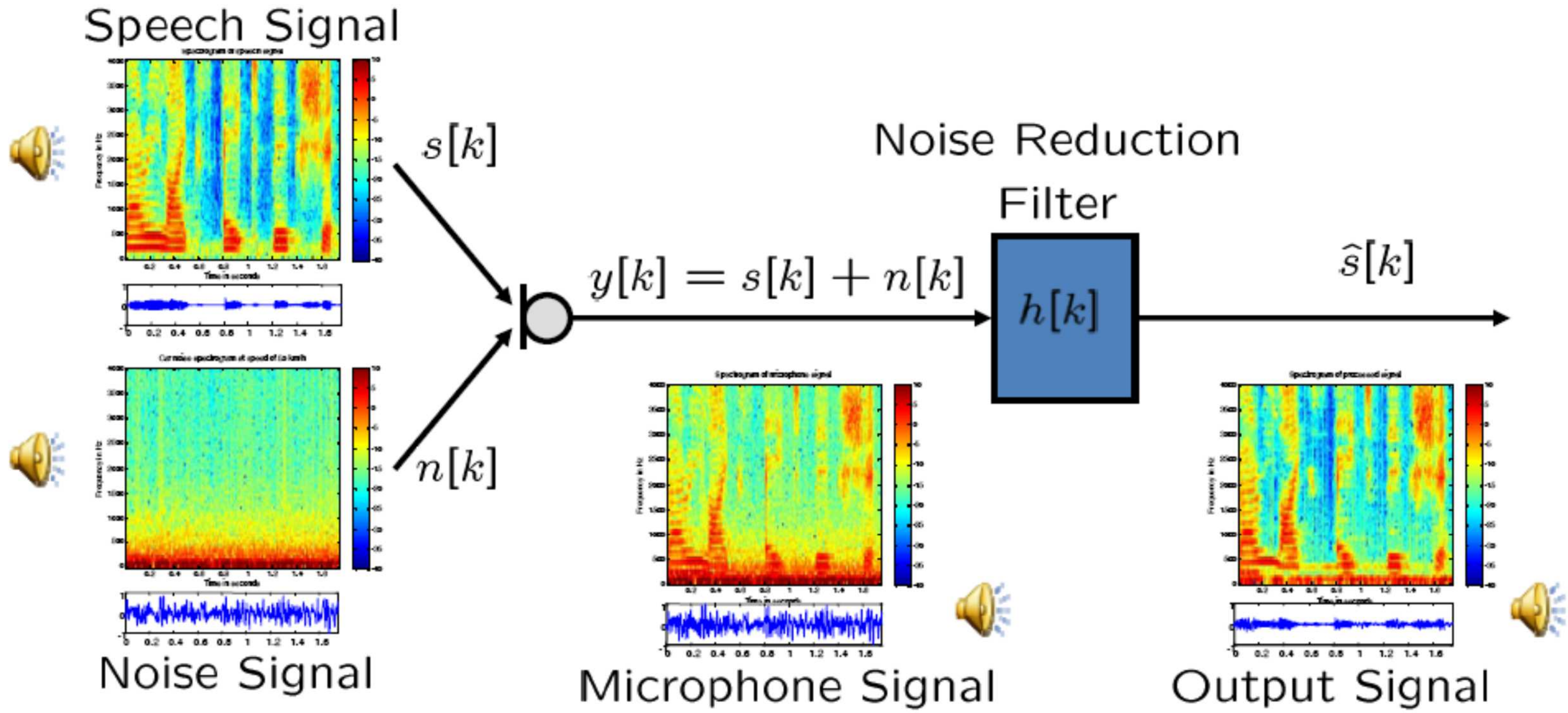




Single-microphone noise reduction

Single-Channel Noise Reduction

- The desired signal $s[k]$ has to be calculated from the microphone signal $y[k]$ which contains a mixture of desired signal and (ambient) noise $n[k]$.
  - Problem: Desired signal and noise may overlap in time, frequency and/or space.



Single-Microphone Noise Reduction

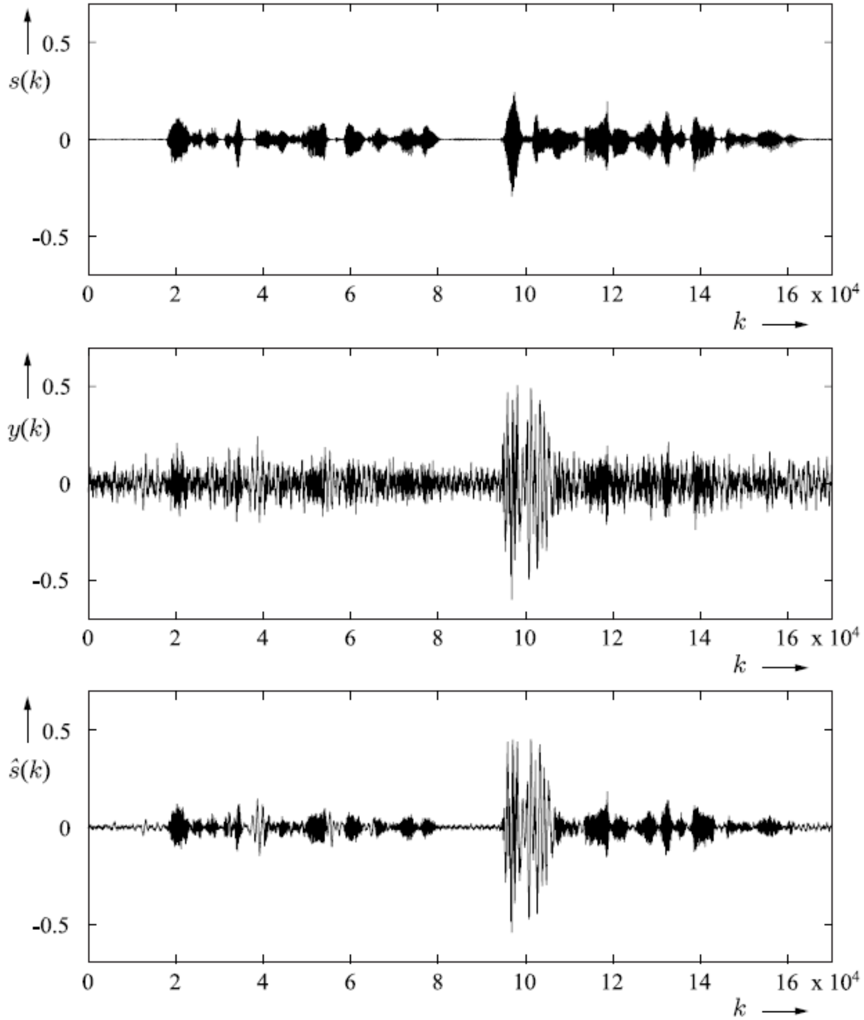


Figure 11.1: Time domain waveforms of a clean (top), a noisy (middle), and an enhanced (bottom) speech signal

$$\boxed{y[k] = s[k] + n[k]}$$

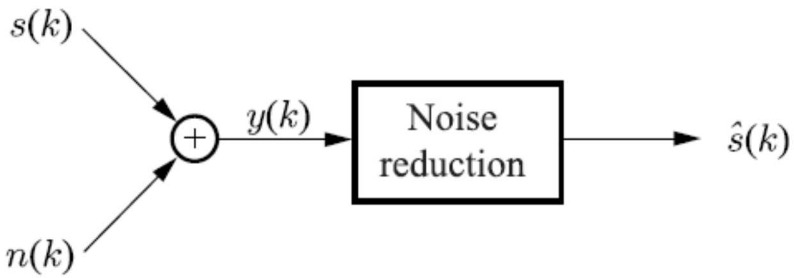


Figure 11.2: The noise reduction filter

Single-Microphone Noise Reduction

- STFT-based techniques (overlap-add)

$$Y(e^{j\Omega}) = \sum_{k=-\infty}^{\infty} y[k] e^{-jk\Omega} \longrightarrow Y[k,l] = \sum_{n=0}^{N-1} w[n] y[n+lL] e^{-j2\pi nk/N}, k = 0 \ldots N-1$$

frequency frame

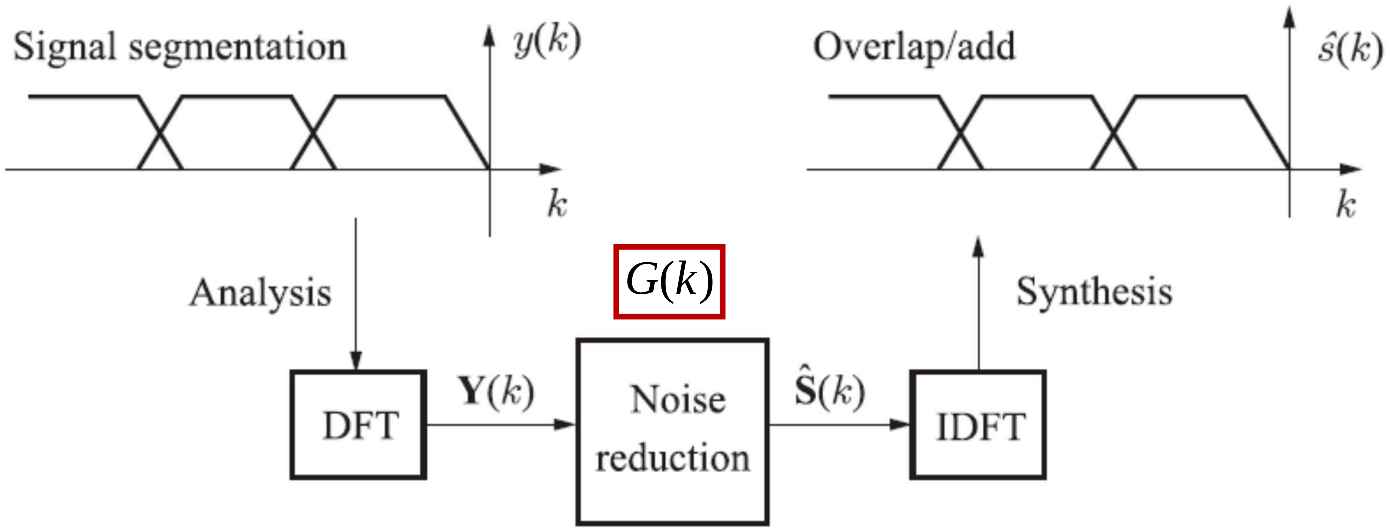


Figure 11.5: DFT domain implementation of the noise reduction filter

Single-Microphone Noise Reduction

- Noisy microphone signal: $Y[k,l] = S[k,l] + N[k,l]$
- Average noise PSD (stationary noise assumption):

$$\sigma_n^2[l] = \frac{1}{M} \sum_{M \text{ noise frames}} \left| N[k,l] \right|^2$$

$\rightarrow$ Estimate clean speech spectrum *S[k,l]* (for each frame), using noisy speech spectrum *Y[k,l]* (for each frame, i.e. short-time estimate) + estimated average noise PSD $\sigma_n^2[l]$:

based on **real-valued gain function:** $\boxed{\hat{S}[k,l] = G[k,l]\, Y[k,l]}$

$$G[k,l] = f(Y[k,l], \sigma_n^2[l])$$

Spectral Enhancement: Gain Functions

- Example: **Wiener Filter**
  - Goal:
    find filter *G[k,l]* such that MSE is minimized : $E\left\{\left|S[k,l]-G[k,l].Y[k,l]\right|^2\right\}$
  - Solution:

$$G[k,l]=\frac{E\left\{Y[k,l].S^*[k,l]\right\}}{E\left\{Y[k,l].Y^*[k,l]\right\}}=\frac{P_{sy}[k,l]}{P_{yy}[k,l]}$$

<- **cross-correlation** in l-th frame

<- **auto-correlation** in l-th frame

Assuming that speech *s[k]* and noise *n[k]* are uncorrelated, then...

$$G[k,l]=\frac{P_{ss}[k,l]}{P_{yy}[k,l]}=\frac{P_{yy}[k,l]-P_{nn}[k,l]}{P_{yy}[k,l]}=1-\frac{P_{nn}[k,l]}{P_{yy}[k,l]}=1-\frac{\sigma_n^2[l]}{\left|Y(k,l)\right|^2}$$

SNR high $\rightarrow$ G[k,l] $\approx$ 1
SNR low $\rightarrow$ G[k,l] $\approx$ 0

Spectral Enhancement: Gain Functions

- Example: **Wiener Filter**

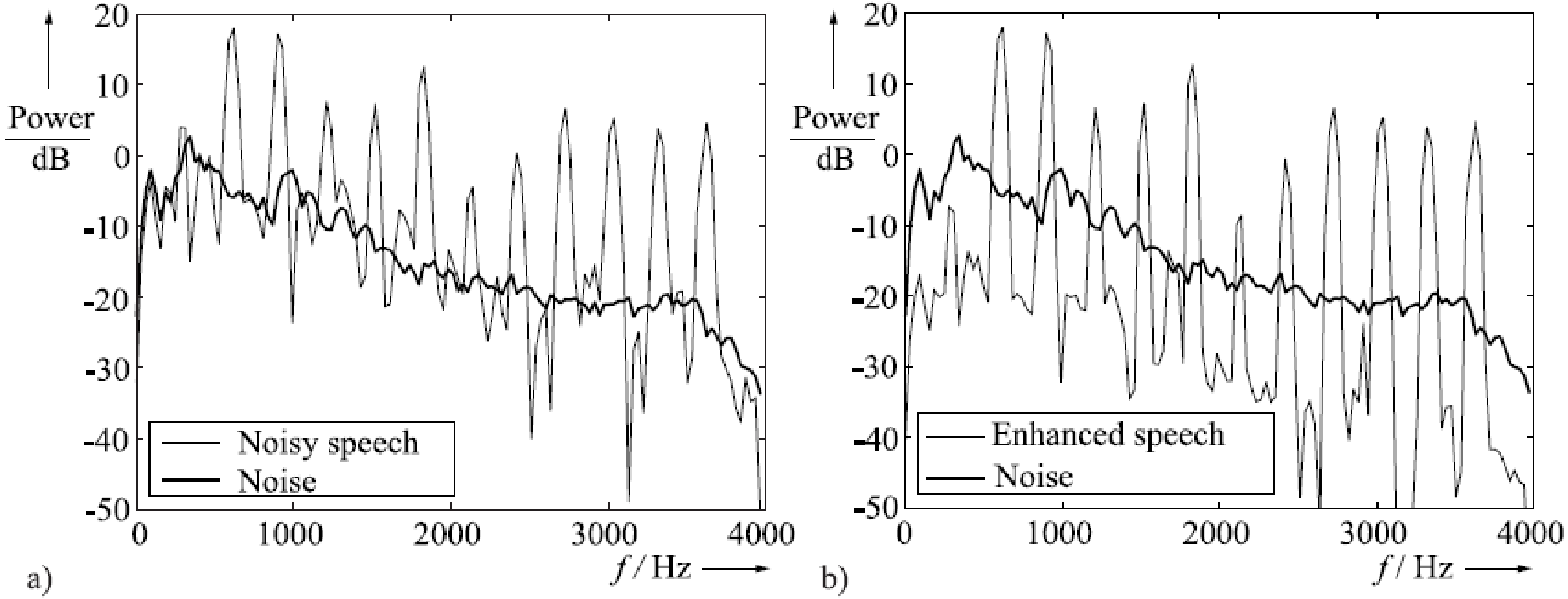


Figure 11.6: Principle of DFT-based noise reduction

a) Short-time spectrum of noisy signal and the estimated noise PSD

b) Short-time spectrum of the enhanced signal and the estimated noise PSD

Spectral Enhancement: Musical Noise

- Audio demo: **car noise**

$$y[k] \longrightarrow \boxed{\text{Wiener filter}} \longrightarrow \hat{s}[k]$$

- Artifact: musical noise
  - Estimation errors in the frequency-domain: usage (subtraction) of **average noise PSD** $\sigma_n^2[l]$ with **short-time estimates** Y[k,l]

    $\rightarrow$ randomly fluctuating noise floor
    $\rightarrow$ spurious peaks in spectral representation of the enhanced signal
    $\rightarrow$ statistical analysis shows that broadband noise is transformed into signal composed of **short-lived tones with randomly distributed frequencies (= musical noise)**

$$\hat{S}[k,l] = G[k,l]\,Y[k,l]$$

average

instantaneous

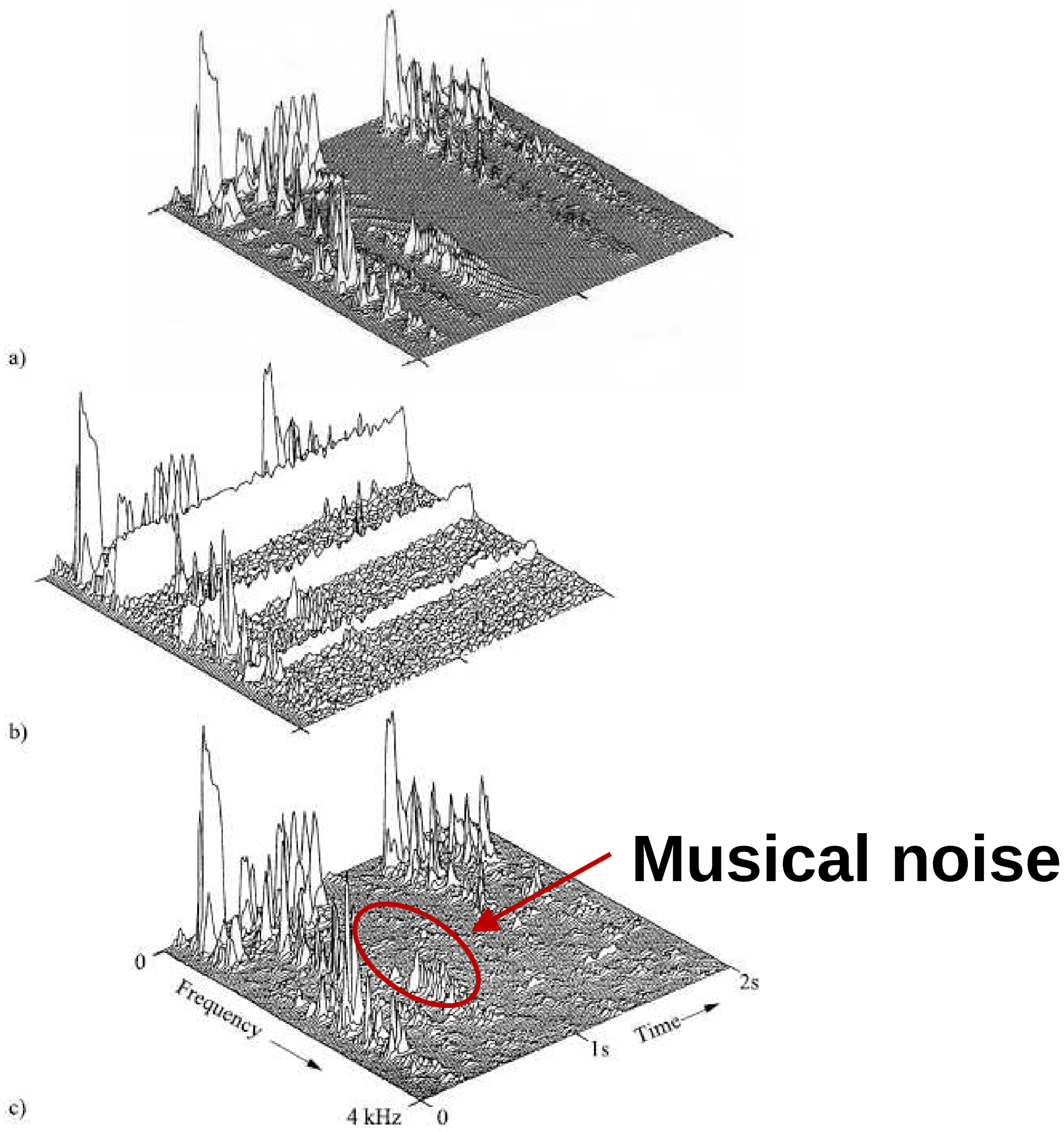


Figure 11.8: Short-term magnitude spectra vs. time and frequency
a) of a clean speech signal,
b) of the clean signal with additive white noise and harmonic tones,
c) of the enhanced signal using magnitude subtraction

Spectral Enhancement: Musical Noise

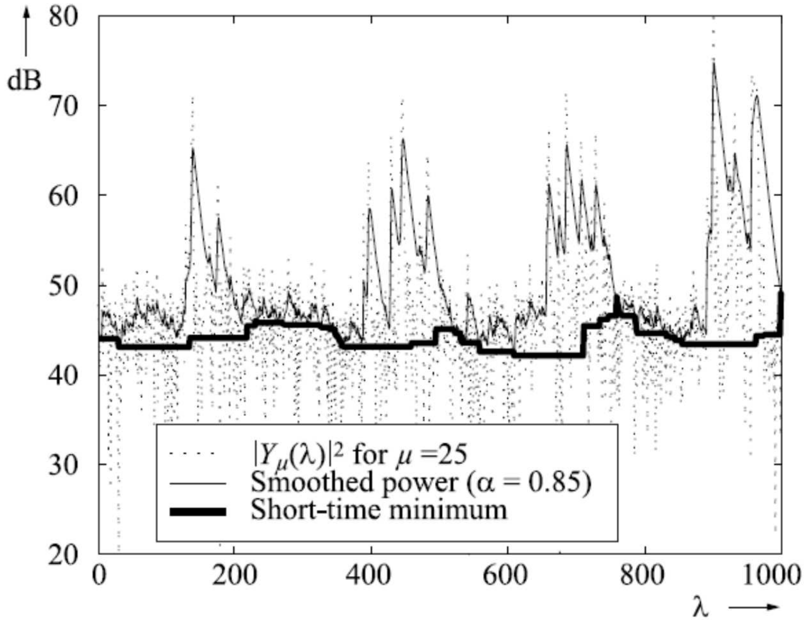
- Counter-measures:
  - Half-wave rectification: put negative values of G[k,l] to 0
  - Better suppression rules: e.g. Ephraim-Malah suppression rule
  - Magnitude averaging: replace *Y[k,l]* in calculation of *G[k,l]* by a local average over frames
  - Noise over-subtraction: increase the estimated noise PSD in order to reduce the amplitude of the random spectral peaks

$$\sigma_n^2[l] \rightarrow O\,\sigma_n^2[l], \quad \text{with } O = 1 \ldots 2$$

  - Spectral floor: impose lower limit $\beta\sigma_n^2[l]$ on magnitude squared enhanced DFT coefficients (trade-off noise reduction vs. musical noise, ß = 0.1...0.4)
  - Cepstral smoothing

Noise PSD estimation

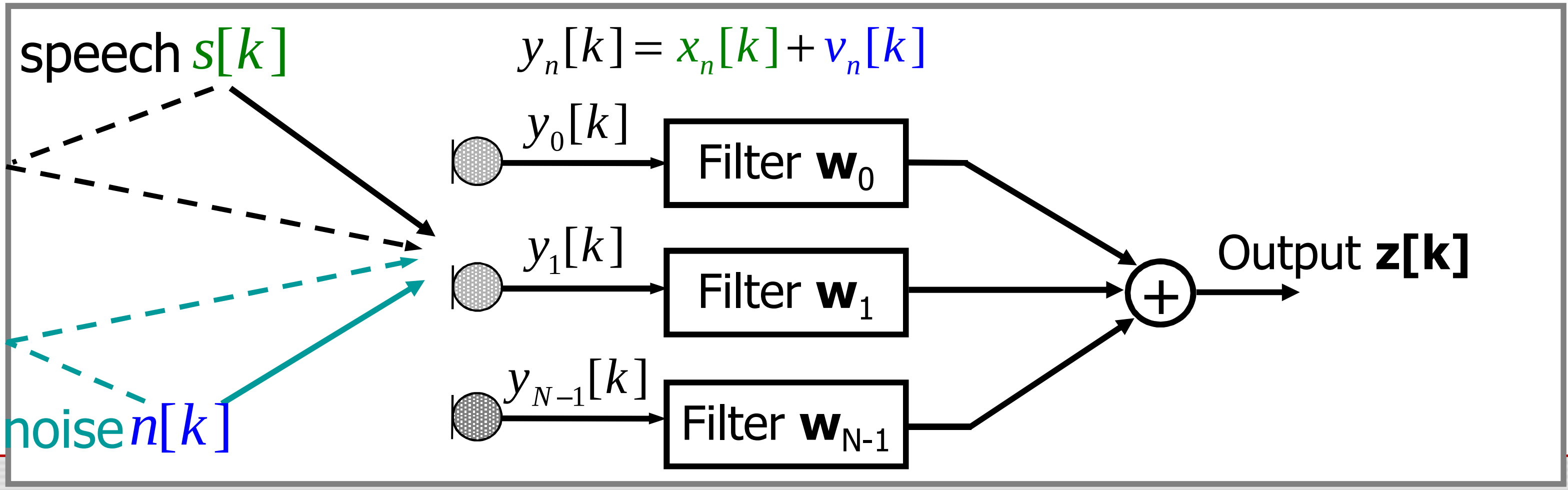
- **Noise PSD is generally** time-varying **and** not known a-priori
- Estimation of average noise PSD $\sigma_n^2[l]$:
  - Based on **VAD (Voice Activity Detection):**
    - Hard decision between speech and noise
    - sample noise in speech pause prior to speech and keep estimate fixed during speech activity
    - Works well for stationary noise at moderate to high SNRs (above 0 dB)
  - Based on **„Minimum Statistics“:**
    - Soft-decision
    - Relies on observation that power of noisy speech signal frequently decays to power level of disturbing noise (gaps/dips in speech PSD)
    - Allows to update estimated noise PSD also during speech activity
    - Works better for non-stationary noise



Background noise reduction

- **Single-microphone techniques:**
  - only temporal and spectral information $\rightarrow$ limited performance
  - spectral subtraction, Kalman filter, subspace-based

- **Multi-microphone techniques:**
  - exploit spatial information
  - *Fixed beamforming*: fixed directivity pattern
  - *Adaptive beamforming*: adapt to different acoustic environments $\rightarrow$ improved performance

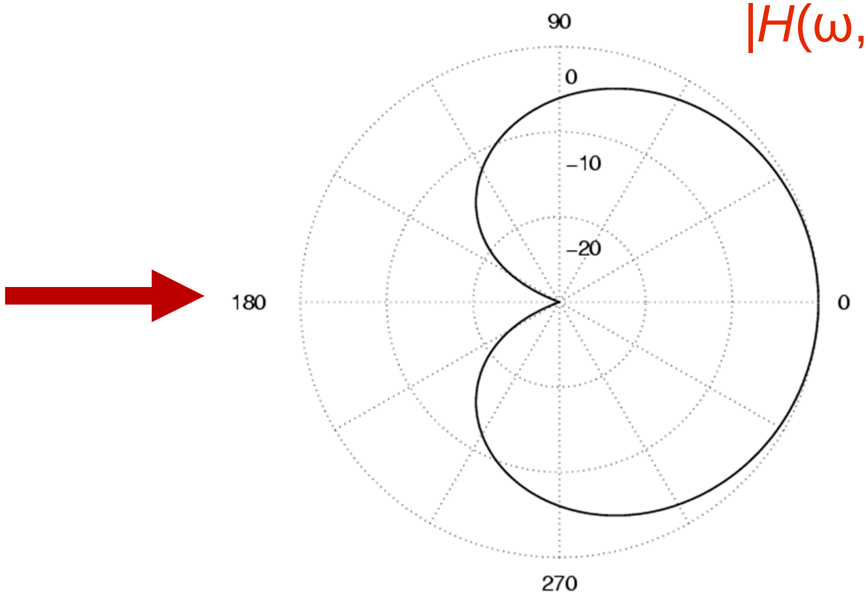


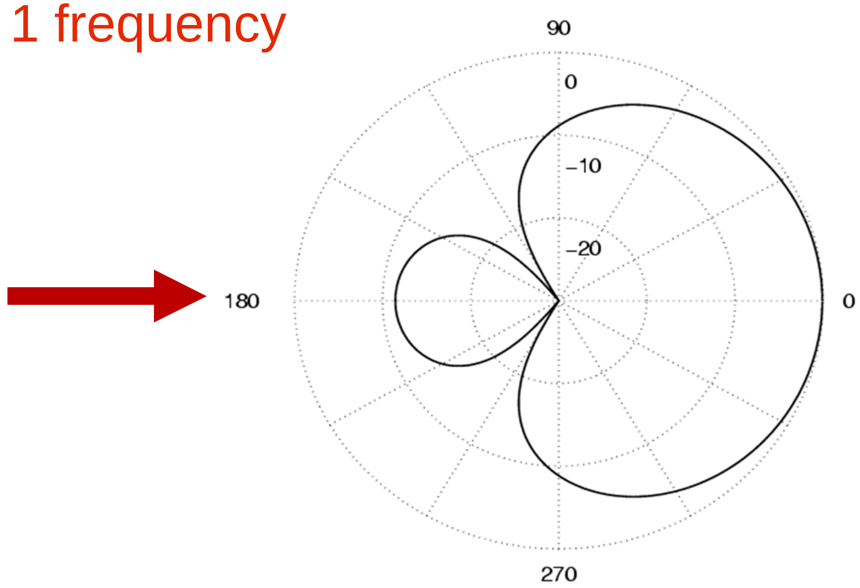
Multi-microphone noise reduction

Introduction: directional microphone

- A (directional) microphone is characterized by a directivity pattern, which specifies the gain (+ phase shift) that the microphone gives to a signal coming from a certain direction θ
- Directivity pattern *H*(ω,θ) is also function of frequency (ω)
- Directivity pattern of directional microphone (e.g. cardioid, supercardioid) is fixed and defined by physical microphone design

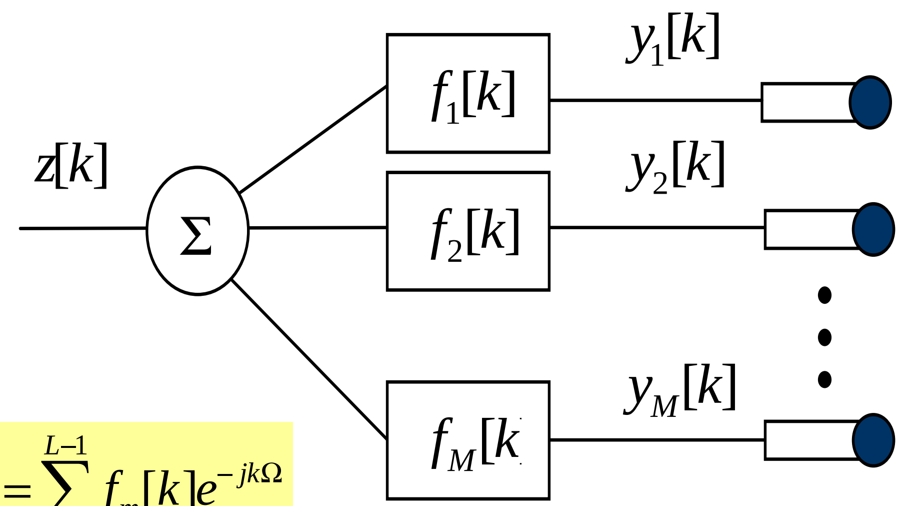
|*H*(ω,θ)| for 1 frequency



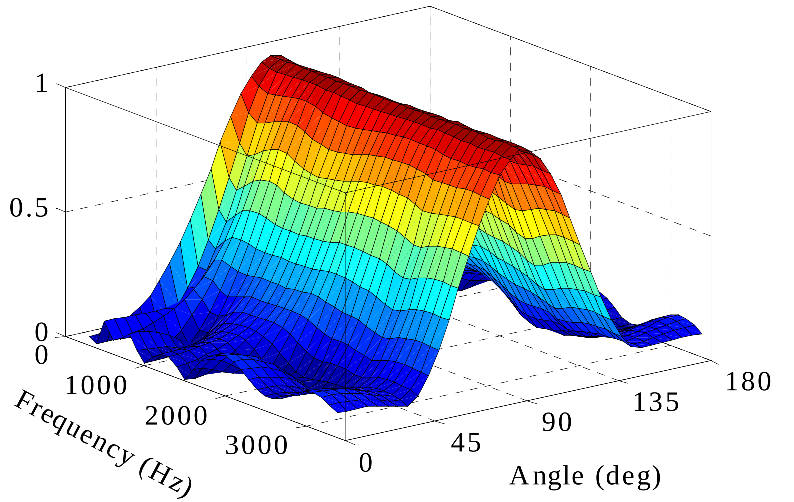


Filter-and-sum beamforming

- By weighting or filtering (= frequency-dependent weighting) + summing the signals from **microphones at different positions**, the aim is to produce a (software-controlled) \`virtual' directivity pattern' (= weighted sum of individual directivity patterns)



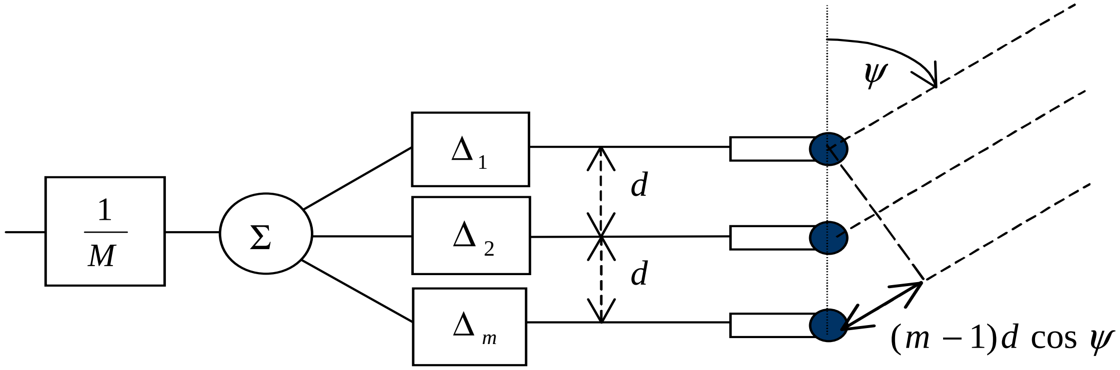
$$F_m(e^{j\Omega}) = \sum_{k=0}^{L-1} f_m[k] e^{-jk\Omega}$$



- This is referred to as \`spatial filtering' and \`spatial filter design', with correspondences to traditional (spectral) filter design

Fixed beamforming: delay-and-sum beamforming

- Principle: Microphone signals are delayed and then summed together



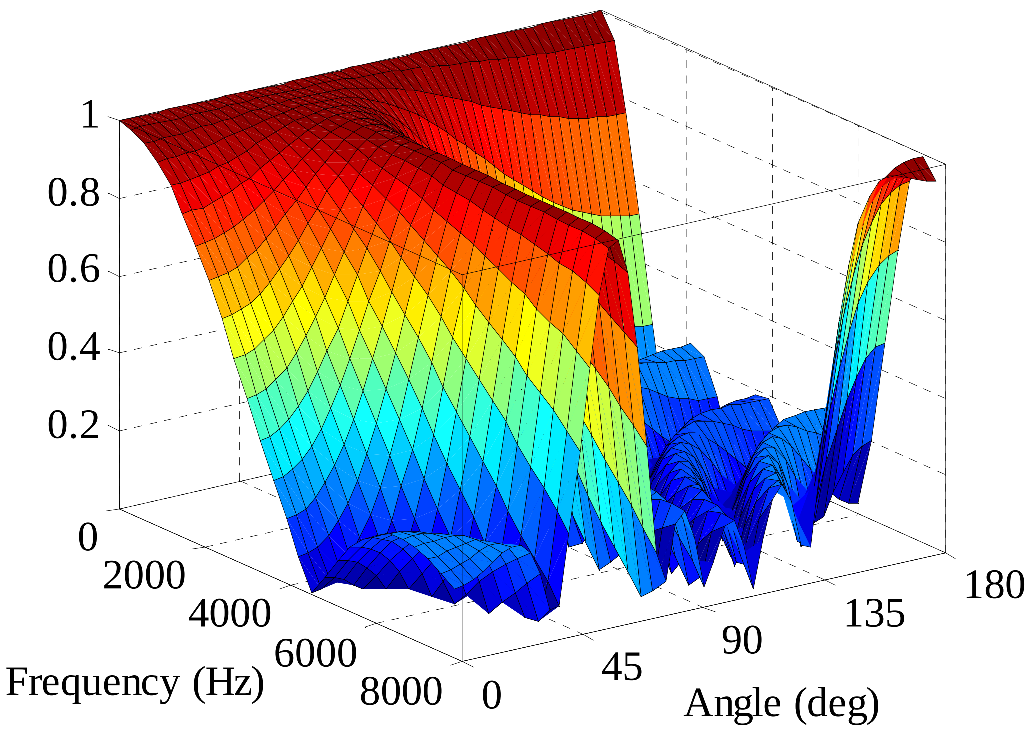
$$z[k] = \frac{1}{M} \cdot \sum_{m=1}^{M} y_m[k + \Delta_m]$$

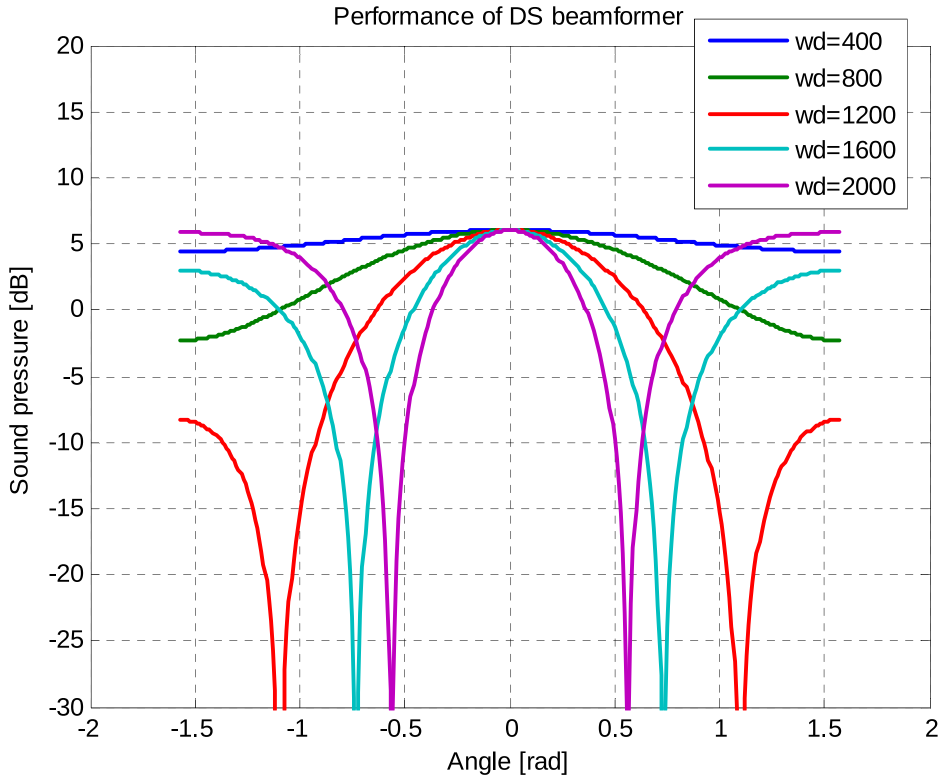
$$F_m(\omega) = \frac{e^{-j\omega\Delta_m}}{M}$$

- Based on **coherent / incoherent interference** *:*
  e.g. for 2 microphones

$$Gain = 2\left(1 + \cos\left(\frac{\omega d \cos\theta}{c}\right)\right)$$

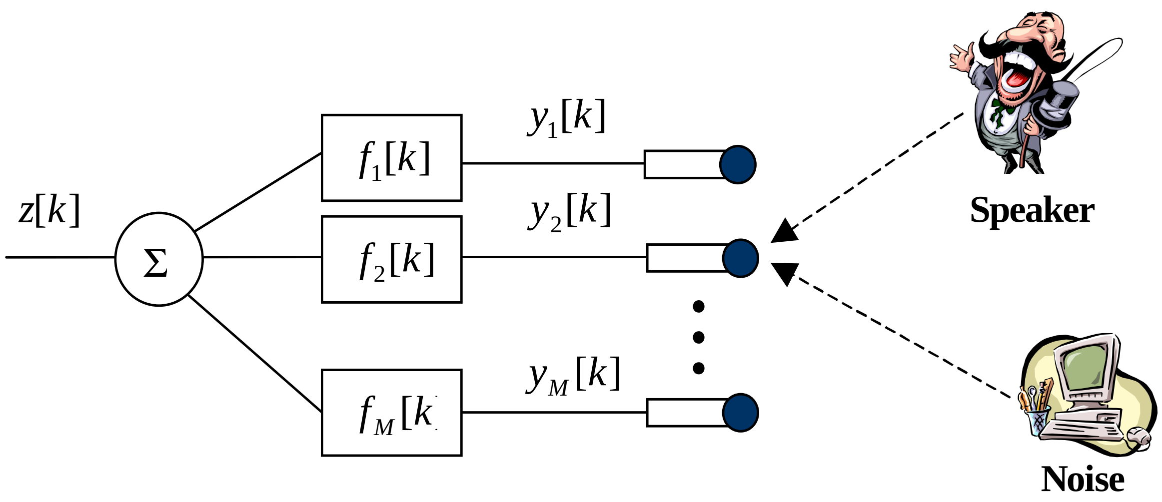
Fixed beamforming: delay-and-sum beamforming

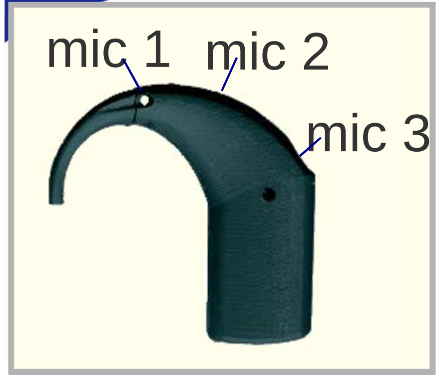




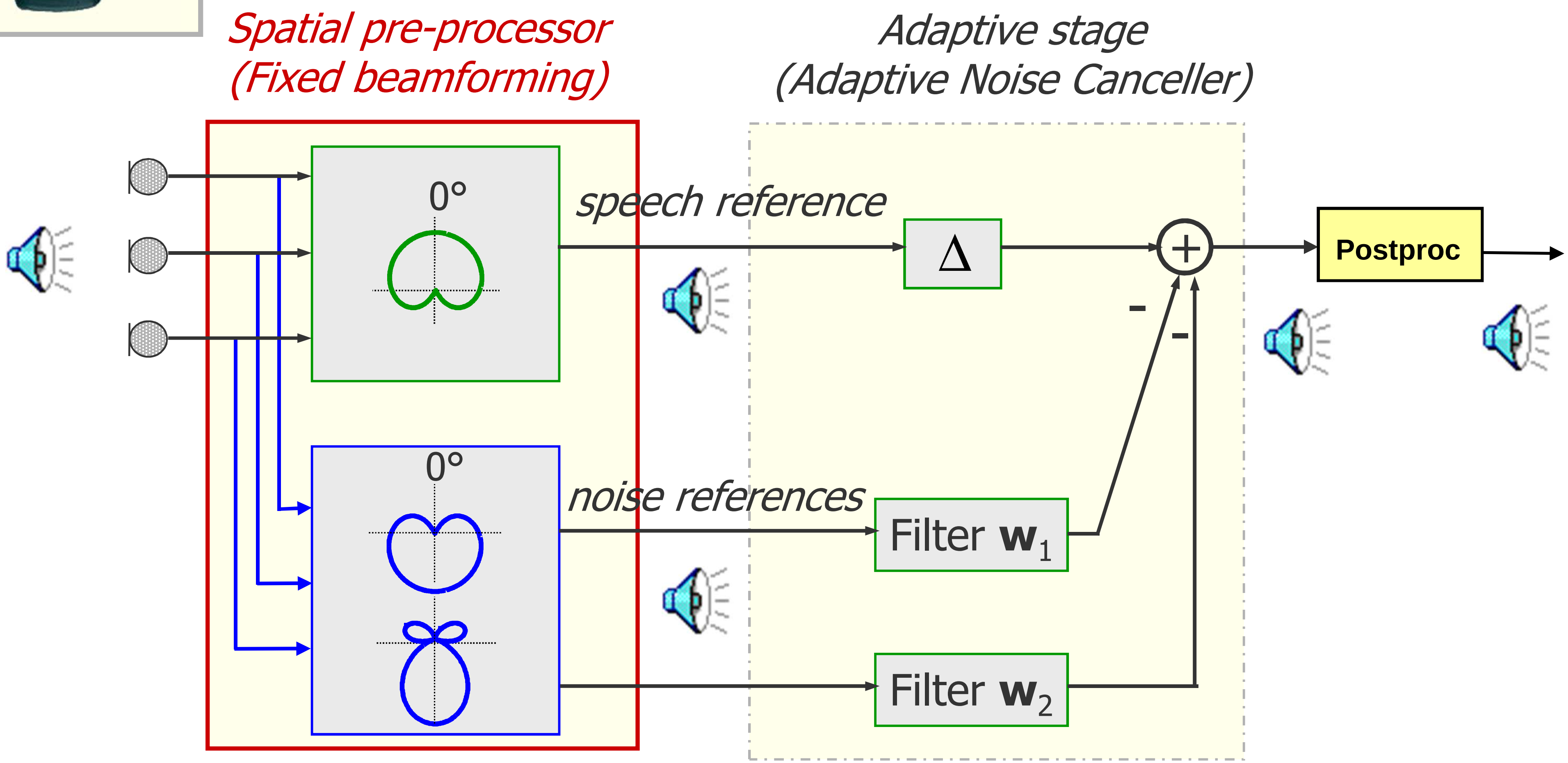
Adaptive beamforming

- **Adaptive filter-and-sum structure:**
  - Aim is to minimize noise output power, while maintaining a chosen frequency response in a given look direction (typically front direction in hearing aids)
  - This is similar to a delay-and-sum beamformer (in white noise), but now the noise field is **unknown** and can change over time
  - Implemented as **adaptive filter** (e.g. constrained LMS algorithm)



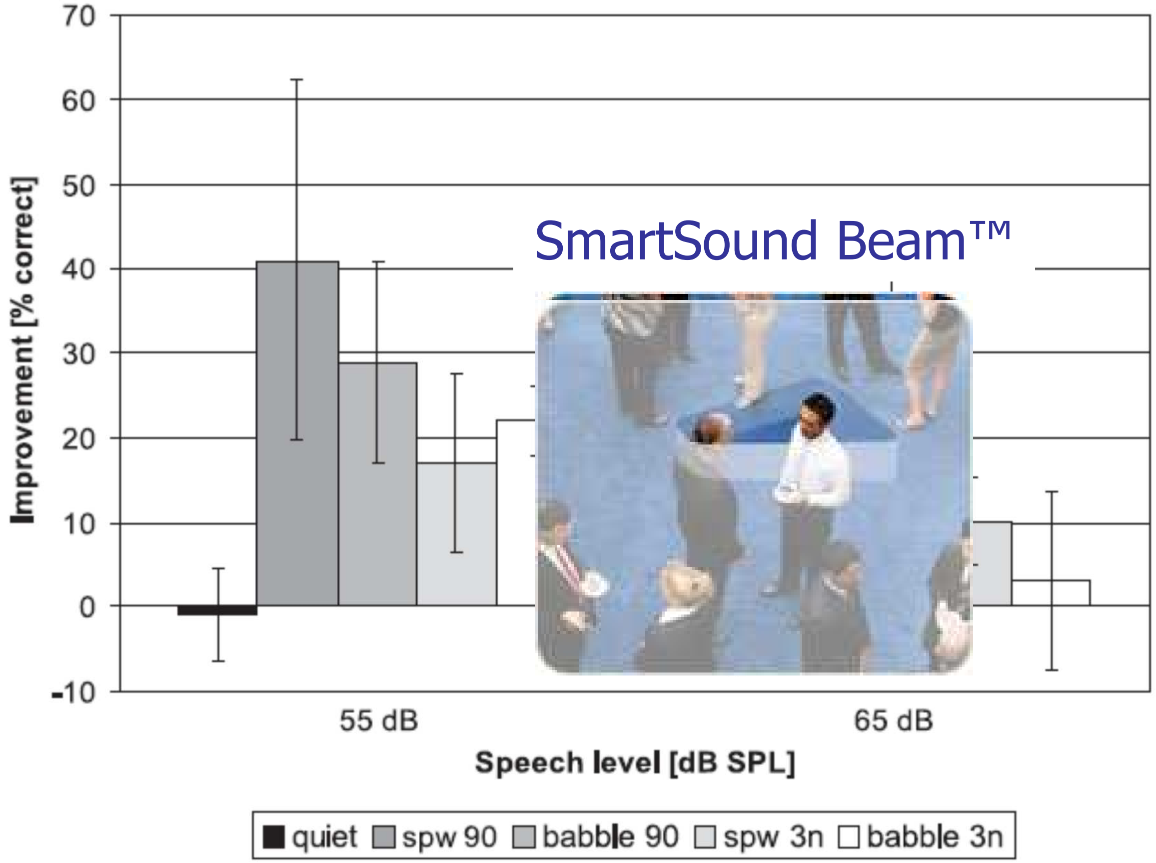


Adaptive beamforming - GSC



Clinical trial

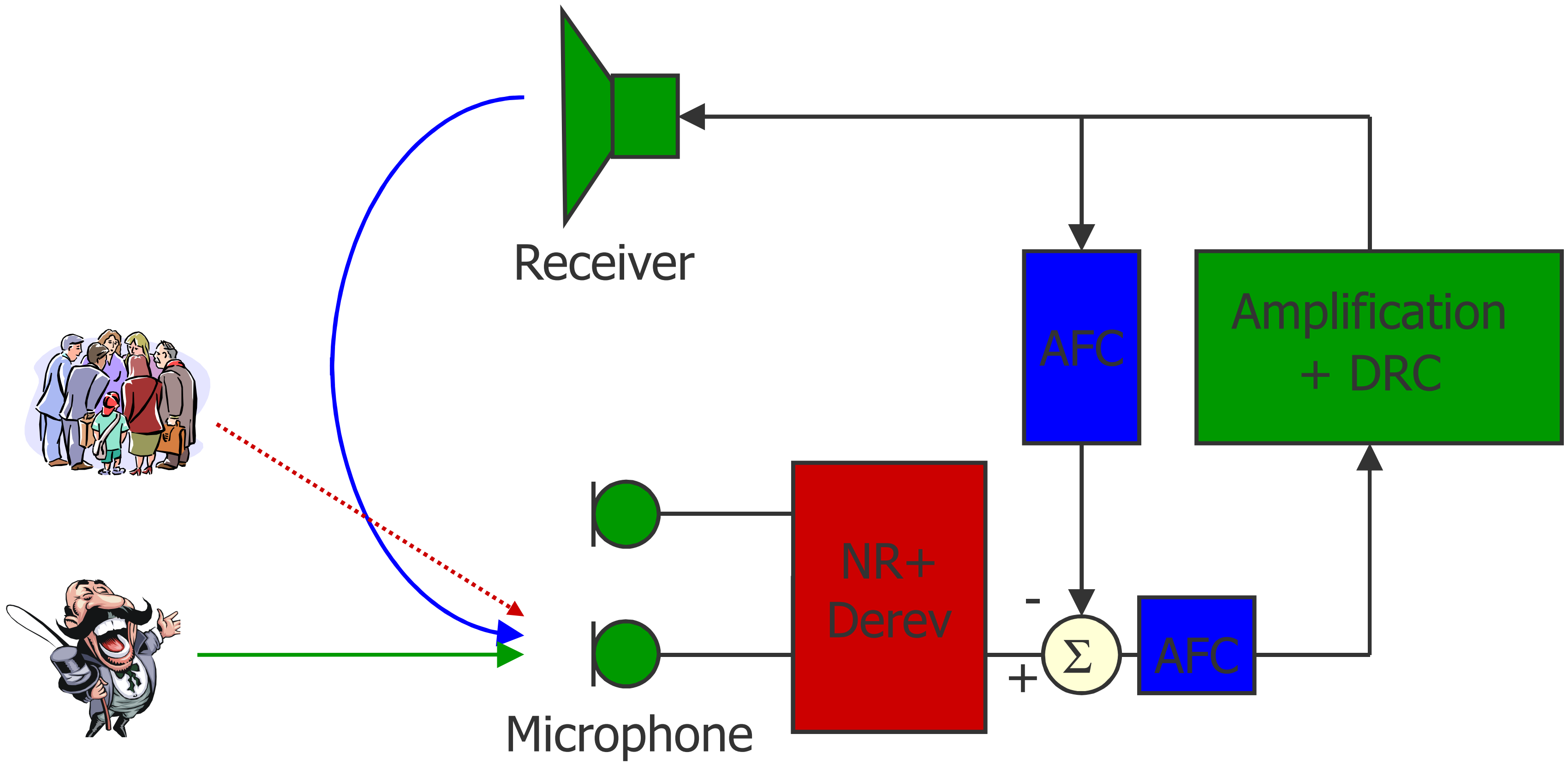
- Implementation on commercial Cochlear Nucleus Freedom device
- 5 CI users, 2 week field test, lab measurement
- Adaptive beamformer vs. fixed directional microphone
- SRT measurements (fixed procedure at SNR = -5dB / +5dB)
- Noise material: stationary speech-weighted (spw) and babble noise: S0N90, S0N90/180/270



Conclusions

- **Single-channel noise reduction**
  - Only <u>spectral filtering</u>
  - can only exploit differences in spectra between speech and noise:
    - noise reduction at expense of speech distortion
    - achievable noise reduction may be limited
    - musical noise
  - Noise PSD estimation is difficult for non-stationary noise
- **Multi-microphone noise reduction:**
  - In addition <u>spatial filtering</u>
  - Can exploit position differences between speech and noise source (also for non-stationary noise)
  - *Fixed beamforming*: fixed directivity pattern
  - *Adaptive beamforming*: adapts to unknown noise fields

- Basic processing: acoustic amplification and dynamic range compression (frequency-selective)
- Due to acoustic coupling between receiver and microphone (large amplification): acoustic feedback control
- Increase speech intelligibility in background noise: single- or multi-microphone noise reduction and dereverberation



Questions ?

House of Hearing, Oldenburg