Listening Enhancement for Mobile Phones

– How to Improve the Intelligibility in a Noisy Environment –

Bastian Sauert and Peter Vary

The Listening Talker Workshop

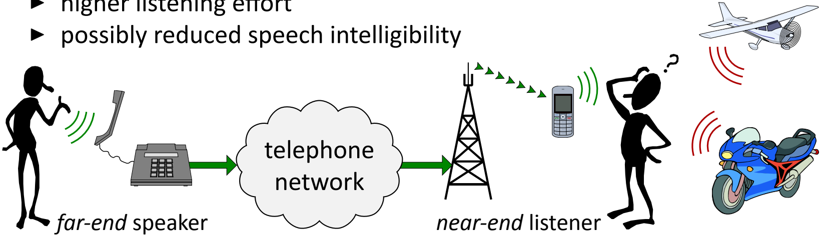
Edinburgh, 3 May 2012

Institute of Communication Systems and Data Processing
Prof. Dr.-Ing. Peter Vary

RWTH AACHEN UNIVERSITY

Introduction: Near-End Listening Enhancement

- **Near-end listener in background noise experiences**
  - higher listening effort
  - possibly reduced speech intelligibility

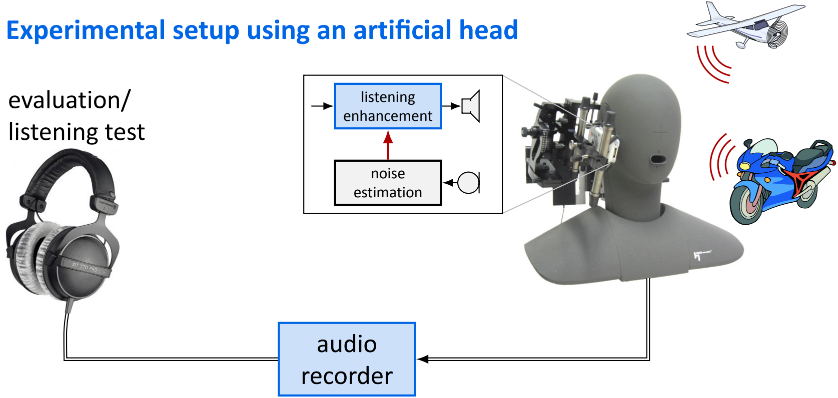


- **Approach**
  - Manipulate clean far-end speech
    depending on local acoustical background noise

- **System constraints**
  - Human ear: psycho-acoustics, damage
  - Thermal load of loudspeaker during playback

Realistic Binaural Evaluation for Telephony Scenario

- Recording of real-world noise environments using an artificial head and a mounted phone
- Experimental setup using an artificial head



Non-Exhaustive Bibliogr., Noise Dependent Processing

- **Energy distribution on phoneme- and/or formant-level**
  - Kretsinger and Young 1960
  - Thomas, Niederjohn, et al. 1968–1972
  - Skowronski and Harris 2002–2006
  - Yoo, Tantibundhit, Rasetshwane, et al. 2005–2009
  - Chanda and S. Park 2007
- **Energy distribution over time and/or frequency**
  - **Our approach: Sauert et al. 2006–2012**
  - **Brouckxon et al. 2008**
  - H. Park 2010
  - Tang and Cooke 2010–2011
  - **Taal et al. 2012**
- **Mimic Lombard effect**
  - Huang 2010
- **Amplification based on partial loudness**
  - **J. W. Shin et al. 2007–2009**
  - **H. S. Shin et al. 2010**

Our Approach Uses Speech Intelligibility Index

- **Objectives**
  - Enhancement of intelligibility of speech
  - Depending on background noise characteristics
  - Preservation of colour of tone not required
  - Delay constraint

- **Maximize *Speech Intelligibility Index* (SII)**
  given the current noise spectrum
  by redistribution of power over frequency
- **Subject to power constraints**
  - Human ear
  - Loudspeaker
  - Amplifier

Speech Intelligibility Index (SII), ANSI S3.5-1997

- **Spectrum levels in critical frequency bands**
  - *Speech spectrum level* E_i

$$E_i = 10\log\{\Phi_{\text{speech},i} \,/\, \Delta f_i\}$$

  - *Disturbance spectrum level* D_i

$$D_i = 10\log\{\Phi_{\text{noise},i} \,/\, \Delta f_i + \text{masking}\}$$

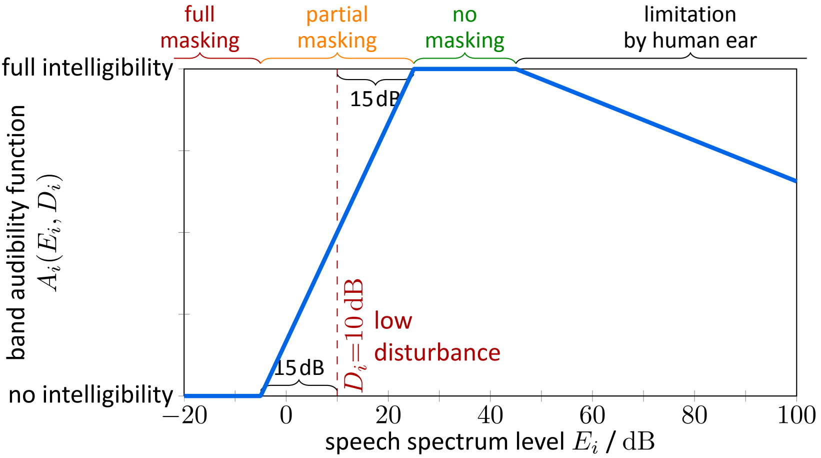
  - Frequency band index i
  - Width of frequency band Δf_i
- ***Band audibility function*** $A_i(E_i, D_i)$ **for each critical frequency band**
- ***Speech intelligibility index*** $S(\underline{E}, \underline{D})$

$$S(\underline{E}, \underline{D}) = \sum_i I_i \cdot A_i(E_i, D_i),$$

I_i: band importance function (fixed weights, see ANSI standard)

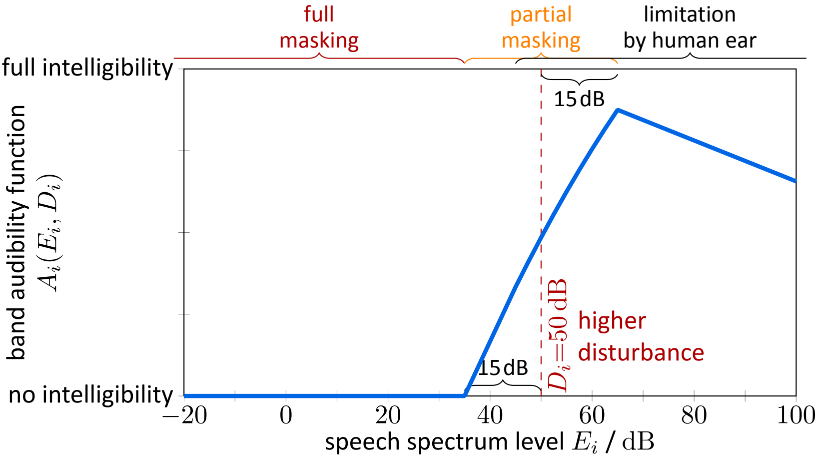
Band audibility function $A_i(E_i, D_i)$

- **Frequency band i**



Band audibility function $A_i(E_i, D_i)$

- Frequency band i



Main Steps of Algorithm

1. **Find optimum output speech spectrum level $\underline{E}_{\text{opt}}(\kappa)$**
   which maximizes the Speech Intelligibility Index
   given the current disturbance spectrum level $\underline{D}(\kappa)$

$$\underline{E}_{\text{opt}}(\kappa) = \arg\max_{\underline{E}^{\text{out}}} S(\underline{E}^{\text{out}}, \underline{D}(\kappa))$$

2. **Calculate spectral weights**
   to reach optimum output speech spectrum level
   with far-end (input) speech signal

$$W_i(\kappa) = 10^{[E_{\text{opt},i}(\kappa) - E_i^{\text{in}}(\kappa)]/20}$$

3. **Apply weights to far-end speech signal**

Optimization Problem

- **Find optimum output speech spectrum level $\underline{E}_{\text{opt}}(\kappa)$**

$$\underline{E}_{\text{opt}}(\kappa) = \arg\max_{\underline{E}^{\text{out}}} \sum_i I_i \cdot A_i\left(E_i^{\text{out}}, D_i(\kappa)\right)$$

given the current disturbance spectrum level $D_i(\kappa)$, subject to

$$\text{short-term output power} \overset{!}{\leq} \text{power constraint}$$

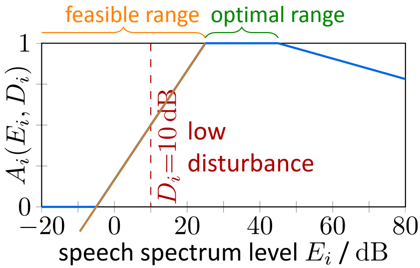
$$\Rightarrow \sum_i \hat{\Phi}_{ss,i}^{\text{out}} = \sum_i \Delta f_i \cdot 10^{E_i^{\text{out}}/10} \overset{!}{\leq} \text{power constraint}$$

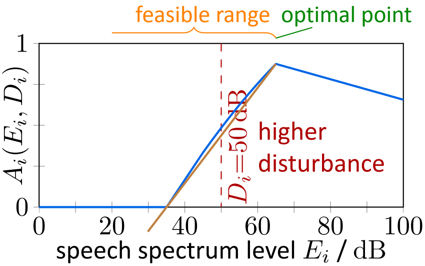
and—to protect the listener's ear—

$$E_i^{\text{out}} \leq 10 \log\left\{\frac{\Phi^{\max}}{\Delta f_i}\right\} \quad \text{with} \quad 10\log\{\Phi^{\max}\} = 94\,\mathrm{dBSPL}$$

- **Non-convex quadratic optim. problem with exponential constraint**

Two Modes of Optimization



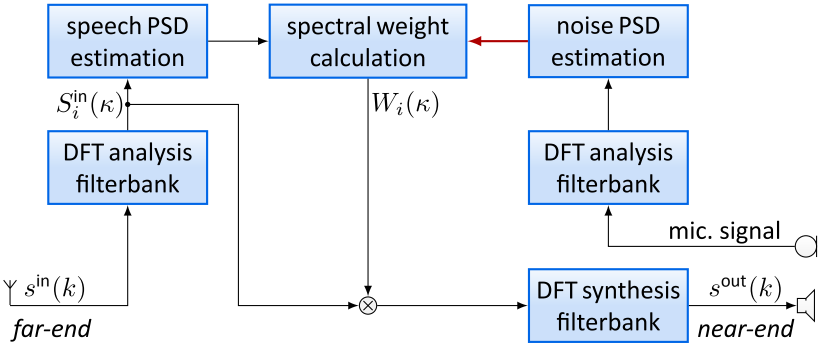


1. **Power constraint *is* sufficiently high:**
   - Optimal range/point can be reached in all frequency bands
   - SII is maximized for any speech spectrum level in optimal range/point
   - Power budget can be used to minimize distortions
2. **Power constraint *is not* sufficiently high:**
   - Only feasible range can be reached
   - Quadratic optimization problem with exponential constraint
   - Recursive closed-form solution
     with linear approximation and restriction to feasible range

Microphone Signal of Mobile Phone

- **Microphone signal $=$ near-end noise $+$ near-end speech**
- **Double-talk: if near-end speech would be considered as noise,**
  1. algorithm would amplify far-end speech to drown out near-end speech
  2. near-end listener would instinctively speak louder
  3. algorithm would further amplify far-end speech
  4. ...
- **At least distracting and annoying**
- **Noise estimation necessary to ignore near-end speech**

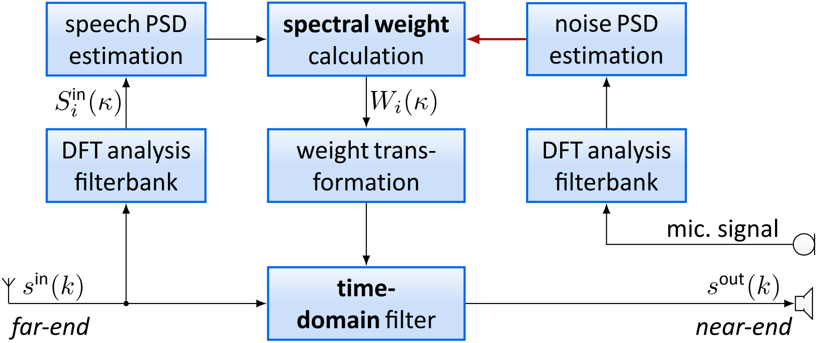
Implementation A: DFT Analysis-Synthesis Filterbank



- **Spectral weight calculation for each block**
  1. Find optimum output speech spectrum level $\underline{E}_{\text{opt}}(\kappa)$
  2. $W_i(\kappa) = 10^{[E_{\text{opt},i}(\kappa) - E_i^{\text{in}}(\kappa)]/20}$

- **Non-uniform frequency resolution possible by subband merging**

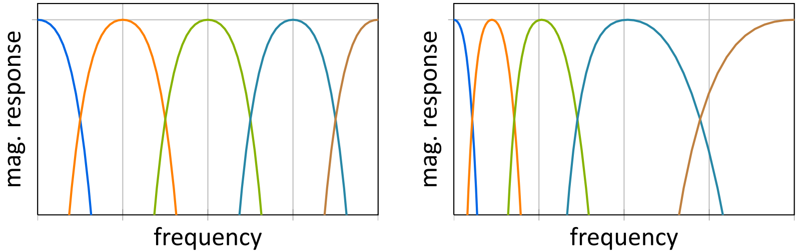
Implementation B: Uniform Filterbank Equalizer



- **Filterbank equalizer [Löllmann & Vary 2007]**
- **Lower delay than DFT analysis-synthesis filterbank**
- **Further delay reduction possible by FIR or IIR filter approximation**

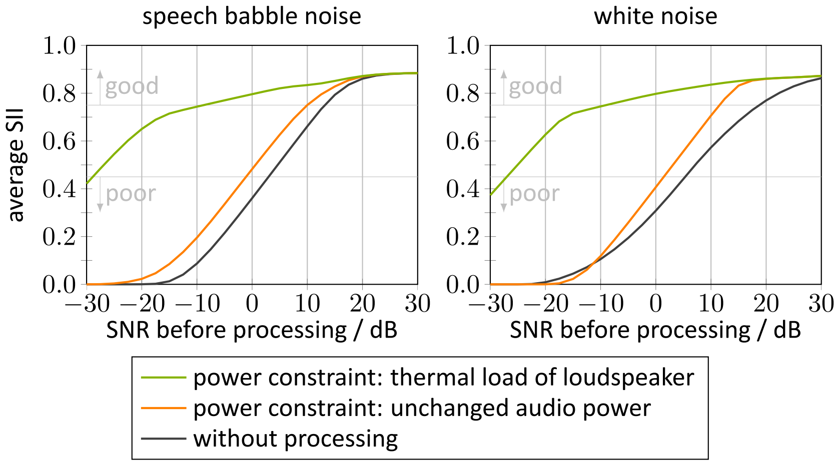
Implementation C: Non-Uniform Filterbank Equalizer

- **Non-uniform frequency bands by means of allpass-transformation**
  - Filterbank equalizer **[Löllmann & Vary 2007]**



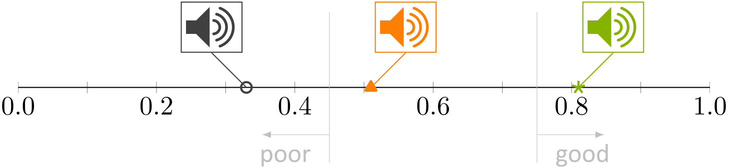
- **Bark-scaled frequency resolution**
- **DFT size:** 32–34 instead of 256 at 8 kHz sampling rate

Simulation Results



- Speech: TIMIT database (5.4 h) at sample rate 8 kHz with level 62.35 dB
- Noise from NOISEX-92 database with level according to SNR

Audio Samples

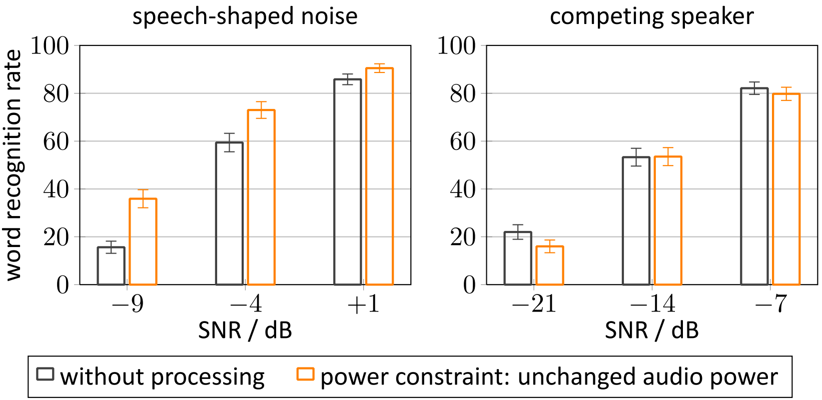


Speech Intelligibility Index

- ○ without processing
- ▲ power constraint: unchanged audio power
- ∗ power constraint: thermal load of loudspeaker

- Real-world noise scenario, recorded with artificial head and mounted phone
- Signal-to-noise ratio before processing $\approx$ 0 dB

The Hurricane Challenge – Preliminary Results



- **Significant intelligibility improvement for speech-shaped noise**
- **Competing speaker poses a systematic problem**
- **After all, algorithm was developed for traffic noise, etc.**
  - Room for further improvements

Application in Digital Hearing Aids

- **Open fitting for modern hearing aids**
  - Does not seal ear channel
  - Improves wearing comfort
  - For customers with low to moderate hearing loss
  - Allows background noise to get through to the ear

- **Near-end listening enhancement can help to improve intelligibility of**
  - noise reduced microphone signal
  - external audio signals from, e.g., a phone, music player, or television

- **Additional Challenges**
  - Overall signal delay less than 10–15 ms
  - Low complexity

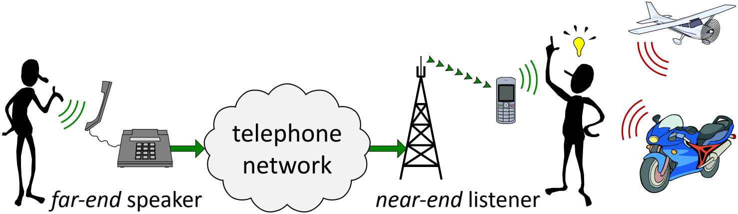
Application in Car Environments

- **Automatic volume control for car radio** depending on
  - speed
  - road surface
  - type of wheels
  - weather condition

- **Intelligibility improvements for in-car communication systems**

- **Additional Challenges**
  - Noise PSD estimation must contain basic echo cancellation
  - Less change of colour of tone for music

Summary



- **Enhance speech intelligibility for near-end listener**
  - by processing of received clean far-end speech signal
  - depending on local near-end acoustical background noise
  - with Speech Intelligibility Index as optimization criterion
  - considering the limits of the loudspeaker and the listener's ear
- **Implementation with different structures and spectral resolutions**
- **Further applications for near-end listening enhancement**

`http://www.ind.rwth-aachen.de/~bastian-sauert/`