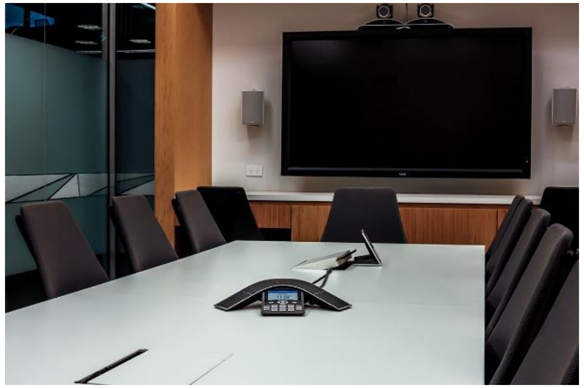
Perception and Binaural Signal Processing

EE4715 Array Processing | 7 June ‘24

TUDelft

Audio Signal Processing Applications



TU Delft

MVDR Beamforming

- STFT domain (short-time stationarity assumption)
- narrowband assumption
- minimise output noise power
- target preservation constraint

$$\boldsymbol{x}(k,l) = \boldsymbol{a}(k,l)s(k,l) + \boldsymbol{n}(k,l)$$

$$y(k,l) = \boldsymbol{w}^{\mathrm{H}}(k,l)\boldsymbol{x}(k,l)$$

$$\underset{\boldsymbol{w}(k,l)}{\text{minimise}} \quad \boldsymbol{w}^{\mathrm{H}}(k,l)\boldsymbol{R}_{\boldsymbol{n}}(k,l)\boldsymbol{w}(k,l)$$

$$\text{subject to} \quad \boldsymbol{w}^{\mathrm{H}}(k,l)\boldsymbol{a}(k,l) = 1$$

$$\boldsymbol{w}(k,l) = \frac{\boldsymbol{R}_{\boldsymbol{n}}^{-1}(k,l)\boldsymbol{a}(k,l)}{\boldsymbol{a}^{\mathrm{H}}(k,l)\boldsymbol{R}_{\boldsymbol{n}}^{-1}(k,l)\boldsymbol{a}(k,l)}$$

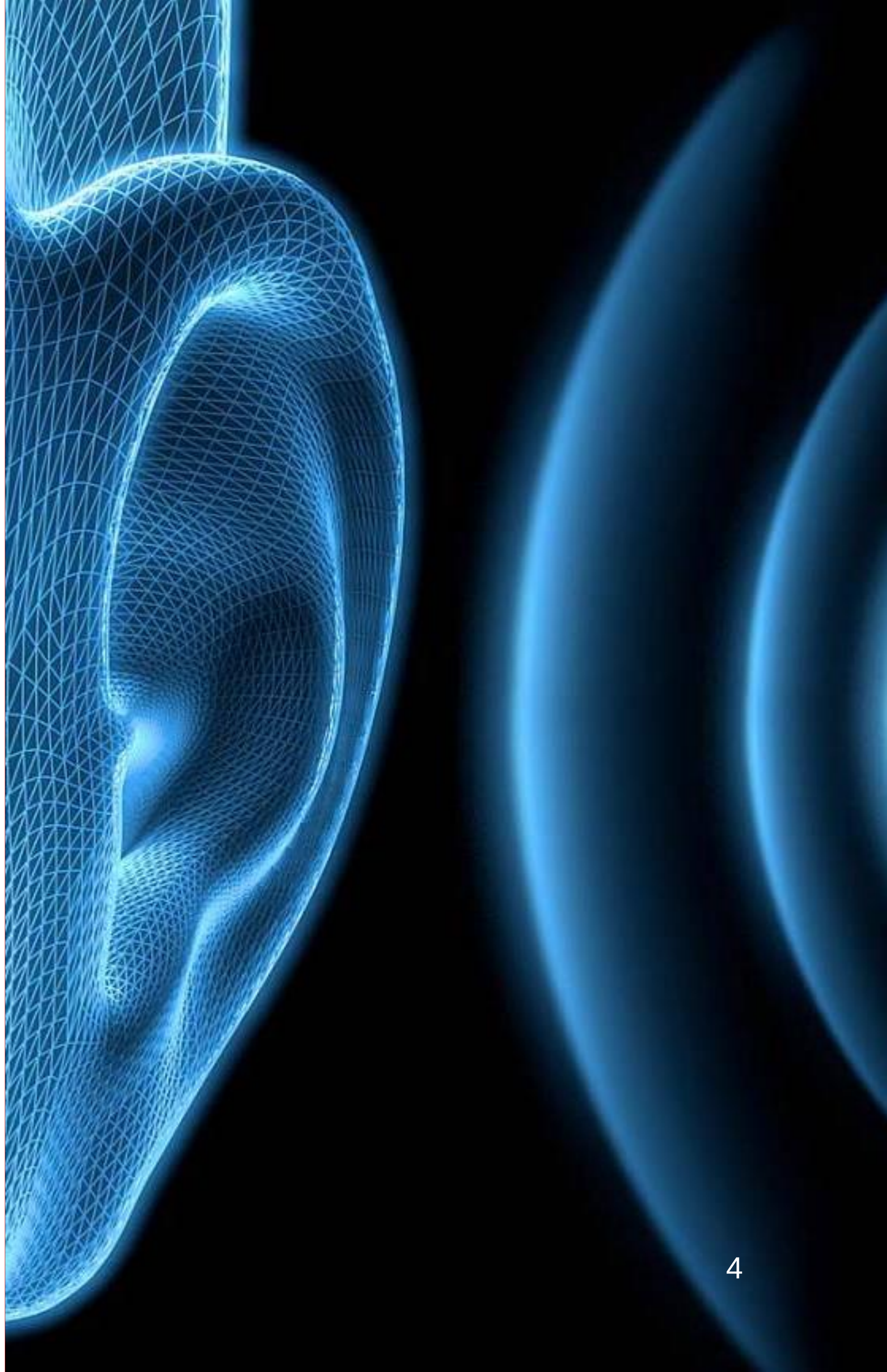
TU Delft

Speech Intelligibility in Noise

‘Measure of how comprehensible speech is in given conditions’

Factors that influence speech intelligibility:

- energetic masking
- informational masking
- attention/listening effort
- cognitive abilities/language proficiency
- hearing impairment
- reverberation
- spatial distribution (auditory stream segregation)

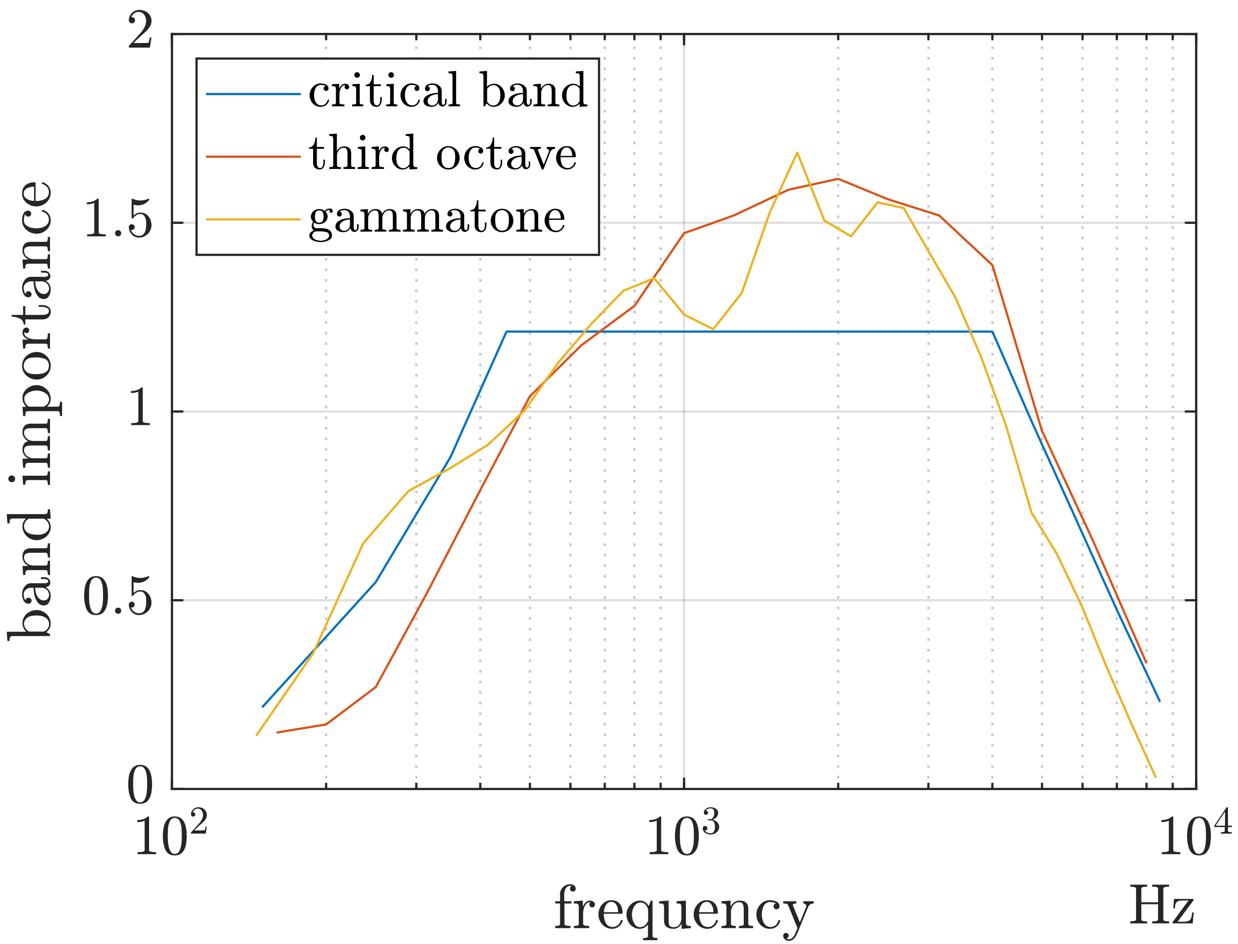


Measuring Intelligibility

- matrix test
- speech reception threshold (SRT)
- audiogram

Speech Intelligibility Index (SII)

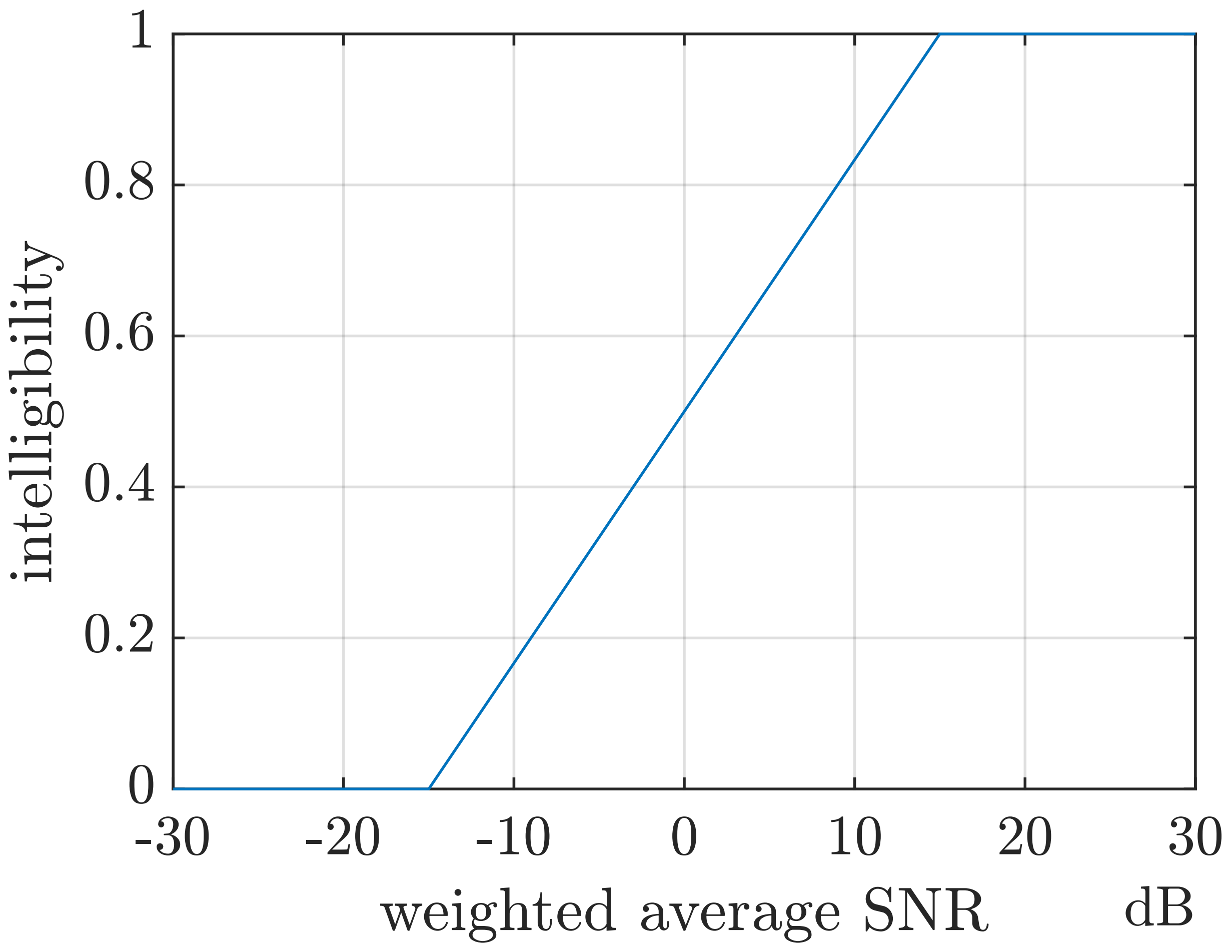
- objective measure, ANSI standard
- started as articulation index (AI), 1947
- filter speech and noise into frequency bands
- perceptually weighted average of band SNRs
- convert weighted average SNR to intelligibility score



TU Delft

Speech Intelligibility Index (SII)

- objective measure, ANSI standard
- started as articulation index (AI), 1947
- filter speech and noise into frequency bands
- perceptually weighted average of band SNRs
- convert weighted average SNR to intelligibility score
- internal noise to model hearing thresholds
- SRT measure



TUDelft

Other Intelligibility Measures

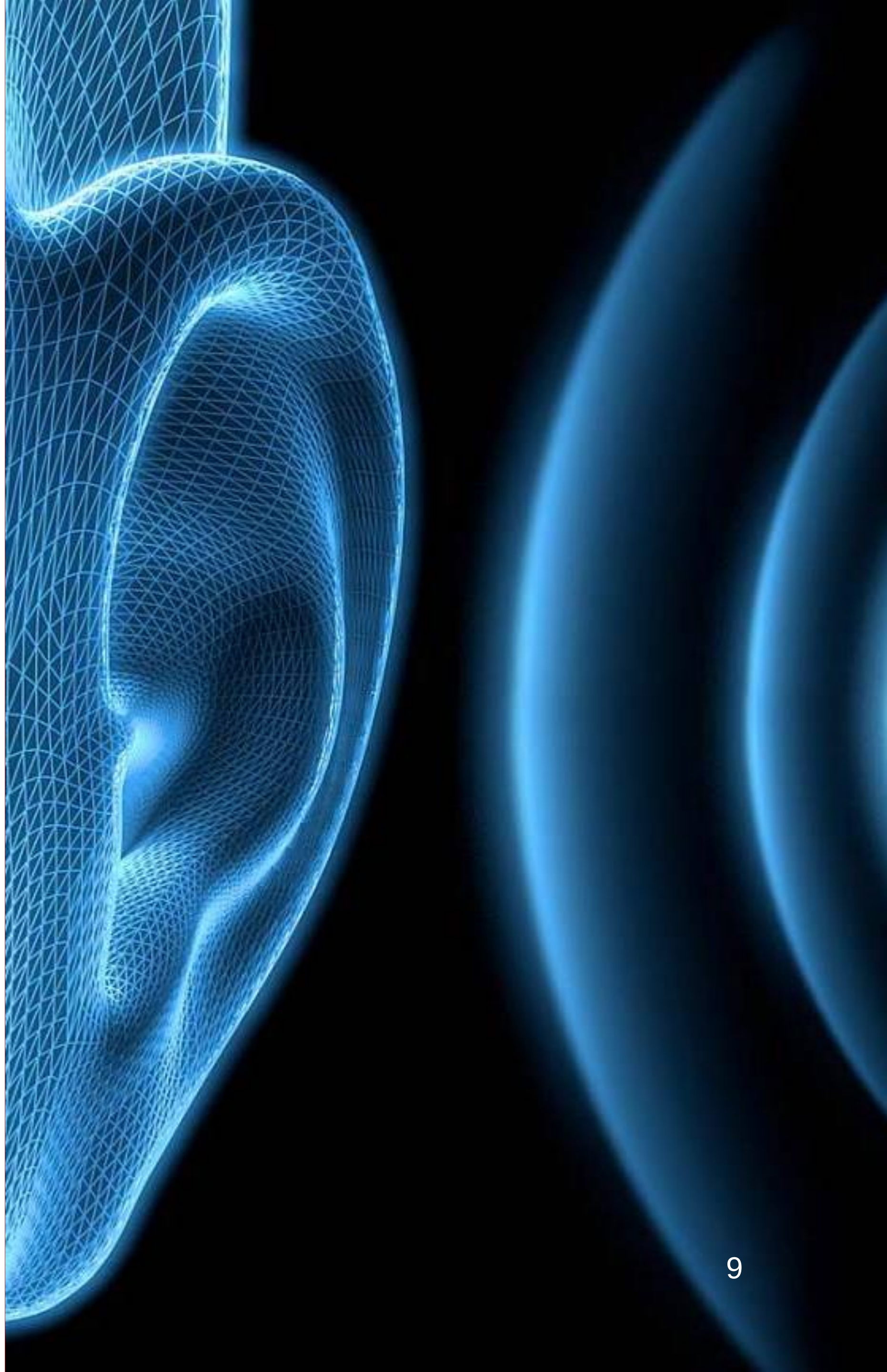
- speech transmission index (STI)
- coherence speech intelligibility index (CSII)
- short-time objective intelligibility (STOI)
- hearing-aid speech perception index (HASPI)
- data-driven measures

Speech Intelligibility in Noise

‘Measure of how comprehensible speech is in given conditions’

Factors that influence speech intelligibility:

- energetic masking
- informational masking
- attention/listening effort
- cognitive abilities/language proficiency
- hearing impairment
- reverberation
- spatial distribution (auditory stream segregation)



Spatial Release from Masking

two main binaural cues:

- interaural time difference (ITD) mainly < 1000 Hz
- interaural level difference (ILD) mainly > 1500 Hz

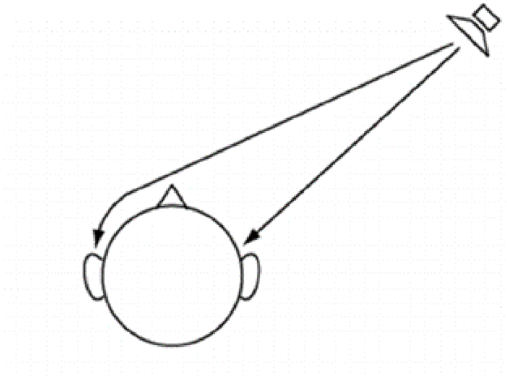
interaural transfer function (ITF)

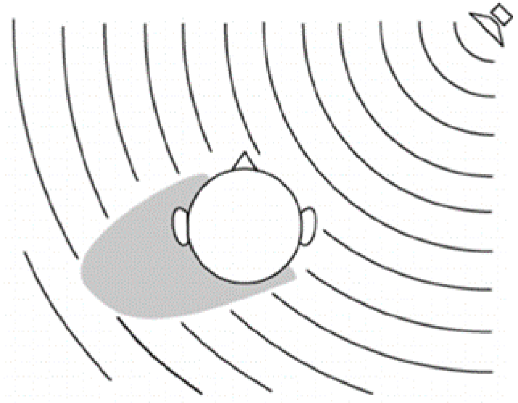
binaural cue consistency across frequency

$$\mathrm{ITF} = \frac{a_{\mathrm{L}}}{a_{\mathrm{R}}}$$

$$\mathrm{ITD} = \frac{\angle\,\mathrm{ITF}}{\omega} = \frac{\angle\,\mathrm{ITF}}{2\pi f}$$

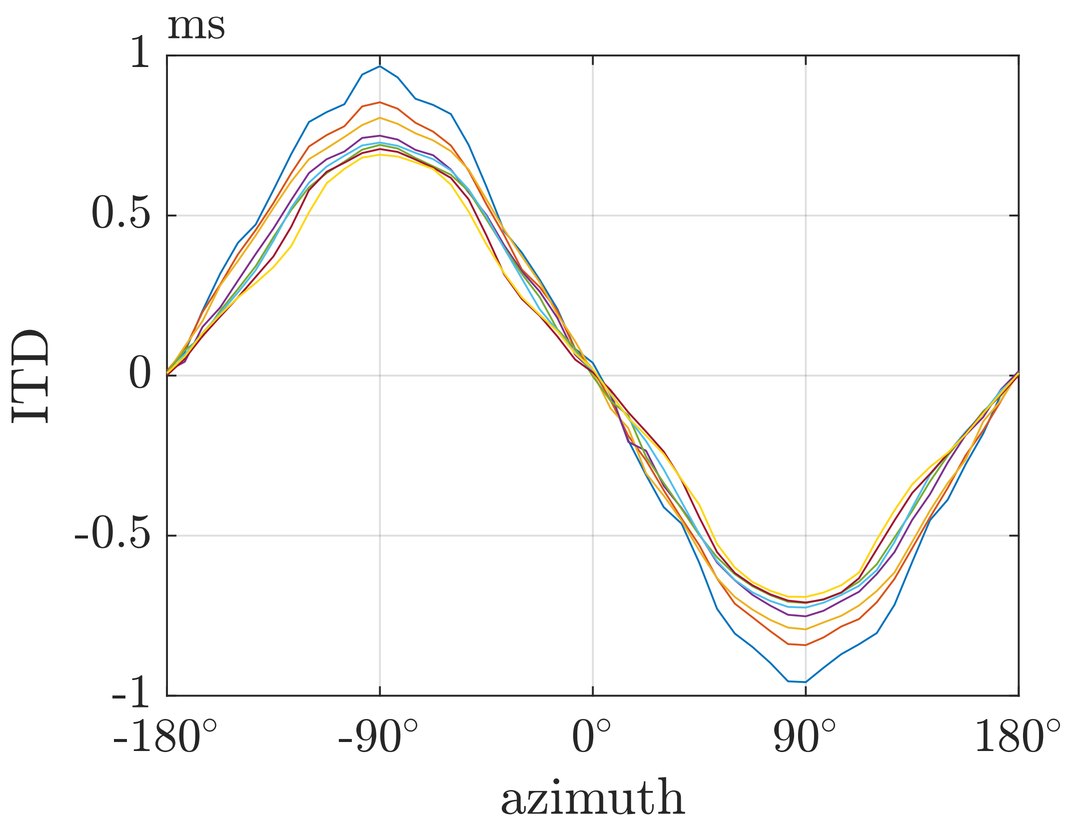
$$\mathrm{ILD} = |\mathrm{ITF}|$$

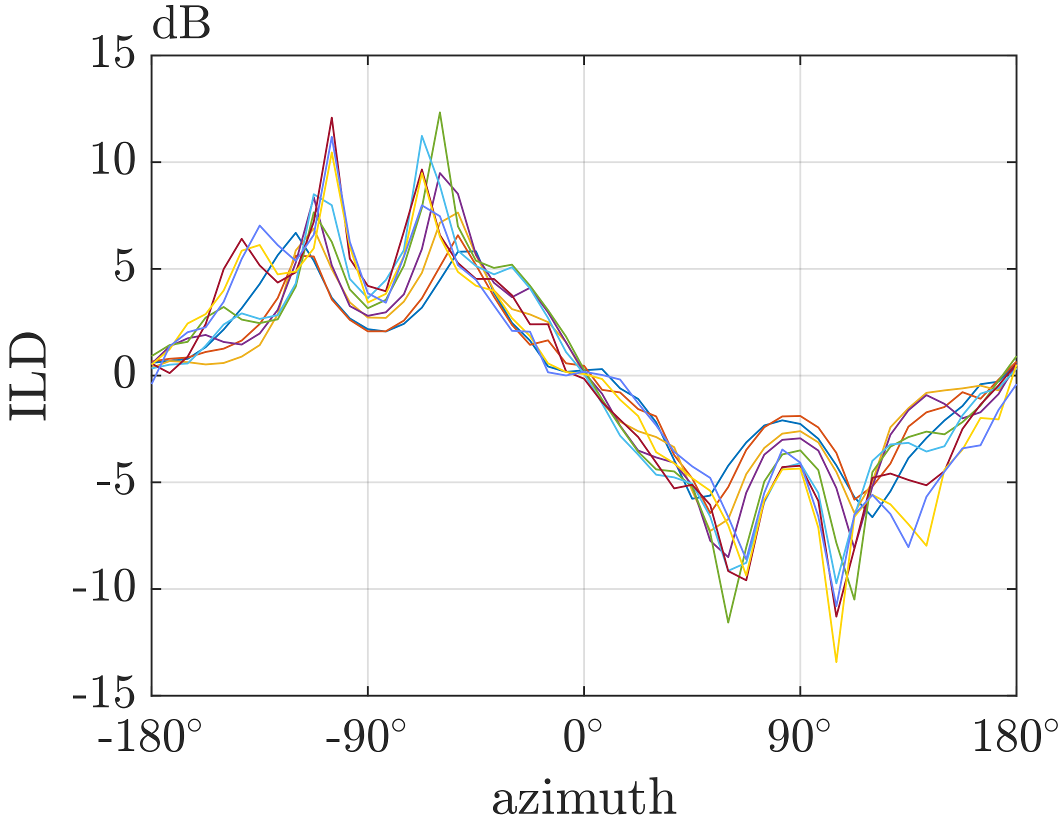




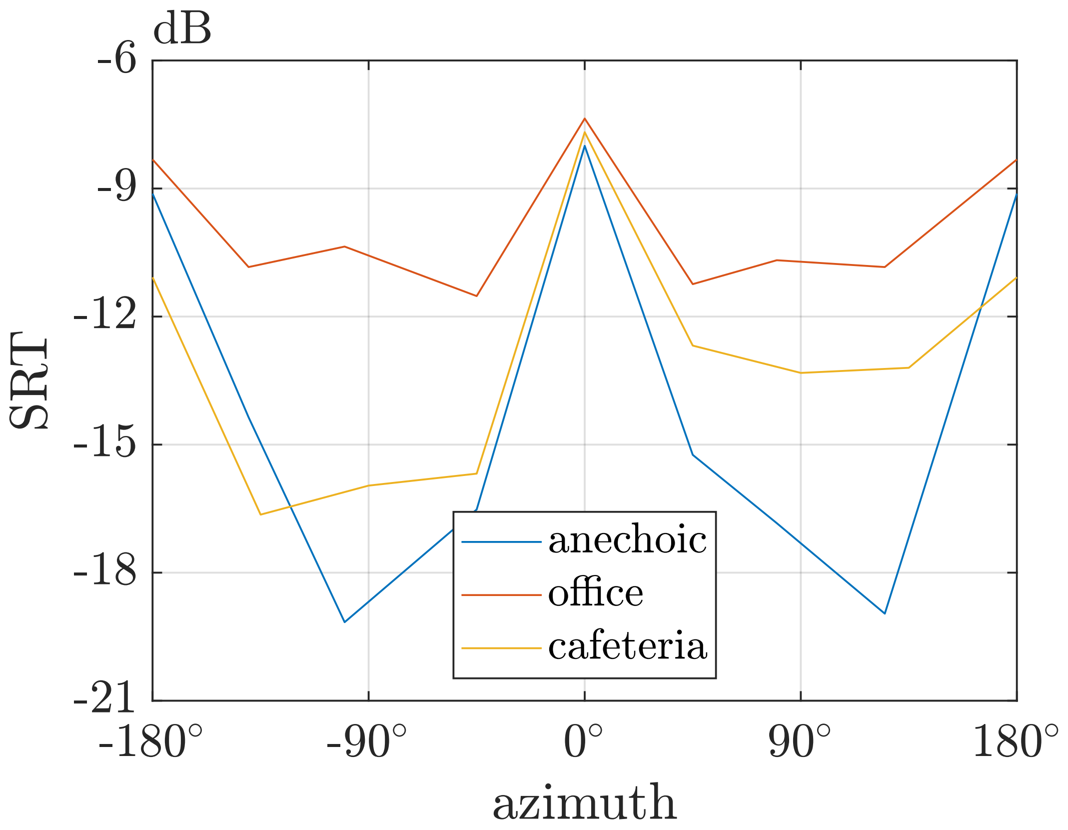
TU Delft

Spatial Release from Masking





Spatial Release from Masking



TU Delft

Binaural Intelligibility Models

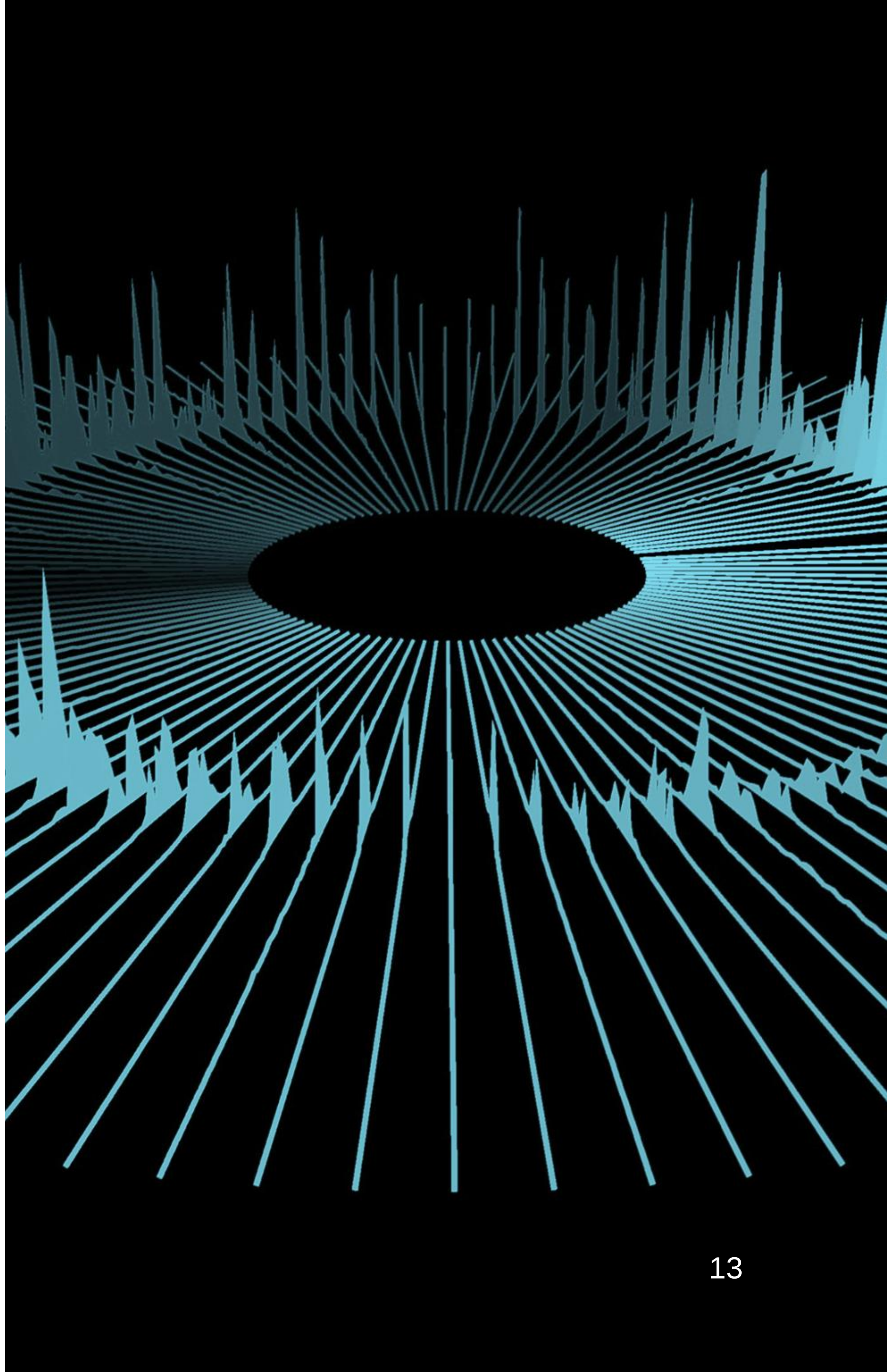
information used:

- SNR (energy)
- modulation
- correlation
- glimpsing

input availability (blindness, intrusiveness)

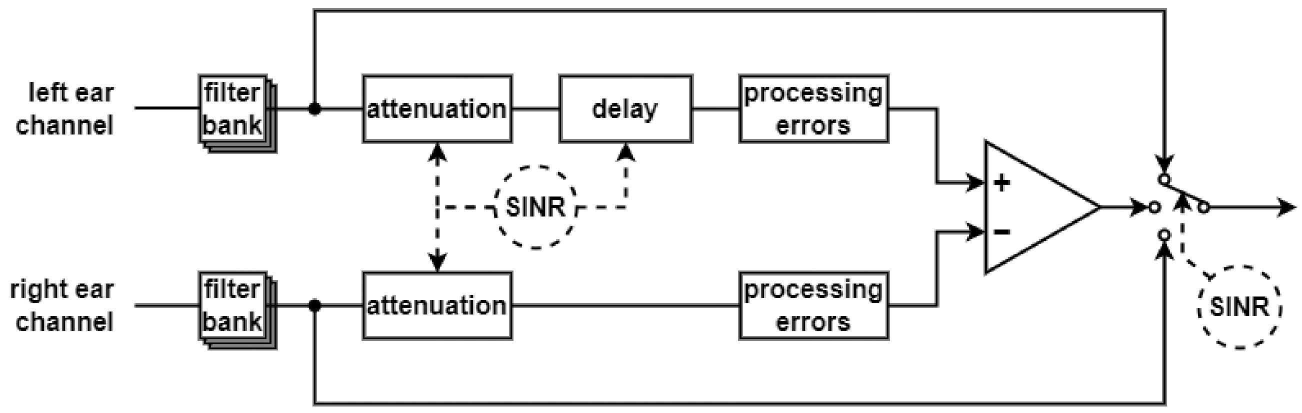
offline/real time

physiological interpretability



Binaural Speech Intelligibility Model

1. gammatone filter bank
2. internal masking noise
3. equalisation–cancellation (Durlach, 1963)
4. artificial processing errors
5. better ear listening
6. monaural intelligibility measure (SII)



TUDelft

Binaural Speech Intelligibility Model

$$y_{\mathrm{EC}}(t) = \mathrm{e}^{\frac{1}{2}\gamma} y_{\mathrm{L}}(t + \tfrac{1}{2}\tau) - \mathrm{e}^{-\frac{1}{2}\gamma} y_{\mathrm{R}}(t - \tfrac{1}{2}\tau)$$

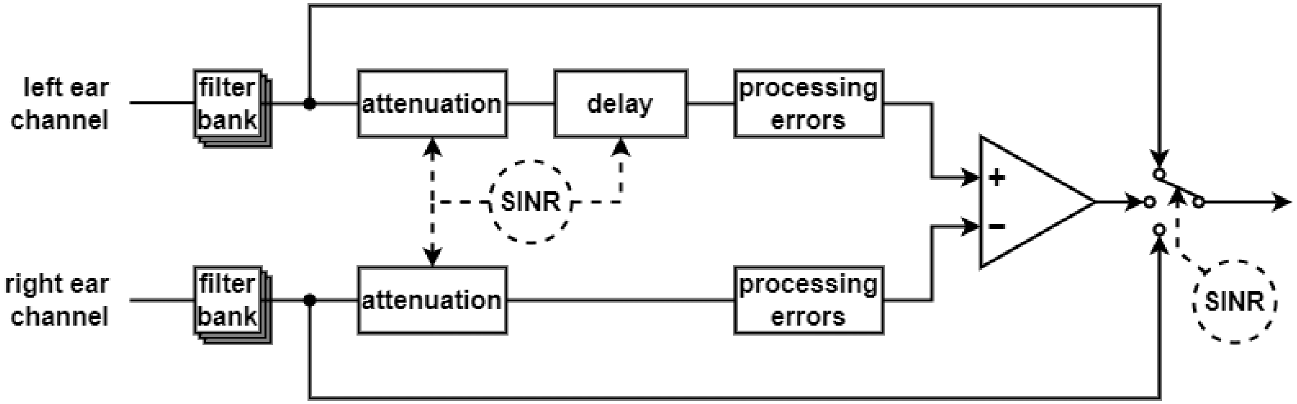
$$y_{\mathrm{EC}}(\omega) = \mathrm{e}^{\frac{1}{2}(\gamma + j\omega\tau)} y_{\mathrm{L}}(\omega) - \mathrm{e}^{-\frac{1}{2}(\gamma + j\omega\tau)} y_{\mathrm{R}}(\omega)$$

$$\boldsymbol{y} = \begin{pmatrix} y_{\mathrm{L}} \\ y_{\mathrm{R}} \end{pmatrix} \qquad \boldsymbol{v} = \begin{pmatrix} \mathrm{e}^{\frac{1}{2}(\gamma - j\omega\tau)} \\ -\mathrm{e}^{-\frac{1}{2}(\gamma - j\omega\tau)} \end{pmatrix}$$

$$\mathrm{SNIR}(y_{\mathrm{EC}}) = \frac{\boldsymbol{v}^{\mathrm{H}} \boldsymbol{P_s} \boldsymbol{v}}{\boldsymbol{v}^{\mathrm{H}} \boldsymbol{P_n} \boldsymbol{v}}$$

$$y_{\mathrm{EC}} = \boldsymbol{v}^{\mathrm{H}} \boldsymbol{y}$$

$$\boldsymbol{P_y} = \int \mathrm{E}\left(\boldsymbol{y}\boldsymbol{y}^{\mathrm{H}}\right) d\omega$$



Binaural MVDR

binaural signal model

reference transfer function constraint

What happens to the binaural cues?

- Target binaural cues are preserved.
- Noise sources virtually move to target!

$$\begin{cases} \boldsymbol{x} = \boldsymbol{a}s + \boldsymbol{n}_0 + \sum_i \boldsymbol{b}_i n_i \\ y_\mathrm{L} = \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{x} \\ y_\mathrm{R} = \boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{x} \end{cases}$$

$$\underset{\boldsymbol{w}_\mathrm{L}}{\text{minimise}} \quad \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{R}_{\boldsymbol{n}}\boldsymbol{w}_\mathrm{L}$$

$$\text{subject to} \quad \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{a} = a_\mathrm{L}$$

$$\boldsymbol{w}_\mathrm{L} = \frac{\boldsymbol{R}_{\boldsymbol{n}}^{-1}\boldsymbol{a}}{\boldsymbol{a}^\mathrm{H}\boldsymbol{R}_{\boldsymbol{n}}^{-1}\boldsymbol{a}} a_\mathrm{L}^*$$

$$\mathrm{ITF}_\mathrm{out}(s) = \frac{\boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{a}}{\boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{a}} = \frac{a_\mathrm{L}}{a_\mathrm{R}} = \mathrm{ITF}(s)$$

$$\mathrm{ITF}_\mathrm{out}(n_i) = \frac{\boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{b}_i}{\boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{b}_i} = \frac{a_\mathrm{L}}{a_\mathrm{R}} = \mathrm{ITF}(s)$$

TUDelft

Joint Binaural LCMV

add binaural cue preservation constraint(s)

couple left and right problems

closed-form solution exists

reduced degrees of freedom for SNR maximisation!

$$\begin{cases} \boldsymbol{x} = \boldsymbol{a}s + \boldsymbol{n}_0 + \sum_i \boldsymbol{b}_i n_i \\ y_\mathrm{L} = \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{x} \\ y_\mathrm{R} = \boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{x} \end{cases}$$

$$\frac{\boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{b}_i}{\boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{b}_i} = \frac{b_{i\mathrm{L}}}{b_{i\mathrm{R}}} \quad \Leftrightarrow \quad \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{b}_i b_{i\mathrm{R}} = \boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{b}_i b_{i\mathrm{L}}$$

$$\begin{aligned} \underset{\boldsymbol{w}_\mathrm{L},\boldsymbol{w}_\mathrm{R}}{\text{minimise}} \quad & \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{R}_{\boldsymbol{n}}\boldsymbol{w}_\mathrm{L} + \boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{R}_{\boldsymbol{n}}\boldsymbol{w}_\mathrm{R} \\ \text{subject to} \quad & \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{a} = a_\mathrm{L}, \\ & \boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{a} = a_\mathrm{R}, \\ & \boldsymbol{w}_\mathrm{L}^\mathrm{H}\boldsymbol{b}_i b_{i\mathrm{R}} = \boldsymbol{w}_\mathrm{R}^\mathrm{H}\boldsymbol{b}_i b_{i\mathrm{L}} \end{aligned}$$

TU Delft

Perception-Based Beamformer

- include BSIM to account for SRM
- maximise ‘perceived’ SNR

Binaural MVDR is one of the solutions!

$$\begin{cases} \boldsymbol{x} = \boldsymbol{a}s + \boldsymbol{n}_0 + \sum_i \boldsymbol{b}_i n_i \\ \boldsymbol{y} = \boldsymbol{W}^{\mathrm{H}}\boldsymbol{x} = \begin{pmatrix} \boldsymbol{w}_{\mathrm{L}} & \boldsymbol{w}_{\mathrm{R}} \end{pmatrix}^{\mathrm{H}} \boldsymbol{x} \\ z = \boldsymbol{v}^{\mathrm{H}}\boldsymbol{y} \end{cases}$$

$$\underset{\boldsymbol{v},\,\boldsymbol{W}}{\text{maximise}} \quad \frac{\boldsymbol{v}^{\mathrm{H}}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{a}\boldsymbol{a}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{H}}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{n}}\boldsymbol{W}\boldsymbol{v}}$$

$$\text{subject to} \quad \boldsymbol{W}^{\mathrm{H}}\boldsymbol{a} = \boldsymbol{a}_{\mathrm{ref}}$$

TU Delft