AAT Seminar

Perceptual quality evaluation of spatial sound

Pedro Lladó

pedro.llado@aalto.fi

4.3.2024

Aalto-yliopisto
Aalto-universitetet
Aalto University

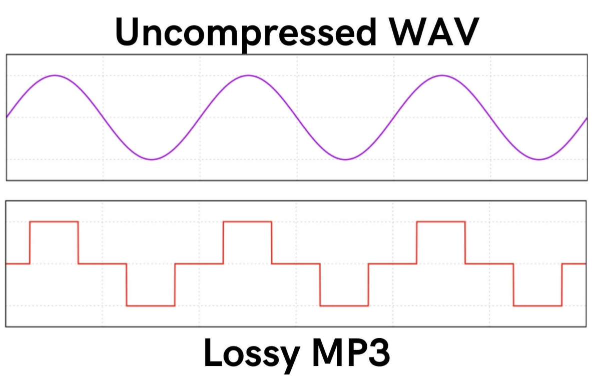
"We define* (perceptual) *quality as the outcome of an individual's comparison and judgment process.

It includes perception, reflection about the perception, and the description of the outcome".

Le Callet et al. (2012)

"We define* (perceptual) *quality as the outcome of an individual's comparison and judgment process.

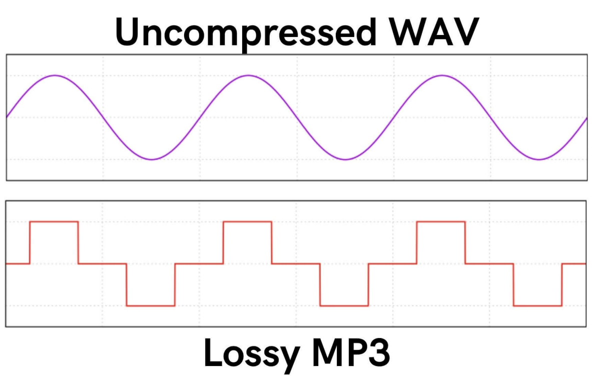
It includes perception, reflection about the perception, and the description of the outcome".



Le Callet et al. (2012)

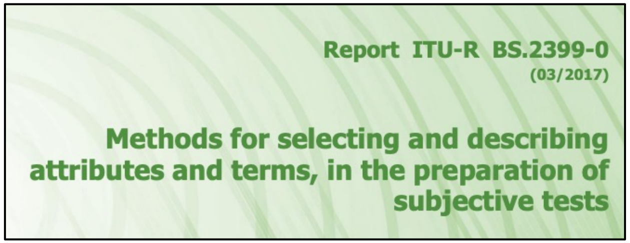
"We define** (perceptual) **quality as the outcome of an individual's comparison and judgment process.

It includes perception, reflection about the perception, and the description of the outcome".



Le Callet et al. (2012)

Sound quality depends on several attributes

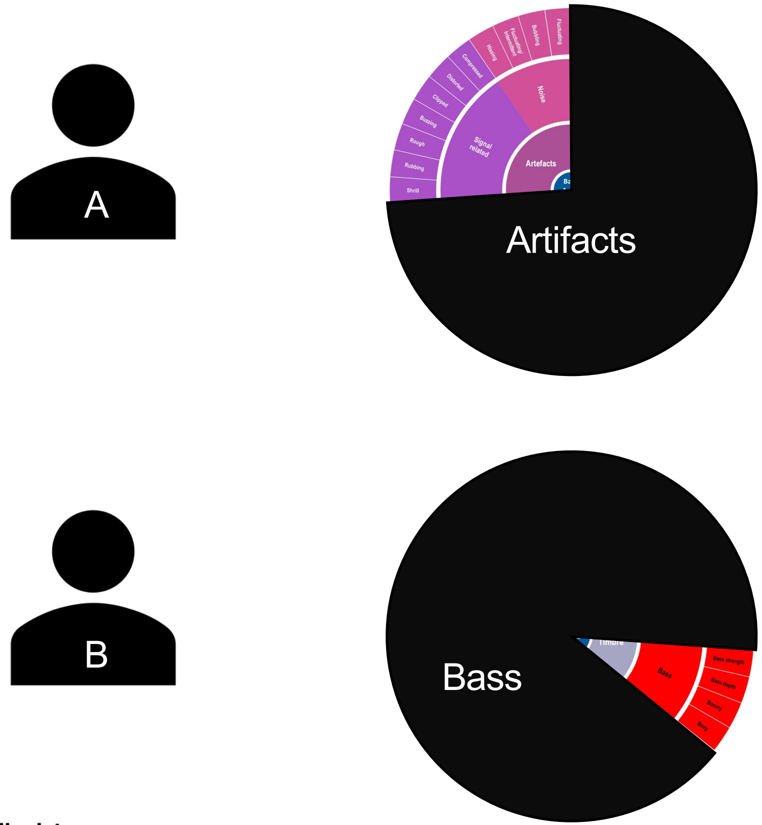

Report ITU-R BS.2399-0
(03/2017)

Methods for selecting and describing attributes and terms, in the preparation of subjective tests

Descriptive sensory analysis

Rating overall sound quality is hard and unstable

day 1

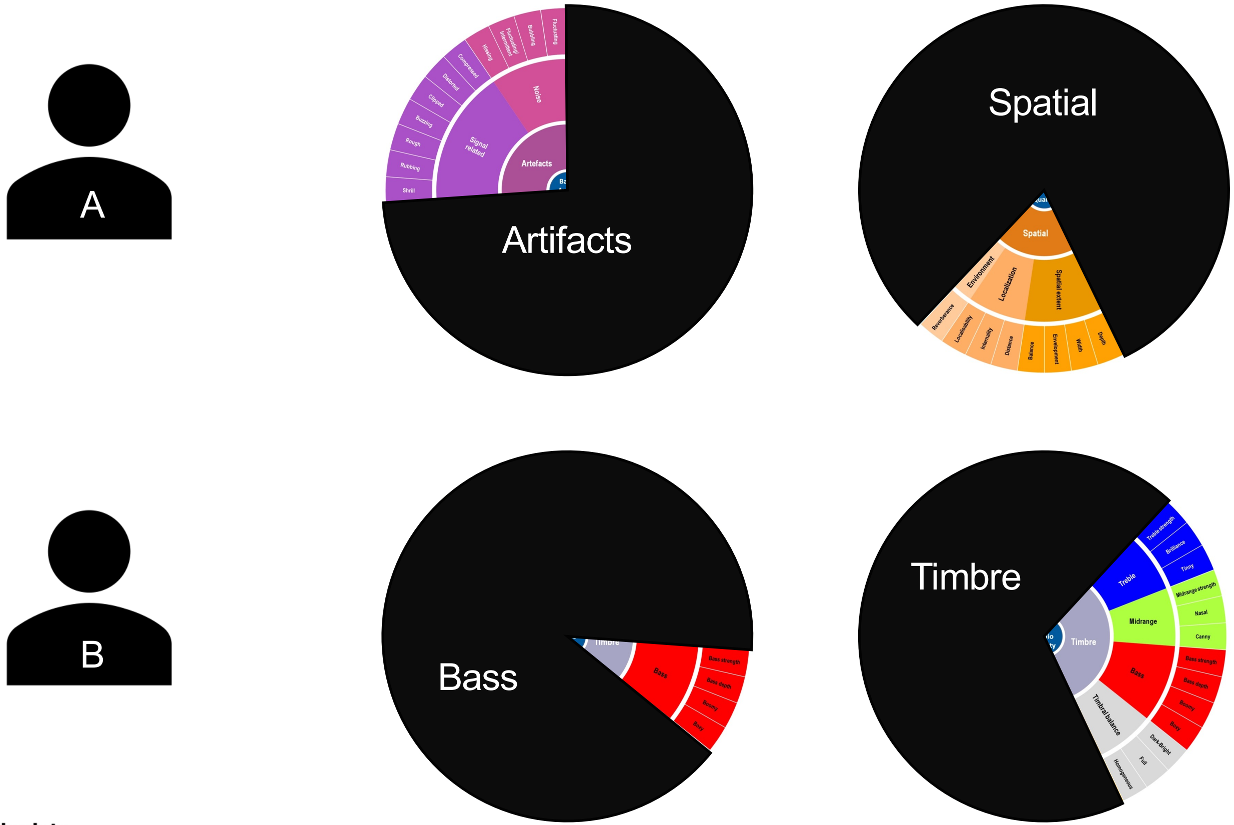


Aalto-yliopisto
Aalto-universitetet
Aalto University

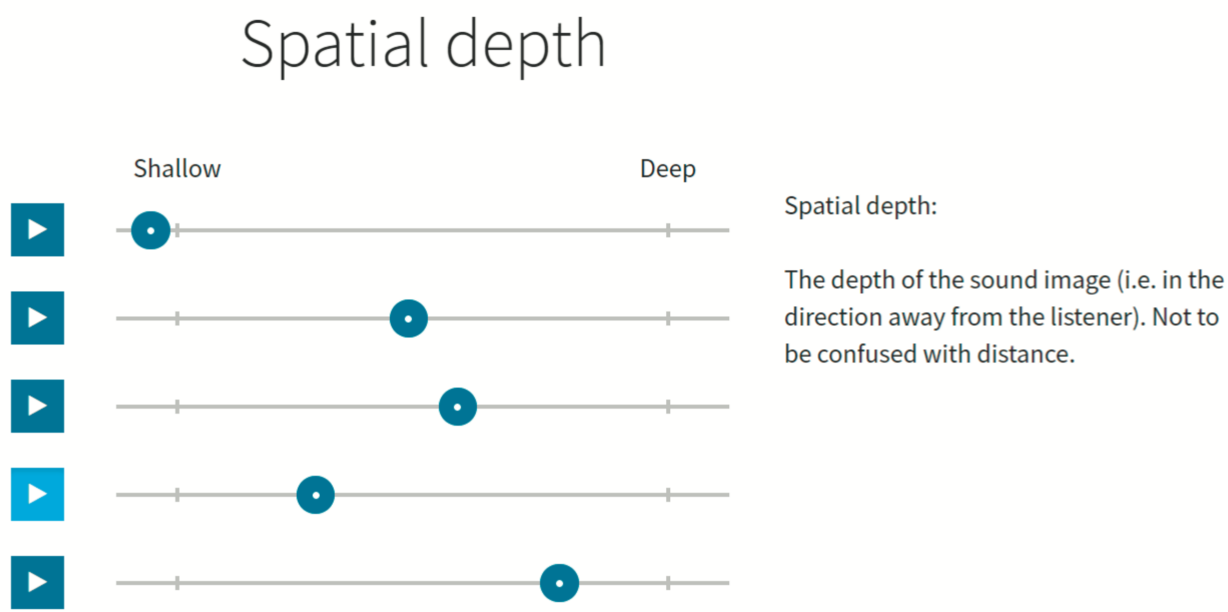
Rating overall sound quality is hard and unstable

day 1

day 2



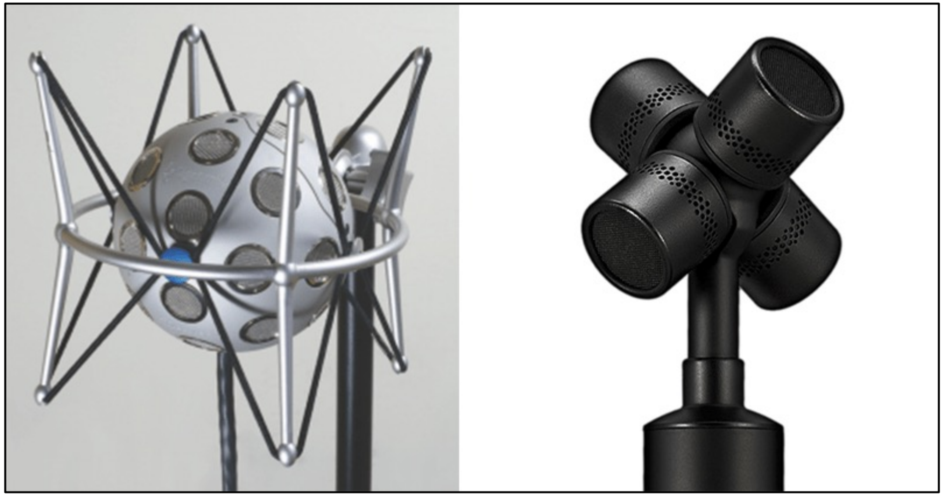
Comparing sounds within a single attribute is relatively easy and repeatable

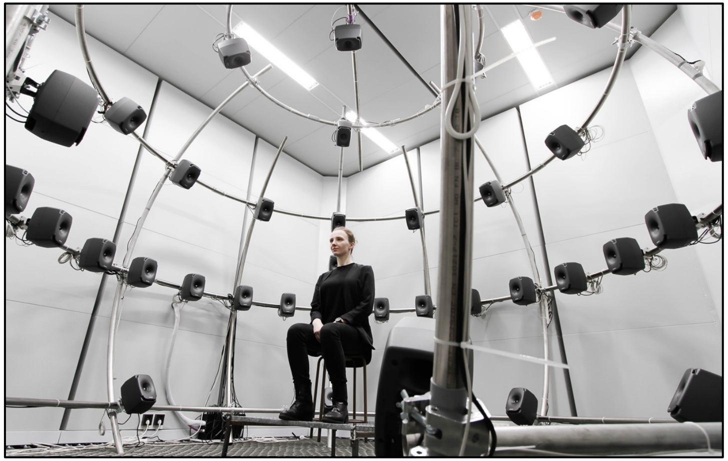


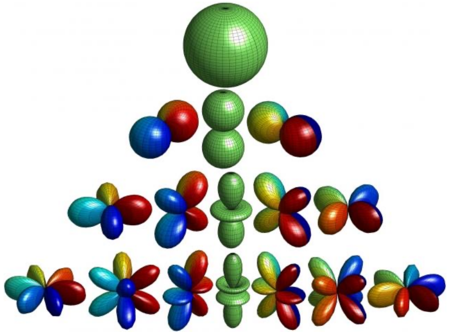
Aalto-yliopisto
Aalto-universitetet
Aalto University

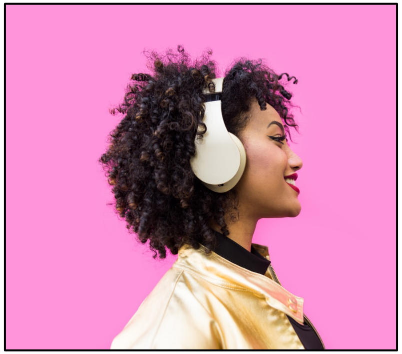
Zacharov et al. (2012)

Spatial audio recording, processing and reproduction is prone to introduce artifacts



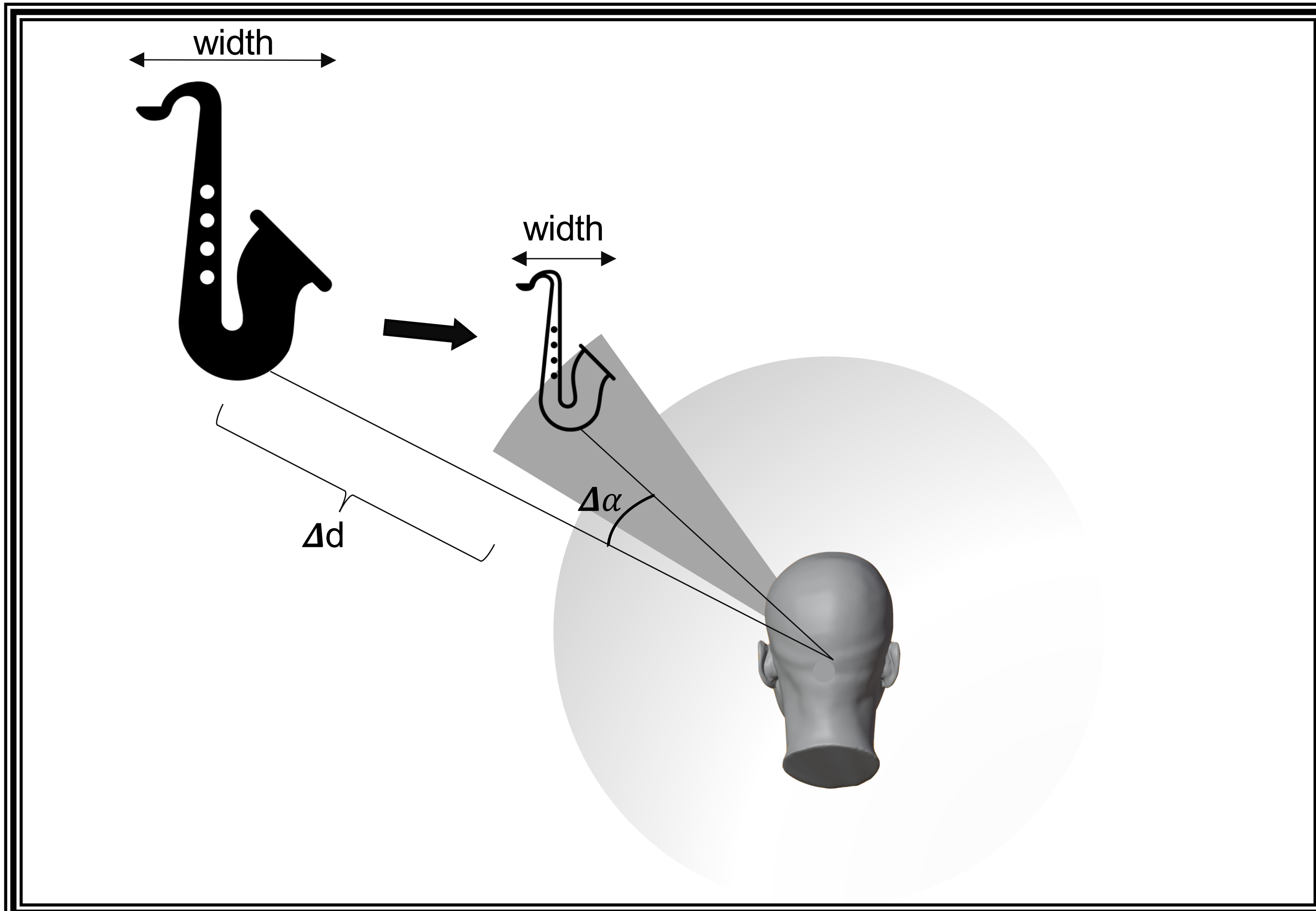






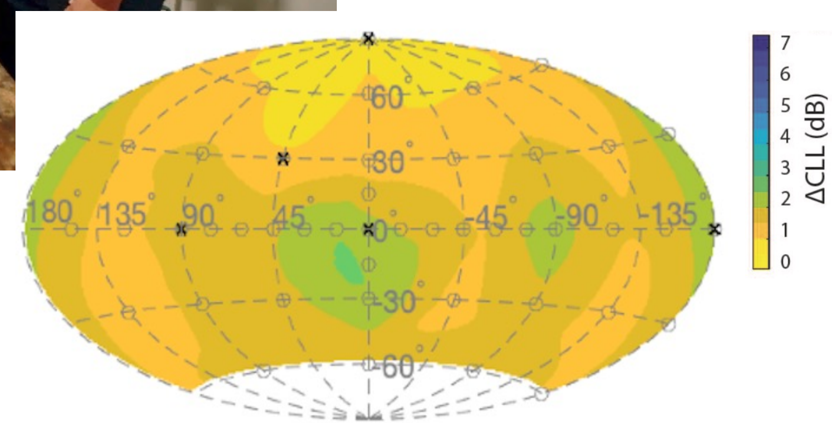
Aalto-yliopisto
Aalto-universitetet
Aalto University

Spatial perception is often distorted



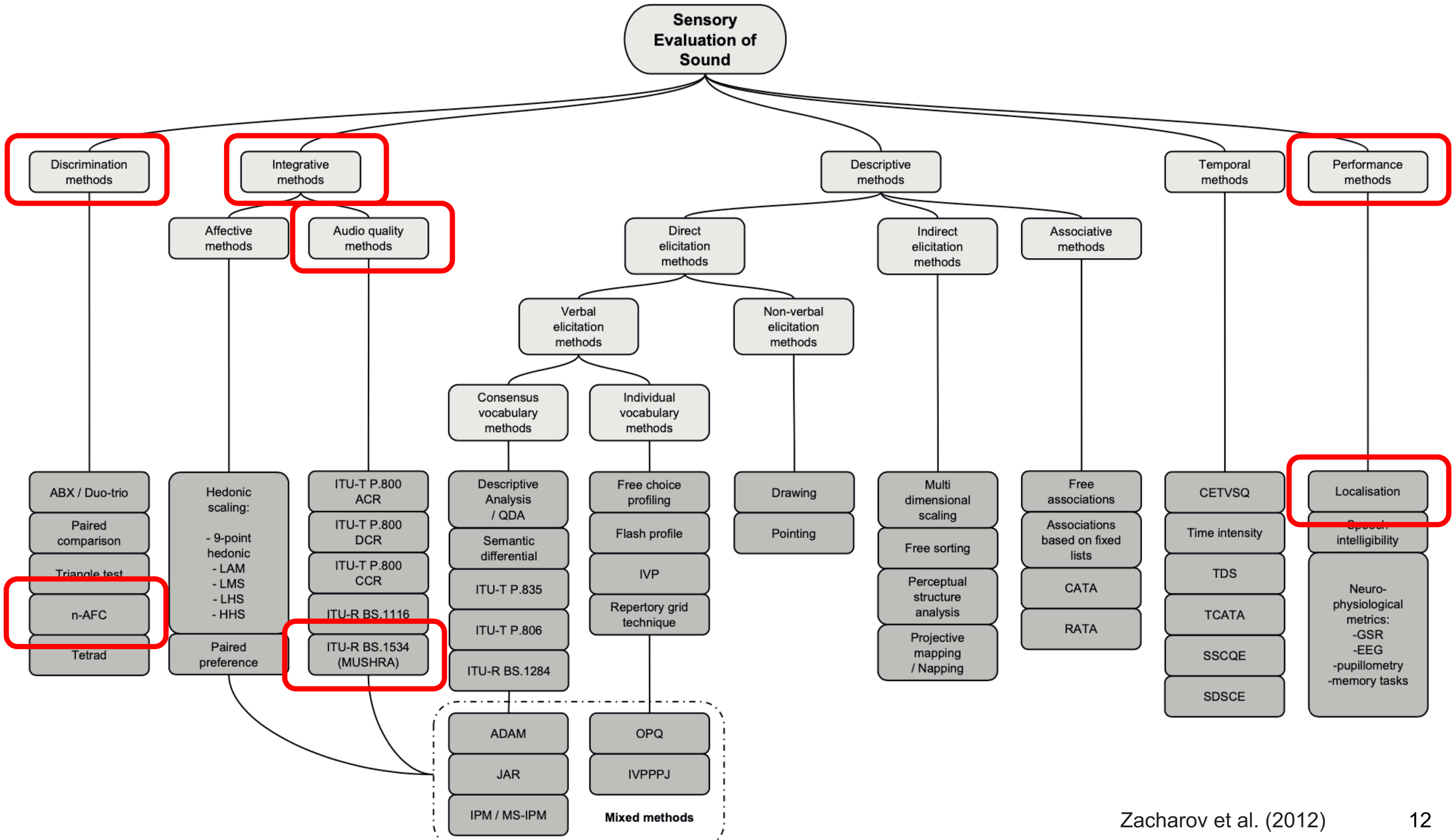
Timbral differences are specially problematic when we have a reference, e.g. augmented reality

Direction dependent colouration



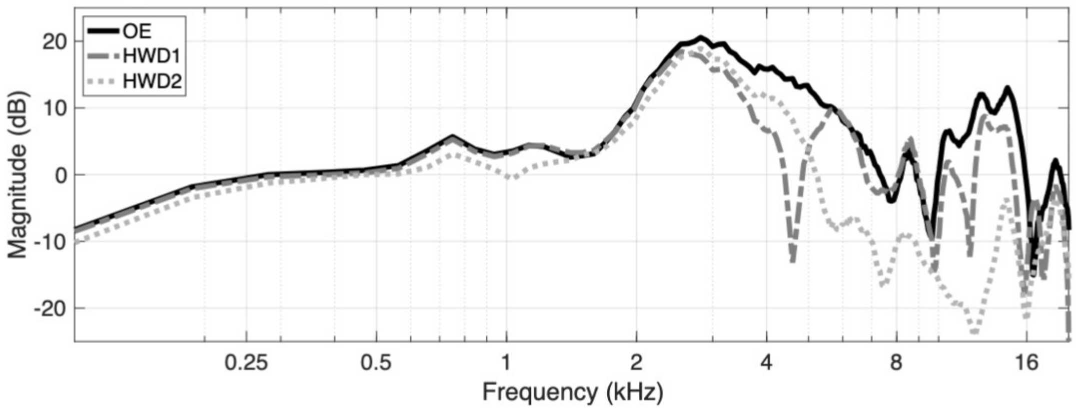
Lladó et al. (2022)

Listening experiments are the best option to achieve <u>objective</u> results



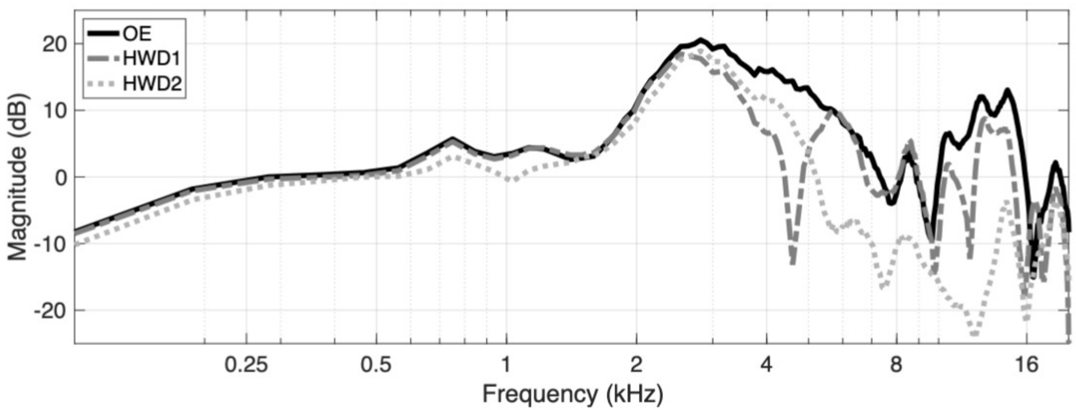
Zacharov et al. (2012)

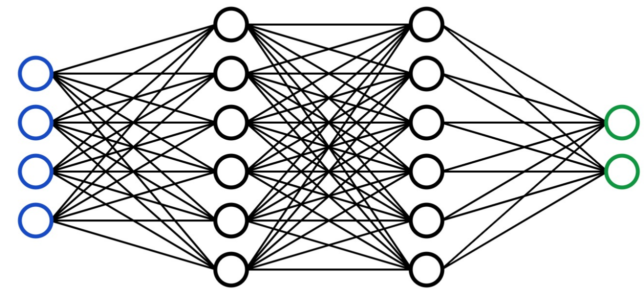
Listening experiments are not always feasible



Majdak et al. (2022)

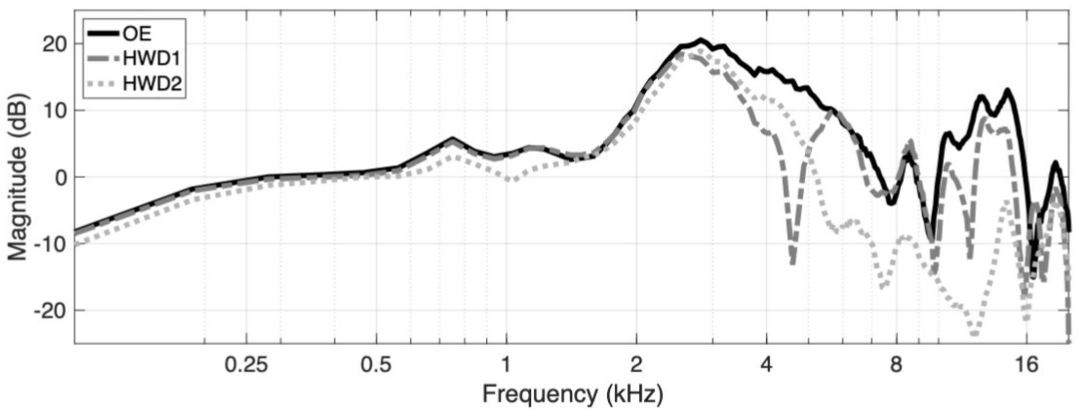
Listening experiments are not always feasible

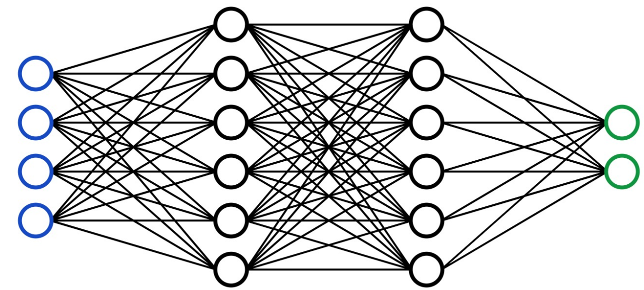


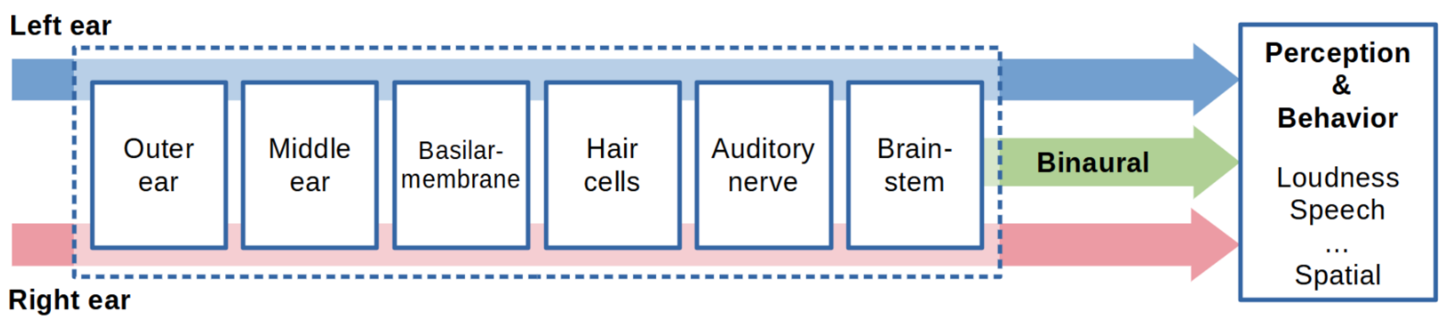


Majdak et al. (2022)

Listening experiments are not always feasible







The Auditory Modelling Toolbox

EXP_BAUMGARTNER2014 - Results from Baumgartner et al. (2014)

Usage:

```
data = exp_baumgartner2014(flag)
```

Description:

`exp_baumgartner2014(flag)` reproduces figures of the study from Baumgartner et al. (2014).

The following flags can be specified

`'fig2'` Reproduce Fig.2: Binaural weighting function best fitting results from Morimoto (2001) labeled as [1] and Macpherson and Sabin (2007) labeled as [2] in a least squared error sense.

`'fig3'` Reproduce Fig.3: Prediction examples. Actual responses and response predictions for three exemplary listeners when listening to median-plane targets in the baseline condition. Actual response angles are shown as open circles. Probabilistic response predictions are encoded by brightness according to the color bar to the right. Actual (A:) and predicted (P:) quadrant error rates (QE) and local polar RMS errors (PE) are listed above each panel.

`'fig4'` Reproduce Fig.4: Model parametrization. Partial and joint prediction residues as functions of the degree of selectivity and the motoric response scatter. Residuum functions are normalized to the minimum residuum obtained for the optimal parameter value.

The Auditory Modelling Toolbox

Categories
- Main
- Core functions
- Models
- Model stages
- Demos
- **Experiments**
- Common functions
- Plot
- Signals
- Data
- Auxdata
- Cache

EXP_BAUMGARTNER2014 - Results from Baumgartner et al. (2014)

Usage:

```
data = 
```

Descript

`exp_baum`

The followi

'fig2'

'fig3'

'fig4'

Categories
- Main
- Core functions
- **Models**
- Model stages
- Demos
- Experiments
- Common functions
- Plot
- Signals
- Data
- Auxdata
- Cache

See also
- baumgartner2014_pmv2ppp
- baumgartner2014_virtualexp

BAUMGARTNER2014 - Localization in saggital planes (robust, linear periphery)

Usage:

```
[p,rang] = baumgartner2014( target,template )
[p,rang,tang] = baumgartner2014( target,template )
[p,rang,tang] = baumgartner2014( target,template, varargin )
[err,pred] = baumgartner2014( target,template,errorflag )
```

Input parameters:

`target`	binaural impulse response(s) referring to the directional transfer function(s) (DFTs) of the target sound(s). Option 1: given in SOFA format -> sagittal plane DTFs will be extracted internally. Option 2: binaural impulse responses of all available listener-specific DTFs of the sagittal plane formatted according to the following matrix dimensions: time x direction x channel/ear
`template`	binaural impulse responses of all available listener-specific DTFs of the sagittal plane referring to the perceived lateral angle of the target sound. Options 1 & 2 equivalent to *target*.

Output parameters:

`p`	predicted probability mass vectors for response angles with respect to target positions 1st dim: response angle 2nd dim: target angle
`rang`	polar response angles (after regularization of angular sampling)
`tang`	polar target angles (usefull if sagittal-plane HRTFs are extracted directly from SOFA object)
`err`	predicted localization error (acc. to performance measure defined in *errorflag*
`pred`	structure with fields *p*, *rang*, *tang*

Aalto-yliopisto
Aalto-universitetet
Aalto University

The Auditory Modelling Toolbox

Peripheral models
Gammatone filterbank
Linear filtering for monaural masking (basic)
Linear filtering for monaural masking (improved)
Invertible Gammatone filterbank
Dual-resonance nonlinear filterbank (DRNL)
Fast acting compression (CARFAC) model
Cochlear transmission-line model (basic)
Cochlear transmission-line model (improved)
Cochlear transmission-line model (improved, incl. brainstem)
Auditory-nerve spike generation
Auditory-nerve filterbank (basic)
Auditory-nerve filterbank (improved)
Auditory nerve filterbank (improved, ready for brainstem)
Compression in the simultaneous masker phase effect

Binaural processing
Binaural processing
Binaural masking level difference
Binaural masking level difference (dynamic sources)
Binaural activity (based on cross-correlation)
Binaural signal detection
Binaural detection model based on interaural coherence
ITDs of hearing-aid users
Binaural activity map

Temporal-modulation sensitivity
Brainstem processing (CN and IC)
Auditory brainstem responses
Modulation filterbank (based on EPSM)
Modulation filterbank (based on nonlinear processing)
Modulation filterbank (based on DRNL)
Modulation (leaky-integrator model)
Non-linear adapation network

Monaural speech perception
Intelligibility in noise
Intelligibility in noise
Intelligibility with harmonic-cancellation
Short-time objective intelligibility

Binaural speech perception
Blind equalization-cancellation model
Binaural intelligibility in stationary noise (from BRIRs)
Binaural intelligibility in stationary noise
Binaural intelligibility of a reverberated speech target
Binaural intelligibility in non-stationary noise considering audibility
Binaural intelligibility in non-stationary noise (NH listeners only)

Perceptual similarity
Monaural perceptual similarity
Binaural perceptual similarity
Binaural perceptual similarity

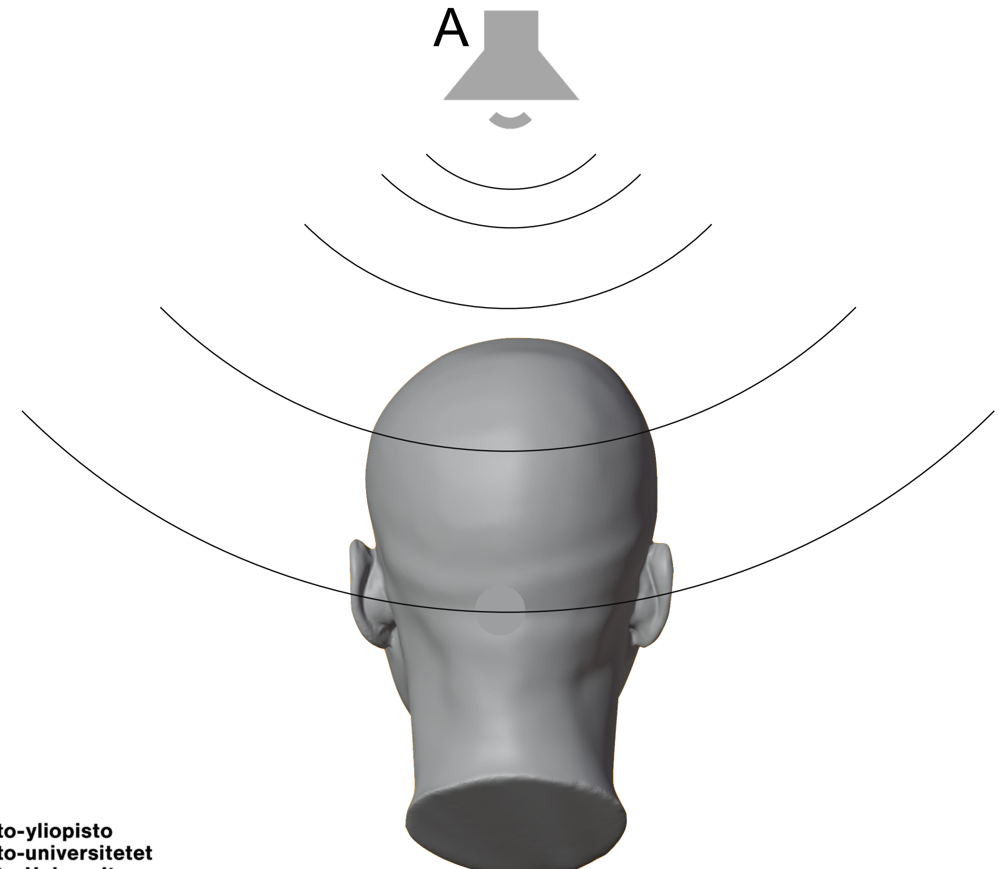
The Auditory Modelling Toolbox

Loudness models
Stationary sounds
Time-varying sounds
Binaural hearing impaired
Binaural loudness
Spatial models
Sound lateral direction
Lateralization, supervised training
´Lateralization in cochlear-implant listeners
Contextual lateralization based on interaural level differences
Median-plane localization
Vertical-plane localization (simple)
Sagittal-plane localization (simple)
Sagittal-plane localization (robust)
Sagittal-plane localization (nonlinear, for hearing impairements)
Sound externalization (ILD based)
Sound externalization (multi-cue)
Sound externalization (reverberant spaces)
Distance perception
Bayesian spherical sound localization (basic)
Bayesian spherical sound localization (multi-feature)
Bayesian sound localization (dynamic, ITD-based)
Lateralization in sound reproduction systems
Directional time-of-arrival (on-axis only)
Directional time-of-arrival in HRTFs (off-axis, robust)

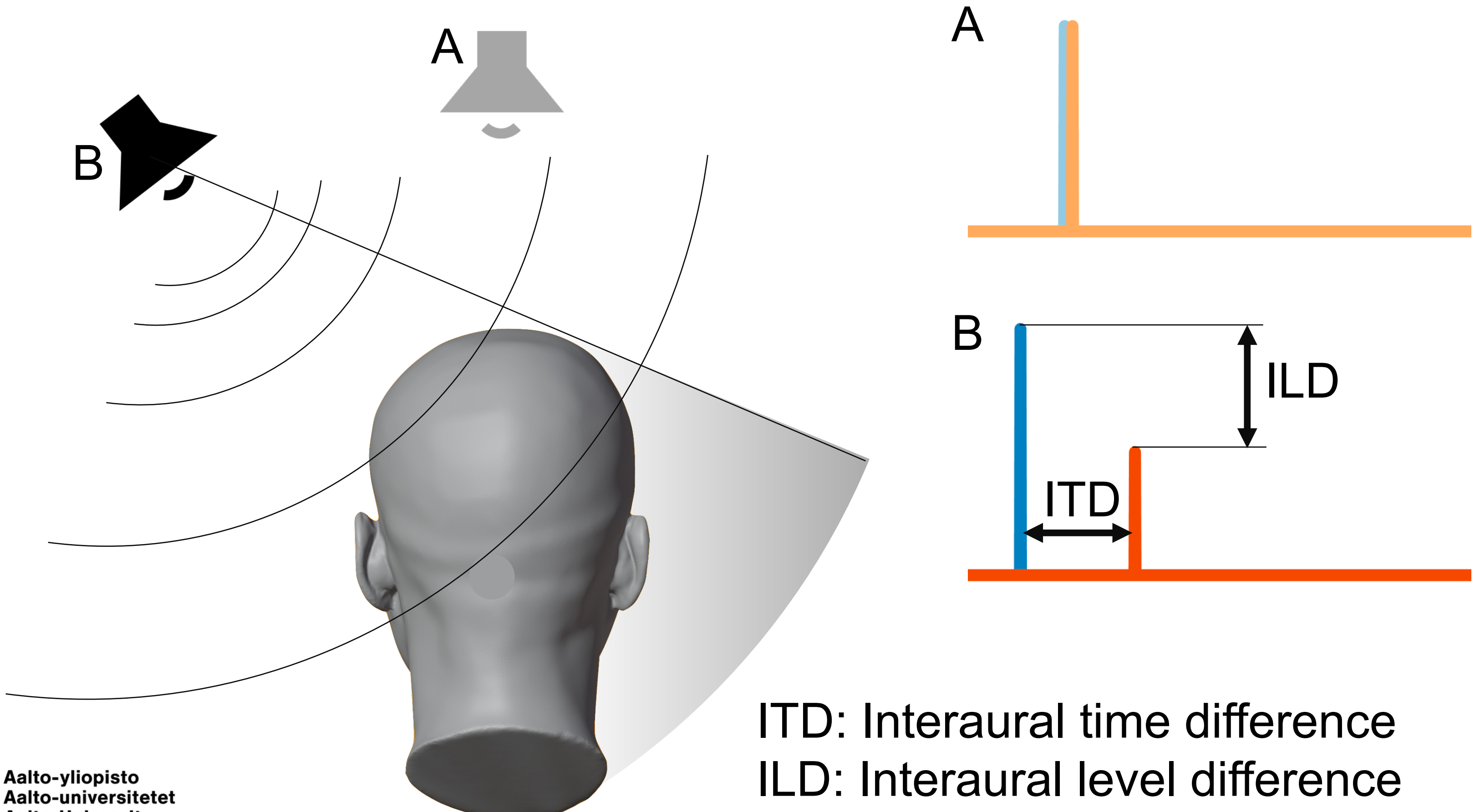
Aalto-yliopisto
Aalto-universitetet
Aalto University

Assessment of spatial attributes

The position of a sound source in the left-right dimension generates interaural differences



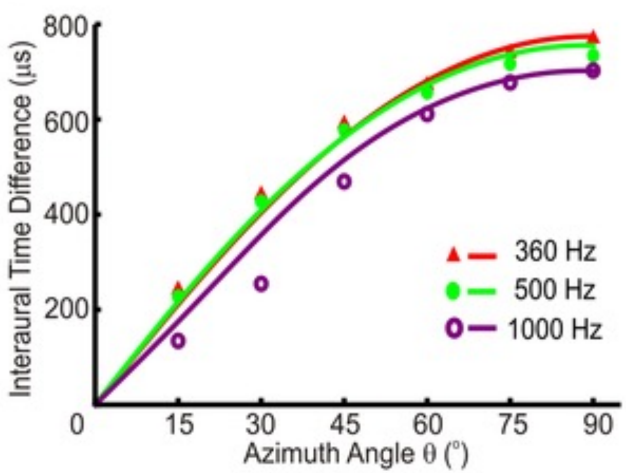
The position of a sound source in the left-right dimension generates interaural differences



ITD: Interaural time difference
ILD: Interaural level difference

Interaural time differences are dominant at lower frequencies (< ~ 1.5 kHz)

JND*: ~ 10 μs
ITD for lateral sources: ~ 600 - 700 μs

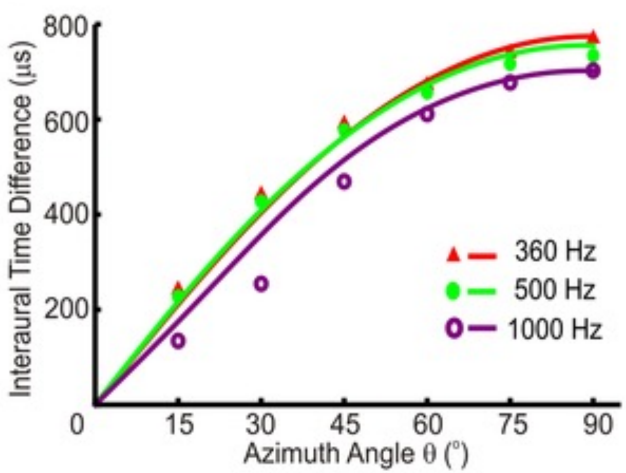


* JND: Just noticeable difference

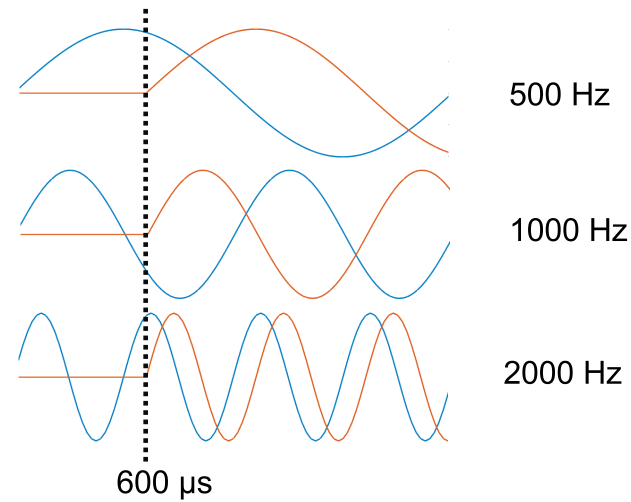
Smith et al. (2014)

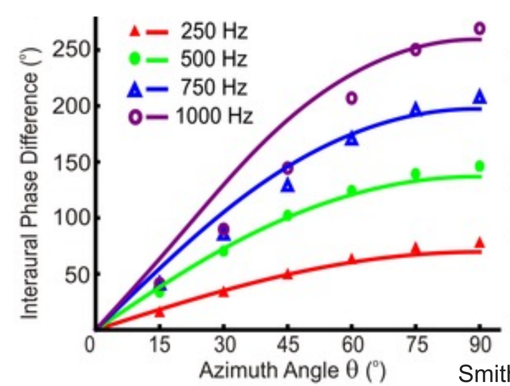
Interaural time differences are dominant at lower frequencies (< ~ 1.5 kHz)

JND*: ~ 10 μs
ITD for lateral sources: ~ 600 - 700 μs



Interaural phase difference (IPD)



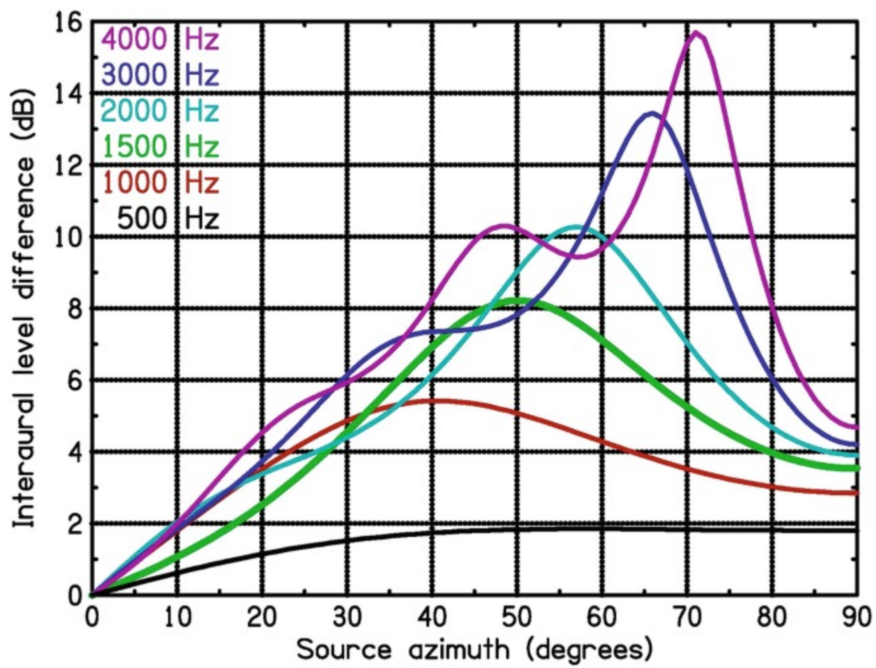


Smith et al. (2014)

* JND: Just noticeable difference

The importance of interaural level differences grows with frequency

They are dominant above ~ 1.5 kHz
JND: ~ 0.5 - 1 dB



These interaural cues are not enough to resolve the location in the three-dimensional space

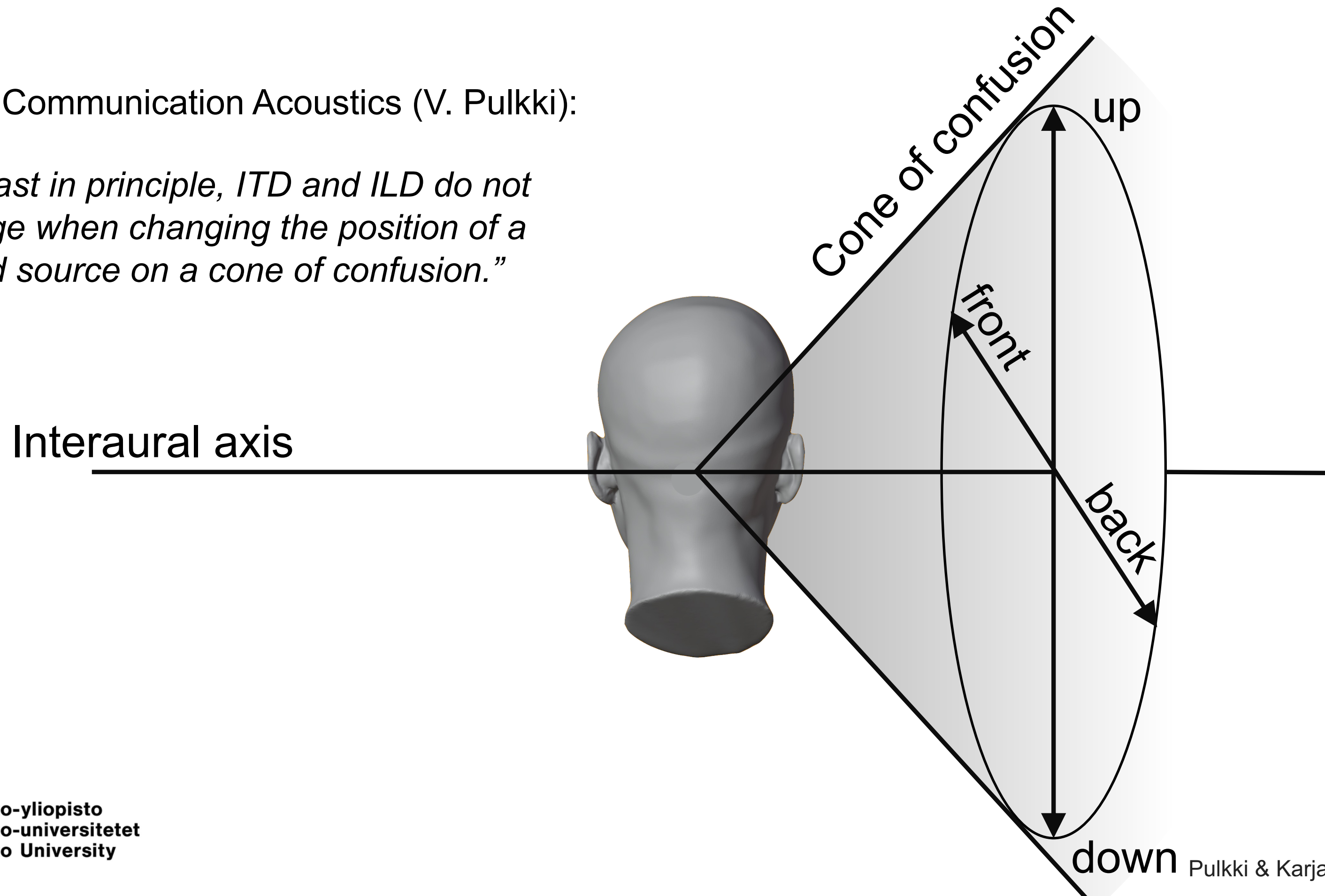
From Communication Acoustics (V. Pulkki):

“*At least in principle, ITD and ILD do not change when changing the position of a sound source on a cone of confusion.*”

These interaural cues are not enough to resolve the location in the three-dimensional space

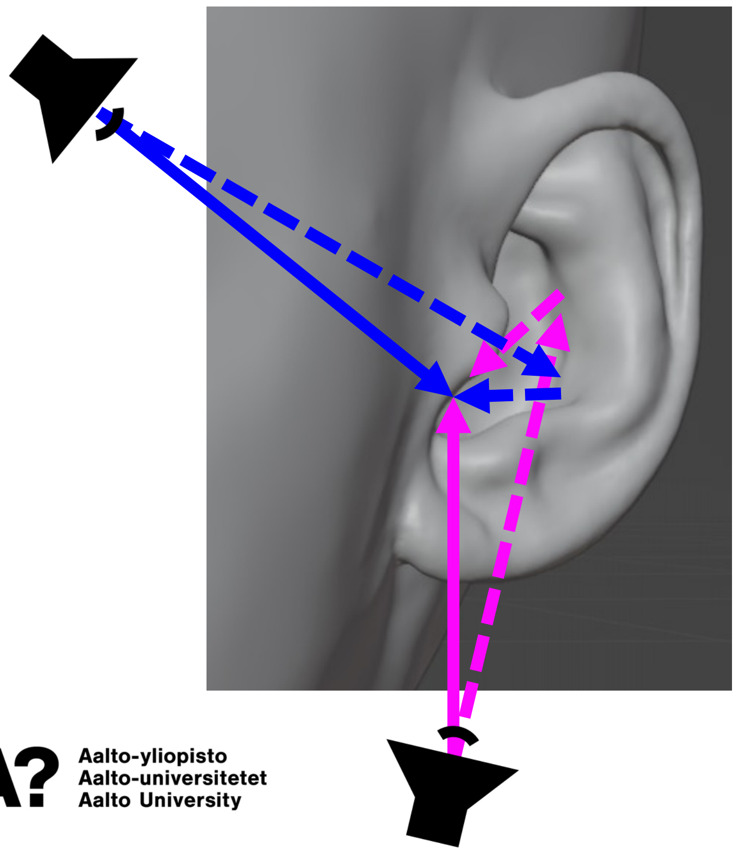
From Communication Acoustics (V. Pulkki):

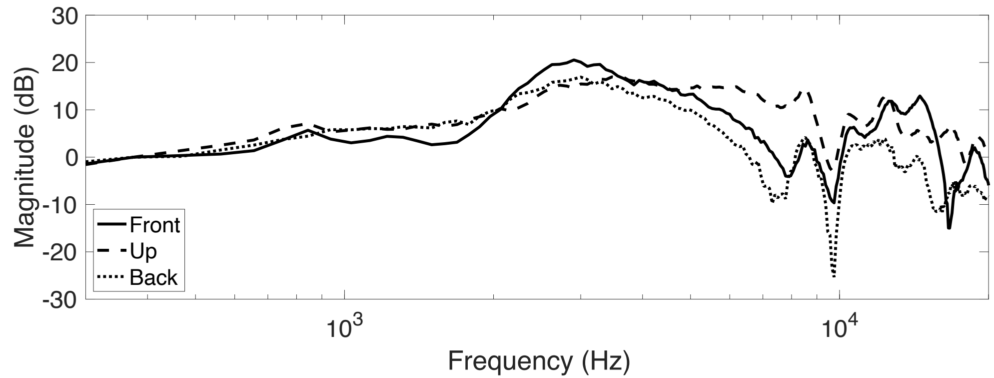
"*At least in principle, ITD and ILD do not change when changing the position of a sound source on a cone of confusion.*"



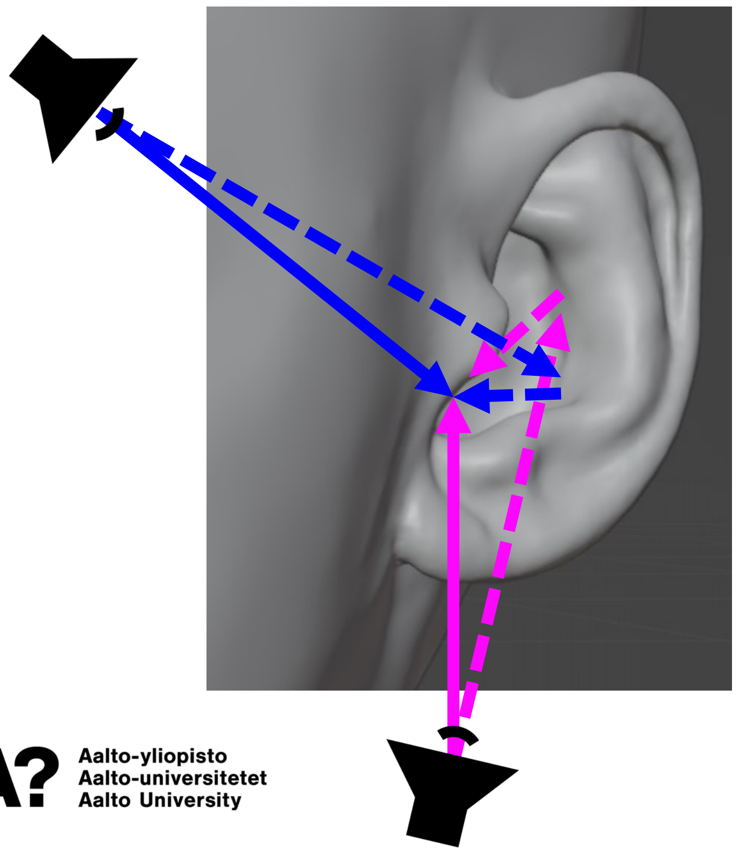
Pulkki & Karjalainen (2015)

Moving a up-down or front-back on the cone of confusion modifies the spectrum of the sound

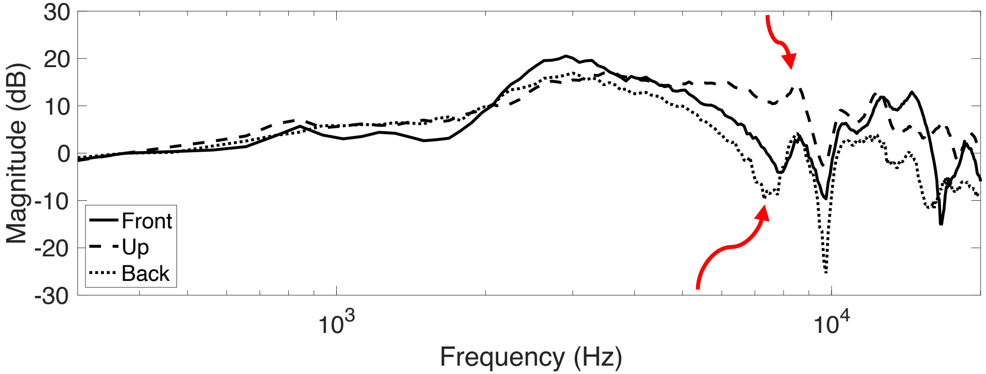




Moving a up-down or front-back on the cone of confusion modifies the spectrum of the sound



The frequency location of main peaks and notches seems crucial

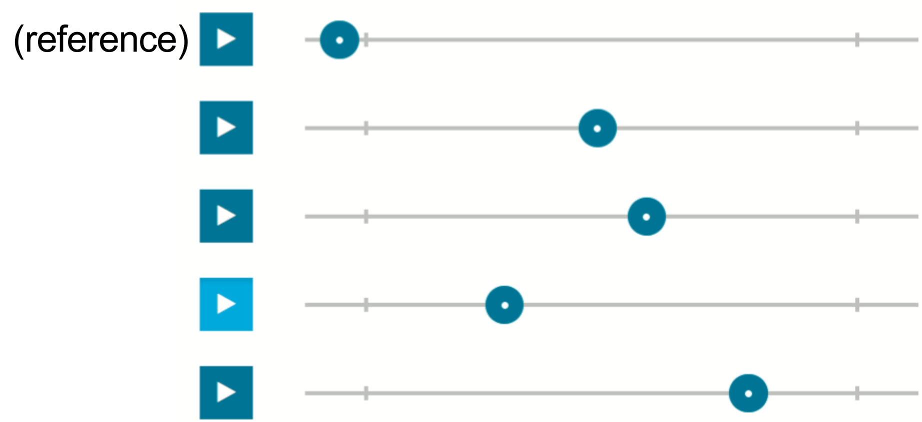


Aalto-yliopisto
Aalto-universitetet
Aalto University

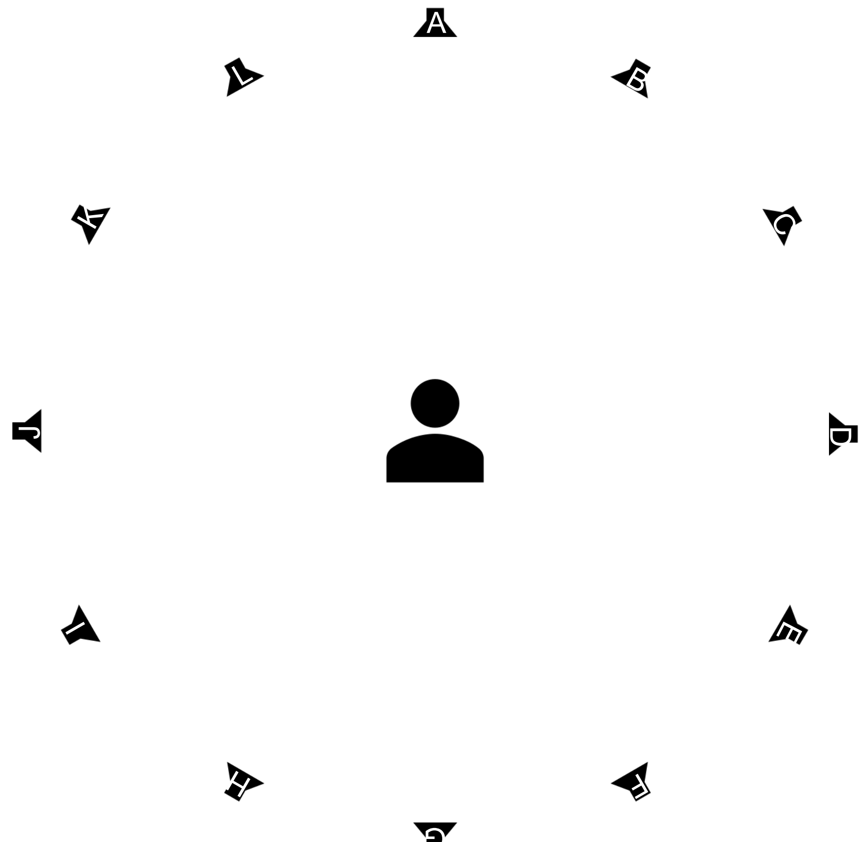
The spatial fidelity can be measured using a more hollistic approach

Rate the spatial fidelity of this samples compared to the reference

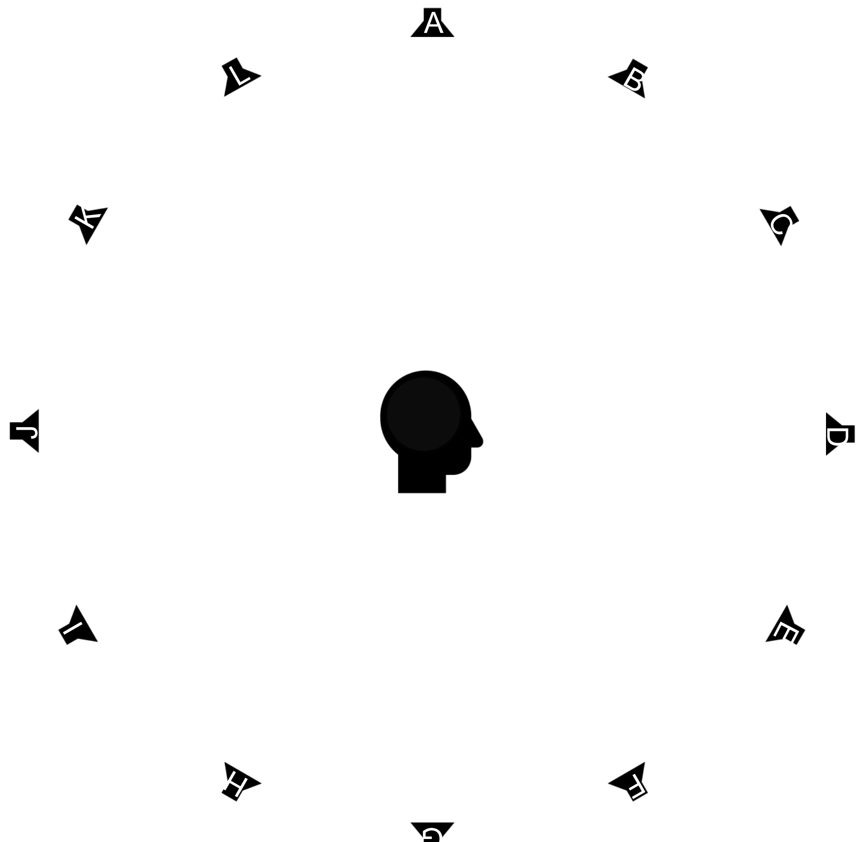
10 (very good) 0 (very bad)



Localisation(-like) tasks are useful to measure artifacts in spatial reproduction



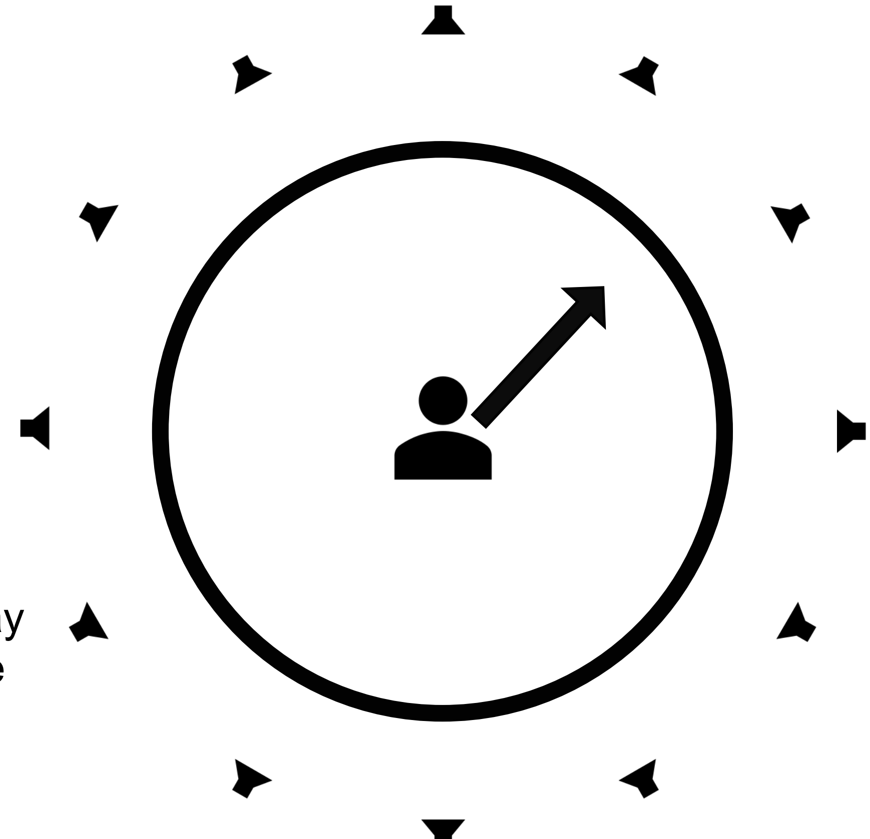
Localisation(-like) tasks are useful to measure artifacts in spatial reproduction



Localisation(-like) tasks are useful to measure artifacts in spatial reproduction

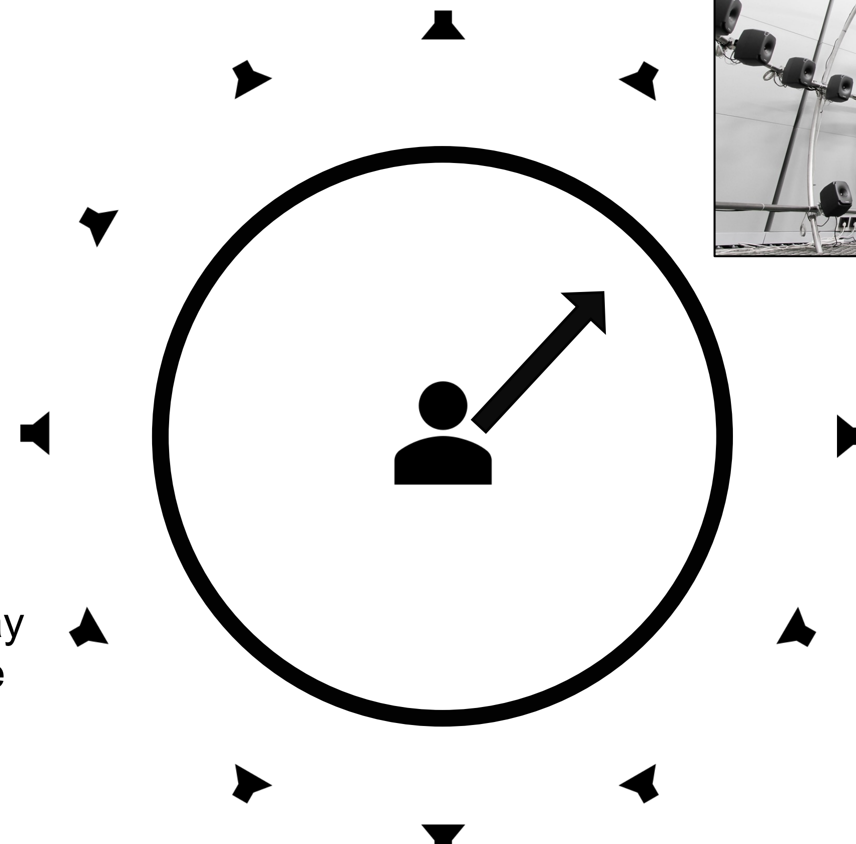
Front/back discrimination

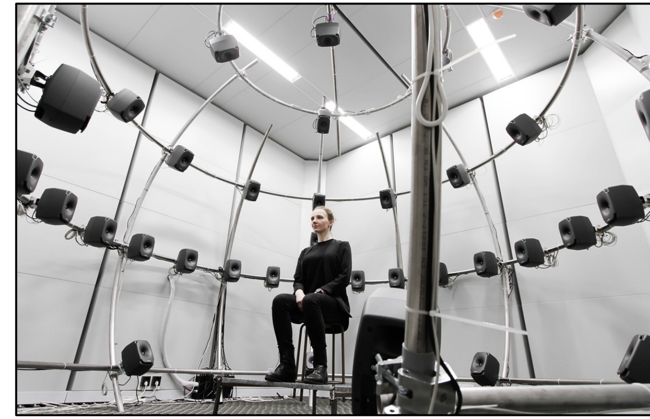
The classic localisation test



The pointing method may change significantly the results

The classic localisation test

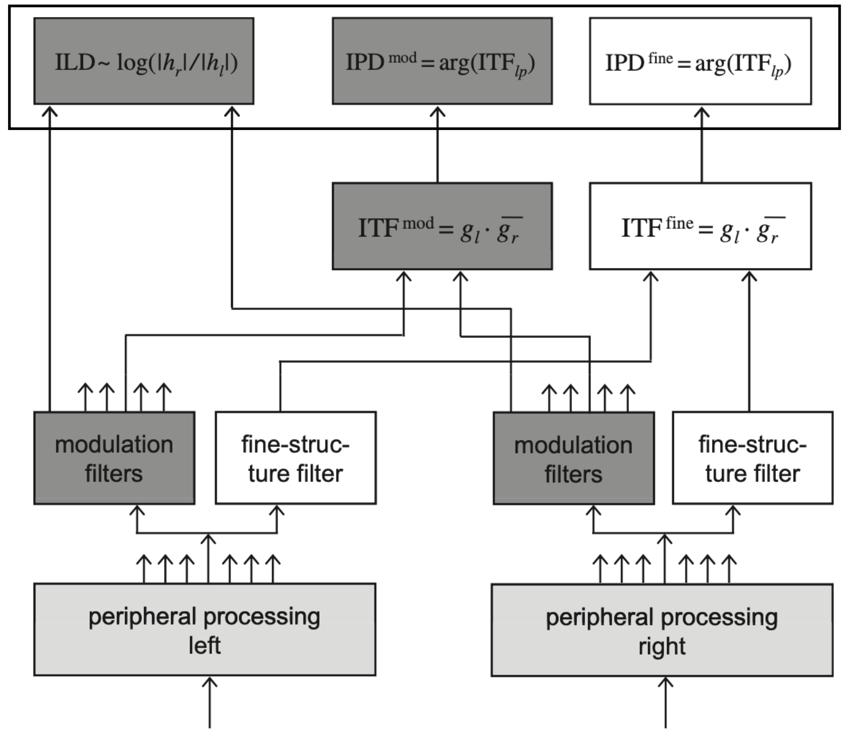




The pointing method may change significantly the results

Auditory models of human localisation estimate experimental data collected in localisation tests

DIETZ2011 - Sound lateral direction

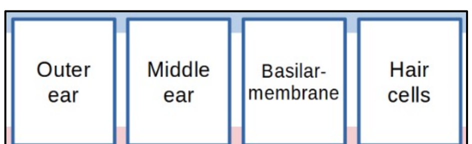


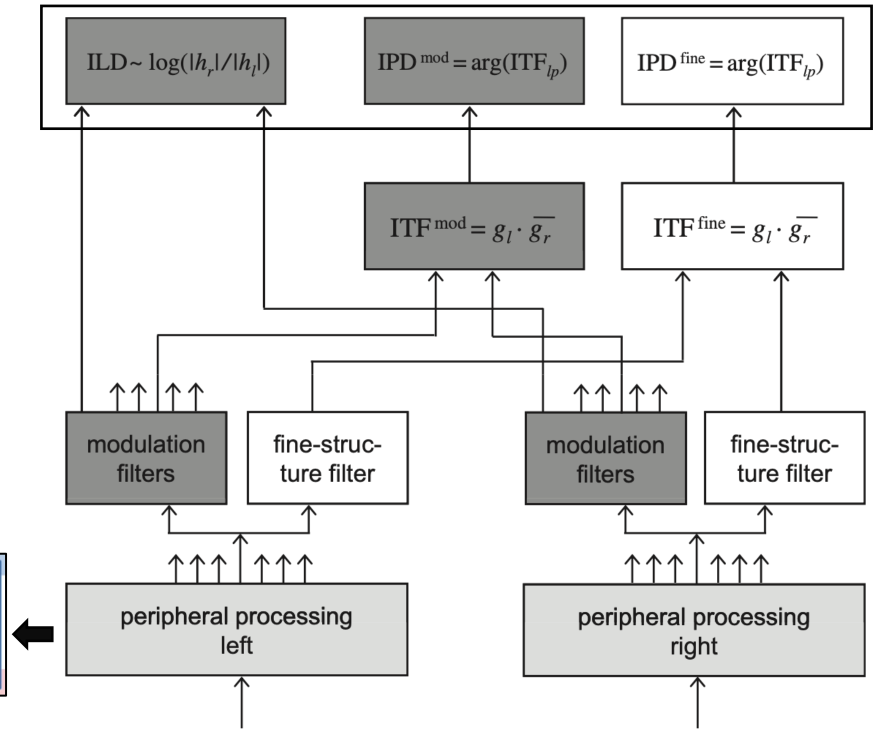
Binaural input signal

Dietz et al. (2011)

Auditory models of human localisation estimate experimental data collected in localisation tests

DIETZ2011 - Sound lateral direction





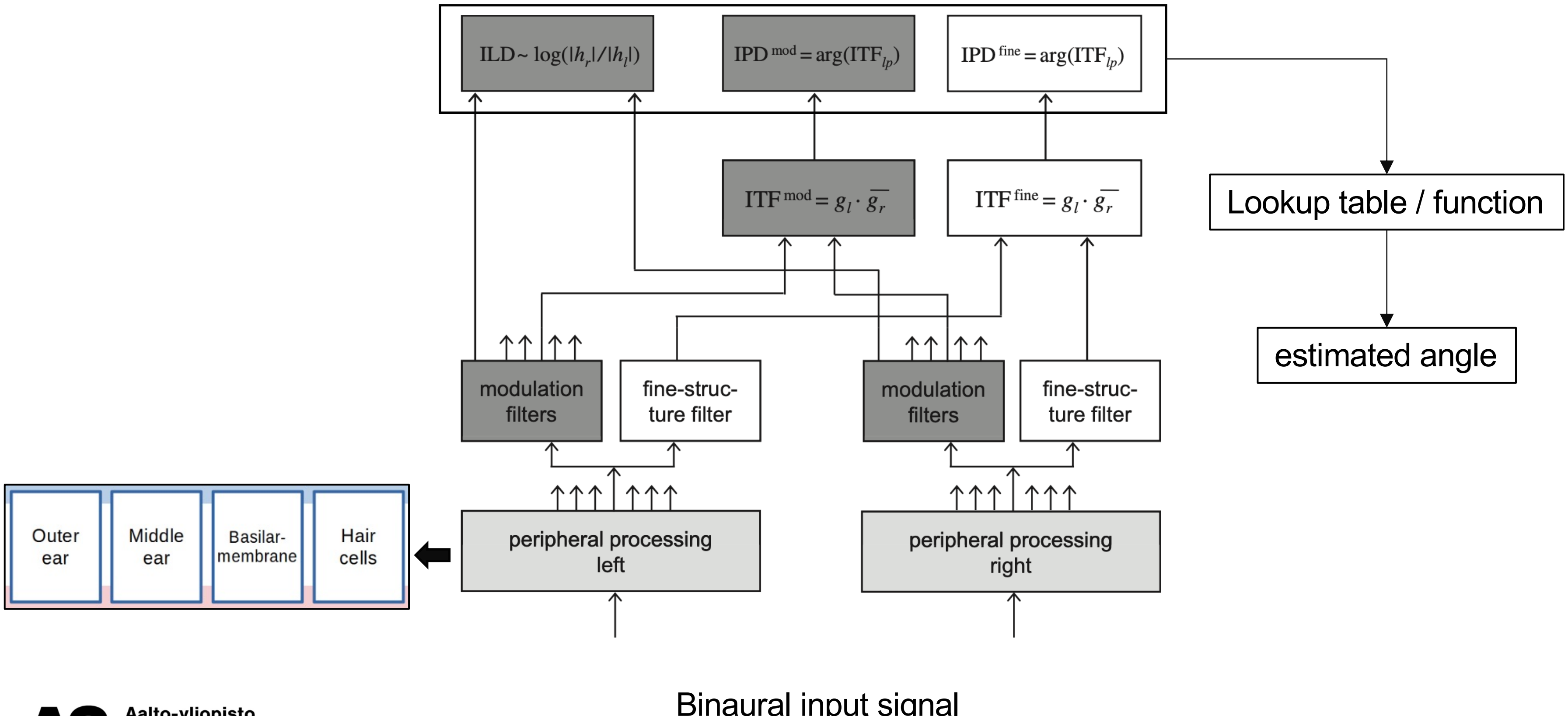
Binaural input signal

![Aalto University logo]

Dietz et al. (2011)

Auditory models of human localisation estimate experimental data collected in localisation tests

DIETZ2011 - Sound lateral direction

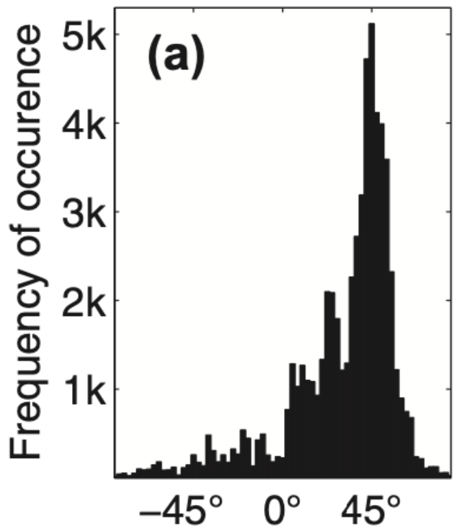


Dietz et al. (2011)

Auditory models of human localisation estimate experimental data in localisation tests

DIETZ2011 - Sound lateral direction

One speaker

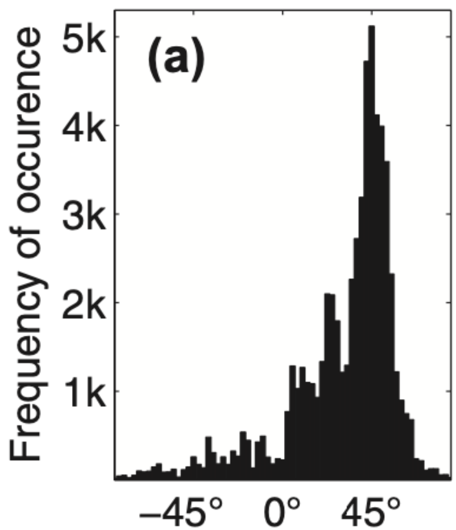


Dietz et al. (2011)

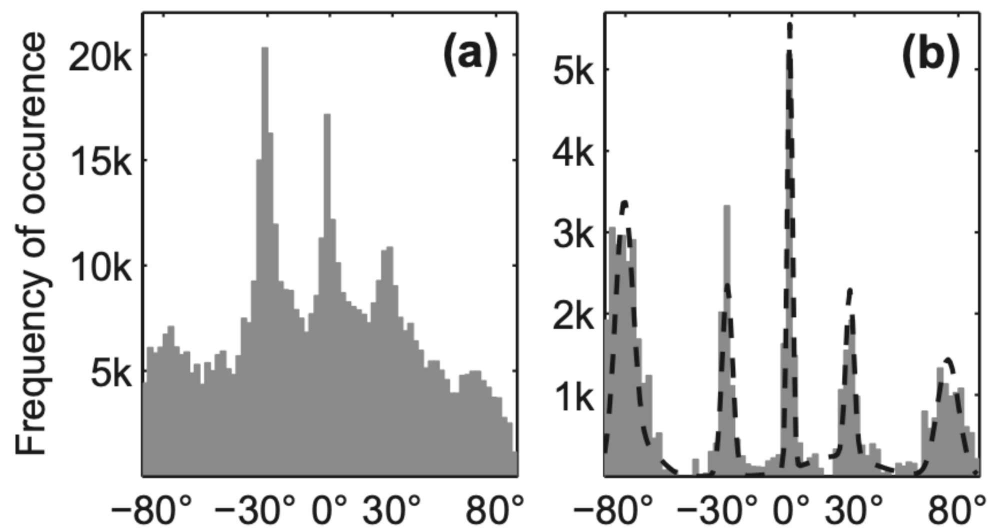
Auditory models of human localisation estimate experimental data in localisation tests

DIETZ2011 - Sound lateral direction

One speaker



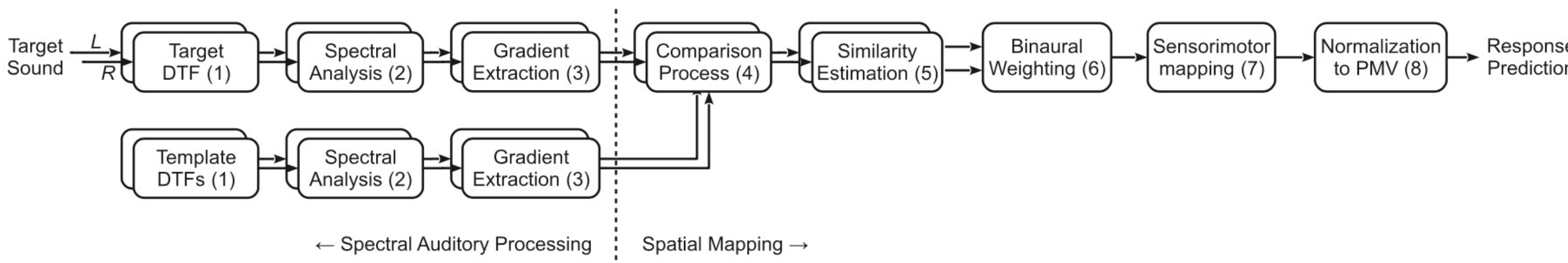
Five speakers



Dietz et al. (2011)

Models of sagittal plane localisation is often used both in research and in industry

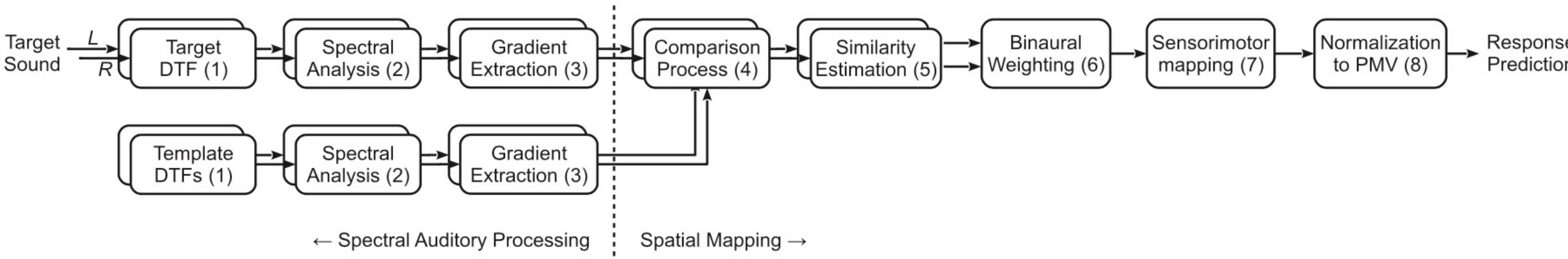
BAUMGARTNER2014 - Localisation in sagittal planes



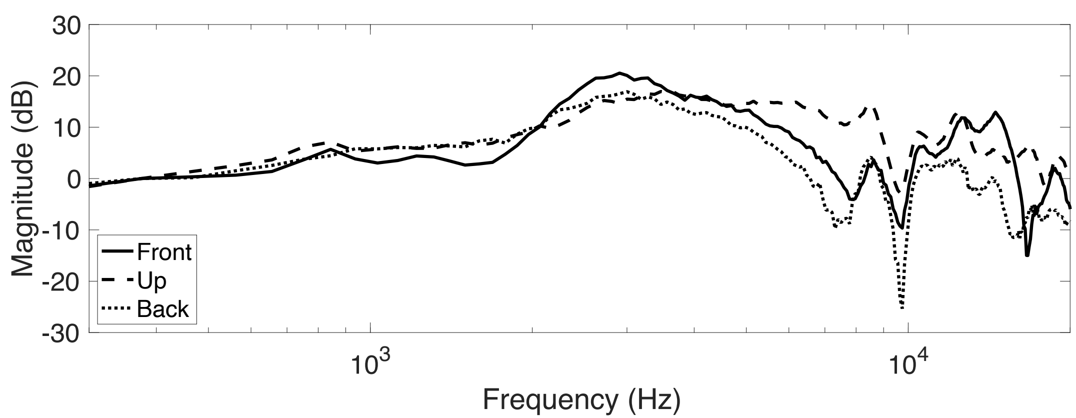
Baumgartner et al. (2014)

Models of sagittal plane localisation is often used both in research and in industry

BAUMGARTNER2014 - Localisation in sagittal planes



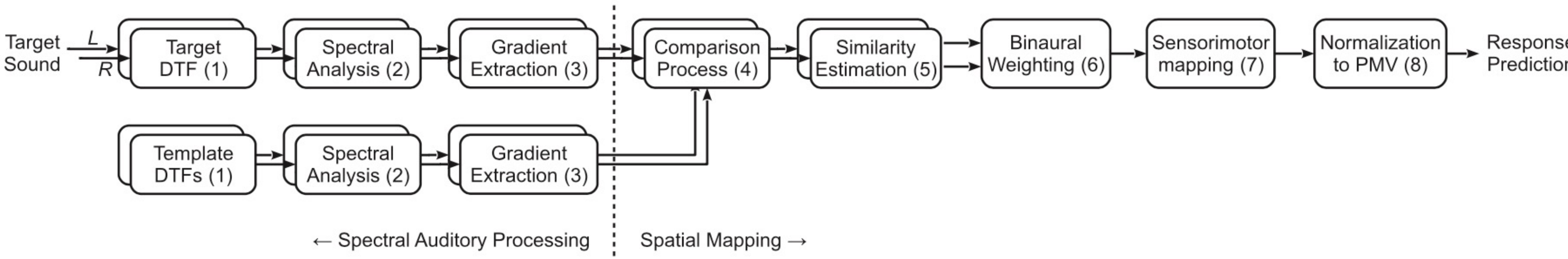
Template



Baumgartner et al. (2014)

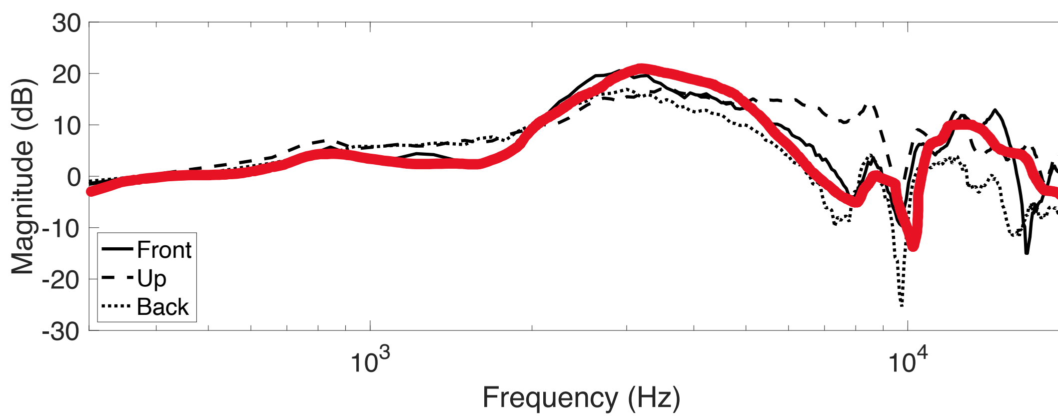
Models of sagittal plane localisation is often used both in research and in industry

BAUMGARTNER2014 - Localisation in sagittal planes



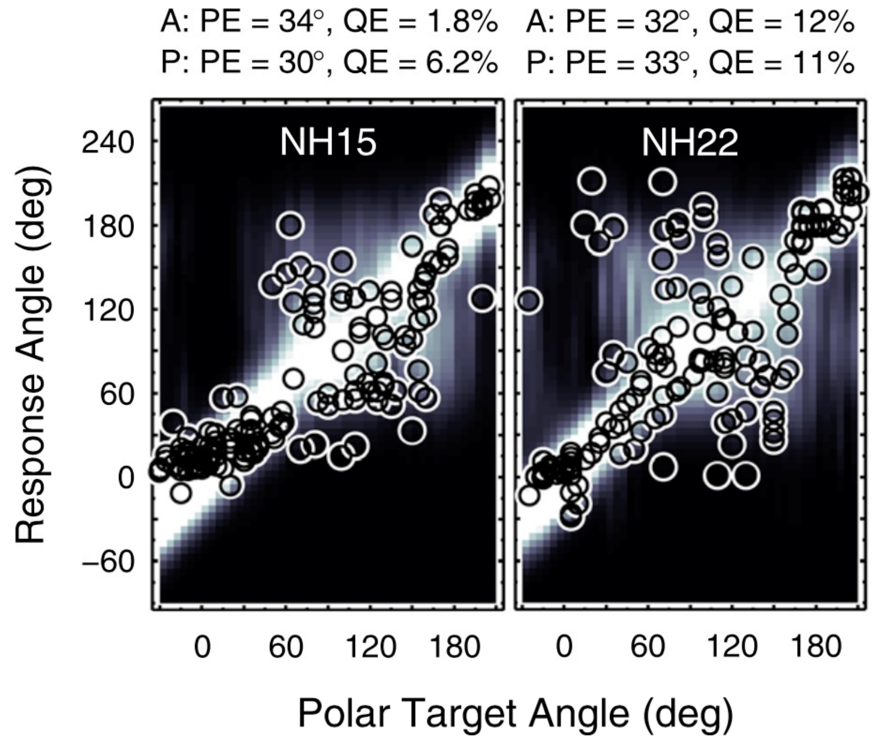
Template

Target sound



Models of sagittal plane localisation are used both in research and in industry

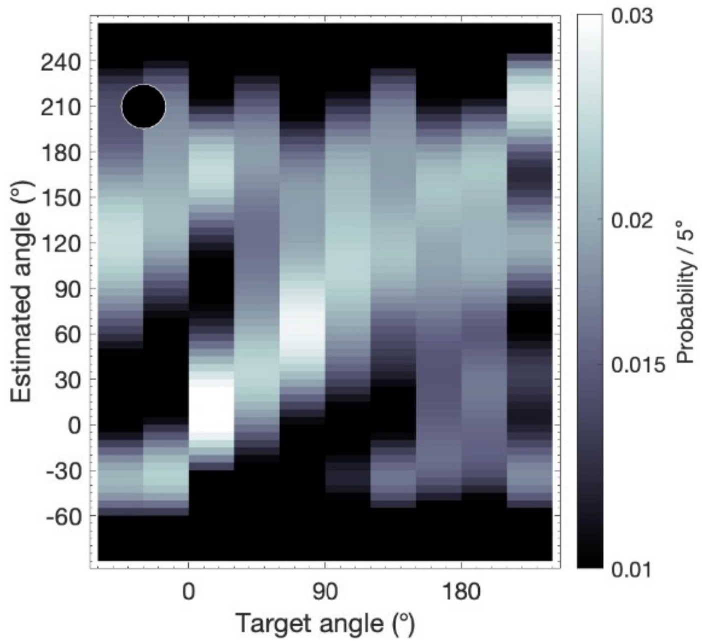
BAUMGARTNER2014 - Localisation in sagittal planes

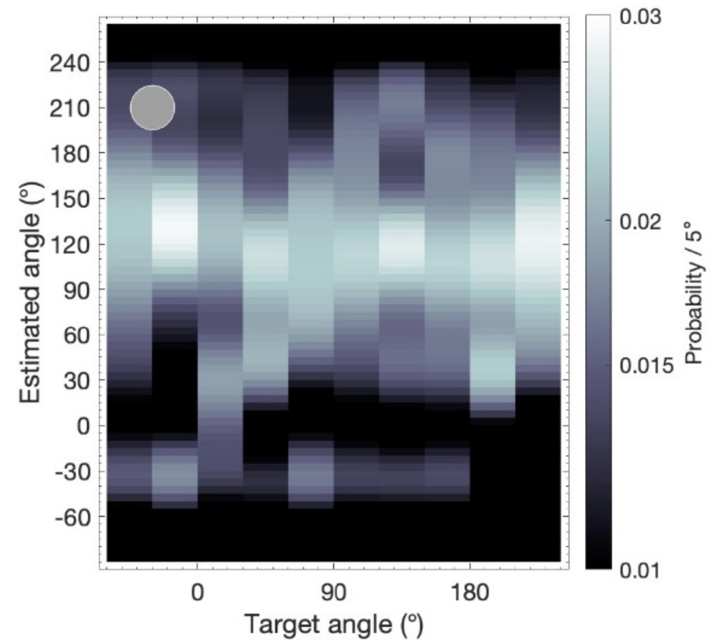


Baumgartner et al. (2014)

They help us estimate the availability of monaural spectral cues

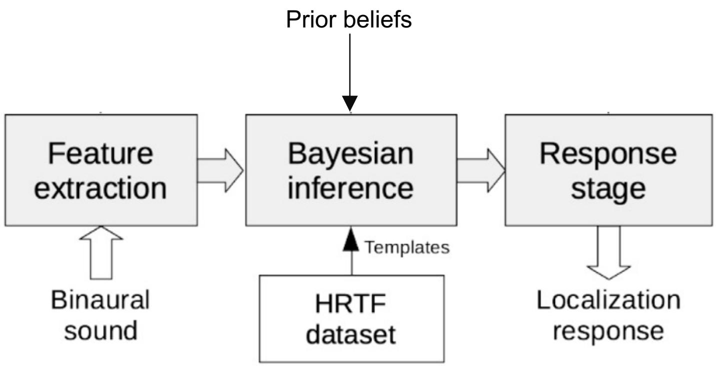
BAUMGARTNER2014 - Localisation in sagittal planes

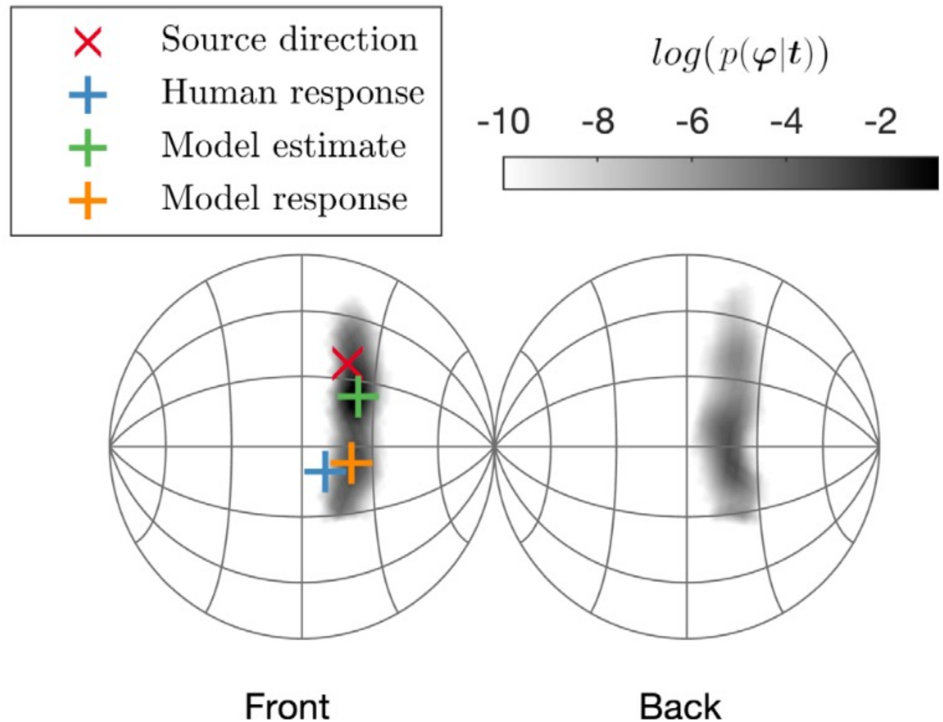




Models of spherical localisation combine the results of the two dimensions

BARUMERLI2023 - Bayesian spherical sound localization model

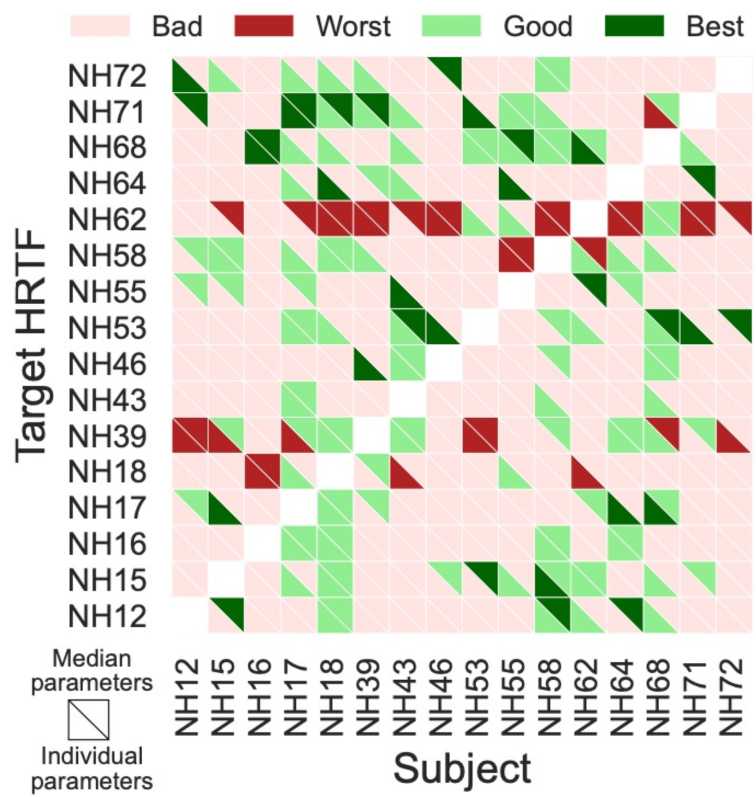




Barumerli et al. (2023)

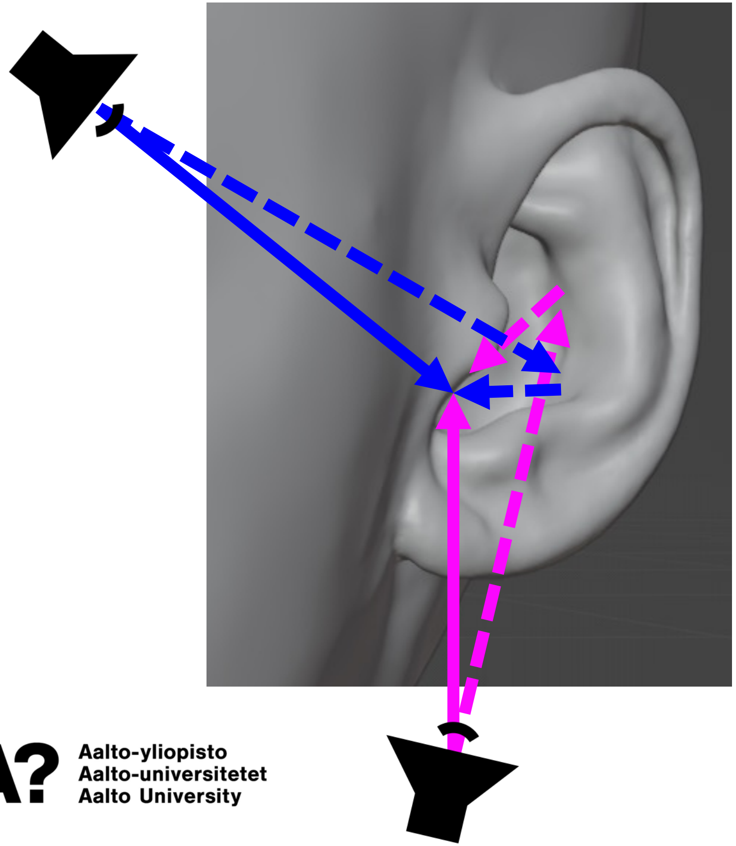
Example of model-based analysis of a dataset for non-individual HRTF selection

BARUMERLI2023 - Bayesian spherical sound localization model

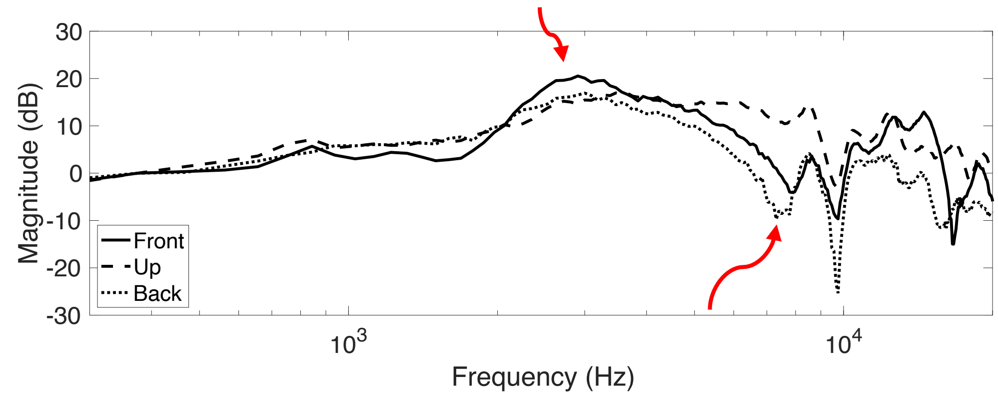


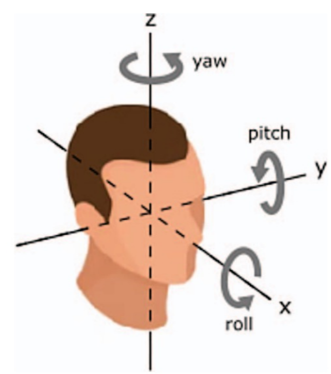
Daugintis et al. (2023)

Artifacts in sound spatial fidelity are less important when the listener can move



The frequency location of main peaks and notches seems crucial

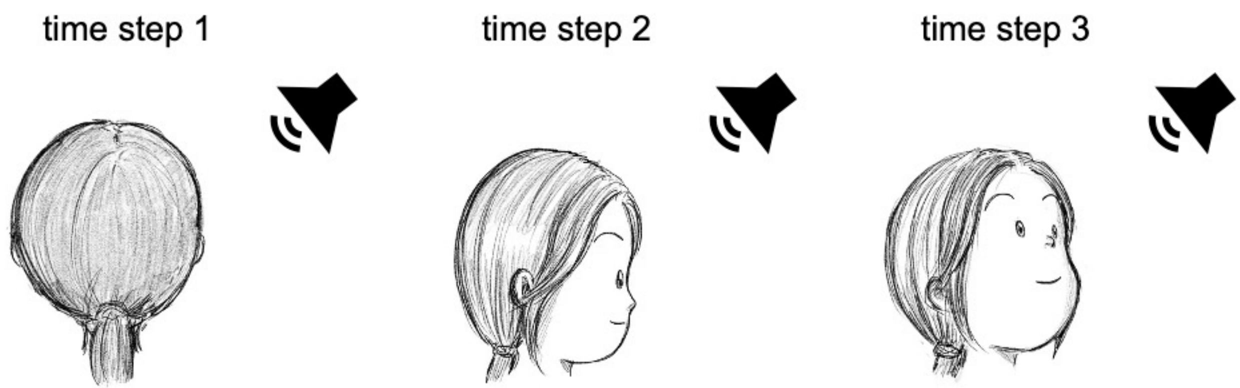




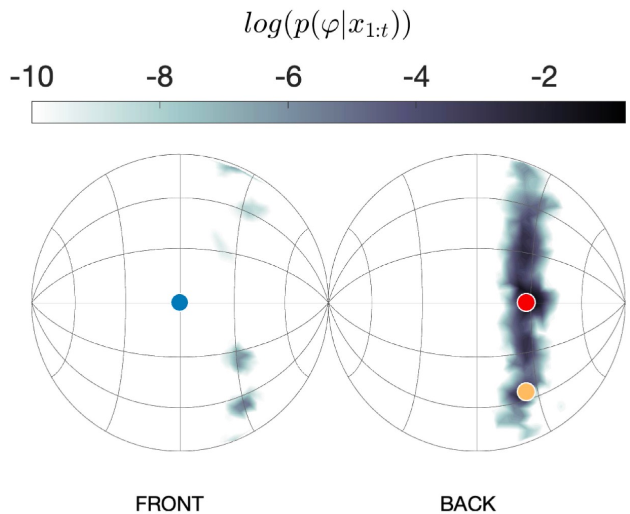
Head motion helps resolving the source location in the cone of confusion, particularly important for front-back discrimination

McLachlan et al. (2021)

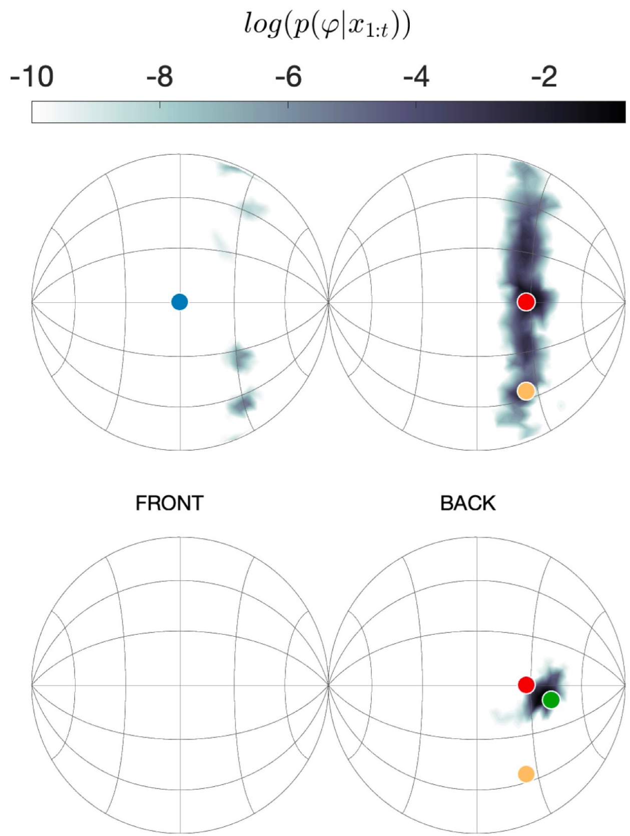
New models that account for head motion



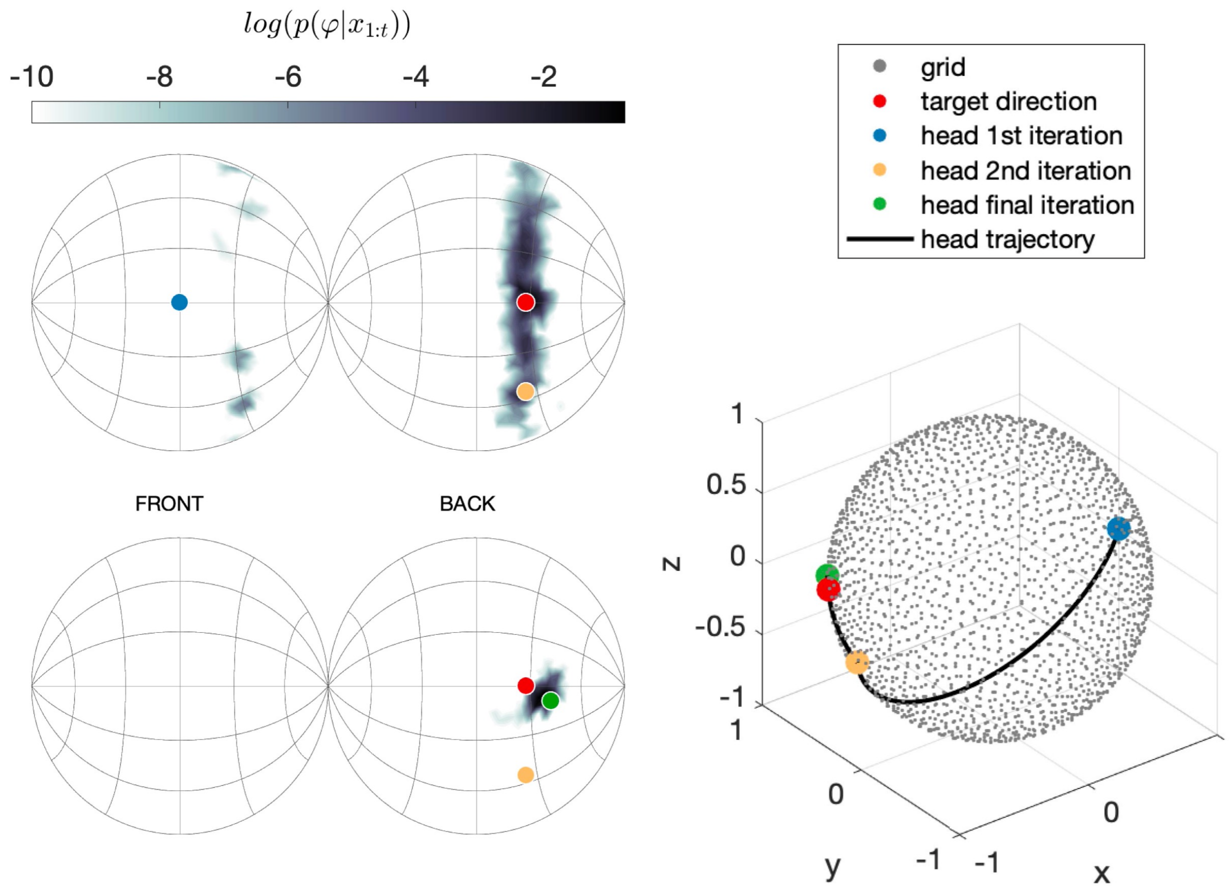
New models that account for head motion



New models that account for head motion

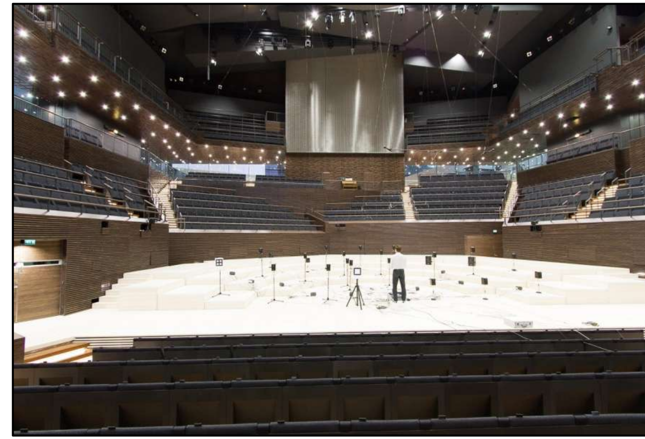


New models that account for head motion

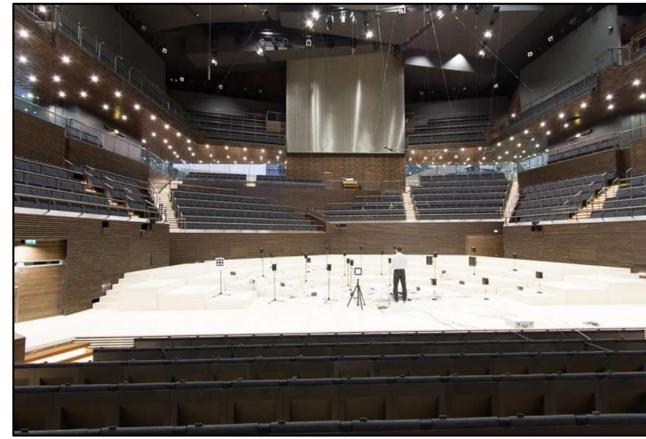


Aalto-yliopisto
Aalto-universitetet
Aalto University

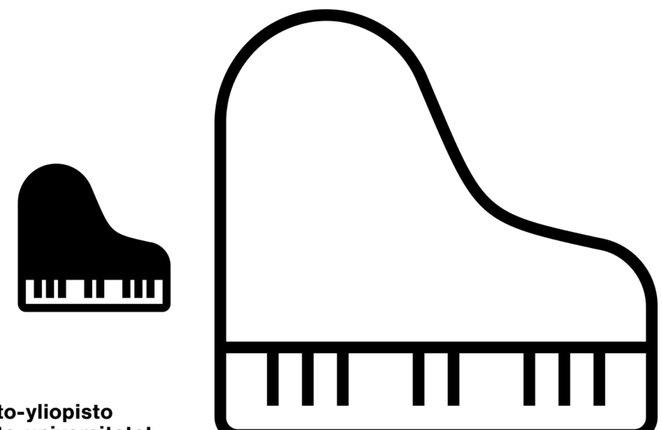
The spaciousness and how to assess it

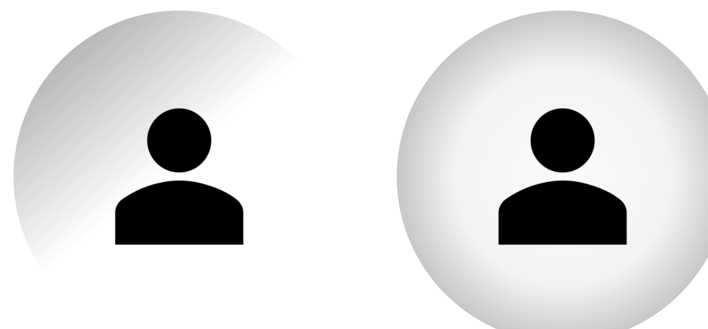


The spaciousness and how to assess it



Apparent source width (ASW) and **Envelopment**





The spaciousness and how to assess it

Which of these samples have more
[**Apparent source width (ASW) / Envelopment**]?

A B

▶ ▶

☐ ☒

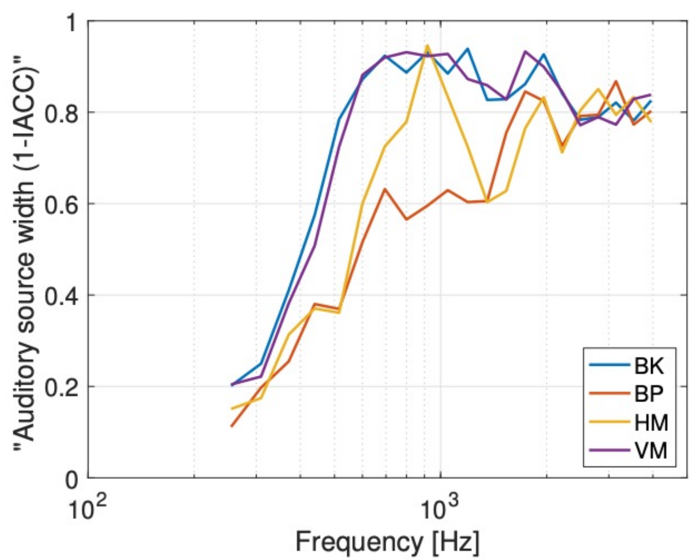
Lokki et al. (2023)

The spaciousness and how to assess it

Which of these samples have more
[**Apparent source width (ASW) / Envelopment**]?

A ☐

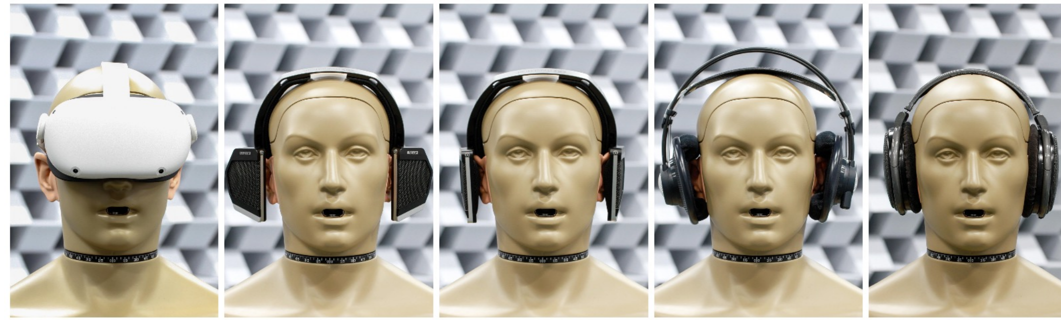
B ☒

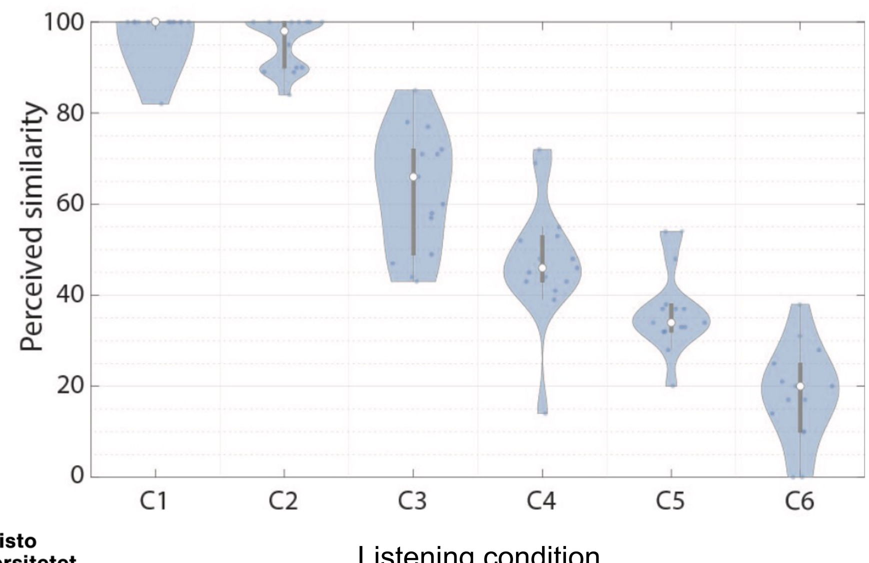


Lokki et al. (2023)

Colouration assessment

Colouration as differences in timbre compared to a reference signal

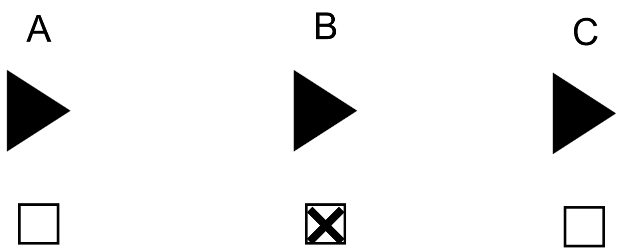


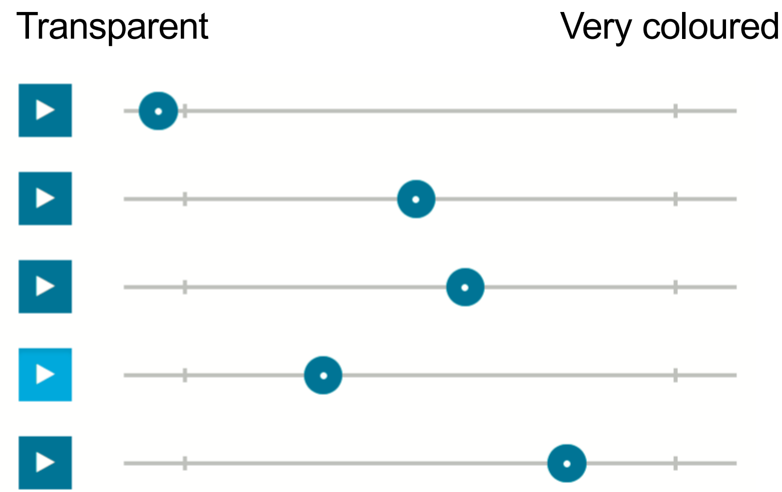


Lladó et al. (2022)

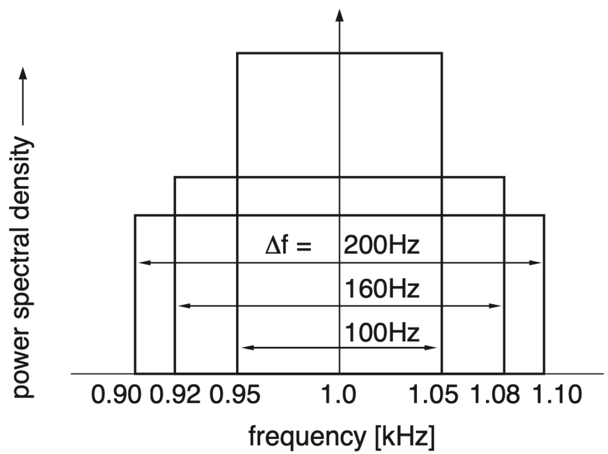
Colouration can be assessed using several methods

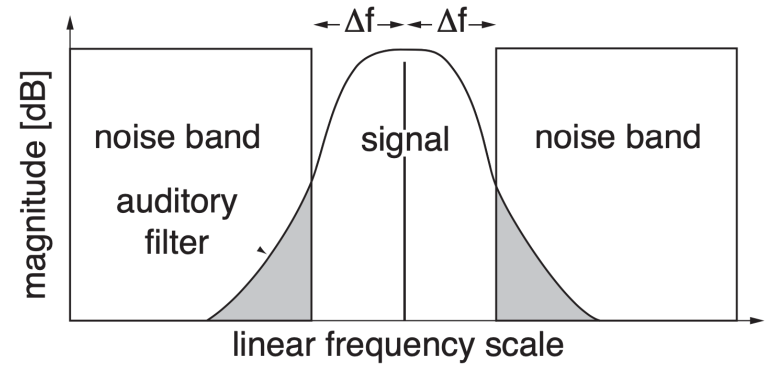
Which of these signals is different?



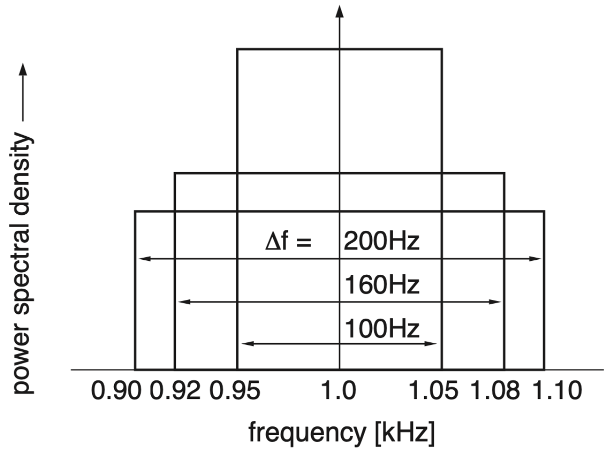


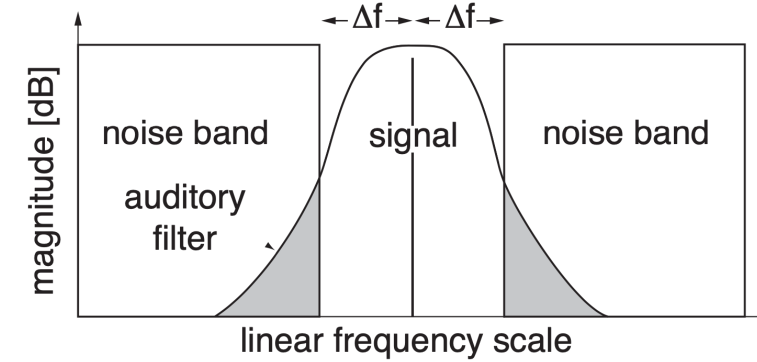
Colouration can be derived from basic concepts related to loudness and auditory filters

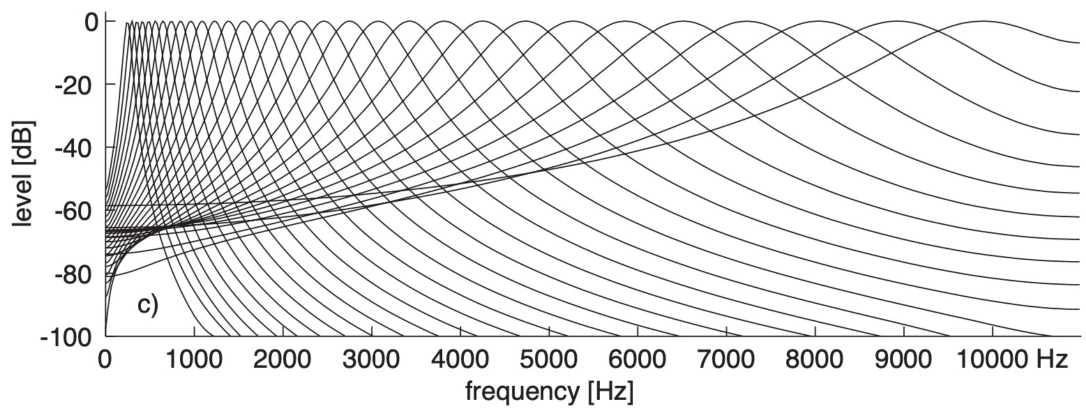




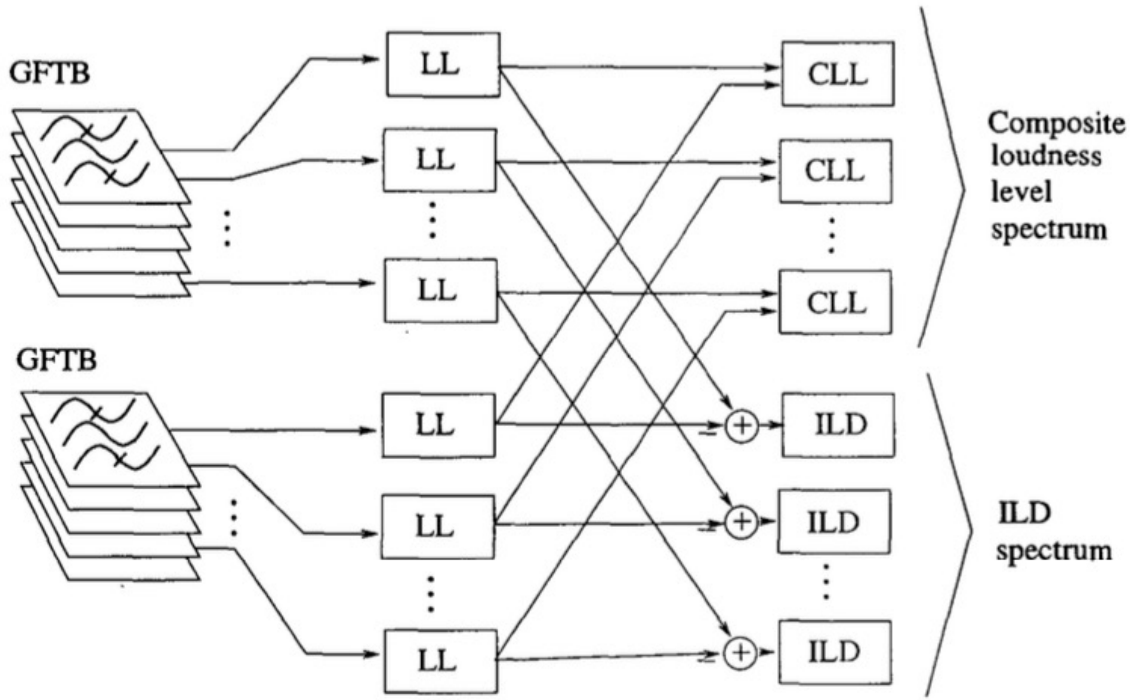
Colouration can be derived from basic concepts related to loudness and auditory filters







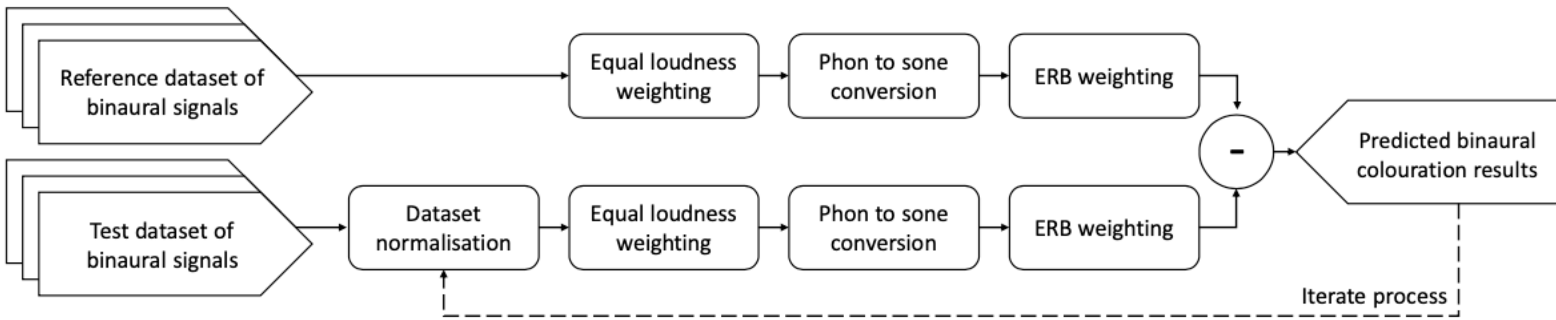
The specific loudness is connected to the perceived colouration



Pulkki et al. (1999)

Model of colouration

MCKENZIE2022 - Binaural perceptual similarity

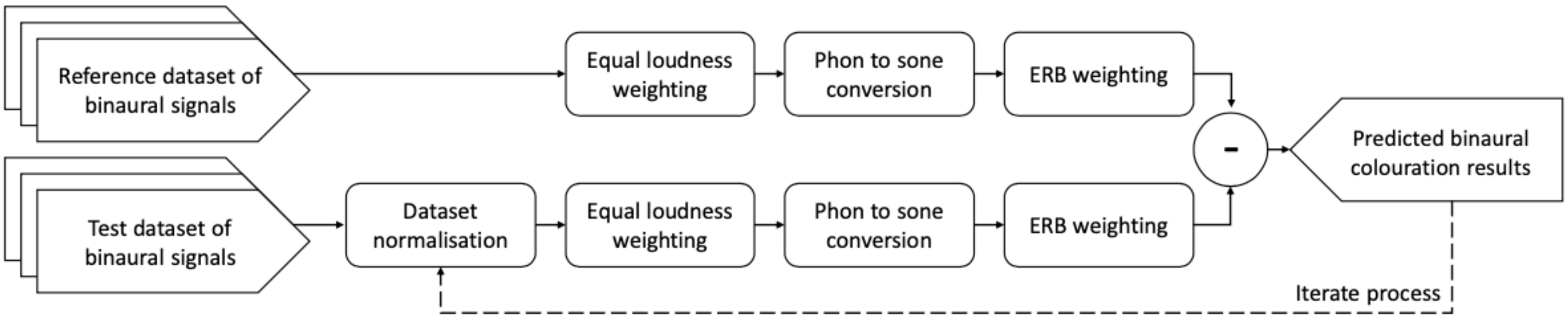


Aalto-yliopisto
Aalto-universitetet
Aalto University

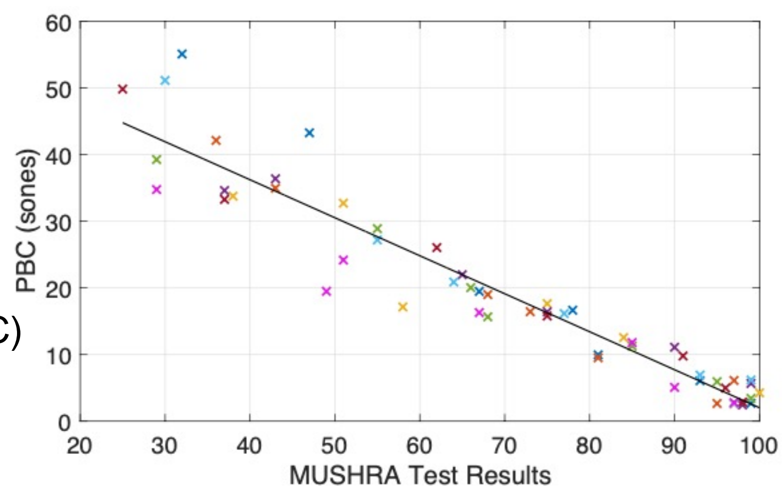
McKenzie et al. (2023)

Model of colouration

MCKENZIE2022 - Binaural perceptual similarity



Predicted binaural colouration (PBC)

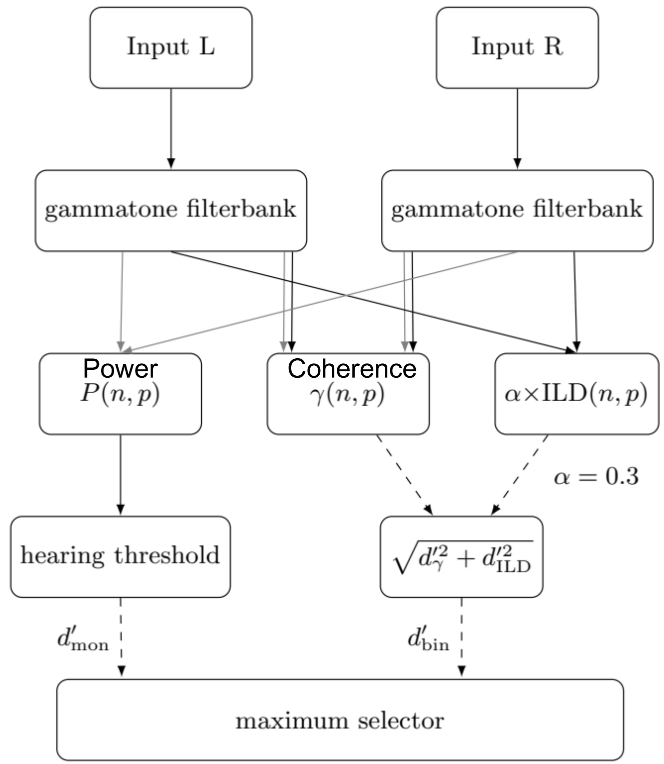


Aalto-yliopisto
Aalto-universitetet
Aalto University

McKenzie et al. (2023)

Assessing the overall quality

Overall binaural quality as a combination of binaural and monaural attributes

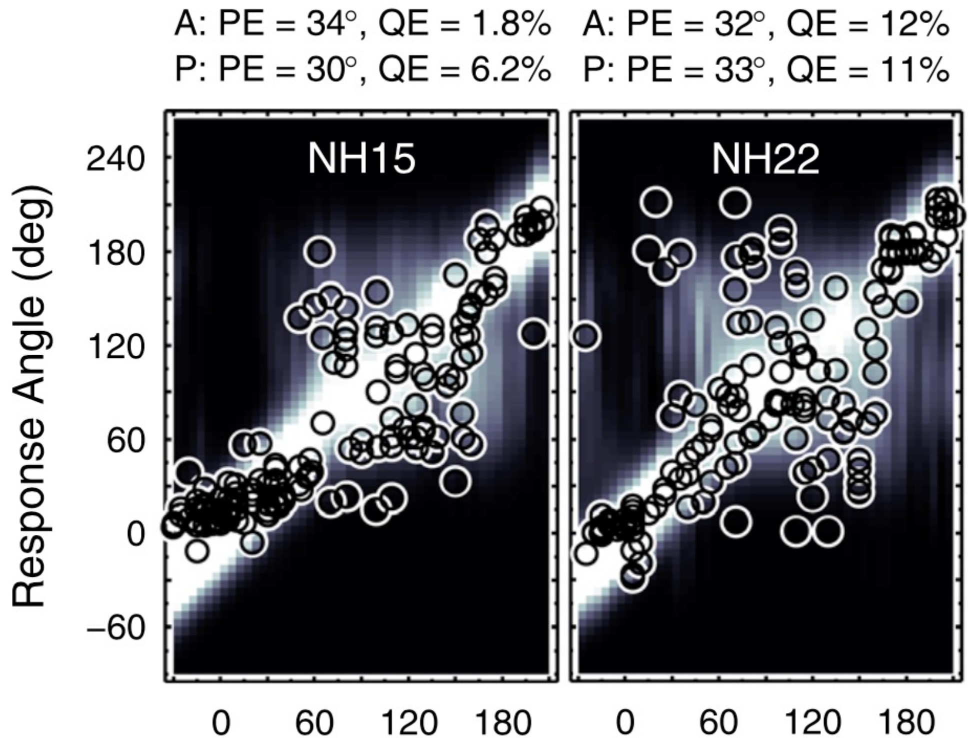


Eurich et al. (2023)

Summary

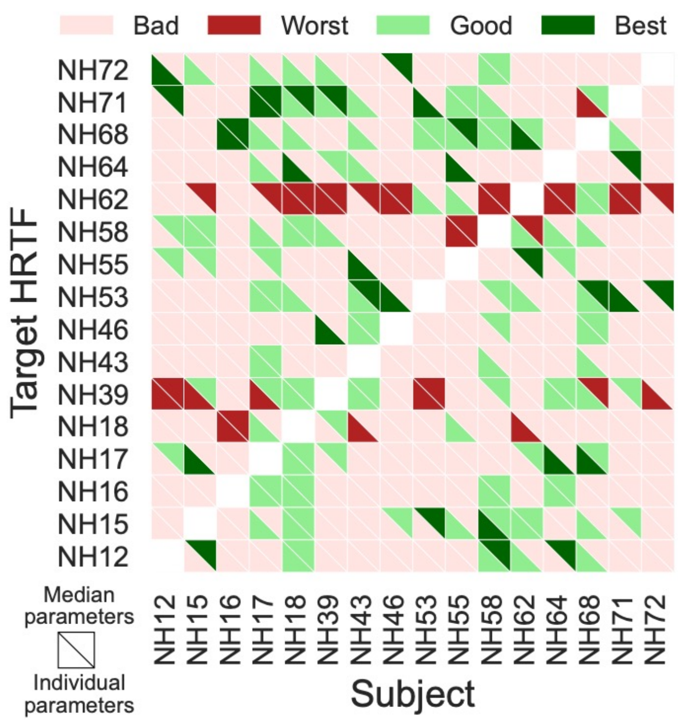
Listening experiments are the best option to assess sound quality

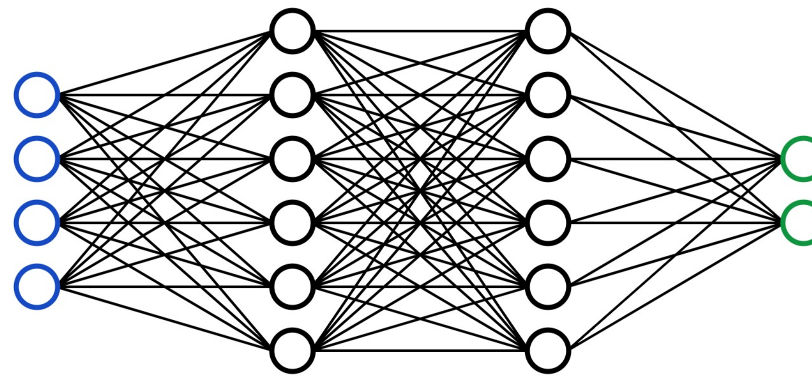
We can estimate experimental data obtained from listening experiments using auditory models



Aalto-yliopisto
Aalto-universitetet
Aalto University

Once these models have been validated, we *can* rely on them





pedro.llado@aalto.fi

References

[1] Le Callet, Patrick, Sebastian Möller, and Andrew Perkis. "Qualinet white paper on definitions of quality of experience." *European network on quality of experience in multimedia systems and services (COST Action IC 1003)* 3.2012 (2012).
[2] Zacharov, N. (Ed.). (2018). Sensory evaluation of sound. CRC Press
[3] Llado, P., McKenzie, T., Meyer-Kahlen, N., & Schlecht, S. J. (2022). Predicting perceptual transparency of head-worn devices. Journal of the Audio Engineering Society, 70(7/8), 585-600.
[4] Majdak, P., Hollomey, C., & Baumgartner, R. (2022). AMT 1. x: A toolbox for reproducible research in auditory modeling. Acta Acustica, 6, 19.
[5] Smith, R. C., & Price, S. R. (2014). Modelling of human low frequency sound localization acuity demonstrates dominance of spatial variation of interaural time difference and suggests uniform just-noticeable differences in interaural time difference. PloS one, 9(2), e89033.
[6] Hartmann, W. M. (2021). Localization and lateralization of sound. Binaural Hearing: With 93 Illustrations, 9-45.
[7] Pulkki, V., & Karjalainen, M. (2015). Communication acoustics: an introduction to speech, audio and psychoacoustics. John Wiley & Sons.
[8] Majdak, P., Baumgartner, R., & Jenny, C. (2020). Formation of three-dimensional auditory space. The technology of binaural understanding, 115-149.
[9] Dietz, M., Ewert, S. D., & Hohmann, V. (2011). Auditory model based direction estimation of concurrent speakers from binaural signals. Speech Communication, 53(5), 592-605.
[10] Baumgartner, R., Majdak, P., & Laback, B. (2014). Modeling sound-source localization in sagittal planes for human listeners. The Journal of the Acoustical Society of America, 136(2), 791-802.
[11] Barumerli, R., Majdak, P., Geronazzo, M., Meijer, D., Avanzini, F., & Baumgartner, R. (2023). A Bayesian model for human directional localization of broadband static sound sources. Acta Acustica, 7, 12.
[12] Daugintis, R., Barumerli, R., Picinali, L., & Geronazzo, M. (2023, June). Classifying Non-Individual Head-Related Transfer Functions with A Computational Auditory Model: Calibration And Metrics. In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1-5). IEEE.
[13] McLachlan, G., Majdak, P., Reijniers, J., & Peremans, H. (2021). Towards modelling active sound localisation based on Bayesian inference in a static environment. Acta Acustica, 5, 45.
[14] Lokki, T., & Lladó, P. (2023, September). Perception of musical dynamics: orchestra spectra combined with auditory modeling. In International Conference on Auditorium Acoustics. Institute of Acoustics.

References

[15] Pulkki, V., Karjalainen, M., & Huopaniemi, J. (1999). Analyzing virtual sound source attributes using a binaural auditory model. Journal of the Audio Engineering Society, 47(4), 203-217.
[16] McKenzie, T., Armstrong, C., Ward, L., Murphy, D. T., & Kearney, G. (2022). Predicting the colouration between binaural signals. Applied Sciences, 12(5), 2441.
[17] B. Eurich, T. Biberger, S.D. Ewert, M. Dietz (2023). Towards a Computationally Efficient Model for Combined Assessment of Monaural and Binaural Audio Quality. Forum Acusticum (Torino, Italy)