Lecture 3: Auditory Perception

Part 3 of 3

Localisation and Spatialisation

Relevance

Geronazzo, M., Sikström, E., Kleimola, J., Avanzini, F., De Götzen, A. and Serafin, S., 2018, October. The impact of an accurate vertical localization with HRTFs on short explorations of immersive virtual reality scenarios. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 90-97). IEEE.

Bălan, O., Moldoveanu, A., Moldoveanu, F. and Dascălu, M.I., 2014, October. Navigational 3D audio-based game-training towards rich auditory spatial representation of the environment. In *2014 18th International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 682-687). IEEE.

Schreuder, M., Blankertz, B. and Tangermann, M., 2010. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. *PloS one*, *5*(4), p.e9813.

Nambu, I., Ebisawa, M., Kogure, M., Yano, S., Hokari, H. and Wada, Y., 2013. Estimating the intended sound direction of the user: toward an auditory brain-computer interface using out-of-head sound localization. *PloS one*, *8*(2), p.e57174.

Bujacz, M., Skulimowski, P. and Strumillo, P., 2012. Naviton—a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired. *Journal of the Audio Engineering Society*, *60*(9), pp.696-708.

Learning Objectives

To provide an understanding of monaural and binaural processing of spatial sound information.

To provide a brief introduction to the head-related transfer function (HRTF) varies with ear, head, and torso dimensions as well as sound source location.

To provide a comparison between using individually measured HRTFs, generic HRTFS, and indirect derivation of the HRTF.

To provide examples of research and applications using HRTFs, spatialized sound, and pointing towards future directions.

Learning Outcomes

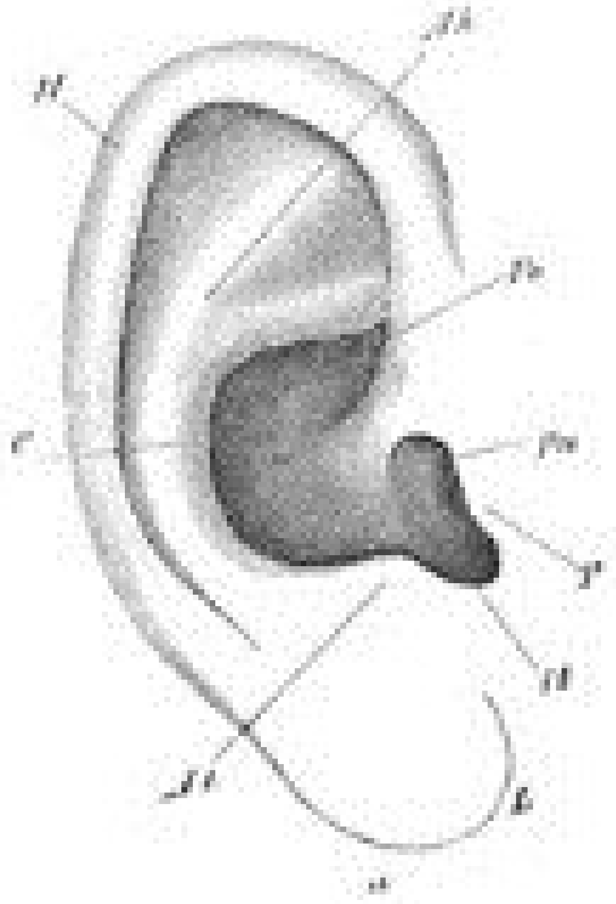
To be able to describe monaural and binaural processing of sound.

To be able to describe how the HRTF can be derived and the causes of the differences in HRTF response across individuals.

To be able to describe the differences between using individually measured HRTFs, generic HRTFs, and indirectly derived HRTFs.

To be able to provide examples of research and applications using HRTFs, and its relevance in audio applications in different research areas.

Monaural Sound Localisation cues

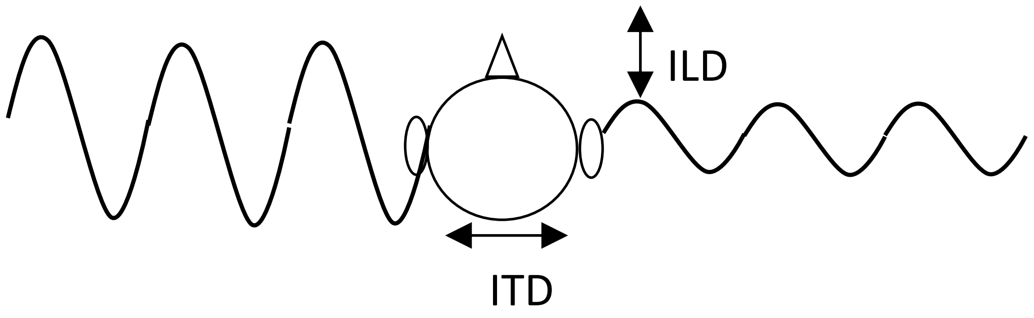


Primary monaural cue for localisation is the spectral cue, since the information for localisation is contained in the differences in distribution of frequencies (or spectrum) that reach the ear from different locations.

These differences arise from reflections from the shape and folds of the ear, as well as head and shoulders.

Ear (pinna) shape is important for determining elevation (sounds from above and below).

Binaural Sound Localisation cues

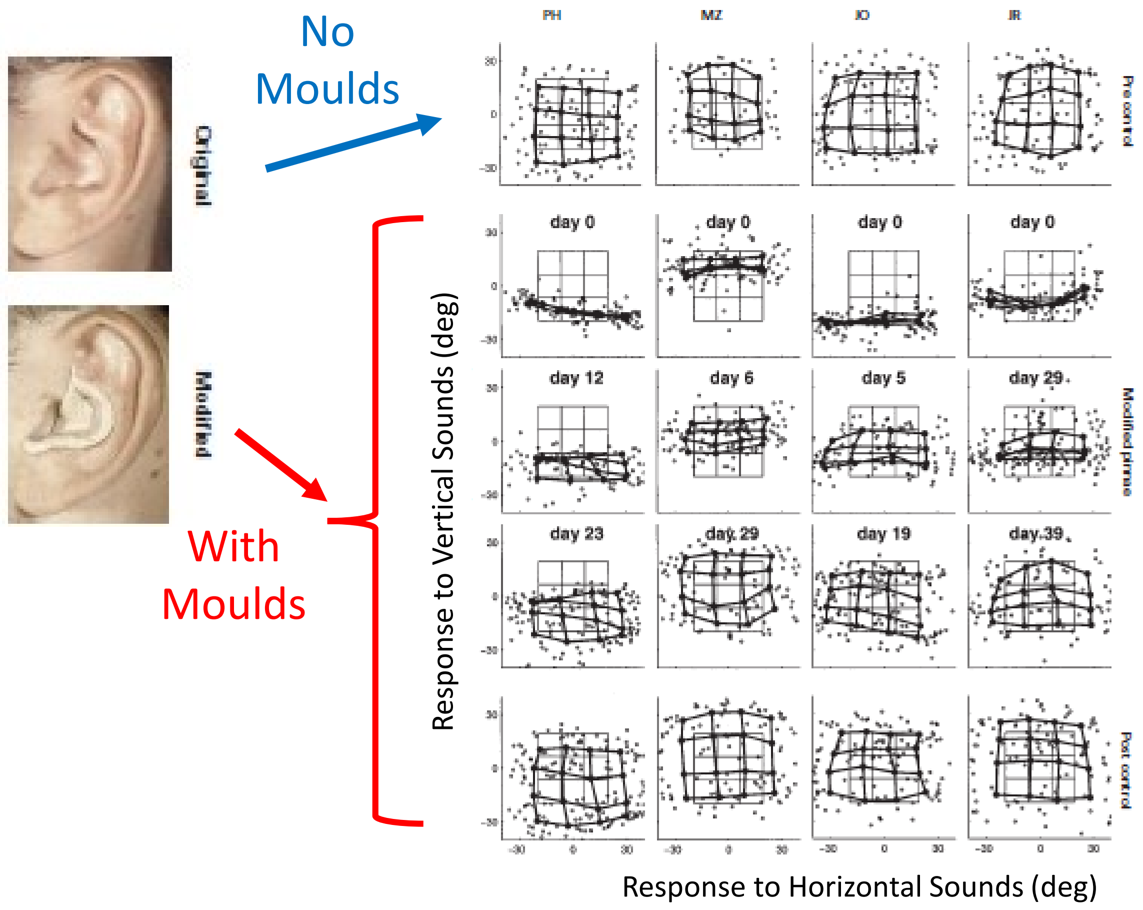


There are two binaural cues for sound localisation -both based on a comparison of the sound signal reaching the left and right ears.

Difference between time of arrival of sound at two ears - Interaural Time Difference (ITD).
Sound wave reaches ear near source before reaching the other ear. Timing differences are needed to locate low frequency sounds.
[Humans $\pm$ 700 µs]
ITD is useful at low frequencies

Difference between loudness of sounds at two ears (head casts a shadow to airborne sound waves)
- Interaural Level Difference (ILD).
Works best for high frequencies (low frequencies wrap around head - no shadow)
In humans > 2kHz
ILD is useful at high frequencies

Adaptation to Altered Localisation Cues



Localisation can be affected by changing outer ear (pinna) shape (changes spectral cues for localising sounds).

Hofman et al. first measured localisation ability without any moulds (top row).
Then moulds were used to change contours of ear.
Participants wore these for several weeks.

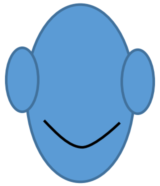
Day 0 with moulds = Localisation poor for elevation (vertical) cues.
Day 5 with moulds = Localisation improving.
Day 19 with moulds = Localisation reasonably good, had learned to use modified cues.
After moulds were removed = Localisation was still good.

2 mappings were created that remained available for localising sounds (like learning 2 languages).

Hofman et al. (1998). Nature Neuroscience.

How do we Provide These Spectral Cues for Localisation if Using Headphones (which cover the pinna)?

Spectral cues are available



Head and ear forms a complex direction-dependent filter (i.e., it shapes the spectrum of the incoming sound before it enters the ear canal). This provides vital cues about the location of a sound.

Spectral cues are lost as the ears are covered by headphones

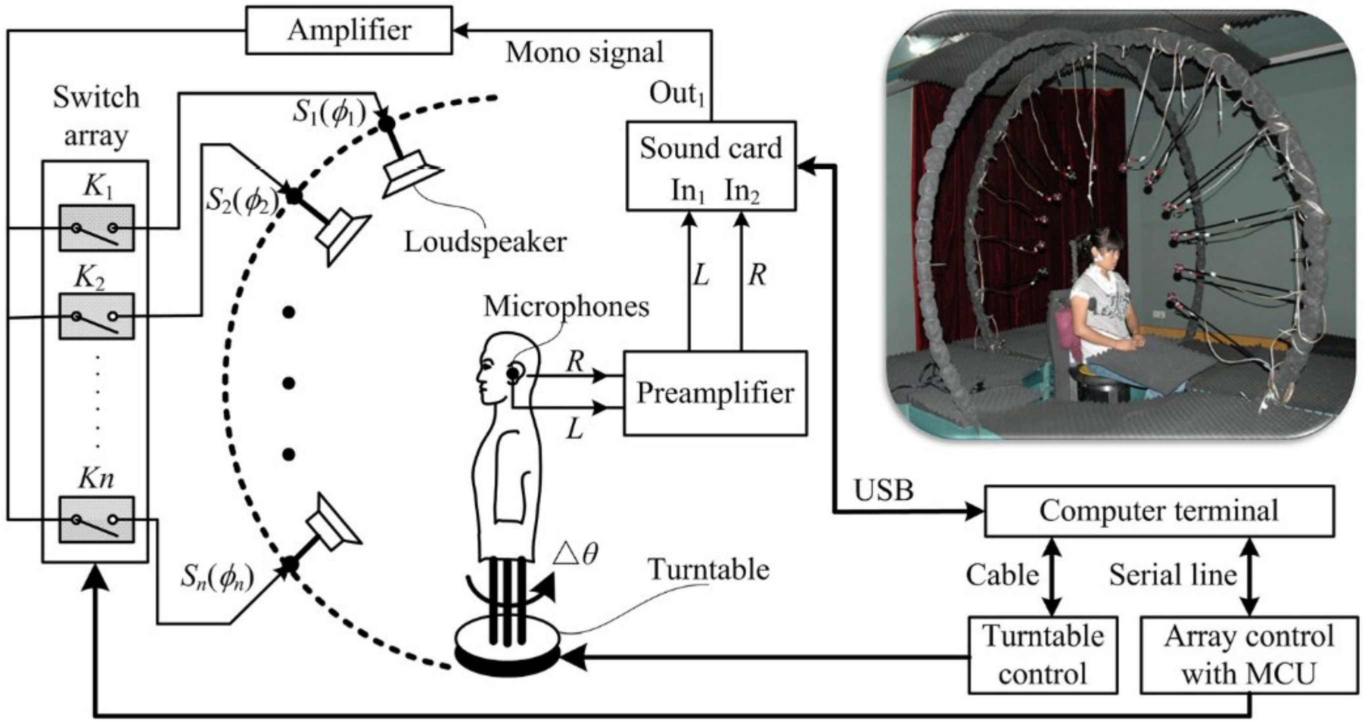
Headphones will obscure this information, so will not achieve a realistic sound experience. I.e., the sounds will not be adequately spatialized.

SO, if we can measure the filtering effect of the ear and head from the individual before they wear the headphones and use that information to shape the sounds presented over headphones they should receive an adequate spatialized sound experience.

Pre-measure spectral cues then add then to the sounds presented over headphones

BUT how do we measure the spectral cues to add in?

Measuring The Head-Related Transfer Function



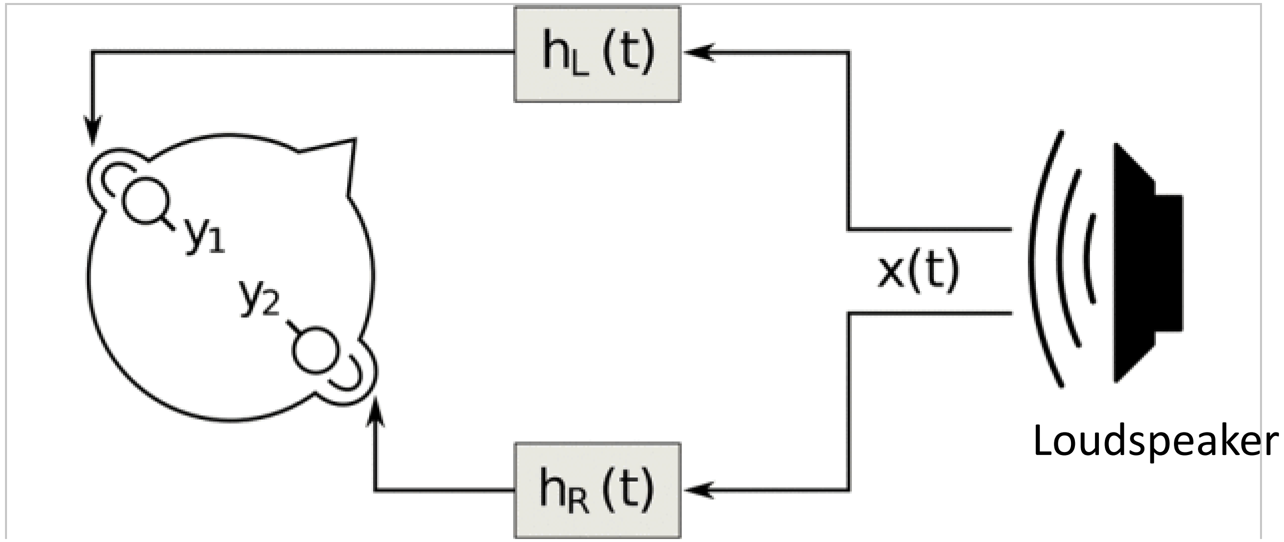
The spectral cues for localisation are created by reflections by the ear, head (and torso). This filtering by the ear and head can be measured.

Loudspeaker array -Presents sounds (anechoic room)

Record from microphones positioned inside the ears.

Schematic and photograph of near-field head-related transfer function (HRTF) measurement system for human subjects. From Yu, G., Wu, R., Liu, Y. and Xie, B., 2018. Near-field head-related transfer-function measurement and database of human subjects. *The Journal of the Acoustical Society of America*, *143*(3), pp.EL194-EL198.

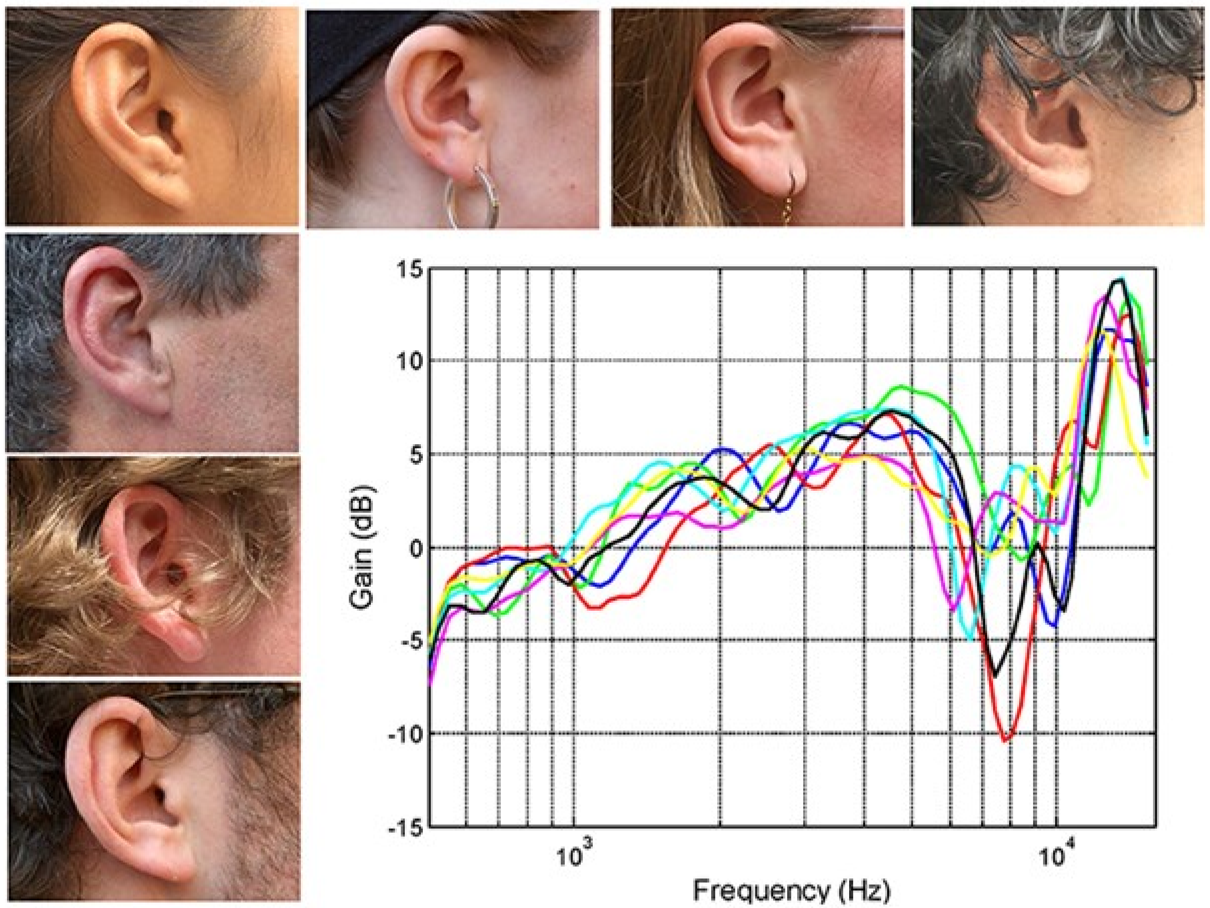
The Head-Related Transfer Function (HRTF)



These acoustical paths can be modelled by linear systems. The transfer functions $h_L(t)$ and $h_R(t)$ which describe these systems are called Head-Related Transfer Functions.

Rothbucher, M., Habigt, T., Habigt, J., Riedmaier, T. and Diepold, K., 2010, December. Measuring anthropometric data for HRTF personalization. In *2010 Sixth International Conference on Signal-Image Technology and Internet Based Systems* (pp. 102-106). IEEE.

HRTF Varies with Ear Shape



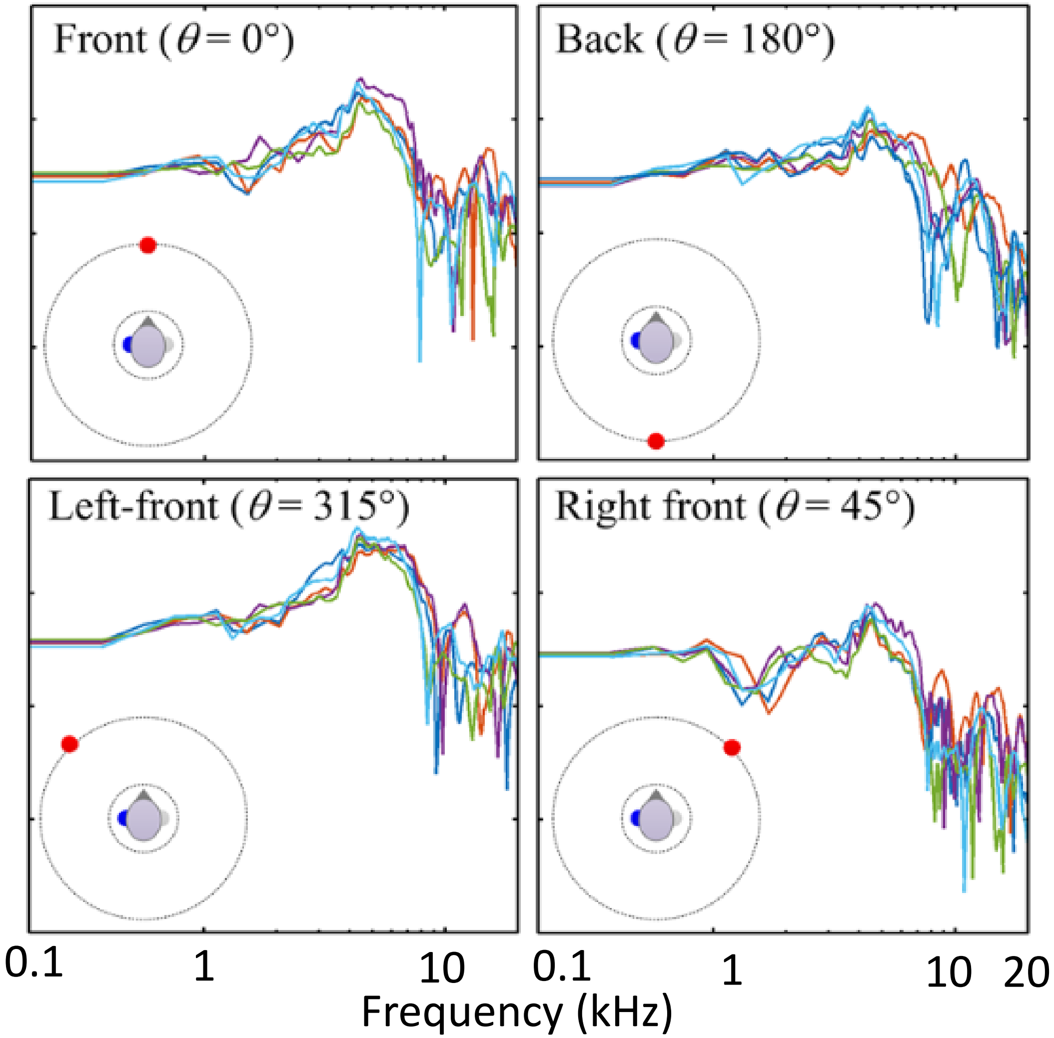
This spectrum shaping by the head and torso as seen in the HRTF is dependent on an individuals' ear shape, head and torso shape.

HRTF- peaks and dips at certain frequencies. Greatest difference across different ears in the effect on high frequencies.

HRTFs are unique to the individual.

The right ears of seven subjects together with their associated head-related transfer functions (HRTFs). Carlile, S., 2014. The plastic ear and perceptual relearning in auditory spatial perception. *Frontiers in neuroscience*, *8*, p.237

HRTF Varies with Sound Source Location



For the same person-The HRTF shows a complex pattern of peaks and dips which varies with direction of a sound source relative to the head and which is unique for each direction in space.

Horizontal HRTF magnitudes for different positions of the sound source. From Yu, G., Wu, R., Liu, Y. and Xie, B., 2018. Near-field head-related transfer-function measurement and database of human subjects. *The Journal of the Acoustical Society of America*, *143*(3), pp.EL194-EL198.

Obtaining and Using HRTFs

HRTFs are unique to the individual.

HRTFs can be used to modify the sounds to be presented over headphones so that they incorporate the special cues for localisation and sounds are adequately spatialised.

BUT special labs and lab equipment are required to measure individual HRTFs, so how can these localisation cues be obtained for everyday use-e.g., listening to sound effects, computer gaming environment, experiencing virtual reality etc.

Option 1- Using Generic HRTFs

Berger, C.C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D. and Zhang, Z., 2018. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Frontiers in neuroscience*, *12*, p.21.

Can use generic HRTFs (HRTFs measured from a population published by researchers - database of values).

So do not need to measure a HRTF- just use someone else's measure.

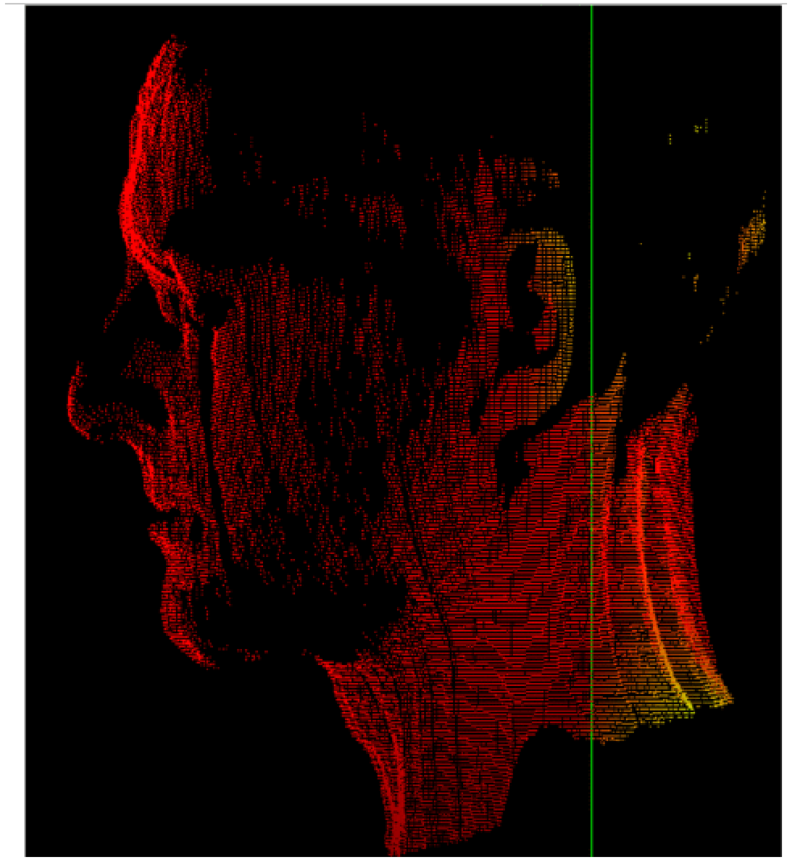
As these are NOT the user's own HRTF, users will take time to adapt to the generic HRTFs.

Various studies looking at how the training period to generic HRTFs can be reduced.

Berger et al. (2018) study. Participants were equipped with a VR headset and reported source of sounds originating from five different locations. Paired the auditory stimulus with a visual stimulus (in the same locations).

Training with this stimulus set enabled users of generic HRTFs to improve subsequent auditory source localization.

Option 2 - Anthropometrics



Rothbucher, M., Habigt, T., Habigt, J., Riedmaier, T. and Diepold, K., 2010, December. Measuring anthropometric data for HRTF personalization. In *2010 Sixth International Conference on Signal-Image Technology and Internet Based Systems* (pp. 102-106). IEEE.

Can estimate HRTFs (spectral shaping) from the geometric features of the ear and head obtained from 3D scans of the ear and head.

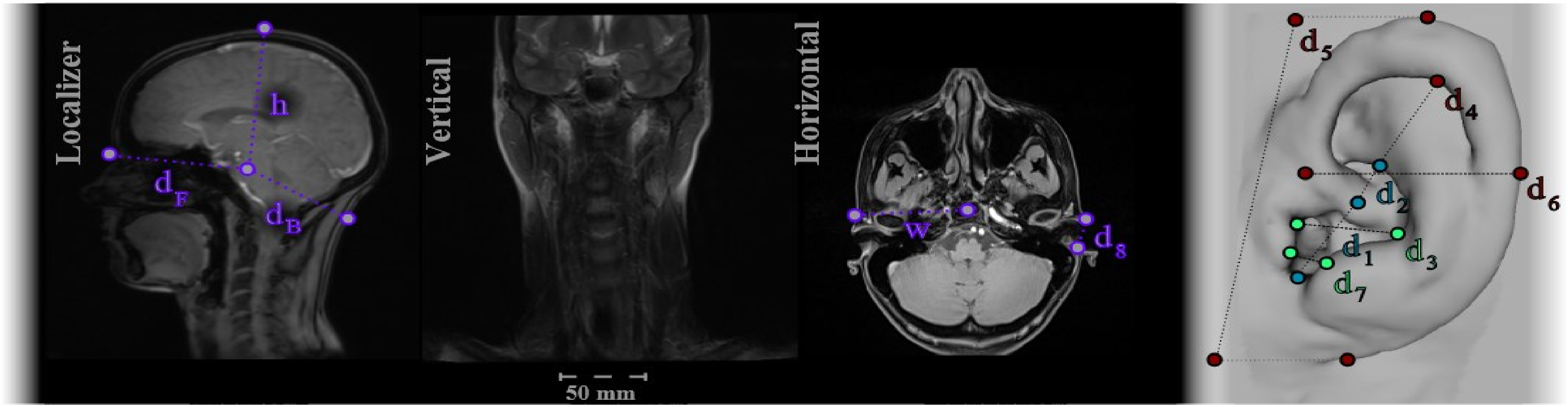
Require a collection of HRTF datasets of various individuals with the corresponding anthropometric data.

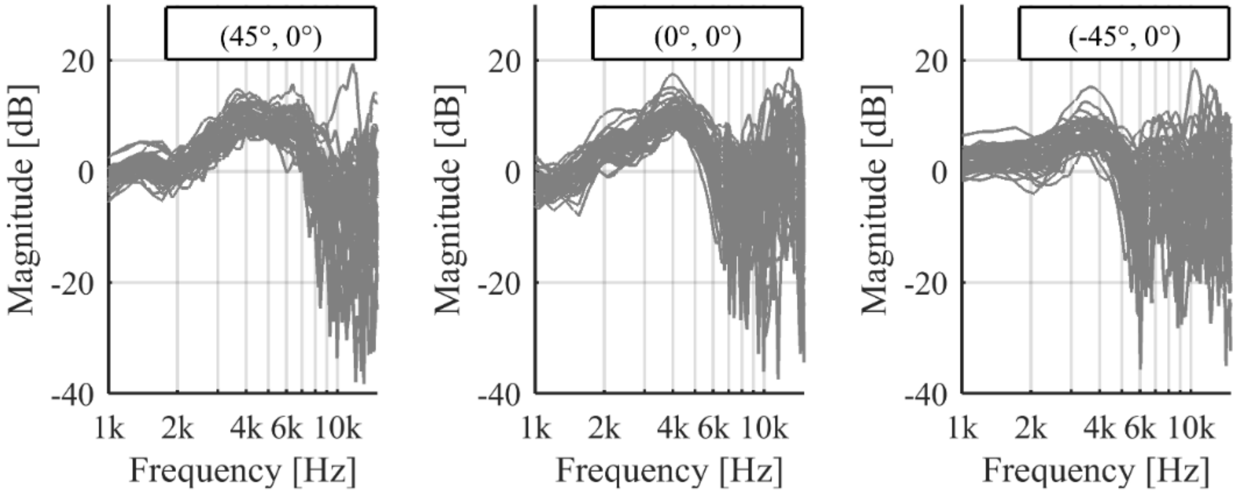
Can then select the best matching HRTF based on the known ear/head shape.

The well-known CIPIC database] provides a set of measured HRTFs with the corresponding geometric features of the listener.

Also,
- not restricted to predefined anthropometric parameters to use, i.e., can use regression models.
- 3D models can be used for ray-tracing approaches to synthesize HRTFs.

Anthropometrics - the CIPIC database





Bomhardt, et al., 2016 presents a database of 48 HRTF datasets, magnetic resonance imaging (MRI) head scans and the corresponding three-dimensional ear mesh models of individual subjects.

This database can be used to select or model an HRTF dataset for binaural synthesis and auditory reproduction.

Bomhardt, R., de la Fuente Klein, M. and Fels, J., 2016, November. A high-resolution head-related transfer function and three-dimensional ear model database. In *Proceedings of Meetings on Acoustics 172ASA* (Vol. 29, No. 1, p. 050002). Acoustical Society of America.

Questions [10 mins]

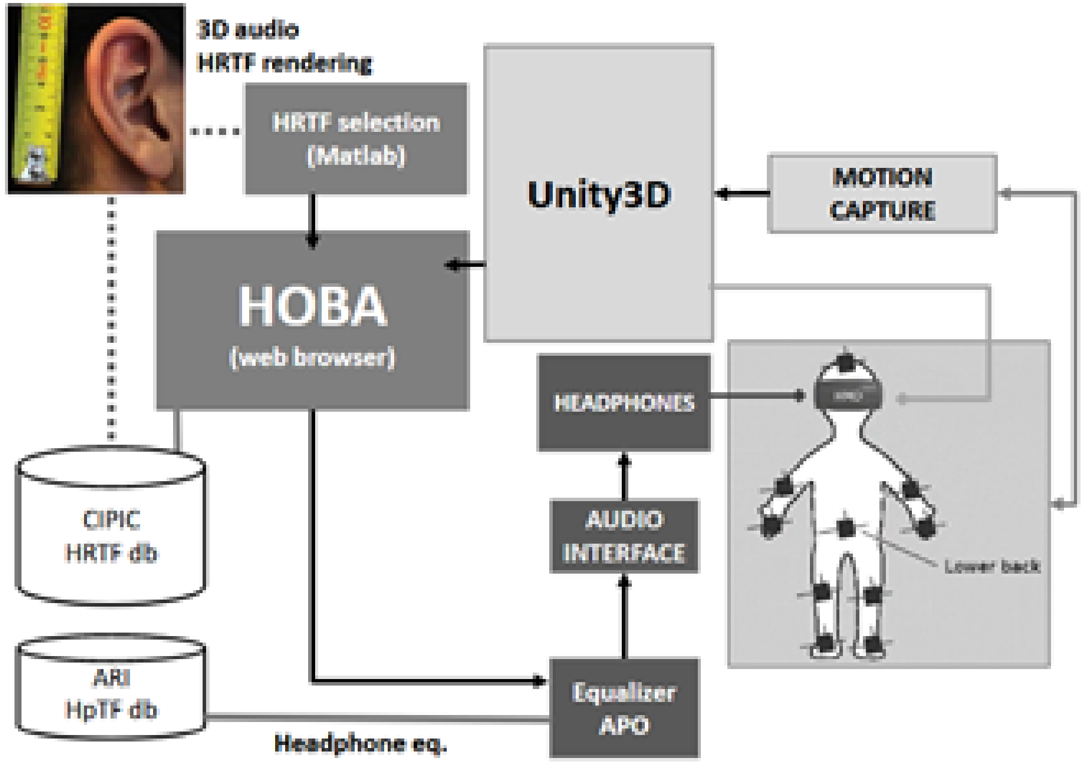
How may vision influence audio localisation, and how may this affect design considerations?

How may spatialized sound be useful in teleconferencing?

What other factors can affect training with generic HRTFs?

Applications

Applications

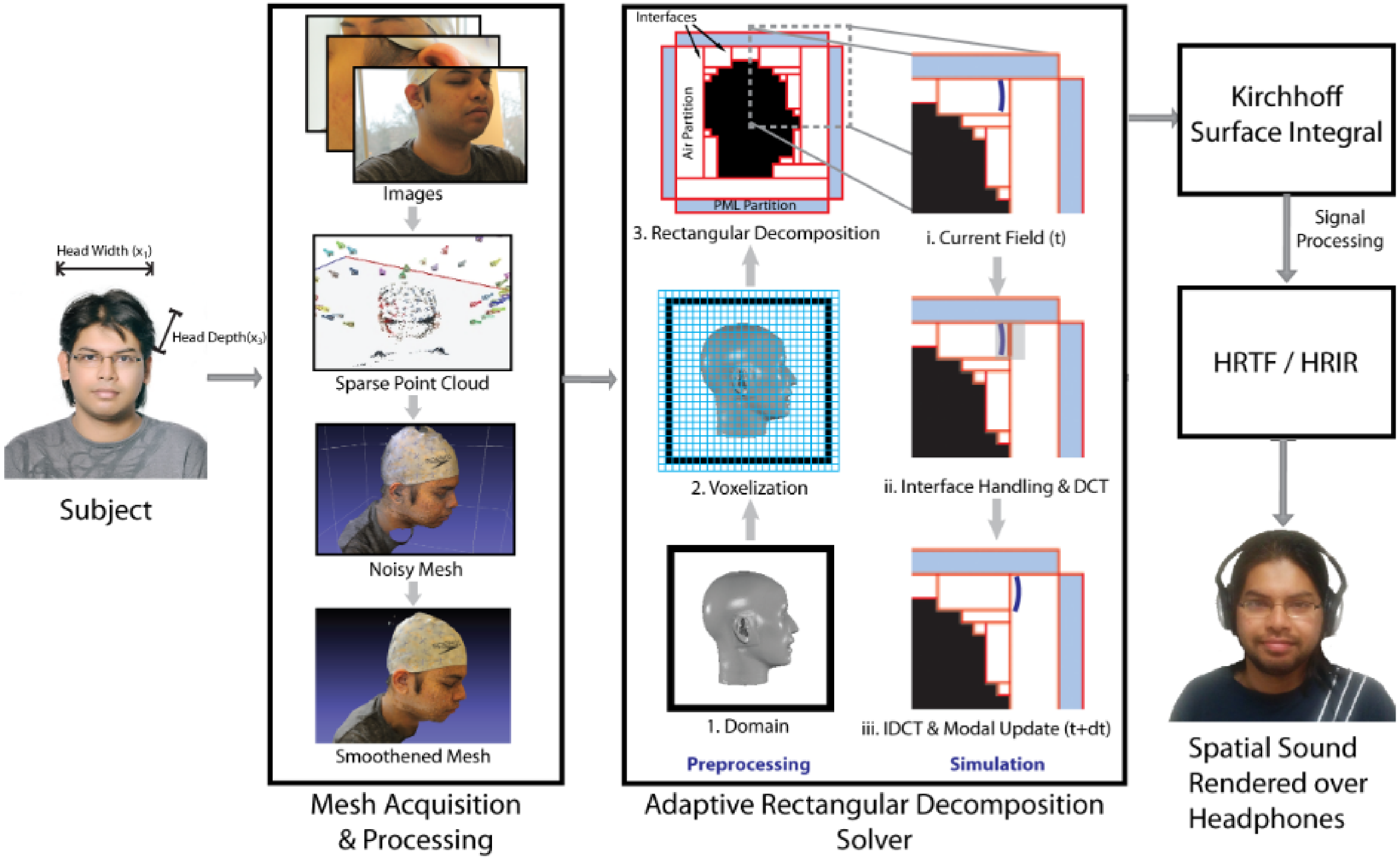


Geronazzo et al. (2018) investigated the connection between listener sensitivity in vertical sound localization cues and experience of presence, spatial audio quality, and attention in a VR environment.

Found that good and bad sound localizers perceived audio latency differently- a consideration in the designing immersive VR experiences.

Geronazzo, M., Sikström, E., Kleimola, J., Avanzini, F., De Götzen, A. and Serafin, S., 2018, October. The impact of an accurate vertical localization with HRTFs on short explorations of immersive virtual reality scenarios. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 90-97). IEEE.

Applications



Meshram et al 2014- An approach to compute personalized HRTFs for any individual using a method that combines image-based 3D modelling with an efficient numerical simulation pipeline.

Meshram, A., Mehra, R., Yang, H., Dunn, E., Franm, J.M. and Manocha, D., 2014, September. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 53-61). IEEE.

Overall Summary

We are able to make use of monaural and binaural localisation cues.

Can also successfully adapt to new localisation cues.

The head-related transfer function (HRTF) and its measurement.

HRTF varies with ear shape (as well as head and torso dimensions) and sound source location.

Instead of directly measuring the HRTF from an individual can uses databases of HRTFs obtained from a population (generic HRTFs) or anthropometric data/other approaches.

Covered some examples of applications and future directions.

Resources

Essential:
Sensation and Perception 8th edition (book), 291-309.

Supplementary:

Geronazzo, M., Sikström, E., Kleimola, J., Avanzini, F., De Götzen, A. and Serafin, S., 2018, October. The impact of an accurate vertical localization with HRTFs on short explorations of immersive virtual reality scenarios. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 90-97). IEEE.

Bălan, O., Moldoveanu, A., Moldoveanu, F. and Dascălu, M.I., 2014, October. Navigational 3D audio-based game-training towards rich auditory spatial representation of the environment. In *2014 18th International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 682-687). IEEE.

Nambu, I., Ebisawa, M., Kogure, M., Yano, S., Hokari, H. and Wada, Y., 2013. Estimating the intended sound direction of the user: toward an auditory brain-computer interface using out-of-head sound localization. *PloS one*, *8*(2), p.e57174.

Berger, C.C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D. and Zhang, Z., 2018. Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity. *Frontiers in neuroscience*, *12*, p.21.

Bomhardt, R., de la Fuente Klein, M. and Fels, J., 2016, November. A high-resolution head-related transfer function and three-dimensional ear model database. In *Proceedings of Meetings on Acoustics 172ASA* (Vol. 29, No. 1, p. 050002). Acoustical Society of America.

Bomhardt, R., de la Fuente Klein, M. and Fels, J., 2016, November. A high-resolution head-related transfer function and three-dimensional ear model database. In *Proceedings of Meetings on Acoustics 172ASA* (Vol. 29, No. 1, p. 050002). Acoustical Society of America.

Meshram, A., Mehra, R., Yang, H., Dunn, E., Franm, J.M. and Manocha, D., 2014, September. P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 53-61). IEEE.

Serafin, S., Geronazzo, M., Erkut, C., Nilsson, N.C. and Nordahl, R., 2018. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE computer graphics and applications*, *38*(2), pp.31-43.