Non-rigid pinna registration for the calculation of head-related transfer functions

Exposé: Katharina Pollack

Every day, millions of people worldwide utilize headphones – for listening to music, watching movies, or making phone calls – and usually the sound created by the headphones is perceived as sound sources located within the head. This is a rather unnatural phenomenon that is, because of the lack of alternatives, widely tolerated despite the loss of a realistic perception. Humans can localise natural sound sources, i.e., assign direction and distance to a perceived auditory event (Blauert, 1997). For that, broadband interaural[1] time and level differences play a great role in the horizontal plane (Macpherson & Middlebrooks, 2002). For the distance, vertical direction attribution, as well as the differentiation between front and back, spectral properties have been found to be a significant contributor. It is currently understood that these properties are responsible for the externalisation effect of sound sources, i.e., the perception of sound sources at their natural source outside of the head (Baumgartner et al., 2017).

Under laboratory conditions, it is possible to reproduce realistic sound sources via headphones. In such a setup, personalised binaural[2] signals are generated enabling an instinctive perception of the acoustic landscape around the person. The field of <u>b</u>inaural <u>v</u>irtual <u>ac</u>oustics (BVA) relates to the development and investigation of new methods for realistic headphone-based applications, employing these personalised signals. Binaural signals emerge through a filtering caused by torso, head, and pinna[3] of a person, with the latter being distinctive for individuals. Thus, the shape and structure of the pinna substantially affects the quality of a BVA system. The personalised filtering can be described with so-called <u>he</u>ad-<u>r</u>elated <u>t</u>ransfer <u>f</u>unctions (HRTFs)(Møller, Sørensen, Hammershøi, & Jensen, 1995), that characterise the spatial filtering of a sound considering the position of the sound source relative to the listener.

Although the measurement process of HRTFs endures only a quarter of an hour (Majdak, Balazs, & Laback, 2007), it requires professional equipment and laboratory conditions, e.g., the room reverberation must approximate an anechoic chamber sufficiently, the loudspeakers, microphones and

1 *interaural:* lat. "inter" = between, "auris" = ear

2 *binaural signals* describe acoustic cues that arrive at both eardrums, containing information about the source position.

3 *pinna*: outer ear

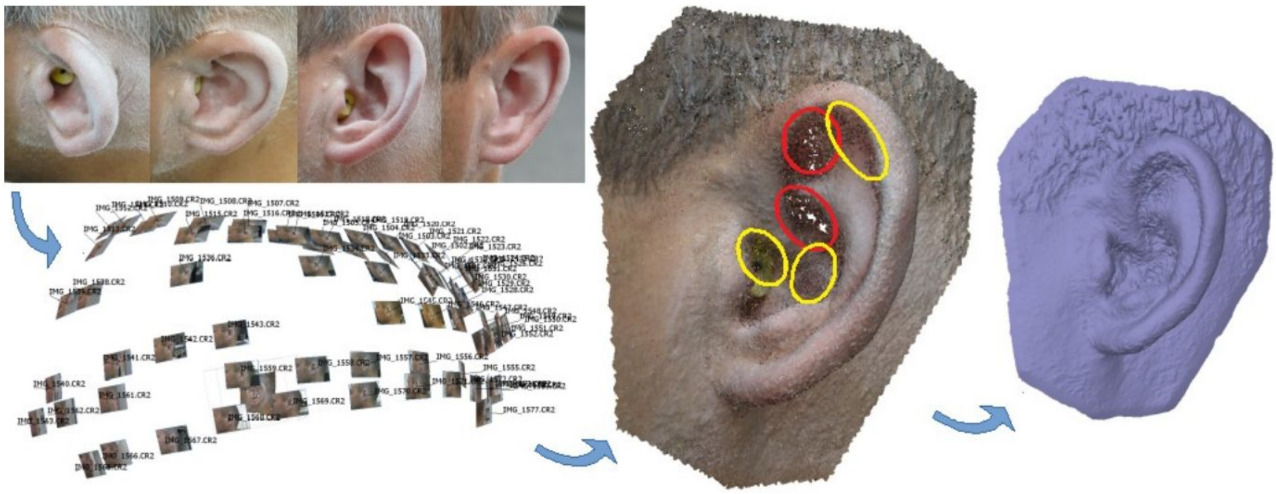

Figure 1: Photogrammetrical acquisition of a pinna geometry. F.l.t.r.: pinna photos, reconstruction of camera positions, 3D pinna model (point cloud), and 3D pinna geometry (mesh). Problematic regions of the pinna are framed in red (holes) and yellow (outliers).

amplifiers must offer a very high signal-to-noise ratio, etc. Various setups for measuring HRTFs have been proposed. For example, in the setup implemented at the ARI, two small microphones are placed at the ear-canal opening of the subject, who sits on a chair that rotates within the horizontal plane and stops every 2.5°. An array of 22 loudspeakers with a resolution of 5° between -30° and +80° in the median plane is placed on a circle around the subject. For every direction in the horizontal plane, multiple exponential sweeps, carefully nested within one another, are played from eleven loudspeakers on average and build up to over 1500 measurements. However, these measurements can only classify one distance of the sound source. When multiple distances are desired to measure, it is connected to an enormous effort (Yu, Wu, Liu, & Xie, 2018) and is not practicable for a commercial use.

Because of the rather elaborate acoustical HRTF measurement process, a numerical calculation of HRTFs provides many advantages: independence of acoustic laboratory equipment (photos with a quality a state-of-the-art smartphone), and convenience in data acquisition (the comfort of not having to sit still for the whole measurement process). Thus, an easy method for numerical calculation of HRTFs would enable access to the HRTF calculation to the public. One state-of-the-art approach origins in the field of geodesy: Photogrammetry describes the retrieval of physical properties from photographic 2D images (see Fig. 1), i.e., processing 2D information to create one large 2D image (maps from satellite data) or to extract features indicated by 3D contours, resulting in a 3D object model, the so-called 3D mesh. With such a 3D mesh calculated from 2D photos of the pinna, the simulation of the HRTF measurement can yield personalised HRTFs of a listener (Ziegelwanger, Kreuzer, & Majdak, 2015).

However, certain regions of the pinna are curled and thus cannot be easily covered by photos with conventional equipment, which results in local perforations in the mesh. Small deviations in these regions, especially within the concha and the cymba (Mokhtari, Takemoto, Nishimura, & Kato, 2016), can result in significant changes of the HRTFs. Additionally, inaccurately captured structures cause distortions in the rather smooth mesh surface. Both issues require manual post-processing of the pinna model in order to meet the precision requirements for HRTF calculations. Therefore, photogrammetry is not yet applicable in automatic HRTF calculations.

In my doctoral thesis, I will systematically investigate the applicability of non-rigid registration (NRR) algorithms for an automatic reconstruction of pinna models retrieved via photogrammetry.

The thesis bridges **computer science** with **audio engineering** because the process of HRTF calculations needs to be evaluated both from the geometric perspective, i.e., quantifiable geometric differences between pinna models, and the auditory perspective, i.e., by considering the properties of human spatial hearing in terms of auditory modelling and psychoacoustic experiments. The findings will be essential for BVA systems rendering realistic sounds in personalised virtual reality and augmented reality (VR/AR, respectively) systems in the future. Still, my investigations will be general and systematic such that any findings will extend the general knowledge of NRR algorithms and pinna models, regardless their future implications for an automatic construction of high-quality pinna geometries.

Stages of the PhD

The thesis consists of three stages, each of them will be closed by submission of a manuscript (journal or conference) for publication.

Stage 1: Examination and adaptation

The first stage will examine the state-of-the-art on NRR algorithms followed by their implementation in order to be able to investigate their performance on pinna-model alignment. To exclude problems resulting from photogrammetry, these NRR algorithms will be used to align template pinnae with perfect targets (provided from CT scan). In order to avoid a trivial alignment, the target pinna will be different from the template pinna and the targets will be changed systematically, e.g., Euclidean transformations (translation, rotation and isotropic scaling) of the target and the geometric distance between template and target. The difference between the NRR-aligned and the reference pinna will be analysed: the smaller the difference, the better the selected NRR algorithm is able to handle details of the pinna geometry. For further adaptation to my problem, the chosen algorithm will be modified to include the information encoded in the pinna texture: less visible areas of the pinna are darker than the exposed ones, thus, the darker regions of the model will be weighted lower than others in the alignment process. Then, HRTFs will be calculated and the perceptive variation will be modelled using auditory models. The evaluation will consider acoustic HRTF measurements, HRTF calculations, calculated and perceived errors between these two (focusing on polar errors and quadrant error rate (Ziegelwanger, Majdak, & Kreuzer, 2015)) and analysis of auditory features obtained with the help of auditory models (Marelli, Baumgartner, & Majdak, 2015). By juxtaposing the various NRR algorithms, I will aim at finding a suitable and adapted candidate algorithm for the following stages.

Stage 2: Error types and robustness

In this stage, photos and photogrammetrically obtained pinna models will be constructed, and with statistical approaches, the types of errors (outliers, holes, deviations from actual geometry) will be analysed and classified. These errors will then be manually imprinted on reference models serving as disrupted targets for the alignment. The manual addition of the errors will enable a systematic evaluation of the effect of the error type on the HRTF quality and estimate the limits of the NRR algorithm's robustness. I will test the algorithm's robustness in two configurations: the alignment of a template to a disrupted mesh of the same template, as well as the alignment of a template to a disrupted mesh of a different template. The handling of the error types by the algorithm will be evaluated as in stage #1 by means of HRTF calculations and auditory modelling.

Stage 3: Psychoacoustic quality evaluation

In the last stage, the achievable quality of the alignment process will be investigated. Different perfect templates will be aligned with the photogrammetrically obtained pinna targets in order to analyse the potential advantage of the choice of the template pinna. The results will be evaluated geometrically and compared to features obtained from auditory models in order to describe perceptually based error. A selected configuration will be verified in a psychoacoustic experiment testing 3D sound localization in normal-hearing listeners.

Personal prerequisites and professional development

I have recently finished my studies in electrical engineering and audio engineering at both Technical University and University of Music and Dramatic Arts Graz and strive for a scientific career. During the process of my education I have spread my interests widely, with my theses and internships at universities, institutes, and companies, but one thing attracted my attention sustainably: binaural audio and its understanding from the scientific point of view. This was triggered by my internship at the Acoustics Research Institute of the Austrian Academy of Sciences in Vienna where I was working on pinna modelling. I would like to contribute to this field substantially and become an expert on spatial hearing, starting with the proposed doctoral thesis.

References

Baumgartner, R., Reed, D. K., Tóth, B., Best, V., Majdak, P., Colburn, H. S., & Shinn-Cunningham, B. (2017). Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias. *Proceedings of the National Academy of Sciences*, *114*, 9743–9748.

Blauert, J. (1997). *Spatial hearing. The Psychophysics of Human Sound Localization* (Revised edition). Cambridge, MA: The MIT Press.

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, *111*, 2219–2236.

Majdak, P., Balazs, P., & Laback, B. (2007). Multiple exponential sweep method for fast measurement of head-related transfer functions. *J Audio Eng Soc*, *55*, 623–637.

Marelli, D., Baumgartner, R., & Majdak, P. (2015). Efficient Approximation of Head-Related Transfer Functions in Subbands for Accurate Sound Localization. *IEEE Transactions on Audio, Speech, and Language Processing*, *23*, 1130–1143.

Mokhtari, P., Takemoto, H., Nishimura, R., & Kato, H. (2016). Vertical normal modes of human ears: Individual variation and frequency estimation from pinna anthropometry. *The Journal of the Acoustical Society of America*, *140*, 814–831.

Møller, H., Sørensen, M. F., Hammershøi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *J Audio Eng Soc*, *43*, 300–321.

Yu, G., Wu, R., Liu, Y., & Xie, B. (2018). Near-field head-related transfer-function measurement and database of human subjects. *The Journal of the Acoustical Society of America*, *143*, EL194–EL198.

Ziegelwanger, H., Kreuzer, W., & Majdak, P. (2015). Mesh2HRTF: Open-source software package for the numerical calculation of head-related transfer functions. *Proceedings of the 22nd International Congress on Sound and Vibration*, 1–8. Florence, IT.

Ziegelwanger, H., Majdak, P., & Kreuzer, W. (2015). Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization. *The Journal of the Acoustical Society of America*, *138*, 208–222.