Project Lab-course in Audio Information Processing (SS2022)

Instructor:	Prof. Dr.-Ing.Bernhard U. Seeber	Room N 6504
Assistant:	Lubos Hladek, PhD. (organization) AIP team (individual projects)	Room N 6503 Tel:089/289-282564 Ema il: lubos.hladek@tum.de
Time &Place:	Kick-off meeting: 25.04.2022, 14:00 Presentations: to be agreed	N0116
ECTS:	6	

Exam

Every student is responsible for proper registration at the Prüfungsamt/TUMonline so that we can upload the final grade. The overall grade is composed of graded labwork (40%), your individual report of the lab results (40%) and the (group) presentation of the results (20%). Group presentations are 20 min, depending on group size, followed by a discussion. In the presentation, focus on the achieved results of the project.

Suggested Topics (one to choose per group of 2-5 students)

1. Extending static coloration models to moving sources

(Matthieu Kuntz)

The coloration of a sound source can be understood as a timbral deviation from a reference sound. The study of coloration can be approached from different perspectives: The coloration of a concert hall, where reflections on the walls change the spectrum of the sound source over time as they overlap with the direct sound, the ability to detect coloration in different scenarios, or as a quality metric for sound reproduction systems and methods, ranging from small living room Hi-Fi setups to loudspeaker arrays.

In this project, we will consider the third aspect. At the AIP, we are using a loudspeaker array in an anechoic chamber to run a variety of psychoacoustic experiments. The sound reproduction methods used to auralize moving sources all color the sound source and introduce artefacts, which were assessed in a listening experiment. The modeling of coloration perception has only been done for static sources so far. The goal of this project is to implement models (starting with the proposed ones) to predict the coloration of a static sound source and to apply them for moving sound sources in a time-frame based approach to investigate their relevance in that scenario.

The project can be summarized in the following goals and work distributions:

1. Familiarize yourself with the literature and the existing models (Pulkki, 2001; Fleßner *et al.*, 2019; McKenzie *et al.*, 2022)
2. Use the implemented models to predict collected data of coloration of static sources in the free-field. How do they perform? Is it safe to assume that they work in the free-field, although they were only verified over headphone experiments?
3. Modify the implemented models to work in a time-frame based manner and try to predict collected data of coloration of moving sound sources.

Going further: Both models are weighting each frequency-bin by the bandwidth of the ERB at that center frequency. Extend the models to base their loudness estimation on an actual loudness model.

Introductory literature:

Fleßner, J.-H., Biberger, T., and Ewert, S. D. (**2019**). "Subjective and Objective Assessment of Monaural and Binaural Aspects of Audio Quality," IEEE/ACM Transactions on Audio, Speech, and Language Processing **27**, 1112-1125.

McKenzie, T., Armstrong, C., Ward, L., Murphy, D. T., and Kearney, G. (**2022**). "Predicting the Colouration between Binaural Signals," Applied Sciences **12**.

Pulkki, V. (**2001**). "Coloration of Amplitude-Panned Virtual Sources," in *110th AES Convention* (Amsterdam).

2. Tone in noise detection with head movements compared to detection of moving stimuli

(Norbert Kolotzek)

Detecting a sound source in noise is an important feature for most everyday life situations. It is well known from the literature that differences in interaural correlation of the noise masker and the target signal are beneficial for detecting the signal of interest compared to a situation where both, masker and signal, have the same interaural correlation. Such differences can be caused in the free field by spatial separation of the noise and the target, introducing different interaural time and level differences. Moving target sounds also achieve better detection thresholds than static collocated sources, which is caused by dynamic interaural cue changes during the motion.

Head movements can also introduce such dynamic changes and are beneficial for sound localization as well as for speech intelligibility. The aim of the project is to investigate whether source motion or head rotations result in different detection thresholds when the dynamic interaural cue changes stay the same. This should be investigated in two pilot experiments, one with moving sources, one with instructed head movements.

The project can be summarized with the following goals and work distributions:

Implementing two detection experiments

1. Moving sound sources and static listener
2. Listener with instructed head movements, resulting in comparable interaural cue changes ensured by head tracking data
3. Analyzing and comparing both experimental results
4. If possible, recording ear signals with in-ear-microphones for both experimental conditions for later analysis.

Introductory literature:

Durlach, N.I. (1963). “Equalization and Cancellation Theory of Binaural Masking-Level Differences,” J. Acoust. Soc. Am., 35, 1206-1218.

Grantham, D. W., and Wightman, F. L. (1979). "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," Journal of the Acoustical Society of America 65, 1509-1517.

Grange, J.A., and Culling, J.F. (2016). “The benefit of head orientation to speech intelligibility in noise,” J. Acoust. Soc. Am., 139, 703-712.

Perrett, S., and Noble, W. (1997). “The contribution of head motion cues to localization of low-pass noise,” Perception & Psychophysics, 59, 1018-1026.

3. Interactive characters for virtual reality for hearing research

(Lubos Hladek)

People with hearing impairment struggle with communication in daily tasks. Despite attempts in the past, the research has been struggling pinpointing the situations in daily life which are the most problematic and in which the hearing technologies fail to improve the listening and communicative situation.

The aim of this project is to study the problematic communication situations during conversations in a realistic scenarios with realistic acoustic and visual cues in which something unexpected happen, for instance, when somebody is addressing a group of people who are engaged in conversation. In this project, such scenario should be created in virtual reality using full-body motion tracking mapped on virtual characters.

The milestones of the project are as follows:

1. Create a virtual character using fullbody motion tracking. Map motion to the character and display in Unreal Engine.
2. Record several types of motion and use it to create controllable behavior. (i.e., character stands, character walks towards the player, character addresses the group, character says something to the group, …). The behavior can controlled by a remotely.
3. Using the protocol and experimental code from Lubos Hladek, record a standing participant inside of the Simulated Open Field Environment reacting on the interactive character.
4. Write a report of the achieved goals and answer the question whether and how the approach can be used to study communication of people with hearing impairments.

4. Clustering Tendency of Auto-Encoded 3D pinna scans vs CIPIC features over subjects

(Payman Azaripasand)

The auditory system ability to localize sound in surrounding space is highly depend on the anthropometric measures of human body. Specifically, the pinna shape in the outer ear has the most effect on the localization at higher frequency sounds by directionally filtering the incoming sound with Head Related Transfer Functions (HRTFs) [1]. Since the pinna shape is very individualized to a person, the HRTFs are also very subject specific. Algazi et al [2] proposed 27 different anthropometric measures from head, torso and pinna, which is still in use in studies.

Recently, researchers capture head and pinna 3D scans mainly by MRI [3] and 3D scanners [4] to capture a 3D representation of human head that enables extracting different anthropometric measures including CIPIC measures. In this project, the students need to encode 3D pinna shapes into the latent space features using Convolutional Auto-Encoders such that the encoded data have less clustering tendency over subjects than CIPIC related features.

The project can be summarized in the following goals and work distributions:

1. Familiarize yourself with the literature and the available data.
2. Design and run a Convolutional Auto-Encoder to find an encoded representation of 3D ear shape.
3. Compare the encoded representation with the CIPIC features of the same dataset in terms of clustering tendency over subjects.

Introductory literature:

[1] Xie, Bosun. Head-related transfer function and virtual auditory display. J Ross (2013).

[2] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The CIPIC HRTF database," Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575), 2001, pp. 99-102, doi: 10.1109/ASPAA.2001.969552.

[3] Ramona Bomhardt, Matias de la Fuente Klein, and Janina Fels , "A high-resolution head-related transfer function and three-dimensional ear model database", Proc. Mtgs. Acoust. 29, 050002 (2016) https://doi.org/10.1121/2.0000467.

[4] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A Cross-Evaluated Database of Measured and Simulated HRTFs Including 3D Head Meshes, Anthropometric Features, and Headphone Impulse Responses," J. Audio Eng. Soc., vol. 67, no. 9, pp. 705-718, (2019 September.). doi: https://doi.org/10.17743/jaes.2019.0024.