SONICOM | EXPERIENCE | CAROUSEL | TOUCHLESS | GUESTXR
HORIZON2020 FET-PROACT COLLABORATION

Artificial Intelligence for Extended Social Interactions
SECTG early career researchers (ECR) meeting
23rd of November 2023 – Imperial College London

SECTG Meeting Agenda

9:00 am - 9:15 am: Introduction and Welcome Remarks from the host (L. Picinali)

9:15 am - 10:15 am: The SONICOM Project
- The Binaural Rendering Toolbox BRT
- Scattering Delay Networks
- On the implementation of model-based analysis to predict listener behaviour with Bayesian inference

10:15 am - 11:15 am: The EXPERIENCE Project
- Explainable artificial intelligence for mental state classifications
- Enhancing VR Content Creation: The Power of CV and DL in Virtualization
- Exploring neural codes and intersubjective alignment of time experience

11:15 am - 11:30 am: Coffee Break

11:30 am - 12:30 pm: The CAROUSEL Project
- Interactive Characters for Virtual Reality
- Simulating Fatigue and Recovery in Virtual Characters
- Dance and Social Media Workshop

12:30 pm - 2:00 pm: Lunch Break

2:00 pm - 3:00 pm: The TOUCHLESS Project
- How Haptic Experiences are Made
- Emotional textile design through multisensory interaction design
- A closed-loop, real-time, predictive model of human response to ultrasonic tactile stimulation using data-driven approaches

3:00 pm - 4:00 pm: The GUEST XR Project
- Empathetic haptics: feel the stress or a virtual agent
- Influence of prior acquaintance on the shared VR experience
- Deep learning for acoustic matching in the VR context

4:00 pm – 4:15 pm: Coffee Break

4:15 pm - 5:00 pm: TBA

5:00 pm - 6:00 pm: Ice-breaking/social event

1.   SONICOM

The Binaural Rendering Toolbox-BRT

María Cuevas-Rodríguez and Daniel González-Toledo
Departamento de Tecnología Electrónica, University of Malaga, Spain

The Binaural Rendering Toolbox-BRT (https://github.com/GrupoDiana/BRTLibrary) is a set of software libraries, applications, and definitions aimed as a virtual laboratory for psychoacoustic experimentation. The BRT is being developed in the framework of the SONICOM project (https://www.sonicom.eu/) and includes the algorithms developed in the 3D Tune-In Toolkit (https://github.com/3DTune-In/3dti_AudioToolkit) in a new open, extensible architecture. At the core of the BRT Toolbox, a library provides C++ implementations of listener models, source models, and environment models, including a standalone application controlled via the Open Sound Control (OSC) protocol. In addition, there are plans to deploy a collection of portings to different audio frameworks such as PureData, MaxMSP and VST plugins, by means of the Avendish library.
In this presentation, the architecture of the BRT, its main features, and its application to reproducible psychoacoustics experiments are described. This includes examples of how the toolbox provides a complete trace of the experiment, including the delivered binaural audio, annotated with the listener and source movements, by means of a new SOFA convention to store dynamic measurements. It will give an update to the current status of the development and will show practical examples of use.

Scattering Delay Networks

Giorgio Presti and Marco Fontana
Department of Computer Science, University of Milano, Italy

Our current research is focused on the artificial reverberation of virtual sources to improve co-immersion in mixed-reality scenarios, minimising the computational cost and system complexity of the final solution.
To achieve this goal different sub-tasks have been addressed. First, we developed a VST audio plugin implementing a real-time artificial reverberator using the Scattering Delay Networks (SDN) approach. It simulates the reverberation in a shoe-box room of variable size, with variable listener and emitter positions. The option to use custom absorption filters for each wall in the room has been added as well as multiple audio output modalities (such as Ambisonics B-format, and binaural output; the latter implemented using the Binaural Rendering Toolbox).
Secondly, we set up a reverberation-matching algorithm to automatically parametrise the SDN VST to match any given Room Impulse Response (RIR). The algorithm is based on a Bayesian optimization using Gaussian Processes, aimed at minimising the log-spectral distance between a given RIR and the SDN-generated RIR.
Finally, to reduce the number of parameters that the optimiser needs to match when dealing with the absorption coefficients for each of the 6 walls, we developed a data-driven parametrization of the coefficients based on the relaxation of a 2D manifold intersecting the 6D data points (representing the 6 frequency bands absorption coefficients) of a dataset of actual materials.

In the future the proposed SDN model will be integrated in the BRT as one of the environment models natively available in the toolkit, thus providing an option for simulating reverberation in BRT.

On the implementation of model-based analysis to predict listener behaviour with Bayesian inference

Roberto Barumerli
Acoustics Research Institute, Austrian Academy of Sciences, Austria

Auditory models provide a common methodology to investigate how listeners understand space through sound. Because of their quantitative nature, these models are a powerful tool to study behavioural measurements recorded during psychoacoustic experiments or to predict listener behaviour in applications that consider virtual or actual sounds. However, similar statistical methods and problem-specific implementations limit reproducibility and cross-study comparisons. In our talk, we describe a flexible methodology grounded in probability theory to perform model-based analysis of experimental data. Without loss of generality, we focus on a model based on Bayesian inference to predict listener performances in a static sound localization task. First, we consider the model's implementation by identifying the fundamental processing steps, such as the extraction of perceptually relevant features, the integration of prior beliefs, and the decision-making strategy. Second, we show how to estimate the model likelihood function, opening the possibility to evaluate the model implementation and to perform statistical analyses on experimental data. Importantly, our example shows how the framework for auditory models based on Bayesian inference (FrAMBI) and the Auditory Modelling Toolbox (AMT) can be valuable tools to implement such analysis. By providing an integrated methodology for model-based analysis, we aim to promote reproducible research in the hearing science field.

2.   Experience

Explainable artificial intelligence for mental state classifications

Antonio Luca Alfeo
Department of Information Engineering, University of Pisa, Italy

Explainable artificial intelligence (XAI) is a technology that can enhance trust in mental state classifications by providing explanations for the reasoning behind artificial intelligence (AI) model outputs, especially for high-dimensional and highly-correlated brain signals. Feature importance and counterfactual explanations are two common approaches to generating these explanations, but both have drawbacks. While feature importance methods, such as Shapley additive explanations (SHAP), can be computationally expensive and sensitive to feature correlation, counterfactual explanations only explain a single outcome instead of the entire model. To overcome these limitations, we consider the frequency with which a change in each feature in isolation leads to a change in classification outcome for the AI model, to build a novel robust feature importance measure.
Experimental results on synthetic data and real publicly available fMRI data from the Human Connect project show that the proposed BoCSoR measure is more robust to feature correlation and less computationally expensive than state-of-the-art methods. Additionally, it is equally effective in providing an explanation for the behavior of any AI model for brain

signals. These properties are crucial for medical decision support systems, where many different features are often extracted from the same physiological measures and a gold standard is absent. Consequently, computing feature importance may become computationally expensive, and there may be a high probability of mutual correlation among features, leading to unreliable results from state-of-the-art XAI methods.

Enhancing VR Content Creation: The Power of CV and DL in Virtualization

Clara Garcia Moll
Laboratory of Immersive Neurotechnologies (LabLENI) – Institute Human-Tech, Universidad Politécnica de Valencia, Spain

Over the past few decades, Virtual Reality (VR) has emerged as a popular topic in a wide range of fields, such as information technology and psychology, among others. One reason for its importance was due to the ability to virtualize real-world scenes. This process typically involves capturing data from those scenes to generate accurate, detailed, and immersive 3D models. However, the creation of virtual content from real-world scenes has traditionally relied on manual techniques, as well as photogrammetry or Computer Vision (CV) algorithms. This frequently yields time-intensive, less accurate, intricate, and semi-automatic results.

To tackle these limitations, we present a novel framework. It employs CV and Deep Learning (DL) techniques to swiftly and seamlessly virtualize any 3D scenario from real indoor environments. Our approach consists of three main stages. Firstly, a 3D reconstruction is conducted using an adapted BundleFusion method to acquire the indoor scene for examination. Secondly, state-of-the-art algorithms are utilized for 3D scene understanding, including 3D instance detection based on Mask3D, scene segmentation employing O-CNN, a bespoke layout strategy, and an alignment technique to map 3D CAD models from ShapeNet based on a novel solution named ScanNotate. Finally, the third stage is the integration and consolidation of each method into a single Unity3D application, facilitating concurrent visualization and customization of the analysed 3D scene.

The results showcase the method's capability to produce highly realistic and visually appealing virtual environments. By eliminating the need for manual intervention and offering an automatic approach, this framework has the potential to significantly facilitate the virtualization of 3D indoor scenarios from real scenes, making the process easier, more precise, unified, consistent, automated, and effective for a broad spectrum of VR applications.

In conclusion, the proposed method not only addresses the limitations of traditional approaches but also opens up new possibilities for the development of highly immersive and interactive VR experiences.

Exploring neural codes and intersubjective alignment of time experience

Camille Grasso
Cognition & Brain Dynamics, NeuroSpin, UNICOG, CEA, France

In this presentation, I will be sharing the progress made in two projects focused on the perception and production of durations.

The first project aims at uncovering how the brain represents time and how this representation is influenced by environmental constraints. To this end, we manipulated

environmental constraints in virtual reality (e.g., size of the room, height of the ceiling) and combined behavioural measures of duration perception and production with electrophysiological recordings to characterise the underlying neural activity. Behavioural results revealed that participants produced longer durations in large environments, as compared to smaller ones. We then used time-resolved multivariate decoding approaches, which consists in training classifiers at specific time points and testing them at every other time points to understand the dynamic neural code underlying these effects. By using electrophysiological measurements and decoding algorithms, we sought to identify the cerebral topography associated with time perception.

The second project aims at aligning the subjective temporal experiences of different individuals. Inspired by recent work on colour qualia, we collected similarity judgments between pairs of durations and constructed a similarity matrix. We then applied multidimensional scaling to this (dis)similarity matrix to represent each participants' psychological space of duration perception. We then used an unsupervised learning algorithm (Gromov-Wasserstein optimal transport) to "align" the subjective experiences of durations (i.e., to find the objective durations that produce the same subjective experience across individuals).

Both projects promise valuable insights into the neural and cognitive processes involved in temporal perception, the impact of environmental constraints, and foster a deeper understanding of intersubjective experiences of time.

3.   CAROUSEL

Interactive Characters for Virtual Reality

Adas Slezas
Aalto University

The process of crafting, simulating and imbuing 3D characters with interactivity is an important part in games, movies and computer graphics. With the emerging popularity of Virtual Reality (VR) this facet of creation has taken on heightened importance, dictating the quality of immersive experiences. This presentation serves as an introductory guide to character creation within VR applications, coupled with the intricate art of animating these characters using cutting-edge motion matching techniques to seamlessly synchronize their actions with the virtual environment, thus fostering a heightened sense of reactivity and realism.

Simulating Fatigue and Recovery in Virtual Characters

Noshaba Cheema
DFKI

This presentation introduces our novel reinforcement learning-framework which -- for the first time in literature -- generates control-policies for full-body physically simulated agents aware of cumulative fatigue, without the need for (potentially dangerous) fatigued motion capture data. For this we make use of a Generative Adversarial Imitation Learning (GAIL) framework and a Three-Compartment Controller (3CC) model from biomechanics literature. The work has been conditionally accepted to Siggraph Asia 2023.

Dance and Social Media Workshop

Alexandra Schmucklermann
DFKI

Wrapping up a tiring day filled with presentations and screen time, we aim to conclude on a revitalizing note by incorporating exercise and an enjoyable exploration of content creation for social media and marketing. Alex is a professional dancer and acrobat with many years of experience in choreographing and performances. Additionally, her skills in marketing and social media growth has lead her to gain a following on Instagram and TikTok with over 60k followers.

4.   TOUCHLESS

How Haptic Experiences are Made

Tor-Salve Dalsgaard
University of Copenhagen, Denmark

Haptic technologies have enormous potential to mediate experiences in the human mind. However, it remains difficult to determine the interconnection between haptic stimulus, sensation, perception, and experience. This is unfortunate, as understanding these connections would allow hapticians to design haptic experiences based on a theoretical foundation.
In this talk, I will present my naïve theory on how haptic experiences are made. I will present evidence for some aspects of the theory, explain my assumptions and limitations, and poke holes into the theory. Please join me in strengthening this theory by tearing it apart!

Emotional textile design through multisensory interaction design

Jing Xue
UCL

The interaction between fabric and our body or hand plays an important role in our everyday lives. We often select fabric products (e.g. clothes, blankets) based on how the fabric feels like (e.g. soft, rough). This kind of interaction with fabrics are not only sensory experiences but also an emotional one. However, while such experiences could be easily made in the shop, there is a still a limitation in the digital space (e.g. online shopping and VR space). This kind of limitation is mainly due to the lack of understanding in mapping the tactile experiences of fabrics' tactile experiences to emotional responses during the interaction. Although there is already a substantial body of knowledge about the tactile perception of fabrics, the relation between the physical properties of the fabric, the touch sensation, and its related emotional responses is still widely under explored within Human-Computer Interaction (HCI), and especially with regards to digital interaction design.
The aim of my Ph.D is to explore the mapping between tactile perception and emotional responses of fabrics through the use of mid-air haptics, an emerging digital technology that enables the design of novel touchless experiences and interactions.
I am particular interested in exploring the coupling of mid-air haptic stimulation with visual and other sensory stimuli to design multisensory emotional fabric experiences. I will address

the question of how we can create new digital textile experiences to convey fabric's perceptual and emotional responses to users. The outcome of this Ph.D thesis will aim to inform the design of future multisensory textile experiences in immersive environments, such as virtual reality.

A closed-loop, real-time, predictive model of human response to ultrasonic tactile stimulation using data-driven approaches

Zhouyang Shen
UCL

Contactless interaction is an emerging field that has gained increasing interest in recent years. One of the main approaches used in this field is mid-air haptic stimuli, which can be delivered using air-jet. Among these methods, ultrasonic mid-air haptics (UMH) has become increasingly popular due to its high spatial and temporal resolutions. To achieve immersive contactless interactions, many researchers have explored how UMH affects the tactile experiences (i.e., perceptions and emotions) users can develop from the stimuli. Despite that, currently, no comprehensive model can accurately predict the tactile experiences users perceive from UMH devices or determine the required stimulation parameters to control users' perceptions and emotions under different interaction contexts. This research addresses this gap by developing a model to comprehensively predict and control perceptions and emotions humans can develop from UMH stimulus. We accomplish this by 1) reviewing the existing literature on the stimulation parameters and potential perceptions inducible by the haptic device, 2) characterizing user perceptions and emotions through subjective and objective measures, 3) using data-driven approaches to link stimulation parameters with perceptual and emotional responses, and 4) developing a real-time closed-loop control system to induce desired perceptions and emotions through adjusting UMH stimulation parameters. The contributions of this research enable the development of more advanced UMH systems that can provide highly immersive and personalized experiences. It also offers significant insights into integrating human-computer interaction and artificial intelligence, demonstrating the potential of data-driven approaches to enhance contactless interaction systems.

5.   GUEST XR

Empathetic haptics: feel the stress or a virtual agent

Jeanne Hecquard
Inria, France

We study the promotion of positive social interactions in VR by fostering empathy with other users present in the virtual scene. For this purpose, we propose affective haptic feedback to reinforce the connection with another user by transmitting their physiological state.
We developed a virtual meeting scenario wherein a human user attends a presentation with several virtual agents. Stressor events make the virtual presenter increasingly stressed and anxious as the presentation proceeds, but the participant can help the presenter and reduce her stress level. The participant directly receives stress via twop physiologically based

affective haptic interfaces. A compression belt and a vibrator simulate the stressed breathing and heart rate, respectively.

Influence of prior acquaintance on the shared VR experience

Esen Küçüktütüncü
University of Barcelona, Spain

Shared Virtual Reality has been around since the early 1990s, where copresence was achieved using avatars, tracker, and later on lip-sync. However, most of these meetings were with experts in the field. With our latest experiment, we wanted to know, how non-experts respond, and to better understand it, go beyond the questionnaires, and assess the participants' sentiments. In this talk will give a presentation of our latest exploratory study where our main question was: Does prior acquaintance have an influence on the experience of the user in a shared VR scenario? I will give a brief introduction to the tools we use for copresence and how we make use of ML algorithms to assess participant sentiments.

Deep learning for acoustic matching in the VR context

Joanna Luberadzka
Eurecat, Spain

To achieve a homogeneous acoustic space in VR, the sound captured by individual users' microphones needs to be transformed so that its acoustic properties match the virtual scene. This problem can be divided into two sub-tasks where the first one is to de-reverberate a signal and the second one is to convolve the de-reverberated signal with a room impulse response (RIR) to obtain a perceptually different space. However, the desired RIR is usually not explicitly given. In this work, we are solving the task of estimating the RIR from speech recorded in a reverberant acoustic space. Our approach is to use deep learning, specifically an encoder-decoder network, which takes a time domain signal at the input and estimates the RIR (or parameters required to approximate an RIR) at the output.