Towards a Synesthesia Laboratory: Real-time Localization and Visualization of a Sound Source for Virtual Reality Applications

Ahmet Kose, *Member, IEEE,* Aleksei Tepljakov, *Senior Member, IEEE,* Sergei Astapov, Dirk Draheim, Eduard Petlenkov, and Kristina Vassiljeva

***Abstract*—In this paper, we present our findings related to the problem of localization and visualization of a sound source placed in the same room as the listener. The particular effect that we aim to investigate is called synesthesia—the act of experiencing one sense modality as another, e.g., a person may vividly experience flashes of colors when listening to a series of sounds. Towards that end, we apply a series of recently developed methods for detecting sound source in a three-dimensional space around the listener. We also apply a Kalman filter to smooth out the perceived motion. Further, we transform the audio signal into a series of visual shapes, such that the size of each shape is determined by the loudness of the sound source, and its color is determined by the dominant spectral component of the sound. The developed prototype is verified in real time. The prototype configuration is described and some initial experimental results are reported and discussed. Some ideas for further development are also presented.**

***Index Terms*—virtual reality, synesthesia, acoustic localization, microphone array, therapy**

I. Introduction

VIRTUAL Reality (VR) provides novel means to visualize and interact with virtual environments including complex visualizations [1], [2]. The advent of low-cost head-mounted display (HMD) devices such as Oculus Rift [3] and HTC Vive [4] made this technology accessible at large, whereas before it was only available to a few specialized laboratories, private companies, and military bodies [5]. Present day advances in VR technology allow to induce a persistent effect of presence in a visual world created using advanced user real-world position and orientation tracking in head-mounted displays (HMDs). Virtual environments, including those already adopted to make use of VR, have proven effective in a number of scientific, industrial and medical applications [6], [7], [8], [9], [10], [11] including but not limited to: underground cave analysis and archaeology, architecture, paleontology, geographic information systems, geosciences, shape perception, physics, organic chemistry, education, MRI and brain tumor analysis, rehabilitation and therapy. Besides that, VR applications also aim to alleviate obstacles of physical environments, particularly for education purposes. Assuring quality of practice in facilities has become rather difficult problem with correlation between complexity of equipment and its cost. Moreover, the substantial increase in the number of students enforces the limited capacity of facilities. Facing inconvenient conditions have convinced researchers to employ meaningful replicated tools in virtual environments. They have also accomplished several real-time VR applications to avoid if not relief continuous difficulties to engage students with experiments adequately [12], [13], [14], [15].

Synesthesia—the act of experiencing one sense modality as another—is an interesting phenomenon that provides many exciting opportunities when applied to VR. For example, some promising results on cross-modal sensory integration in virtual environments were reported in [16]. The ultimate goal of the project, the development of which is documented in this paper, is to provide means for inducing voluntary synesthetic experiences through the VR environment. In our earlier work [17], we have reported initial findings related to acoustic localization and sound processing based on prerecorded data. In this work, we describe the revised technical solution meant to deliver the synesthetic experience to the listener in real time.

We now outline the main contribution of the present paper. First, we provide a firm motivation for a full-scale synesthesia laboratory to be developed during the course of related research activities—this largely complements the contribution described in [18]. Then, we describe the first technological contribution towards the development of such a laboratory. Namely, we investigate the application of an acoustic localization method to the problem of locating the sound source in a room. We also apply a Kalman filter to reduce motion noise generated by the uncertainty of sound source location prediction. Next, for consistency, we summarize the method for extracting dominant features from the audio spectrum of the sound source and mapping those to the object representing the sound source in the VR environment [17]. Then, we provide the description of the proposed VR system prototype and the complete experimental configuration and present the novel developments related to the real-time implementation thereof. Finally, we report and analyze the initial findings related to the synesthetic experiences and outline some items for future research.

Manuscript received November 15, 2017; revised February 10, 2018. Date of publication March 15, 2018. Prof. Mladen Russo has been coordinating the review of this manuscript and approved it for publication.

Authors are with the Department of Computer Systems, Department of Software Science, Tallinn University of Technology, Estonia (e-mails:{ahmet.kose, aleksei.tepljakov, sergei.astapov, dirk.draheim, eduard.petlenkov, kristina.vassiljeva}@ttu.ee).

Digital Object Identifier (DOI): 10.24138/jcomss.v14i1.410

The structure of the paper is as follows. Section II is dedicated to outline of general concept of the Synesthesia Laboratory for initial experiments. In Section III the proposed acoustic localization and feature extraction method is described. The creation of VR environment is explained and data communication is pointed out in Section IV. In Section V the reader is introduced to the developed real-time VR system prototype. Experimental results are discussed in Section VI. Finally, conclusions are drawn in Section VII.

II. Towards a Synesthesia Laboratory

A. Sense Swapping

The current achievement of turning 3D sound occurrences into 3D visual signals represents a step of a larger programme. The ideal target is the ability to completely swap senses, i.e., turning all audio signals that can be heard by an individual human being, from arbitrarily many sources into video signals and, vice versa, turning everything that can be seen by a person into a sound carpet. Given a preset sound-to-light mapping: how would Gustav Mahler's 2nd symphony look like [19]? We consider the ideal concept of *sense swapping* as a research challenge to drive the development of increasingly better tools in sense-crossing and sense-mapping. In particular, if integrated into augmented reality scenarios, real-life applications for such technologies can be found immediately in numerous domains: from architecture over education, supportive technology for the elderly and people with special needs, therapy and rehabilitation, to rescue systems and all kinds of mission-control systems. Furthermore, and this is particularly important for us, we see such technologies as enabling technologies for foundational research in synesthesia, as we will argue in due course in Sects II-B and II-C.

B. The Synesthesia Laboratory and Knowledge Base

In a broad sense synesthesia [20], [21] is about merging different senses' stimuli. The concrete meaning of synesthesia is usually said to be the experience of a so-called cross-modal sense perception, i.e., the experience that a stimulus triggers a perception with regard to a sense different from the one that it usually belongs to. People that regularly or intensively experience synesthesia are called synesthetes, compare with Fig 1 that shows how synesthete Alexander László perceives certain chords as colors if played on a piano.

When does a person start to be a synesthete? This question is hard to answer. Actually, due to lack of a joint reference framework, it is impossible to decide this question. For example, we can test whether a person has the perfect pitch, but we do not know how to test whether he or she is a synesthete. Such considerations will lead us to a discussion of the theoretical foundations of synesthesia later in Sect. II-C. For the moment, it is important to see that all people experience synesthesia to a certain degree at some level. We all know that we would associate some *music* pieces with a warm *temperature* and with an orange or red *color*, whereas we would judge other music as cold and maybe green or blue. It is that day-to-day form of synesthesia that makes the concept so relevant. We have seen massive efforts in synesthesia throughout the ages,

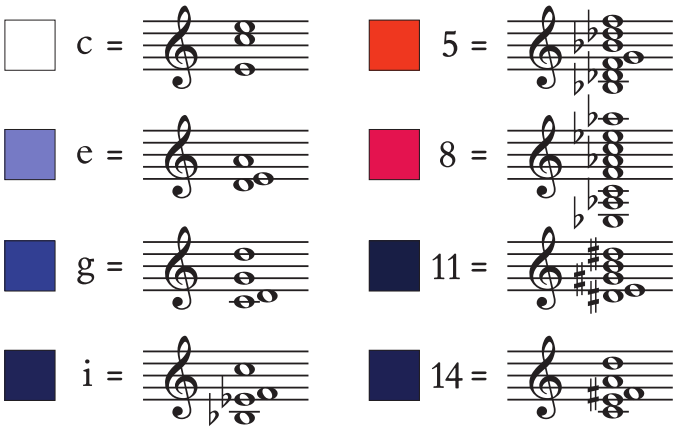


Fig. 1. How synesthete Alexander László perceives certain chords if played on piano, compare with Fig. 18 in 'Die Farblichtmusik' [22].

ranging from the holistic design of the medieval cathedrals to the different forms of multimedia in contemporary pop music.

To explain the relevance and importance of synesthesia let us discuss the Weimarer Bauhaus. It was the Bauhaus that aimed at the renaissance of the cathedrale's Bauhütte. Here, the Bauhaus vision was a holistic approach to architecture, painting, and sculpture in service of the *complete building* [23]. However, a different viewpoint on the Bauhütte is possible, i.e., as (i) an even more holistic approach to all fine arts, including music[1], that is (ii) not in service of the complete building, but in service of the *synesthetic experience*, i.e., the cathedral can also be considered a music instrument—*form ever follows function* [25].

In practice and also in research synesthesia is often used, rather loosely, to refer also to the design of synesthetic scenarios, i.e., to refer to multimedia design efforts. This does not harm, but in order to explain our vision of the synesthesia laboratory better, it is helpful to delve into the differences. If we want to approach terminology accurately, synesthesia stands for cross-modal perceptions. In our endeavors, we also want to use it for cross-modal associations, which applies to more people and is at the same time less contradictory. Multimedia design efforts, or just multimedia for short, are intentional multi-modal compositions – the different simultaneous channels mutually support each other, created harmonies, as well as tensions between them are in service of a holistic experience.

Given its relevance, synesthesia has been subject of investigation from different perspective, including music theory, psychology, and neurology [21], [26]. Consequently, music in virtual reality forms a quickly emerging field with a plethora of exciting projects [27]. Also, we see first endeavors in systematic reasarch on cross-modality in virtual environments, e.g., [28], [16].

Artists and multimedia designers rely on artistic inspiration to create multimedia art [29], [19], [30]. Here is where our vision of a *synesthesia laboratory* enters the scene. The vision might be best described as a *synesthesia experience*

[1]Note that architecture, painting, and sculpture are incorrectly translated as fine arts in the English version of the Bauhaus Manifesto [24]

factory. It is about creating and providing leading-edge tools for the systematic generation, comparison and assessment of cross-modal experiences involving virtual reality. It aims at foundational insights in cross-modal experience, building, step-by-step, a *synesthesia knowledge base*. As an example, consider case no. 5 in Fig. 1. It tells of that Alexander László perceives the chord B♭m7 as an orange color. Does the color change if the chord is played on a guitar rather than a piano? What is the color if the notes of the chord are not played simultaneously but one by one? With appropriate tools we can start experimenting. Which kind of sound-to-color mapping do we like best for a melody floating through a 3D virtual world? Which sound-to-color mapping fits best to visualization of a whole symphony orchestra, a choir or a Jazz band. The single individual can experiment, but from there we could start seeking for group-specific synesthesia or asking whether there exist anthropological constants. Is Alexander László the only one, who perceives B♭m7 as an orange color or is that rather the rule? Perceive people with different cultural background, age, gender the chord differently?

The vision of a *synesthesia laboratory* is crucial building block of the encompassing Re:creation Virtual and Augmented Reality laboratory [31] vision, compare with [32].

C. Monad Theoretical Foundation

An original trigger for *swapping senses* has been the Leibniz anniversary year 2016. The intention is to create a living metaphor for Leibniz' monad theory [33], [34], primarily for its cognitive aspects that are concerned with *communication and perception* and not so much for its meta-physic conclusion that are about the concept of *pre-stabilized harmony* [35]. For this purpose, the concrete way how senses are mapped to each other is not important. What is important is the fact that the senses are exchanged at all. Crucial and at the same time a challenge is that all information is conserved by concrete exchanges. In Leibniz' theory monads are the smallest building blocks of mind. Monads interact only via their senses. Opinions about impressions are just learned. They are just negotiated between individuals. Actually, it turns out that it even makes no sense to ask how others materialize sensual impressions in their mind, because of the lack of a joint reference framework. We believe that the experience of exchanged senses can serve as a great didactical device to help understanding the cognitive part of monad theory.

III. Proposed Method for Sound Localization, Processing, and Visualization

In what follows, we outline each stage of the method in a separate subsection.

A. Acoustic Localization

The acoustic data related to the sound source of interest is acquired by an array of spatially distributed audio sensors (microphones). This paper considers a conical configuration of sensors, however, the approach presented below can be applied to other configurations with minimal modifications.

The task of three-dimensional acoustic localization in spherical coordinates lies in estimating the parameters (r, θ, ϕ) of the sound source, where r is the distance to the source, θ is elevation, and ϕ is azimuth. For our VR simulation we assume that r is known, as discussed in Section VI, therefore, the task of localization comprises estimation of the Direction of Arrival (DOA) consisting of angles ϕ and θ. For DOA estimation we apply the Steered Response Power with Phase Transform (SRP-PHAT) method, which is highly robust and tolerant to reverberation [36].

The SRP $P(\boldsymbol{a})$ is a real-valued functional of a spatial vector $\boldsymbol{a}$, which is defined by the Field of View (FOV) of a specific array. The maxima of $P(\boldsymbol{a})$ indicate the direction to the sound source. $P(\boldsymbol{a})$ is computed as the cumulative Generalized Cross-Correlation with Phase Transform (GCC-PHAT) across all pairs of sensors at the theoretical time delays, associated with the chosen direction. Consider a pair of signals $x_k(t)$, $x_l(t)$ of an array consisting of M microphones. The time instances of sound arrival from some point $a \in \boldsymbol{a}$ for the two microphones are $\tau(a,k)$ and $\tau(a,l)$, respectively. Hence the time delay between the signals is $\tau_{kl}(a) = \tau(a,k) - \tau(a,l)$. The SRP-PHAT for all pairs of signals is then defined as

$$P(a) = \sum_{k=1}^{M} \sum_{l=k+1}^{M} \int_{-\infty}^{\infty} \Psi_{kl}(\beta) X_k(\omega) X_l^*(\omega) \mathrm{e}^{j\omega\tau_{kl}(a)} \mathrm{d}\omega, \tag{1}$$

where $X_i(\omega)$ is the spectrum (i.e., the Fourier Transform) of signal $x_i(t)$, $X_i^*(\omega)$ is the conjugate of that spectrum and $\Psi_{kl}(\beta)$ is the β-PHAT weight, defined as

$$\Psi_{kl}(\beta) = \left(|X_k(\omega) X_l^*(\omega)|\right)^{-\beta}. \tag{2}$$

The PHAT is used for eliminating reverberation effects, though, it can over-sharpen the SRP. Applying the more flexible β-PHAT weight allows to adjust to specific reverberation levels. We use the coefficient $\beta = 0.8$ in our experiments.

Though SRP-PHAT is effective and robust, it requires significant amounts of computational power if computed in the frequency domain and applied to large spatial vectors $\boldsymbol{a}$. In order to be able to compute the SRP in real-time, we calculate the GCC in the time domain using an integer delay step beamformer and also reduce the spatial vector. To reduce vector $\boldsymbol{a}$, it is proposed to perform horizontal and vertical DOA estimation separately. In this manner the horizontal plane is divided into n_h and the vertical plane—into n_v possible discrete angles, respectively. The points are chosen in the volumetric FOV along a spherical surface with radius r_{FOV}. The horizontal evaluation is performed along a circumference of a half circle, which covers the front FOV of the array, i.e., in the angle interval of $[0, \pi]$. The discrete angle step is calculated as $\phi_h = \frac{\pi}{n_h}$. The horizontal evaluation is performed for the points $a_h(i) = (x_h(i), y_h(i), 0)$:

$$\begin{aligned} x_h(i) &= r_{FOV} \cos\left(i\phi_h\right), \quad (0 \le i \le n_h), \\ y_h(i) &= r_{FOV} \sin\left(i\phi_h\right), \quad (0 \le i \le n_h). \end{aligned} \tag{3}$$

The azimuth ϕ is estimated in the directions of elevated SRP values. For a single source case it is equal to

$$\phi = \arg\max\left(P(\boldsymbol{a}_h)\right) \cdot \phi_h. \tag{4}$$

After obtaining ϕ, the vertical SRP-PHAT evaluation is performed over the vertical half-circumference from the positive z-axis downward, i.e. $[0, \pi]$, with a discrete angle step of $\theta_v = \frac{\pi}{n_v}$, in the direction of established azimuth ϕ for the points $a_v(i) = (x_v(i), y_v(i), z_v(i))$:

$$\begin{aligned} x_v(i) &= r_{FOV} \cos(\phi) \sin(i\theta_v), \quad (0 \le i \le n_v), \\ y_v(i) &= r_{FOV} \sin(\phi) \sin(i\theta_v), \quad (0 \le i \le n_v), \\ z_v(i) &= r_{FOV} \cos(i\theta_v), \quad (0 \le i \le n_v). \end{aligned} \tag{5}$$

The elevation angle is estimated in the direction of elevated SRP, and also brought to a more comprehensive interval $\left[\frac{\pi}{2}, -\frac{\pi}{2}\right]$ from the positive z-axis downward:

$$\theta = \frac{\pi}{2} - \arg\max\left(P(\boldsymbol{a}_v)\right) \cdot \theta_v. \tag{6}$$

In our experiments a single degree angle resolution is chosen for both azimuth and elevation, therefore, parameters are set as $n_h = 180$, $n_v = 180$; the radius is set to $r_{FOV} = 0.5$ m.

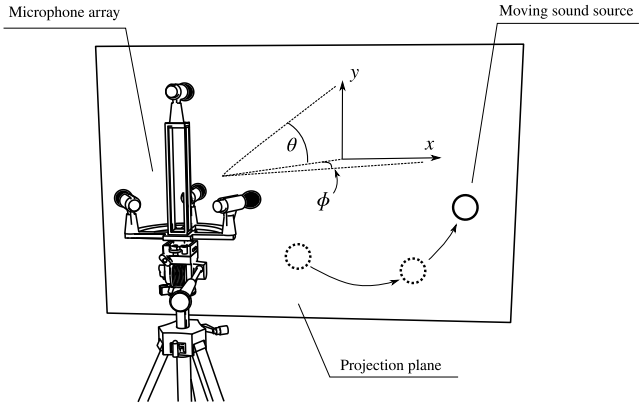


Fig. 2. Localization principle: The sound source is assumed to be moving in a plane

B. Kalman Filter for Motion Tracking

The discrete time Kalman filter (KF) is a linear quadratic estimator [37], which provides the closed form recursive solution for a linear discrete-time dynamic system of the form:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_{k-1}\mathbf{x}_k + \mathbf{r}_{k-1} \end{aligned}, \tag{7}$$

where $\mathbf{x}_k$ is the system state vector at time step k, $\mathbf{y}_k$ is the measurement vector at k, $\mathbf{A}_{k-1}$ is the transition matrix of the dynamic model, $\mathbf{H}_{k-1}$ is the measurement matrix, $\mathbf{q}_{k-1} \sim \mathrm{N}(0, \mathbf{Q}_{k-1})$ is the process noise with covariance $\mathbf{Q}_{k-1}$ and $\mathbf{r}_{k-1} \sim \mathrm{N}(0, \mathbf{R}_{k-1})$ is the measurement noise with covariance $\mathbf{R}_{k-1}$. Kalman filtering consists of a prediction step, where the next state of the system is predicted given the previous measurements, and an update step, where the current state is estimated given the measurement at that time instance. The prediction step is characterized by the following equations:

$$\begin{aligned} \hat{\mathbf{x}}_k^- &= \mathbf{A}_{k-1}\hat{\mathbf{x}}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{A}_{k-1}\mathbf{P}_k\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \end{aligned}, \tag{8}$$

where $\hat{\mathbf{x}}_k^-$ and $\mathbf{P}_k^-$ are the system *a priori* (i.e., before observing the measurement at time k) state and covariance estimates, and $\hat{\mathbf{x}}_k$, $\mathbf{P}_k$ are *a posteriori* (i.e., after observing the measurement) estimates. The update step is performed as:

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \\ \hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_k^- + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-\right) \\ \mathbf{P}_k &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_k^- \end{aligned}, \tag{9}$$

where $\mathbf{K}_k$ is the Kalman gain of prediction correction at time instance k. KF is optimal for a linear system with Gaussian measurement and process noise [38], [37], which applies to our situation.

Acoustic object movement is described as a discrete Wiener process velocity model [39] with the state vector defined as $\mathbf{x}_k = \begin{bmatrix} x_k & y_k & \dot{x}_k & \dot{y}_k \end{bmatrix}^T$, where (x_k, y_k) denotes object position and $(\dot{x}_k, \dot{y}_k)$ — the velocity in a two-dimensional Cartesian space. The transition and measurement matrices for model (7) are then defined as:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{10}$$

where Δt is the time interval between consecutive estimates in seconds. The process and measurement noise variance is specified by matrices

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{3}\Delta t^3 & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & \frac{1}{3}\Delta t^3 & 0 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & 0 & \Delta t & 0 \\ 0 & \frac{1}{2}\Delta t^2 & 0 & \Delta t \end{bmatrix} q,$$

$$\mathbf{R} = diag(r_1, r_2), \tag{11}$$

where q and r_k are the power spectral densities of process and measurement noise, respectively. For our conditions the parameters are set as $\Delta t = 0.1$ s, $q = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.2 \end{bmatrix}^T$, $r_1 = 0.1$, $r_2 = 0.3$.

C. Acoustic Feature Extraction and Mapping

Mel-Frequency Cepstral Coefficients (MFCC) is a popular method for extracting features from an audio signal and is used in speech detection and recognition. An interesting property of this approach is that it is based on psychophysical studies that reveal that human perception of the sound frequency contents for speech signals does not follow a linear scale [40]. A Mel scale is used instead:

$$f_m = 2595 \log_{10}\left(1 + \frac{f}{700}\right),$$

where f_m is the subjective pitch in Mel units corresponding to a frequency f in Hz. Thus, applying this method can be beneficial in improving the interplay of sound and visual excitation in a psychophysically coherent way. Moreover, studies also show that MFCC can also be applied to modeling music [41] which can also be advantageous in enhancing the experience of synesthesia. The computation of the MFCC comprises the following steps [41]. Starting from a waveform:

1) Convert waveform to frames;
2) Take Discrete Fourier Transform (DFT);
3) Take Log of amplitude spectrum;

4) Apply Mel-scaling and smoothing;
5) Apply Discrete Cosine Transform (DCT).

Once the procedure is complete, MFCC features are obtained. In this work, we consider the auditory spectrum portion of the features, hereinafter denoted as A_{spec}. To produce the visualization, the acoustic features of the sound source must be examined. At the moment, a simple approach is employed, such that the incoming sound waves are visualized as spheres moving towards the listener. The color, size, velocity of travel, and sampling rate for generating the spheres can be determined experimentally. The incoming waveforms are broken down into frames and analyzed as discussed previously. Currently, the following mappings are in effect:

- The size of a single sphere is determined by the scaled maximum amplitude of the sampled waveform in the frame;
- The color of the sphere is determined by the dominant feature in A_{spec} obtained by applying the MFCC approach. A transform is thus defined as $\xi : \mathscr{I} \rightarrow \mathscr{C}$, where $\mathscr{I} \subset \mathbb{N}$ is the index of the dominant feature in A_{spec}, and $\mathscr{C} \subset \mathbb{R}^3$ is the color specification in a particular color space. For this work, we consider the Red-Green-Blue (RGB) color space and jet color mapping, the latter shown in Fig 3.

Fig. 3. Jet Color Mapping

The signal processing described above is completely done in MATLAB environment and sent to the visualization engine for further processing.

IV. Real-Time Application to a Virtual Reality Environment

The VR environment conducted with synesthetic experiences is advanced for various purposes such as bridging Cyber-Physical Systems (CPS) to VR interface, self-learning activities, physiological and psychological aspects of VR based real-time simulation. Although the immersive environment is unique and subsections are slightly relevant, the process for creating the virtual environment had been familiar with ordinary VR applications. The process is as follows:

(a) Environmental Creation and System Design,
(b) Modeling and Texturing,
(c) Materials and Hardware Integration,
(d) Software Integration,
(e) Optimization.

Unreal Engine 4 was employed to create the virtual environment as the engine is well known, feature-complete and capable solution for VR development purposes [42]. Robustness, compatibility and low maintenance cost could be considered as some of prior advantages for the preferred pyhsics engine. Overall, using primary game engines is relevant to benefit prominent VR development methods such as the concept of object oriented programming, process of computer generated graphics, reusable code with libraries[43]. Those benefits are also used to create communication tool between software linked to the project.

Regardless, successful immersive applications based on realistic environment minimize difference among physical and virtual environments. Hence, authors considered to create the virtual environment based on physical conditions significantly. Additionally, the floor plan of facility is utilized used for accurate virtual environment based on physical assets [44]. Those assets have been rendered and modeled in three dimensions (3D) by using Autodesk Maya3D animation, modeling, simulation, and rendering software [45]. 3D models of physical assets located at laboratory were developed using polygon shape and modified using functions such as extrudes, append etc. The physics engine can respectfully identify the material components and map file (Textures) with Filmbox format (FBX). We also aimed to maximize impact of synesthesia during the experiment. Whereas, existing objects in virtual environment should be perceived in minimum acceptable level. Therefore, we preferred to avoid emissive color contents for replicated objects. Fundamental parts such as walls, columns etc. are inserted in physics engine according to suggested scale compatibility of physics engine (1 unreal unit is equal to 1 cm) [46]. The chosen engine allows to create realistic environments and simulate realistic affects. Hence, users can feel lifelike interaction during experiencing the application. Moreover, teleportation feature of VR technology is implemented for the application. That feature allows users to move independently and effortless in computer simulated realistic laboratory during experiencing synesthesia effect. Furthermore, independent movement may also grant users to sense self-learning activities in VE. The user can still ascertain sound source localization during continuous movements. The principle logic of the application is referred to predefined VR class of the engine: Motion Controller Pawn, HUD, VR GameMode, Player Controller.

User Datagram Protocol (UDP) provides a procedure for application programs to send messages to other programs with a minimum of protocol mechanism [47]. The lightweight procedure does not acquire specific requirements when running in numerous platforms. UDP transport protocol provided us a straight method to transfer packets over local network among MATLAB and the physics engine.

The Blueprints Visual Scripting system in Unreal Engine is a complete gameplay scripting system based on the concept of using a node-based interface to create gameplay elements from within Unreal Editor [48]. The Blueprint system allowed us to create a UDP interface [49] to communicate between MATLAB and the based on a custom C++ class. To avoid problems with Blueprint multithreading in Unreal Engine 4, the implementation uses a custom class variable to transfer data between threads to avoid a racing condition in requesting/getting new data from the UDP socket. With this approach, reliable communication via a UDP socket at high sampling rates such as f_s=1kHz is easily achieved.

The complete visualization thus comprises the following components:

- A C++ class based UDP socket implementation available to the UE4 Blueprint scripting system;

- All necessary animations are scripted in UE4 Blueprints and are based on the information received via the UDP socket from MATLAB software;
- The room where the prototype is located is recreated as a Virtual Reality environment.

Since HTC Vive is used, the corresponding UE4 VR template is employed and thus the user can navigate the VR room. The main idea here is to synchronize the user real-world location and that in the VR environment.

V. Experimental Setup

The proposed prototype comprises several components: a microphone array with four Behringer C2 microphones, an USB audio card—Focusrite Scarlett 18i20 Gen 2—for sound data acquisition, a personal computer running MATLAB/Simulink environment and the VR sound visualizer, the HMD device—HTC Vive, and the emulated sound source represented by a Creative T15 Bluetooth speaker. The user wearing the HMD device and the microphone array are assumed to be stationary. The distance between the user and the conical array is constant and the corresponding VR environment position offset may be computed and applied within the VR room scale. The complete prototype configuration is depicted in Fig. 4.

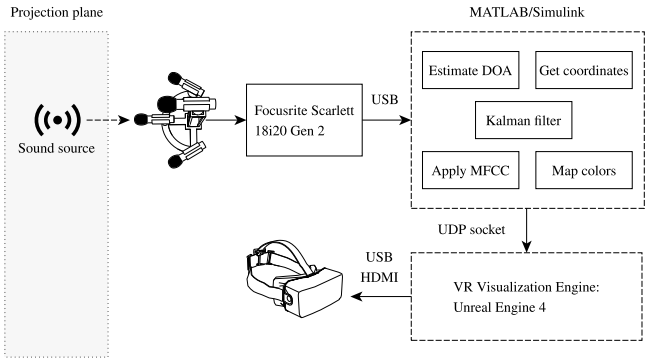


Fig. 4. Experimental configuration and signal flow

Compared to the original configuration described in [18], the following changes to the prototype configuration have been made. We have revised the implementation of the acoustic localization algorithm such that instead of directly using Eq. (5), we adopted a similar approach as in Eq. (3). This means that we skip information about the horizontal location and use only the three vertical microphones to determine the vertical location. This resulted in a significant improvement in accuracy of acoustic localization.

We use MATLAB software due to the availability of necessary toolboxes for real-time data acquisition and processing. We have previously successfully used MATLAB/Simulink for rapid prototyping of computationally intensive real-time control algorithms [50]. Once the data is processed within MATLAB, the necessary information (spatial coordinates of the sphere, its size and color) are sent via a socket connection in real time to the visualizer application that drives the HMD.

Developing with the engine and running the serious game and MATLAB software at once prompts the use of specific hardware and software requirements. The applications run on a PC equipped with a 3.20GHz Intel i5-4570 processor, 8GB RAM, and an NVidia GTX 980 graphic card driving the HTC Vive HMD.

VI. Experimental Results

For verification of the designed prototype we consider two types of experiments. First, we observe how well the acoustic tracking part works. Second, we verify the real-time VR application. In both cases, the distance from the sound source to the microphone array is $r = 1.5$m. An audio clip with modern music is used as audio such that has no distinct spectral features. The real-life experimental layout is shown in Fig. 5.

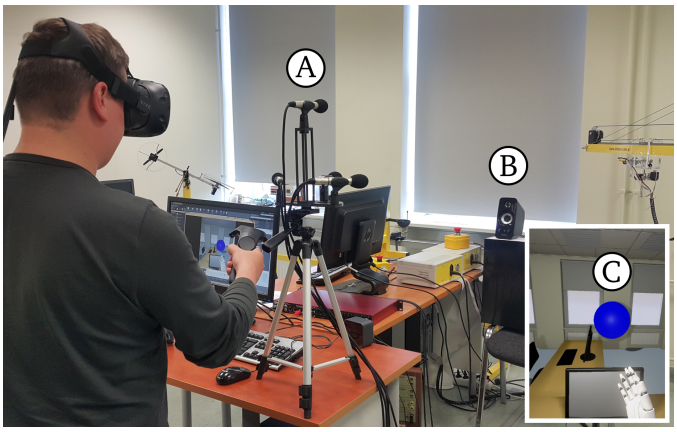


Fig. 5. Experimental real-time prototype setup. Elements on image: A—Microphone array; B—Emulated sound source (Bluetooth speaker); C—Spherical visualization as it appears in the recreated room VR environment

For the first experiment, the sound source, represented by the PC speaker, is moved in a rectangular motion in the projection plane for a 100 seconds. The sampling rate is 10Hz. The resulting scaled circles with colors obtained using the procedure described above as well as corresponding trajectories are shown in Fig. 6. It can be seen that although the manually delivered sound source motion trajectory is approximately rectangular, the location is sometimes misclassified in the vertical direction which results in erroneous upward and downward motions. The motion is smooth due to the additional Kalman filtering step. The larger circles represent a louder section of the audio clip in the lower frequency range. Yellow circles represent features in the higher frequency range.

The second experiment was concerned with initial impressions from using the prototype in VR. The following observations can be made:

- When two people are talking in to each other in the room, the sphere correctly identifies the speaker location. While plain speech does not by itself give rise to special features, a sharp difference in color can be observed when the “s” sound is produced.
- The current implementation introduces a noticeable delay of about 500ms. This is distracting, though does not break the effect of immersion.

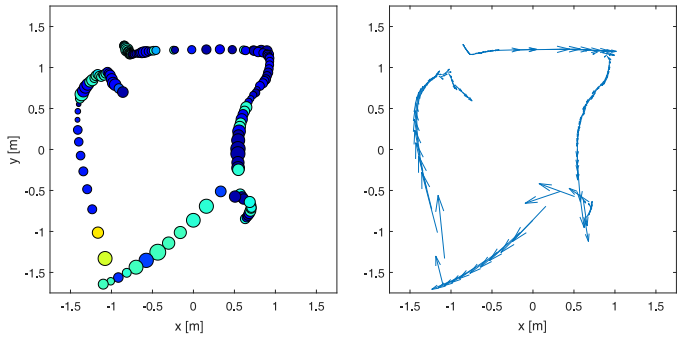


Fig. 6. Features and spatial dynamics of the emulated sound source captured during a real-time experiment with an improved configuration

- Colors produced by playing simple musical instruments fall inside an almost indistinguishable color spectrum (shades of blue, since jet is used).
- The overall effect seems interesting, however, as mentioned above certain technical limitations have to be overcome before further investigation with a control group is conducted.

To provide a further illustration of the visualization part, we moved the prototype to Re:creation Virtual and Augmented Reality laboratory [31]. Therefore, the virtual environment is recreated to represent the actual laboratory. In Fig.7 one can observe a snapshot of the visualization resulting in a speaker in the real environment pronouncing the word "hats". It can be noticed, as mentioned previously, the sound "s" with the *jet* color mapping results in a yellow sphere being produced.

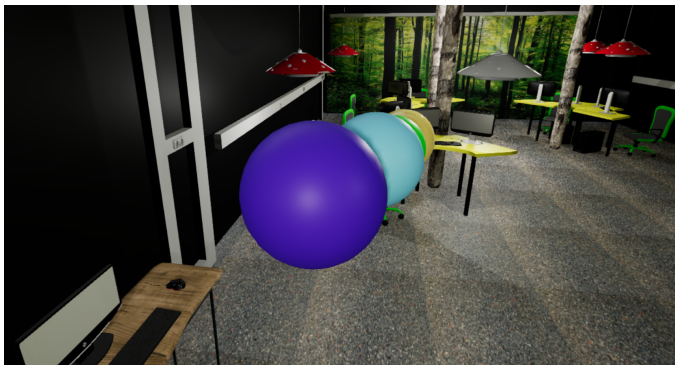


Fig. 7. Visualization resulting from pronouncing the word "hats"

In what follows, several ideas related to the improvement of the proposed solution are provided:

- A graphical user interface for selecting various parameters of the sound processing algorithm should be implemented for convenience. Most importantly, color mapping should be made easily selectable. Furthermore, since we know that Mel's scale is nonlinear, the color mapping may also follow a nonlinear scale to improve color variation in the visualization.
- Depth (third coordinate) detection should be available. To achieve this, it is possible to place HTC Vive standalone trackers directly on the sound source. This can be used for initial experiments as well as calibration of the developed hardware prototype.
- Finally, once the technical limitations are overcome, we shall proceed with the assembly of the control group and do subject based testing. Towards that end, subjects who experience synesthesia naturally should be interviewed, their responses documented and analyzed and corresponding color mapping methods implemented in the prototype. An ethics committee approval should be sought beforehand.

Most importantly, the experiments allowed us to gain deeper insight into the effect generated by introducing sound visualization with source localization in a VR environment. It is also of interest to repeat the experiment in an Augmented Reality (AR) environment where the spheres would overlap an existing real-life environment.

VII. Conclusions

In this work, the general framework towards developing a full-scale synesthesia laboratory has been introduced. Then, the first technological contribution in the scope of this laboratory was presented—we have described a prototype for acoustic sound localization, processing, and visualization for the purpose of inducing a synesthetic experience in a VR environment. The proposed prototype and corresponding methods have been described. In this contribution, the real-time solution was implemented in a VR environment. Initial experiments were successful and provided important insights into the development of synesthetic experiences. While the acoustic localization aspect has some issues, these can be resolved by means of revising the physical configuration and localization algorithm as it was shown in the present effort. Moreover, these tracking issues do not strongly affect the VR immersion aspect, especially due to the application of the Kalman filter which smooths out the motion and enhances the experience, hence the improvement of tracking is not presently seen as top priority. Rather, efforts are being put into research and development of a multiple sound source localization algorithm and a coherent visualization mechanism. The developed application can be used in subject based testing following an ethics committee approval. Since the prospective application is envisioned to be used for real-life medical and artistic applications, further development efforts must also be exhibited towards an embedded system prototype. Furthermore, due to its importance in medical and industrial applications, implementing the synesthetic experience in augmented reality will be investigated since the resulting application will complement the actual real-life environment.

Acknowledgment

The authors would like to express their thanks to undergraduate student Ralf Anari for his valuable assistance in developing reliable UDP socket based communication in Unreal Engine 4.

REFERENCES

[1] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson, "Immersive analytics," in *2015 Big Data Visual Analytics (BDVA)*, Sept 2015, pp. 1–8.

[2] M. Teras and S. Raghunathan, "Big Data visualisation in immersive virtual reality environments: Embodied phenomenological perspectives to interaction," *ICTACT Journal on Soft Computing*, vol. 5, no. 4, 2015.

[3] Oculus VR, LLC. (2017) Oculus Rift. Retrieved on 20.04.2017. [Online]. Available: https://www.oculus.com/rift/

[4] HTC Corporation. (2017) HTC Vive. Retrieved on 20.04.2017. [Online]. Available: https://www.vive.com/eu/

[5] G. F. Welch, "History: The use of the Kalman filter for human motion tracking in virtual reality," *Presence*, vol. 18, no. 1, pp. 72–91, Feb 2009.

[6] J. Psotka, "Immersive training systems: Virtual reality and education and training," *Instructional science*, vol. 23, no. 5-6, pp. 405–431, 1995.

[7] A. A. Rizzo and G. J. Kim, "A SWOT analysis of the field of virtual reality rehabilitation and therapy," *Presence*, vol. 14, no. 2, 2005.

[8] A. K. Shah, J. L. Patton, S. Pacini, N. Hsu, F. Zollman, E. B. Larson, and A. Y. Dvorkin, "Visuo-haptic environment for remediating attention in severe traumatic brain injury," in *Virtual Rehabilitation (ICVR), 2013 International Conference on*. IEEE, 2013, pp. 242–247.

[9] C. Donalek, S. G. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, S. Davidoff, J. S. Norris, and G. Longo, "Immersive and collaborative data visualization using virtual reality platforms," in *2014 IEEE International Conference on Big Data*, Oct 2014, pp. 609–614.

[10] I. R. Draganov and O. L. Boumbarov, "Investigating Oculus Rift virtual reality display applicability to medical assistive system for motor disabled patients," in *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2015 IEEE 8th International Conference on*, vol. 2. IEEE, 2015, pp. 751–754.

[11] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas, "Immersive collaborative analysis of network connectivity: CAVE-style or head-mounted display?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 441–450, Jan 2017.

[12] Y. Liang and G.-P. Liu, "Design of large scale virtual equipment for interactive HIL control system labs," *IEEE Transactions on Learning Technologies*, pp. 1–1, 2017.

[13] F. J. Badesa, R. Morales, N. M. Garcia-Aracil, J. M. Sabater, L. Zollo, E. Papaleo, and E. Guglielmelli, "Dynamic adaptive system for robot-assisted motion rehabilitation," *IEEE Systems Journal*, vol. 10, no. 3, pp. 984–991, sep 2016.

[14] P. Donner and M. Buss, "Cooperative swinging of complex pendulum-like objects: Experimental evaluation," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 744–753, jun 2016.

[15] S. Khan, M. H. Jaffery, A. Hanif, and M. R. Asif, "Teaching tool for a control systems laboratory using a quadrotor as a plant in MATLAB," *IEEE Transactions on Education*, vol. 60, no. 4, pp. 249–256, nov 2017.

[16] F. Biocca, J. Kim, and Y. Choi, "Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions," *Presence*, vol. 10, no. 3, pp. 247–265, 2001.

[17] A. Tepljakov, S. Astapov, E. Petlenkov, K. Vassiljeva, and D. Draheim, "Sound localization and processing for inducing synesthetic experiences in virtual reality," in *2016 15th Biennial Baltic Electronics Conference (BEC)*, Oct 2016, pp. 159–162.

[18] A. Kose, A. Tepljakov, and S. Astapov, "Real-time localization and visualization of a sound source for virtual reality applications," in *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, sep 2017.

[19] J. Deutsch, "Synaesthesia and synergy in art — gustav mahler's "symphony no. 2 in c minor" as an example of interactive music visualization," in *Sensory Perception – Mind an Matter*, F. G. Barth, P. Giampieri-Deutsch, and H.-D. Klein, Eds. Wien New York: Springer, 2012, pp. 215–235.

[20] R. E. Cytowic and D. M. Eagleman, Eds., *Wednesday Is Indigo Blue – Discovering the Brain of Synesthesia*. Cambridge London: MIT Press, 2009.

[21] J. Simner and E. Hubbard, Eds., *Oxford Handbook of Synesthesia*. Oxford University Press, 2013.

[22] A. László, *Die Farblichtmusik*. Leipzig: Breitkopf & Härtel, 1925.

[23] W. Gropius, Ed., *Bauhaus-Manifest*. Die Leitung des Staatlichen Bauhauses in Weimar, April 1919.

[24] ——, *Bauhaus Manifesto*. The administration of the Staatliche Bauhaus in Weimar, April 1919.

[25] L. H. Sullivan, "The tall office building artistically considered," *Lippincott's Magazine*, vol. 57, pp. 403–409, March 1896.

[26] G. F. F. B. ca, J. ao Gabriel Marques Fonseca, and P. Caramelli, "Synesthesia and music perception," *Dementia & Neuropsychologia*, vol. 9, no. 1, pp. 16–23, March 2015.

[27] S. Whiteley and S. Rambarran, Eds., *Oxford Handbook of Music and Virtuality*. Oxford University Press, 2013.

[28] M. Cohen, S. Aoki, and N. Koizumi, "Augmented audio reality – telepresence/vr hybrid acoustic environments," in *Proceedings of the 2nd IEEE International Workshop on Robot and Human Communication*. IEEE, 1993, pp. 361–364.

[29] A. László, "Die farblichtmusik," *Die Musik*, vol. 12, no. 9, pp. 680–683, June 1925.

[30] H. kon Austbø, "Visualizing visions – the significance of Messiaen's colours," *Music & Practice*, vol. 2, 2015.

[31] Tallinn University of Technology. (2018) Official website of Re:creation Virtual and Augmented Reality Laboratory. Retrieved on 01.03.2018. [Online]. Available: https://recreation.ee/

[32] A. Tepljakov, E. Petlenkov, and K. Vassiljeva, Eds., *Re:creation – Virtual and Augmented Reality Laboratory (White Paper)*. Re:creation Laboratory, a Division of Alpha Control Systems Research Laboratory, Tallinn University of Technology, 2016.

[33] G. W. Leibniz, *La Monadologie (1714)*. C. Delagrave, 1881.

[34] ——, "La monadologie," in *Opera philosophica, quae exstant Latina Gallica Germanica omnia*, G. W. Leibniz, Ed. Berlin: J.E. Erdmann, 1820, pp. 700–712.

[35] A. Lamarra, "Contexte génétique et première réception de la monadologie – leibniz, wolff et la doctrine de l'harmonie préétablie," *Revue de Synthese*, vol. 128, pp. 311–323, September 2007.

[36] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Brown University, 2000.

[37] T. Adali and S. Haykin, *Adaptive Signal Processing: Next Generation Solutions*, ser. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control. Wiley, 2010.

[38] S. Vaseghi, *Multimedia Signal Processing: Theory and Applications in Speech, Music and Communications*. Wiley, 2007.

[39] Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[40] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for MFCC feature extraction," in *Signal Processing and Communication Systems (ICSPCS), 2010 4th Intl. Conf. on*. IEEE, 2010, pp. 1–5.

[41] B. Logan, "Mel frequency cepstral coefficients for music modeling," in *ISMIR*, 2000.

[42] Epic Games. Unreal Engine. Retrieved 03.06.2016. [Online]. Available: https://www.unrealengine.com/what-is-unreal-engine-4

[43] C. M. Torres-Ferreyros, M. A. Festini-Wendorff, and P. N. Shiguihara-Juarez, "Developing a videogame using unreal engine based on a four stages methodology," in *2016 IEEE ANDESCON*. Institute of Electrical and Electronics Engineers (IEEE), oct 2016, pp. 1–4.

[44] A. Kose, E. Petlenkov, A. Tepljakov, and K. Vassiljeva, "Virtual reality meets intelligence in large scale architecture," in *Augmented Reality, Virtual Reality, and Computer Graphics*, 2017, pp. 297–309.

[45] Autodesk Maya Software. (2017) Features. Retrieved on 25.05.2017. [Online]. Available: http://www.autodesk.com/products/maya/features/all

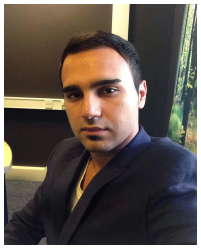
[46] World of Level Design. (2016) UE4/Maya LT: Set Up Grid in Maya LT/Maya to Match Unreal Engine 4. Retrieved on 25.05.2017. [Online]. Available: http://www.worldofleveldesign.com/categories/ue4/ue4-set-up-maya-grid-to-match-unreal-engine4.php

[47] H. Pranoto and A. Ulvan, "Retransmission issue of SIP session over UDP transport protocol in IP multimedia subsystem - IMS," in *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*. IEEE, nov 2013, pp. 273 – 277.

[48] Epic Games. (2017) Blueprints Visual Scripting. Retrieved on 25.05.2017. [Online]. Available: https://docs.unrealengine.com/latest/INT/Engine/Blueprints/

[49] D. Madhuri and P. C. Reddy, "Performance comparison of TCP, UDP and SCTP in a wired network," in *2016 International Conference on Communication and Electronics Systems (ICCES)*. IEEE, oct 2016, pp. 1 – 6.

[50] A. Tepljakov, E. Petlenkov, and J. Belikov, "Implementation and real-time simulation of a fractional-order controller using a MATLAB based prototyping platform," in *Proc. 13th Biennial Baltic Electronics Conference*, 2012, pp. 145–148.



Ahmet Kose received his B.S. (2012) in electrical and electronics engineering from Erciyes University and the M.S. (2015) in computer and systems engineering from Tallinn University of Technology in Tallinn, Estonia. He works as an Early Stage Researcher in TUT's Department of Computer Systems, where he is also currently doing his PhD. His research interests include machine learning, artificial and computational intelligence, system modelling and identification, virtual and augmented reality.

Aleksei Tepljakov received his B.Sc and M.Sc in Computer and Systems Engineering from Tallinn University of Technology in 2009 and 2011, respectively, and his Ph.D. in Information and Communication Technology from Tallinn University of Technology in 2015. Dr. Tepljakov is a member of IEEE since 2011 and a member of IEEE Control Systems Society since 2012. From January 2016 Dr. Tepljakov holds a Research Scientist position at the Department of Computer Systems, School of Information Technologies, Tallinn University of Technology. His main research interests include fractional-order modeling and control of complex systems, fractional-order filter based analog and digital signal processing. Dr. Tepljakov is also interested in developing efficient mathematical and 3D modeling methods for Virtual Reality applications in medicine and education. From March 2017 he serves as head of Re:creation Virtual Reality laboratory in Mektory Business and Innovation Centre.

Sergei Astapov was born in 1988. He received his M.Sc degree in the field of Computer System Engineering at the Tallinn University of Technology in 2011, and his PhD degree in Information and Communication Technology from Tallinn University of Technology in 2016. Dr. Astapov is a member of the Laboratory for Proactive Technologies. His research interests include object tracking using wideband signal analysis, classification tasks and distributed computing in embedded multi-agent systems. His recent research concerns object localization and identification in open environments and acoustic signal based diagnostics of industrial machinery.

Dirk Draheim is full professor of information society technologies and head of the large-scale systems group at Tallinn University of Technology. From to 1990 to 2006 he worked as an IT project manager, IT consultant and IT author in Berlin. In summer 2006, he was Lecturer at the University of Auckland and from 2006-2008 he was area manager for database systems at the Software Competence Center Hagenberg as well as Adjunct Lecturer in information systems at the Johannes-Kepler-University Linz. From 2008 to 2016 he was head of the data center of the University of Innsbruck and, in parallel, from 2010 to 2016, Adjunct Reader at the Faculty of Information Systems of the University of Mannheim. Dirk is author of the Springer books "Business Process Technology", "Semantics of the Probabilistic Typed Lambda Calculus" and "Generalized Jeffrey Conditionalization".

Eduard Petlenkov was born in 1979. He received his B.Sc, M.Sc and PhD degrees in computer and systems engineering from Tallinn University of Technology. Currently, he is an Associate Professor at the Department of Computer Systems at Tallinn University of Technology. His main research interests lie in the domain of nonlinear control, system analysis and computational intelligence.

Kristina Vassiljeva was born in 1979. She received her M.Sc degree in Computer and Systems Engineering from Tallinn University of Technology in 2003 and her PhD degree from in Information and Communication Technology from Tallinn University of Technology in 2012. Currently, she is an Associate Professor at the Department of Computer Systems in Tallinn University of Technology. Her main research interests are virtual reality, artificial intelligence, and knowledge based control.