OPEN ACCESS Freely available online
R. L. Pedersen, L. Picinali, N. Kajs and F. Patou,
"Virtual-Reality-Based Research in Hearing Science: A Platforming Approach"
J. Audio Eng. Soc., vol. 71, no. 6, pp. 374–389, (2023 June).
DOI: https://doi.org/10.17743/jaes.2022.0083

Virtual-Reality-Based Research in Hearing Science: A Platforming Approach

RASMUS LUNDBY PEDERSEN,[1] **LORENZO PICINALI,**[2] *AES Member*, **NYNNE KAJS,**[1] **AND**
(raun@demant.com) (l.picinali@imperial.ac.uk) (cocokajs@mail.dk)
FRANÇOIS PATOU[1]
(francois.patou@gmail.com)

[1] *Oticon Medical, Research & Technology, Smørum, Denmark*
[2] *Dyson School of Design Engineering, Imperial College London, United Kingdom*

The lack of ecological validity in clinical assessment, as well as the challenge of investigating multimodal sensory processing, remain key challenges in hearing science. Virtual Reality (VR) can support hearing research in these domains by combining experimental control with situational realism. However, the development of VR-based experiments is traditionally highly resource demanding, which places a significant entry barrier for basic and clinical researchers looking to embrace VR as the research tool of choice. The Oticon Medical Virtual Reality (OMVR) experiment platform fast-tracks the creation or adaptation of hearing research experiment templates to be used to explore areas such as binaural spatial hearing, multimodal sensory integration, cognitive hearing behavioral strategies, auditory-visual training, etc. In this paper, the OMVR's functionalities, architecture, and key elements of implementation are presented, important performance indicators are characterized, and a use-case perceptual evaluation is presented.

0 INTRODUCTION

Advances in hearing science and related applications are more and more dependent on the vertical integration of scientific insights, from the understanding of lower-level perceptual phenomena to the exploration of higher-level sensory processing, multimodal sensory integration, and other top-down mechanisms involved in making sense of the acoustic world. Hearing scientists also increasingly insist on the importance of bringing more ecological validity to hearing research (i.e., to validate our theories in more lifelike scenarios) if we are to build more relevant contraptions and design more effective interventions to help restore or enhance hearing. One example of a field where these concerns have been highlighted recently is that of Cochlear Implant research.

Cochlear Implants (CI) are the most successful example of neural prosthesis [1] with more than 700,000 registered devices having been implanted worldwide so far [2]. Despite undeniable success, there are still significant limitations that have to be tackled. Examples are high variability in outcome in speech intelligibility [3–6], especially in complex environments, music appreciation [7], as well as fatigue and its adverse effects [8].

Systemic issues such as lack/inconsistency in rehabilitation, the role of nonsensory cognitive aspects of hearing, as well as the behavioral peculiarities of CI users in their approach to the auditory modality, remain elusive problems without effective solutions. These challenges/problems have at least one element in common: the lack of understanding of some of the putative mechanisms underlying the poor performances (or failure) of CI therapy to meet up current expectations. Investigating these matters represent a significant scientific challenge overspanning the exploration and interpretation of behavior and cognition, using both subjective and objective assessments, all ideally in more ecologically valid settings.

The design of experimental paradigms capable of integrating these elements is complex and challenging [9]. Although tools such as mobile apps can help gather invaluable real-life acoustic behavioral insights, they do not allow the control of experimental conditions that are generally required to explore systematically hearing-related questions. Task-evoked, controlled experimental paradigms remain key to disentangling cause–effect mechanisms.

Virtual Reality (VR) offers a venue to conciliate controlled experimental testing with ecological validity. VR-based research has already been proposed in other domains of science [10]. It can, however, be complex and time-consuming to set up and orchestrate, with multiple input/outputs requiring synchronization or with the need to design specific virtual environments. Besides, VR-based hearing science experimentation requires not only the capability to design and render visual scenes but also to render virtual *acoustics* scenes [11] synchronously and with high fidelity. Platforming has been proposed as a way to streamline experiment design including using VR [10, 9], thereby lowering the barrier to entry to VR-based experimentation. Platforming has, however, not been available for VR-based hearing research until now.

In this paper, the authors present the first VR-enabled research platform aiming to facilitate the design and execution of experimental hearing research studies involving the exploration of cognition behavior, multimodal sensory integration, real-life listening strategies, spatial hearing, and other relevant aspects of the emerging field of cognitive hearing science.

This platforms allows the rapid design of custom VR-based hearing research experiments through the customization of prespecified experimental templates. It allows the interlacing of preexisting or custom visual scenes and objects with acoustics objects rendered spatially over headphones or over an arbitrary configuration of loudspeakers, which makes the system suitable for studies involving ear-worn devices. A consumer-grade VR head-mounted display (HMD) provides head positional tracking and eye-tracking, facilitating the acquisition of head movement, gaze, and pupil dilation data in ecologically valid simulated conditions while user input data can be recorded synchronously with the other input streams. The key features of this platform are introduced, along with behavioral validation data using a virtual reproduction of a Spatial Speech in Noise (SSiN) test introduced in [12]. Finally, key performance metrics, including system latency, are presented. The Oticon Medical Virtual Reality (OMVR) platform is not publicly available, but it is being shared with external research partners outside of Oticon Medical and options to open the tool up for the hearing research community further are currently being investigated.

1 BACKGROUND

VR provides a dynamic immersive audiovisual experience that can be leveraged to simulate realistic environments while maintaining a high degree of control over experimental conditions. Early uses of VR for hearing research include the work of Cubick and Dau [13] who proposed a comparison of real-life against virtual acoustic evaluations delivered over a multiloudspeaker system. Normal hearing subjects were assessed looking at Speech Reception Threshold (SRT) performances using a hearing aid (HA) beamformer, comparing results in a real versus virtual environment. The results showed that VR-based hearing tests could indeed be a valuable tool for assessing HA benefits. Mansour et al [14] recently reported results from a standard headphone-based speech audiometry test compared with the same assessment carried out in virtual acoustics conditions using a 3D loudspeaker-based system.

Hohmann and colleagues [15] had carried out a similar study looking at HA performances in laboratory conditions versus more realistic conditions based on virtual acoustic environments, aiming at identifying the identify the factors affecting HA performance across various tested environments. For certain metrics (e.g., speech intelligibility) significant differences could be found between lab tests and virtual acoustic tests. These differences were very similar to the ones found when comparing lab tests with real life tests, underlying the need for novel, more lifelike hearing assessment methods, whether real or virtual.

Beside speech intelligibility testing, VR has understandably been coveted to assess and train sound source localization skills. In a recent study, Ahrens et al. [16] assessed audiovisual perception in realistic environments looking specifically at the effect of the HMD and of different visual information on loudspeaker-based sound localization performances. The HMD condition resulted in higher localization error compared with the head-free condition (i.e., looking at the actual loudspeakers in the room), but this was negligible when virtual loudspeakers matching the position of their physical counterpart in the room were displayed within the virtual environment. Visual information of hand location and room dimensions allowed for overall better localization performances.

Steadman and colleagues [17] looked at the short-term effects of sound localization training, showing how procedural and perceptual learning can happen through short (12 minutes) localization training sessions using nonindividual Head-Related Transfer Functions (HRTF) within a VR context. This study confirmed earlier results suggesting that speech perception and production can improve after VR-based audiovisual training [18].

Multimodal (audiovisual) integration has also been subject to VR-based research recently. Visual cues in audio-based tasks within a VR context have shown to be beneficial to intelligibility. Devesse et al., for instance [19], found that the visual presence of the target speaker resulted in an SRT improvement of 1.5 to 2 dB. In a study by Hendrikse and colleagues [20], audio-only rendering was compared with audiovisual presentation within VR-based speech intelligibility and sound localization tasks, looking specifically at changes in head and eye movement behaviors. Movement behavior, task performance, and overall perception were all influenced by the presentation of visual cues.

VR has been sparsely (but robustly) used with CI users from its early days, perhaps because there is substantial evidence that CI users are better multisensory integrators compared with normal hearing individuals and that the appropriate simulation of both visual and acoustic cues when performing VR-based training and assessment can better relate to the real life challenges of CI users [21]. More recently, Briggs et al. [22] outlined VR as a potential vector to drive further development in the field of CI.

Sechler et al. [23] designed and implemented a custom VR tool for measuring sound localization performances with bilateral CI users. In this case, localization performances using VR were lower when compared with previous CI localization studies, possibly explained by the added complexity of the virtual task compared with the standard method. Majdak and colleagues [24] carried out an HRTF-

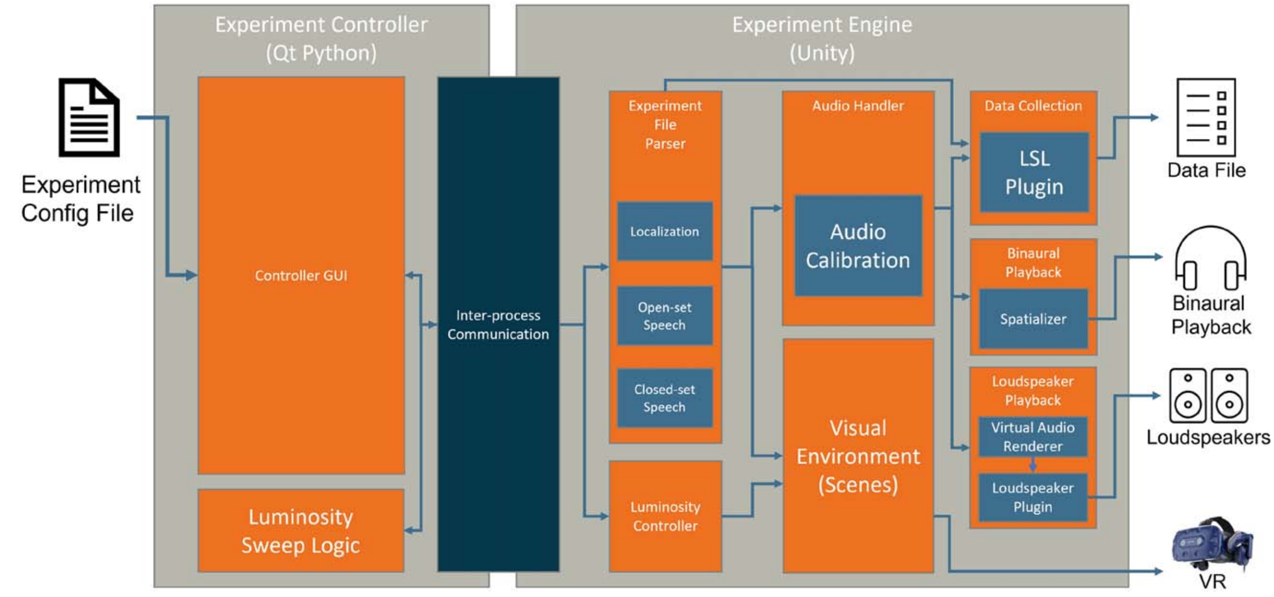

Fig. 1. System block diagram of the OMVR platform. The OMVR platform is split in two software processes, more specifically an OMVR Controller, used by the researcher to initiate, stop, and control the flow of the experiment, and an OMVR Engine, which parses the user-defined ECF and renders the visual and acoustics elements of the VR experience to the subject during the test.

based localization study to assess general sound localization performances of CI users, specifically those related with sound sources elevation. Significantly larger localization errors were found for CI listeners compared with normal-hearing individuals, both in terms of lateralization and elevation. Nevertheless, Majdak et al. report that both interaural and spectral cues, as well as head movements, are significant contributors to sound localisation for CI users.

Virtual sound auralization or spatialization engines have also been a focus of their own in VR-based auditory assessment/training such as in the 3D Tune-In project [25]. Eastgate et al. report how VR and videogames can be used to support individuals with HAs. They propose a series of immersive simulations aimed to demonstrate the difficulties experienced by hearing impaired people in performing everyday tasks, such as listening to a conversation in a noisy restaurant. In a similar manner, Pausch and colleagues [26, 27] developed a binaural real-time auralization system designed for HAs and hearing loss (HL) research, and more recently, the BEARS (Both Ears) project began looking at developing a package of VR video games to train spatial hearing in young people (8–16 years) with bilateral cochlear implants [28]. As part of the project, a virtual acoustics version of the SSiN [29] test was developed and validated [12].

These various works have all reasserted and demonstrated the potential of VR-based, controlled (albeit more ecologically valid) hearing research. A significant caveat to all the experimental paradigms underlying the aforementioned research is that they needed to be custom made. The generalization of VR-based experimentation in hearing science requires the simplification and the streamlining of experimental design as well as the means to explore populations of normal-hearing and hearing-impaired populations alike. The OMVR platform provides such capability.

2 OTICON MEDICAL VIRTUAL REALITY

The OMVR platform implements two software processes, namely the OMVR Controller and the OMVR Engine (Fig. 1). The OMVR Controller and OMVR Engine are two separate software processes that can be configured to run on the same machine or on two network-connected computers. The processes continuously communicate with each other using the interprocess communication protocol ZeroMQ, which is a lightweight messaging kernel build on socket interfaces [30].

The network details of the connection are configured using the OMVR Controller GUI. Practically, an operator supervising an experiment primarily interacts with the OMVR Controller as it is used to load new Experiment Configuration Files (ECF) and control various aspects of the OMVR experiment run-time elements, i.e., system calibrations and network settings, while the OMVR Engine embeds the background processes that run on a high-end performance PC. The OMVR Engine is connected to the VR hardware equipment of choice. The two software processes continuously communicate with each other using the ZeroMQ [30] interprocess communication protocol, and they can easily be configured to connect across a network while running on separate computers.

2.1 Experiment Configuration File

The ECF is a central element of our platform-based design. The ECF is a human-readable, structured text file that specifies the experiment parameters of a given preexisting experiment protocol *template*. These parameters include, for instance, the virtual scene the test subject will be exposed to during the test and the definitions of one or more unique trials that must be completed by the test subject. The

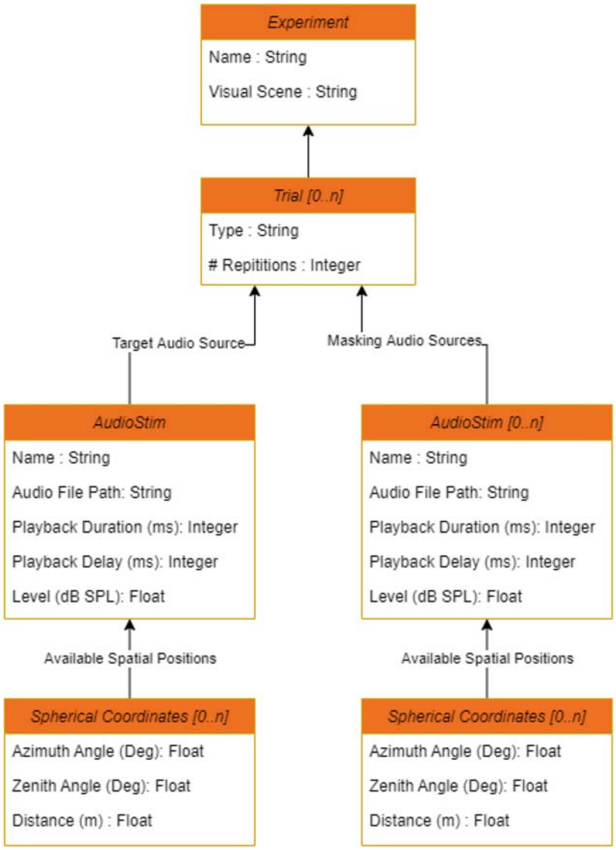

Fig. 2. Diagram visualizing the simplified hierarchical structure of the ECF. The ECF is implemented as a JSON-structured text file.

ECF is loaded by the operator at the beginning of an experiment using the OMVR controller's graphical user interface (GUI) and is passed to the OMVR Engine process where it is parsed to initiate the run-time procedure. The OMVR Engine is responsible for synchronous stimuli presentation, i.e., spawning visual and sound objects in the environment. The ECF is structured in a JSON format with nested components that are parsed and read by the OMVR Engine process, the hierarchical structure of the ECF illustrated on Fig. 2, where the *Experiment* type is observed as the top level element of the structure.

The *Experiment* type contains a number of *Trial* types, which contain the parameters for one or more unique trials that can be repeated an arbitrary number of times. The *Trial* type contains the definition of a single target audio source of a given trial as well as an arbitrary number of masking sources that will be spawned around the test subject. Each audio source is defined in the ECF as an *AudioStim* component, which is characterized by a number of parameters, including the path to the audio file that will be played by the audio source, as well as the playback duration, delay, and level.

The playback duration can be configured to play the audio for a set amount of time, for the full extent of the loaded audio file, or in an infinitely repeating loop that is stopped once the given trial is finished. The playback delay defines the time from when the trial begins until the audio starts playing and the playback level defines the targeted sound pressure level that the audio will play at, assuming that the audio level calibration process of the OMVR platform has been completed.

2.2 OMVR Controller

The primary purpose of the OMVR Experiment Controller is to load an ECF at the beginning of the execution of an experiment, control the flow of the running experiment and initialize various calibration and setup processes. The Experiment Controller is implemented in the Python programming language, supported by the Qt software development framework [31].

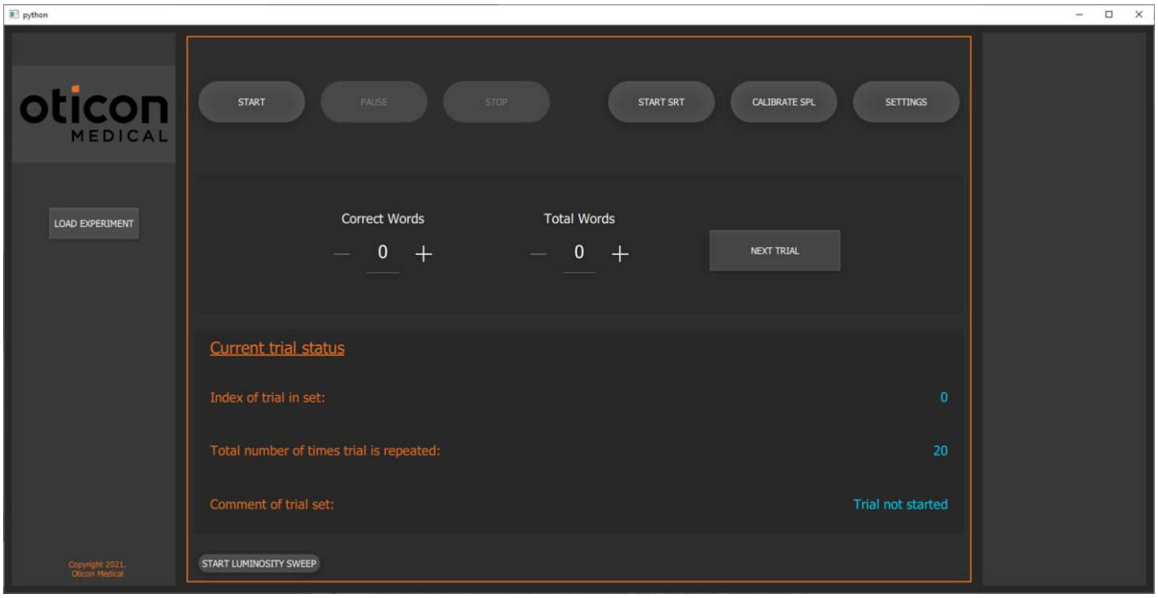

Fig. 3. The OMVR Controller GUI main window. The GUI elements are used to change platform settings such as network information and audio playback method, load ECFs, control the flow of the experiment, and provide run-time information to the researcher during experiment execution.

The GUI consists of a main window where the user, typically the researcher/operator, can find controls to load an ECF, configure the system settings and start or stop an active loaded experiment. System settings include the network information of the PC running the OMVR Engine (Defaults to 'localhost' if the OMVR Engine and OMVR Controller are running on the same machine) as well as the targeted audio output format, as described in SEC. 2.5. Additionally, the GUI is used to initiate and perform system calibration procedures, such as audio level calibration or a pupillometry sweep sequence for an active experiment scene as described in SEC. 2.6.

2.3 OMVR Engine

The purpose of the OMVR Engine is mainly to parse the loaded ECF and render both the visual environments and acoustic elements of a VR experiment. The OMVR Engine is implemented as an executable process using the real-time development engine Unity, developed by Unity Technologies [32]. Unity is traditionally used as a platform for game development and VR applications, which makes it a powerful tool in the creation of software applications for experimental research projects in human behavior [16, 33]. Like most game engines, Unity includes tools for audio reproduction and also offers the option to include custom software plugins that augment or replace the built-in audio processing mechanisms in charge of rendering the virtual audio environment. Details are provided in SEC. 2.5.

Unity offers mechanisms that can be used to control the timing of events in what are internally defined as *scenes*, and these are used to maintain the order and timing of events that are initiated in an experiment trial. The execution of an experiment trial depends on the type of the active trial, or *template*, although all experiment trials share the following basic event-blocks:

1. Load acoustic and visual objects into memory.
2. Play the loaded audio and display the visual objects per the timings defined for the given trial.
3. Wait for subject task completion.
4. Gather task completion data and push it to a Lab Streaming Layer (LSL) stream (See SEC. 2.4).
5. Reset the scene.
6. Repeat the same trial or continue to the next.

The different built-in *templates* are mostly distinguished in the way they handle the timing of the audio playback and the subject tasks included in the given trial, and the current available *templates* are detailed in SEC. 2.4. The handling of trial performance data, including data format structuring and sample time-synchronization, is described in SEC. 2.4.

2.4 Experiment Templates

The purpose of experiment *templates* are to distinguish experiment paradigms that differs significantly in terms of stimuli timing requirements, subject task, and relevant data collection streams. A few versatile experiment *templates* have been created to cover a wide range of needs while staying specific to a set of requirements. These *templates* include the following:

- *Localization*: The *Localization template* consists of a sequence of events where a target sounds source is spawned somewhere in the virtual environment, alongside an arbitrary number of masking sound sources, or none at all. Playback of the various audio sources starts per the exact timing details included in the ECF. Then, the process waits for the subject to point somewhere in the room with a virtual laser pointer and indicate where they believed the target sound source originated from. The coordinates of the believed location are saved and passed to an LSL stream alongside the true location, the time elapsed from playback to the subject providing their response, and additional content- and timing-information about the audio sources.
- *Open-Set Speech Reception and Source Localization*: The *Open-Set Speech Reception and Source Localization template* presents a number of sound sources to the subject, typically playing a speech file, and the researcher is prompted to word-score the subject's response, which will trigger the start of the next trial. The *template* can also be configured to prompt the subject to point out in space where they thought the sound source originated from, or even combine the two tasks.
- *Closed-Set Speech and Spatial Discrimination*: The *Closed-Set Speech and Spatial Discrimination template* is a VR-based implementation of an existing non-VR study, where two audio sources are played from somewhere in the virtual environment, with a small azimuth angular displacement [29]. The task for the user is to select the words they heard from a closed-set selection that are displayed in front of them, on a virtual panel, as well as whether they thought the second target sound came from the left or the right, relative to the first. This *template* was used as part of the platform validation efforts, and the full experiment is described in SEC. 3.

Data on the task performance of the subject during trial execution and information about their physical movements, such as head tracking, are gathered throughout the lifetime of an experiment using the LSL system. LSL is an open-source tool for collecting measurement time-series in research experiments and handles networking, time-synchronization, and data structuring and is supported by a large number of third-party software applications. The currently supported tools are listed in the *Apps* section of LSL's github repository, such as PupilLabs for pupillometry measurements and gaze tracking, and g.Tec for EEG measurements [34–36].

All data samples are timestamped before they are pushed to the so-called *streams*, and LSL synchronizes these timestamps across *streams* when they are processed for storage or statistical analysis. The data sample time-stamping is typically done automatically by the third-party applications, but it is also possible to use the LSL library to timestamp and push data samples to custom *streams*, which the OMVR does for the subjects' trial performance data and VR head-tracking. Data collected using the LSL protocol include the following:

- *Trial Performance Data*: Subject performance and feedback data are gathered at the end of each trial and pushed to an LSL stream unique for the given experiment trial *template*. Stimulus and subject feedback events such as audio playback initiation and subject selections are timestamped manually using the LSL library and structured in a custom format, which is then pushed to the LSL stream as a single sample at the end of the trial. Thus, the researcher can analyze the events of each trial and match the timestamps to additional streams such as the LSL stream of a pupillometry-measurement process generated by the PupilLabs software tools.
- *Head Tracking*: The OMVR Engine collects and time-stamps head-tracking data, such as head positioning and rotation, continuously throughout trial execution and the samples are pushed to a custom LSL stream for later processing.
- *Eye-Gaze and Pupillometry*: PupilLabs add-ons can be inserted in an HTC Vive or Vive Pro HMD and serve for the collection of gaze-tracking and pupillometry data [37]. Additionally, open-source LSL plugins have been created for the PupilLabs software tools making it trivial to integrate with the OMVR platform for studies including pupillometry measures.
- *Third-Party Data Collection*: The stand-alone, open-Source LabRecorder software application can be used for gathering LSL stream *outlets* and saving the data on the file-system [38]. The LabRecorder is configured by the researcher to collect data from one or more streams and the application is responsible for generating the data file in an Extensible Data Format (XDF [39]) file format, which is a structured file that can be opened in Matlab or Python using third-party tools.

2.5 Visual and Acoustic Stimulus

The OMVR platform includes a set of virtual environments that can be selected during the design of OMVR experiments. Currently, a few virtual scenes are available for use with the OMVR platform. This includes the following:

- *Sound Studio*: The *sound studio* scene seen on Fig. 4(a) is a virtual sound studio similar to what one would normally expect to find at universities and in-

Fig. 4. Demonstration of a subset of virtual environments that is available in OMVR platform. Each environment is developed or selected to either provide an immersive experience, a visually neutral background for studies using pupillometry, or to mimic real physical environments, such as the replication of one of the sound studio located at Oticon Medical headquarters. (a) Virtual sound studio scene, custom made to look like the Germany sound studio in Demant's facilities at Kongebakken, Denmark. (b) Virtual classroom scene [58]. (c) Virtual luxury apartment [59]. (d) Empty neutral grayscale room for experiments with minimal amount of visual distractions.

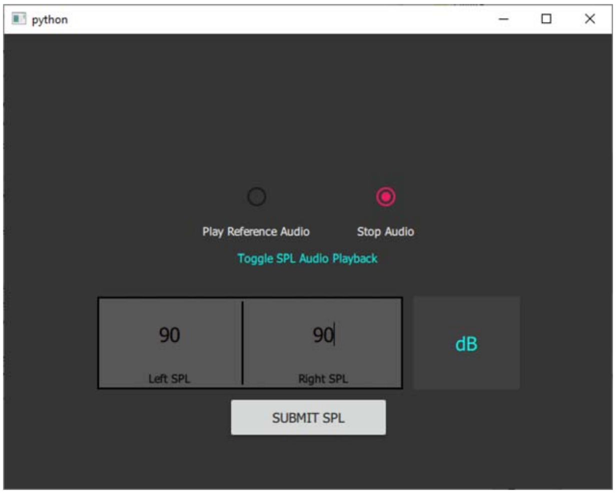

Fig. 5. Screenshot of the OMVR Controller GUI window used to perform audio playback level calibration. The radio buttons are used to toggle the audio playback, for which the researcher measures the resulting SPL with a sound pressure meter and inserts the measured values in the GUI before submitting the measurements.

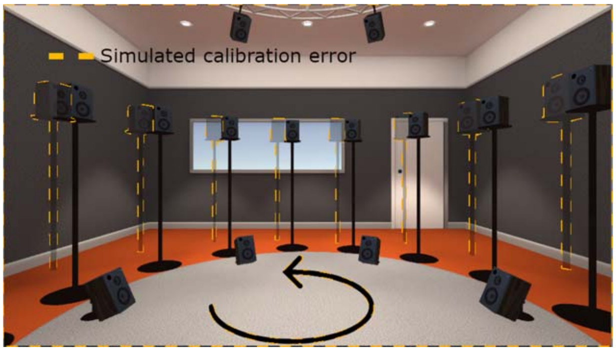

Fig. 6. Visualization of the effect of rotational calibration error, when the integrated SteamVR calibration process is done improperly. Positional and rotational alignment processes of the OMVR platform can help mitigate the effects of calibration errors generated through human error.

dustrial complexes where experimental research in hearing science in carried out.

- *Classroom*: The *classroom* scene seen on Fig. 4(b) resembles an ordinary middle-school classroom, which may be suitable for experiments including young subjects as it can represent a traditionally noisy environment. The room is empty by default but can be populated with desks and chairs as needed.
- *Apartment*: The *apartment* scene seen on Fig. 4(c) is meant as a visually immersive environment that can be used to mimic daily life at home in comfortable settings.
- *Grey Room*: The OMVR platform also offers an empty neutral gray room seen on Fig. 4(d) that can be used for experiments that requires minimal visual distractions and studies that include sensitive pupillometry measures.

The desired virtual scene is specified in the ECF. More visual scenes will be added and integrated over time.

The OMVR platform supports two main methods of audio playback: binaural playback through headphones with relative audio source positional tracking and loudspeaker playback through an arbitrary number of loudspeakers in a static audio environment.

2.5.1 Binaural Headphone Playback

Targeting normal-hearing test subjects, the binaural audio technique allows to render immersive audio through a pair of headphones, emulating the acoustic cues associated with spatial hearing, such as Interaural Time-Differences (ITDs), Interaural Level-Differences (ILDs), and monoaural spectral cues. These localization cues are stored as HRTF, formatted in the Spatially Oriented Format for Acoustics (SOFA) format, which are loaded by the OMVR Engine during run-time to deliver convincing and immersive spatial audio to the test subject [40]. The spatial cues are continuously updated based on the relative position and rotation between the listener's head and the individual audio sources, given VR engine's capability of tracking head-movements. This means that the test subjects can move around in the virtual environment while the audio sources are locked to the surrounding space.

The Unity engine includes a built-in spatialization plugin, but it is possible to include third-party software plugins that may expand upon the processing capabilities of the default plugin, with one example being the 3DTune-In Unity plugin [41, 42]. The 3DTune-In plugin proposes additional features such as HL simulation, HA simulation, and the ability to dynamically load custom HRTF files.

2.5.2 Loudspeaker Array Playback

The OMVR platform is meant to be used with normal-hearing test subjects as well as hearing device users. However, members of the latter group are typically unable to wear stereo headphones, which are required to use the platform's binaural playback capabilities. A custom audio playback plugin has been implemented for multichannel loudspeaker setups, enabling the use of arbitrary loudspeaker configurations in terms of channel count and physical speaker placement. A desired loudspeaker configuration is selected through the OMVR Controller GUI before an experiment is started, and the researcher has an option to specify to a dynamic auralization rendering option or to include a prerendered and auralized *impulse response* audio file with the audio input:

- *Dynamic Auralization*: Each mono-channel audio source instance described in the ECF is associated with positional information, which contains the desired spatial coordinates of the audio source in the virtual space. The OMVR Engine automatically auralizes the audio source per the positional information and the physical locations of the loudspeakers in the selected speaker configuration, generating a multichannel sound file that is played back by the OMVR

Engine through an ASIO audio interface. The TASCAR software is used as Virtual Audio Renderer to render the auralized audio [43].

- *Pre-rendered Impulse Response*: Each mono-channel audio source instance described in the ECF is associated with an *impulse response* audio file that has been auralized for the selected loudspeaker configuration. The OMVR engine convolutes the input audio with the impulse response (IR) in real time, generating the multichannel output audio buffers that are played through the connected ASIO audio interface. In addition to the auralization and speaker channel mapping, this also allows the researcher to include custom IRs that contain reverberations and room acoustics.

2.6 System Calibrations and Features

The OMVR platform includes functionalities for calibrating playback audio levels and virtual environment luminosity for studies that includes pupillometry measurements.

- Audio Levels: It is possible to define the audio levels of individual audio sources in the ECF. The OMVR platform includes a procedure that is used to calibrate the audio output level of a given playback configuration. The calibration process is initiated through the OMVR controller, which instructs the OMVR Engine to start playing a sound source with a known decibels relative to full scale (dBFS) value through the connected audio interface. The researcher uses a sound pressure level meter to measure the resulting sound pressure at the point where the subject is expected to receive the sound and enters the measured value in dB in the OMVR controller's GUI. This value is saved in the file system and on the PC's memory, and the output playback level of loaded audio source from this point is adjusted in accordance with an associated known dBFS value that is defined together with the input audio files. The audio level calibration procedure also supports stereo channel balancing for binaural audio playback through headphones, in case the playback hardware channels are imbalanced.
- Luminosity Sweep: A benefit of using VR for research in human behavior is the option to utilize the measure of pupil dilation with hardware such as the PupilLabs VR Inserts [35]. Research has shown that dynamic change in pupil dilation can be used as an indicator for change in cognitive load during mentally demanding tasks [44, 45]. However, the pupil size and pupil dilation of individuals are typically sensitive to light exposure, which means there is a risk that a test subject's pupil size becomes saturated or has a baseline size that is outside the linear region of the *tanh* fitting models of pupil dilation as a function of light exposure, or luminosity [46].

  Most VR engines intrinsically includes tools to control for lighting conditions, either in the form of placing semitransparent screens in front of subjects' eyes to limit the light they receive from their surroundings or by adjusting the light exposure of the virtual cameras in the experiment scenes. The OMVR platform utilizes these tools and implements an adaptive procedure, a *luminosity sweep*, that can be used to sweep the light exposure of an experiment scene, record the resulting pupil dilation, and fit the response to a mathematical model which is used to find the point of light exposure where the subject's baseline pupil diameter is at 50% of its full range of size change. Currently, the luminosity sweep procedure is only supported by the PupilLabs hardware described in SEC. 2.4.

Positional and rotational alignment between the VR-tracked virtual environments and the real-world is critical for the success of experiments carried out using loudspeakers as audio output. Virtual-to-Real misalignment can result is audiovisual positional incongruency, meaning that visual anchors in the subject's virtual environment may not match with loudspeaker positions in the real world.

The SteamVR-supported VR hardware should always be calibrated per the instruction specified by the manufacturer, and although the integrated calibration procedure provided by Valve is technically accurate, there is a significant chance that human error during the process may result in a misalignment between the VR-tracked virtual environment and the intended references in the real world [47]. Therefore, the OMVR platform implements an automatic realignment procedure every time a new experiment is loaded. The OMVR platform tracks the location of a independent SteamVR Tracker unit that is placed in a known location in the real-world testing location and the OMVR platform compares the internal positional data of the tracker with the correct manually input tracker position and realigns the virtual environment according to the error.

2.7 System Requirements

The main performance requirements comes from the graphics-intensive visual rendering of the VR environments in the OMVR Engine that requires a high-end graphical processing unit (GPU) in order to generate a visually immersive environment. The PC machines listed in Table 1 have successfully been used to develop, test, and validate the performances of OMVR.

Currently, the OMVR Controller and OMVR Engine must be installed on a Windows 10 PC. The OMVR platform has been built using cross-platform libraries, but certain internal software processes relies on Windows 10 architecture so the compatibility is limited to this operating system (OS). Cross-platform support may eventually be implemented.

The OMVR platform supports VR hardware that utilizes the SteamVR engine for tracking and rendering [47]. Most modern tethered VR equipment is supported by SteamVR.

Table 1. Listed computers that has successfully been configured to run the OMVR Engine.

PC	Type	CPU	GPU	RAM
Alienware M15	Laptop	Intel i7-9750H	GeForce RTX 2070 Mobile	16 GB
MSI GS65 Stealth	Laptop	Intel i7 8750H	GeForce GTX 1070 Mobile	32 GB
Custom	Desktop	Intel i7-9700K	GeForce RTX 2070	16 GB

Currently, the following VR hardware has been tested with the OMVR platform:

- HTC Vive (HTC Corporation, New Taipei City, Taiwan, 2016)
- HTC Vice Pro (HTC Corporation, New Taipei City, Taiwan, 2018)
- Meta Quest 2 (Tethered) (Meta Platforms, Inc., Menlo Park, California, USA, 2020)

The OMVR platform supports traditional stereo audio interfaces for binaural audio playback and ASIO interfaces for loudspeaker playback. Using binaural audio playback, the user can chose to connect a pair of stereo headphones directly to the internal sound card PC running the OMVR Engine application or through an external audio interface using an optical or USB connection, as long as the stereo audio interface can be selected as a default playback device in the Windows 10 OS. For loudspeaker playback, the OMVR Engine PC must be connected to the loudspeakers through an audio interface that supports the ASIO sound card driver protocol [48].

3 PLATFORM VALIDATION

The behavior of OMVR was assessed through a comparative study with a loudspeaker-based dual-task hearing test developed by Bizley et al., by replicating their ”Experiment 2” paradigm in a VR environment [29]. The purpose of the original study was to create an experiment paradigm for simultaneous assessment of speech identification and spatial discrimination using a small loudspeaker array for audio playback.

The main differences between this VR-implemented version and the original study were the introduction of VR visuals, the number of recruited participants (10 subjects were tested by Bizley et al.), and finally, the number of trials in the test battery. Although the original study had 16 trials per speaker position, the authors opted to reduce this to four trials, in order to make the session shorter overall while ensuring that each of the four word-groups from the speech set were represented at each position once.

3.1 Participants

Twenty normal-hearing adults between the ages of 20 and 59 participated in the study. All participants had normal-hearing thresholds as assessed by pure tone audiometry or self-reporting and had no reported neurological disorders. All participants had good or corrected vision on both eyes. Because of technical difficulties, the data from two participants had to be discarded.

Fig. 7. Screenshot of the virtual environment the test participants would be presented to during the *Simultaneous Assessment of Speech Identification and Spatial Discrimination* experiment.

3.2 Test Setup

The participants were seated in the middle of a sound-treated sound studio with all potential unwanted noise sources removed from the room, because they may disrupt the results of the experiment, even with the use of headphones for audio playback. Ideally, an anechoic chamber is preferred, but because the authors did not have one available at the time of testing, a sound-treated room with minimal reverberations was deemed adequate for the purpose of the experiment.

The visuals were presented on a VR HMD (HTC Vive Pro system). The virtual scene was a visual approximation of a traditional classroom, presented in Fig. 4(b), where the participants were placed in the middle of the room with chairs representing the possible speaker locations positioned around and facing the participant. The virtual setup including the virtual button panel for trial task feedback can be seen on Fig. 7.

The stereo audio was spatialized using a generic KEMAR HRTF [49], which was generated in an anechoic environment and presented binaurally from a pair of headphones (Sennheiser HD 280 Pro, Wedermark, Germany). No reverberation effects were added to the audio playback because the original study was carried out in an anechoic chamber. All audio was presented virtually at head height, and the HRTF used in by Corbetto et al. was utilized in this study, just as the same spatializer engine was used to apply it [12, 42].

3.3 Stimuli

All stimuli were generated and presented at a sampling frequency of 44.1 kHz. Stimuli were monosyllabic word tokens from the Chear Auditory Perception Test spoken by a single female British English talker [50]. The stimuli presentation was identical to the one presented in the comparison study, except for the fact that the participant would not indicate the feedback on a physical touch panel but rather on a virtual control GUI using a VR controller as a virtual laser pointer [29, 12]. Details about the audio playback timing and the general design of the auditory stimulus is described in the original paper [12].

The level of the background babble (several incoherent sources) was set to 52 dB SPL A-weighted, and the target speech level was determined by an adaptive Speech Recognition Threshold (SRT) test to determine an individualized target-babble Signal-to-Noise Ratio (SNR) at which 50% of the spoken words are understood. This test was implemented as a VR-version of the adaptive procedure that was developed as part of the Oticon Medical Experiment Building Platform by Sulas et al. (OMEXP) [9].

The procedure was adapted to use the speech material from the study by Bizley et al., *Simultaneous Assessment of Speech Identification and Spatial Discrimination* (SSIN) [29]. An additional 6 dB was added to the SNR generated from the SRTs to offset the potential increase in difficulty from using binaural audio playback and including VR, as it was also done by Corbetto et al. in a study that sought to replicate the original SSIN study using spatialized binaural audio but without VR [12]. The test participants were instructed to look straight ahead at a reference position in the virtual environment at the beginning of each trial and to keep their posture while the trial's audio was playing.

3.4 Validation Results

The subjects' spatial and word discrimination scores are plotted on Fig. 8(a) and 8(b), respectively. Two subjects' performance scores were discarded from the spatial discrimination data set due to being statistical outliers.

The speech identification scores seen on Fig. 8(b) show a slightly irregular U-curve, indicating the speech identification performance is higher when speech is coming from more lateral positions. An inverted U-shape can be observed for the spatial discrimination scores on Fig. 8(b), indicating relative spatial discrimination performance is higher when the audio source is located closer to the front of the test participant.

A Student's t test was performed for all speaker positions to compare the subject performances between the results found in this study and the results obtained by Bizley et al. in the original study. The resulting statistical significance levels are displayed above each positional angle on Fig. 8(a) and 8(b). Both performance curves are seen to be similar to the results obtained by Bizley et al. in terms of curve shape.

An expected bias is observed for the overall performance curve of the word discrimination scores because of the higher SNR employed in our study. All but the performance scores at 60° and 90° show a statistically significant difference between the results of this study and the findings by Bizley et al., likely due to the high variance of the scores, which seems to be a result of the low number of repetitions for each angle.

Little to no difference in bias is observed for the spatial discrimination between our data and the results obtained by Bizley et al. seeing that there are no significant differences between the angle-dependent performance scores. This is as expected, based on the findings by Corbetto et al. who showed that variations in SNR from a baseline of a 50% SRT SNR had little effect on sound source location discrimination [12].

4 PERFORMANCE CHARACTERIZATION

Key OMVR performances were characterized to ensure the accuracy and precision of the system's components involved in generating or recording audiovisual stimuli. This includes the sound levels of the platform's audio playback but most importantly, addresses the timing of stimuli onset and offset, combined with the timestamping of trial events and user performance. Therefore, this section seeks to highlight the performance characteristics of the temporal and spatial audio playback components. The spatial audio playback performance relates to the virtual-to-real world alignment that is performed for setups including loudspeaker playback, for which the virtual environment is visually shifted to align with a real-world reference space that is typically tied to the loudspeaker positional arrangement. The techniques used to characterize these performance indicators were inspired by similar efforts by Bridges et al. [51], Chénéchal et al. [52], and Tachibana et al. [53].

4.1 Audio Latency

A system with low audio playback timing accuracy and high precision can largely be compensated for during the postprocess efforts of an experimental study, but it requires that the latency is characterized beforehand. A system with high accuracy but low precision can result in audio onsets with unacceptable amounts of jitter that may render some biometric data collection modalities unsuitable for use with the given platform, namely electroencephalography, which sometimes requires accuracy and precision in the range of a couple of milliseconds. OMVR's audio playback timing has been assessed for both spatialized binaural audio over headphones and for spatialized loudspeaker playback.

A MAYA44 USB+ soundcard was used to output the audio and was configured so that its two first output channels were connected directly to its two first input channels such that the audio could be recorded back with minimal hardware latency [54]. For the evaluation of playback latency and onset delay, audio bursts of 300 ms in duration (1 kHz tones) were played repeatedly. The onset time and playback durations were controlled by the OMVR Engine software. The OMVR platform records the timestamp at which it attempts to start each pulse and pushes it to one LSL stream, *time-stream*, while the Open-Source LSL-supported software AudioCapture was used to record the audio samples from the MAYA44 USB+ sound-card and automatically record each audio sample's timestamp to push it to a second LSL stream, *audio-stream*, where the timestamps are synchronized [55].

The recorded audio from the *audio-stream* was split in segments, each segment matching the duration of the 300 ms audio-pulse plus a small buffer for the potential delay

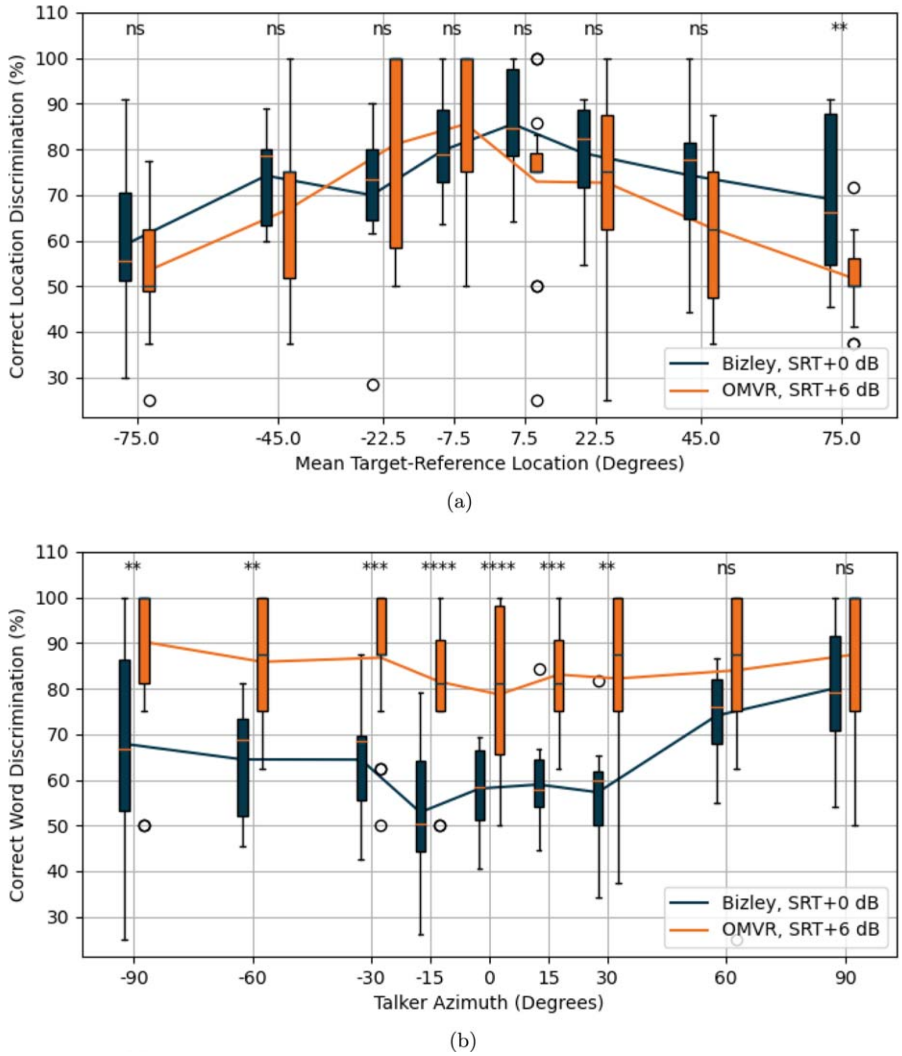

Fig. 8. Boxplots comparing the results of the study presented here to the results obtained by Bizley et. al. [29]. Line plots represents the subjects' mean performances for each position. The statistical significance of the difference between the two studies are plotted above each positions (Student's *t* test), where "ns" marks a nonsignificant difference between the two data sets. (a) SSiN-VR performance for relative localization, or spatial discrimination, for each mean target reference location (the subject would select a target speaker between two adjacent potential speakers). (b) SSiN-VR performance for word discrimination for each speaker position.

before the actual audio plays, and the start of each segment aligning with the onset-timestamps recorded in the LSL *time-stream*. A short time-window with zeroed-out audio-samples is inevitably expected at the beginning of each audio segment, between the first sample of the epoch and the sample where the audio is detected to start. This time-delay represents the difference between the believed time the audio was started and the actual time at which it was started.

- Binaural Spatialized Stereo Playback. Audio playback latency was investigated for audio spatialized by Unity and played through stereo headphones. The latency will partially be influenced by the computational efficiency of the Unity software implementing the experiment logic, the spatializer plugin, and the internal or external sound card that is configured to provide the analog audio output, including the audio buffer used for the spatialization signal processing. The spatializer plugin and sound card will likely be the main contributors to latency bias (accuracy) while Unity's software backend and experiment logic implementation is likely the primary influence to latency jitter (precision).

The 3D Tune-In Unity plugin was used to spatialize the audio during the test using a KEMAR HRTF configuration from the Sadie II database [49]. The spatializer plugins works by selecting an IR from a list of IRs in a given HRTF, based on the relative angle between the spatial location of an audio source and the orientation of the user's head, which is then convoluted with an audio buffer from the same audio source and passed to the stereo audio output of the Unity process, all in near real-time. This process may introduce additional latency to the audio output in the form of a short bias as well as a slight vary-

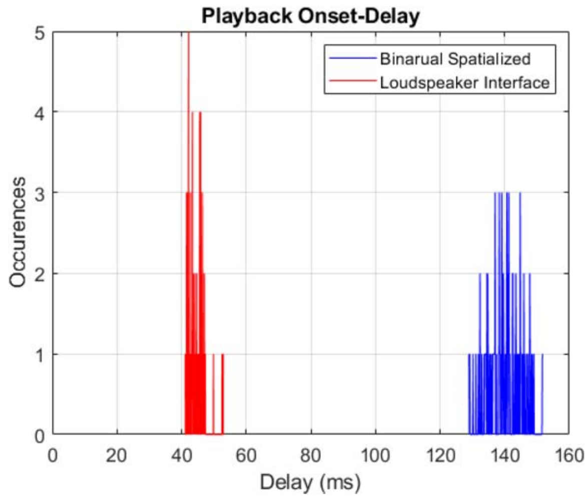

Fig. 9. Histogram of the pulse onset-delays for audio played back binaurally and through an ASIO-based interface.

Table 2. The average values and standard deviations of the audio playback onset-delays and delta target durations, for the 3D-TuneIn Unity-based binaural stereo playback and the custom ASIO-based loudspeaker audio-plugin, respectively. Delta duration are the differences between the measured pulse durations and the target 300 ms pulse duration.

	Avg (ms)	Std (ms)
Onset-Delay		
Binural Spatialized Stereo	139.9	5
Loudspeaker ASIO Interface	44.5	2.2
Delta Duration		
Binaural Spatialized Stereo	−2.1	6.6
Loudspeaker ASIO Interface	−0.8	1.2

Avg: Average, Std: Standard Deviation.

ing delay, depending on the implementation of the plugin. These aspects are equally relevant for any audio duration inaccuracies and imprecision, because the Unity-based OMVR engine uses the integrated timing functions to track the duration of any playing audio, rather than letting the audio play for a specific amount of samples.

- Loudspeaker Array Audio Interface. The audio output buffer of the ASIO-based loudspeaker interface is fixed at 2,048 samples in order to accommodate intensive signal processing such as the multichannel convolution required for multispeaker playback and reverb IRs. At a default sampling-rate of 44.1 kHz, this guarantees a playback latency of at least 46 ms, in addition to any latency introduced by the OMVR engine. The precision of the onset-delay will largely be governed by the timing of the Unity-backed OMVR engine; however, the inaccuracy and imprecision of target playback duration will likely be negligible because the loudspeaker software interface tracks the duration of any playing audio in samples, rather than the timing approximation Unity-based binaural stereo playback.

4.2 Characterization Results

A histogram of the pulse onset-delay for audio played back binaurally and through the ASIO-based interface can be seen in Fig. 9. The average values and standard deviations of the onset-delays and playback duration deltas for the two audio output configurations can be seen in Table 2. The delta time-duration values presented in Table 2 represents the differences between the duration of recorded audio pulses and the ones that were played back. The duration is simply a measure of the time between the actual pulse onset and its offset in milliseconds.

Given the high playback latency observed for both spatialized binaural playback and the loudspeaker-interface in Fig. 9 and Table 2, care must be taken to ensure that the latency does not affect the results of the given trial, specifically in applications that include low-latency evoked response EEG paradigms. The variance of the playback latency is quite low for loudspeaker playback, meaning that it should be possible to compensate for the latency during post-processing of the trial data or in the timing details of the experiment configurations. However, future efforts should be focused on reducing the latency variations of the latency observed in spatialized binaural playback. The results obtained and illustrated here are obviously dependent on the 3D Tune-In spatializer and one would likely get different results if the OMVR platform was reconfigured to use another spatializer plugin.

As discussed in the publication of the OMEXP platform, typical modalities such as pupillometry measurements used in cognitive hearing science experiment paradigms can accept audio playback timing jitter of less than 10 ms, which attests to the OMVR platform's general usability for VR implementations of these paradigms with its sub-10-ms jitter illustrated in Table 2 [9]. However, the authors are not confident to generally recommend the OMVR platform for use in EEG studies, until sub-1-ms audio playback latency jitter has been achieved.

5 CONCLUSIONS

The OMVR experiment-building platform enables clinicians and researchers to design and implement VR-based hearing research experiments with significantly reduced time and effort igcompared with off-the-shelf solutions. The platform provides an interface that requires little to no experience in software development. It offers time-synchronized data collection, spatialized binaural audio playback, a loudspeaker array audio interface, and visual rendering for commercially available VR hardware and accessories. Furthermore, evaluation and characterization data have been presented for validating the platform and, possibly, compare it with other existing tools.

Future work on OMVR will include the implementation of additional modules and GUI components, the creation of a larger number of experiment templates, as well as further validations, including experiments carried out in clinical settings. Furthermore, considerations will be made whether to openly share the platform to the wider research community, in order to allow OMVR to contribute to the standardization of more ecologically valid, multimodal research endeavors in the field of sensory and cognitive hearing science.

6. ACKNOWLEDGMENTS

The authors would like to show appreciation to the Open-Source software community, specifically the developers and contributors behind the projects 3D Tune-In [42], TASCAR [43], and RtAudio [56]. The authors would like to thank Kirsten Marie Jensen Rico for her inputs on OMVR *template* designs and Borgný Súsonnudóttir Hansen and Camille Olsen Hald-Andersen for their help on collecting data for the platform validation efforts. Additionally, the authors would like to thank Pierre-Yves Hasan for his help on implementing the pupillometry-related features of the platform.

7 CONFLICT OF INTEREST

During the development of the OMVR platform, Rasmus Lundby Pedersen, Nynne Kajs, and Francois Patou were employees at Oticon Medical. Lorenzo Picinali is a Reader at the Dyson School of Design Engineering, Imperial College London, and declares no conflict of interest.

8 DATA AVAILABILITY

The data and materials for the experiments reported here may be made available upon request. The experiment presented was not preregistered. The data used for the validation study presented in SEC. 3 can be found at https://data.mendeley.com/datasets/69f6w6scdg/draft?a=1dff679a-da47-472d-a5cb-3298a45f49e4 [57].

9 REFERENCES

[1] B. S. Wilson, M. F. Dorman, M. G. Woldorff, and D. L. Tucci, “Cochlear Implants: Matching the Prosthesis to the Brain and Facilitating Desired Plastic Changes in Brain Function,” *Prog. Brain Res.*, vol. 194, pp. 117–129 (2011 Jul.). https://doi.org/10.1016/B978-0-444-53815-4.00012-1.

[2] National Institute on Deafness and Other Communication Disorders, “What Are Cochlear Implants for Hearing?,” NIH Pub. No. 00-4798 (2016 Feb.).

[3] J. Hillyer, E. Elkins, C. Hazlewood, S. D. Watson, J. G. Arenberg, and A. Parbery-Clark, “Assessing Cognitive Abilities in High-Performing Cochlear Implant Users,” *Front. Neurosci.*, vol. 12, paper 1056 (2019 Jan.). https://doi.org/10.3389/fnins.2018.01056.

[4] T. N. Tamati, C. Ray, K. J. Vasil, D. B. Pisoni, and A. C. Moberly, “High-and Low-Performing Adult Cochlear Implant Users on High-Variability Sentence Recognition: Differences in Auditory Spectral Resolution and Neurocognitive Functioning,” *J. Am. Acad. Audiol.*, vol. 31 , no. 5, pp. 324–335 (2020 May). https://doi.org/10.3766/jaaa.18106.

[5] K. Kestens, S. Degeest, and H. Keppler, “The Effect of Cognition on the Aided Benefit in Terms of Speech Understanding and Listening Effort Obtained With Digital Hearing Aids: A Systematic Review," *Am. J. Audiol.*, vol. 30 , no. 1, pp. 190–210 (2021 Mar.). https://doi.org/10.1044/2020_AJA-20-00019.

[6] A. C. Moberly, J. H. Lewis, K. J. Vasil, C. Ray, and T. N. Tamati, “Bottom-Up Signal Quality Impacts the Role of Top-Down Cognitive-Linguistic Processing During Speech Recognition by Adults With Cochlear Implants,” *Otol. Neurotol.*, vol. 42, no. 10S, pp. S33–S41 (2021 Dec.). https://doi.org/10.1097/MAO.0000000000003377.

[7] K. Gfeller, “Musical Backgrounds, Listening Habits, and Aesthetic Enjoyment of Adult Cochlear Implant Recipients,” *J. Am. Acad. Audiol.* 11, no. 7, 390–406 (2000 Jul./Aug.).

[8] S. Alhanbali, P. Dawes, S. Lloyd, and K. J. Munro, “Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults,” *Ear Hear.*, vol. 38, no. 1, (2017 Jan./Feb.). https://doi.org/10.1097/AUD.0000000000000361.

[9] E. Sulas, P.-Y. Hasan, Y. Zhang, and F. Patou, “Streamlining Experiment Design in Cognitive Hearing Science Using OpenSesame," *Behav. Res. Methods* (2022 Jul.). https://doi.org/10.3758/s13428-022-01886-5.

[10] WorldViz, " WorldViz: Virtual Reality for Training and Research,” http://www.worldviz.com (accessed Jun. 27, 2022).

[11] M. Vorländer, “Virtual Acoustics,” *Arch. Acoust.*, vol. 39, no. 3, pp. 307–318 (2014 Mar.). https://doi.org/10.2478/aoa-2014-0036.

[12] M. Salorio-Corbetto, B. Williges, W. Lamping, L. Picinali, and D. Vickers, “Evaluating Spatial Hearing Using a Dual-Task Approach in a Virtual-Acoustics Environment,” *Front. Neurosci.*, vol. 16 (2022 Mar.). https://doi.org/10.3389/fnins.2022.787153.

[13] J. Cubick and T. Dau, “Validation of a Virtual Sound Environment System for Testing Hearing Aids," *Acta Acust. united Acust.*, vol. 102, no. 3, pp. 547–557 (2016 May/Jun.). https://doi.org/10.3813/AAA.918972.

[14] N. Mansour, M. Marschall, T. May, A. Westermann, and T. Dau, “Speech Intelligibility in a Realistic Virtual Sound Environment," *J. Acoust. Soc. Am.*, vol. 149 no. 4, pp. 2791–2801 (2021 Apr.). https://doi.org/10.1121/10.0004779.

[15] G. Grimm, B. Kollmeier, and V. Hohmann, “Spatial Acoustic Scenarios in Multichannel Loudspeaker Systems for Hearing Aid Evaluation,” *J. Am. Acad. Audiol.,* vol. 27, no. 7, pp. 557–566 (2016 Jul.). https://doi.org/10.3766/jaaa.15095.

[16] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau, “Sound Source Localization With Varying Amount of Visual Information in Virtual Reality,” *PLoS One* , vol. 14, no. 3, paper e0214603 (2019 Jul.).

[17] M. A. Steadman, C. Kim, J.-H. Lestang, D. F. Goodman, and L. Picinali, “Short-Term Effects of Sound Localization Training in Virtual Reality,”

Sci. Rep., vol. 9, no. 1, paper 18284(2019 Dec.) https://doi.org/10.1038/s41598-019-54811-w.

[18] D. W. Massaro and J. Light, “Using Visible Speech to Train Perception and Production of Speech for Individuals With Hearing Loss,” *J. Speech Lang. Hear. Res.*, vol. 47, no. 2 (2004 Apr.). https://doi.org/10.1044/1092-4388(2004/025).

[19] A. Devesse, A. Dudek, A. van Wieringen, and J. Wouters, “Speech Intelligibility of Virtual Humans,” *Int. J. Audiol.*, vol. 57, no. 12, pp. 914–922 (2018 Sep.). https://doi.org/10.1080/14992027.2018.1511922.

[20] M. M. Hendrikse, G. Llorach, G. Grimm, and V. Hohmann, “Influence of Visual Cues on Head and Eye Movements During Listening Tasks in Multi-Talker Audiovisual Environments With Animated Characters,” *Speech Commun.*, vol. 101, paper 70–84 (2018 Nov.).

[21] J. Rouger, S. Lagleyre, B. Fraysse, S. Deneve, O. Deguine, and P. Barone, “Evidence That Cochlear-Implanted Deaf Patients Are Better Multisensory Integrators,” *Proc. Nat. Acad. Sci.*, vol. 104, no. 17, pp. 7295–7300 (2007 Apr.). https://doi.org/10.1073/pnas.0609419104.

[22] R. J. Briggs, “Future Technology in Cochlear Implants: Assessing the Benefit,” *Cochlear Implants Int.,* vol. 12, Sup. 1, pp. S22–S25 (2011 May.). https://doi.org/10.1179/146701011X13001035752291.

[23] S. D. Sechler, A. L. Valdes, S. M. Waechter, C. Simoes-Franklin, L. Viani, and R. B. Reilly, “Virtual Reality Sound Localization Testing in Cochlear Implant Users,” in *Proceedings of the 8th International IEEE/EMBS Conference on Neural Engineering (NER),* pp. 379–382 (Shanghai, China) (2017 Jun.). https://doi.org/10.3389/frvir.2021.691984.

[24] P. Majdak, M. J. Goupell, and B. Laback, “Two-Dimensional Sound Localization in Cochlear Implantees,” *Ear Hear.*, vol. 32 no. 2, pp. 198–208 (2011 Mar.). https://doi.org/10.1097%2FAUD.0b013e3181f4dfe9.

[25] R. Eastgate, L. Picinali, H. Patel, and M. D’Cruz, “3D Games for TUNing and lEarnINg About Hearing Aids,” *Hear. J.*, vol. 69 no. 4, pp. 30–32 (2016 May). https://doi.org/10.3030/644051.

[26] F. Pausch, L. Aspöck, M. Vorländer, and J. Fels, “An Extended Binaural Real-Time Auralization System With an Interface to Research Hearing Aids for Experiments on Subjects With Hearing Loss,” *Trends Hear.,* vol. 22, paper 2331216518800871 (2018 Jan./Dec.). https://doi.org/10.1177/2331216518800871.

[27] F. Pausch and J. Fels, “MobiLab–A Mobile Laboratory for On-Site Listening Experiments in Virtual Acoustic Environments,” *Acta Acust. united Acust.*, vol. 105, no. 5, pp. 875–887 (2019 Jul.). https://doi.org/10.3813/AAA.919367.

[28] D. Vickers, M. Salorio-Corbetto, S. Driver, et al., â€œInvolving Children and Teenagers With Bilateral Cochlear Implants in the Design of the BEARS (Both EARS) Virtual Reality Training Suite Improves Personalization," *Front. Digital Health*, vol. 3, paper 759723 (2021 Nov.). https://doi.org/10.3389/fdgth.2021.759723.

[29] J. K. Bizley, N. Elliot, K. C. Wood, and D. V. Vickers, “Simultaneous Assessment of Speech Identification and Spatial Discrimination: A Potential Testing Approach for Bilateral Cochlear Implant Users?,” *Trends Hear.,* vol. 19 (2015 Dec.). https://doi.org/10.1177/2331216515619573.

[30] ZeroMQ, “ZeroMQ,” http://wiki.zeromq.org/ (accessed Feb. 17, 2022).

[31] Qt Group, "Qt: Tools for Each Stage of Software Development Lifecycle," https://www.qt.io/ (accessed Jan. 13, 2022).

[32] Unity Technologies, "Unity Real-Time Development Platform," https://unity.com/ (accessed Feb. 3, 2022).

[33] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq, “Studying Human Behavior With Virtual Reality: The Unity Experiment Framework,” *Behav. Res. Methods* , vol. 52 (2019 Apr.) https://doi.org/10.3758/s13428-019-01242-0.

[34] T. Stenner, “Labstreaminglayer,” https://github.com/sccn/labstreaminglayer (2018). (accessed May 24, 2023).

[35] Pupil Labs GmbH, "Eye Tracking Technology - Gain Insight Into Human Behavior - Pupil Labs," https://pupil-labs.com/ (accessed Jan. 6, 2022).

[36] g.tec medical engineering GmbH, "Brain-Computer Interface & Neurotechnology," https://www.gtec.at/ (accessed Feb. 4, 2022).

[37] PupilLabs VR, “HTC Vive Binocular Add-on,”. https://docs.pupil-labs.com/vr-ar/htc-vive/ (accessed May 24, 2023).

[38] C. Kothe, “Labrecorder,” https://github.com/labstreaminglayer/App-LabRecorder (accessed Feb. 17, 2022).

[39] A. Ojeda and C. Kothe, “Extensible Data Format,” https://github.com/sccn/xdf (accessed Feb. 17, 2022).

[40] Research Institute Acoustics, “SOFA (Spatially Oriented Format for Acoustics),” https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics) (accessed May 24, 2023).

[41] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, et al., “3D Tune-In Toolkit: An Open-Source Library for Real-Time Binaural Spatialisation,” *PLoS ONE*, vol. 14, no. 3, paper e0211899 (2019 Mar.). https://doi.org/10.1371/journal.pone.0211899.

[42] M. Cuevas-Rodriguez, T. Murray-Browne, C. Garre, D. Gonzalez-Toledo, L. Molina-Tanco, and Rodríguez-Rivero Ángel, “Unity Wrapper for 3D-Tune-In Toolkit,” https://github.com/3DTune-In/3dti_AudioToolkit_UnityWrapper (2018). (accessed May 24, 2023).

[43] HörTech gGmbH, “Tascar,” https://github.com/HoerTech-gGmbH/tascar. (accessed May 24, 2023)

[44] J. Beatty, “Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources,” *Psychol. Bull.*, vol. 91, no. 2, pp. 276–292 (1982 Mar.).

[45] R. D. Gavas, S. R. Tripathy, D. Chatterjee, and A. Sinha, “Cognitive Load and Metacognitive Confidence Extraction From Pupillary Response," *Cognit. Syst. Res.*, vol. 52, pp. 325–334 (2018 Dec.). https://doi.org/10.1016/j.cogsys.2018.07.021.

[46] A. B. Watson and J. I. Yellott, "A Unified Formula for Light-Adapted Pupil Size," *J. Vision*, vol. 12, no. 10, paper 12 (2012 Sep.). https://doi.org/10.1167/12.10.12.

[47] Valve, "SteamVR," https://www.steamvr.com/en/ (accessed Feb. 17, 2022).

[48] Steinberg Media Technologies GmbH, "Audio Stream Input/Output (ASIO)," https://en.wikipedia.org/wiki/Audio_Stream_Input/Output (accessed May 24, 2023).

[49] University of York, "SADIE Binaural Measurements: Head Related Transfer Function Measurements for Virtual Loudspeaker Rendering," https://www.york.ac.uk/sadie-project/database_old.html (2018) (accessed May 24, 2023).

[50] R. C. Seewald and J. S. Gravel, *Sound Foundation Through Early Amplification. Proceedings of the Second International Conference* (Phonak AG/Immediate Proceedings, Stäfa, Switzerland, 2002).

[51] D. Bridges, A. Pitiot, and J. W. P. Michael R. MacAskill, "The Timing Mega-study: Comparing a Range of Experiment Generators, Both Lab-Based and Online," *PeerJ*, vol. 8, paper e9414 (2020 Jul.) https://doi.org/10.7717/peerj.9414.

[52] M. L. Chénéchal and J. C. Goldman, "HTC Vive Pro Time Performance Benchmark for Scientific Research," in *Proceedings of the International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pp. 82–84 (Limassol, Cyprus) (2018 Nov.). https://doi.org/10.2312/egve.20181318.

[53] R. Tachibana and K. Matsumiya, "Accuracy and Precision of Visual and Auditory Stimulus Presentation in Virtual Reality in Python 2 and 3 Environments for Human Behavior Research," *Behav. Res. Methods*, pp. 729–751 (2021 Apr.). https://doi.org/10.3758/s13428-021-01663-w.

[54] ESI, "Maya44 USB+ 4-in/4-out USB Audio Interface," https://www.esi-audio.com/products/maya44usb+/ (accessed May 24, 2023).

[55] C. Boulay, "Audiocapture," https://github.com/labstreaminglayer/App-AudioCapture (2020).

[56] G. P. Scavone, "RtAudio," https://www.music.mcgill.ca/gary/rtaudio/ (2021).

[57] R. L. Pedersen, L. Picinali, and F. Patou, "OMVR SSiN Validation Xdf Dataset," *Mendeley Data*, V1 (2022). https://doi.org/10.17632/69f6w6scdg.1.

[58] C. Seux, "Classroom Render," (2020) https://www.blender.org/download/demo-files/ (accessed Oct. 27, 2020).

[59] Rendwork Studios, "PBR Modern Room," https://assetstore.unity.com/packages/3d/props/interior/pbr-modern-room-115341#publisher (accessed Feb. 17, 2022).

THE AUTHORS

Rasmus Lundby Pedersen

Lorenzo Picinali

Nynne Kajs

François Patou

Rasmus Lundby Pedersen holds a Master's degree in electrical engineering from the Technical University of Denmark. Primarily working with research and development of medical technologies, he has experience creating software and signal processing solutions for EEG-based Brain-Computer Interfaces, Hearing Implants, Virtual Reality, and Hearing Instruments. He currently works at Oticon A/S as Senior Discovery Engineer, where he is responsible for scouting, developing, and evaluating novel technological innovations for hearing instrument users.

•

Lorenzo Picinali is a Reader in Audio Experience Design at Imperial College London. In the past years, he worked in Italy, France, and the UK on projects related with 3D binaural sound rendering, interactive applications for visually and hearing impaired individuals, audiology, and hearing aids technology, audio, and haptic interaction, and more generally, acoustical virtual and augmented reality. More information about the projects in which Lorenzo is involved can be found here https://www.axdesign.co.uk/

•

Nynne Kajs has recently graduated with a master's degree in human-centered artificial intelligence from the Technical University of Denmark, specializing in VR and computer graphics, and has worked on a number of research projects in this field. She is currently hired as a programmer in a Danish video games company.

•

François Patou holds a PhD from the Technical University of Denmark. He has held various positions in the Medical Device industry including that of Senior Translational Research Manager at Oticon Medical where he was responsible for the conceptualization, implementation, and operation of research platforms for cognitive and behavioral research in cochlear implant users.