*Journal of Universal Computer Science, vol. 27, no. 6 (2021), 582-608*
*submitted: 28/11/2020, accepted: 19/4/2021, appeared: 28/6/2021 CC BY-ND 4.0*

# Multimodality for Passive Experience: Effects of Visual, Auditory, Vibration and Draught Stimuli on Sense of Presence

**Fabian Honegger**
(Renomation AG, Neuenhof, Switzerland
faby@bluewin.ch)

**Yuan Feng***
(Northwestern Polytechnical University, Xi'an, China;
Eindhoven University of Technology, Eindhoven, the Netherlands
https://orcid.org/0000-0001-9968-4817, freyafengtue@gmail.com)

**Matthias Rauterberg**
(Eindhoven University of Technology, Eindhoven, the Netherlands
https://orcid.org/0000-0002-2596-395X, g.w.m.rauterberg@tue.nl)

**Abstract:** Adequate use of multimodal stimuli plays a crucial role in help forming the sense of presence within a virtual environment. While most of the presence research attempts to engage more sensory modalities to induce a higher sense of presence, this paper investigates the relevance of each sensory modality and different combinations on the subjective sense of presence using a specifically designed scenario of a passive experience. We chose a neutral test scenario of "waiting at a train station while a train is passing by" to avoid the potential influence of story narrative on mental presence and replicated realistic multimodal stimuli that are highly relevant to our test setting. All four stimuli - visual, auditory, vibration, and draught - with 16 possibilities of combinations were systematically evaluated with 24 participants. The evaluation was performed on one crucial aspect of presence – "realness" to reflect user presence in general. The perceived realism value was assessed using a scalometer. The findings of main effects indicate that the auditory stimuli had the most significant contribution in creating the sense of presence. The results of interaction effects suggest the impact of draught stimuli is significant in relation to other stimuli - visual and auditory. Also, the gender effects revealed that the sense of presence reported by female participants is influenced by more factors than merely adding more sensory modalities.

**Keywords:** Multimodal stimuli, Passive experience, Presence, Realism, Scalometer, Gender effects
**Categories:** H.5.1; H.5.2
**DOI:** 10.3897/jucs.68384

## 1 Introduction

The development of "presence science" is a process of constantly evolving and understanding of the origin and nature of presence and the factors on which it depends [Hartmann, 2008, Riva et al., 2014]. The phenomenon of presence was first discussed in 1980, when "telepresence" was introduced to describe the experience when operators controlled machines remotely in a simulated environment as if they were there [Minsky, 1980]. This feeling of "being there" was considered to be conducive to effective task performance and was therefore valued and further researched [Riva et al., 2003]. Modern presence research trying to understand the theoretical construct of presence and explain the underlying mental mechanism that enables humans to feel presence. Nakatsu and colleagues [Nakatsu et al., 2005] proposed a new framework for studying presence and categorized human activities in relation to user presence using a dimension of two poles: passive versus active experience. Passive experience is mainly correlated with "mental presence", for instance, utilizing one's mental imagination when watching a movie. Active experience, such as participating in sports, emphasize involvement in a dynamic situation in which a user senses "physical presence" through personal body movement.

In responses to passive and active experience, two streams of factors emerged from literature and were reported to influence the formation of sense of presence [Regenbrecht and Schubert, 2002, Schubert, 2009]. The first cluster relates to the affordances of mediated environment that engage users, for instance, the extent of sensory information provided and multiple parameters of various media presentations [Ijsselsteijn et al., 1998]. The second cluster relates to "the action that is possibly happening within the virtual environment", for example, factors such as virtual body representation, body engagement, or medium interactivity [Sanchez-Vives and Slater, 2005]. The latter stream of research points out that the possibility to interact and move the virtual body, or even the illusion of movement in the virtual environment, is one of the successful factors for enhancing user presence. Due to this action-centered view, most efforts of presence research within recent years dedicated to studies involving active experience, while few studies have been published on studying user presence under passive experience [Danieau et al., 2012a, Danieau et al., 2012b, Danieau et al., 2014]. Among these, the majority were conducted within the cinematography field, where story presence was investigated when induced by storyline narratives [Dining, 2017].

Industries have long been interested in improving user presence during passive experiences. Around a decade ago, there was booming popularity and heavy investment in the adoption of stereoscopy technique in personal entertainment devices such as three-dimensional (3D) monitors, televisions, and laptops, due to its significant benefit than other visual quality improvements such as fidelity [Basdogan and Loftin, 2009, Obrist et al., 2013, Cummings and Bailenson, 2016, Rotter, 2017]. However, this emerging technology shrunk rapidly after a short time of dynamic growth, and the market encountered an unforeseen rapid collapse years later. Although suggested by Rotter [Rotter, 2017] that the research and development in 3D technology proved unprofitable and had no real demand from the user perspective, this huge need for the enhanced sense of presence still exists. Besides the potential benefits for personal entertainment industries, there were also reported needs for enhancing user presence within healthcare domains [Vincent et al., 2009], for instance, under circumstances in which the "possibility to interact" was disabled or limited. The research team of Philips company proposed the Ambient Experience of Magnetic Resonance Imaging (MRI) through the use of a soothing visual-audio experience that can be viewed while in-bore (via an easy-to-position mirror and headphones) [Anastos, 2007]. Enhanced user presence within this calming virtual environment can improve patients' experience and reduce potential claustrophobia during an MRI scan. Another case concerns user groups with mobility restraints, such as the elderly with dementia. People with dementia commonly undergo multisensory stimulation therapies to reduce the risk of sensory deprivation, challenging behaviors, and maintain cognitive functions [Collier and Jakob, 2017]. Within a virtual multisensory environment, an increased sense of presence can lead to enhanced therapeutic effects, therefore, contributing to improved well-being [Strong, 2020].

Ongoing research on multimodal stimuli for presence holds great promise in envisioning the next generation of passive experiences [Basdogan and Loftin, 2009]. As early as 1962, Heilig attempted to simulate real-life experience realistically using multiple sensory channels (e.g., visual, audio, breeze, odor, and vibrations/jolts), and built an individual cinema box named "sensorama simulator" [Heilig, 1962, Jones and Dawkins, 2018]. With nowadays technological advancement in mediated displays, acoustics, and haptics, we are able to create the experience of "being in a virtual world" to the next level [Velasco and Obrist, 2020]. Although the realistic and natural presentation of a mediated environment with more human sensory channels engaged were known can lead to an increased sense of presence [Ranasinghe et al., 2018]. It is surprising that there still is a lack of experimental evidence on the extent to which each sensory stimulus is relevant, and how different stimuli would impact the subjective sense of presence instead of task performance [Wei et al., 2019, Gonçalves et al., 2019].

Therefore, in contradiction to majority efforts within presence research conducted during active experiences, this paper aims to (a) study user presence under passive experience by trying to elicit influential factors induced by the possible body movements within the virtual environment. Furthermore, (b) we attempt to fill the research gap by gathering more evidence on the contributions of each sensory modality used both alone and in various combinations for enhancing the subjective sense of presence instead of task performance.

To achieve the objectives mentioned above, the following section reviewed the related work of multimodality research for enhancing user presence and challenges for assessing subjective sense of presence. Next, we present our experimental study that examines the main, interaction, and gender effects of multimodal stimuli - visual, auditory, vibration, and draught stimuli - on user presence through a specifically designed scenario of a passive experience. A physical "cave-like" test environment was chosen for simulating the experiences instead of the Head-Mounted Displays (HMD) was due to: (I) users wearing HMD will naturally turn around to check the visual content, and such active user engagement (head movements) is what we aim to avoid during passive experience presence research; (II) literature also suggests that non-worn approach has advantages such as providing more intuitive and natural exploration of the visual environment compared to a user encumbered by a head-worn device [Havig et al., 2011]. A neutral scene of "waiting at a train station while a train is passing by" was selected. Four realistic sensory stimuli that are highly relevant to the test scenario were replicated: (a) visual, the video of the train passing by; (b) auditory, the sound of the train, and the noise of the wheels clattering on the rails; (c) draught, the wind of the train hitting the skin; and (d) the vibration that trains triggers on the platform. Different stimuli were presented in all 16 combinations to be tested with 24 participants. Evaluation of user presence was performed on one crucial aspect of presence - "realness" - to reflect user presence in general. This perceived realism value was assessed using a scalometer. In addition, reported specification of the preferable sensory stimuli combination was collected during interviews along with other open comments.

This paper contributes to the related field by (a) providing evidence to better understand important contributing factors to user presence under passive experience; (b) benefiting future multimodal immersive interface design by providing insights to support optimal choices when combining sensory modalities and presenting simulated stimuli, especially when limited sensory channels are available.

## 2 Related Work

### 2.1 Multimodal Stimuli for Enhanced Presence

Presence induced during passive experience, where active participation is not involved, relies largely on the extent of sensory information in assisting with the formation of "mental" presence [Ijsselsteijn, 2002, Nakatsu et al., 2005]. Hence, proper use of multimodal stimuli in system design is vital for enhancing the presence of passive experience. Up until now, most recent immersive experiences using multimedia presentations still focus on audio-visual feedback, as it is relatively easy to reproduce such perception through high fidelity video and audio content [Cummings and Bailenson, 2016]. The most intensively studied sensory modality that provides critical information regarding spatial cues is visual stimuli. Visual stimuli were suggested to be the dominant sense in perceiving spatial information [Nesbitt, 2003, Hecht and Reiner, 2009]. Recent studies suggest that the use of stereoscopic visuals and broader fields of view of visual displays can significantly improve the sense of presence compared to other factors. Such as fidelity of visual content and realistic mapping, which are most known for enhancing the immersive experience [IJsselsteijn and Riva, 2003, Zerroug et al., 2008, Cummings and Bailenson, 2016]. Similar evidence was found with auditory modality. The spatialized sound is associated with a higher reported sense of presence than no sound or a non-spatialized sound [Larsson et al., 2005]. Therefore, our study adopts a wider field of view and spatialized sound for presenting audio-visual stimuli in a vivid and naturalistic way.

Despite most existing works that focus on audio-visual aspects, other sensory stimuli were explored and researched on the basis of traditional virtual experiences. Vibration is another kind of feedback that has been widely adopted in fields such as gamification and cinematography for increasing perceived sense of presence [Danieau et al., 2014, Guillotel et al., 2016]. Vibration calls on multiple human senses besides what we have commonly known as the tactile sense. It also utilizes the vestibular sense (located in the semi-circular canals in the inner ear for perceiving the movement and position of the body) and skin senses (feeling of pressure caused by vibration). This feedback was suggested can improve perception of spatial cues in virtual reality environments [Kreimeier et al., 2019, Makin et al., 2019]. Limited works have been reported on exploring the effects of adding vibration stimuli based on audio-visual contents on the sense of presence during passive experience [Danieau et al., 2012a, Danieau et al., 2014]. More recently, other displays such as olfactory [Suzuki et al., 2014], wind [Zhou, 1999, Moon and Kim, 2004], wind with thermal [Ranasinghe et al., 2017], and wind, thermal, olfactory presentations were also investigated [Ranasinghe et al., 2018]. Findings of the above studies show that adding such displays can significantly affect the sense of presence compared to traditional virtual reality experience [Chen and Ding, 2019, Farooq et al., 2020].

Although it has been well acknowledged that proper system designs with displays of more sensory modalities were correlated with a higher sense of presence, only a limited amount of studies addressed the contribution of different modalities and the effects of their combinations. And the evaluations were mostly based on task performance instead of the subjective sense of presence. Moreover, to our best knowledge, none has been done during the passive experience. The interaction effects are of particular interest because current studies imply that different combinations of sensory stimuli can have a significantly different influence on task performance and sense of presence [Cooper et al., 2018]. And this is further explained in the following section.

### 2.2 Measuring User Presence

During the early stage of presence research, the sensation of "being there" was observed can optimal task performance. Thus many studies (i.e., [Welch, 1999, Nash et al., 2000, Stevens and Kincaid, 2015]) supported this idea of a positive correlation between presence and performance, and systems were designed to increase the sense of presence in the hope of improving user' task performance. Consequently, evaluation of the effectiveness of contributing factors, including different sensory modalities, were mostly based on indicators of task performance such as task efficacy (e.g., completion and reaction time), accuracy (e.g., error rate), and secondary task performance [Zimmons and Panter, 2003, Sanchez-Vives, 2005, Jia et al., 2012]. For instance, Burke [Burke et al., 2006] conducted a meta-analysis with 43 studies to examine combinations of visual-auditory and visual-tactile feedback compared to visual feedback alone on task performance. Findings suggest that adding modalities improve performance overall. Meanwhile, different modalities perform dissimilarly according to task type, workload, and the number of tasks. Visual-auditory feedback is most effective when a single task is being performed under normal workload conditions, while the visual-tactile combination is more effective when workload is high during multiple tasking conditions. This consensus has also been confirmed by a later systemic review conducted by Sigrist and colleagues [Sigrist et al., 2013], in which they studied augmented visual, auditory, and haptic feedback on motor learning. And conclude that although multimodal feedback can enhance motor learning, each modality within certain specific task requirements presents different advantages. Therefore, multimodal feedback designs should take advantage of each modality and fully consider the task requirements.

However, more recent evidence regarding the causation of presence and performance was mixed. Studies suggest that in some cases, a lower level of presence can also have a better performance [Bormann, 2006]. Therefore, the sense of presence is then considered should be measured independently from task performance. Assessment of subjective sense of presence that is reliable, valid, and robust is essential for designing multimodal media from the user perspective. In order to find an appropriate method for qualitative measuring user presence during passive experience, we examined the possibilities for adopting existing methodologies.

In literature, questionnaire-based presence assessment is the most effective and the largest category by far [Insko, 2003]. There were many questionnaires developed and continuously refined based on the improvements of questionnaires design methodologies and theoretical understanding of presence experience. The variety of presence assessment is partially due to the lack of consensus on how to conceptualize presence [Hein et al., 2018]. Among which, Schubert [Schubert, 2009] concluded previous works and proposed the well-known multidimensional presence construct that involves three individual components termed as: "spatial presence" – the feeling of being in a virtual environment; "involvement" - the degree of attention focused on the virtual environment; and "realness" – one's judgment of the realness of a virtual environment compared to reality. The above three dimensions were used to evaluate user presence (e.g., IGroup Presence Questionnaire [Schubert et al., 1999, Schubert et al., 2001]). The reported limitations of questionnaires mostly concern their intrinsically limited scope (e.g., participants without proper training may assess aspects of the sense of presence based on their own interpretations) [IJsselsteijn, 2004]; post-experiment use (e.g., the recall of previous experiences) [Van Baren and IJsselsteijn, 2004]; and the potential circularity problem (i.e., the fact of asking questions about presence itself may bring bias on presence assessment) [Slater, 2004]. Other often-used methods include behavioural measures and physiological measures (e.g., heart rate and skin conductance). Limitations of the above methods in use of studies during passive experiences concern inapplicable due to limited bodily movements (e.g., evaluations of behavioural measures) or its intrusive nature (e.g., accurate physiological measures are commonly assessed through multiple wearable sensors, which may risk reducing the sensation of presence).

In this paper, given considerations of practical limitations - the large number of sessions a participant needs to experience and the total length of the study, we attempt to find a less time-consuming method than repeatedly filling out the questionnaires after each session. And focuses on one important aspect of presence construct that is highly relevant to our test setting – "realness". Perceived realism concerns the degree to which a particular medium can reproduce seemingly accurate representations of real-life experience [Van Baren and IJsselsteijn, 2004, Lombard et al., 2009]. Inspired by the work of IJsselsteijn [IJsselsteijn et al., 1998], in which he proposed the use and design of a scalometer and asked the subjects to move a slide potentiometer along a scale to indicate their perceived realism. With a scalometer, participants can indicated thier perceived realism value effectively, immediately, and conveniently after exposure to each test condition. The sensitivity, reliability, and validity of this method were confirmed by his later works [IJsselsteijn, 2004]. We used a scalometer to quantify realism value and avoid participants' potential boredom and fatigue that may influence the evaluation's reliability.

# 3 Materials and Methods

## 3.1 Test Environment and Stimuli

### 3.1.1 Test Environment

The immersive scenario of "sitting on a platform while a subway train is passing by" was chosen due to the following reasons: (a) waiting for a train is almost a neutral setting and similar to dark rides [Langhof and Güldenberg, 2019]. It tries to avoid the influence of narrative storyline on user's sense of presence [Dining, 2017]; (b) this specific scenario is familiar to most potential participants, and therefore, it limits the influence of pre-experience on user presence; (c) the test session of experiencing this scenario is short enough (about 1 minute) to avoid potential boredom [Aart van et al., 2010] and simple enough for easy repetition; and (d) four major perceived representative sensory stimulations – visual, auditory, draught, and vibration, are relatively easy to replicate using existing technologies (see Figure 1).

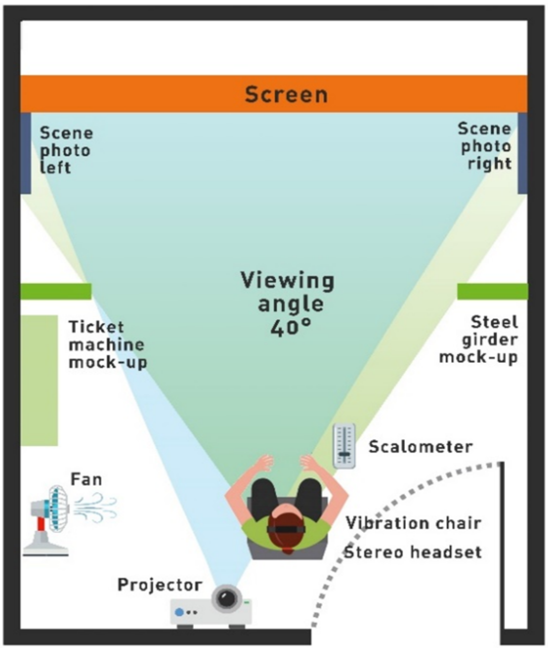


*Figure 1: Floorplan and layout of the silent chamber as the test environment.*

We transformed the audiology anechoic test chamber (at the Institute for Hygiene and Applied Physiology-IHA, ETH Zurich) into our test room, so that it shuttered outside noise, and inside, it visually appears to be more like a train station platform. As shown in Figure 2, the field of vision from the perspective of the test subject is filled with the projected video on a screen, a printed left/right scene photo, and built mock-ups to form a panorama view setting from centre to side, accordingly, see also Figure 3. The printed scene photo, both left and right, covered the floor to ceiling with coherent images that were shot from the same actual location of the video footage. In addition, the part of ceiling that is in view is also covered with digital images of the platform roof. This should provide an impression that the space extends along the rails.

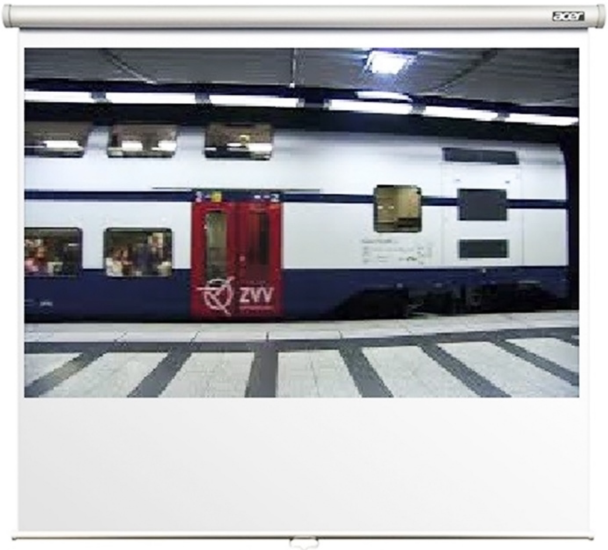


*Figure 2: Front view onto projection screen with train.*


*Figure 3: Front view left side with fan and right side; on both sides with mock-up decorations.*

#### 3.1.2 Simulation of Four Stimuli

The visual content was recorded (Panasonic S-VHS Videorecorder) at the Swiss train station platform of Zurich-Oerlikon and projected onto a screen via a projector. Stereo sound was recorded with the help of an artificial head microphone at the same location as the video. The artificial head microphone is an artificial head shape, in which the microphones are embedded in the ears. These human-like reflections of sound waves result in a highly realistic stereo sound effect. The sound is reproduced via stereo headphones (Stax SR-Lambda Pro). Headphones were chosen because the stereo effect (train runs from left to right) with surround sound was not sufficiently effective during a test trial. When wearing headphones, we assume no loss of realism, since many people today are used to being in public spaces with headphones for their mobile devices. Also, audio speakers would have had to be camouflaged somehow, otherwise visual realism would be compromised. The test room is soundproof so that no noise can enter the chamber from outside.

In addition to the video and audio content, the vibration is simulated by a high-quality balanced vibration chair. A vibrating high precision coil is attached to a plate and can be vibrating in different strengths. The chair is connected to the plate with a spring system, which then vibrates strongly at 50 Hz. The vibration data was also recorded authentically on site at the Oerlikon train station platform. Although we tried to include the olfactory modality (i.e., the smell of the train station) through odour display technology, but we discovered a trade-off with draught (see also [Barfield and Danas, 1996]). Due to technical constraints (i.e., no recordings of odour) and the overall upcoming importance of "wind displays", we decided to exclude the olfactory but use the draught stimuli. The draught is simulated by a commercial fan [see Figure 3]. The fan is outside the field of vision of test subjects, so the visual perception of the presented scene would not be affected. The fan switches on slowly and turns off slowly, together with the train on the screen. All these four stimuli were controlled and synchronized by a computer and programmed to provide the required stimuli according to each test condition. Stimuli were presented in high-quality and consistent ways to investigate the contributions of different modalities alone and in all possible combinations on perceived realism.

### 3.2 Study Design

The same scenario "waiting at a train station while a train is passing by" was used for all experimental sessions. There are 15 test conditions in total, which are generated by the combination possibilities of all four stimuli (including video, sound, vibration, draught, 6x bi-stimuli combinations, 4x tri-stimuli combinations, and 1x all stimuli together), except the one when no stimulus is presented. A within-subjects experiment with repeated measures was performed, so that each participant will experience all test conditions. To control for learning effects, a counterbalancing approach was adopted. Since testing with all permutations of the sequential order of all combinations is not feasible, we randomly assigned the participants to different sequential orders of test conditions.

As the experiment was designed to study the main and interaction effects of all stimuli alone and in combination, the following variables are used:

- **Independent variables:** test scenario with the full factorial combinations of video, sound, vibration, and draught.
- **Dependent variable:** evaluation of user presence in terms of perceived realism values.
- **Control variables:** demographic characteristic of participants (i.e., gender), previous experience, and ratings of the Eigenstate Scales [Nitsch, 1976].

### 3.3 Participants

The participant sample consisted of 24 test subjects (aged 17-54 years, 12 women and 12 men). The average age was 26.8 years (SE 1.8; SD 8.68) [see Table 1]. Fourteen (58%) participants had an academic educational background, and ten participants

(48%) did not. The professions were diverse. Table 1 shows all participant socio-demographics and data collected from the pre-experiment questionnaire. Using a t-test, no statistical difference in age between genders was found ($t$ = -1.054, $df$ = 22, $p$ < .303).

| **Variable** | **N** | |
|---|---|---|
| Age | | Mean (SD) |
| Female | 12 | 24.92 (9.45) |
| Male | 12 | 28.67 (7.90) |
| Total | 24 | 26.80 (8.68) |
| Eyesight | | Percent % |
| Good | 8 | 33.3 |
| Average | 16 | 66.7 |
| TV watching frequency per week | | |
| <2 hours | 9 | 37.5 |
| 2-5 hours | 7 | 29.2 |
| 5-10 hours | 7 | 29.2 |
| >10 hours | 1 | 4.2 |
| Cinema visiting frequency | | |
| 1x week | 1 | 4.2 |
| 1x month | 5 | 20.8 |
| 2x year | 15 | 62.5 |
| Rarely | 3 | 12.5 |
| Train travel frequency | | |
| Daily | 7 | 29.2 |
| Weekly | 6 | 25.0 |
| Monthly | 4 | 16.7 |
| Rarely | 7 | 29.2 |
| Sitting frequency while waiting for a train | | |
| Always | 3 | 12.5 |
| Often | 10 | 41.7 |
| Rarely | 7 | 29.2 |
| Never | 4 | 16.7 |

*Table 1: Demographics and pre-experiences such as eyesight, TV watching frequency, cinema visiting frequency, train travel frequency and sitting habit of participants.*

### 3.4 Measures

#### 3.4.1 Scalometer for Measuring the Perceived Realism Value

We evaluate the main dependent variable user presence through measuring the perceived realism of the presented scene compared to a real-life scenario. For pre- and post-experiment and each test session, we have to record the realism value of a scene, a numerical value that indicates how realistic a virtual scene looks [Hamberg and Ridder, 1996, Pearson, 1996]. In order to evaluate it statistically, we use our interval or a rational "realism" scale, measured via the scalometer – a linear potentiometer in the form of a slider [Series, 2012]. With our scalometer, we can continuously measure the realism value from "very unrealistic" (0) to "very realistic" (10,000). The scalometer consists of a fixed handle and a sliding handle, which automatically moves back to the starting position ("very unrealistic") when it is released (see Figure 4). The distance between the two handles is measured with a potentiometer. On the right side of the chair, participants positioned the two handles of the scalometer according to their impression, and then pressed the mouse on the left side to save this realism value. With this scalometer, the individual sensations were measured with regard to their perceived realism.

*Figure 4: The chair with a scalometer positioned at the right side of the participants.*

#### 3.4.2 Pre-experience Questionnaire and the Eigenstate Scale

Two questionnaires were filled out by all participants to assess their pre-experience (pre-experience questionnaire) and current states of mood (Eigenstate Scale, also referred to as EZ scale/questionnaire) [Nitsch, 1976]. Both questionnaires aim to control for a possible bias induced by individual experience or our experimental procedure, to make our further data analysis more robust. The pre-experience questionnaire was used to assess whether there is a bias induced by the pre-experience of participants. And the EZ scale assesses whether the participation of all 15 test conditions caused a carryover effect, such as user fatigue or mood variations.

The pre-experience questionnaire consists of eight questions regarding three aspects: visual and auditory abilities (wear glasses or not, eyesight, hearing ability, and hearing aid use); screen use experience (TV watching frequency and cinema visiting frequency); and train waiting experience (including frequency of train rides and whether they sit down when waiting). The EZ scale is a questionnaire designed by Nitsch [Nitsch, 1976] and can be used to determine the two factors of the current state (F1:stress level and F2:motivation). The EZ scale consists of 40-items as 6-point Likert scales. Each state descriptive item is rated from 1 ("hardly applicable") to 6 ("completely applicable") based on participant's personal evaluation of his/her current state. The 40 items are divided into eight first-level subscales (Mood, Sleepiness, Social Recognition, Self-confidence, Willingness to Communicate, Willingness to Work, Fatigue, and Tension) and calculated by summing the respective value of each questionnaire item. Moreover, those eight subscales can be integrated into second-level subscales (Activation, Efficiency, Affect, Deficiency), and further into the third level (Motivation and Stress) using a specific method of data analysis called Binary Structural Analysis (the combination of basic decision theory and factor analysis [Nitsch, 1974]).

### 3.5 Procedure

All participants were invited into the silent test chamber and were informed about the procedure of experiment and instructions of questionnaires before signing the informed consent forms. Subsequently, the experimenter explained how to use our scalometer. The experiment had three stages with different measurements taken.

*Stage-1 pre-experiment.* Demographic data were collected by asking the participant to first fill out the pre-experience questionnaire, the EZ scale, and then assess the realism value of test room setting (without presenting any stimuli) using the scalometer. During the pre-experiment, the experimenter remained present in case clarification was needed on questionnaires or scalometer use, and then he exited the room.

*Stage-2 during experiment.* The test subject was instructed to sit on the chair, to put the stereo headphones on, and to look ahead at the vertically positioned screen. The experimenter presented the 15 test conditions with different modality combinations in random order. After each of those 15 sessions, the participant was asked to assess how realistic it was compared to real-life experience. If the visual content is not presented, the test room is completely dark.

*Stage-3 post-experiment.* The realism value of the whole test setting was assessed again without any stimuli presented. The EZ scale was filled out a second time to assess the participant's current personal state. In the end, the experimenter interviewed the participants and asked questions including their favourite modality combination and other open comments regarding the whole experimental experience. Then he thanked them for their participation.

### 3.6 Data Analysis

IBM SPSS Statistics Version 25 was used for data entry and statistical computation. There was no missing data as all participants finished all experimental sessions. Data collected from the socio-demographics, pre-experience, and EZ questionnaires were compared between genders using the t-test for continuous variables and the chi-square for categorical variables. The non-parametric sign test and the Wilcoxon signed-rank test were implemented where applicable for analyzing the EZ scale pre- and post-experiment. The multivariate analysis of variances with the realism value as the dependent variable, and four stimuli as independent variables was performed to determine the main and interaction effects. The critical P value was set at 0.05 (=5% alpha error).

# 4 Results

## 4.1 Results of the Pre-experience Questionnaire

Data collected using the pre-experiment questionnaire was summarized using the means, standard deviations (SD), or percentage where applicable [see also Table 1]. The results show thirteen participants (54%) were wearing glasses or contact lenses. No participants (0%) were wearing a hearing aid. The self-rating of eyesight quality was average to good. The TV watching and cinema visiting behaviour did not deviate from normal TV watching and cinema visiting behaviour in Switzerland [King et al., 1996). About 2/3 (70.8%) of all participants were used to taking trains. About half (54.2%) of all participants were used to sitting while waiting for trains. In addition, chi-square tests show there is no significant association between TV usage and gender $X^2$ (3, 24) =5.21, $p$ =.157, cinema experiences and gender $X^2$ (3, 24) =1.60, $p$ =.659, train experiences and gender $X^2$ (3, 24) =.29, $p$ =.963, or the experience of sitting while waiting for a train and gender $X^2$ (3, 24) =.88, $p$ =.831. We also tested all questions in the pre-experience questionnaire (dependent variables as ordinal data) with non-parametric tests for independent samples (gender); none was significant. These results indicate that our convenience sample [Sousa et al., 2004] will not bias our statistical analysis regarding the realism measures [King et al., 1996].

## 4.2 Results of the Eigenstate Scale Pre- and Post-experiment

All 24 participants filled in the EZ scale pre- and post-experiment. We analysed the differences of eight first-level subscales using two different non-parametric tests: the sign test and the Wilcoxon signed-rank test. Significant results were found for the subscales “social recognition” (sign test $p$ = .041; Wilcoxon signed-rank test $p$ = .039) and “self-confidence” (sign test $p$ = .027; Wilcoxon signed-rank test $p$ = .028), indicating the test subjects were significantly more confident and familiar with the context after the experiment than before. This is normal, as participants are usually unfamiliar with the situation before participating in the experiment and feel more confident in control afterwards. No significance was found in other subscales between pre- and post-experimental conditions, meaning no significant carry-over effects between pre- and post-experiment was discovered.

### 4.3 Assessment of the Perceived Realism

#### 4.3.1 Results of Baseline Realism Value Pre- and Post-experiment of the Test Room

The mean and SD of both pre- and post-experiment realism values measured without any modal stimuli were adopted as the baseline for examining the test condition when no stimuli were presented, as shown in Table 2. We performed a MANOVA test on realism value using gender as between-subject factor and *pre-post* as the within-subject factor. We did not find a significant effect for *pre-post* $F(1, 22) = .13$, $p = .720$, partial $\eta^2 = .006$; *gender* $F(1, 22) = .67$, $p = .421$, partial $\eta^2 = .030$. The results show that regardless of gender, the total mean realism value after the experiment is higher than before. However, it is not significant. Therefore, we treat both total means as equal and use the average of both for our no-stimuli condition in the full factorial design (i.e., 15 sessions + 1 average =16 test conditions).

| Variable | Gender | Mean (SD) | N |
|---|---|---|---|
| Realism value pre-experiment | Female | 1494.33 (1881.15) | 12 |
| | Male | 1087.33 (1342.16) | 12 |
| | Total | 1290.83 (1611.58) | 24 |
| Realism value post-experiment | Female | 1666.58 (1900.75) | 12 |
| | Male | 1239.33 (1424.18) | 12 |
| | Total | 1452.96 (1656.97) | 24 |

*Table 2: Means and standard deviation of the baseline realism measure for different combinations of pre-, post-experiment and genders without designed stimuli presented.*

#### 4.3.2 Results of Realism Value under Experiment Conditions

To answer the main research question, we tested all combinations of stimuli (full factorial design) for the scalometer data measuring realism. It should be determined whether each individual stimulus (i.e., modality) has an impact on the assessment of the perceived realism, and whether particular interactions between individual stimuli can be determined. The size of the analysed data is 24 participants * 16 test conditions =384 data points. The multivariate analysis of variances with repeated measurements revealed several main and interaction effects. Each of the significant effects is described and presented in Table 3. Overall, we found partial eta-squared, as indicators of effect size, of medium and large sizes ($\eta^2 \geqslant .06$). Suggested norms for partial eta-squared according to Cohen's guidelines are small $\leqslant .01$; medium $\approx .06$; large $\geqslant .14$.

| **Effect** | **F** | **Sig. p value** | **Partial Eta Squared** | **Observed Power** |
|---|---|---|---|---|
| Video | 6.435 | **.019** | .226 | .679 |
| Sound | 102.045 | **<.001***** | .823 | 1.000 |
| Vibration | 18.386 | **<.001***** | .455 | .984 |
| Draught | 4.645 | **.042** | .174 | .540 |
| Video * Sound | .136 | .716 | .006 | .064 |
| Video * Vibration | 2.404 | .135 | .099 | .317 |
| Video * Draught | 5.783 | **.025** | .208 | .633 |
| Sound * Vibration | .595 | .449 | .026 | .114 |
| Sound * Draught | 8.024 | **.010**** | .267 | .773 |
| Vibration * Draught | 1.655 | .212 | .070 | .234 |
| Video * Sound * Vibration | 1.942 | .177 | .081 | .266 |
| Video * Sound * Draught | 2.674 | .116 | .108 | .346 |
| Video * Vibration * Draught | .287 | .597 | .013 | .081 |
| Sound * Vibration * Draught | 2.736 | .112 | .111 | .353 |
| Video * Sound * Vibration * Draught | 1.515 | .231 | .064 | .218 |
| Video * Gender | .078 | .782 | .004 | .058 |
| Sound * Gender | 1.067 | .313 | .046 | .167 |
| Vibration * Gender | 2.244 | .148 | .093 | .299 |
| Draught * Gender | 5.600 | **.027** | .203 | .619 |
| Video * Sound * Gender | .045 | .835 | .002 | .055 |
| Video * Vibration * Gender | 3.713 | .067 | .144 | .453 |
| Sound * Vibration * Gender | .003 | .956 | .000 | .050 |
| Video * Draught * Gender | 1.062 | .314 | .046 | .167 |
| Sound * Draught * Gender | 4.925 | **.037** | .183 | .564 |
| Vibration * Draught * Gender | 1.132 | .299 | .049 | .175 |
| Video * Sound * Vibration * Gender | 1.069 | .312 | .046 | .167 |
| Video * Sound * Draught * Gender | .360 | .555 | .016 | .089 |
| Video * Vibration * Draught * Gender | 2.116 | .160 | .088 | .285 |
| Sound * Vibration * Draught * Gender | .013 | .909 | .001 | .051 |
| Video * Sound * Vibration * Draught * Gender | .032 | .860 | .001 | .053 |

Note: a. Bold values are p < .05, ** p <.01, *** p <.001.

*Table 3: Results of the multivariate analysis of variances with the realism value as the dependent variable, and various combinations of four stimuli and gender as independent variables to disclose all the significant main and interaction effects.*

*Main Effects.* Significant main effects were exhibited with all four provided stimuli, meaning that each modality had a significant increase of user presence in terms of realism. In particular, the results (as shown in Table 3) showed significant main effects of *video* $F(1, 22) = 6.44$, $p = .019$, partial $\eta^2 = .226$, *sound* $F(1, 22) = 102.05$, $p \leq .001$, partial $\eta^2 =.823$, *vibration* $F(1, 22) = 18.39$, $p \leq .001$, partial $\eta^2 =.455$, and *draught* $F(1, 22) = 4.65$, $p = .042$, partial $\eta^2 =.174$. The estimated marginal means of realism value when each modality was presented (On) and not presented are reported in Table 4. With the video on, the realism value increased overall with an increasing percentage of 24% than the condition with the video off (calculated using the mean value presented in Table 4; INC = (3084-2493)/2493*100%). With *auditory* stimuli (sound On versus Off), the average perceived realism demonstrated a sharp increase by INC =169%. The growth rate for stimuli *vibration* is INC =32%, and *draught* by INC =15%.

| **Stimuli** | **Mean** | **SE** | **95% Confidence interval** | |
|---|---|---|---|---|
| | | | **Lower bound** | **Upper bound** |
| Video | | | | |
| Off | 2492.82 | 293.09 | 1884.98 | 3100.65 |
| On | 3083.79 | 367.14 | 2322.39 | 3845.19 |
| Sound | | | | |
| Off | 1511.23 | 250.55 | 991.62 | 2030.84 |
| On | 4065.38 | 403.42 | 3228.74 | 4902.13 |
| Vibration | | | | |
| Off | 2402.14 | 283.76 | 1815.66 | 2992.62 |
| On | 3172.46 | 359.29 | 2427.33 | 3917.59 |
| Draught | | | | |
| Off | 2596.70 | 304.70 | 1964.78 | 3228.62 |
| On | 2979.90 | 341.35 | 2271.98 | 3687.82 |

*Table 4: Mean and standard error (SE) for the main effect video, sound, vibration, and draught of realism measures.*

*Interaction Effects.* In our MANOVA analysis, we also found a significant two-way interaction effect between *video x draught* $F(1, 22) = 5.78$, $p = .025$, partial $\eta^2 = .208$. This indicates that the impact of the draught stimuli on perceived realism is dependent on the visual modality. In different words, only when the visual stimuli were presented, adding the draught would result in a significant increase on the realism value.

When the video was off, adding the draught stimuli increased the realism value by INC =8% (calculated using the mean values presented in Table 5); but in the condition when the video was on, perceived realism increased by INC =21%. The joint effect of video and draught has a significantly greater impact than the sum of each part alone. There was also a significant two-way interaction effect regarding *sound x draught* $F(1, 22) = 8.02$, $p = .010$, partial $\eta^2 = .267$. When the sound was off, the realism value was not increased, whether draught was provided or not (INC $\approx$0%). But when the sound was on, a scene was perceived as more realistic with the draught by an increase of INC

=21%. This outcome indicates that the draught gives an additional boost to the sense of reality when the sound is on. However, we could not find a significant two-way interaction effect between *vibration x draught* $F(1, 22) = 1.66$, $p = .212$, partial $\eta^2 = .070$. The effect sizes for all found that the significant effects above are large to extra-large. The post-hoc observed test power is negatively related to the p-value. No further significant main or interaction effects for the scalometer data *realism* could be identified.

| **Stimuli** | | **Mean** | **SE** | **95% Confidence interval** | |
|---|---|---|---|---|---|
| | | | | **Lower bound** | **Upper bound** |
| Video | Draught | | | | |
| Off | Off | 2400.35 | 313.73 | 1749.72 | 3050.98 |
| | On | 2585.28 | 308.16 | 1946.21 | 3224.36 |
| On | Off | 2793.05 | 339.36 | 2089.26 | 3496.85 |
| | On | 3374.52 | 413.83 | 2516.3 | 4232.75 |
| Sound | Draught | | | | |
| Off | Off | 1523.16 | 240.32 | 1024.78 | 2021.55 |
| | On | 1499.29 | 284.34 | 909.61 | 2088.98 |
| On | Off | 3670.24 | 397.34 | 2846.21 | 4494.27 |
| | On | 4460.51 | 454.78 | 3517.35 | 5403.68 |

*Table 5: Mean and standard error (SE) for the two-way interaction effects video x draught and sound x draught of realism measures.*

| **Gender** | **Stimuli** | **Mean** | **SE** | **95% Confidence interval** | |
|---|---|---|---|---|---|
| | | | | **Lower bound** | **Upper bound** |
| Female | Draught | | | | |
| | Off | 2668.90 | 430.92 | 1775.23 | 3562.57 |
| | On | 3472.88 | 482.74 | 2471.73 | 4474.02 |
| Male | Draught | | | | |
| | Off | 2524.50 | 430.92 | 1630.83 | 3418.17 |
| | On | 2486.93 | 482.74 | 1485.78 | 3488.07 |

*Table 6: Mean and standard error (SE) for the two-way interaction effect gender x draught of measure realism.*

*Gender Effects.* We also examined the main and interaction effects using *gender* as an additional independent variable. Main effect of *gender* was not significant $F(1, 22) = .83$, $p = .374$, partial $\eta^2 = .036$. The MANOVA output displayed a significant two-way interaction effect regarding *gender x draught* $F(1, 22) = 5.60$, $p = .027$, partial $\eta^2 = .203$. Female participants experienced higher realism with the draught stimulus On than Off (by an increase of INC=30%, calculated using the mean value presented in Table 6), while male participants rated similarly with or without draught. There was also a three-way interaction effect among *gender x sound x draught* $F(1, 22) = 4.93$, $p = .037$, partial $\eta^2 = .183$. For female participants, whether the sound was on or off, adding the draught stimuli increased the realism value. Especially in the condition sound is on, a scene is perceived more realistic with draught by an increase of INC=41% than without, calculated using the mean value presented in Table 7. On the contrary, independent of the sound stimuli, adding draught did not contribute to male participants' perceived realism.

| Gender | Stimuli | | Mean | SE | 95% Confidence interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower bound | Upper bound |
| | Sound | Draught | | | | |
| Female | Off | Off | 1624.22 | 339.86 | 919.40 | 2329.04 |
| | | On | 1702.23 | 402.12 | 868.29 | 2536.17 |
| | On | Off | 3713.58 | 561.92 | 2548.23 | 4878.94 |
| | | On | 5243.52 | 643.16 | 3909.69 | 6577.36 |
| Male | Off | Off | 1422.10 | 339.86 | 717.28 | 2126.93 |
| | | On | 1296.35 | 402.12 | 462.42 | 2130.29 |
| | On | Off | 3626.90 | 561.92 | 2461.54 | 4792.26 |
| | | On | 3677.50 | 643.16 | 2343.66 | 5011.34 |

*Table 7: Mean and standard error (SE) for the three-way interaction effect gender x sound x draught of measure realism.*

### 4.4 Findings from Interviews

All participants were interviewed and asked for their favorite combination of stimuli modalities. In order to simplify the choices, we provided three options: the *platform experience*, meaning the combination of all four stimuli together, which was intended to simulate the actual waiting experience; the *tunnel experience*, indicating the combination of sound, draught, and vibration, without video to simulate the experience of waiting in the dark (hence "tunnel"); and *others*, the remaining possibilities. It was surprising to find out that half of the participants (N = 12, female participants n = 4, male participants n = 8) feel the *tunnel experience* combination to be most realistic instead of the combination of all output media. Twenty-six percent of participants (N = 7, female participants n = 3, male participants n = 4) reported the *platform experience* as their favorite combination, and 21% (N = 5, all female participants) preferred the *others* combinations. The collected open comments indicated a great job in transforming the environmental setting and the good quality of the experiment materials, especially the stereo sound.

## 5 Discussion

### 5.1 Contributions of Different Modalities on User Presence

The analysis results of realism value confirmed with previous work that all four provided stimuli contribute to the perceived realism and therefore have an influence on sense of presence. Since the mean realism value was higher in conditions (a) with sound stimuli than without (increase by 169%), (b) with vibration stimuli than without (increase by 32%), (c) with video stimuli than without (increase by 24%), and (d) with draught stimuli than without (increase by 15%). As we can see, the contribution of each stimulus is different, meaning that they have different effect sizes. We found that the main effects with or without video and with or without draught are relatively small, while the main effect with sound is highly significant (partial $\eta^2 = .823$, observed power = 1.0). The vibration stimuli bring a considerable increase in the sense of realism as well. The realism value is more than doubled with sound provided in means compared to the other stimuli. Therefore, we can say that the auditory stimuli contribute immensely to the subject's sensation of realism.

Moreover, differentiated from previous multimodality research work, the significant interaction effects of different combinations of sensory modalities provide new insights in designing towards an increased sense of presence. Specifically, findings regarding significant two-way interaction effects *sound x draught* and *video x draught* suggest that the above two combinations are particularly impactful on top of the already known main effects for each modality separately. Moreover, results show that draught stimuli were significantly impactful only when audio or visual content was presented simultaneously. This could be explained by the process of how presence is experienced. Suggested by Wirth and colleagues [Wirth et al., 2007] that a user first needs to perceive the mediated environment through the spatial cues as a plausible space, then s/he is able to experience as being actually located within such a perceived space. In our case, video and audio stimuli contain sufficient information for participants to recognize the experience of train waiting at a platform, which draught or vibration stimuli did not. We could not find other significant 2-, 3-, or 4-way interaction effects. More findings regarding draught stimuli concerning gender effects are discussed in the later section. In summary, our results provide strong support for all research in developing wind simulation technologies (e.g., [Moon and Kim, 2004, Kulkarni et al., 2015, Ito et al., 2019]).

### 5.2 How Darkness Induces the Sense of Presence

Although the realism data shows that the combination of video, sound, vibration, and draught was perceived as the most realistic scene. An unexpected outcome is that, when asked for the most realistic combination during the final interview, only a quarter of participants opted for the *platform experience* (with all four stimuli presented). Half of them rated the *tunnel experience* (without video and therefore dark) as the most realistic scene. This raises the following question: Why did participants report answers in their interviews differently than scalometer measures regarding the most realistic scene?

One possible explanation could be that sense of presence in a virtual scene is much stronger in complete darkness. There is still a feeling of uncertainty and surprise. Some participants reported of a small sign of fear, although the EZ scale for the "Tension" subscale did not show any effect here. Several participants said that they had the feeling of sitting at a subway station, which is likely to be able to simulate the described combination of sensory stimuli almost perfectly, especially with our high quality of stimuli presentations. Our results support and complement the work of Claudio and colleagues [Claudio et al., 2015], in which the sense of presence inside a feared virtual tunnel was explored. A different reason is probably because the visual presentation might not be immersive enough in terms of the presentation setting, video quality, and content for natural scene comprehension. Since we have decided to run the experiment in an audiology anechoic test chamber, and the actual size of this chamber constrained our visual viewing size. So that the video screen only occupied a small part of the field of view. This also limits the participants' head movement as we may naturally turning around when we hear the train comes on a platform. If the entire experiment had been carried out in a 360° immersive cinema, where the screen covers the entire field of view, immersion into the virtual scene would probably be intensified. Another problem is the creation of the impression of a large platform width, so the participants would feel as though they were on a real platform. Our printed pictures covering the side and top of the screen clearly could not achieve this impression. With the current setting, the participant perhaps felt more as though they were in a cinema or in front of a TV than in the real world.

### 5.3 Gender Differences in Perceived Realism

The revealed significant gender differences over the draught stimuli also deserve some comments, as few existing designs of wind displays for enhancing sense of presence during virtual reality experience have considered genders as an influential factor. Only female participants rated significantly higher realism value when the draught stimuli were presented than without. Moreover, independent of the sound stimuli, adding draught contributed to increased perceived realism of female participants, however, not male participants. This is particularly interesting in comparison to the discovered interaction effects that were regardless of gender. The two-way interaction effect regarding *sound x draught* showed that draught contributed to an increased sense of presence significantly only when sound stimuli were provided as well.

To understand the possible reasons behind this, we first need to know two essential elements in order to receive and process sensory-based information: (I) the ability to detect sensory stimuli, and this is done by various sensory organs as receptors including eyes, ears, nose, mouth, skin, muscles and joints; and (II) sensory systems to deliver such information to the brain for processing and integration to make sense of the sensory experience. Therefore, one possible reason for explaining the above could be due to the gender differences in the bio-physical properties of the skin, that women are more sensitive to the variations such as regional thermal changes caused by the wind hits the skin, see references like [Firooz et al., 2012, Matsukura et al., 2013]. Another possible explanation could be the differences when processing the sensory information by the central nervous system. Neuropsychological studies and theories have been long consistent with the findings that women show more bilateral activation during information processing than men [Kemp et al., 2004]. Hence, when draught stimuli were presented without sound, while male participants may follow relatively logical thinking and deciding whether the space as a plausible space; female participants may also value other factors during the recall of previous experiences to compare with the present one. And this could also explain why when asked for their favorite combination of stimuli modalities while all men chose the two expected combinations of *tunnel* and *platform experience* as their favourite combination, almost half of the women (N = 5) chose *others* as the most realistic. Future designs should take gender differences into account during multimodal display designs to increase the perceived sense of presence of users.

### 5.4 Limitations and Future Work

Apart from the above already discussed possible visual presentation limitations, another limitation concerns the experimental design. For a full factorial experiment of within-subject design, we did not run all permutation possibilities of the stimuli combinations due to the large number of possibilities. And the order of the presentation clearly has an influence on realism value, as shown by the pre- and post-experiment measure of realism value without stimuli presented. The post-experiment data is in general higher than before, although no significance is discovered. Also, participants who were initially presented with the generally poorly rated combinations might have rated higher on average than those who were initially presented with the generally highly rated combinations. In addition, we could not include smell or odour, so the effects of this modality in combination with others still remain unknown [Matsukura et al., 2013, Seah et al., 2014]. Further studies that adopt visual presentation media that covers a larger field of view, like 180°-360° IMAX screens, would be valuable for a more highly immersive experience. Evaluation of user presence could combine measurements of different aspects of presence other than realness alone (e.g., spatial presence, involvement, and realness) for a more comprehensive understanding. Moreover, our study could be replicated in a different narrative setting to confirm the validity of the main and interaction effects we identified. More modalities (e.g., taste and odour) could be involved in future works for a better understanding of multimodal stimuli on user presence under passive experience. In addition, according to gender effects, women are more enthusiastic about other factors than activating as many sensory stimuli as possible, which could be interesting for further investigations. Lastly, a further exploration could be conducted to investigate the effects of darkness on the sense of presence.

## 6 Conclusion

This paper aims to study the contribution of four modalities – *visual* (video), *auditory* (sound), *draught*, and *vibration* and their combinations on user presence under a neutral scene for a passive user experience. The main and interaction effects were systematically examined through a full factorial experimental design using four simulated stimuli as independent variables and perceived realistic feeling as the dependent variable to reflect user presence. The analysis of realism data collected through a scalometer shows that all four modalities have significant main effects on increasing the perceived realism. S*ound* stimulus alone has the highest impact on the sense of presence, and additionally in combination with *draught* but not with *video* stimuli. In addition, two significant two-way interaction effects indicate that the *draught* stimuli have a positive impact on the sense of presence in combination with *sound* or *video*, but not with *vibration*. Gender effects were also found over draught stimuli. This means that "wind displays" have a promising future, however gender differences need to be taken into account.

To increase user presence, proper selection of sensory modalities and presentation media in a meaningful manner is of great importance (see also [Rauterberg and Szabó, 1995]). In particular, the TV and other personal entertainment industries are advised to shift focus from the visual to the audio presentation. In addition, tactile stimulation (i.e., wind display) is relevant in combination with *auditory* and *visual* modalities. Therefore, future Ambient Experience designs of the healthcare domain could add wind display to increase patient comforts and satisfaction. Although *vibration* had a significant contribution to the sense of presence alone, it does not combine with any other modality (i.e., audio, video, draught). Finally, not only the form (sensory modalities) as such, but also the content (scene), supported by the fitting modalities through natural mappings, has to be considered in future designs (see, i.e., [Jung et al., 2015]).

**Acknowledgments**

We thank the Institute for Hygiene and Applied Physiology of Swiss Federal Institute of Technology, the Center for User System Interaction Research (IPO) of Eindhoven University of Technology, and China Scholarship Council for their support in this study.

## References

[Aart van et al., 2010] Aart Van, J., Bartneck, C., Hu, J., Rauterberg, M., and Salem, B. (2010). How to behave as Alice in Wonderland–about boredom and curiosity. *Entertainment Computing* 1**,** 125-137.

[Anastos, 2007] Anastos, J.P. (2007). The ambient experience in pediatric radiology. *Journal of Radiology Nursing*, 26(2), 50-55.

[Barfield and Danas, 1996] Barfield, W., and Danas, E. (1996). Comments on the use of olfactory displays for virtual environments. *Presence: Teleoperators & Virtual Environments* 5**,** 109-121.

[Basdogan and Loftin, 2009] Basdogan, C., and Loftin, R.B. (2009). Multimodal display systems: Haptic, olfactory, gustatory, and vestibular. *The PSI handbook of virtual environments for training and education***,** 116-134.

[Bormann, 2006] Bormann, K. (2006). Subjective performance. *Virtual Reality*, 9(4), 226-233.

[Burke et al., 2006] Burke, J.L., Prewett, M.S., Gray, A.A., Yang, L., Stilson, F.R., Coovert, M.D., Elliot, L.R., and Redden, E. (2006). "Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis," in *Proceedings of the 8th International Conference on Multimodal Interfaces*. (Banff, Alberta, Canada: ACM), 108-117.

[Chen and Ding, 2019] Chen, C., and Ding, S. (2019). How the Skin Thickness and Thermal Contact Resistance Influence Thermal Tactile Perception. *Micromachines* 10**,** 87.

[Claudio et al., 2015] Claudio, A.P., Carmo, M.B., Gomes, I.L., Esteves, F., and Gaspar, A. (2015). "Sense of presence inside a feared (virtual) tunnel", in: *Proceedings of 10th Iberian Conference on Information Systems and Technologies (*CISTI*)*: IEEE), m1-m7.

[Collier and Jakob, 2017] Collier, L., and Jakob, A. (2017). The multisensory environment (MSE) in dementia care: examining its role and quality from a user perspective. HERD: Health Environments Research & Design Journal, 10(5), 39-51.

[Cooper et al., 2018] Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., and Meyer, G. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PloS one* 13**,** 1-25.

[Cummings and Bailenson, 2016] Cummings, J.J., and Bailenson, J.N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology* 19**,** 272-309.

[Danieau et al., 2012a] Danieau, F., Fleureau, J., Cabec, A., Kerbiriou, P., Guillotel, P., Mollet, N., Christie, M., and Lécuyer, A. (2012a). "Framework for enhancing video viewing experience with haptic effects of motion," in *IEEE Haptics Symposium (HAPTICS)*. (Vancouver, BC, Canada: IEEE), 541-546.

[Danieau et al., 2014] Danieau, F., Fleureau, J., Guillotel, P., Mollet, N., Christie, M., and Lécuyer, A. (2014). Toward haptic cinematography: Enhancing movie experiences with camera-based haptic effects. *IEEE MultiMedia* 21**,** 11-21.

[Danieau et al., 2012b] Danieau, F., Lécuyer, A., Guillotel, P., Fleureau, J., Mollet, N., and Christie, M. (2012b). Enhancing audiovisual experience with haptic feedback: a survey on HAV. *IEEE transactions on haptics* 6**,** 193-205.

[Dining, 2017] Dining, A. (2017). "A user study of story presence in an immersive narrative experience tested with variant levels of immersion", in: *SMPTE Annual Technical Conference and Exhibition*: SMPTE), 1-8.

[Farooq et al., 2020] Farooq, A., Coe, P., Evreinov, G., and Raisamo, R. (2020). "Using dynamic real-time haptic mediation in VR and AR environments," in *Human Interaction and Emerging Technologies (IHIET),* eds. T. Ahram, R. Taiar, S. Colson & A. Choplin. (Cham: Springer Nature Switzerland AG), 407-413.

[Firooz et al., 2012] Firooz, A., Sadr, B., Babakoohi, S., Sarraf-Yazdy, M., Fanian, F., Kazerouni-Timsar, A., Nassiri-Kashani, M., Naghizadeh, M.M., and Dowlati, Y. (2012). Variation of biophysical parameters of the skin with age, gender, and body region. *The Scientific World Journal* 2012.

[Gonçalves et al., 2019] Gonçalves, G., Melo M., Vasconcelos-Raposo J., Bessa M. (2019). "Impact of different sensory stimuli on presence in credible virtual environments," *IEEE transactions on visualization and computer graphics* 26 (11), 3231-3240.

[Guillotel et al., 2016] Guillotel, P., Danieau, F., Fleureau, J., Rouxel, I., Christie, M., Galvane, Q., Jhala, A., and Ronfard, R. (2016). "Introducing basic principles of haptic cinematography and editing," in *Eurographics Workshop on Intelligent Cinematography and Editing (WICED),* eds. M. Christie, Q. Galvane, A. Jhala & R. Ronfard. (Lisbon, Portugal: The Eurographics Association), 15-21.

[Hamberg and Ridder, 1996] Hamberg, R., and Ridder, H. (1996). Continuous assessment of perceptual image quality. *Journal of the Optical Society of America. A, Optics, image science, and vision* 12**,** 2573-2577.

[Hartmann, 2008] Hartmann, T. (2008). "Presence," in *The International Encyclopedia of Communication,* ed. W. Donsbach. (Oxford, UK: Blackwell Publishing).

[Havig et al., 2011] Havig, P., McIntire, J. and Geiselman, E. (2011). Virtual reality in a cave: limitations and the need for HMDs? *Head-and Helmet-Mounted Displays XVI: Design and Applications*. International Society for Optics and Photonics 8041, 804107.

[Hecht and Reiner, 2009] Hecht, D., and Reiner, M. (2009). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research* 193**,** 307-314.

[Heilig, 1962] Heilig, M.L. (1962). *Sensorama simulator*. US patent application 81,864.

[Hein et al., 2018] Hein, D., Mai, C., and Hußmann, H. (2018). "The usage of presence measurements in research: a review," in *Proceedings of the 17th International Workshop on Presence*. (Prague, Czech Republic: The International Society for Presence Research), 1-28.

[Ijsselsteijn et al., 1998] Ijsselsteijn, W., De Ridder, H., Hamberg, R., Bouwhuis, D., and Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. *Displays* 18**,** 207-214.

[Ijsselsteijn, 2002] Ijsselsteijn, W.A. (2002). Elements of a multi-level theory of presence: Phenomenology, mental processing and neural correlates. *Proceedings of PRESENCE* 2002**,** 245-259.

[Ijsselsteijn, 2004] Ijsselsteijn, W.A. (2004). *Presence in Depth.* PhD Dissertation, Technische Universiteit Eindhoven.

[Ijsselsteijn and Riva, 2003] Ijsselsteijn, W.A., and Riva, G. (2003). "Being there: The experience of presence in mediated environments," in *Being there: Concepts, Effects and Measurement of User Presence in Synthetic Environments,* eds. G. Riva, F. Davide & W.A. Ijsselsteijn. (Amsterdam Washington Tokyo: IOS Press-Omsha), 3-16.

[Insko, 2003] Insko, B.E. (2003). "Measuring presence: Subjective, behavioral and physiological methods," in *Being There: Concepts, Eeffects and Measurement of User Presence in Synthetic Environments,* eds. G. Riva, F. Davide & W.A. Ijsselsteijn. (Amsterdam Washington Tokyo: IOS Press-Omsha), 110-118.

[Ito et al., 2019] Ito, K., Ban, Y., and Warisawa, S.-I. (2019). "Manipulation of the perceived direction of wind by cross-modal effects of wind and three-dimensional sound", in: *Proceedigns of IEEE World Haptics Conference (WHC)*: IEEE), 622-627.

[Jia et al., 2012] Jia, D., Bhatti, A., Nahavandi, S., and Horan, B. (2012). Human performance measures for interactive haptic-audio-visual interfaces. *IEEE Transactions on Haptics* 6**,** 46-57.

[Jones and Dawkins, 2018] Jones, S., and Dawkins, S. (2018). "The sensorama revisited: evaluating the application of multi-sensory input on the sense of presence in 360-degree immersive film in virtual reality," in *Augmented reality and virtual reality*. Springer), 183-197.

[Jung et al., 2015] Jung, T., Chung, N., and Leue, M.C. (2015). The determinants of recommendations to use augmented reality technologies: the case of a Korean theme park. *Tourism Management* 49**,** 75-86.

[Kemp et al., 2004] Kemp, A.H., Silberstein, R.B., Armstrong, S.M., and Nathan, P.J. (2004). Gender differences in the cortical electrophysiological processing of visual emotional stimuli. *NeuroImage* 21**,** 632-646.

[King et al., 1996] King, A., Wold, B., Tudor-Smith, C., and Harel, Y. (1996). *A report of the 1993-94 survey results of Health Behaviour in School-Aged Children: A WHO cross-national study.* Copenhagen: World Health Organization.

[Kreimeier et al., 2019] Kreimeier, J., Hammer, S., Friedmann, D., Karg, P., Bühner, C., Bankel, L., and Götzelmann, T. (2019). "Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality," in *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. (Rhodes, Greece: ACM), 289-298.

[Kulkarni et al., 2015] Kulkarni, S.D., Fisher, C.J., Lefler, P., Desai, A., Chakravarthy, S., Pardyjak, E.R., Minor, M.A., and Hollerbach, J.M. (2015). A full body steerable wind display for a locomotion interface. *IEEE Transactions on Visualization and Computer Graphics* 21**,** 1146-1159.

[Langhof and Güldenberg, 2019] Langhof, J.G., and Güldenberg, S. (2019). Pirates, ghosts and customer loyalty–Reviewing the dark ride experience. *Tourism Management Perspectives* 31**,** 398-420.

[Larsson et al., 2005] Larsson, P., Väjamäe, A., Västfjäll, D., and Kleiner, M. (2005). "Auditory-induced presence in mediated environments and related technology," in *Handbook of Presence Research: Cognition, Measurement, and Interface Design,* eds. F. Biocca, W.A. Ijsselsteijn & J. Freeman. (Mahwah, NJ, USA: Lawrence Erlbaum), m1-m16.

[Lombard et al., 2009] Lombard, M., Ditton, T.B., and Weinstein, L. (2009). "Measuring presence: the temple presence inventory", in: *Proceedings of the 12th Annual International Workshop on Presence*: International Society for Presence Research), m1-m15.

[Makin et al., 2019] Makin, L., Barnaby, G., and Roudaut, A. (2019). "Tactile and kinesthetic feedbacks improve distance perception in virtual reality," in *Proceedings of the 31st Conference on l'Interaction Homme-Machine*. (Grenoble, France: ACM), 2:1-2:9.

[Matsukura et al., 2013] Matsukura, H., Yoneda, T., and Ishida, H. (2013). Smelling screen: development and evaluation of an olfactory display system for presenting a virtual odor source. *IEEE Transactions on Visualization and Computer Graphics* 19**,** 606-615.

[Minsky, 1980] Minsky, M. (1980). Telepresence. *OMNI magazine***,** 44-52.

[Moon and Kim, 2004] Moon, T., and Kim, G.J. (2004). "Design and evaluation of a wind display for virtual reality", in: *Proceedings of the ACM symposium on Virtual reality software and technology*), 122-128.

[Nakatsu et al., 2005] Nakatsu, R., Rauterberg, M., and Vorderer, P. (2005). "A new framework for entertainment computing: from passive to active experience," in *Entertainment Computing (ICEC),* eds. F. Kishino, Y. Kitamura, H. Kato & N. Nagata. (Sanda, Japan: Springer), 1-12.

[Nash et al., 2000] Nash, E.B., Edwards, G.W., Thompson, J.A., and Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of human-computer Interaction*, 12(1), 1-41.

[Nesbitt, 2003] Nesbitt, K. (2003). *Designing multi-sensory displays for abstract data.*, University of Sydney.

[Nitsch, 1974] Nitsch, J.R. (1974). Die hierarchische Struktur des Eigenzustandes-ein Approximationsversuch mit Hilfe der Binärstrukturanalyse (BISTRAN). *Diagnostika* 20**,** 142-164.

[Nitsch, 1976] Nitsch, J.R. (1976). "Die Eigenzustandsskala (EZ)-Ein Verfahren zur hierarchisch mehrdimensionalen Befindlichkeitsskalierung," in *Beanspruchung im Sport: Beitraege zur psychologischen Analyse sportlicher Leistungssituation,* eds. J.R. Nitsch & I. Udris. (Bad Homburg: Limpert), 81-102.

[Obrist et al., 2013] Obrist, M., Wurhofer, D., Meneweger, T., Grill, T., and Tscheligi, M. (2013). Viewing experience of 3DTV: An exploration of the feeling of sickness and presence in a shopping mall. *Entertainment computing* 4**,** 71-81.

[Pearson, 1996] Pearson, D. (1996). "Selective use of model-based coding for large moving objects," in *Proceedings of 8th European Signal Processing Conference (EUSIPCO).* (Piscataway: IEEE), 1-4.

[Ranasinghe et al., 2017] Ranasinghe, N., Jain, P., Karwita, S., Tolley, D., and Do, E.Y.-L. (2017). "Ambiotherm: enhancing sense of presence in virtual reality by simulating real-world environmental conditions", in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*: ACM), 1731-1742.

[Ranasinghe et al., 2018] Ranasinghe, N., Jain, P., Thi Ngoc Tram, N., Koh, K.C.R., Tolley, D., Karwita, S., Lien-Ya, L., Liangkun, Y., Shamaiah, K., Eason Wai Tung, C. and Yen, C.C. (2018). "Season traveller: Multisensory narration for enhancing the virtual reality experience", in: *Proceedings of the CHI Conference on Human Factors in Computing Systems*: ACM), 1-13.

[Rauterberg and Szabó, 1995] Rauterberg, M., and Szabó, K. (1995). "A design concept for N-dimensional user interfaces," in *Proceedings of 4th International Conference INTERFACE to Real & Virtual Worlds*. (Montpellier, France: EC2&Cie), 467-477.

[Regenbrecht and Schubert, 2002] Regenbrecht, H., and Schubert, T. (2002). Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 11(4), 425-434.

[Riva et al., 2003] Riva, G., Davide, F., and Ijsselsteijn, W.A. (eds.). (2003). *Being There : Concepts, Effects and Measurements of User Presence in Synthetic Environments.* Amsterdam Washington Tokyo: IOS Press-Ohmsha.

[Riva et al., 2014] Riva, G., Waterworth, J., and Murray, D. (eds.). (2014). *Interacting with Presence: HCI and the Sense of Presence in Computer-mediated Environments.* Warsaw-Berlin: De Gruyter Open Ltd.

[Rotter, 2017] Rotter, P. (2017). Why did the 3D revolution fail?: The present and future of stereoscopy. *IEEE Technology and Society Magazine* 36**,** 81-85.

[Sanchez-Vives and Slater, 2005] Sanchez-Vives, M.V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332-339.

[Schubert et al., 1999] Schubert, T., Friedmann, F., and Regenbrecht, H. (1999). "Embodied presence in virtual environments," in *Visual Representations and Interpretations,* eds. R. Paton & I. Neilson. (London: Springer), 269-278.

[Schubert et al., 2001] Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments* 10**,** 266-281.

[Schubert, 2009] Schubert, T.W. (2009). A new conception of spatial presence: Once again, with feeling. *Communication Theory*, 19(2), 161-187.

[Seah et al., 2014] Seah, S.A., Martinez Plasencia, D., Bennett, P.D., Karnik, A., Otrocol, V.S., Knibbe, J., Cockburn, A., and Subramanian, S. (2014). "SensaBubble: a chrono-sensory mid-air display of sight and smell", in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM), 2863-2872.

[Series, 2012] Series, B. T., (2012). "Methodology for the subjective assessment of the quality of television pictures (BT.500-13)", in: *BT Series - Broadcasting service (television).* (Geneva, Switzerland: International Telecommunication Union (ITU)-Radiocommunication (R)).

[Sigrist et al., 2013] Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic Bulletin & Review* 20**,** 21-53.

[Slater, 2004] Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 13**,** 484-493.

[Slater, 2009] Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364**,** 3549-3557.

[Sousa et al., 2004] Sousa, V.D., Zauszniewski, J.A., and Musil, C.M. (2004). How to determine whether a convenience sample represents the population. *Applied Nursing Research* 17**,** 130-133.

[Stevens and Kincaid, 2015] Stevens, J.A., and Kincaid, J.P. (2015). The relationship between presence and performance in virtual simulation training. *Open Journal of Modelling and Simulation*, 3(02), 41.

[Strong, 2020] Strong, J. (2020). Immersive Virtual Reality and Persons with Dementia: A Literature Review. *Journal of gerontological social work*, 63(3), 209-226.

[Suzuki et al., 2014] Suzuki, R., Homma, S., Matsuura, E., and Okada, K.-I. (2014). "System for presenting and creating smell effects to video", in: *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI)*: ACM), 208-215.

[Van Baren and Ijsselsteijn, 2004] Van Baren, J., and Ijsselsteijn, W. (2004). "Measuring presence: A guide to current measurement approaches", in: *OmniPres project IST-2001-39237* (Eindhoven, Netherlands: Eindhoven University of Technology).

[Velasco and Obrist, 2020] Velasco, C., and Obrist, M. (2020). *MULTISENSORY EXPERIENCES: Where the Senses Meet Technology.* OXFORD University Press.

[Vincent et al., 2009] Vincent, E., Battisto, D., Grimes, L., and Planning, D. (2009). The effects of presence and influence in nature images in a simulated hospital patient room. *Therapeutic Benefits of Nature Images on Health*, 167.

[Wei et al., 2019] Wei, W., Qi, R., and Zhang, L. (2019). Effects of virtual reality on theme park visitors' experience and behaviors: a presence perspective. *Tourism Management* 71**,** 282-293.

[Welch, 1999] Welch, R.B. (1999). How can we determine if the sense of presence affects task performance?. *Presence*, 8(5), 574-577.

[Wirth et al., 2007] Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., and Gouveia, F.R. (2007). A process model of the formation of spatial presence experiences. *Media psychology* 9**,** 493-525.

[Zerroug et al., 2008] Zerroug, A., Cassinelli, A., and Ishikawa, M. (2008). "Spatial coverage vs. sensorial fidelity in VR", in: *of: Devices that alter perception workshop (dap 2008).(Cited on page 2.)*: Citeseer), 34-37.

[Zhou, 1999] Zhou, G. (1999). *Human perception of air movement: Impact of frequency and airflow direction on draught sensation.* PhD Dissertation, Technical University of Denmark.

[Zimmons and Panter, 2003] Zimmons, P., and Panter, A. (2003). "The influence of rendering quality on presence and task performance in a virtual environment", in: *Proceedings of IEEE Virtual Reality*: IEEE), 293-294.