



OPEN ACCESS

EDITED BY

Augusto Sarti,
Politecnico di Milano, Italy

REVIEWED BY

Franco Penizzotto,
Facultad de Ingeniería–Universidad
Nacional de San Juan, Argentina
Eleftherios Triantafyllidis,
The University of Edinburgh,
United Kingdom

*CORRESPONDENCE

Verity McIntosh,
verity.mcintosh@uwe.ac.uk

SPECIALTY SECTION

This article was submitted to
Technologies for VR,
a section of the journal
Frontiers in Virtual Reality

RECEIVED 25 March 2022

ACCEPTED 07 July 2022

PUBLISHED 23 August 2022

CITATION

Bremner P, Mitchell TJ and McIntosh V
(2022), The impact of data sonification
in virtual reality robot teleoperation.
Front. Virtual Real. 3:904720.
doi: 10.3389/frvir.2022.904720

COPYRIGHT

© 2022 Bremner, Mitchell and
McIntosh. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

The impact of data sonification in virtual reality robot teleoperation

Paul Bremner¹, Thomas J. Mitchell^{2,3} and Verity McIntosh^{3,4*}

¹Bristol Robotics Laboratory, University of the West of England, Bristol, United Kingdom, ²Creative Technologies Lab, University of the West of England, Bristol, United Kingdom, ³Digital Cultures Research Centre, University of the West of England, Bristol, United Kingdom, ⁴Bristol VR Lab, University of the West of England, Bristol, United Kingdom

Virtual Reality (VR) is being increasingly used to provide a more intuitive and embodied approach to robotic teleoperation, giving operators a sense of presence in the remote environment. Prior research has shown that presence can be enhanced when additional sensory cues such as sound are introduced. Data sonification is the use of non-speech audio to convey information and, in the context of VR robot teleoperation, it has the potential to 1) improve task performance by enhancing an operator's sense of presence and 2) reduce task load by spreading data between sensory modalities. Here we present a novel study methodology to investigate how the design of data sonification impacts on these important metrics and other key measures of user experience, such as stress. We examine a nuclear decommissioning application of robotic teleoperation where the benefits of VR in terms of spatial reasoning and task performance are desirable. However, as the operational environment is hazardous, a sense of presence may not be desirable as it can lead to heightened operator stress. We conduct a study in which we compare the effects of *diegetic* sounds (literal and established sonifications) with *abstract* sounds (non-established sonifications). Our findings show that the diegetic sounds decrease workload, whereas abstract sounds increase workload, and are more stressful. Additionally, and contrary to expectations, sonification does not impact presence. These findings have implications for the design of sonic environments in virtual reality.

KEYWORDS

virtual reality, teleoperation, sonification, presence, robotics, auditory display, digital twin, human robot interaction

1 Introduction

From underwater surveillance (García et al., 2017) to disaster recovery (Schwarz et al., 2017), immersive technologies are increasingly being deployed for teleoperation applications, enabling operators to take action in a virtual reality simulation, or Digital Twin environment (Fuller et al., 2020). Operators can use virtual reality to phenomenologically project themselves into a remote environment in order to observe or embody a robot at work. Whilst the immersed operator is not necessarily physically present at the site of activity, their actions have direct physical consequences in the connected physical space.



FIGURE 1

A robot teleoperation workstation at Sellafield. Artistic adaptation from (Temperton, 2016).

A pervasive goal of immersive systems research is to engage an individual's senses to the extent that they perceive a virtual setting as the one in which they are consciously present (Cummings and Bailenson, 2015). Presence, in this respect, is typified by the psychological experience of 'being there' and previous VR studies have shown that presence is enhanced when a virtual environment is represented using multi-sensory cues (Cooper et al., 2018). Furthermore, as an individual's sense of presence increases, their ability to perform remote or simulated tasks has been shown to improve (Grassini et al., 2020). Consequently the sense of presence and relatedly a heightened awareness of the remote environment has the potential to enhance task performance and safety.

In this paper we focus on immersive robotic teleoperation applications in which a sense of "being there" may not be desirable, i.e., contaminated and hazardous environments where VR and teleoperation are used to complete tasks deemed too dangerous to perform in person. We explore an existing nuclear decommissioning teleoperation scenario in collaboration with experienced robot teleoperation practitioners at Sellafield nuclear facility: Europe's most contaminated and hazardous environment. We then conduct a study, in which these practitioners, along with a professional sound designer, collaboratively explore how sound can be used to support teleoperation tasks in a nuclear decommissioning scenario. In this application, operators are currently required to remotely control a demolition and excavation robot arm located

in a hazardous environment, a task which is currently undertaken using a multi-screen array, joystick and button controls, as shown in Figure 1. It is being considered for conversion to immersive teleoperation, and here we aim to understand whether sonification techniques might enable users to operate more efficiently and safely.

In Section 2 we review the literature on robotic teleoperation, presence and sonification in the context of VR, before setting out the study hypotheses in Section 3. In Section 4 we co-design both the VR robot teleoperation tasks, systems and sonification strategies, with implementation details provided in Section 5. We then compare the effect that different data presentation formats have on the sense of presence, workload and task performance for users of our VR-based teleoperated robotics simulation in Section 6, before closing the paper with discussions and conclusions in Sections 8, 9 respectively.

2 Background

Teleoperating a robot is a complex, high cognitive load task and in a nuclear decommissioning context it requires highly skilled operators to attend to multiple camera views and controls in order to carry out tasks effectively and safely Figure 1. In collaboration with operators at Sellafield Ltd., we are exploring Virtual Reality (VR) and immersive sound as an alternative teleoperation interface to explore

the potential advantages that the technology offers in terms of enhanced spatial reasoning and embodied interaction.

2.1 Presence

One of the key affordances of virtual reality is its ability to give users a sensation of presence: “*that peculiar sense of “being there” unique to virtual reality*” (Bailenson, 2018). Here we are discussing ‘presence’ as a subjective experience, as distinct from the more mechanistic qualities of “immersion” as disambiguated by Sanchez, Vives and Slater:

“The degree of immersion is an objective property of a system that, in principle, can be measured independently of the human experience that it engenders. Presence is the human response to the system, and the meaning of presence has been formulated in many ways.” (Sanchez-Vives and Slater, 2005).

Slater (2009) offers a helpful distinction between Place Illusion, “*the sensation of being in a real place*”, and Plausibility Illusion, “*the illusion that the scenario being depicted is actually occurring*”. This distinction gives us scope to consider each as a dynamic that could be measured and potentially modified. As Slater points out, users reporting a strong sense of presence in virtual reality should not be imagined to be fully convinced of the veracity of their circumstances. “*In the case of both (Place Illusion) and (Plausibility Illusion) the participant knows for sure that they are not “there” and that the events are not occurring*” (Slater, 2009). Murray (1997) suggests that in virtual reality “*we do not suspend disbelief so much as we actively create belief*”, doing so as a conscious act of attentional transfer. With this in mind, it may be helpful to consider a definition proposed by Nunez (2004) regarding “presence as the condition when a virtual environment becomes more salient as a source of cognition for a user than the real environment”.

It is worth noting that whilst users generally recognise and understand the artifice of their experience (Slater, 2009), in our robot teleoperation scenario, actions taken by the user in virtual reality are simultaneously enacted in the real world, i.e., translate directly in real time to the movements a robot arm in a high-risk, nuclear decommissioning environment. In this instance, the actions of the immersed operators have substantive consequences that could impact on commercial outcomes and public safety.

Presence is enhanced when a virtual environment is represented using visual and auditory cues (Cooper et al., 2018). Furthermore, user task performance has been shown to improve as their sense of presence increases (Grassini et al., 2020). A recent meta-analysis additionally concluded that a heightened sense of presence in VR can enhance both knowledge, skill development, and long-term learning effects (Wu et al., 2020).

Makransky and Petersen (2021) developed the Cognitive Affective Model of Immersive Learning (CAMIL) to explore the causes of this apparent improvement in performance of those in immersive vs. non-immersive environments. They conclude that “*heightened levels of situational interest, intrinsic motivation, self-efficacy, embodiment, and self-regulation and lower levels of cognitive load can have positive effects on learning outcomes*”.

When compared with a flat screen equivalent, VR has been shown to improve task performance as a result of the immersive 3D visual environment which provides an enhanced sense of space, depth and relative motion (Bowman and McMahan, 2007). Users demonstrate better and swifter spatial reasoning and improved motor skills in VR; although, interestingly, users consistently under-perceive relative distances (Kelly et al., 2022) which can present challenges for those designing for interactivity.

In presence research, immersive experiences are frequently mapped onto the Migram Kishino Reality-Virtuality (RV) continuum (Milgram et al., 1995) as a means to describe the extent to which a user’s perception is mediated by technology. This spectrum ordinarily places “real world” activity and “full immersion” experiences at opposite ends of a spectrum. In VR-enabled robotic teleoperation such as the use case being explored in this study, the user sits uncomfortably across the full spectrum: simultaneously present in, and cognitively prioritising, a constructed virtual environment, whilst physically implicated and materially impacting a digitally coupled, physical environment. The psychological consequences of traversing the reality-virtuality continuum in this way are not yet well understood.

2.2 Embodiment

Connected to presence is the experience of embodiment, wherein users report not just Place Illusion and Plausibility Illusion but a strong sense of themselves as physically present entities. This is sometimes referred to as “self-presence” (Lee, 2004; Ratan, 2012).

Madary and Metzinger (2016) discuss the significance of the Phenomenal Self-Model, and suggest that in VR this model can become extended or supplanted, such that the user regards the actions and experiences of the avatar that they embody as their own. This often leads to “*autophenomenological reports of the type “I am this!”*”

A sense of embodiment appears not to be limited to realistic, or personally recognisable avatars (Nunez, 2004). Studies have shown users capable of rapidly adapting their body schema when embodied within entities significantly different to their own. Some reference the “Proteus effect” in which subjects start to exhibit the characteristics they associate with their assigned body, e.g., people embodied as taller avatars behave with more confidence in a negotiation than those assigned shorter

avatars (Yee and Bailenson, 2007). Extreme examples have included participants reporting strong levels of self-presence when embodied as cows in an abattoir (Bailenson, 2018), or when embodied as an avatar with additional functioning limbs (Won et al., 2015). In the latter example, participants exhibit homuncular flexibility by “remapping movements in the real world onto an avatar that moves in novel ways”. This radical remapping appears to occur quickly and with no discernible impact on the participant’s sense of presence, something they attribute to the cortical plasticity of the human mind (Won et al., 2015).

Madary and Metzinger (2016) identify this as a potential risk of virtual reality technology, suggesting that “the mind is plastic to such a degree that it can misrepresent its own embodiment”. They caution that “illusions of embodiment can be combined with a change in environment and context in order to bring about lasting psychological effects in subjects” the main risk being that “long-term immersion could cause low-level, initially unnoticeable psychological disturbances involving a loss of the sense of agency for one’s physical body”.

Some have expressed concern that the deprioritising of one’s immediate surroundings in order to give cognitive priority to the virtual world could lead to depersonalization and derealization dissociative disorders, particularly with prolonged usage (Madary and Metzinger, 2016; Spiegel, 2018).

“Depersonalization involves a sense of detachment or unreality of one’s own thoughts, feelings, sensations, or actions, while derealization is marked by a sense of detachment or feeling of unreality with respect to one’s environment” (Spiegel, 2018).

Our industrial use case, and others like it, presume that operators will spend significant proportions of the working day in virtual reality; tele-operating robots, conducting reconnaissance, assessing risk, performing tasks, collaborating with other human and robotic entities. All the while cognitively “present” in a hazardous virtual environment, the physical analogue of which has been deemed unsafe for their human bodies. In this case, maximising the user’s “perceptual proximity” (Rubo et al., 2021) through presence might not be in the interests of employee wellbeing, nor may it be compliant from a health and safety perspective as operator workload and stress are key issues. It seems likely that in the near future, work of this nature will become directly identified by unions and policymakers and will be addressed within worker’s rights legislation.

2.3 Managing presence

Most presence research presumes the inherent desirability of presence for immersive environments, identifying technical and design features that can be optimised to keep users “in the moment”. However, in our use case, and others like it, there may be an argument against maximising presence, and for exploring alternative techniques that could modulate presence.

One approach might be to wilfully disrupt the user’s sense of presence at particular intervals. Whilst they view it as an undesirable quality, Sanchez, Vives and Slater speak about observed Breaks In Presence or BIPs.

“A BIP is any perceived phenomenon during the VE exposure that launches the participant into awareness of the real world setting of the experience, and, therefore, breaks their presence in the VE.” (Sanchez-Vives and Slater, 2005).

Nunez (2004) suggests that a break in presence tends to occur when the sensory input received does not match the expectations that the user has of the virtual world. This may be brushing against a piece of furniture in the real world that is not present in the virtual world, or when the design language of the virtual world switches and becomes inconsistent with the user’s expectations. Steed and Slater classify these as “External” and “Internal” BIPs (Slater and Steed, 2000).

In order to provoke an internal BIP (Nunez, 2004) guards against the use of surprising or fantastical elements, asserting that immersed users can accommodate a range of unusual or unrealistic scenarios in virtual reality such as magical creatures or physics-defying action without it leading to a BIP, so long as the user can attribute this to the diegetic reality of the virtual environment. Instead, an internal BIP might best be achieved by the confounding of diegetic expectations such as incongruous sound design or ludonarrative dissonance.

2.4 Immersive sound for robotic teleoperation

When performing nuclear decommissioning tasks in a remote environment, operators are required to attend to multiple streams of visual and non-visual information including radiation levels, proximity to obstacles, the state of robot and the position of objects being manipulated. Representing and conveying information in a way that does not increase the cognitive load of the operator is a significant challenge. Additional visual information is a potential solution; however, this risks dividing the operators attention away from the task they are completing. One possible approach is data sonification, where information is conveyed using the auditory channel by representing data features with sound. The human auditory system has a high temporal resolution, wide bandwidth, and is able to localise and isolate concurrent audio streams within an audio scene (Carlile, 2011). These features make hearing an indispensable channel in film and gaming, and consequently a promising medium to explore for robot teleoperation and, in particular, nuclear decommissioning.

Inspired by studies showing that multisensory integration of vision with sound improves our ability to accurately process information (Proulx et al., 2014), we are investigating how sound

might be used to complement visual feedback for robot arm teleoperation in VR.

2.4.1 Sonification

Sonification is defined as “*the use of non-speech audio to convey information*” (Kramer et al., 2010) and has been used effectively in a wide range of applications: sensory substitution (Meijer, 1992), medical diagnosis (Walker et al., 2019), peripheral process monitoring (Vickers, 2011) and visual decongestion (Brewster, 2002). A significant factor differentiating sonification from music is the intent to convey information by delegating some aspect of the aural fabric to a chosen data value (Sinclair, 2011).

Parameter Mapping Sonification (Grond and Berger, 2011) is one of the most widely used techniques for converting data streams to sound and involves creating manual connections between data features and auditory parameters. While this approach is simple to implement and offers great flexibility, data to sound mappings must be considered carefully as choices directly impact on the efficacy of an auditory display (Walker, 2002). The most widely used auditory parameters in the sonification literature are pitch, loudness, duration, panning and tempo (Dubus and Bresin, 2013), all of which represent perceptually salient auditory parameters that can be easily controlled using simple one-to-one parameter mappings.

The sonification of data features has been shown to improve learning of complex motor tasks. Sigrist et al. (2015) demonstrate this capability using a virtual reality rowing simulator for athlete training. They demonstrate the utility of splitting data between visual and audio channels, and that it aids user understanding of the presented information. Similarly Frid et al. (2019) demonstrate that splitting data presentation between visual, audio and haptic modalities improves task performance.

An ongoing challenge for sonification is the need to create auditory data representations that are transparent; that is, the data points must be rendered in such a way that they can be interpreted correctly by listeners. Kramer (1994) suggests that the transparency of an auditory display can be enhanced by selecting mappings that complement the metaphorical and affective associations of listeners. In this context, metaphorical associations refer to mappings in which a data variable change (i.e., rise in temperature) is represented by a metaphorically related change in an auditory variable (i.e., rise in pitch) (Walker and Kramer, 2005). Affective associations consider the attitudes that listeners have towards given data values, i.e., trends that might be considered “bad”, such as rising global poverty and CO2 emissions, could be mapped onto auditory features that are known to increase perceived noise annoyance (Di et al., 2016) or stress (Ferguson and Brewster, 2018).

Walker (2002) has shown that domain knowledge is an important factor in understanding how sonifications are

interpreted, implying that listeners invoke auditory expectations of how given data features should be expressed as sound. Sonifications that correlate auditory dimensions with listener expectations have been shown to be effective in a number of studies (Walker, 2002; Walker and Kramer, 2005; Ferguson and Brewster, 2019). Ferguson and Brewster (2018) refer to this correlation as *perceptual congruency* and have demonstrated that the psychoacoustic parameters roughness and noise correlate with the conceptual features danger, stress and error. Despite these efforts to support the design and evaluation of transparent and effective sonification, the majority of sonification practitioners take an intuitive, ad-hoc approach, making “unsupported design decisions” when creating mappings (Pirhonen et al., 2006; MacDonald and Stockman, 2018).

Consequently, numerous guidelines have emerged that are intended to support the sonification design process. For example, Frauenberger et al. (2006) define sonification design as “the design of functional sounds” and present a set of sonification design patterns to help novice practitioners create effective auditory displays. de Campo (2007) later presented a Sonification Design Space Map along with an iterative design process interleaving stages of implementation and listening in order to make salient data features more easily perceptible. Further guidelines have also been presented by Grond and Sound (2014) which link established listening modes from acousmatic music with the intended purposes of an auditory display. Sonifications were categorised as goal-orientated (normative) or exploration-orientated (descriptive) with design and aesthetic recommendations that were intended to enhance both functionality and usability. Participatory design methods have also been used to help define more rigorous data-sound representations, working with a variety of participants to rapidly explore and co-design potential solutions. Typically, this approach involves the use of practical design workshops in which stakeholders and independent experts engage in time constrained concept development activities from which design guidelines can be defined. For example, Droumeva and Wakkary (2006) iterated the designs for an interactive game over the course of two workshops in which developers and participants reviewed and refined sonification prototypes. Similarly, Goudarzi et al. (2015) developed sonification prototypes for climate data by bringing together domain scientists and sonification experts within a 2-day participatory workshop.

2.4.2 Sonification in human robot interaction

In the context of robotics, Zahray et al. (2020) use data sonification to provide information on arm motion such that an observer need not see the arm to understand how it is moving. Similarly, Robinson et al. (2021) have used sound as a way to provide additional information about the motion of a robot arm. In both of these studies they show that by using different sounds, perception of the motion and the robot capabilities can be altered.

Sonification has also been used to supplement non-verbal gestures to convey emotion. Both mechanical sounds (Zhang et al., 2017), and musical sonifications (Latupeirissa et al., 2020), have been utilised. While Bramas et al. (2008) use a wide variety of sounds to convey emotion. They found that perceptual shifts of the emotions being conveyed in gesture where possible, i.e., shifts in valence, arousal and intensity.

Hermann et al. (2003) use sound to convey some of the information in their complex process monitoring system of a cognitive robot architecture. By splitting the representation of different data features between the visual and auditory channels, operators were shown to establish a better understanding of system operation. This is conceptually similar to the aims of the work presented here.

Lokki and Gröhn (2005) demonstrate that users can use sonified data to navigate a virtual environment. They compare audio, visual, and audiovisual presentation of cues, and find audiovisual cues to perform best, demonstrating multimodal cue integration. However, it is worth noting that participants were still able to complete some of the task in the other conditions.

Triantafyllidis et al. (2020) evaluate the performance benefits of stereoscopic vision, as well as haptic and audio data feedback on a robot teleoperation pick and place task; both haptic and audio feedback were collision alerts rather than for data representation as evaluated here. They found that stereoscopic vision had the largest benefit, with audio and haptics having a small impact. This result gives credence to our use of stereoscopic VR for robot teleoperation.

It is clear that data sonification is an under explored area in HRI and telerobotics, with relatively few studies across a small number of application domains. Hence, the work we present here represents an early step in understanding the utility of sonification in HRI. We are interested in understanding the utility of sonification in this novel problem domain, consequently we collaboratively design and compare two different sets of sonifications and systematically investigate their impact on task performance and workload of teleoperators. The two sets of sounds are “diegetic”, i.e., more literal, calling on established representations of these parameters, and “abstract”, with no established connection to established sound representations.

3 Hypotheses

Based on the literature, we hypothesised that the inclusion or exclusion of different kinds of sounds would alter the operator’s perceptual model, potentially impacting performance, presence and comfort. We designed a study to address the following hypotheses:

1) Compared with visual representation, auditory data representations will:

- a) Improve usability.
 - b) Reduce workload.
 - c) Improve task performance¹.
- 2) Adding sonification to the VR environment will increase the user’s sense of presence.
- 3) Compared with the abstract sounds the diegetic sounds will:
- a) Improve usability.
 - b) Reduce workload.
 - c) Increase presence.
- 4) Compared with the diegetic sounds the abstract sounds will:
- a) Decrease presence without negatively impacting task performance².
 - b) Be less stressful than diegetic sounds.

4 Task definition and sonification design

To understand the domain, a remote discussion was conducted between the authors and two domain experts who were members of the Alpha Operations team at Sellafield Ltd. who use robotic teleoperation systems on a day-to-day basis. The aim of the consultation was to generate an understanding of the tasks that operators perform, the challenges they face, and to identify a set of data features for sonification.

A pick and place task was selected, as the most common activity performed, in which the teleoperated robots are used to sort objects in the hazardous environment. We also identified the three most salient data features that operators rely upon to complete this task:

- Radiation (risk of exposure).
- Proximity (risk of collision).
- Load (risk of stressing or overbalancing the robot arm).

To design the auditory representations of the above data features, we recruited a professional sound designer (specialising in VR audio) to take part in two co-creative workshops scheduled 2 weeks apart. Each workshop lasted a full day, and was attended by the sound designer, the two industry experts from Sellafield, and the authors. The aim of the first workshop was to generate two sets of sonifications for each of the above data features: one “diegetic” and the other “abstract”.

At the start of the workshop the domain experts introduced the telerobotics work they do at Sellafield, before describing the meaning and significance of the data features selected for sonification.

¹ See Section 6.3.1 for the relevant measures.

² The caveat of not negatively affecting task performance differentiates H4.b. from H3.c.

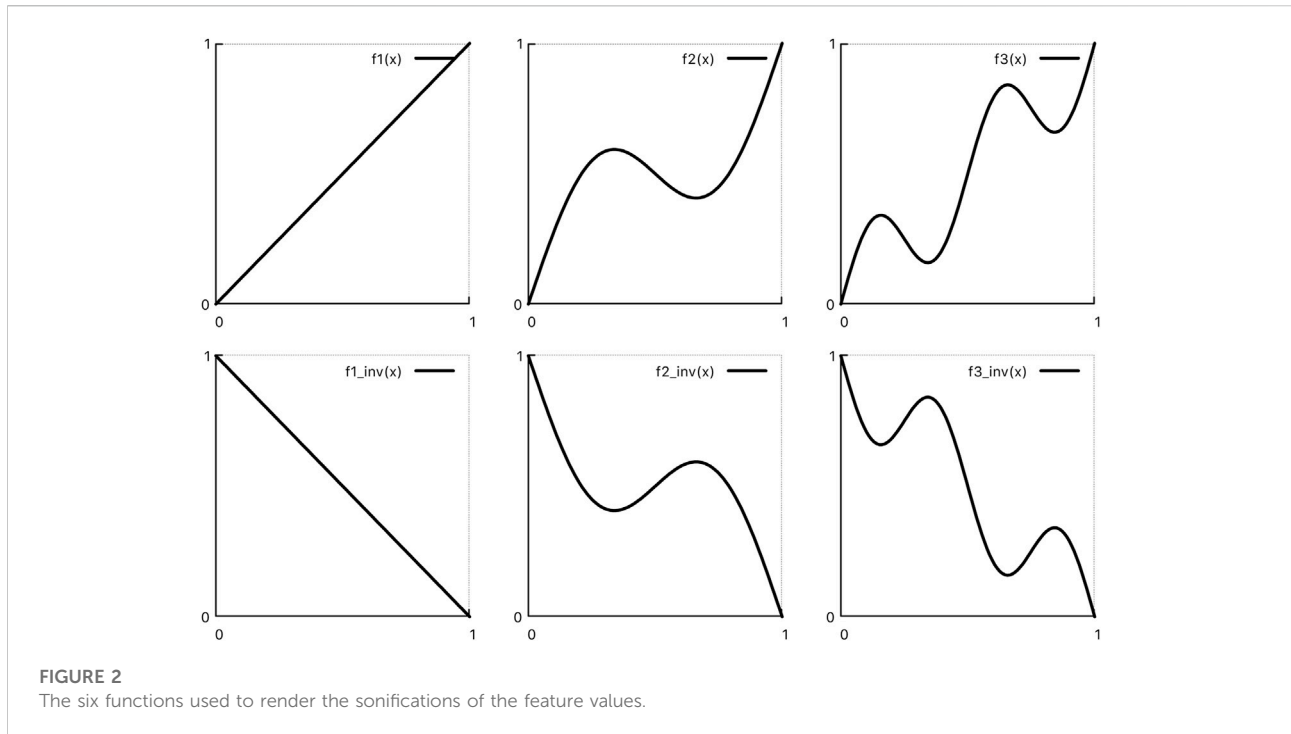


FIGURE 2
The six functions used to render the sonifications of the feature values.

For the diegetic sound set, the sound designer brief was specific. For radiation, the auditory representation was required to emulate the sound of a Geiger counter, one of the most notable sonification devices (Rutherford and Geiger, 1908). For proximity, the *pulsed tone sonification* technique should be used to convey distance in the manner of parking assistance systems found in cars (Plazak et al., 2017). For load, the sonification should emulate the sound of a motor with differing levels of strain.

As the design of the “abstract” sounds did not call on established auditory representations, the workshop participants and facilitators completed an (online) sticky note exercise to rapidly generate ideas for how these data features could sound. The sound designer then worked independently for the remainder of the day before presenting their initial sonifications to the rest of the group.

To give the sound designer some feedback on the efficacy of their sonifications before the second workshop, an online survey was conducted with 40 participants (18 Male, 22 Female, Age $M25.41 \pm 5.52$), recruited via Prolific Academic, and compensated £2.45. The aim of the survey was to identify whether the sonification was correctly identified as the feature it represents (perceptual congruence), and to gauge how precisely the value of data feature could be identified from the sound (transparency). Each sonification was rendered as the underlying data feature was varied as a function of x as it increased linearly from 0.0 to 1.0 over a period of 10 s. Three functions were defined represented by the equations $f_1 = x$, $f_2 = x + 0.3\sin(2\pi x)$ and $f_3 =$

TABLE 1 Table showing the proportion of data features that the participants associated with each abstract sound.

Sound	Identified as...		
	Radiation	Load	Proximity
Radiation	14%	24%	62%
Load	30%	32%	38%
Proximity	38%	20%	42%

$0.2\sin(4\pi x)$, and a further three functions were generated by inverting these curves, where, for example, the inversion of f_1 was taken as $1 - f_1$. The resulting curves are shown in Figure 2.

Rendering the sonifications for each curve resulted in a total of 18 audio files which were presented to the participants in a randomised order. Participants could listen to the sound as many times as they liked as they responded to three questions. Q1 was a multiple choice question asking the participant select the data feature they felt the sound best portrayed (radiation, load or proximity). In Q2, participants were asked to explain their choice for the answer to Q1. Q3 comprised a six part multiple choice question where participants were asked to select the curve from Figure 2 that most accurately represented how the sound was perceived to change as a function of time. From the answers to Q3 it was possible to identify whether the general trend of the function was identified correctly (i.e., descending for the inverted curves and ascending otherwise), as well as the specific function.

TABLE 2 Table showing the proportion of participants correctly identifying the trend and specific function from the abstract sounds.

Sound	Trend (%)	Function (%)
Radiation	80	40
Load	80	38
Proximity	81	37

TABLE 3 The mean distance from the target location for each of the data parameters. Hence, the lower the distance the better the set performed.

Data feature	M distance from max \pm SD
Load	0.0094 \pm 0.08
Contamination	0.087 \pm 0.11
Proximity	0.094 \pm 0.15

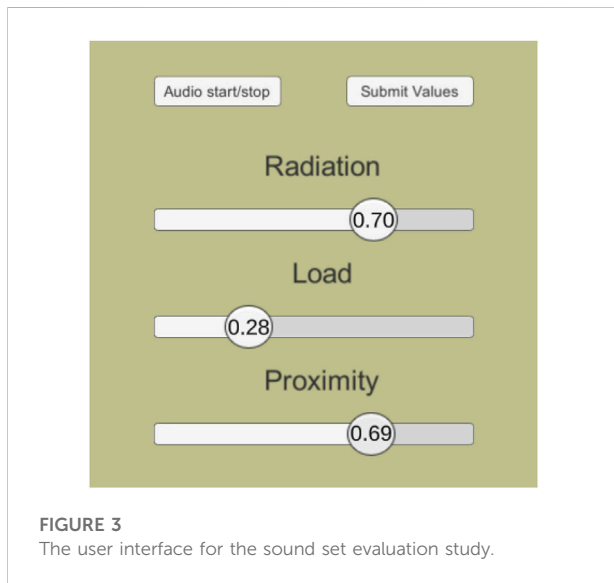


FIGURE 3
The user interface for the sound set evaluation study.

Results are shown in [Tables 1, 2](#) which, along with the qualitative written responses to Q2, were shared with the sound designer in advance of the second workshop to inform the second iteration of their sonification designs.

In the second workshop the survey results were reviewed, and the sound designer was then tasked with combining and enhancing their work so that the three data features, and their corresponding sounds, worked together and could be conveyed simultaneously. In the remaining time the sound designer produced both the diegetic and abstract sound sets that would be used in our subsequent study. The results in [Tables 1, 2](#) were purely intended to provide feedback on the initial sonification attempt. Following the discussion the sound designer was left to interpret and incorporate this information into their final sonification. Most notable from the results was that the radiation sound was identified as proximity 62% of the time. Interestingly, at the end of the workshop the sound designer commented that they “*went with it*” and used this sound to represent proximity instead.

Please see the [Supplementary Materials](#) for videos of the diegetic (Video 1) and abstract sound sets (Video 2) produced at the end of the workshop.

4.1 Abstract sound set evaluation

To ensure the abstract sounds created in the second workshop transparently represented the underlying data features, we ran a short study with 18 participants (13 male, 5 female, age 29 ± 2.3 [$M \pm SD$]).

A desktop application was built in Unity consisting of three sliders (range 0–1), one for each data type which together represented a position in 3D space ([Figure 3](#)). For each sound set, a peak value for the data was randomly assigned a location for each data type with the data feature value decreasing with distance from this peak according to a Gaussian function. For each sound set, all three sounds played simultaneously, starting with the slider at position 0, so that each data type was initially at a random level dependent on where the peak was located. Participants were tasked with finding the peak for all three data types simultaneously, and to click a button when they thought all three sliders were in the right place. Distance to the peak location was logged to a text file. The distance to the peak value of all three data types was calculated, and used to calculate a mean distance score of 0.092 with a standard deviation of 0.06, demonstrating that participants were able to consistently find the target within approximately 10%. The data for all three data types is shown in [Table 3](#).

5 Virtual reality system design

To evaluate and compare the “diegetic” and “abstract” sonifications, we have developed a simulated VR teleoperation system to control a robot arm for pick and place operations. This task models the nuclear decommissioning activity of sorting and segregation of nuclear waste. Designated target objects must be moved to target zones, avoiding radiation hazards and collisions with the environment.

The VR environment and teleoperation system was built using the Unity game engine. Assets were utilised that include a model of the Franka-Panda robot arm, industrial textures and objects, and the BioIK inverse kinematics solver³. Our VR

³ <https://assetstore.unity.com/packages/tools/animation/bio-ik-67819>.

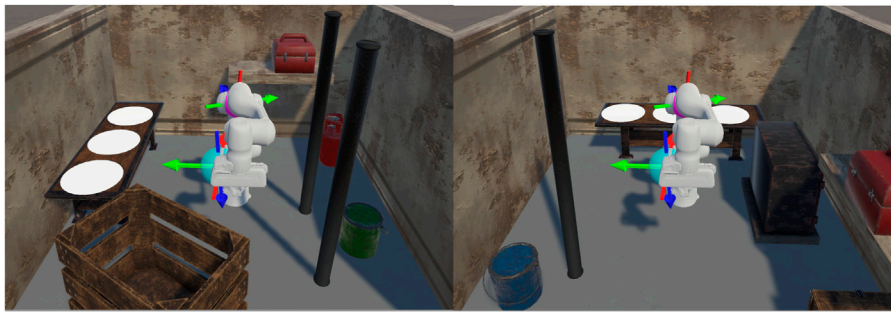


FIGURE 4

The two experimental scenes. The robot control targets are the coloured spheres, one for the end effector in cyan, and one just visible at the “elbow” in magenta.

application is displayed on an HTC Vive Pro HMD, and utilises the Vive controllers for robot teleoperation.

5.1 The VR teleoperation system

The robot arm teleoperation system has been designed to be easily used, such that a novice user could carry out the defined tasks after a short period of practice. Further, an intuitive control system aims to minimise the cognitive load of the operator, such that we can more effectively evaluate the impact of data sonification on the cognitive load of task performance and data understanding.

Our control scheme uses an inverse kinematics solver to configure the robot to reach two targets, one for the end effector and one for a central link (Figure 4). The operator moves the target points by holding down the trigger on the VR controllers where one hand controls each target. The targets copy the relative motion of the users hands but only when the triggers are pressed, further, the central target is only utilised when the corresponding trigger is pressed. This clutched motion facilitates easy movement of the targets within the whole environment, and allows the user to locate themselves anywhere in the environment while maintaining a clear one-to-one spatial mapping of where the targets will move. This allows the user to move around the environment as needed to facilitate good visibility of target objects and obstacles. The central target is present to allow the user more control over the configuration of the robot, to aid in obstacle avoidance, and to better align the end effector with objects to be grasped.

Objects to be grasped are outlined in red, and this outline goes green when the fingers of the gripper overlap the object, to indicate the object may be grasped. Grasping/dropping of objects is done with a single button press, for simplicity of control, grasp points and grasp physics are not modelled.

To try and prevent object overlap collision avoidance is implemented. When a collision is detected between the robot

(or its payload if one is attached) and an obstacle, the end effector target point is shifted a short distance away from the collision along its normal. While this does not prevent all overlaps, this method models realistic behaviour without compromising the control scheme.

As the control scheme is not the focus of this paper, the intuitiveness of it was informally evaluated with a few pilot participants. All pilot participants were able to successfully complete the pick and place task after a short practice period.

5.2 Visual and auditory data representation

The data features identified in the earlier workshops with the teleoperation experts from Sellafield Ltd. were implemented into the VR system. Radiation is measured as the highest level of radiation the robot is exposed to at a given moment, hence, it is determined by the point on the robot closest to a radiation source. The proximity measure is how close the robot is to the nearest obstacle, like radiation it is measured as the closest point on the robot to any obstacle, the closer an obstacle the higher the proximity. Load on the robot is calculated as $\text{payload_weight} * \text{distance from robot base}$.

For auditory data representation we sonified the above data features. To do so we used the diegetic and abstract sound sets that were produced by the sound designer during the second co-design workshop (Section 4).

For the diegetic sound set we used: a Geiger counter for radiation, where click frequency increases with radiation; a parking sensor beep for proximity, where beep frequency increases with proximity; an electric motor noise for load on the arm, as force increases volume and motor strain increase. By using established data to sound mappings we could be reasonably sure that the data features could be interpreted with minimal training. This approach to sound selection follows the principles of perceptually congruent sound design similar to that found in the literature

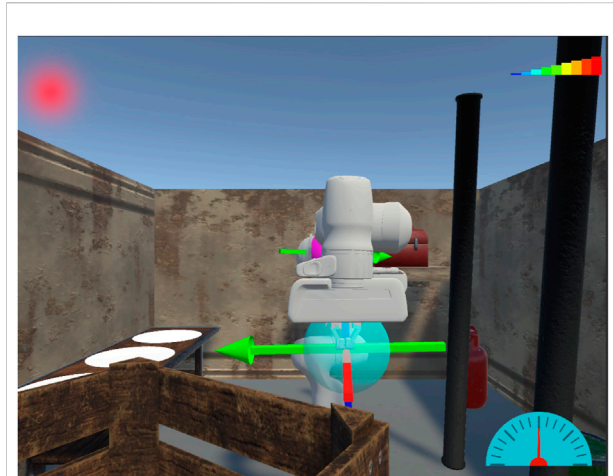


FIGURE 5

The heads up display elements. Proximity is displayed using the red dot in the top left, the maximum value is displayed. Radiation is displayed using the coloured bars in the top right, the maximum value is displayed. Load is displayed on the gauge in the bottom right, a middle value is shown.

(Walker, 2002; Walker and Kramer, 2005; Ferguson and Brewster, 2019).

For the abstract sound set we used: for radiation, a vibrato whine overlaying rhythmic clicks, as radiation increased the clicks decreased in frequency and the volume and pitch of the whine increased; for proximity, ambient musical chimes which increased in frequency and volume as proximity increased; for load, a whirring sound with a background rhythmic beat, where volume and pitch, and rhythmic beat frequency increase as load increased. To clarify the sounds that were used, a video is available in the [Supplementary Materials](#) that shows the data levels changing over time for each sound. The aim of the sound design was to have unexpected but transparent representations of the data, that diverged from the diegetic sound set.

As a comparative condition, we developed a visual representation of the data features, i.e., a Heads-Up-Display (HUD) of graphical elements that would overlay the users vision of the VR environment (Figure 5). Proximity is displayed using the red dot in the top left, as proximity increases the red dot expands. Radiation is displayed using the coloured bars in the top right, as radiation increases more of the bars become visible from left to right. Load is displayed on the gauge in the bottom right, as force increases the needle moves from left to right.

As with the control scheme, our data presentation choices were informally evaluated by pilot participants. They reported that they were able to easily understand the approximate data levels for the three data features, both with the data presented as sound and using the HUD.

6 User study evaluation

To evaluate our proposed hypotheses, we conducted a mixed design user study with data representation of ancillary data as our within subjects variable (visual or audio) and sound set (diegetic or abstract) as our between subjects variable. 40 participants (29 male, 10 female, 1 non-binary, age $M32.9 \pm SD8.71$) took part. Our participant pool was drawn principally from staff and students in the Bristol Robotics Lab, with approximately one third of participants from the wider university population. Across the whole sample there was a broad range of VR experience, though for the majority their experience was minimal. Participants were assigned pseudo-randomly to the diegetic or abstract sound condition such that 20 were assigned to each. The task took approximately 30 min to complete, and participants were remunerated with a £5 voucher. The study was approved by the University of the West of England ethics committee.

6.1 Materials

The study consisted of three VR environments: a tutorial environment, where participants could learn how to operate the robot and understand the data features; two experimental scenes (5) designed to be as similar as possible in terms of task completion difficulty, and obstacle (physical and radiation) avoidance. Each experimental scene had two invisible spherical radiation zones, radiation level is at a maximum at the centre of the sphere, and decreases with distance from the centre. Each experimental scene also had three objects of different weights that had to be transported to target zones.

Task performance data was logged automatically using a logging script in Unity, questionnaire data was recorded using the Qualtrics⁴ survey platform.

6.2 Procedure

Participants were tasked with moving designated target objects on to target zones on a table in the environment. The target zones turned from white to green when an object was placed on them to show the objective was complete.

Prior to the experimental trials participants undertook a tutorial to familiarise themselves with operating the robot, the data feature representations, and the tasks they were expected to complete. The tutorial consisted of a series of pre-recorded voice-overs that were triggered using the VR controller. Participants were able to experiment within the tutorial environment between

⁴ <https://www.qualtrics.com>.

each voice-over to ensure they understood all the necessary information. In the tutorial the data features were presented with both the HUD and as sound, two target objects of different weights were utilised in the tutorial; additionally, a zone of radiation was made visible to facilitate understanding. Hence, participants were able to understand how each of the data features varied with different data levels through experience in the tutorial.

Each experimental trial was a VR scene with three objects to be moved to target zones. Participants undertook two trials one for each data presentation condition (i.e., one visual and one audio). Condition scene assignment and condition ordering was randomised between participants. When all three objects had been moved to the target zones an “Objectives Completed” message was displayed, performance data was logged, and the scene exited. Prior to each experimental trial participants were reminded that they should move the objects to the target zones as quickly as they are able while remaining safe, i.e., avoiding collisions and radiation. They were also instructed that collisions were more problematic when grasping heavier objects. After each experimental scene, participants completed a questionnaire.

6.3 Measures

We have used a variety of objective and subjective measures to evaluate the data presentation approaches.

6.3.1 Performance measures

The objective measures are based on task performance. A Unity script records the following measures:

- Collision penalty: cumulative value incremented by $1 + \text{payload_weight}$ each time a collision occurs, hence higher weight collisions have a larger effect on the penalty score.
- Radiation exposure: cumulative radiation value, added to every frame, hence the longer and more intense the exposure the higher the value.
- Distance penalty: cumulative value incremented every frame with the distance of the main control point from the end effector, moving the control point faster and further results in less control of the robot motion.
- Task completion time.

6.3.2 Questionnaires

There are a number of subjective measures of participant experience that we have used to evaluate the impact of data sonification. The System Usability Scale (SUS) (Sauro, 2011) was used to evaluate differences in usability. The IGroup Presence Questionnaire (IPQ) evaluates presence and immersion in the VR environment (Schubert et al., 2001). The NASA TLX (Hart and Staveland, 1988) evaluates user workload during the tasks.

TABLE 4 Questions in the stress/comfort questionnaire. Rated on a 5 point Likert scale from *Strongly Disagree* to *Strongly Agree*.

Label	Statement
Stress1	I felt relaxed throughout the exercise
Stress2	I found the experience stressful
Stress3	I found aspects of the experience frustrating
Stress4	At times I felt unsafe during the exercise

We also wrote a set of questions to evaluate user comfort and stress levels during task performance, Table 4. The majority of questions were assessed using a 5 point Likert scale (strongly disagree → strongly agree), some questions were rephrased to fit this answer pattern. NTLX was assessed using a nine point Likert scale, and weighting of the different elements was established by having participants rank the features.

7 Results

As the study was a mixed design a two-way mixed ANOVA should be appropriate for analysis. However, almost all measures failed the Shapiro-Wilk test for normal distribution of residuals, and have some outliers. Hence, we have analysed our data using the **bwtrim** R package that calculates a two-way mixed ANOVA with trimmed means that is robust to both non-normality and outliers.

7.1 Questionnaire data

The questionnaire data was processed according to the guidelines in the literature from which they are drawn. Results are presented in Table 5.

For the IPQ four factors are defined: *sense of being there* (PRES), *spatial presence* (SP), *involvement* (INV) and *experienced realism* (REAL). Each factor is a mean of the related items from the instrument. Statistical analysis is reported in Table 6. None of the factors were statistically significant.

The SUS results are calculated by summing the likert scale values (scored from 0–4), and multiplying by 2.5. This gives a score between 0 and 100 with the *de facto* score of 68 for a good system. All conditions fall just below the “good” threshold. Statistical analysis is reported in Table 7. No significant effects were found.

The NASA TLX results are calculated as a weighted mean of all six factors. Factors are weighted according to the order of importance for each participant, ranking the six factors gives weights from 5 (most important) to 0 (least important). The workload score was calculated by Formula 1. Where W_n is the weight assigned to the n th feature, F_n . Statistical analysis is reported in Table 8. There was a significant interaction effect ($p = 0.0492$):

TABLE 5 Questionnaire Data. Each result is reported as M ± SD.

Measure	Visual diegetic	Sound diegetic	Visual abstract	Sound abstract
IPQ PRES	4.25 ± 0.44	4.3 ± 0.47	3.9 ± 0.97	4.35 ± 0.59
IPQ SP	4.08 ± 0.42	4.05 ± 0.6	4.09 ± 0.57	4.13 ± 0.67
IPQ INV	3.85 ± 0.61	3.98 ± 0.62	3.6 ± 0.66	3.75 ± 0.64
IPQ REAL	3.0 ± 0.8	2.98 ± 0.58	2.87 ± 0.76	2.92 ± 0.76
SUS	64.38 ± 16.6	66.0 ± 17.91	64.38 ± 16.16	61.75 ± 21.93
NTLX	12.46 ± 3.31	11.84 ± 2.95	11.26 ± 3.29	13.26 ± 2.8
Stress1	3.75 ± 1.02	3.65 ± 1.04	3.5 ± 1.0	3.05 ± 1.39
Stress2	1.9 ± 0.79	1.9 ± 0.97	2.45 ± 1.1	2.8 ± 1.32
Stress3	2.85 ± 1.14	3.0 ± 1.17	3.35 ± 1.04	3.65 ± 1.14
Stress4	1.1 ± 0.45	1.0 ± 0.0	1.2 ± 0.52	1.35 ± 0.67

TABLE 6 Statistical analysis of IPQ data.

Effect	PRES		SP		INV		REAL	
	value	p	value	p	value	p	value	p
Main Effect of Data Representation	1.544	0.229	0.600	0.450	2.833	0.107	0.342	0.565
Main Effect of Sound Set	0.156	0.697	0.003	0.958	3.634	0.070	0.001	0.978
Interaction Effect of D. Rep*S. Set	0.386	0.542	0.007	0.932	0.023	0.880	0.071	0.793

TABLE 7 Statistical analysis of SUS data.

Effect	Value	p
Main Effect of Data Representation	0.304	0.588
Main Effect of Sound Set	0.046	0.833
Interaction Effect of D. Rep*S. Set	0.045	0.834

TABLE 8 Statistical analysis of NTLX data.

Effect	Value	p
Main Effect of Data Representation	0.482	0.495
Main Effect of Sound Set	0.289	0.597
Interaction Effect of D. Rep*S. Set	4.335	0.049

Significant results shown in bold.

workload decreased in the diegetic sound condition compared to visual data, while it increased with the abstract sounds compared to the visual data. This effect can be observed in Figure 6.

$$Workload = \sum_{n=1}^6 W_n F_n \tag{1}$$

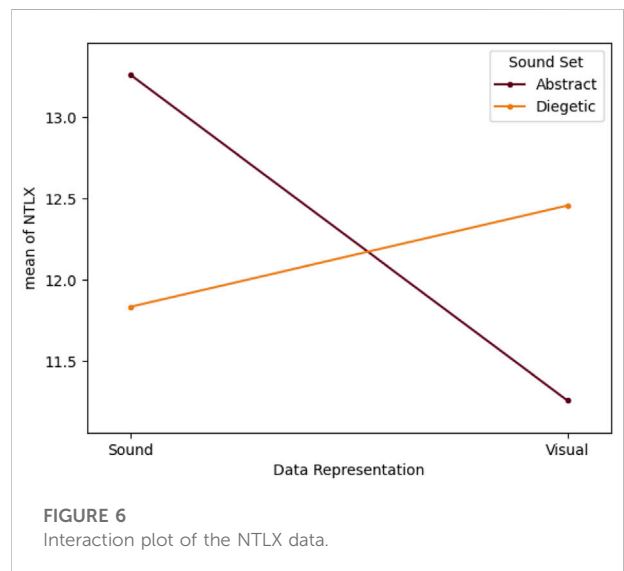


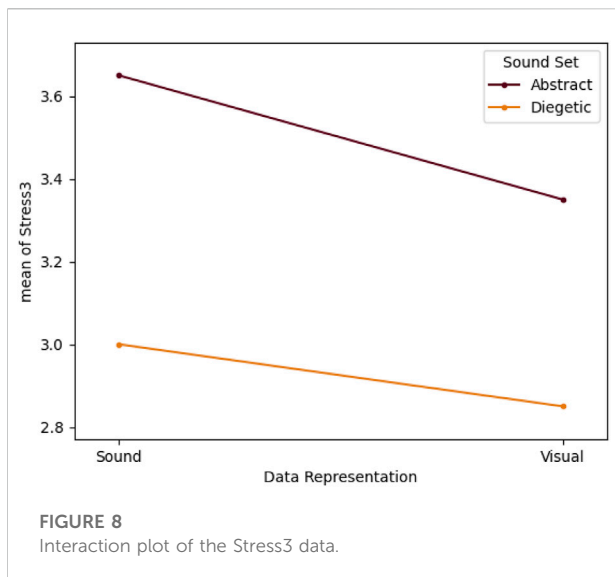
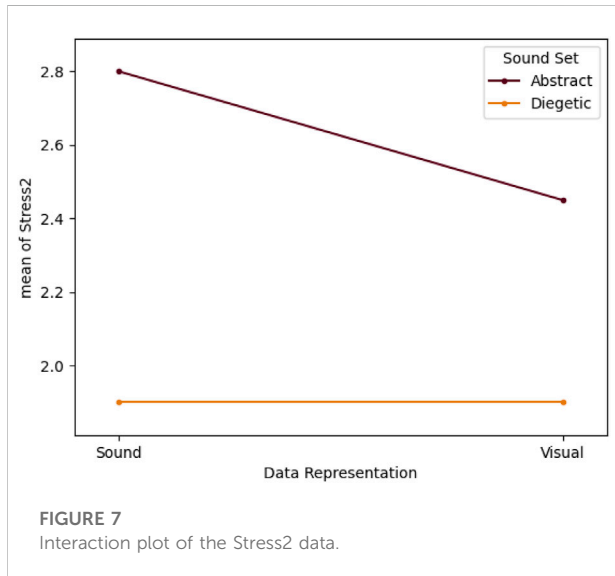
FIGURE 6 Interaction plot of the NTLX data.

As the stress/comfort questions were not from a validated questionnaire, and are few in number, each question was tested separately. Statistical analysis is reported in Table 9. A significant main effect of sound set was found for Stress2 and Stress3. These effects can be observed in Figures 7, 8.

TABLE 9 Statistical analysis of stress data.

Effect	Stress1		Stress2		Stress3		Stress4	
	value	<i>p</i>	value	<i>p</i>	value	<i>p</i>	value	<i>p</i>
Main Effect of Data Representation	1.355	0.261	0.056	0.816	0.575	0.457	0.244	0.631
Main Effect of Sound Set	2.524	0.131	5.412	0.032	6.377	0.023	0.244	0.631
Interaction Effect of D. Rep*S. Set	0.691	0.418	1.402	0.253	0.144	0.708	0.244	0.631

Significant results shown in bold.



7.2 Performance measures

No main effects or interaction effects were found to be significant for any of the performance measures. The main effect of data representation approached significance for distance penalty ($p = 0.06$). Results are reported in Table 10. Statistical analysis is reported in Table 11.

8 Discussion

The results from the study overall refuted or failed to confirm many of our hypothesis, offering valuable insights into the under-explored interplay between virtual reality, teleoperation and sonification.

Based upon Sigrist et al. (2015), we predicted that the use of sonification would divert information load from the operator’s visual system to their auditory system, thus improving usability (H1a) and reducing workload (H1b). Their findings also led to the hypothesis that sonification would also have a positive effect on task performance (H1c). These hypotheses were largely refuted by our results, with no main effect for data representation on the SUS (usability), NTLX data (workload) or task performance measures.

The main reason for the workload finding was that the diegetic sounds were shown to reduce workload in a way that the abstract sounds did not. This may suggest that the overall impact of sonification on workload was masked by an increase in workload in the abstract sound condition.

Whilst auditory data representation had no discernible impact on usability, it is worth noting that the usability in all conditions was close to the threshold for a good system (SUS score > 68).

Additionally, whilst in general there was no main effect for data representation on the task performance data, the main effect for data representation on the distance penalty measure approached significance ($p = 0.06$), this suggest some support for hypothesis H1c, but does not confirm it. The distance penalty

TABLE 10 Task performance data. Each result is reported as $M \pm SD$.

Measure	Visual diegetic	Sound diegetic	Visual abstract	Sound abstract
Collision P	141.93 \pm 60.31	163.81 \pm 77.0	192.76 \pm 254.69	177.73 \pm 129.67
Radiation P	817.53 \pm 355.01	933.31 \pm 482.55	1104.41 \pm 799.79	1051.34 \pm 556.63
Distance P	44.96 \pm 24.33	49.31 \pm 30.22	52.71 \pm 52.84	55.5 \pm 32.56
Task Time	195.61 \pm 61.99	190.69 \pm 61.21	218.77 \pm 167.83	216.19 \pm 90.65

TABLE 11 Statistical analysis of performance measure data.

Effect	Collision P		Radiation P		Distance P		Task time	
	value	<i>p</i>	value	<i>p</i>	value	<i>p</i>	value	<i>p</i>
Main Effect of Data Representation	1.849	0.190	3.950	0.060	1.442	0.249	1.562	0.226
Main Effect of Sound Set	0.119	0.735	0.019	0.892	1.624	0.249	0.024	0.879
Interaction Effect of D. Rep*S. Set	0.014	0.909	1.459	0.240	0.028	0.869	0.008	0.929

is higher when the data is represented using sound, indicating that in the sonification conditions, participants moved the main control point further and faster. One possible explanation for this is that participants felt they had a clearer perception of the dangers (collisions and radiation) and trusted the robot to make larger control motions when they perceived that the robot to be safe.

Based on Nunez (2004), Walker (2002), Walker and Kramer (2005), and Ferguson and Brewster (2018), we predicted that sounds with more “perceptual congruence” with the environment, i.e., sounds that met operator’s expectations with familiar sonic metaphors would improve usability (H3a) and reduce workload (H3b). This prediction was accurate in relation to workload. As mentioned above, the diegetic sound condition resulted in a reduced workload, whereas the abstract sound condition resulted in a higher workload. This may be because people have a familiarity with the meaning of the diegetic sounds and can apply them more readily to the task at hand, allowing them to benefit from the distribution of salient information between visual and auditory stimuli. In the case of the abstract sounds, participants may have had to work harder in the first instance to infer meaning from the unfamiliar sounds. Our H3a assumption that diegetic sounds would similarly improve usability was not supported as there was no interaction effect for the SUS data.

Hypothesis H4b proposed that the abstract sounds would be less stressful than the diegetic sounds. This built on the concerns raised by Madary and Metzinger (2016) that a heightened sense of presence in spaces that would cause stress in the physical world could cause operators similar issues in the virtual environment. We predicted that if abstract sounds could be used to lessen the operator’s sense of presence in a potentially stress-inducing

scenario, abstract sounds would induce less stress than the diegetic sounds. Further, as the diegetic sounds are designed to model established data to sound mappings relating to hazards, they are implicitly designed to be stressful (Ferguson and Brewster, 2018).

The results directly contradict our hypothesis, with users who experienced the abstract sound condition reporting higher levels of stress and frustration than those who experienced the diegetic sounds. It is plausible that the diegetic sounds are less stressful than abstract sounds. An alternative explanation, based on our findings on workload, is that frustration and stress may be attributed to the additional effort of learning, and making sense of, new auditory representations that do not use established reference points such as the Geiger counter or parking sensor.

The results may also be explained by the semi-artificial nature of our user study. Whilst we made participants aware that we were testing a virtual reality control system for a robotic teleoperation system within a nuclear decommissioning context, they were aware that this was a simulation only, and that their actions and errors would have no real-world consequences.

Based on the literature, our hypothesis about the cause of stress or otherwise in this experience was linked to the sensation of being embodied in an unpredictable, hazardous, and contaminated environment. Something not possible to fully realise in this context. Further study, digitally coupling the virtual reality simulation with an active robot, would be a useful step towards understanding the impact of different sonification approaches on operator’s comfort and stress levels.

For our remaining hypotheses we were particularly interested in factors that might modulate presence, and, based on Slater

(2009), Nunez (2004) and Walker (2002), hypothesised that adding sonification would increase presence (H2). Further, diegetic sounds would have a stronger effect in increasing the user's sense of presence (H3c), whereas less congruous sensory information such as our abstract sound condition might serve to reduce an operator's sense of presence (Slater and Steed, 2000; Nunez, 2004; Sanchez-Vives and Slater, 2005) (H4a). We additionally predicted that the lessening of presence could be achieved without negatively impacting task performance (H4a). If found to be correct, and with further study this could have signalled a means of insulating operators from the risks of long-term immersion such as depersonalization and derealization dissociative disorders (Madary and Metzinger, 2016; Spiegel, 2018).

Interestingly, there was no effect of adding sound on the IPQ data (presence measures), moreover, there was no interaction effect either, refuting all three hypotheses related to the impact of sonification on presence (H2, H3c, H4a). Adding sound to our VR environment did not increase presence, as we might expect from the literature (Cooper et al., 2018), which could suggest that sonification, unlike sounds that are perceived to originate from the environment, does not increase presence. The fact we found no changes in presence for either sound set suggests that the diegetic sounds were not sufficiently congruent to increase presence, and the abstract sounds were not sufficiently incongruent to decrease the sense of presence. An alternative explanation is that the auditory mappings used for sonification does not have a significant bearing on presence in our use case. This explanation marries up with our finding there is no impact of sonification on task performance in immersed teleoperation as we expected, in part due to the findings in the literature that increased presence results in increased task performance (Cooper et al., 2018). Given the potential range of sonification approaches as yet unexplored, and the limited scale of this study, we would recommend that significant further research be undertaken to better understand this dynamic.

9 Conclusion

In this paper we share a novel, interdisciplinary methodology for evaluating the impact of data sonification on presence, task performance, workload and comfort in immersed teleoperation, key to understanding its performance in this context. Further, we demonstrate the implications of different sonification sound design choices on these metrics. Our findings demonstrate the improved utility of using diegetic sounds for sonification due to their positive impact on stress and workload. Additionally, our findings suggest that sonification does not impact presence as we might expect from the engagement of an additional sensory modality.

These outcomes have implications for the design of data sonification for use in simulated hazardous environments, suggesting that diegetic sounds are preferable to abstract sonifications. However, this work only compared one abstract sound set with diegetic sounds, more work is required investigating different sound design choices and design methodologies.

The persistence and resilience of operator presence observed in this study, seemingly immune to our attempts to influence it with different visual and auditory conditions has wider significance. Researchers and designers seeking to reduce presence may need to consider different or stronger measures to overcome the tenacity of presence in immersed teleoperation scenarios.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by University of the West of England ethics committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

PB, VM, and TM contributed to the workshop, study design, implementation and preparation of the paper. PB built the VR simulation and ran the main user study.

Funding

This work was funded by the Robotics for Nuclear Environments (EPSRC grant number EP/P01366X/1), Isomorph (Apex Award Grant No. APX/R1/180118), the University of the West of England VCIRCF and virtually there (TAS, UKRI).

Acknowledgments

We would like to thank Mandy Rose, Joe Mott, Lisa Harrison, Stephen Jackson, Alex Jones, Helen Deeks, Joseph Hyde, David Glowacki, Oussama Metatla as well as members

of the Bristol VR lab, Creative Technologies Lab and the Bristol Robotics Lab for their advice, discussions and participation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

References

- Bailenson, J. (2018). *Experience on demand: what virtual reality is, how it works, and what it can do*. First edn. New York: W.W. Norton Company.
- Bowman, D. A., and McMahan, R. P. (2007). Virtual reality: How much immersion is enough?. *Comput.* 40 (7), 36–43. doi:10.1109/MC.2007.257
- Bramas, B., Kim, Y. M., and Kwon, D. S. (2008). "Design of a sound system to increase emotional expression impact in human-robot interaction," in 2008 International conference on control, automation and systems, Seoul, South Korea, 2008, 2732–2737. ICCAS 2008. doi:10.1109/ICCAS.2008.4694222
- Brewster, S. (2002/2002). Overcoming the lack of screen space on mobile computers. *Personal Ubiquitous Comput.* 6 (3), 188–205. doi:10.1007/S007790200019
- Carlile, S. (2011). "Psychoacoustics," in *The sonification handbook* (Berlin, Germany: Logos Publishing House), 41–61. chap. 3.
- Cooper, N., Milella, F., Pinto, C., Cant, I., White, M., Meyer, G., et al. (2018). The effects of substitute multisensory feedback on task performance and the sense of presence in a virtual reality environment. *PLoS One* 13, e0191846. doi:10.1371/JOURNAL.PONE.0191846
- Cummings, J. J., and Bailenson, J. N. (2015). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychol.* 19, 272–309. doi:10.1080/15213269.2015.1015740
- de Campo, A. (2007). "Toward a data sonification design space map," in Proceedings of the international conference on auditory display, Montreal, Canada, 2007.
- Di, G. Q., Chen, X. W., Song, K., Zhou, B., and Pei, C. M. (2016). Improvement of Zwicker's psychoacoustic annoyance model aiming at tonal noises. *Appl. Acoust.* 105, 164–170. doi:10.1016/J.APACOUST.2015.12.006
- Droumeva, M., and Wakkary, R. (2006). "The role of participatory workshops in investigating narrative and sound ecologies in the design of an ambient intelligence audio display," in Proceedings of the international conference on auditory display, London, United Kingdom (London, UK: Georgia Institute of Technology).
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS One* 8, e82491. doi:10.1371/JOURNAL.PONE.0082491
- Ferguson, J., and Brewster, S. A. (2018). "Investigating perceptual congruence between data and display dimensions in sonification," in Conference on human factors in computing systems, Montreal, QC, Canada, 2018-April. doi:10.1145/3173574.3174185
- Ferguson, J., and Brewster, S. (2019). "Modulating personal audio to convey information," in Conference on human factors in computing systems - proceedings, Glasgow, Scotland, United Kingdom (Association for Computing Machinery). doi:10.1145/3290607.3312988
- Frauenberger, C., Stockman, T., Putz, V., and Höldrich, R. (2006). "Design patterns for auditory displays," in *People and computers XIX—The bigger picture*. London: Springer. doi:10.1007/1-84628-249-7_30
- Frid, E., Moll, J., Bresin, R., and Sallnäs Pysander, E.-L. (2019). Haptic feedback combined with movement sonification using a friction sound improves task performance in a virtual throwing task. *J. Multimodal User Interfaces* 13, 279–290. doi:10.1007/s12193-018-0264-4
- Fuller, A., Fan, Z., Day, C., and Barlow, C. (2020). Digital twin: enabling technologies, challenges and open research. *IEEE Access* 8, 108952–108971. doi:10.1109/ACCESS.2020.2998358
- García, J. C., Patrão, B., Almeida, L., Pérez, J., Menezes, P., Dias, J., et al. (2017). A natural interface for remote operation of underwater robots. *IEEE Comput. Graph. Appl.* 37, 34–43. doi:10.1109/MCG.2015.118
- Goudarzi, V., Vogt, K., and Höldrich, R. (2015). "Observations on an interdisciplinary design process using a sonification framework," in Proceedings of the 21st international conference on auditory display, Graz, Styria, Austria, 2015.
- Grassini, S., Laumann, K., and Rasmussen Skogstad, M. (2020). The use of virtual reality alone does not promote training performance (but sense of presence does). *Front. Psychol.* 11, 1743. doi:10.3389/fpsyg.2020.01743
- Grond, F., and Berger, J. (2011). "Parameter mapping sonification," in *The sonification handbook* (Berlin, Germany: Logos Publishing House).
- Grond, F., and Sound, T. H. (2014). Interactive Sonification for Data Exploration: how listening modes and display purposes define design guidelines. *Org. Sound.* 19, 41–51. doi:10.1017/S1355771813000393
- Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi:10.1016/S0166-4115(08)62386-9
- Hermann, T., Niehus, C., and Ritter, H. (2003). "Interactive visualization and sonification for monitoring complex processes," in Proceedings of the 2003 international conference on auditory display, Boston, MA, United States, 247–250.
- Kelly, J. W., Doty, T. A., Ambourn, M., and Cherep, L. A. (2022). Distance perception in the oculus quest and oculus quest 2. *Front. Virtual Real.* 3, 850471. doi:10.3389/frvir.2022.850471
- Kramer, G. (1994). "Some organizing principles for representing data with sound," in *Auditory display: sonification, audification and auditory interfaces* (Reading, MA: Addison-Wesley), 285–221.
- Kramer, G., Walker, B., Bonebright, T., Cook, P., and Flowers, J. (2010). *Sonification report: status of the field and research agenda 444*. University of Nebraska, Lincoln: Faculty Publications.
- Latupeirissa, A. B., Panariello, C., and Bresin, R. (2020). "Exploring emotion perception in sonic HRI," in Proceedings of the sound and music computing conferences, Torino, Italy, 2020, 434–441.
- Lee, K. M. (2004). Presence, explicated. *Commun. Theory* 1, 27–50. doi:10.1111/j.1468-2885.2004.tb00302.x
- Lokki, T., and Gröhn, M. (2005). Navigation with auditory cues in a virtual environment. *IEEE Multimed.* 12, 80–86. doi:10.1109/MMUL.2005.33

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2022.904720/full#supplementary-material>

SUPPLEMENTARY VIDEO S1

Diegetic sonification sound set.

SUPPLEMENTARY VIDEO S2

Abstract sonification sound set.

- MacDonald, D., and Stockman, T. (2018). "SoundTrAD, A method and tool for prototyping auditory displays: can we apply it to an autonomous driving scenario?," in International conference on auditory display, Michigan, United States (Michigan Technological University: International Community for Auditory Display), 145–151. doi:10.21785/ICAD2018.009
- [Dataset] Madary, M., and Metzinger, T. K. (2016). Real virtuality: a code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. *Front. Robot. AI* 3, 3. doi:10.3389/frobt.2016.00003
- Makransky, G., and Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): a theoretical research-based model of learning in immersive virtual reality. *Educ. Psychol. Rev.* 33, 937–958. doi:10.1007/s10648-020-09586-2
- Meijer, P. B. (1992). An experimental system for auditory image representations. *IEEE Trans. Biomed. Eng.* 39, 112–121. doi:10.1109/10.121642
- Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented reality: a class of displays on the reality-virtuality continuum. *Telemanipulator Telepresence Technol. (SPIE)* 2351, 282–292. doi:10.1117/12.197321
- Murray, J. (1997). *Hamlet on the holodeck: the future of narrative in cyberspace*. New York: Free Press.
- Nunez, D. (2004). "How is presence in non-immersive, non-realistic virtual environments possible?," in Proceedings of the 3rd international conference on computer graphics, virtual reality, visualisation and interaction in africa Language: (ACM), Stellenbosch, South Africa, 83–86.
- Pirhonen, A., Murphy, E., McAllister, G., and Yu, W. (2006). "Non-speech sounds as elements of a use scenario: a semiotic perspective," in Proceedings of the international conference on auditory display, London United Kingdom (London, UK: Georgia Institute of Technology).
- Plazak, J., Drouin, S., Collins, L., and Kersten-Oertel, M. (2017). Distance sonification in image-guided neurosurgery. *Healthc. Technol. Lett.* 4, 199–203. doi:10.1049/HTL.2017.0074
- Proulx, M. J., Ptito, M., and Amedi, A. (2014). Multisensory integration, sensory substitution and visual rehabilitation. *Neurosci. Biobehav. Rev.* 41, 1–2. doi:10.1016/j.neubiorev.2014.03.004
- Ratan, R. (2012). "Self-presence, explicated: body, emotion, and identity extension into the virtual self," in *Handbook of research on technoself: identity in a technological society*. Editor R. Luppincini (Pennsylvania, United States: IGI Global), 321–335. doi:10.4018/978-1-4666-2211-1.ch018
- Robinson, F. A., Velonaki, M., and Bown, O. (2021). "Smooth operator: tuning robot perception through artificial movement sound," in ACM/IEEE international conference on human-robot interaction, Boulder, CO, United States (IEEE Computer Society), 53–62. doi:10.1145/3434073.3444658
- Rubo, M., Messerli, N., and Munsch, S. (2021). The human source memory system struggles to distinguish virtual reality and reality. *Comput. Hum. Behav. Rep.* 4, 100111. doi:10.1016/j.chbr.2021.100111
- Rutherford, E., and Geiger, H. (1908). An electrical method of counting the number of α -particles from radio-active substances. *Proc. R. Soc. Lond. Ser. A, Contain. Pap. a Math. Phys. Character* 81, 141–161. doi:10.1098/RSPA.1908.0065
- Sanchez-Vives, M. V., and Slater, M. (2005). From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6 (4), 332–339. doi:10.1038/nrn1651
- Sauro, J. (2011). *A practical guide to the system usability scale: background, benchmarks & best practices*. Scotts Valley, California, United States: CreateSpace.
- Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: factor analytic insights. *Presence (Camb)*. 10, 266–281. doi:10.1162/105474601300343603
- Schwarz, M., Rodehutsors, T., Droschel, D., Beul, M., Schreiber, M., Araslanov, N., et al. (2017). NimbRo rescue: solving disaster-response tasks with the mobile manipulation robot momaro. *J. Field Robot.* 34, 400–425. doi:10.1002/ROB.21677
- Sigrist, R., Rauter, G., Marchal-Crespo, L., Riener, R., and Wolf, P. (2015). Sonification and haptic feedback in addition to visual feedback enhances complex motor task learning. *Exp. Brain Res.* 233, 909–925. doi:10.1007/s00221-014-4167-7
- Sinclair, P. (2011). Sonification: what where how why artistic practice relating sonification to environments. *AI Soc.* 27 (2), 173–175. doi:10.1007/S00146-011-0346-2
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Phil. Trans. R. Soc. B* 364, 3549–3557. doi:10.1098/rstb.2009.0138
- Slater, M., and Steed, A. (2000). A virtual presence counter. *Presence (Camb)*. 9, 413–434. doi:10.1162/105474600566925
- Spiegel, J. S. (2018). The ethics of virtual reality technology: social hazards and public policy recommendations. *Sci. Eng. Ethics* 24, 1537–1550. doi:10.1007/s11948-017-9979-y
- [Dataset] Temperton, J. (2016). *Inside sellafield: how the UK's most dangerous nuclear site is cleaning up its act*. <https://www.wired.co.uk/article/inside-sellafield-nuclear-waste-decommissioning>
- Triantafyllidis, E., MCGreavy, C., Gu, J., and Li, Z. (2020). Study of multimodal interfaces and the improvements on teleoperation. *IEEE Access* 8, 78213–78227. doi:10.1109/access.2020.2990080
- Vickers, P. (2011). "Sonification for process monitoring," in *The sonification handbook* (Berlin, Germany: Logos Publishing House), 455–491.
- Walker, B. N., and Kramer, G. (2005). Mappings and metaphors in auditory displays. *ACM Trans. Appl. Percept.* 2, 407–412. doi:10.1145/1101530.1101534
- Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *J. Exp. Psychol. Appl.* 8, 211–221. doi:10.1037/1076-898X.8.4.211
- Walker, B. N., Rehg, J. M., Kalra, A., Winters, R. M., Drews, P., Dascalu, J., et al. (2019). Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: laboratory and prospective. *EBioMedicine* 40, 176–183. Elsevier. doi:10.1016/j.ebiom.2019.01.028
- Won, A. S., Bailenson, J., Lee, J., and Lanier, J. (2015). Homuncular flexibility in virtual reality. *J. Comput. Mediat. Commun.* 20, 241–259. doi:10.1111/jcc4.12107
- Wu, B., Yu, X., and Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: a meta-analysis. *Br. J. Educ. Technol.* 51, 1991–2005. doi:10.1111/bjet.13023
- Yee, N., and Bailenson, J. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x
- Zahray, L., Savery, R., Syrkett, L., and Weinberg, G. (2020). "Robot gesture sonification to enhance awareness of robot status and enjoyment of interaction," in 29th IEEE international conference on robot and human interactive communication, RO-MAN 2020, Naples, Italy (Institute of Electrical and Electronics Engineers Inc.), 978–985. doi:10.1109/RO-MAN47096.2020.9223452
- Zhang, R., Barnes, J., Ryan, J., Jeon, M., Park, C. H., and Howard, A. M. (2017). "Musical robots for children with ASD using a client-server architecture," in Proceedings of the 22nd international conference on auditory display (ICAD2016), Canberra, Australia (International Community for Auditory Display), 83–89. doi:10.21785/icad2016.007