"Who said that?" Applying the Situation Awareness Global Assessment Technique to Social Telepresence

ADAM K. COYNE, KESHAV SAPKOTA, and CONOR MCGINN, Trinity College Dublin, Ireland

As with all remotely-controlled robots, successful teleoperation of social and telepresence robots relies greatly on operator situation awareness, however existing situation awareness measurements, most being originally created for military purposes, are not adapted to the context of social interaction. We propose an objective technique for telepresence evaluation based on the widely-accepted Situation Awareness Global Assessment Technique (SAGAT), adjusted to suit social contexts. This was trialled in a between-subjects participant study ($n = 56$), comparing the effect of mono and spatial (binaural) audio feedback on operator situation awareness during robot teleoperation in a simulated social telepresence scenario. Subjective data was also recorded, including questions adapted from Witmer and Singer's Presence Questionnaire, as well as qualitative feedback from participants. No significant differences in situation awareness measurements were detected, however correlations observed between measures call for further research. This study and its findings are a potential starting point for the development of social situation awareness assessment techniques, which can inform future social and telepresence robot design decisions.

CCS Concepts: • **Human-centered computing** → **Usability testing**; *Collaborative and social computing design and evaluation methods*; Auditory feedback.

Additional Key Words and Phrases: telepresence, social robotics, situation awareness, robot teleoperation, spatial audio

ACM Reference Format:
Adam K. Coyne, Keshav Sapkota, and Conor McGinn. 2023. "Who said that?" Applying the Situation Awareness Global Assessment Technique to Social Telepresence. *ACM Trans. Hum.-Robot Interact.* 1, 1, Article 1 (January 2023), 23 pages. https://doi.org/10.1145/3592801

1 INTRODUCTION

With the ongoing COVID-19 pandemic, remote social interaction has shown itself to be an essential part of modern life, with remote work and videoconferencing becoming widely accepted alternatives to their in-person equivalents. With this, the phenomenon of "Zoom fatigue" has also emerged [50], where the excessive use of virtual communication can result in decreased energy and motivation [47]. This has been attributed in part to the increased difficulty perceiving and conveying non-verbal information (social cues, such as turn-taking during conversations) through a conventional monitor-webcam setup [4]. In response to the limitations of existing solutions, new technologies are emerging to facilitate remote interaction such as robots or virtual reality applications. Their informed use may help mitigate the above communication issue through changes in interaction paradigms, such as allowing remote social interaction to more closely mirror in-person interaction. Self-effacing natural user interfaces such as this can allow for higher exchange of non-verbal information without the added cognitive load.

Authors' address: Adam K. Coyne, kavana21@tcd.ie; Keshav Sapkota, sapkotak@tcd.ie; Conor McGinn, c_mcginn@tcd.ie, Trinity College Dublin, Dublin, Ireland, D02PN40.

Manuscript submitted to ACM

Teleoperated social robots are suitable platforms to address this in practice, notably telepresence robots or social humanoid robots. The former are already in use commercially [40], providing the operator the experience of being socially present at a remote location. The latter, while still mostly restricted to research applications, could also act as a natural user interface for remote interactants on behalf of the operator, providing more human social cues (body language etc) through humanoid embodiment. As with any teleoperated robot, operator situation awareness (SA) is important for adequate performance and low cognitive load. Situation awareness is defined by Endsley as "the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their future status" [11]; with the scope and subtleties of non-verbal information during social interaction [35, 52], SA may be all the more important for telepresence robots.

Situation awareness as a whole has not been studied in the context of social robot teleoperation such as telepresence. Existing methods for its measurement are anchored to the task-oriented military origins of the concept, and do not translate easily to the social domain; SA has been shown to improve for non-social robot teleoperation through representative instrumentation (e.g. minimaps, haptic feedback, etc), but social cues are less feasibly abstracted (possibly requiring the *human* information feedback of a natural user interface to be easily perceived). We propose that, if properly placed in the context of social interaction, situation awareness could be viewed as analogous to a low-level social awareness, which may be assisted or hindered by the teleoperation interface. By bridging *conventional* and *social* robot teleoperation with situation awareness methods, telepresence robot designers could measure and evaluate *social* situation awareness, potentially leading to design choices that increase operator performance and decrease mental and social fatigue.

In this paper, we aim to develop a repeatable methodology for measuring *social situation awareness* (SSA) by adapting existing SA tools. An initial trial of this methodology will be performed on a simulated social telepresence interface, examining the potential benefits of spatial audio feedback for social situation awareness alongside several auxiliary metrics and qualitative analysis.

After a review of related work in the next section, section 3 details the development of the SSA measurement technique, by adapting the widely-used Situation Awareness Global Assessment Technique (SAGAT). Following this, section 4 presents the methodology of the trial study employing this technique. A summary of experimental results are then presented in section 5, followed by their discussion in section 6 and a conclusion on the findings.

2 PRIOR WORK

2.1 Situation awareness measurement techniques

While seldom if ever applied to social teleoperation, techniques for situation awareness (SA) evaluation are well-established for non-social tasks and contexts. A number of such techniques would be considered standard in human-robot interaction research [53]:

SAGAT. Endsley's Situation Awareness Global Assessment Technique (SAGAT) [10] is possibly the most commonly-used SA evaluation technique. It requires a simulated remote environment for the teleoperated robot. Periodically, the simulation is paused and questions are asked to the operator about the situation. SAGAT questions are divided into three categories, each measuring a particular degree of awareness (immediate, recall and future estimation). This reproducible, empirically validated technique results in an objective score measuring situation awareness, with the caveat of being incompatible with real-world robot teleoperation. While intended for the piloting of unmanned aerial vehicles, it has also been applied to medical simulations [15]. SAGAT has also been adapted to use real-time querying

of the operator instead of pauses [25] in tools such as the Situation Present Assessment Method (SPAM) [33]. These tools can notably be administered during real-world robot operation (outside of simulation), but recent studies have called their validity into question [12].

SART. Unlike the objective measurement techniques above, Taylor's Situation Awareness Rating Technique (SART) [44] is a commonly-used subjective measurement. A form with three sets of experiential Likert-scale questions is completed by the operator after the teleoperation session, resulting in three scores describing different facets of awareness, which combine into a composite overall score. It is vulnerable to the same disadvantages as other forms of subjective assessment, such as bias from participants' feelings and emotional state [54], and was shown to be less accurate than SAGAT in particular [13], but it has the additional advantage of easy, universal administration.

Measuring other qualities. The connection between situation awareness and the distinct but similar concept of *presence* [36] has been pointed out in the literature [17, 41], linking conventional robot teleoperation and social robotics. Goodrich et al. [17] explain that while the two do not share the same definition, a high sense of presence implies the operator has sufficient awareness of the remote situation to feel present. While no explicit techniques have been developed for evaluating situation awareness in social contexts, subjective questionnaires have been proposed for the measurement of qualities such as presence, one of the most commonly used being Witmer and Singer's Presence Questionnaire [49].

2.2 Spatial audio in teleoperation

With its adoption in interactive media such as video games [8], spatial audio has been employed in the past for a number of computer-mediated social interactions. The potential benefits of spatial audio during conference calls was investigated by Inkpen et al. [24] and Ahrens et al. [2], and this approach to conference calls was extended to that of a shared virtual room in recent years by Wong et al [51].

Existing proposals of spatial audio feedback in robotics have predominantly been to enhance telepresence. This includes the technical work of Keyrouz and Diepold [27] and Saraiji et al. [42]; both concerned the teleoperation of a humanoid robot, using spatial audio in order to enhance the operator's sense of *presence*. A lack of spatial audio [29, 30] or otherwise unnatural audio behaviour [32, 37] has been expressed as an issue during teleoperation for some robot telepresence platforms.

Combining audio directionality with telepresence has seen applications outside of social robotics. While not involving an embodied teleoperated robot, a telepresence simulation concept by Torrejon et al. was enabled in part by spatial audio [45]. Telepresence was also employed for industrial machinery operation by Ahn et al. [1], using spatial audio feedback on the predicate that it would enhance immersion and situation awareness.

3 DEVELOPING AN ASSESSMENT TECHNIQUE OF SOCIAL SITUATION AWARENESS

With caution taken to ensure reliability, this section details the development of a measurement of situation awareness (SA) when applied to social contexts, which we will name *social situation awareness* (SSA).

3.1 Situation awareness and social interaction

A defining feature of social robotics, and by extension social telepresence, is the use of a *social interface* for human-robot interaction [22, 34]. The social interface can be seen as the communication medium used for everyday human-human social interaction, which includes channels such as speech, gesture and affect expression.

According to Endsley's model of situation awareness [11], SA is defined strictly relative to performance in a task, towards one or multiple goals and objectives. Social interaction on the other hand is seldom goal- or task-driven, however the design exercise of social telepresence has goals, one of which is *to simulate the social interface as closely as possible to the experience of in-person interaction.* SA in social telepresence can be therefore be informed by this goal of successful social interaction.

3.2 Awareness traits of interest

Social awareness it is not a concretely-defined concept, but its mention in psychosocial literature [16, 46] as well as technology-related publications [5] often concerns high-level awareness during social interactions, such as awareness of emotional states (as provided by empathy) or of socioeconomic factors. While it could be argued that these high-level concepts are more important for social interaction, it is difficult to conceptualise a consistent measurement of awareness thereof, as it possibly depends much more on the operator than the teleoperation interface.[1] Furthermore, if an operator fails to pick up on more basic cues (e.g. who is saying a phrase), then their ability to pick up on more complex ones (e.g. who is surprised to hear the phrase) is likely impacted nonetheless.

Because of this, we will target a low-level form of social situation awareness: knowledge of basic non-emotional information about a social situation that can be obtained only through the social interface. This can include for example the names or explicit social roles of people interacted with, but not information about their appearance (which can be acquired non-socially). Targeting this low level of awareness will minimise the complexity of the problem at hand, potentially increasing the reliability of the technique, while also allowing for shorter, more practical procedures for participant testing.

3.3 Existing technique selection

To maximise reliability, an existing situation awareness assessment technique will be adapted to measure awareness of the above traits. Despite being one of the most commonly used subjective techniques, it is judged that the abstract experiential questions of Taylor's Situation Awareness Rating Technique (SART) may be confusing to non-expert participants[2] and are overall unsuited for measuring situation awareness for social interaction. Although the questions of Witmer and Singer's Presence Questionnaire may be of value for its relevance to telepresence studies, an objective technique is preferred for consistency.

While real-time variant techniques of SAGAT may provide additional information of interest (workload measurements in the case of SPAM [33]), it is unknown whether the potentially intrusive nature of the questions might introduce a confounding factor during social interaction. This risk, combined with the widespread acceptance and stronger empirical validation of SAGAT, makes the latter the most reliable choice.

3.4 Adapting SAGAT for social contexts

The body of literature on SAGAT details requirements and guidelines for its administration. However, as its original intended use concerned the piloting of unmanned aerial vehicles [10] there are complications adapting it to social interaction. As many of SAGAT recommendations as possible will be replicated, with some changes for it to pertain to social situation awareness.

[1] Measurement of the awareness of these higher-level concepts can take entirely different approaches that do not involve the teleoperation interface, for example through Baron-Cohen's Empathy Quotient [6].
[2] A SART questionnaire includes queries such as "How many variables are changing within the situation?" and "How much information have you gained about the situation?"

For one, it is instructed that SAGAT queries be created through a goal-directed task analysis [9] i.e. based on the operator's goal, but as mentioned in 3.1 the average real-world social situation is seldom intentionally goal-driven, and the telepresence goal of *engaging in social interaction comparable to face-to-face interaction* is too broad to analyze in this manner. We propose instead that analysis of the *context* of the social interaction determine the selection of important information to query.

SAGAT divides awareness into three levels:

(1) **Perception of Data**
(2) **Comprehension of meaning**
(3) **Future projection**

Each SAGAT query concerns awareness of one of the three levels.

The first level, *Perception of data*, translates well to the social domain. Originally, this would consist of awareness about the position and speed of remote aircrafts. Adapted to social interaction, this data can incorporate most of what was targeted in subsection 3.2—data explicitly expressed through the social interface, such as names.

Comprehension of meaning, while slightly more complex to translate, can still be adapted to social interaction. One potential consideration is the awareness of emotional states based on secondary social cues such as speech prosody [48] but expressing, perceiving and even defining emotional states in this manner is potentially difficult [43]. We will instead consider Level 2 awareness to concern implicit contextual *objective* data, requiring understanding of the social situation to identify. For instance, while awareness of the information "Brian wishes to leave" is a deeper comprehension of implicit social meaning, it is subjective and may be difficult to pick up on, so is considered unreliable for participant testing. Contextual information such as "Brian is chairing the meeting", however, is more suitable; it is objective while remaining implicitly expressed and socially relevant, and it requires awareness and synthesis of lower-level social signals to comprehend.

The third level, *Prediction of future events*, is less applicable to social contexts. Social interaction can be inherently unpredictable, and while predictions during a conversation might occur intuitively, it is assumed that conscious attempts to anticipate the outcome of an interaction to be unneccesary for the majority of cases (akin to considering social interaction goal-driven). This level shall therefore not be considered during generation of SAGAT queries.

Beyond the selection of queries, a substantial issue that may exist when applying SAGAT socially may be that of *demand characteristics* [38]. While awareness of socially-relevant information is maintained passively, participants that are explicitly aware that they are being queried on that awareness may attempt to overcompensate, deliberately seeking pieces of information they would not otherwise. This could drastically bias results in a positive direction, particularly in the case of a repeated-measures trial, where participants may overcompensate in this manner during the second test condition. This foregrounds the need for measures mitigating demand characteristics and order effects; it is recommended that the study be conducted in between-subjects design, and that the intent to measure situation awareness is obscured or distracted from.

3.5 Practical considerations

SAGAT requires the situation be simulated, so that it can be paused to administer queries. The simulation used for this must be a sufficiently valid representation of reality, in this case a valid representation of social interaction. The ubiquitous nature of remote video communication in modern life (and the acceptance of video chats as a valid form of

social interaction) makes this easier to accomplish; the simulator does not need to have a similar level of immersion to in-person interaction, as long as it is at least as immersive as a video chat.

Communication through social interaction can nonetheless include dozens of channels [19], which cannot feasibly all be accounted for, therefore to construct a simulation of social interaction and faithfully convey as many natural social signals as possible, live video recordings of actors must be used rather than computer-generated images. Consequentially, conditions cannot be easily randomized, limiting experimental design. While the conditions of a flight simulator can be changed in software (aircraft positions, velocities etc), with potentially no two identical situations across all participants, for a social simulator every video used must be captured in advance. Although variations can be generated by editing videos together programmatically, two issues prevent them from being viable for use. For one, to maintain immersion and realism, individual phrases should not be interrupted by a cut in the video, which means that the information provided in each utterance cannot be altered, drastically decreasing the random sample space, and with it the utility of programmatic editing. But even without splitting sentences, cuts in the video may interrupt the natural flow of the overall interaction portrayed by actors, possibly destroying subtler social signals.[3] This limits the available test conditions to the amount of videos the researcher can feasibly record. While requiring more resources, having a larger pool of different videos to draw from for each participant will mitigate confounding factors, however it may also decrease the sensitivity of the trial. Regardless of the quantity of videos used, they should all concern the same type of social encounter and the same SAGAT queries should apply across all, but with different ground-truths determining the correct answer.

3.6 Summary

The final technique developed to assess social situation awareness (SSA) is as follows:

- A simple, common social context for the study is chosen, depending on the research focus
- The information to query participants about is chosen through analysis of this context, along the first two of the three SAGAT levels of awareness:
  - **Perception of data**, information explicitly available via the social interface
  - **Comprehension of meaning**, implicit information deduced using first-level information and through context
- This information must be:
  - socially relevant
  - objective in nature
  - non-emotional
- A simulation is created using recorded footage of actors, simulating social interaction in the chosen context
- SAGAT is administered as appropriate during testing, using queries on the selected information

Section 4 below details our trial use of this technique to assess the effect of spatial audio feedback on social situation awareness.

[3]The use of *live acting* for participant testing is a hypothetical alternative allowing for shuffled test conditions, in which actors engage with each participant in real-time during the experiment as opposed to through pre-recorded videos. While this approach would allow for a near-perfect simulation of natural social interaction, it is would require rigorous caution to ensure consistency across participants, and its replicability would presumably be unsound.

4 TRIAL STUDY: SPATIAL AUDIO FEEDBACK

A design choice that may enhance social situation awareness through the paradigm of a natural user interface is via spatial audio feedback. The human brain can instinctively localize where sounds are coming from by combining several different techniques, one of the most significant being difference in sound arrival times between the ears [27]. By mathematically modelling the human head and using a pair of headphones, sounds can be made to be perceived from different locations - this is often used in interactive media such as video games [8] to increase operator immersion in a virtual environment. The binaural nature of hearing is also a major contributor to the "cocktail party effect", [21] our ability to distinguish and focus on an auditory source in noisy environments, including speech when multiple people are talking simultaneously, however this can occupy substantial mental resources (the "cocktail party problem") [14]. As spatial audio feedback increases the quantity of information a user receives through the natural user interface paradigm, it would seem logical that its inclusion over mono audio feedback would increase situation awareness of the remote location.

As an initial trial of the technique proposed in the previous section to measure social situation awareness (SSA), we applied it to a telepresence experiment wherein spatial audio feedback was compared to mono audio feedback during social robot teleoperation. A telepresence simulation was created with the robot at the centre of a social scene; participants were asked to follow along as though teleoperating the robot remotely. One group of participants received mono audio feedback as a baseline, the other received spatial audio, and were therefore capable of localising sounds in the virtual environment.

4.1 Stimuli creation

To simulate a social interaction as closely as possible while relying on pre-recorded video, omnidirectional footage was recorded of the portrayed scene.

Social context. The form of social interaction chosen for this study is a classroom quiz scenario. A scene like this is a structured, systemic way to represent a social interaction, as other typical social interactions can be more complex. This situation is also a potential use-case for social robot teleoperation, in the form of robot-mediated remote learning or teaching.

Design. Actors were recruited (two men and four women), to portray one quiz master and five contestants. So that the simulation would adequately represent robot teleoperation by contrast to a conventional video call, actors were seated at desks in a circle around the camera (as shown in Figure 1)—this was to encourage the operator to look around the virtual environment rather than simply view it like a static camera feed, so that the workload would incorporate the control input component.[4] Quiz questions that the quiz master asked of the contestants were explicitly chosen to be extremely difficult or very vaguely worded, so although the questions and answers seemed conventional in a quiz, it was unlikely that contestants would be able to answer using prior knowledge. Two videos were recorded using two distinct scripts following the same format. Actor positions were shuffled, and a different set of quiz questions, character roles and names were used between the two. Quiz questions and answers for both videos are detailed in Appendix A.

[4]A secondary reason for this robot-centric layout (as opposed to including the robot in the social circle with the actors) was motivated by an informal observation made some years prior by the researchers. It was observed that when a telepresence robot brings a substantial degree of novelty to a social gathering, it can draw the attention of groups of people, who observe and interact with it by surrounding it in a circle in this manner, requiring a mentally-taxing 360 degrees of social awareness for the operator.

Fig. 1. Side view of the set used for recording stimuli. Omnidirectional camera and ambisonic microphones can be seen in the centre.

Duration. One recommendation for the administration of SAGAT [9] is that a minimum of 3 minutes must have elapsed from the start of the simulation before the first pause, and that further pauses must be spaced apart by a minimum of 1 minute. In order to keep simulation time relatively short, both for ease of testing with participants and to facilitate video recording, it was decided that SAGAT pauses be administered twice per video; first at a random time between 2 and 4 minutes, and then at a random time at least one minute after the first. Each video would therefore last 5 minutes total.

Script. The quiz master began by introducing the experiment to the operator, initiating a round of introductions where every contestant spoke their name. This provided the operator a chance to hear every contestant name, and also encouraged them to visually explore the room. Then the main loop of the scene began: every 15 seconds, the quiz master would ask a question to a specific contestant, addressing them by name (e.g. “Which European city hosted the 1936 Summer Olympics? Annie?”). The chosen contestant would begin by saying “I think the answer is-” to prime the operator for the answer. Then both the contestant and another "distractor" contestant would speak different answers simultaneously, of which one at random was correct, talking over one another (Annie: “I think the answer is...London”; Maria: “Berlin”). The quiz master would then congratulate the contestant who answered correctly (“That’s right, Maria"), pause for the remainder of the 15 seconds, and continue with the next question. This main loop of the script would continue until the video time reached 5 minutes. A complete list of quiz questions and answers is provided in Appendix A.

Recording. Video was recorded using a Ricoh Theta Z1 omnidirectional camera. The camera captures videos in a spherical format using its dual fish-eye lens, which are converted to a conventional 360-degree format (MPEG-4) using Ricoh Theta proprietary software. Audio was captured using a Sennheiser AMBEO VR Mic, a ambisonic microphone array consisting of four high-fidelity microphones in tetrahedral arrangement, via the Zoom H6 Audio Recorder. Sennheiser’s AMBEO A-B proprietary software was used to convert the recording to standard Ambisonics-B format (WAV).

4.2 Implementation

Simulator. For its support in the literature for use in human-robot interaction simulators [3, 31] as well as its ease in handling multimedia content, Unity was chosen as a framework to create the simulator. The 360-degree MPEG-4 videos were projected on the interior of a Sphere object in Unity, with a Camera object at its centre to represent the teleoperated robot. The open-source plugin Resonance Audio [18] was used to generate spatial audio in real time based on the ambisonic recordings, conveying both interaural level and time differences using head-related transfer functions.

Teleoperation interface. A conventional desktop PC setup was used as a teleoperation interface, for its widespread familiarity and ease of access. Audio feedback was provided through a pair of high-fidelity over-ear headphones. Pressing the left and right arrow keys of the keyboard would rotate the camera object, as if rotating the camera feed of a robot—the spatial audio feed would rotate accordingly.

4.3 Task

In order to mitigate SAGAT query demand characteristics as explained in 3.4, as well as to engage and maintain focus on the scene, a simple task was created; participants were asked to follow along with the video shown to them, and to identify for each question the correct answer. After a question was answered, both the correct and incorrect answers that were uttered were displayed at the bottom of the screen, and the participant chose which one they believed was correct by pressing a key on the keyboard. In this manner, the task pertained to the situation and awareness thereof, without any direct overlap with the content of the SAGAT queries.

4.4 Manipulations

Two videos were recorded, and two levels of the independent variable (audio feedback) were to be evaluated, resulting in four test condition permutations. Each participant would view one of the four permutations in between-subjects experimental design, resulting in two groups; one having received Mono audio feedback, the other Spatial audio feedback.

The increased sensitivity of conducting a within-subjects trial a.k.a. repeated measures would have been advantageous (presenting one of each video in random order with one for each audio feedback condition). However it was judged based on informal testing of the simulator that ordering effects could be significant between the two audio feedback conditions, potentially introducing confounding factors, so the decision was made to use independent measures.

4.5 Measures

While the SAGAT score of each participant is the primary metric of interest to this study, measuring situation awareness, secondary metrics were also recorded and analyzed to gain further understanding of test results.

SAGAT score. This score of social situation awareness (SSA) is measured using the technique devised in Section 3. Queries were created that pertained socially to the quiz scenario, such as "Which contestant last answered correctly?". No queries required the participant to identify the correct answer to a quiz question, as this task was already asked of participants (see 4.3). The full pool of questions, randomized between both SAGAT pauses, is shown in Table 1.

Each query allowed the participant to select from multiple response options—six options available, with a single correct answer, as well as an additional "I don't know" option. Two SAGAT pauses were administered per participant of five queries each, resulting in a final SAGAT score out of 10 (a score of 10 showing that every question was answered

Level of awareness	Queries
1. Perception of Data	*"Identify one person who was in your field of view just before the simulation paused."* *"What was the most recent question?"* *"Who is sitting left/right of the quiz master?"* *"Who is sitting two seats to the left/right of the quiz master?"* *"What is the colour of the quiz master's shirt?"*
2. Comprehension of Meaning	*"Who was asked to answer a question most recently?"* *"Who is the quiz master?"* *"Who last answered correctly?"* *"Who last answered incorrectly?"* *"What kind of social event is taking place here?"*

Table 1. Pool of all potential queries delivered during SAGAT pauses.

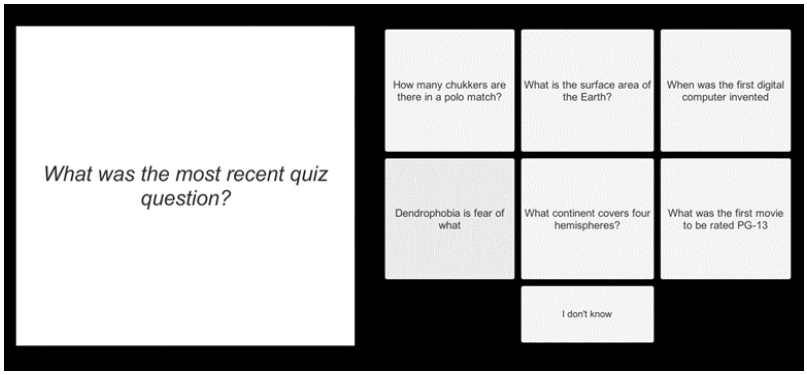

Fig. 2. A sample view of the teleoperation screen during a SAGAT pause, during which the participant was queried about their awareness of aspects of the scene.

correctly, indicating high situation awareness). A sample SAGAT query screen displayed during the simulation is shown in Figure 2

Presence Questionnaire. With the value of *presence* in relation to situation awareness, particularly where telepresence is concerned, of a subset of questions from Witmer and Singer's Presence Questionnaire [49] was administered to participants. Each response was provided through a 7-point Likert scale. Question labels were slightly adjusted and extended for clarity, as shown in Table 2. The sum of scores for all component questions constitutes an overall presence score—for 6 questions, this will be a score out of 42.

Positional metrics. Spatial audio can enable an operator to localise sounds without needing to see the source. It is therefore predicted that mono audio feedback will incite participants to visually pan around the simulation more by comparison. To investigate this, the following metrics were derived using positional data and user input logs from the teleoperation interface:

- *Mean answer time* - the mean time in seconds for each participant to select which of the two answers to a quiz question they deemed correct.
- *Ratio of time in motion* - the ratio of experiment time spent turning the camera to total experiment time.
- *Mean viewing angle* - the mean angle in degrees (where an angle of 0 is facing the quiz master).

Label	Original number [49]	Original question [49]	Adjusted for clarity
P1	5	*"How much did the visual aspects of the environment involve you?"*	*"How much did the visual aspects of the environment involve you? In other words, how much did the visual component of the experience contribute to the awareness of the situation?"*
P2	6	*"How much did the auditory aspects of the environment involve you?"*	*"How much did the auditory aspects of the environment involve you? In other words, how much did the audio component of the experience contribute to the awareness of the situation?"*
P3 *	15	*"How well could you identify sounds?"*	*"How well could you identify sounds?"*
P4	16	*"How well could you localize sounds?"*	*"How well could you tell where sounds were coming from? In other words, how well could you localize the direction of sound?"*
P5 *	12	*"How much did your experiences in the virtual environment seem consistent with your real-world experiences?"*	*"How much did your experiences in the virtual environment seem consistent with your real-world experiences?"*
P6 *	23	*"How involved were you in the virtual environment experience?"*	*"How involved were you in the virtual environment experience?"*

* Question unchanged from original questionnaire

Table 2. List of questions used from Witmer and Signer's Presence Questionnaire [49], adjusted for clarity and administered post-experiment.

- *Heading angle variance* - we wish to evaluate the angular range of motion employed for each participant, but it is assumed participants will pan to view the entire scene (360 degrees) at least once during the experiment. The variance of the heading angle over time for each particpant can represent the variability of angles in the given timespan; a low variance indicates an generally smaller angular range was viewed, while a high variance indicates the participant was more willing to cover wider ranges over the course of the experiment.

Task performance. A lesser metric is that of task performance; the total amount of correct quiz answers identified. Every time a question is asked by the quiz master to a contestant, the participant is prompted on-screen to identify which answer was correct, as shown in Figure 3. The task was not created with the intent to measure performance but rather to mitigate demand characteristics (drawing attention away from the SAGAT queries), and with no particular precendent in the literature for the use of such a metric, it was considered auxiliary to the others.

Qualitative feedback. The post-experiment questionnaire includes the following optional fields for positive and negative open feedback from participants:

- *"Is there anything in particular that you liked about the experience?"*
- *"Is there anything in particular that you disliked about the experience?"*

4.6 Participants

The goal was to recruit a representative sample of the lay population, controlled for English language fluency, hearing or spatial awareness issues, and colourblindness. This was done through random recruitment of the footfall in semi-public locations.

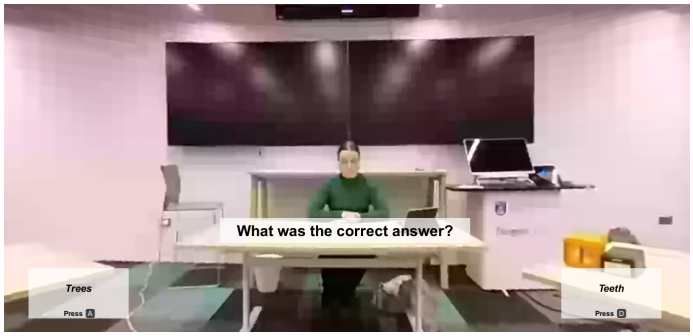

Fig. 3. A view of the simulator screen shortly after a quiz question was asked, illustrating the task asked of the participant to identify the correct answer.
Screenshot captured after testing was completed, with inferior video resolution than what was used during data collection.

4.7 Procedure

Participants were recruited one by one. After providing informed consent, the participant would begin a 2-minute trial of the experiment during which no data was recorded, which included one SAGAT pause. The participant was encouraged to ask the researcher any necessary questions during the trial. Once complete, the screen would fade to black, and the participant informed that they could begin the experiment proper when ready, which began once they accepted through the simulator interface. After the simulation, participants completed a questionnaire based on their experience, and were thanked for their participation.

4.8 Analysis

Data was pre-processed using Python scripts. Statistical analysis was performed using R. [39]

We wish to evaluate whether the difference in means for SAGAT scores between the two groups is significant. To first determine parametricity, a Shapiro-Wilk normality test was conducted. A two-tailed independent-measures t-test would be conducted for normally distributed data, a Wilcoxon rank-sum test if not normally distributed. The same process was followed to determine a difference in means between the answers given in the presence questionnaire. The chosen significance level in all cases was $\alpha = 0.05$.

A Pearson correlation matrix was also calculated across all quantitative variables, to investigate and evaluate their research value. Finally, qualitative analysis was conducted to identify any overarching themes in participant feedback.

5 RESULTS

A total of 56 participants were recruited from two separate locations, a college building (28) and an office block (28). Participant age ranged from 18 to 60 ($\mu = 28.9$, $\sigma = 11.5$), and the male-to-female ratio was 31:25. All participants reported to be fluent in the English language, and none reported any form of colourblindness, hearing issues or spatial awareness issues.

5.1 Quantitative results

SAGAT score data across all participants was not normally distributed ($W = 0.926, p = 0.002$). A Wilcoxon rank-sum test showed that the group that received spatial audio feedback did not yield significantly higher situation awareness scores compared to the group with only mono audio ($W = 338.5, p = 0.373$). Indeed, the median score was the same for

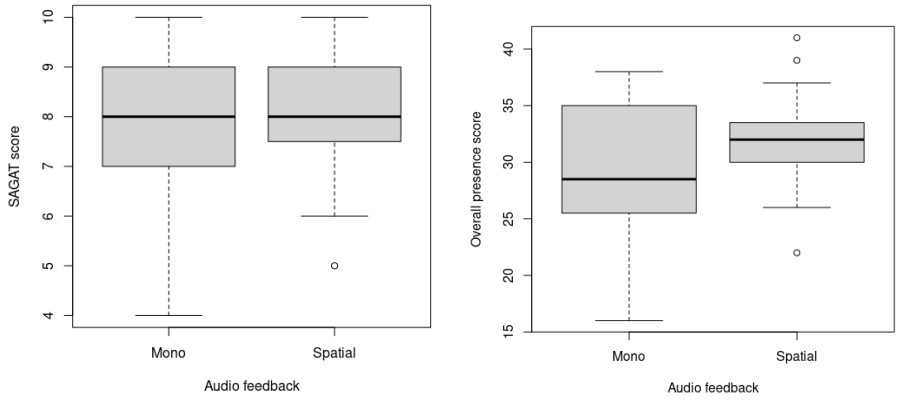

Fig. 4. Box plot summaries of SAGAT scores and overall presence scores for both Mono ($n_1 = 28$) and Spatial ($n_2 = 28$) audio feedback groups.

Fig. 5. Density of SAGAT scores observed for both Mono ($n_1 = 28$) and Spatial ($n_2 = 28$) audio feedback groups.

both groups (8). No instance of a participant selecting the "I don't know" answer was recorded for any of the queries. SAGAT score data is summarised graphically in Figure 4 and numerically in Table 4, with detailed information on responses to each query shown in 3.

Although composite overall responses to the Questionnaire formed a normal distribution ($W = 0.976, p = 0.331$), individual responses to Presence Questionnaire were not normally distributed ($p \leq 0.004$). Cronbach's alpha showed that the six items of the Presence Questionnaire were poorly internally consistent ($\alpha = 0.577$). Wilcoxon rank-sum tests for each individually did not show any significantly different means between the spatial audio and mono audio groups ($p \geq 0.104$). These results are summarised in Table 5.

The other quantitative measures (task performance, time to answer etc) are summarised in Table 4. Pearson correlations between all measures are shown in Table 6, identifying a number of significant correlations.

	Mono audio			Spatial audio			Overall			
SAGAT Query	n	C	r	n	C	r	n	C	r	Δr
"What was the most recent quiz question?"	27	23	0.85	33	28	0.85	60	51	0.85	0
"Who is the quiz master?"	21	20	0.95	22	22	1.00	43	42	0.98	+ .05
"Who is sitting two seats to the [l/r] of the quiz master (your [l/r])?"	29	20	0.69	33	22	0.67	62	42	0.68	− .02
"What is the colour of the quiz master's shirt?"	33	23	0.70	28	24	0.86	61	47	0.77	+ .16
"Who was asked to answer a question most recently?"	29	24	0.83	32	27	0.84	61	51	0.84	+ .01
"Who last answered incorrectly?"	28	14	0.50	32	20	0.62	60	34	0.57	+ .12
"Who last answered correctly?"	27	22	0.81	26	22	0.85	53	44	0.83	+ .04
"What kind of social event is taking place here?"	21	21	1.00	19	16	0.84	40	37	0.93	− .16
"Who is sitting [l/r] of the quiz master (your [l/r])?"	28	19	0.68	27	23	0.85	55	42	0.76	+ .17
"Identify one person who was in your field of view just before the simulation paused."	37	30	0.81	28	23	0.82	65	53	0.82	+ .01

Table 3. SAGAT scores detailed by each individual SAGAT query. As query selection was random, not every query was equally represented. With n the number of times the query appeared, C the number of times the query was answered correctly, r the ratio of correct answers, Δr the difference between the correct answer ratios of spatial and mono audio conditions.

	Mono audio		Spatial audio			
	μ	σ	μ	σ	W	p
SAGAT scores	7.71	1.61	8.11	1.31	338.5	0.373
Task performance	9.11	1.17	9.25	0.89	377.5	0.804
Mean answer time (s)	2.99	0.70	3.19	1.33	385	0.916
Ratio of time in motion	0.14	0.09	0.15	0.07	371	0.739
Mean viewing angle (°)	2.81	13.7	-0.43	10.6	475	0.178
Variance of viewing angle	69.0	20.4	79.0	20.1	296	0.118

Table 4. Statistical summary of objective results, with results of Wilcoxon rank-sum tests between both Mono ($n_1 = 28$) and Spatial ($n_2 = 28$) audio feedback groups

		Mono group		Spatial audio			
Question	Content	μ	σ	μ	σ	W	p
P1	"Visual involvement"	5.04	1.48	5.19	1.30	358.5	0.742
P2	"Auditory involvement"	5.79	1.37	5.85	1.32	364	0.808
P3	"Sound identification"	5.21	1.34	5.56	1.01	334	0.445
P4	"Sound localisation"	4.21	1.89	5.00	1.62	283	0.104
P5	"Real-world consistency"	4.64	1.31	4.74	1.70	337.5	0.492
P6	"General involvement"	4.89	1.71	5.56	1.25	296.5	0.163
	Overall presence score	29.8	5.51	31.9	4.22	306.5	0.230

Table 5. Statistical summary of results of Presence Questionnaire questions, with results of Wilcoxon rank-sum tests between both Mono ($n_1 = 28$) and Spatial ($n_2 = 28$) audio feedback groups. Shown here with reminder phrases for question content (full questions as seen by participants shown in Table 2)

5.2 Qualitative results

32 items of positive feedback and 20 items of negative feedback were provided through the post-experiment questionnaire. Thematic analysis identified several themes of interest, shown in Table 7. Individual answers to positive and negative open feedback questions are detailed in Appendix B along with their thematic relevance.

	(1)	(2)	(3)	(4)	(5)	(6)	(P1)	(P2)	(P3)	(P4)	(P5)	(P6)
(1) SAGAT score	1											
(2) Task performance	0.28*	1										
(3) Mean answer time	-0.21	-0.49***	1									
(4) Ratio of time in motion	0.23	0.1	0.02	1								
(5) Mean viewing angle	0.11	0.26	-0.13	-0.05	1							
(6) Variance of viewing angle	-0.02	-0.12	0.15	0.19	-0.2	1						
(P1) "Visual involvement"	-0.02	0.04	0.11	0.09	-0.06	0.23	1					
(P2) "Auditory involvement"	0.14	0.02	-0.14	0.1	0	-0.03	0.04	1				
(P3) "Sound identification"	0.06	0.35**	-0.12	0.08	-0.09	-0.17	0.25	0.36**	1			
(P4) "Sound localisation"	-0.04	0.25	-0.06	-0.06	0.17	-0.29*	0.12	0.01	0.33*	1		
(P5) "Real-world consistency"	-0.11	0.33*	-0.17	-0.12	0.05	0.13	0.21	0.17	0.38**	0.33*	1	
(P6) "General involvement"	0.04	0.05	-0.02	0.06	-0.13	0	0.18	-0.01	0.39**	-0.01	0.21	1

*$p < .05$
**$p < .01$
***$p < .001$

Table 6. Pearson correlation matrix of all quantitative measures. (P1) through (P6) represent Presence Questionnaire answers, shown here with reminder phrase (full questions as seen by participants shown in Table 2)

Themes identified	Feedback categories	Mono audio	Spatial audio
	Positive feedback		
Immersion	Immersion / Ability to look around	7	7
	Spatial audio feedback	0	6
Entertainment	Challenging / Gamelike	3	2
	Novel experience	1	2
	Other	3	1
	Total	**14**	**18**
	Negative feedback		
Boredom, impatience	No vertical camera control	0	2
	Slow camera panning speed	1	2
	It was repetitive / boring	0	2
Difficulty, frustration	Audio was poor (increasing difficulty)	6	2
	Field of view too small	2	0
	Other	2	1
	Total	**11**	**9**

Table 7. Categorised summary of open feedback; with number of items from respective groups, and identified themes. *Positive feedback* consists of responses to the question "Is there anything in particular that you liked about the experience?" and *negative feedback* responses to the question "Is there anything in particular that you disliked about the experience?".

"Immersion" and "Entertainment" in overall positive feedback. The majority of feedback given across both groups (14 items total) was of appreciation for the immersion of the interface, in particular the ability to look around. A smaller number of items (8) across both groups reported appreciation for the entertainment of the experience.

"Sound localisation" in Spatial group positive feedback. Participants of the Spatial audio group reported appreciation for the spatial nature of audio feedback and the ability to localise sounds (6 items).

"Difficulty, Frustration" in Mono group negative feedback. Negative feedback provided by the group having received mono audio feedback describes experiencing *difficulty and overall frustration* with the task and interface (8 items, 6 more than in the Spatial audio group).

"Boredom, Impatience" in Spatial group negative feedback. Negative feedback provided from the spatial audio feedback group reports impatience with wait times between questions or a desire for a faster, more dynamic interface (6 items, 5 more than in the Mono group).

6 DISCUSSION

6.1 Quantitative findings

While researchers predicted quantitative differences in favour of spatial audio, no statistically significant differences were found between both groups (for a global significance threshold of $\alpha = 0.05$) for situation awareness measurements, presence questionnaire results, or any secondary metrics. Although it is possible that spatial audio feedback has no effect on any of these factors by comparison to mono audio, the sensitivity of the study may have been poor for a number of reasons. For one, as shown in Figure 5, greater SAGAT scores were observed with higher densities, with 75% of SAGAT scores between 7 and 10. With data such as this grouped around the maximum, the *ceiling effect* may have occurred, whereby the upper limit placed on the measure reduces the meaningfulness of the data, potentially obscuring an effect that would otherwise be observable. Another reason is the low statistical power of the study (estimated at 0.45).

Table 6 shows several correlations of note, a number of which may be of interest for improving the SSA assessment technique, or the experimental design of subsequent studies. Several were identified among component questions of the Presence Questionnaire—with how the Questionnaire was designed, along with the similarity of the component questions and the method they are administered, any significant correlations between them are to be expected. A strongly significant negative correlation is that between *Task performance* and *Mean answer time*, which is also to no surprise—participants that are more confident of their answers (or simply more focused on the simulator) would be more likely to input the answers quickly. The positive correlation between *Task performance* and the Presence Questionnaire component on *Sound identification* can be similarly explained; the task itself required identifying sounds.

A weaker positive correlation is that of *SAGAT score* and *Task performance*. This reflects the similar, albeit non-overlapping nature of both SAGAT queries and the quiz task; both require information obtained through social signals, and the success rate of identifying the correct answer to a question based on events in the scene could be considered a very focused measurement of situation awareness. Finally, the positive correlation of *Variance of viewing angle* with *Sound localisation* shows how participants would look around the room more to compensate for difficulty localising sounds. This also shows initial promise for the use of the viewing angle variance as an objective measure, either of ease of sound localisation directly, or of workload related to difficulty localising sounds.

6.2 Qualitative findings

Examining the Table 7, negative feedback provided by the Mono audio group predominantly pertained to the *difficulty and frustration* theme identified, while negative feedback from the Spatial audio group pertained to the theme of *boredom and impatience*. This illustrates a substantial difference in how the trial was experienced between both groups. What participants disliked about the Mono audio experience was the workload, particularly citing the audio feedback as a

cause.[5] By comparison, what participants expressed they disliked about the Spatial audio experience was tedium, a desire for it to be more dynamic, and to engage more in the interface. In tandem with the fact that a substantial amount of positive feedback for the Spatial group was on the spatial nature of the audio, it can be induced that introducing spatial audio over mono likely *decreased* the workload for the task at hand. This is in accordance with the neurological basis of the "cocktail party problem" [14]; attempting to distinguish one thread of speech spoken simultaneously with others can heavily engage the brain, to the point of decreasing performance at concurrent tasks. The binaural hearing of spatial audio is known to be a major contributor to our ability to distinguish speech in this manner [21], so it can be intuited this would decrease the associated workload.

For the above reasons, along with the overall feedback distribution (the spatial audio group having provided 29% more positive feedback items and 10% less negative feedback items than the mono audio group), it can be concluded that spatial audio was a qualitatively superior experience to mono audio, likely due in most part to the potential workload alleviation. This is consistent with observations made in the literature around spatial audio; participants often report appreciation for being able to localize sounds in prior studies. Finally, the praise given by both groups to the ability to look around using the interface underlines the potential value of telepresence over conventional videoconferencing.

6.3 Limitations and recommendations

The most substantial limitation of this evaluation was its statistical power, estimated at 0.45 for an assumed effect size of 0.5.[6] An increase in sample size could overcome this, as well as a redesign of the experiment to allow for repeated measures over both test conditions for each participant, although as explained in Section 3, it may be challenging to adapt to within-subjects design because of ordering effects and demand characteristics.

While the robot being placed at the centre of the group of actors was relevant to some teleoperation situations, it may be less representative of the majority of real-life interactions. Drawing on the domain of proxemics can be used to improve on this in future studies, such as through the use of Kendon's F-formations [23, 26].

Although care was taken to preserve sound directionality in the audio pipeline, a complementary study to validate the spatial audio of the setup could ensure with full confidence that spatial audio was properly conveyed, such as that performed by Kiselev et al. [28]. The setup of this experiment used first-order ambisonics—improvements can be made to the fidelity of spatial audio by increasing the ambisonic order of the microphone (using a larger microphone array).

The potential observed ceiling effect on SAGAT scores is another limitation of note. Adjusting for this in future studies can be done based on the data in Table 3, which breaks down the individual scores for each SAGAT query type. This can be used to assess the relative difficulty of future queries, and if a similar "quiz" scenario is repeated in a future study, can be prioritized for re-use.

7 CONCLUSION

This study set out to develop a methodology for measuring social situation awareness through a novel application of the Situation Awareness Global Assessment Technique (SAGAT) to the domain of social interaction, evaluated by investigating whether spatial audio feedback during robot teleoperation would increase operator awareness of socially relevant details.

While qualitative analysis of the trial study showed tangible usability benefits to spatial audio feedback, the quantitative data of the trial study was mostly inconclusive, although correlations between measurements indicate potential

[5]This can be seen in the full list of feedback items in Appendix B.
[6]as informed by [7].

value in the approach; more work is required to improve and validate the methodololgy. It is hoped that these findings can create a starting point for further social situation awareness studies, as well highlight the value of qualitative analysis as a complement to quantitative statistics.

Much work remains to be done towards developing and perfecting social situation awareness measurement. The reliability of the technique detailed in this paper depends on that of SAGAT—future methods papers may consider moving away from a reliance on SAGAT and/or conducting assessments of consistency and sensitivity. It would also be of value to develop a method for it to be used in within-subjects trials, perhaps introducing alternative solutions to mitigate ordering effects and demand characteristics. Finally, its queries could be generalised to more easily apply the technique to any social context.

Concerning spatial audio in telepresence, future work might begin by investigating more aspects of spatial audio during social robot teleoperation, notably a formal workload evaluation using tools such as NASA-TLX [20]. A subsequent study could more closely focus on the cocktail party effect during robot teleoperation, studying the effect of spatial audio feedback in contrast to mono while also varying the level of interfering noise from other speakers. This could be also applied in a contrasting social context, such as a longer, more drawn-out explanation or storytelling session, to examine different forms of social awareness.

ACKNOWLEDGMENTS

We thank Tangent as well as The Digital Hub for enabling data collection. We also thank Maya Vizel-Schwartz (University College Dublin) for discussions on qualitative methods, and Ann Bell (The Digital Hub) for general helpfulness.

REFERENCES

[1] Ho Seok Ahn, Samuel McArdle, Greg Sumner, and Bruce A. MacDonald. 2014. Development of user interfaces for an internet-based forklift teleoperation system with telepresence. In *Proceedings of Australasian Conference on Robotics Automation*, Vol. 02-04-December-2014. ARAA, Melbourne, Australia.

[2] Jens Ahrens, Matthias Geier, Alexander Raake, and Claudia Schlegel. 2010. Listening and conversational quality of spatial audio conferencing. In *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. Audio Engineering Society, Tokyo, Japan.

[3] Almohannad Albastaki, Marius Hoggenmüller, Frederic Anthony Robinson, and Luke Hespanhol. 2020. Augmenting Remote Interviews through Virtual Experience Prototypes. In *32nd Australian Conference on Human-Computer Interaction*. ACM, Sydney NSW Australia, 78–86. https://doi.org/10.1145/3441000.3441057

[4] Jeremy N. Bailenson. 2021. Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior* 2, 1 (Feb. 2021). https://doi.org/10.1037/tmb0000030

[5] M. Cecília C. Baranauskas. 2014. Social awareness in HCI. *Interactions* 21, 4 (July 2014), 66–69. https://doi.org/10.1145/2621933

[6] Simon Baron-Cohen and Sally Wheelwright. 2004. The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders* 34, 2 (2004), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00 arXiv: 0162-3257/04/0400-0163/0 ISBN: 0162-3257.

[7] Madeleine E. Bartlett, C. E. R. Edmunds, Tony Belpaeme, and Serge Thill. 2022. Have I Got the Power? Analysing and Reporting Statistical Power in HRI. *ACM Transactions on Human-Robot Interaction* 11, 2 (June 2022), 16 pages. https://doi.org/10.1145/3495246

[8] James Broderick, Jim Duggan, and Sam Redfern. 2018. The importance of spatial audio in modern games and virtual environments. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, Galway, Ireland, 1–9.

[9] Mica R. Endsley. 1988. Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors Society Annual Meeting* 32, 2 (Oct. 1988), 97–101. https://doi.org/10.1177/154193128803200221

[10] Mica R. Endsley. 1988. Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*. IEEE, Dayton, Ohio, USA, 789–795. https://doi.org/10.1109/NAECON.1988.195097

[11] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (March 1995), 32–64. https://doi.org/10.4324/9781315092898-13

[12] Mica R. Endsley. 2021. A Systematic Review and Meta-Analysis of Direct Objective Measures of Situation Awareness: A Comparison of SAGAT and SPAM. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 63, 1 (Feb. 2021), 124–150. https://doi.org/10.1177/0018720819875376

[13] Mica R. Endsley, Stephen J. Selcon, Thomas D. Hardiman, and Darryl G. Croft. 1998. Comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics Society* 1 (1998), 82–86. https://doi.org/10.1177/154193129804200119

[14] S Evans, C McGettigan, ZK Agnew, S Rosen, and SK Scott. 2016. Getting the cocktail party started: masking effects in speech perception. *Journal of cognitive neuroscience* 28, 3 (March 2016), 483–500. https://doi.org/10.1162/jocn_a_00913

[15] Aimee K. Gardner, Matthew Kosemund, and Joseph Martinez. 2017. Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simulation in Healthcare* 12, 1 (Feb. 2017), 17–21. https://doi.org/10.1097/SIH.0000000000000181

[16] Barney G Glaser and Anselm L Strauss. 1964. Awareness Contexts and Social Interaction. *American Sociological Review* 29, 5 (1964), 669. https://doi.org/10.2307/2091417

[17] Michael A. Goodrich, Jacob W. Crandall, and Emilia Barakova. 2013. Teleoperation and Beyond for Assistive Humanoid Robots. *Reviews of Human Factors and Ergonomics* 9, 1 (2013), 175–226. https://doi.org/10.1177/1557234X13502463

[18] Marcin Gorzel, Andrew Allen, Ian Kelly, Julius Kammerl, Alper Gungormusler, Hengchin Yeh, and Francis Boland. 2019. Efficient Encoding and Decoding of Binaural Sound with Resonance Audio. In *2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society, York, UK, 1–12. https://www.aes.org/e-lib/browse.cfm?elib=20446

[19] Marine Grandgeorge. 2020. Communication Between Humans: Towards an Interdisciplinary Model of Intercomprehension. In *Human-Robot Interaction: Evaluation Methods and Their Standardization*. Springer, 3–19. https://doi.org/10.1007/978-3-030-42307-0_1

[20] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (Jan. 1988), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9 arXiv: 1011.1669v3.

[21] Monica L. Hawley, Ruth Y. Litovsky, and John F. Culling. 2004. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *The Journal of the Acoustical Society of America* 115, 2 (2004), 833–843.

[22] Frank Hegel, Claudia Muhl, Britta Wrede, Martina Hielscher-Fastabend, and Gerhard Sagerer. 2009. Understanding Social Robots. In *2009 Second International Conferences on Advances in Computer-Human Interactions*. IEEE, Cancun, Mexico, 169–174. https://doi.org/10.1109/ACHI.2009.51

[23] Helge Hüttenrauch, Kerstin Severinson Eklundh, Anders Green, and Elin A. Topp. 2006. Investigating spatial relationships in human-robot interaction. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Beijing, China, 5052–5059.

[24] Kori Inkpen, Rajesh Hegde, Mary Czerwinski, and Zhengyou Zhang. 2010. Exploring spatialized audio & video for distributed conversations. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, Savannah, Georgia, USA, 95–98.

[25] Debra G. Jones and Mica R. Endsley. 2000. Can real-time probes provide a valid measure of situation awareness?. In *Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*. SA Technologies, Savannah, Georgia, USA, 245–250.

[26] Adam Kendon. 1990. Spatial organization in social encounters: The F-formation system. In *Conducting interaction: Patterns of behavior in focused encounters*. Cambridge University Press.

[27] Fakheredine Keyrouz and Klaus Diepold. 2007. Binaural source localization and spatial audio reproduction for telepresence applications. *PRESENCE: Teleoperators and Virtual Environments* 16, 5 (2007), 509–522.

[28] Andrey Kiselev, Mårten Scherlund, Annica Kristoffersson, Natalia Efremova, and Amy Loutfi. 2015. Auditory Immersion with Stereo Sound in a Mobile Robotic Telepresence System. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. ACM, Portland Oregon USA, 55–56. https://doi.org/10.1145/2701973.2702034

[29] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction* 2013 (2013), 17 pages.

[30] Annica Kristoffersson, Silvia Coradeschi, Kerstin Severinson Eklundh, and Amy Loutfi. 2011. Sense of presence in a robotic telepresence domain. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, Orlando, Florida, USA, 479–487.

[31] D Krupke, S Starke, L Einig, J Zhang, and F Steinicke. 2018. Prototyping of immersive HRI scenarios. In *Human-Centric Robotics- Proceedings of the 20th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines, CLAWAR 2017*. CLAWAR Association Ltd., Porto, Portugal, 537–544. https://doi.org/10.1142/9789813231047_0065

[32] Min Kyung Lee and Leila Takayama. 2011. "Now, i have a body" uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, Vancouver, British Columbia, Canada, 33–42.

[33] Shayne Loft, Daniel B. Morrell, and Samuel Huf. 2013. Using the situation present assessment method to measure situation awareness in simulated submarine track management. *International Journal of Human Factors and Ergonomics* 2, 1 (2013), 33. https://doi.org/10.1504/ijhfe.2013.055975

[34] Conor McGinn. 2019. Why Do Robots Need a Head? The Role of Social Interfaces on Service Robots. *International Journal of Social Robotics* 12 (June 2019), 15 pages. Issue 1. https://doi.org/10.1007/s12369-019-00564-5

[35] Albert Mehrabian. 1971. *Silent messages*. Vol. 8. Wadsworth Belmont, CA.

[36] Marvin Minsky. 1980. Telepresence. *OMNI magazine* June (1980). https://web.media.mit.edu/~minsky/papers/Telepresence.html

[37] Carman Neustaedter, Gina Venolia, Jason Procyk, and Daniel Hawkins. 2016. To Beam or not to Beam: A study of remote telepresence attendance at an academic conference. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, San Francisco, USA, 418–431.

[38] Martin T. Orne. 1996. Demand characteristics and the concept of quasi-controls. In *Introducing psychological research*. Springer, 395–401.

[39] R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

[40] Arsénio Reis, Márcio Martins, Paulo Martins, José Sousa, and João Barroso. 2019. Telepresence robots in the classroom: The state-of-the-art and a proposal for a telepresence service for higher education. In *Technology and Innovation in Learning, Teaching and Education*, Vol. 993. Springer International Publishing, Thessaloniki, Greece, 539–550. https://doi.org/10.1007/978-3-030-20954-4_41

[41] Jennifer M. Riley, David B. Kaber, and John V. Draper. 2004. Situation awareness and attention allocation measures for quantifying telepresence experiences in teleoperation. *Human Factors and Ergonomics In Manufacturing* 14, 1 (2004), 51–67. https://doi.org/10.1002/hfm.10050

[42] MHD Yamen Saraiji, Charith Lasantha Fernando, Yusuke Mizushina, Youichi Kamiyama, Kouta Minamizawa, and Susumu Tachi. 2014. Enforced telexistence: teleoperating using photorealistic virtual body and haptic feedback. In *SIGGRAPH Asia 2014 Emerging Technologies*. ACM, Shenzen, China, 1–2.

[43] Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Vol. 142. Citeseer.

[44] R M Taylor. 1989. Situational Awareness Rating Technique (SART): The development of a tool for aircrew systems design. In *Proceeding of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478*. NATO AGARD, Neuilly-sur-Seine.

[45] Alfonso Torrejon, Vic Callaghan, and Hani Hagras. 2013. Selectable Directional Audio for Multiple Telepresence in Immersive Intelligent Environments. In *2013 9th International Conference on Intelligent Environments*. IEEE, Athens, Greece, 181–187.

[46] Daniel M. Wegner and Toni Giuliano. 1982. The Forms of Social Awareness. In *Personality, Roles, and Social Behavior*. Springer New York, New York, NY, 165–198. https://doi.org/10.1007/978-1-4613-9469-3_6

[47] Brenda K. Wiederhold. 2020. Connecting through Technology during the Coronavirus Disease 2019 Pandemic: Avoiding "Zoom Fatigue". *Cyberpsychology, Behavior, and Social Networking* 23, 7 (2020), 437–438. https://doi.org/10.1089/cyber.2020.29188.bkw

[48] Deirdre Wilson and Tim Wharton. 2006. Relevance and prosody. *Journal of Pragmatics* 38, 10 (Oct. 2006), 1559–1579. https://doi.org/10.1016/j.pragma.2005.04.012

[49] Bob G. Witmer and Michael J. Singer. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* 7, 3 (June 1998), 225–240. https://doi.org/10.1162/105474698565686

[50] Carolyn Reinach Wolf. 2020. Virtual Platforms Are Helpful Tools but Can Add to Our Stress. https://www.psychologytoday.com/blog/the-desk-the-mental-health-lawyer/202005/virtual-platforms-are-helpful-tools-can-add-our-stress

[51] Matthew Wong and Ramani Duraiswami. 2021. Shared-Space: Spatial Audio and Video Layouts for Videoconferencing in a Virtual Room. In *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, Bologna, Italy, 1–6.

[52] Philip Yaffe. 2011. The 7% rule: fact, fiction, or misunderstanding. *Ubiquity* 2011, October (2011), 1.

[53] Holly A Yanco and Jill Drury. 2004. "Where am I?" Acquiring situation awareness using a remote robot platform. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, Vol. 3. IEEE, The Hague, Netherlands, 2835–2840. https://doi.org/10.1109/ICSMC.2004.1400762

[54] Megan Zimmerman, Shelly Bagchi, Jeremy Marvel, and Vinh Nguyen. 2022. An Analysis of Metrics and Methods in Research from Human-Robot Interaction Conferences, 2015–2021. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, Sapporo, Japan, 644–648. https://doi.org/10.1109/HRI53351.2022.9889595

A QUIZ QUESTIONS AND ANSWERS

Tables 8 and 9 are the complete sets of quiz questions asked by the quiz master in simulation for each video used, along with the two answers spoken simultaneously; that of the contestant targeted by the question, and of the "distractor" (see Section 4).

Quiz question	Target	Target's answer	Distractor	Distractor answer
Which European city hosted the 1936 Summer Olympics?	Annie	London	**Maria**	**Berlin**
How far is the target in Olympic Archery?	John	70 metres	**Annie**	**80 metres**
When is World Literacy Day celebrated?	John	October 7	**Susan**	**September 8**
What was invented by James Dewer in 1872?	John	Steam pump	**Susan**	**Thermos flask**
Where is the largest bowling alley?	**Elizabeth**	**Japan**	Maria	Minnesota
Who was the first Tsar of Russia?	**Susan**	**Ivan the Terrible**	John	Catherine the Great
Where is the oldest tree in the world?	Elizabeth	Colorado	**Maria**	**California**
What is the tallest mountain in Canada?	Annie	Mount Columbia	**John**	**Mount Logan**
Who wrote Flowers for Algernon?	**Elizabeth**	**Daniel Keyes**	Maria	Isaac Asimov
What colour is the Mallow flower?	Susan	Pink	**John**	**Purple**

Table 8. Quiz questions asked during Video 1. Correct contestant and answer for each question shown in **bold**.
Quiz master: "Carlos", Contestants: "John", "Elizabeth", "Annie", "Susan", "Maria"

Quiz question	Target	Target's answer	Distractor	Distractor answer
What country was previously called Formosa?	**Amy**	**Taiwan**	Paul	Indonesia
How many chukkers are there in a polo match?	Amy	4	**Michael**	**6**
Dendrophobia is the fear of what?	**Carol**	**Trees**	Michael	Teeth
What was the first movie to be rated PG-13?	Michael	Gone with the Wind	**Carol**	**Red Dawn**
What is the surface area of the Earth?	**Paul**	**510 million km**	Emma	197 million km
When was the first digital computer invented?	Amy	1960	**Michael**	**1946**
What species of land animal has the largest eyes?	Paul	Elephant	**Emma**	**Ostrich**
Who was the first performer at Woodstock?	Emma	Fleetwood Mac	**Paul**	**Richie Havens**
Where are the most Trappist breweries?	Michael	Denmark	**Carol**	**Belgium**
How many countries are in the Southern Hemisphere?	Emma	35	**Paul**	**32**

Table 9. Quiz questions asked during Video 2. Correct contestant and answer for each question shown in **bold**.
Quiz master: "Megan", Contestants: "Michael", "Emma", "Amy", "Carol", "Paul"

B OPEN PARTICIPANT FEEDBACK

The feedback in the following tables was provided via post-experiment questionnaire; positive feedback in Table 10, negative in Table 11. This feedback was used for quantitative analysis in 5.2.

Feedback item	*Category*
Mono group	
The ability to replicate a real world environment by aiding the auditory senses with visual to create a more complete picture	Immersion / Ability to look around
Fun to try and work your way around the posed challenge in creative ways	Challenging/Gamelike
Questions and answers very easy to follow, visuals clear, instructions and questions easily understood - no confusion throughout the experiment as to what I had to do	Other
Sound very clear, felt that I had good awareness of what was happening	Other
Being able to see whos talking	Immersion / Ability to look around
the instructions were very clear	Other
I liked being able to look around the room.	Immersion / Ability to look around
a bit like a game for me	Challenging/Gamelike
it was cool being able to move it around like you are in the room	Immersion / Ability to look around
being able to look around	Immersion / Ability to look around
I liked being able to look around at who was talking, having control over that felt immersive.	Immersion / Ability to look around
ability to see who is speaking	Immersion / Ability to look around
wanted to concentrate more to get answers right	Challenging/Gamelike
I learnt things!	Novel
Spatial group	
A new concept I've never experienced	Novel
The directional audio	Spatial audio feedback
The controls, being able to witness the surroundings	Immersion / Ability to look around
I really wanted to try to get the answer correct and that made me look around the room a lot.	Challenging/Gamelike
I liked the way you could move around, although I don't think it was necessary to answer the questions.	Immersion / Ability to look around
Was interesting that it was possible to tell from which direction the sound was coming.	Spatial audio feedback
In was an interesting idea	Novel
the audio came from different directions	Spatial audio feedback
The quiz questions	Challenging/Gamelike
The clarity of direction where the sound come from	Spatial audio feedback
real time 360 field of view, very useful for quiz/ conference viewing	Immersion / Ability to look around
the way I could identify where the voice was coming from even though the camera was pointing somewhere else	Spatial audio feedback
easy to navigate	Other
immersion	Immersion / Ability to look around
Questions were unpredictable which made this more engaging	Challenging/Gamelike
Spatial audio let me locate the speaker much easier.	Spatial audio feedback
Getting to see the whole room and everyone involved	Immersion / Ability to look around
Got more used to looking around the room as the experiment continued	Immersion / Ability to look around

Table 10. Positive feedback provided in answer to the question *"Is there anything in particular that you liked about the experience?"*

Feedback item	*Category*
Mono group	
Not having a wider field of view - not being able to see more than one person at a time	Field of view too small
Not having a 360 degree view and taking the time to shift between people.	Field of view too small
Was a bit strange being in the centre of a group, almost like I was being looked at from every direction	Other
difficult to differentiate the girl voices, I would need to see who speaks	Audio was poor (increasing difficulty)
sound quality!	Audio was poor (increasing difficulty)
sound had a bit of an echo to it	Audio was poor (increasing difficulty)
Sometimes I tried to lipread the participant that was asked to answer and the video lagged slightly making that difficult.	Audio was poor (increasing difficulty)
using the arrows to move left and right. in a real environment I would just position my body, move my head or just my eyes	Other
when answered together, I could only hear one answer rather than both	Audio was poor (increasing difficulty)
frustrating when cant make out answers as someones talking over	Audio was poor (increasing difficulty)
Found it rather annoying cause I like to see who is answering while they answer, but it takes a while to shift focus to a person and figure out who answered till the teachers says it out loud.	Slow camera panning speed
Spatial group	
slow panning speed of the visuals	Slow camera panning speed
The wait time between the questions was quite long. If the video paused until the questions was answered, and then resumed straight away. I think it would've been slightly better.	It was repetitive / boring
No vertical control	No vertical camera control
you cant see up and down	No vertical camera control
the overlay of two answers required a lot of concentration and often the choices were not clear until the graphic with the answer selection appeared on screen.	Audio was poor (increasing difficulty)
the talking over each other so it was hard to concentrate on correct answers	Audio was poor (increasing difficulty)
wish the arrows would move quicker	Slow camera panning speed
It was somewhat repetitive	It was repetitive / boring
How I experienced volume differed to how I experience volume normally. I expected the people behind me to be quieter.	Other

Table 11. Negative feedback provided in answer to the question *"Is there anything in particular that you disliked about the experience?"*