2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

A comparison of audible, visual, and multi-modal communication for multi-robot supervision and situational awareness

Richard Attfield[1] , Elizabeth Croft[2] , and Dana Kulić[3]

Abstract— **Multi-robot supervision becomes increasingly cognitively demanding as the ratio of robots to human supervisors rises, potentially leading to situational awareness (SA) losses and robot system failures. Nonverbal cues have been employed to direct supervisor attention and prevent awareness loss in diverse human-computer interaction (HCI) settings. This paper compares the effects of uni-modal and multi-modal audiovisual nonverbal cues on supervisor SA in a multi-robot supervision task. In a simulation-based navigation scenario, 50 participants monitored a multi-robot mission and responded to supervision requests from the robots. We evaluated supervisor SA using response speed and the situational awareness global assessment technique. Results demonstrate that supervisor awareness hinges on the communication method employed by the robots, with greater significance observed at higher awareness levels and when the robot-to-human ratio is higher. Findings also indicate the utility of sonification mapping in human-multi-robot interactions and the benefits of multi-modal cues for sustaining awareness during multi-robot supervision.**

I. INTRODUCTION

Human supervision of a multi-robot system (HSMRS) can be cognitively demanding, especially as the ratio of robots to humans increases. To ensure effective supervision, it is important to consider the mental resources of the human supervisor. Improper accounting of these resources can result in an unsatisfactory experience for the supervisor, automation misuse [1], and robot neglect [2]. One of the most critical dimensions of human-robot interaction (HRI) design that can affect the quality of interaction and a supervisor's cognitive load is the method of communication [3]. While experts have an enhanced ability to predict and act in anticipation of future events [4], reliance on expert users limits the system's applicability. As multi-robot systems become more widespread, people from varied backgrounds will be expected to interact with these systems. Therefore, it is important to study how diverse participants (not just highly trained experts) can effectively supervise robot teams. These supervisors will need to maintain situational awareness (SA), even during periods when their input may not be required and their focus may waver. Thus, appropriate communication strategies must convey pertinent information in easily understood ways, even to non-experts, and effectively orient supervisor attention following periods of boredom or distraction.

There are many ways in which communication may be designed considering human mental resources. One important aspect is the utilised modality, or modalities, with visual, auditory, and tactile modes most widely used in HRI [5]. For the task of robot supervision, previous methods have used screen-based [6], audible [7], haptic [8], mixed reality [9], and multimodal [10] interfaces. While there are benefits to multi-modal communication [3], [11], these advantages are not universal to all task domains [12]. Within supervisory human-computer interaction (HCI) applications, nonverbal cues have been used to effectively direct human attention while considering their mental resources [13]. While many of these works have designed their communication methods to target human cognitive factors, there has been little research to date focusing on the comparative effects of different nonverbal communication strategies, both uni- and multi-modal, specifically during HSMRS. In particular, the effect of different modalities on the supervisor's SA and the impact of the robots-to-human ratio on this effectiveness are understudied. This paper aims to address this gap by comparing the effects of audible, visual, and audiovisual nonverbal communication on supervisor performance during varied simulated HSMRS scenarios, focusing on the supervisor's SA at different cognitive levels.

For this study, four communication strategies were created based on audio-only, visual-only, and audiovisual status signals and implemented within simulated HSMRS scenarios based on a group navigation task. As well as investigating different modalities, two audible methods are compared based on different sonification mappings. Sonification mapping is the nonverbal representation of robot states or actions using sounds, often parameterized to relate a characteristic of a robot's status (e.g., valence) proportionally to an aspect of communication (e.g., volume). We recruited 50 participants, who were asked to monitor the pre-recorded scenarios and to respond to supervision requests. Users were also asked to answer questions regarding the robots' statuses using the Situational Awareness Global Assessment Technique (SAGAT) [14]. The metrics used to analyze the communication methods considered awareness at different levels, from lower-level awareness of a supervision request to a higher-level understanding of the origin and meaning of an alert. The results show that the communication method did not significantly contribute to differences in the time taken by participants to respond to supervision requests, representing the lowest level of SA. For higher levels of SA, however, such as the number and identities of robots requesting supervision, the communication type was found to significantly affect the participants' situational awareness,

*Richard Attfield and Dana Kulić are with the Faculty of Engineering at Monash University, richard.attfield@monash.edu, dana.kulic@monash.edu.

*Elizabeth Croft is with the Faculty of Engineering and Computer Science at the University of Victoria, ecroft@uvic.ca.

*D. Kulić is supported by an Australian Research Council Future Fellowship (FT200100761).

with the audiovisual strategy showing the best performance on all metrics. The effects of the different communication methods were also found to be more substantial when supervising a larger team of robots, when the challenge of maintaining awareness of multiple agents is greater.

II. RELATED WORK

Human supervision of multi-robot systems. Multi-robot teams possess several advantages over single-robot systems including efficiency, cost, and fault-tolerance [15]. Their ability to spatially distribute a range of sensors and actuators makes them well-suited to tasks such as exploration and search and rescue [4]. While improvements in robot autonomy have increased the functionality of these teams, full autonomy is not always desirable in safety-critical or complex situations. By incorporating a human in the system, HSMRS has been shown to increase team versatility and reliability [4]. This supervision can, however, be challenging due to the high mental workload associated with monitoring multiple agents, which can lead to loss of SA [16] and degraded team performance [17]. SA has been identified as a critical metric related to human cognition during multi-agent supervision [18]. In one of the most significant test applications of HSMRS to date, the 2019 DARPA Subterranean (SubT) Challenge, SA was highlighted as one of the most considerable challenges during the supervised navigation task [19].

Situational Awareness. SA was defined first by Endsley as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future", in the general context of humans interacting with automated systems [20]. Endsley also developed several tests of SA, including the SAGAT and Situational Awareness Rating Technique (SART) tests [14]. SA may be compromised by information overflow, over-reliance on automation, or cognitive factors [21]. Cognitive factors, such as distractions and interruptions, are major factors in human error [22] and can contribute to change-detection failure and a loss of awareness [23]. Several approaches have been used to mitigate awareness losses. Visualization approaches, such as Delaunay triangulation, have focused on the spatial arrangement of data within interfaces to maintain awareness [24]. For multi-robot supervision, Petersen proposed the concept of Situational Overview (SO), which seeks to avoid overload by focusing on high-level information that applies to the overall team goals and forgoing lower-level information [25]. Alternatively, Avetisyan et al. showed that communication approaches have different effects at different SA levels, advocating a variable approach to interactions during autonomous vehicle supervision depending on the targeted SA level [26]. For the application of HSMRS, other approaches have considered what or when to communicate to avoid awareness losses, e.g. in [25]. Less attention has been paid to how to choose communication modalities and their effects on different awareness levels.

Nonverbal communication. While screens remain a major tool for communication interfaces within HRI [16], interaction methods have also been developed using expressive lights [27], gestures [28], tactile sensors [29], and nonverbal sounds [7], [30]. While often explicit, implicit communication has also been used to maintain constant awareness while supervising a robot, e.g., using virtual shadows [31]. Gunn et al. showed that for vigilance-requiring tasks, e.g. monitoring a robot's status, sensory displays, such as nonverbal abstract cues, are more effective than cognitive alternatives that may require symbolic interpretation [32]. Within HCI supervision tasks, visual, audible, and tactile nonverbal cues have been effectively utilised [13]. While these methods are less explicit and therefore prone to ambiguity [33], they can be used to convey information with greater efficiency, appeal, and universality in appropriate contexts [28], [34]. One method applied to communication using nonverbal sounds in HRI is sonification mapping. This is the process of representing robot state information using sounds, and can be used to communicate states, actions, and emotions [35]. Parameterized sonification mapping, the process of identifying continuous variables of a sound that can be altered to convey different meanings, has been effectively used in HRI [36], [37].

Alert systems and multi-modal communication. In critical supervision situations, only a short window is available to appropriately orient the supervisor's attention [38]. Supervising multiple robots, which may require a user to quickly identify a specific robot whilst simultaneously giving a command, benefits from interaction schemes that utilise multi-modal inputs from the human to the robots [39]. Multiple interfaces have been designed accordingly [40], including audiovisual cues [41]. There has been less focus, however, on the comparison of different robot-to-human communication modalities, particularly targeting the important metric of SA. In another critical supervision situation where humans are the recipients of communication, namely, semi-autonomous driving, multi-modal communication has been found to produce fewer errors and missed prompts than uni-modal communication [42]. Alternatively, Bazilinskyy et al. found that while multi-modal communication produced better results for high-emergency driving alerts, uni-modal audio communication was preferable during low-emergency alerts [43]. Additionally, Chen et al. observed that prompts that included the contextual channel (in the case of driving, the visual surroundings) caused a higher mental workload than prompts delivered using only alternate channels [44]. Similar findings have been obtained in multi-UAV supervision studies, which found that supplementing visual information with audible communication to avoid crowding the visual channel improved task performance [45], [46]. Within HRI, Abich and Barber found that while increasing the dimensionality of communication can improve task performance metrics, it can also lead to an increase in mental workload and subjective effort compared to uni-modal approaches [47].

The above literature review highlights several aspects of communication yet to be explored for HSMRS. While multi-modal communication possesses advantages, there is a risk of oversaturating the interaction channel; these advantages are juxtaposed with the benefits of uni-modal communication under specific contexts. Nonverbal abstract cues have been advocated for monitoring tasks, however, their use for HSMRS remains under-explored, particularly the method of sonification mapping which has been effectively developed for other HRI applications. When targeting SA, interaction approaches that either ignore low-level events or vary the interaction scheme accordingly have been used to mitigate awareness losses during HSMRS and HCI. This paper compares different nonverbal communication methods, using visual communication, audible sonification mappings, and multi-modal audiovisual cues, to determine how effective multi-modal cues are for maintaining SA during HSMRS and how this effect differs among awareness levels.

III. METHODOLOGY

A simple navigation scenario was designed to simulate an exploration or search and rescue task, where a supervisor monitors a robot team via a map display. Map displays, often supplemented with additional communication interfaces, are a standard in robot control interfaces [48]. In such scenarios, supervisors may need to maintain awareness for extended periods without an event that requires their intervention, increasing the risk of awareness loss due to distraction or boredom. The implemented scenario, while simple, was designed to capture the effects of mental fatigue during multi-agent supervision. Consequently, the findings on robot communication's impact on mental workload and awareness are relevant to various multi-robot supervision applications.

In the designed tasks, participants monitored the robots for over 60 seconds before a supervision event occurred, providing ample opportunity for boredom or distraction. The robots' communication was used to supplement the contextual information provided by the map display, alerting the user to important events. The environment, a maze-like structure, featured robots slowly navigating their way from the centre to various exit points. It was implemented in Gazebo with a team of simulated Turtlebot robots, each identified by colour. Robots had designated start and end point within the environment, marked by coloured walls at those locations. Figure 1 shows the layout of the environment.

The maze featured structures and objects designed to impede the robots, including unpredictable environmental changes such as moving or appearing objects, which triggered supervision events requiring participant attention. Two scenarios were recorded: one with a small team of two robots and another with a large team of eight robots. The larger team size was selected to be at the edge of comfort when monitoring multiple agents, as the span of human apprehension is typically limited to $7(\pm 2)$ entities [49]. The smaller team served as a baseline comparison with two robots. The scenarios were designed to be relatively slow and boring, allowing the participants to lose interest or become distracted, as could occur in real supervision scenarios.

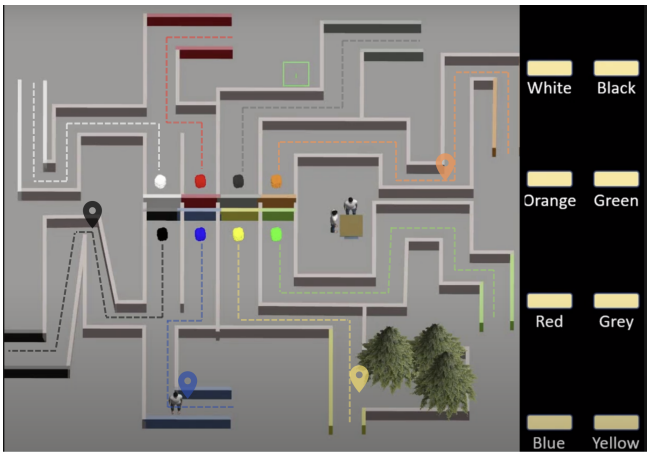

Fig. 1. Starting position of the robots during the large team scenario. A visual interface is included in this example, as seen to the right of the map. The coloured lines and pins are included in this figure to indicate the robots' paths and locations of the supervision requests, respectively, and were not shown to the participants. The supervision events were: (i) the orange robot is blocked suddenly by a spherical object that appears in its path, (ii) the blue robot is blocked by a standing person, (iii) the yellow robot is impeded by a falling tree, and (iv) the black robot approaches a narrow point through which it might not fit. The two-robot scenario included only the white and black robots, with slight differences in the environment.

The two recorded scenarios were embedded in an online survey using Qualtrics. Prior to the first scenario, participants were introduced to the robots and their communication method. They then viewed both scenario videos in a randomized order and responded via key press whenever a new supervision request was made, with *response times* recorded. A SAGAT test assessed their SA by unexpectedly cutting away during the recording. When half of the robots requested supervision (in both cases, after around 90 seconds), the video was stopped and the participants were automatically moved to the next stage of the survey. Participants were first asked how many robots required attention, and, subsequently, which robots required attention.

For the robots' supervision requests, four different communication strategies were developed. The strategies were all designed as status signals as opposed to alert signals, meaning that communication was maintained throughout the interaction. Status signals allow supervisors to maintain awareness of a robot without constant monitoring, which aids in context switching [31]. Additionally, a loss of communication, which would otherwise amount to a significant awareness loss, could be observed. This is a likely scenario in environments where communication bandwidth may be a concern, as was observed in the SubT challenge [6]. Sample recordings of these strategies can be viewed within the supplementary material[1]. Each participant was randomly assigned to one of these communication methods, used for both supervision tasks.

Robot sonification strategy. The audible communication

[1]https://www.youtube.com/@richardhatfield1186/videos

strategies were developed using Audacity. Using a simple looped beep, each robot's need for supervision was proportional to its beep frequency. As a robot approached an impediment to its progress its need for assistance increased along with the frequency of its beep. If completely stuck, the beep frequency would reach its maximum. Such abstract audible communication methods have been previously used effectively within HCI and HRI interfaces [13], [36]. In this approach, although starting with their beeps synchronised, each robot communicated individually regardless of how its communication affected that of the other robots.

Team sonification strategy. This approach was designed to act as a team-wide adaptation of the robot sonification strategy. Using the same sound, the beeps were instead designed to be synchronous with each other when the robots were progressing and asynchronous when some portion of the team was in difficulty. When a robot encountered a problem and required attention, its beep would increase in frequency until reaching a π-shift in its phase, whereupon it would return to the original frequency but in an off-beat rhythm. This phase shift was designed to break the synchrony of the team's overall in-time beeping, and result in the robot requesting supervision to be beeping off-beat with the rest. As all robots used the same beep, the beep volume was a function of the number of robots beeping in phase together. The louder the off-beat beep, the more robots were at their maximum supervision request. This team-wide strategy was intended to be a clearer indication of the overall team status than the individual robot sonification strategy, where all robots requesting supervision would beep at different frequencies and out of time with each other. To account for the case in which the whole team encountered issues, a "baseline" beep, unaffected by the progress of the robots, was included within the interface to provide the necessary contrast.

Visual communication strategy. A simple visual interface was implemented alongside the map interface to provide supplementary information about the robots' statuses, following a common convention [45]. Positioned next to the simulation videos, as shown in Figure 1, each robot was given a stacked status bar that ranged from one to four blocks. A single block above a robot's name indicated that it was progressing without the need for supervision. When a robot encountered a difficulty, its status bar increased to a maximum of four blocks. Colour has also been previously used to represent robot status [50], but status bars were used in this work to avoid confusion as colours are used to identify individual robots. Grouping the robots' status representations alongside the map interface ensured that awareness of the whole team could be maintained, even when the robots were spatially dispersed or not simultaneously viewable on the map.

Audiovisual communication strategy. This strategy was used to test the efficacy of a multi-modal approach. In this method, an audiovisual approach was created by combining the visual communication method with one of the audible strategies. The team sonification method was used, as this was hypothesized to be the superior audible approach.

Recruitment. Ethics approval for this study was provided by Monash University (project ID 37750), with all participants first providing consent. A power analysis predicted a required participant base of 46 people. To accommodate for failures such as improper completion of the survey, a total of 50 participants were recruited. Of the participants, 76% ranged in age from 18 to 29, with the oldest being between 40 and 49. We did not require our participants to have prior experience with robots or robot supervision. On a 5-point Likert scale (1 = no experience, 5 = extensive experience), the majority reported limited previous experience with robots (M = 2.1, SD = 1.1).

Hypotheses. With the designed interaction strategies, we sought to investigate the relative effectiveness of audio and visual communication for maintaining awareness within an HSMRS scenario. Given a null hypothesis that supervisor performance and SA would not be significantly affected by the robots' communication strategy, we hypothesized the following:

H1 Supplementing visual screen-based information with audible communication will improve the supervisor's situational awareness and response times compared to either uni-modal communication.

H2 A sonification model designed to form a team-wide communication channel will improve supervisor awareness, especially as the size of the robot team increases, when compared to a sonification model that only focuses on individual robots.

IV. RESULTS

The participants' performances were measured by the speed of their responses and their awareness of the robots' statuses. The impacts of three factors on these metrics were evaluated: the robots' communication method, the team size, and the scenario order (whether the scenario in question was the participant's first or second task). The significance of these effects was determined using multi-factor analysis of variance (ANOVA), with a threshold of p = 0.05. To confirm the validity of these tests, it was first confirmed that all distributions were Gaussian and that the variances of distributions satisfied the conditions of homogeneity.

The results are discussed in order of ascending awareness requirements. First, the participants' response timeliness to the supervision requests is analyzed. Awareness of these requests necessitated no awareness of the task or robots, and was merely indicated by a response to a prompt. Second, the participants' awareness of the number of robots requiring supervision prior to the video cut-off is investigated. Thirdly, the participants' understanding of which robots were making these requests is analyzed. This level required participants to maintain an awareness of the communication, team-wide status, and individual robots' statuses. The main performance metrics were: i) *response times* to supervision requests, ii) error when selecting the number of robots making supervision requests (*#SR*), and iii) error when selecting the identity of

these robots (*#RI*). Following these analyses, the results of a post-experiment question set regarding challenges associated with the tasks and communication are presented.

A. Responses to supervision requests

Response times were recorded as the time between the beginning of a robot's supervision request and the participant's corresponding keypress. Figure 2 shows the distributions of these times across the four communication strategies. The mixed-modality audiovisual approach exhibits a much tighter distribution and lower average than the other strategies, however, a 3-way ANOVA test did not meet the significance threshold for the communication strategy, as shown in Table I. Both the robot team size and task order did contribute to the participants' performance, with p-values of 1.07e-05 and 0.00176, respectively. Participants were found to respond faster when supervising the larger robot team than the smaller team, with average response times of 3.6 and 5.1 seconds, respectively. During the second task, participants' *response times* increased by an average of 0.8 seconds compared to their first attempt.

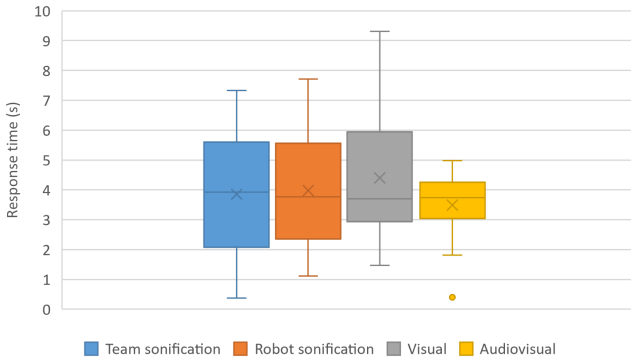

Fig. 2. Distributions of the participants' *response times* across the different communication strategies. The distribution means are indicated by the 'x' on all box plots. The mixed-modality audiovisual approach shows the lowest mean *response time* and the smallest variation.

TABLE I

3-WAY ANOVA TEST BETWEEN SUPERVISION RESPONSE TIMES DISTRIBUTIONS.

Factor	Df	F value	Pr (>F)	Sig.
Communication type	3	1.65	0.173	
Team size	1	21.386	1.07e-05	***
Order	1	10.3	0.00176	**
Team size:Order	1	4.736	0.00306	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

B. Awareness of the number of supervision requests

Two measurements of the participants' SA were obtained using the SAGAT test. The first was the reporting error in the number of robots requesting supervision immediately before the cut-off point (*#SR*). A 3-way ANOVA test revealed that the team size had the strongest impact on this metric (p = 1.9e-10). Participants correctly identified the number of robots requiring supervision with an accuracy of 80% when supervising the smaller team, compared to 38% with the larger team. The task order (whether or not the supervision scenario in question was the first or second attempted by the participant) was found to have no significant effect.

The large effect of team size on *#SR* error created large variances within the distributions of some of the communication types, which violated the requirements of homoscedasticity. The effect of the communication types was therefore analyzed within the smaller and larger team sizes separately. With the smaller team of two robots, no significant difference was found between the different communication methods. With the larger team, however, the communication strategy was found to have a significant effect (p = 5.21e-06). with the audiovisual approach achieving the lowest average error. These distributions are shown in Figure 3.

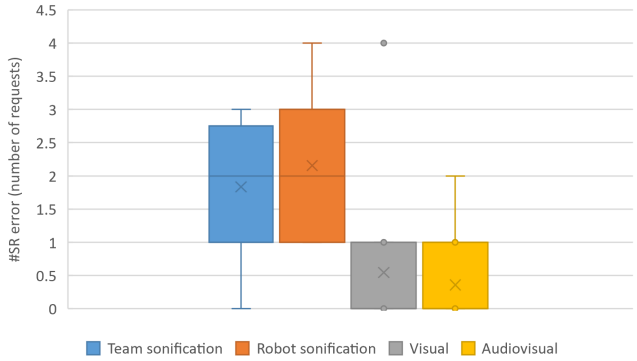

Fig. 3. Distributions of the participants' *#SR* errors when supervising the larger robot team. The methods that included a visual display showed lower errors, with the mixed-modality approach achieving the best results. The audible-only methods show higher error levels, with the team sonification approach outperforming the individual robot sonification approach.

The observed advantage of communication methods that included a visual component was confirmed by a one-way ANOVA test (p = 2.72e-07), with these approaches showing an average *#SR* error of 0.44 compared to an *#SR* error of 2.0 with the audible-only methods. Further post-hoc one-way ANOVA tests were used to test the significance of the differences between the individual communication types, shown in Table II. As multiple tests were performed concurrently, a Bonferroni correction was applied. Six comparisons reduced the significance threshold to p = 0.0083.

TABLE II

POST-HOC ONE-WAY ANOVA TESTS BETWEEN INDIVIDUAL COMMUNICATION STRATEGIES' *#SR* ERROR DISTRIBUTIONS WHEN SUPERVISING THE LARGER ROBOT TEAM.

Factors	p-value	Sig.
Team sonification & Robot sonification	0.3912	
Team sonification & Visual	0.00527	*
Team sonification & Audiovisual	3.2e-05	***
Robot sonification & Visual	0.00165	**
Robot sonification & Audiovisual	1.35e-05	***
Visual & Audiovisual	0.611	

Signif. codes: 0 '***' 0.000167 '**' 0.00167 '*' 0.0083 '.' 0.0167 ' ' 1

C. Awareness of the identity of robots requesting supervision

The second metric of the SAGAT test measured the number of robots requesting supervision that were not correctly identified (*#RI* error). Figure 4 shows the distributions of *#RI* errors across the communication methods, while Table III shows the results of a 3-way ANOVA test. The team size again had the most significant impact (p = <2e-16), with the *#RI* errors of 0.16 and 1.64 robots for the small and large teams, respectively. The communication type used was also found to significantly affect the participants' performance (p = 0.0407). The two methods that included a visual component corresponded to higher awareness than the audible-only methods (p = 0.0288), with average *#RI* errors of 0.76 and 1.04 robots, respectively. Further post-hoc one-way ANOVA tests were used to compare the individual communication strategies, with only the robot sonification and audiovisual methods showing a significant comparison following a Bonferroni correction (p = 0.0008). The robot sonification method produced the highest average error of 1.15 robots, while the audiovisual method yielded the lowest average error of 0.64 robots.

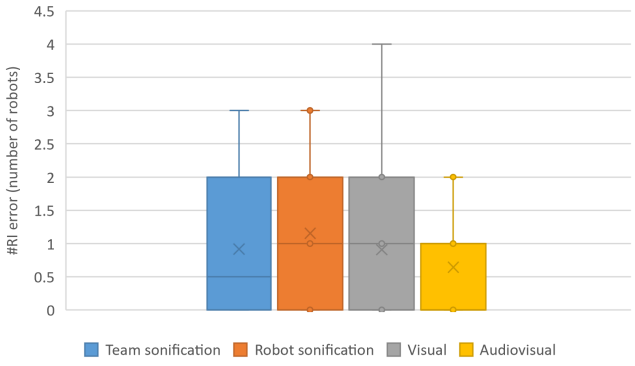

Fig. 4. Participants' ability to correctly identify which robots were requesting supervision following the cut-off point (*#RI*). The mixed-modality method again shows the best performance.

TABLE III

3-WAY ANOVA TEST BETWEEN DISTRIBUTIONS OF ERRORS WHEN IDENTIFYING THE ROBOTS REQUESTING SUPERVISION.

Factor	Df	F value	Pr (>F)	Sig.
Communication type	3	2.881	0.04068	*
Team size	1	113.858	<2e-16	***
Order	1	3.402	0.06863	.
Team size:Order	1	8.517	0.00451	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

D. Post-experiment questions

Upon completion of the tasks, participants were asked to report on their experience of the study. Figure 5 shows three of these questions regarding the ease of supervising the two team sizes and of understanding the robots' communication using 5-point Likert scale questions. The team sonification method is consistently ranked as easier to use than the robot sonification method and was the highest-rated method when supervising the larger team. When supervising the smaller team the audiovisual communication approach was the most favoured.

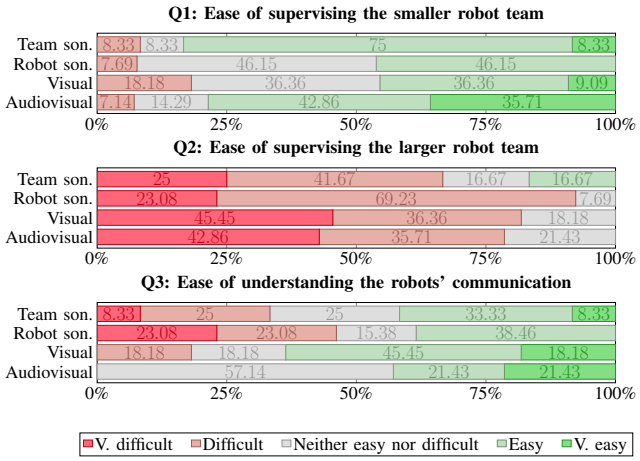

Fig. 5. Post-experiment 5-point Likert scale questions regarding the difficulty of the tasks and understanding the robots' communication. For each communication type, the distributions range from "very difficult" to "very easy" and are displayed as percentages of the total group.

V. DISCUSSION

As expected, team size was the most significant of the three investigated factors. In both SAGAT test metrics, participants performed significantly worse when supervising the larger team. However, *response times* were found to decrease when supervising the larger team. With multiple requests issued by the larger team in quick succession, participants attention likely remained on task after the initial alert, leading to faster responses to subsequent requests. This underscores the influence of distractions on supervisor performance during event-sparse tasks, as responsiveness encompasses orienting the supervisor's attention to the task in addition to their response.

The order of scenarios, whether a result came from the participant's first or second task, significantly impacted their *response times*. While improvements were expected in the second task due to prior experience, results showed a slight performance drop, likely due to a loss of interest. Given the relatively monotonous scenario design, participants may have been less engaged during the second task, leading to reduced effort. This highlights how easily focus can be lost during even short supervisory tasks and underscores the need for communication strategies that effectively reorient attention.

The results indicate that the communication strategy becomes more influential at higher awareness levels. At the lowest level, at which participants had merely to respond to prompts, the communication method did not prove significant. Considering the metric of *#SR*, which required higher SA, the communication strategy did have a significant influence with the visual methods outperforming the audible-only strategies (p = 2.72e-07). The strongest individual differences were observed between the audiovisual method and the two audible-only methods (team sonification: p = 3.2e-05, robot sonification: p = 1.35e-05). At the highest awareness level, regarding awareness of both the team and individual robot

statuses (*#RI*), the communication type was also shown to be significant ($p = 0.041$) with the audiovisual method again showing the lowest average error.

Figure 2 suggests a slight advantage of audible over visual-only communication when re-orienting a supervisor's attention to the task after a prolonged period without a supervision alert. While this highlights the efficacy of audible communication for gaining a supervisor's attention, Figures 3 and 4 show the advantage of utilizing a visual display for maintaining awareness of the robots' statuses. Particularly when considering the overall status of the team, Figure 3 shows a clear superiority of visual displays despite sharing the channel with the contextual information. This division of the visual channel was more consequential at the higher awareness level, where the strong advantage shown by the visual methods diminishes. Participants reported a natural focus on robot statuses over the map interface, leading to accurate tracking of alerts but sometimes losing context. This suggests an advantage in using alternative communication channels. Additionally, the advantages of visual displays could diminish over longer tasks, as it is easier to lose focus of a visual display due to its directional nature. Despite the reported challenges associated with the visual channel, the multi-modal audiovisual approach outperformed all of the uni-modal methods with the lowest average *response times*, *#SR* errors, and *#RI* errors, supporting **H1**.

The varying impacts of the modalities at different awareness levels supports findings from autonomous driving research [26], suggesting that when targeting the lowest awareness level for HSMRS, communication design should consider factors other than SA. User satisfaction, trust in the system, or mental workload are other criteria that could be used to determine an appropriate communication strategy. A similar observation may be made concerning the robot team size, with the communication method shown to be more impactful when supervising a larger team. If the team size is sufficiently small, considerations beyond SA may be used to design the robots' communication. The effects shown by different communication modalities at the higher awareness levels are particularly useful as it is the comprehension of an element in the environment, as opposed to mere perception of it, that is most affected by automation [51].

The results are only partially supportive of **H2**, that a sonification mapping designed to operate on a team-wide as well as individual robot level would enhance performance and awareness compared to an individual-only approach. Participants using the team sonification strategy had lower average *response times* and fewer *#SR* and *#RI* errors; however, the differences were not statistically significant. In post-experiment feedback, participants consistently ranked the team sonification approach as easier to use. When supervising the smaller robot team, 83% of participants within the team sonification group rated the task as "easy" or "very easy", compared to 46% of participants in the robot sonification group. With the larger robot team, these figures were 17% and 0%, respectively. Additionally, 42% of participants found the team sonification method "easy" or "very easy" to understand, compared to 38% for the robot sonification approach. These findings suggest that a multi-robot sonification mapping that includes a team-wide channel may improve user SA and reduce mental workload, although further testing is needed to confirm this.

Although the tasks and scenarios in this study were relatively simple, the observed effects are broadly applicable. As multi-robot systems become more widely used, the cognitive challenges associated with monitoring and reacting to multiple agents simultaneously will remain an important and universal consideration. Two particular aspects of interest will be the inclusion of non-experts in HSMRS and periods of this supervision during which the human supervisor may feel superfluous. Effective communication that is intuitive to diverse users will be necessary to ensure that during such periods, vital awareness is not lost that could impact the performance of the system.

VI. CONCLUSION

This study investigated how communication modality influences human situational awareness during robot team supervision. The results support the hypothesised advantage of the multi-modal audiovisual method compared to the visual and audio-based methods, with the audiovisual approach showing the best supervisor performance and awareness results. The lack of significance found at the lowest awareness level, however, suggests that a less cognitively demanding approach, perhaps utilizing only one modality, would be more appropriate if targeting only this level. Despite achieving slightly superior averages on all three metrics, the team sonification method was not found to be significantly more effective than the robot sonification method, as was hypothesized. Some support for this hypothesis was found in the subjective responses from participants, who consistently ranked the team sonification method as easier to use.

Further exploration into the effects of communication modality on SA during HSMRS could benefit from increased task complexity or duration, potentially allowing for a clearer articulation of a communication method's effects. Using a larger team size well beyond the comfortable range of human apprehension could also provide potentially clearer results. Future work could also aim to address the scalability of the modes of communication. Further improvements could potentially be made by including, for example, a tactile element as well as audible and visual. Additional work could also investigate the benefits of a team of heterogeneous robots using different communication strategies, e.g. different base sounds for a sonification model. An investigation employing physical robots with real-time interactions would also allow for a more grounded substantiation of these findings.

REFERENCES

[1] G. M. Alarcon, A. Capiola, J. Morgan, *et al.*, "Trust Violations in Human-Human and Human-Robot Interactions," in *55th HICSS*, pp. 675–684, 2022.

[2] J. W. Crandall, M. A. Goodrich, D. R. Olsen, *et al.*, "Validating human-robot interaction schemes in multitasking environments," *IEEE SMC A*, vol. 35, pp. 438–449, 7 2005.

[3] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: A survey," *Found. Trends Hum.-Comput. Interact*, vol. 1, no. 3, pp. 203–275, 2007.

[4] D. Bourque, T. Desilva, and N. Otero, "Human Supervision of Multi-Robot Systems," tech. rep., WPI, 2015.

[5] H. A. Frijns, O. Schürer, and S. T. Koeszegi, "Communication Models in HumanRobot Interaction: An Asymmetric MODel of ALterity in HumanRobot Interaction (AMODAL-HRI)," *Int. J. Soc. Robot*, 2021.

[6] K. Otsu, S. Tepsuporn, R. Thakker, *et al.*, "Supervised Autonomy for Communication-degraded Subterranean Exploration by a Robot Team," in *IEEE Aerosp. Conf. Proc*, IEEE, 2020.

[7] L. Zahray, R. Savery, L. Syrkett, *et al.*, "Robot Gesture Sonification to Enhance Awareness of Robot Status and Enjoyment of Interaction," in *29th IEEE RO-MAN*, pp. 978–985, IEEE, 8 2020.

[8] H. I. Son, L. L. Chuang, A. Franchi, *et al.*, "Measuring an Operators Maneuverability Performance in the Haptic Teleoperation of Multiple Robots," *IROS*, pp. 3039–3046, 2011.

[9] M. Walker, H. Hedayati, J. Lee, *et al.*, "Communicating Robot Motion Intent with Augmented Reality," in *ACM/IEEE HRI*, pp. 316–324, 2018.

[10] A. Hong, D. G. Lee, H. H. Bülthoff, *et al.*, "Multimodal feedback for teleoperation of multiple mobile robots in an outdoor environment," *J. Multimodal User Interfaces*, vol. 11, no. 1, pp. 67–80, 2017.

[11] G. Goos, J. Hartmanis, J. Van, *et al.*, *Human Machine Interaction*. Springer, 2009.

[12] S. Oviatt, "Ten myths of multimodal interaction," *ACM*, vol. 42, no. 11, pp. 74–81, 1999.

[13] I. Politis, S. Brewster, and F. Pollick, "Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers," *IJMHCI*, vol. 9, pp. 1–16, 7 2017.

[14] M. R. Endsley, S. J. Selcon, T. D. Hardiman, *et al.*, "Comparative analysis of SAGAT and SART for evaluations of situation awareness," *Proceedings of HFES*, vol. 1, pp. 82–86, 1998.

[15] W. Burgard, D. Fox, R. Simmons, *et al.*, "Collaborative Multi-Robot Exploration," *IEEE ICRA*, vol. 1, pp. 476–481, 2000.

[16] A. Dahiya, A. M. Aroyo, K. Dautenhahn, *et al.*, "A survey of multi-agent HumanRobot Interaction systems," *Rob Auton Syst*, vol. 161, p. 104335, 2023.

[17] N. Schurr, *TOWARD HUMAN-MULTIAGENT TEAMS*. PhD thesis, University of Southern California, 2007.

[18] J. Y. Chen, M. J. Barnes, and M. Harper-Sciarini, "Supervisory control of multiple robots: Human-performance issues and user-interface design," *IEEE SMC C*, vol. 41, pp. 435–454, 7 2011.

[19] T. H. Chung, V. Orekhov, and A. Maio, "Into the Robotic Depths: Analysis and Insights from the DARPA Subterranean Challenge," *Annu. Rev. Control Robot. Auton. Syst. 2023*, vol. 6, pp. 477–502, 2022.

[20] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Hum. Factors*, vol. 37, no. 1, pp. 32–64, 1995.

[21] M. She and Z. Li, "Team situation awareness: A review of definitions and conceptual models," in *LNCS*, vol. 10275 LNAI, pp. 406–415, Springer Verlag, 2017.

[22] D. C. McFarlane and K. A. Latorella, "The scope and importance of human interruption in human-computer interaction design," *INT J HUM-COMPUT INT*, vol. 17, no. 1, pp. 1–61, 2002.

[23] P. J. Durlach, "Change blindness and its implications for complex monitoring and control systems design and operator training," *INT J HUM-COMPUT INT*, vol. 19, no. 4, pp. 423–451, 2004.

[24] C. M. Hoffmann, Y.-J. Kim, R. P. Winkler, *et al.*, "Visualization for Situation Awareness," *CIKM*, pp. 36–40, 2000.

[25] Karen Petersen, *General Concepts for Human Supervision of Autonomous Robot Teams*. PhD thesis, T.U. Darmstadt, 2014.

[26] L. Avetisyan, J. Ayoub, and F. Zhou, "Investigating explanations in conditional and highly automated driving: The effects of situation awareness and modality," *Transp. Res. F: Traffic Psychol. Behav*, vol. 89, pp. 456–466, 8 2022.

[27] K. Baraka, S. Rosenthal, and M. Veloso, "Enhancing human understanding of a mobile robot's state and actions using expressive lights," *25th IEEE RO-MAN*, pp. 652–657, 2016.

[28] C. Breazeal, C. D. Kidd, A. L. Thomaz, *et al.*, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," *IEEE/RSJ IROS*, pp. 708–713, 2005.

[29] A. Haynes, M. F. Simons, T. Helps, *et al.*, "A wearable skin-stretching tactile interface for human-robot and human-human communication," *IEEE RA-L*, vol. 4, pp. 1641–1646, 4 2019.

[30] J. Sterkenburg, M. Jeon, and C. Plummer, "Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons," *LNCS*, vol. 8511, no. PART 2, pp. 633–640, 2014.

[31] A. Boateng, W. Zhang, and Y. Zhang, "Implicit Projection: Improving Team Situation Awareness for Tacit Human-Robot Interaction via Virtual Shadows," in *IROS*, pp. 7945–7952, 2023.

[32] D. V. Gunn, J. S. Warm, W. T. Nelson, *et al.*, "Target Acquisition With UAVs: Vigilance Displays and Advanced Cuing Interfaces," *HFES*, vol. 47, no. 3, pp. 488–497, 2005.

[33] R. Fernandez, N. John, S. Kirmani, *et al.*, "Passive Demonstrations of Light-Based Robot Signals for Improved Human Interpretability," *27th IEEE RO-MAN*, pp. 234–239, 2018.

[34] H. Admoni, T. Weng, B. Hayes, *et al.*, "Robot nonverbal behavior improves task performance in difficult collaborations," *11th ACM/IEEE HRI*, vol. 2016-April, pp. 51–58, 2016.

[35] M. Schmitz, B. C. Fehringer, and M. Akbal, "Expressing emotions with synthetic affect bursts," in *CHI PLAY*, pp. 91–96, Association for Computing Machinery, Inc, 10 2015.

[36] L. Roy, R. Attfield, D. Kulić, *et al.*, "Toward Improving User Experience and Shared Task Performance with Mobile Robots through Parameterized Nonverbal State Sonification," in *Sound and Robotics*, pp. 176–198, CRC Press, 2023.

[37] T. Komatsu, "Toward Making Humans Empathize with Artificial Agents by Means of Subtle Expressions," in *ACII*, pp. 458–465, 2005.

[38] W. Chen, T. Sawaragi, and T. Hiraoka, "Adaptive multi-modal interface model concerning mental workload in take-over request during semi-autonomous driving," *SICE JCMSI*, vol. 14, no. 2, pp. 10–21, 2021.

[39] B. Gromov, L. M. Gambardella, and G. A. Di Caro, "Wearable Multi-modal Interface for Human Multi-robot Interaction," in *IEEE SSRR*, pp. 240–245, 2016.

[40] S. J. Levulis, P. R. DeLucia, and S. Y. Kim, "Effects of Touch, Voice, and Multimodal Input, and Task Load on Multiple-UAV Monitoring Performance During Simulated Manned-Unmanned Teaming in a Military Helicopter," *Hum. Factors*, vol. 60, pp. 1117–1129, 12 2018.

[41] N. Y. Chong, T. Kotoku, K. Ohba, *et al.*, "Multioperator teleoperation of multirobot systems with time delay: Part II - Testbed description," *Presence: Teleoperators and Virtual Environments*, vol. 11, pp. 292–303, 6 2002.

[42] C. Geitner, F. Biondi, L. Skrypchuk, *et al.*, "The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario," *Transp. Res. F: Traffic Psychol. Behav*, vol. 65, pp. 23–33, 8 2019.

[43] P. Bazilinskyy, S. M. Petermeijer, V. Petrovych, *et al.*, "Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays," *Transp. Res. F: Traffic Psychol. Behav*, vol. 56, pp. 82–98, 7 2018.

[44] W. Chen, T. Sawaragi, and T. Hiraoka, "Comparing driver reaction and mental workload of visual and auditory take-over request from perspective of driver characteristics and eye-tracking metrics," *Transp. Res. F: Traffic Psychol. Behav*, vol. 97, pp. 396–410, 8 2023.

[45] C. Y. Wong and G. Seet, "Workload, awareness and automation in multiple-robot supervision," *IJARS*, vol. 14, 6 2017.

[46] S. R. Dixon, C. D. Wickens, and D. Chang, "Mission Control of Multiple Unmanned Aerial Vehicles; A Workload Analysis," *HFES*, vol. 47, no. 3, pp. 479–487, 2005.

[47] J. Abich and D. J. Barber, "The impact of human-robot multimodal communication on mental workload, usability preference, and expectations of robot behavior," *J. Multimodal User Interfaces*, vol. 11, pp. 211–225, 6 2017.

[48] J. Y. Chen, E. C. Haas, and M. J. Barnes, "Human performance issues and user interface design for teleoperated robots," *IEEE SMC C*, vol. 37, pp. 1231–1245, 11 2007.

[49] J. Patel and C. Pinciroli, "Improving Human Performance Using Mixed Granularity of Control in Multi-Human Multi-Robot Interaction," in *29th IEEE RO-MAN*, pp. 1135–1142, IEEE, 2020.

[50] C. M. Humphrey, S. M. Gordon, and J. A. Adams, "Visualization of multiple robots during team activities," *Hum. Factors*, vol. 50, pp. 651–655, 2006.

[51] M. R. Endsley and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Hum. Factors*, vol. 37, pp. 381–394, 6 1995.