https://doi.org/10.17221/73/2024-RAE

Perception of bimodal warning cues during remote supervision of autonomous agricultural machines

Anita Chidera Ezeagba[1], *Cheryl Mary Glazebrook*[2], *Daniel Delmar Mann*[1]*

[1]*Department of Biosystems Engineering, University of Manitoba, Winnipeg, Manitoba, Canada*
[2]*Faculty of Kinesiology and Recreation Management, University of Manitoba, Winnipeg, Manitoba, Canada*

**Corresponding author: danny.mann@umanitoba.ca*

Citation: Ezeagba A.C., Glazebrook C.M., Mann D.D. (2025): Perception of bimodal warning cues during remote supervision of autonomous agricultural machines. Res. Agr. Eng., 71: 69–79.

Abstract: Agricultural machines that are fully autonomous will still need human supervisors to monitor and troubleshoot system failures. Recognising the emergency as soon as possible is crucial to reduce adverse effects. The ability of humans to detect visual, auditory, or tactile cues is usually enabled by warning systems. The effectiveness of different warning cues varies in terms of prompting a quick response. The study's objective was to compare the effectiveness of two bimodal warnings (i.e., visual-auditory and visual-tactile) at eliciting supervisor perception (which equates to level one situation awareness). Twenty-five participants engaged in an autonomous sprayer simulation. Two realistic remote supervision scenarios (i.e., in-field and close-to-field) were used to examine two bimodal warning cues: *(i)* visual-auditory and *(ii)* visual-tactile. The effectiveness of each bimodal warning was assessed based on two measures: *(i)* response time and *(ii)* noticeability. There was no significant difference between the bimodal warning cues in terms of response time when tractor sound was present in the experimental environment (reflecting the in-field remote supervision scenario); however, visual-tactile cues yielded shorter response times than visual-auditory cues when the experimental environment was quiet (reflecting the close-to-field remote supervision scenario). There were no statistically significant differences between visual-auditory and visual-tactile warnings concerning noticeability. Participants' subjective answers indicated they preferred the visual-tactile cues better than the visual-auditory cues. It is concluded that visual-tactile warnings are preferred over visual-auditory warnings to enable perception during remote supervision of autonomous agricultural machines (AAMs).

Keywords: warning systems; situation awareness; human supervision; automated farm machinery

In contemporary society, automation has permeated every facet of existence. Powerful new automation technologies have been launched in many sectors, such as flight management systems for pilots, navigational displays for drivers, diagnostic and surgical aids for physicians, and decision-aiding systems for air traffic controllers (Mouloua et al. 2019; Parasuraman 2000). Numerous advantages have resulted from this technological revolution. Not to be outdone, the agricultural industry uses automation technology (autosteer systems, variable rate technologies, etc.) to carry out various farm tasks. Even with the most recent developments, work is continually being done to increase these machines' operability and efficiency. Currently, the goal of agricultural machine designers is full automation, which would eliminate the need for human intervention for agricultural machines

https://doi.org/10.17221/73/2024-RAE

to navigate and control themselves. It is challenging to completely remove human intervention from the control loop (Schreckenghost et al. 2008), given the variable operating conditions of these autonomous machines (Adamides et al. 2014). Autonomous machines will need an interface like the one outlined by Blackmore et al. (2007) for the human supervisor to communicate with them. These interfaces should give people easily accessible and useful information for their supervisory duties (Dorais & Gawdiak 2003).

According to some research, the human's responsibilities in an autonomous system include assigning tasks, distributing resources, monitoring how tasks are being completed, and intervening through an automation interface in emergencies (Auat Cheein & Carelli 2013; Bechar & Vigneault 2016; Shi et al. 2023). Therefore, fully autonomous agricultural machines (AAMs) must include a human supervisor to oversee the machine's operations (Alexander et al. 2009). Compared with active engagement in that same task, supervision of a task frequently results in decreased situation awareness, increased mental workload, inefficient monitoring, and a worsening capacity for manual control and intervention if automated systems malfunction (Edet & Mann 2021). To act during an emergency, the supervisor must first identify and assess the situation (Peryer et al. 2005). As a result, warning signals must be included in the automation interface to inform the supervisor of what is occurring in their surroundings at any given time. This is known as supporting the supervisor's awareness of the situation.

Situation awareness (SA) is defined as "the perception of the elements in the environment within a span of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley 1995). SA is essential for making decisions and taking actions that work. Regarding a specific task, information is frequently categorised as relevant or important. The self-awareness level is determined by how the individual interprets the data. When operators perceive the information required to complete the task, they reach Level 1 SA (Endsley et al. 2003). To maintain SA, different tasks call for different kinds of information. When attempting to comprehend information pertinent to the task, the operator usually combines their senses of taste, smell, touch, and hearing (Endsley & Garland 2000). When an operator reaches level 2 SA, they clearly understand the significance of perceived information for pertinent objectives (Endsley et al. 2003). To accomplish the current goals, the operator needs to process the data, combine disparate pieces of information, and develop an understanding of the information. Attaining level 3 SA entails seeing information, interpreting it based on pertinent objectives, and forecasting how the situation will unfold (Endsley et al. 2003). To reach level 3 SA, the operator must be thoroughly aware of the existing circumstances and the system's operation. Although all three levels of situation awareness hold significance, the research presented here was limited to examining tactics for bolstering Level 1 situation awareness (or perception).

Non-agricultural devices have used various warning techniques for comparable objectives. Typically, they employ haptic, visual, or auditory modalities (Laughery & Wolgater 2006). The following situations are good candidates for visual stimuli: *(i)* the message is long and complex, and it is anticipated to be seen for some time or referred to later; *(ii)* the individual is stationary; or *(iii)* the surroundings are noisy. Visual stimuli can be delivered as text, images, or bursts of light (Elbert et al. 2018; Edet & Mann 2021). A vocal message, a continuous or periodic tone, or an aural symbol (natural or metaphorical) can all be used as audio warnings (Petocz et al. 2008; Sabic et al. 2017). The omnidirectional nature of auditory warnings makes them useful for drawing and focusing the user's attention on a few different situations: *(i)* when the receiver's position changes, *(ii)* when the message is brief, *(iii)* when the environment is visually cluttered and noisy; and *(iv)* when illumination or barriers obscure the user's vision (Wogalter et al. 2002; Edet & Mann 2021; Chen et al. 2022). An alternative to visual and auditory modalities is tactile (haptic) input. This is particularly useful when the user's eyesight is heavily focused on other task-related activities, the environment is noisy, or the supervisor has visual or hearing impairments (Delavarpour et al. 2019). Also, as tactile warning is less invasive, it is useful in group work environments (Smith et al. 2009).

A variety of criteria, including noticeability, response time, comprehension, recall, hazardousness, perceived urgency, risk, likelihood of injury, likelihood of compliance, and importance, can be used to evaluate the effectiveness of each warning modality (visual, auditory, and tactile) (Wolgater et

https://doi.org/10.17221/73/2024-RAE

al. 2002; Edet & Mann 2021). Response time, which is the amount of time that passes between when the user receives the warning and when the user responds to the warning (through one or more modalities), is the most popular and extensively applied assessment technique (Wolgater et al. 2002; Whelan 2008). A shorter response time would suggest that the warning is more effective than a longer response time. Noticeability, or the capacity to attract attention, is another essential element of warning effectiveness. Warnings must be at least noticed for comprehension and compliance to occur (Young 2002). Saccadic reaction time was used in this study as a measure of noticeability. Rapid eye movements from one focus point to another are called saccades. Since attentional processes impact saccadic eye movements, saccadic reaction time can gauge the attentional state (Braun & Breitmeyer 1988). Compared to a longer saccadic reaction time, a shorter saccadic reaction time would suggest that the warning is more noticeable and effective. Subjective and/or objective measures can be used to measure effectiveness. Open-ended questions, oral interviews, sorting techniques, and evaluations (such as the Likert scale) are examples of subjective approaches, whereas user performance is the basis for objective measurements (Wogalter et al. 1999; Wogalter et al. 2002; Edet & Mann 2021).

Studies have demonstrated that the use of bimodal sensory modalities has advantages when compared to a single modality, particularly when one sensory modality is overloaded due to the primary task or surrounding conditions (Hancock et al. 2013; Haas & Van Erp 2014; White & Hancock 2020). Politis et al. (2014) assessed every combination of auditory, visual, and tactile driver warnings that is multi-modal (i.e., bimodal and trimodal) in two scenarios: *(i)* the lead car braking and *(ii)* the lead car not braking. Their findings show drivers reacted to multi-modal alerts more quickly than unimodal ones. The effectiveness of warning systems (visual, auditory, and tactile, both unimodal and bimodal) in terms of their ability to alert drivers to hazardous situations when faced with various forms of interference (such as devices embedded inside a vehicle, aural noise, and vehicle vibration) was examined by Murata et al. (2013). According to their results, unimodal warnings cause slower reaction times and a lower percentage of right answers than multi-modal warnings. The audio-tactile warning was discovered to be the most successful of all the unimodal and bimodal warning cues. The efficacy of seven warning modalities (i.e., visual, auditory, tactile, visual-tactile, audio-tactile, and audio-visual-tactile) was assessed by Edet and Mann (2021) for four distinct remote location concepts: within-the-field, close-to-the-field, farm office, and outside-the-farmland. Their findings demonstrated that tactile and visual warning modalities (i.e., visual-tactile) produced the shortest response times for remote location concepts with background tractor noise (i.e., within-the-field and close-to-the-field). Compared to unimodal warning signals, the literature consistently shows that employing multiple warnings led to a faster response time. It also showed that the complexity and burden of the task affect the efficacy of bimodal warning signals, leading to higher workload circumstances and better performance when dealing with multiple tasks.

Agricultural machinery typically operates in a dynamic environment, and operators must rely on their senses (i.e., vision, hearing, and touch) to complete their tasks efficiently (Edet & Mann 2021). The most crucial sense operators employ is vision (Macadam 2003; Karimi 2008). It has been reported that the human visual system processes almost 80% of the information required for safe driving (Lee et al. 1998), implying that other senses, including hearing and touch, provide the remaining information. Sensory cue-based signals have been used to alert the operator to potential machine malfunctions. For instance, operators have found that using tactile feedback makes them travel around the field more efficiently (Han et al. 2015; Delavarpour et al. 2019) such that, should there be lateral deviance from the intended course, the steering wheel rattles. The operator perceives this information through their palm, after which they take the appropriate steps to align the machine on the intended course (Edet & Mann 2021). An analogous visual aid is the lightbar, which conveys the machine's lateral deviations to the operator through a horizontal configuration of light-emitting diodes (LEDs) (Ima & Mann 2003). Operators use a variety of auditory input formats to make well-informed choices. One illustration would be the sound the threshing machine produces during harvesting, which varies depending on how much crop is fed into it. By listening to this sound, operators can tell when the threshing unit is overloaded.

https://doi.org/10.17221/73/2024-RAE

Another common source of auditory feedback is the sound produced by a vehicle's engine. Uneven engine load and performance circumstances can cause loud fluctuations in noise (Bilski 2013), and an unusual sound could be a sign of an issue with the engine that needs to be investigated.

With agricultural machinery moving closer to complete autonomy, it would be helpful to identify which bimodal warning method, visual-auditory or visual-tactile, performs best to provide the human supervisor with feedback in a remote supervision task. Therefore, the goals of this study were to *(i)* identify which of the bimodal warning methods, based on response time and noticeability, would be most appropriate to get the attention of an AAM's human supervisor and *(ii)* ascertain whether background noise has an impact on the efficacy of bimodal warning methods.

MATERIAL AND METHODS

Experimental apparatus. The experiment used a tractor cab (Figure 1) in the Agricultural Ergonomics Laboratory at the University of Manitoba to control intrusion from outside noise. The experimental setup consists of two computer monitors positioned one above the other. The bottom monitor displayed the output from a simulation of an agricultural sprayer. In contrast, the top monitor was used to complete the primary internet search task, which will be described in a subsequent section.

Figure 1. A tractor cab was used for the experimental study Participants used the top monitor for the internet search task, while the bottom monitor displayed the simulation of the autonomous agricultural sprayer

An existing simulation of an autonomous agricultural sprayer, described by Edet et al. (2022), was modified to include different warning methods (i.e., visual, auditory, and tactile) and an 'Alert Perceived' button for acknowledging the warnings was added to the interface (Figure 2). The 'Alert Perceived' button was positioned on both sides of the interface, considering left-handed and right-handed individuals (Figure 2A). Two separate bimodal warning methods (i.e., visual-auditory or visual-tactile) were integrated with the operation of the simulated autonomous sprayer.

Experimental protocol. Screening tests, training trials, two experimental sessions, and an end-of-experiment questionnaire made up the experimental protocol. The participants were briefed on the purpose and methods before the experiment. When queries arose, the principal investigator addressed them as clearly as possible. The participants signed a consent form certifying that they had read the terms and circumstances of their participation in the study and provided their consent voluntarily. Ethics approval was obtained from the University of Manitoba's Research Ethics Board.

The participant's visual, auditory, tactile, and comfort levels were evaluated during the screening tests. The visual screening involved showing the participants a sample of the visual warning and asking them to score how well they could see the content. Pure-tone audiometry was performed using a hearing test app (e-audiologia.pl, version 1.1.3) that was obtained from the Google Play store to assess each participant's hearing threshold (≤ 40 dBHL). A tiny, coin-sized 5 VDC motor was used for the tactile assessment. It was placed inside a soft, cushioned band to reduce vibration, and the participant's wrist was wrapped in the band to provide a tactile impression. When the device was purchased, the vibration frequency was measured and found to be approximately 180 ± 10 Hz. This frequency was chosen because it was the optimum frequency for vibrotactile perception (Yim et al. 2007; Edet & Mann 2021).

During the training trials, participants were allowed to become acquainted with the experimental procedures. The training trials involved presenting the warning cues to the participants while they were sitting in the tractor cab, much like in the experimental sessions. After the training, the participants were asked to rate how well they could see, hear, and feel the sensory information presented and

https://doi.org/10.17221/73/2024-RAE

Figure. 2. Altered version of the study's user interface. Visual, auditory, and tactile warning cues were incorporated into the interface to alert the human supervisor to irregularity; this picture displays only the mistake acknowledgement button (Alert perceived) and the visual warning (Alert)

(A) an error-detection visual alert and error acknowledgement button, respectively; (B) a visual display that appears when everything is OK or when the supervisor clicks the Alert perceived button to acknowledge receiving a warning

asked to rate their comfort level in response to the various sensory cues on a post-training evaluation form to make sure they didn't feel uncomfortable in any way that would have introduced bias.

The three steps of the experiment were: *(i)* searching the internet for answers to specific agronomic questions; *(ii)* watching the automation interface's presentation of the simulated autonomous sprayer in action; *(iii)* pressing the 'Alert Perceived' button on the interface screen, the bottom monitor right in front of the participant's seated position, to indicate that an error had been perceived. In a real-world situation, humans managing the autonomous sprayer might be distracted by other things. Therefore, the main goal of the internet search task was to simulate the environment that a human would encounter when monitoring the use of an autonomous sprayer. It also helped determine when participants felt that they had been induced to make an error (i.e. when they looked away from the primary task after the error had been induced). The internet search activity was selected because it accomplished three things: *(i)* it prevented the participant from becoming distracted from the primary screen; *(ii)* it had few distracting elements that would make it difficult for the participant to see the notification; and *(iii)* it was an ecologically valid task (i.e., it is reasonable to expect that farmers would be looking at the internet for farm-related information while remotely supervising their autonomous machines).

During the simulated spraying operation, errors were introduced at random intervals, and the participants were alerted to the faults using bimodal warning cues: visual-auditory or visual-tactile. Subsequently, the participants were required to press the 'Alert perceived' button to indicate that they had noticed the error. An eye tracking device was worn by the participants to track their eye movements (i.e., point of gaze), especially to identify when an error occurred as well as when the participant noticed the error and when clicking the 'Alert perceived' button by the participant. For additional analysis, each participant's response time and level of noticeability (i.e., their ability to identify the error) were considered. The eye tracking setup used a laptop and the SensoMotoric Instruments (SMI) Eye Tracking Glasses 2.0 60 Hz. Previous experiments involving mobile eye tracking have used the SMI system (Caspi et al. 2018;

https://doi.org/10.17221/73/2024-RAE

Hoppe et al. 2018; Niehorster et al. 2020). A three-point calibration and recording were performed using the SMI iViewETG software (version 2.7.1) and a USB cable to connect the glasses to a laptop.

After the eye model adaptation phase of iViewETG was finished, the calibration was initiated. The participant's task was to fixate on the centre of three distinct markers in the stimulus grid: the keyboard, the screen above the participant's seated position, and the screen below the participant's seated position. The principal investigator used the recording laptop's live view of the scene camera to select these areas during each fixation. Each eye camera recorded a video stream at 120 Hz with a 320 × 240 pixels resolution. In comparison, the front-facing scene camera recorded a video stream at 24 Hz with a resolution of 1280 × 960 pixels. It should be noted that the frame rate of the eye cameras does not correspond to the real camera frame rate but rather to the recorded video provided by the iViewETG.

Two sessions of experimentation were completed. No tractor noise was introduced into the tractor cab during the first session, creating a quiet environment, as might be expected for close-to-field remote supervision. It should be noted that the tractor cab was not completely soundproof; an average noise level of 44 dBA was measured. During the second session, 'tractor noise' was introduced into the environment to create the in-field remote supervision scenario. The tractor noise, measured to be 78 dBA, used in this study was previously recorded from an operating John Deere combine. The sound clip was played inside the tractor cab using a computer speaker. The average sound levels for both background noises were measured using a sound meter (Q094168, REED Instruments, USA). The two remote supervision situations modelled for our study are described in Table 1.

There were two trials in each experimental session, each lasting an average of 6 min. To inform the participants of the error, one of the bimodal warning cues was employed in each trial (trial 1 – visual and auditory, trial 2 – visual and tactile). The experimental sessions and trial orders within each session were counterbalanced and randomised among participants to minimise order effects. Following each trial, participants were given a trial questionnaire on which they were asked to rate and provide feedback on the warning cues used. Participants were given a 10 min break between sessions to unwind and move around. After the final session, participants were given an end-of-experiment questionnaire asking about their overall experience and any further recommendations.

Data analysis. Participants' SA was examined in relation to their response times per trial and levels of noticeability. The time stamp specified in the sprayer simulation code computed the response time based on the difference between the error display and acknowledgement times. The level of noticeability was measured as saccadic reaction time. This was determined as the recorded time between error display and saccade (i.e., the rapid eye movement from one gaze point to another) onset such that short saccadic reaction time denotes a high level of noticeability and vice versa. The BeGazeTM Analysis Software (version 3.7.41) estimated the saccadic reaction time. The saccade onset was interpreted as the quick eye movement of the participant from the primary task screen above the participant's seated position to the top centre of the screen below the participant's seated position. The entire top centre of the screen below the participant's seated position was classified as the region of interest as it contains the visual warning indicator, which continuously displays when each of the bimodal warnings (i.e., visual-auditory or visual-tactile) occurs together with the accompanying warning message drawing the attention of the participant to that specific region (Figure 3). Figure 4 shows examples of gaze fixation points relevant to this study.

Data analysis was performed to determine the effectiveness and efficiency of the two bimodal warn-

Table 1. Scenarios for remote supervision and related work environments were used in the study (adapted from Edet & Mann 2021)

Scenario	Background noise	Primary task	Monitoring level
In-field	Tractor noise	Expected to operate another machine while supervising the autonomous agricultural machine	Intermittent
Close-to-field	Little or no tractor noise	Perform another off-field task, e.g., prepare chemical to refill the spray tank	Intermittent

https://doi.org/10.17221/73/2024-RAE

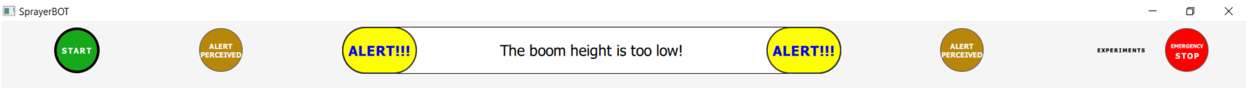

Figure 3. The region of interest associated with the perception of the warning cues

ing modalities tested in the two environments (i.e., noisy and quiet). Outlier responses were evaluated using the 2 × SD criteria for the average number of responses. Statistically significant differences were accepted at the 95% confidence level ($p < 0.05$) for the repeated measures analysis of variance (ANOVA) conducted. To find the most effective bimodal warning technique, the participants' saccadic reaction and response times to the warnings were compared with each other and with two different background noise levels (tractor sound and quiet). The participants' subjective evaluations and comments from the end-of-experiment questionnaire were also considered during the data analysis.

RESULTS AND DISCUSSION

Participant demographics. A total of 25 participants between the ages of 18 and 35 (29 ± 4.9) participated in the study (11 male, 14 female, 2 left-handed). Fourteen participants had farming experience, and ten of the fourteen self-reported having experience driving a tractor or operating an agricultural sprayer. Each participant completed an informed consent form and received an honorarium for their time. During the visual screening, all participants could see the visual cues provided to them. In addition, the findings of the pure-tone audiometry hearing tests showed that none of the participants had any hearing impairments that would influence the study's conclusions (i.e., they could discern between various background noises and the auditory warning).

(A)

(B)

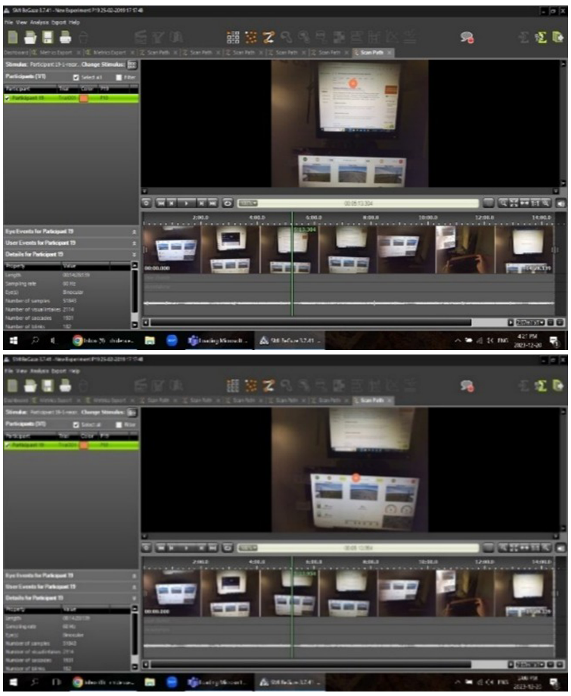

Figure 4. Display of gaze fixation points following saccade onset

(A) gaze cursor on the primary task screen before the error display; (B) gaze cursor on the region of interest at saccade onset

Response time. Figure 5 presents the mean response times for each background condition for the two types of trials. Both bimodal warning cues (i.e., visual-auditory and visual-tactile) were able to draw the attention of participants; however, the visual-tactile warning cue had a lower response time for both the quiet (2 464.0 ± 393.5) and tractor noise (2 452.0 ± 445.8) conditions making it the more effective warning cue. The response times were entered into a 2 (Trial type) × 2 (Background condition) ANOVA, which showed that there was a significant main effect of Trial type, $F_{(1,\ 24)} = 5.36$, $p < 0.05$, and no significant effect of background condition, $F_{(1,\ 24)} = 0.03$, $p > 0.05$ with an interaction between them, $F_{(1,\ 24)} = 6.03$, $p < 0.05$.

The participants' visual and auditory sensory overload may cause these findings (Lee 2015). Most of the participants' time during the tractor sound session was spent using their visual and auditory senses to complete an internet search task and listen to the sound of the tractor. As a result, it is possible that individuals took longer to understand the warning when they were informed of an issue utilising visual and auditory stimuli. When participants shifted their focus from the online search to the sprayer simulation screen, the visual load remained rather consistent, which is why this effect was stronger for the visual-auditory stimuli. Nonetheless, the participants were subjected to tractor noise during the session because there was a visual-auditory warning cue for when the tractor sound was introduced into the background. Therefore,

https://doi.org/10.17221/73/2024-RAE

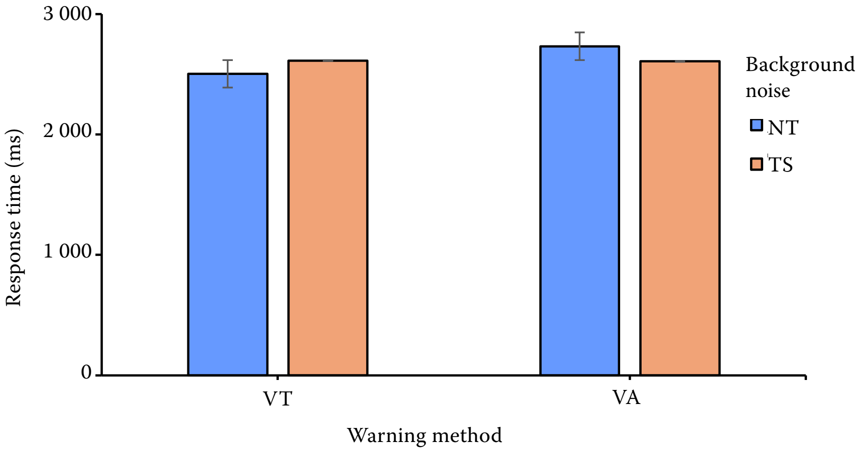

Figure 5. Comparison of participants' response times in relation to background noise

VA – visual-auditory; VT – visual-tactile; NT – no tractor sound; TS – tractor sound; error bars represent standard error

the auditory burden was further enhanced by the audio warning from the visual-auditory warning cue, which might have made the visual-auditory cue less effective. The results displayed in Figure 5 also align with findings from other studies, which indicated that visual-tactile stimuli were more effective than visual-auditory stimuli (Burke et al. 2006; Whang et al. 2007). Hence, visual-tactile warning cues work best for close-to-field remote supervision when background noise is absent.

Noticeability. Out of the 25 participants who participated in the study, nine were excluded from the eye-tracking data analysis because of various problems with the eye movement recording device. Figure 6 reveals the mean saccadic reaction times for each background condition for the two types of trials (trial 1 – visual-auditory; trial 2 – visual-tactile). Notably, visual-auditory warning when background noise was present had the shortest saccadic reaction time (379 ± 185.1), which denotes the highest level of noticeability. The reaction times were entered into a 2 (Trial type) × 2 (Background) ANOVA, which showed that there was no significant main effects of Trial type, $F_{(1, 15)} = 0.53$, $p > 0.05$, and background condition, $F_{(1, 15)} = 2.18$, $p > 0.05$ with no interaction between them, $F_{(1, 15)} = 0.01$, $p > 0.05$.

This outcome agrees with the findings by Corneil and Munoz (1996), which state that an irrelevant auditory cue influences gaze shifts to visual targets differently than an irrelevant visual cue influences gaze shifts to auditory targets in a complex environment. This is consistent with the notion that peripheral or surrounding auditory stimuli are powerful in capturing visual attention when notified by visual-auditory cues (Mazza et al. 2007). Conversely, visual-tactile warning with no tractor sound in the background had the lowest level of noticeability as it had the longest saccadic reaction time (469.0 ± 237.8).

Subjective responses. Participants were able to recognise the bimodal warning cues. According to an analysis of the end-of-experiment questionnaire, they felt at ease receiving error notifications

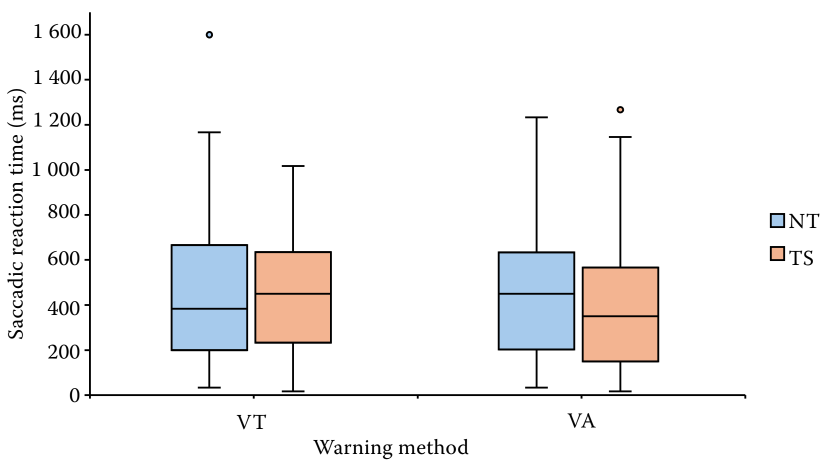

Figure 6. Comparison of participants' saccadic reaction times in relation to background noise; VA – visual-auditory

VA – visual-auditory; VT – visual-tactile; NT – no tractor sound; TS – tractor sound; error bars represent standard error

https://doi.org/10.17221/73/2024-RAE

Table 2. Summary of findings

Scenario	Warning method	Response time	Notice-ability	Subjective response
In-field (noisy)	Visual-auditory	NS	NS	
	Visual-tactile	NS	NS	Preferred
Close-to-field (quiet)	Visual-auditory	NS	NS	
	Visual-tactile	SS	NS	Preferred

NS – not statistically significant, SS – statistically significant

with either the visual-tactile or the visual-auditory bimodal warning cues. The participants also reported that both bimodal warning cues successfully got their attention. However, there were differences in the participants' efficiency levels (i.e., slightly or highly effective) and terms of background noise (quiet scenario and tractor sound).

In general, 24% of participants preferred visual-auditory cues, whereas 76% suggested using visual-tactile cues. No participant suggested using both bimodal warning cues simultaneously. When asked why they made their suggestions, the participants said they found it harder to distinguish between the visual-auditory cues and the background noise; therefore, the visual-tactile cues were less distracting when recognising the warning messages. Six participants also stated that the visual-auditory cues were less annoying than the visual-tactile cues and easily allowed them to think. They stated that the visual-auditory cues seemed more like an alert to them. On average, however, visual-auditory warning cues did not yield the fastest response times for both quiet (2 498.0 ± 617.1) and tractor noise (2 488.0± 597.8) conditions (Figure 5). The subjects' increased mental workload and decreased SA may have contributed to this heterogeneity. According to this, when tractor noise was introduced into the background, for the visual-tactile, the participants had to perceive and process only one auditory stimulus- the background noise- while for the visual-auditory, the participants had to perceive and process two auditory stimuli- the background noise and the auditory warning cue- during the visual-tactile warning cue. This resulted in a higher mental workload and a longer response time before participants could correct the error. Table 2 shows the summary of findings as discussed above.

CONCLUSION

This study aimed to identify the most appropriate bimodal warning cue to alert an AAM's human supervisor. Specifically, this study assessed the supervisor's response based on what is known as Level 1 SA or perception. The perception of two different bimodal warning signals, with and without background tractor noise, was assessed by measuring response time and level of noticeability in four conditions. Visual-tactile and visual-auditory warning cues had the shortest and longest response times, respectively. In terms of noticeability, no statistically significant differences were observed between visual-auditory and visual-tactile, although noticeability improved with background tractor noise. A more significant percentage of the participants preferred the visual-tactile warning method to the visual-auditory one. This suggests that the most appropriate warning cue is visual-tactile in remote supervision situations when humans might not be exposed to tractor noise. Both bimodal warning methods are suitable for remote supervision scenarios that expose humans to tractor noise, as there was no significant effect between the bimodal warnings when tractor sound was present. That said, the visual-tactile was the most consistent, preferred bimodal warning modality, as indicated by participants. This result will help designers select the best modality when designing warning systems for remotely supervised autonomous agricultural machines, minimising hazards experienced by farmers during spraying operations and enhancing their efficiency.

Acknowledgement: The authors want to acknowledge the financial support of Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

Adamides G., Christou G., Katsanos C., Xenos M., Hadzilacos T. (2014): Usability guidelines for the design of robot teleoperation: A taxonomy. IEEE Transactions on Human-Machine Systems, 45: 256–262.

Alexander R.D., Herbert N.J., Kelly T.P. (2009): The role of the human in an autonomous system. Proceedings of the 4th IET System Safety Conference, London, Oct 26–29, 2009: 1–6.

Auat Cheein F.A., Carelli R. (2013): Agricultural robotics: Unmanned robotic service units in agricultural tasks. IEEE Industrial Electronics Magazine, 7: 48–58.

https://doi.org/10.17221/73/2024-RAE

Bechar A., Vigneault C. (2016): Agricultural robots for field operations: Concepts and components. Biosystems Engineering, 149: 94–111.

Bilski B. (2013): Exposure to audible and infrasonic noise by modern agricultural tractor operators. Applied Ergonomics, 44: 210–214.

Blackmore B.S., Griepentrog H., Fountas S., Gemtos T. (2007): A specification for an autonomous crop production mechanisation system. Agricultural Engineering International: The CIGR E-Journal, 9: 1–17.

Braun D., Breitmeyer B.G. (1988): Relationship between directed visual attention and saccadic reaction times. Experimental Brain Research, 73: 546–552.

Burke J.L., Prewett M.S., Gray A.A., Yang L., Stilson F.R., Coovert M.D., Elliot L.R., Redden E. (2006): Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. Proceedings of the 8th International Conference on Multi-modal Interfaces, Banff, Nov 2–4, 2006: 108–117.

Caspi A., Roy A., Wuyyuru V., Rosendall P.E., Harper J.W., Katyal K.D., Greenberg R.J. (2018): Eye movement control in the Argus II retinal-prosthesis enables reduced head movement and better localisation precision. Investigative Ophthalmology & Visual Science, 59: 792–802.

Chen J., Šabić E., Mishler S., Parker C., Yamaguchi M. (2022): Effectiveness of lateral auditory collision warnings: Should warnings be toward danger or toward safety? Human Factors, 64: 418–435.

Corneil B.D., Munoz D.P. (1996): The influence of auditory and visual distractors on human orienting gaze shifts. Journal of Neuroscience, 16: 8193–8207.

Delavarpour N., Eshkabilov S., Bon T., Nowatzki J., Bajwa S. (2019): The performance analysis of tactile and ultrasonic sensors for planting, fertilising, and cultivating cover crops. 2019 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers, Boston, Jul 7–10, 2019: 1.

Dorais G.A., Gawdiak Y. (2003): The personal satellite assistant: An internal spacecraft autonomous mobile monitor. 2003 IEEE Aerospace Conference, Big Sky, Mar 8–15, 2003: 1–348.

Edet U., Mann D.D. (2021): Evaluating warning modalities for remotely supervised automated agricultural machines. Journal of Agricultural Safety and Health, 28: 1–17.

Edet U., Ogidi F., Mann D. (2022): Design and evaluation of a user interface for an autonomous agricultural sprayer. Agricultural Sciences, 13: 221–243.

Elbert K.K., Kroemer H.B., Hoffman A.D.K. (2018): Selection, design, and arrangement of controls and displays. In: Kroemer K.H.E. (ed.): Ergonomics: How to Design for Ease and Efficiency. Englewood Cliffs, NJ: Prentice Hall: 500–557.

Endsley M.R. (1995): Measurement of situation awareness in dynamic systems. Human Factors, 37: 65–84.

Endsley M.R., Bolté B., Jones D.G. (2003): Designing for Situation Awareness: An Approach to User-Centered Design. Taylor & Francis, London.

Endsley M.R., Garland D.J. (2000): Theoretical underpinnings of situation awareness: A critical review. In: Endsley M.R., Garland D.J. (eds): Situation Awareness Analysis and Measurement, 1: 3–21.

Haas E.C., Van Erp J.B. (2014): Multi-modal warnings to enhance risk communication and safety. Safety Science, 61: 29–35.

Han S., Steward B.L., Tang L. (2015): Intelligent agricultural machinery and field robots. In: Precision Agriculture Technology for Crop Farming, CRC Press, Boca Raton: 133–176.

Hancock P.A., Mercado J.E., Merlo J., Van Erp J.B. (2013): Improving target detection in visual search through the augmenting multi-sensory cues. Ergonomics, 56: 729–738.

Hoppe S., Loetscher T., Morey S.A., Bulling A. (2018): Eye movements during everyday behavior predict personality traits. Frontiers in Human Neuroscience, 12: 328.

Ima C.S., Mann D.D. (2003): Lightbar design: The effect of light color, lightbar size, and auxiliary indicators on tracking and monitoring performance. Agricultural Engineering International 3: 5.

Karimi D., Mann D.D. (2008): Role of visual cues from the environment in driving an agricultural vehicle. The Ergonomics Open Journal, 1: 54–61.

Laughery K.R., Wogalter M.S. (2006): Designing effective warnings. Reviews of Human Factors and Ergonomics, 2: 241–271.

Lee C. (2015): Banner blindness: An effect of information overload on the World Wide Web [unpublished manuscript]. Available at: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.501.8980&rep=rep1&type=pdf (accessed Jun 28, 2023)

Lee W.S., Kim J.H., Cho J.H. (1998): A driving simulator as a virtual reality tool. In: Proceedings of the 1998 IEEE International Conference on Robotics and Automation, Leuven, May 16–20, 1998, 1: 71–76.

Macadam C.C. (2003): Understanding and modeling the human driver. Vehicle System Dynamics, 40: 101–134.

Mazza V., Turatto M., Rossi M., Umiltà C. (2007): How automatic are audiovisual links in exogenous spatial attention? Neuropsychologia, 45: 514–522.

Mouloua M., Ferraro J.C., Kaplan A.D., Mangos P., Hancock P.A. (2019): Human factors issues regarding automation trust in UAS operation, selection, and training. In: Human Performance in Automated and Autonomous Systems, CRC Press, Boca Raton: 169–190.

Murata A., Kanbayashi M., Hayami T. (2013): Effectiveness of automotive warning system presented with multiple sensory modalities. In: International Conference on Digital Hu-

https://doi.org/10.17221/73/2024-RAE

man Modeling and Applications in Health, Safety, Ergonomics and Risk Management, Springer, Berlin, Heidelberg: 88–97.

Niehorster D.C., Santini T., Hessels R.S., Hooge I.T., Kasneci E., Nyström M. (2020): The impact of slippage on the data quality of head-worn eye trackers. Behavior Research Methods, 52: 1140–1160.

Parasuraman R. (2000): Designing automation for human use: Empirical studies and quantitative models. Ergonomics, 437: 931–951.

Peryer G., Noyes J., Pleydell-Pearce K., Lieven N. (2005): Auditory alert characteristics: A survey of pilot views. The International Journal of Aviation Psychology, 15: 233–250.

Petocz A., Keller P.E., Stevens C.J. (2008): Auditory warnings, signal-referent relations, and natural indicators: Rethinking theory and application. Journal of Experimental Psychology: Applied, 14: 165.

Politis I., Brewster S.A., Pollick F. (2014): Evaluating multimodal driver displays under varying situational urgency. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Apr 26 – May 1, 2014: 4067–4076.

Sabic E., Mishler S., Chen J., Hu B. (2017): Recognition of car warnings: An analysis of various alert types. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, May 6–11, 2017: 2010–2016.

Schreckenghost D., Fong T., Milam T. (2008): Human supervision of robotic site surveys. AIP Conference Proceedings, Jan 21, 2008. American Institute of Physics. 969: 776–783.

Shi J., Bai Y., Diao Z., Zhou J., Yao X., Zhang B. (2023): Row detection-based navigation and guidance for agricultural robots and autonomous vehicles in row-crop fields: Methods and applications. Agronomy, 13: 1780.

Smith C.A.P., Clegg B.A., Heggestad E.D., Hopp-Levine P.J. (2009): Interruption management: A comparison of auditory and tactile cues for both alerting and orienting. International Journal of Human-Computer Studies, 67: 777–786.

Whang M.C., Hyun H.J., Lim J.S., Park K.R., Cho Y.J., Park J.S. (2007): Effect of different modal feedback on attention recovery. In: International Conference on Usability and Internationalization, Beijing, Jul 22–27, 2017: 631–636.

Whelan R. (2008): Effective analysis of reaction time data. Psychological Record, 58: 475–482.

White T.L., Hancock P.A. (2020): Specifying advantages of multi-modal cueing: Quantifying improvements with augmented tactile information. Applied Ergonomics, 88: 103–146.

Wogalter M.S., Conzola V.C., Vigilante Jr W.J. (1999): Applying usability engineering principles to the design and testing of warning messages. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications, Los Angeles, 43: 921–925.

Wogalter M.S., Conzola V.C., Smith-Jackson T.L. (2002): Research-based guidelines for warning design and evaluation. Applied Ergonomics, 33: 219–230.

Yim M., Salemi B., Rus D. (2007): Modular self-reconfigurable robot systems [Grand challenges of robotics]. IEEE Robotics & Automation Magazine, 14: 43–52.

Young M.S., Stanton N.A. (2002): Attention and automation: New perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science, 3: 178–194.

Received: September 2, 2024
Accepted: February 19, 2025
Published online: April 24, 2025