Int. J. Human–Computer Studies 205 (2025) 103643

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Brief non-spatial signals facilitate visual search and temporal sensitivity in robot supervision

Bora Celebi [a,*,1], Julian Kaduk [b,1], Müge Cavdan [a], Heiko Hamann [b], Knut Drewing [a]

[a] Justus-Liebig University, Giessen, 35390, Germany
[b] University of Konstanz, Konstanz, 78464, Germany

ARTICLE INFO

Keywords:
Human–robot interaction
Human–swarm interaction
Visual search
Attention
Pip-and-pop

ABSTRACT

The human role in human–swarm interaction (HSI) shifts from controller to supervisor, as robots become more autonomous and require efficient search strategies in complex visual environments. Previous research has shown that spatially uninformative brief cues enhance search performance in laboratory environments (namely, "pip-and-pop" effect). Here we examined if these effects can be effectively applicable in HSI. To this end, we conducted two experiments using small mobile robots (Thymio II) to investigate the impact of auditory, tactile, and audiotactile cues on visual search performance and timing judgments. In the first experiment, 20 participants identified a stopped robot among moving robots. The results showed that all cue conditions significantly reduced reaction times (RTs) compared to the no-cue condition, suggesting that brief spatially non-informative signals improve search performance by increasing sensory information accumulation speed. The second experiment involved 12 participants judging the duration of a robot's stop after a tactile cue was presented or not. The findings indicate that tactile cues improve temporal sensitivity without affecting subjective duration judgments. These results highlight the potential of uni- and multisensory cues to enhance HSI performance by facilitating quicker and more accurate human responses, particularly in dynamic environments. The study extends the "pip-and-pop" effect to real-world scenarios, offering insights for designing HSI systems that allow users to interact with robotic swarms more naturally and efficiently.

1. Introduction

With advances in the ability and autonomy of robots, the human role in human–robot interaction (HRI) is changing. Especially in regards to human–swarm interaction (HSI) with large swarms of autonomous robots, the human is more often supervising rather than controlling. The supervision of robots in a swarm requires an overview of a complex and dynamically changing visual environment. Finding an important target in this setting is an exhaustive and time-consuming task which usually involves processing each item and disentangling the correct one from the irrelevant ones (Quinlan, 2003; Barros and Lindeman, 2014). In time-critical situations, it may be important for a human supervisor to quickly identify a specific robot. Therefore, reducing the search time of a human operator in applications that require urgency and precision is a fundamental objective, which can significantly improve performance (Pawlak and Vicente, 1996). Furthermore, in these situations, precisely estimating core aspects such as time becomes crucial, as it directly influences an operator's ability to efficiently estimate the exact time points of important events occurring and their subsequent durations within the environment. In this study, we investigate whether brief and spatially uninformative signals can improve search times and consequently influence subjective time perception in an HSI task with real robots.

In our daily life, we are continually bombarded with sensory cues from the external world and strive to discern the most relevant and important information from these signals. Processing all these inputs is hindered by our limited capacity and we cannot reliably identify and discriminate all available cues at the same time (Wolfe, 2020). To find a target object in our field of vision, we employ visual search to select and further process the desired object among a large collection of redundant ones (Wolfe, 2020). Subsequently, attention is directed to the object, with resources predominantly allocated to processing its features (Chan and Hayward, 2013). Classical visual search studies

* Corresponding author.
E-mail address: boracelebii@gmail.com (B. Celebi).
[1] These authors share joint first authorship and contributed equally to the manuscript.

https://doi.org/10.1016/j.ijhcs.2025.103643
Received 16 December 2024; Received in revised form 1 July 2025; Accepted 13 September 2025
Available online 22 September 2025

suggest that when a target stimulus has distinct features that distinguish it from the redundant distractors, the task of finding it is significantly facilitated (Egeth, 1977). For instance, a red ball amongst green ones can be easily detected and segregated from the rest of the set. The fast detection of a salient target with distinct features is called the pop-out effect.

Search performance also greatly benefits from additional informative cues about the features or location of the target of interest. Detection of a visual target that is spatially and temporally accompanied by a brief auditory cue was facilitated (Bolia et al., 1999; Doyle and Snowden, 1998). Bolia et al. (1999) found that, in a visual search paradigm, auditory cues that are co-located with the visual target improved search performance by decreasing reaction times (RTs). This effect was observed in other tasks, where the covert orientation of attention to a stimulus in one modality also increased the processing of a stimulus in another modality (Shi et al., 2010; Spence and Driver, 1997; Ho et al., 2005). For instance, Gray et al. (2009) have shown that temporally aligned tactile and auditory cues (in close proximity to visual) decreased RTs in a visual localization task. Therefore, in different contexts, cross-modal influences can assist task performance if both the stimulus and the cue are proximally aligned in space by orienting attention. These results highlight how attentional mechanisms, influenced by both spatially and temporally aligned cues, can affect task performance across different sensory modalities.

Since there is no dedicated sensory organ for perceiving time, temporal processing always coincides with sensory events. Therefore, temporal judgments, that is, perceived duration, temporal order, or simultaneity, can be disturbed and modulated by perceptual properties of a stimulus and cognitive functions that are associated with sensory processing (Matthews and Meck, 2016). In this sense, attention is one of the modulating factors of temporal judgments. Covertly or endogenously orienting attention with spatially aligned cues to the location of the stimulus was found to expand the duration judgments (Seifried and Ulrich, 2011; Yeshurun and Marom, 2008; Mattes and Ulrich, 1998). Furthermore, Chica and Christie (2009) found that the temporal resolution increases when the RTs are not constrained. When participants were allowed to respond whenever they wanted, attentional cueing had a facilitating effect on discrimination performance. In addition to spatial cues, attention can be directed to a specific stimulus by temporal cues (Coull and Nobre, 1998). These temporally informative signals orient the attention to a specific window of time and thus enhance the sensory processing of stimuli appearing in that window. Furthermore, it has been found that temporal sensitivity is improved by orienting attention to the relevant time indicated by the temporal cues (Correa et al., 2006). Thus, attentional orientation through spatial and temporal cues influences temporal judgments.

In real-life applications, it may not always be possible to guarantee spatial alignment of the cue with its target. However, improvements in search performance were also observed with spatially uninformative auditory and tactile cues (Van der Burg et al., 2008, 2009, 2010; Zou et al., 2012; Ngo and Spence, 2010a). Van der Burg et al. (2008) found that uninformative auditory tones facilitate visual search by decreasing RTs. In the employed visual search task, participants were instructed to find the horizontal or vertical target among cluttered distractors. The target and distractors were green or red and changed colors at random intervals. Synchronous to the color change, in half of the trials, they presented a spatially uninformative auditory tone. RTs were significantly reduced in the presence of auditory signals. This facilitation, termed the "pip-and-pop" effect, occurs when a brief auditory "pip" triggers a rapid "pop out" of the visual target. They argued that the effect was the product of an audiovisual integration process where the saliency of the visual stimulus increased and subsequently attracted attention (Van der Burg et al., 2008, 2010). Alternatively, Zou et al. (2012) proposed that people employ a "wait-at-beep" strategy where oculomotor scanning behavior is frozen just after the onset of the auditory signal (Vroomen and de Gelder, 2000). They argue that this freezing behavior allows for increased sampling of information from the display by facilitating the search. Furthermore, Ngo and Spence (2012) proposed that synchronizing the visual target with cues can transform the display into an "oddball" scenario, naturally drawing attention. In conclusion, spatially uninformative auditory or tactile cues can orient the attention to the target in cluttered environments and create a "pop-out" effect. Moreover, since these cues orient attention to the target, one would expect "pip-and-pop" effect to alter timing judgments by increasing the temporal sensitivity. In this study, we investigate the effectiveness of spatially uninformative cues in human and robot collaboration in terms of visual search performance and timing.

Incorporating insights from visual search research, which highlights the impact of distinct features and cross-modal cues, provides a foundation for developing multimodal interfaces finely tuned for human operators in HSI and HRI. An example of such a multimodal interface application is the combination of visual, auditory, tactile, and olfactory cues in search and rescue robotics (Barros and Lindeman, 2014). In their setup, they use visual feedback through an interface to provide information on the robot's speed, collisions, the robot's proximity to collisions, and the carbon monoxide levels sensed by the robot. Their results show an improvement in the data perception and performance of the robot operator. In a different example, Reardon et al. (2018) present an Augmented Reality (AR) display to provide the user with additional navigational information in a cooperative search task with a robot. Their results show an increase in situational awareness by assisting a human to navigate to a target with augmented visual cues. However, it is important that the additional cues do not interfere with the main task or overload the human with information. In a recent study, Arend et al. (2023) have shown that haptic search cues with a long delay of 500 ms can affect the cornering performance of the operator in a main driving task, while short delays of 50 ms did not interfere. Other examples cover in-vehicle navigation and warning signals through tactile cues (Erp and Veen, 2004) or the combination of auditory and visual cues emitted by a non-humanoid robot to ask for help (Cha et al., 2016).

Even though the "pip-and-pop" effect has been studied in classic laboratory environments with simpler visual displays, it is unclear whether this effect will persist in a dynamically changing naturalistic environment. Facilitatory effect of non-special cues could have significant implications for improving robot supervision in HSI. In scenarios where multiple robots operate in congested or dynamic environments, the use of non-spatial auditory or tactile cues can serve as a tool for directing human attention to critical information or events. Prior work has not examined whether spatially uninformative cues can improve search or affect temporal judgments in real-world HSI settings. This gap is important to be filled because human operators in time-sensitive HSI tasks must quickly detect important events without the aid of precise spatial indicators. Distinctive auditory or tactile cues could for example trigger a "pop out" effect, capturing the human supervisor's attention and facilitating quick decision-making in response to a critical event. This concept not only improves communication in complex tasks but also opens avenues for designing effective and novel multimodal interfaces, contributing to the seamless collaboration between humans and robots in various real-world applications. However, to our knowledge, this is not yet explored in the literature, nor is it known how the brief and spatially uninformative signals that constitute the "pip-and-pop" effect could affect temporal judgments. In the two experiments of this study, we investigated (1) how auditory, tactile, and audiotactile brief non-spatially aligned cues affect search performance and (2) whether these cues affect temporal estimations and sensitivity in a naturalistic environment with moving robots.

Further insights regarding the visual search process in HSI can be obtained using cognitive models such as Drift Diffusion Modeling (DDM) (Ratcliff and McKoon, 2008). DDM posits that decisions are made by the continuous accumulation of sensory evidence until a

certain decision threshold is reached DDM quantifies key parameters such as the drift rate, decision boundary, and non-decision time which are further explained in the methods section. By analyzing these parameters, the model can elucidate how brief, spatially uninformative cues can enhance the saliency of visual targets and expedite decision-making processes (Ratcliff, 1978; Wiecki et al., 2013). In the context of HSI, DDM can be particularly useful in revealing how these cues might enhance the accumulation of sensory evidence, thereby improving the speed and accuracy of identifying critical targets among distractors.

Building on these findings regarding the influence of spatial and temporal cues on visual search and temporal judgments, this study aims to address the gap concerning the effectiveness of brief, spatially uninformative auditory and tactile signals in dynamic human–swarm interaction environments. Specifically, we ask: Can these non-spatial multisensory cues improve search efficiency and alter subjective time perception during real-world supervision of robot swarms? By answering this, we seek to advance understanding of attentional mechanisms in HSI and inform the design of multimodal interfaces that support rapid and accurate human decision-making in complex, time-sensitive tasks.

2. Experiment outline

2.1. Transparency and openness

The complete dataset, along with the analysis code and research materials, is publicly accessible on the Open Science Framework (OSF) at (https://osf.io/p7tz2/?view_only=5f77d11769274fabae437aebccbf20c9). We employed Thymio II robot platform (Riedo et al., 2013) and Raspberry Pi 3B for wireless connectivity. The analyses were performed using Python (version 3.10.2.) with the PyMC (version 5.9.2.) and psignifit (version 4.0.) packages.

Ethical approval for this research was granted by the ethics committee of (removed for masking), following the guidelines of the Declaration of Helsinki (2013) without preregistration.

2.2. Robot hardware

For both studies, we used the Thymio II (Riedo et al., 2013) robot platform. It is a small differential drive robot with a footprint of 11 cm × 11 cm and a maximum linear velocity of 0.2 m/s. The perception system contains nine infrared (IR) distance sensors with an approximate range of 15 cm. Five sensors are horizontally positioned at the front of the robot, and two are horizontally placed toward the rear. The last two IR sensors are located in the front section underneath the robot, aimed at the ground. For added wireless connectivity and better controls with a Python programming interface, we enhanced it with a Raspberry Pi 3B and an additional battery bank. An image of the fully assembled robot can be seen in Fig. 1.

2.3. Robot behavior

The basic robot behavior can be described as ballistic motion in which the robot drives straight with a constant velocity of $v = 0.1$ m/s until it detects the arena boundary or another robot. For other robots, which are detected by the front-facing horizontal IR sensors, the robot rotates on the spot in a random direction for a random duration between 0.5 and 1.5 s, both sampled from a uniform distribution. When the robot detects the arena boundary, marked by black tape on the ground, using its downward-facing IR sensors, it performs an escape maneuver. It reverses for a random duration between 0.5 and 1.5 s, drawn from a uniform distribution, or until its rear-facing IR sensors detect another robot behind it. Subsequently, it rotates on the spot away from the arena boundary for a random duration between 0.5 and 1.5 s, sampled from a uniform distribution.

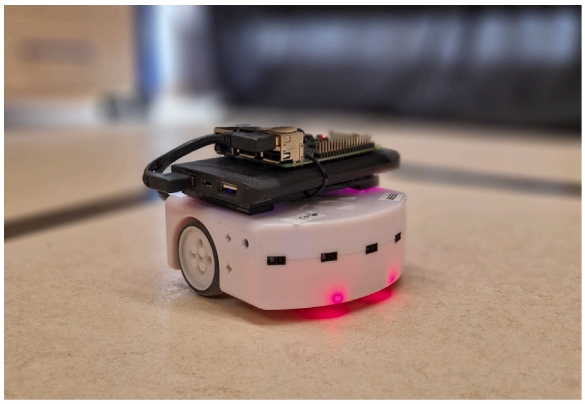

Fig. 1. The robot has a footprint of 11 cm × 11 cm. Four of its 5 front facing infrared (IR) distance sensors can be seen on the image.

3. Experiment 1

In Experiment 1, we investigated the effects of auditory, tactile, and audiotactile cues on visual search performance in a dynamically changing environment populated by moving Thymio II robots. Twenty participants were tasked with identifying and responding to a robot that stopped moving within a 2 m × 2 m arena. The experiment consisted of eight trial conditions combining different cue types and robot behaviors. Next, we present the experiment and results in further detail.

3.1. Method

3.1.1. Participants

A priori power analysis was conducted using G*Power (Faul et al., 2007) based on initial piloting with a medium effect size (Power 80%, medium effect size of partial $\eta^2 = .07$ for repeated measures Analysis of Variance [ANOVA] with four levels, $\alpha = .05$). This analysis indicated a required sample size of 20. Twenty participants (13 females, age range: 19–37, $M = 25.2$, $SD = 4.9$) were recruited through a circular email at Justus-Liebig University, Giessen. None of the participants reported any somatosensory impairments and all reported normal or corrected-to-normal vision. Prior to the experiment, they provided written informed consent and were compensated 8€/h for their participation.

3.1.2. Stimuli and apparatus

In Experiment 1, the robots were placed inside of a 2 m × 2 m robot arena marked with black tape on the ground. The participants were standing adjacent to this arena at a distance of 1.2 m. During the experiment, the participants were holding a 3D printed cube measuring 9 cm × 9 cm with a 24 mm diameter button on top. The button was used to collect the participants' responses. During the experiment, a custom-made multimodal haptic vest (removed for masking) was used to deliver tactile cues. Headphones were used to present white noise to suppress the sound of the robots. Additionally, the participants wore a Polar H10 electrocardiography (ECG) chest strap (Polar, 2024) and a Pupil Labs Core eye tracker (Kassner et al., 2014) which are not subject to the current study. All data was recorded with the Capture software by Pupil Labs via the lab streaming layer (LSL) system (Kothe, 2024) for timestamp time synchronization.

In addition to the basic robot behavior detailed in the experiment outline, we introduced "stop" and "no stop" conditions. In the "stop" condition, a randomly selected robot would stop for 5 s before resuming its movement. The "no stop" condition did not involve a change in robot motion. Additionally, four types of cues were introduced: no cue, a tactile cue delivered via the haptic vest (7.38 m/s^2 root mean

square) on the right and left side of the upper chest area, an auditory cue (300 Hz, sine wave) from the headphones, and an audiotactile cue combining both the auditory and tactile cues. The cues were designed according to the previous literature which demonstrated and explored the 'pip-and-pop effect" (Van der Burg et al., 2008, 2009, 2010; Zou et al., 2012). Furthermore, tactile cues were selected specifically to be above detection threshold and we used similar amplitudes that has been used previous studies which showed attentional effects (Celebi et al., 2025). Each cue lasted for 100 ms and their onsets were synchronized with the "stop" and "no stop" conditions.

3.1.3. Design and procedure

A within-participant design with the effect of cue (four levels: no cue, auditory cue, tactile cue, and audiotactile cue) and stopping (two levels: stop and no stop) was used. At the beginning of the experiment, we thoroughly explained the procedure and the task to the participants. We asked the participants to observe all robots in the arena and identify if any robot had stopped, as in the stopping condition. Their task was to press the button on the handheld interface until the robot started moving again. We also informed them of the different cues which could randomly be triggered during the experiment and highlighted that they were random and not an indication that a robot had stopped. This task was designed to assess the participant's ability to perceive and respond to the robot's change in status under different cues.

Each experimental trial began with a 30-second start offset in which robots move in their basic behavior, followed by the random initiation of trials at intervals ranging between eight and twelve seconds, and concluded with a ten-second end offset. Every participant first absolved a training session, consisting of two repetitions of each trial condition in a random sequence, followed by the main experiment. The main experiment included ten trials for each condition, also in a random order. Overall, the experiment lasted approximately 20 minutes per participant.

During the experiment, we continuously recorded a time-series of the experiment state, including stopping trigger, cue types, and button input. All data were captured with synchronized timestamps to extract the response times for each trial condition.

Hierarchical drift–diffusion modeling. Drift–diffusion modeling (DDM) serves as a well-established method to reveal the underlying stages of processing involved in a specific perceptual decision-making task. These models operate on the assumption that sensory information is progressively gathered over time, and once a predefined boundary is reached in this accumulation process, a decision is made (Ratcliff and McKoon, 2008). In the employed task, individuals gather sensory information from the environment to find the target robot and decide if sufficient information was gathered regarding the stopping. Thus, to further investigate what sensory processes are involved in this effect, we applied hierarchical DDM to the RTs data (Ratcliff, 1978; Wiecki et al., 2013). Here, as we only have RTs for the correct responses and considering the high accuracy of the participants, we fitted a single boundary DDM for the correct responses. To this end, we used the approximation derived by Lee et al. (2006). This approximation involves three parameters: drift rate v, boundary a, and non-decision time t. The drift rate v represents the speed of information accumulation during decision-making. Higher drift rates indicate faster and more confident decisions based on stronger sensory evidence, whereas lower drift rates suggest slower decision-making due to weaker evidence. The boundary parameter a signifies a threshold that must be reached for a decision to be made. A higher boundary reflects a more cautious decision-making criterion. And finally, the non-decision time t parameter accounts for time spent outside the decision-making process, encompassing sensory encoding, response preparation, and execution (Ratcliff and McKoon, 2008; Allenmark et al., 2018). As the model was hierarchical, we estimated these parameters both at group-level and individual levels (Wiecki et al., 2013). To gain insight into which latent processes can explain the empirical data, we employed six hypothetical models that vary in which parameters were allowed to change between cue conditions and then compared these models using the *arViz* package in Python (Kumar et al., 2019). Model comparison results are shown in the Results section. In the model fitting procedure, we sampled all of the group-level parameters from a Truncated Normal distribution bounded by $[0, \infty]$ as we require positive values. We chose weakly informative hyperparameters for these distributions. We generated 40,000 samples for each model using *PyMC* package on Python (Oriol et al., 2023) which employs a Markov Chain Monte Carlo algorithm (Andrieu et al., 2003) to robustly estimate the posterior distribution of model parameters. To ensure convergence, we visually inspected the traces, and their auto-correlation, and checked the R-hat statistics for the reliability and accuracy of the sampling (Wiecki et al., 2013). Finally, we performed Bayesian hypothesis testing on posterior estimates of group-level variables by calculating the probability of difference between parameters being above zero (Kruschke, 2010; Gelman et al., 1995).

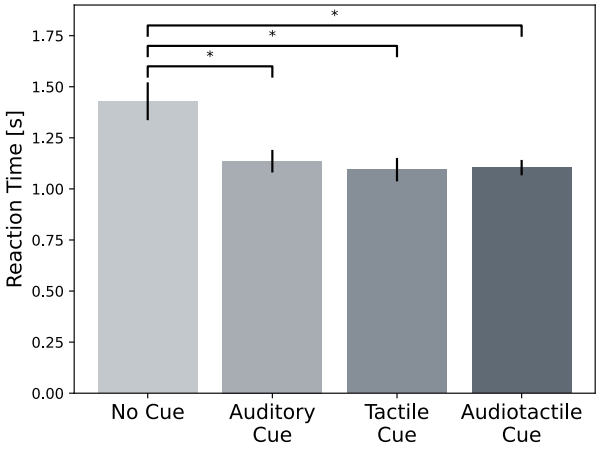

Fig. 2. RTs for each Cue. Error bars depict the standard error of the mean. Asterisks depict significant differences.

3.2. Results

Data from a participant were not included in the statistical analyses because of not properly adhering to task instructions. The accuracy was calculated as the percentage of correct button presses in trials where the robot stopped. The RTs were calculated as the time difference between the robot stopping and the button press. The average accuracy of the responses was high across all conditions (no cue: 95.6%, auditory cue: 98.1%, tactile cue: 96.0%, audiotactile cue: 95.3%). Individual RTs were submitted to a repeated-measures ANOVA to test the effect of cue (no cue, auditory cue, tactile cue, and audiotactile cue) on search performance. Mauchly's test did not indicate a significant violation of the sphericity assumption, χ^2 (5) = 7.37, p = .19. There was a significant main effect of the cue on RTs $F(3,54) = 10.82$, p <.001, partial η^2 = .37 (Fig. 2). Bonferroni corrected pairwise comparisons showed that the no cue condition resulted in significantly higher RTs (M = 1.42, SD = 0.4) compared to the auditory (M = 1.13, SD = 0.24), p <.01, d = .88, tactile (M = 1.09, SD = 0.25), p <.01, d = .99, and audiotactile cue conditions (M = 1.10, SD = 0.16), p <.01, d = 1.05. There were no significant differences between the RTs for auditory and tactile (p > 0.43, d = .17), auditory and audiotactile (p > 0.58, d = .15) as well as tactile and audiotactile (p > 0.87, d = .05) conditions. Overall, this shows that regardless of the cue modality, a cue enhances the performance.

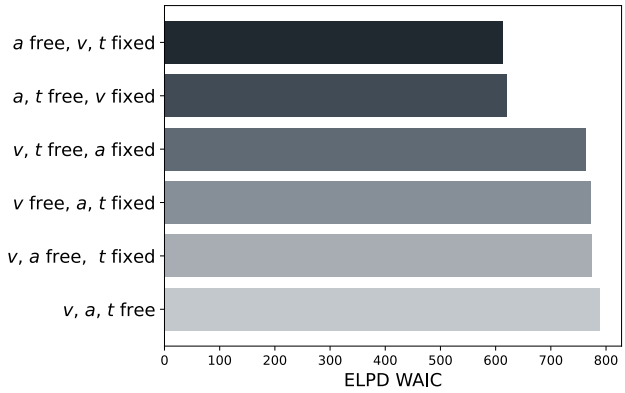

Fig. 3. ELPD WAIC Scores for Each Proposed Model.Y axis labels depict which parameters were fixed or free in different models. v is drift rate, a is boundary, and t is non-decision time.

3.2.1. Hierarchical DDM results

We used expected log-pointwise predictive densities (ELPD) derived by Widely Applicable Information Criterion (WAIC) to compare models. WAIC is a Bayesian criterion for model comparison that estimates a model's expected predictive accuracy. Specifically, it computes the sum of the log-likelihood of the data, averaged over the posterior distribution, while penalizing model complexity by estimating the effective number of parameters (Vehtari et al., 2017). In hierarchical models, parameters are often partially pooled across participants, and the actual complexity of the model is determined not only by the number of parameters, but by how much those parameters are informed by the data. WAIC accounts for this by penalizing models based on the variance of the log-likelihood across posterior samples. Model comparisons revealed that the best model was the one that varied all parameters for each condition (Fig. 3). We conducted Bayesian hypothesis testing in order to examine the differences in estimated model parameters between conditions. In this model, the drift rate for the no cue condition was lower than for auditory $P_{P|D} = 1.0$, tactile $P_{P|D} = 1.0$, and audiotactile cues $P_{P|D} = .99$ revealing that each cue increased the speed of accumulation of sensory information. There were no credible differences between the conditions a cue was present (all $P_{P|D} < .79$). Furthermore, there were no credible differences between the conditions on the boundary parameters (all $P_{P|D} < .84$), and on non-decision time parameters (all $P_{P|D} < .52$) (Table A.1). Thus, even though the model that varied all parameters was the best, brief non-spatial cues resulted in a change only in sensory evidence gathering speed, namely the drift rate (Fig. 4).

4. Experiment 2

In Experiment 1 we studied how different cues affect visual search performance regarding RTs in a dynamically changing environment populated by moving robots. As attentional cueing can affect timing judgments, in experiment 2 we examine how tactile cues influence temporal estimations and sensitivity in a visual search task similar to Experiment 1 coupled with a temporal bisection task. Twelve participants observed a robot stopping within a 2.2 m × 1.6 m arena and judged the duration of its stop as being closer to short or long anchor durations. The following sections present the experiment and results in further detail.

4.1. Methods

4.1.1. Participants

An a priori power analysis was conducted using G*Power (Faul et al., 2007) based on the findings from the first experiment with a large effect size (Power 80%, large effect size of Cohen's $d = .8$ for a one-sided t-test, $\alpha = .05$). This analysis revealed a required sample size of 12. Twelve participants (8 females, age range: 20–40, $M = 25.5$, $SD = 5.2$) were recruited through a circular email at Lübeck University. None of the participants reported any somatosensory impairments and all reported normal or corrected-to-normal vision. Prior to the experiment, they provided written informed consent.

4.1.2. Stimuli and apparatus

For this experiment, we used the same setup as in the first experiment besides a minimally different arena measuring 2.2 m × 1.6 m because it was conducted in a different lab. Additionally, the button interface was extended with two additional small buttons on either side of the main 24 mm diameter button on top.

The general behavior of the robot is described in the outline of the experiment. In each trial, a randomly selected robot would stop moving and would start again with a delay period once the participant signaled detecting it with a button press. This was accompanied by one of two cue conditions: no cue or a tactile cue delivered via the haptic vest. Further details of the robot behavior are explained in the experiment procedure.

4.1.3. Design and procedure

A within-participant design with the effect of cue (two levels: no cue, tactile cue) and delay period (seven levels: 200, 300, 400, 500, 600, 700, 800 ms) was used. The range from 200 ms to 800 ms was chosen, based on previous research on temporal bisection in the sub-second range by Wearden and Ferrara (1995). Before the experiment, we thoroughly explained the procedure, the interface, and the task to the participants. The participants were tasked with localizing the stopped robot and then pressing the large button on their handheld

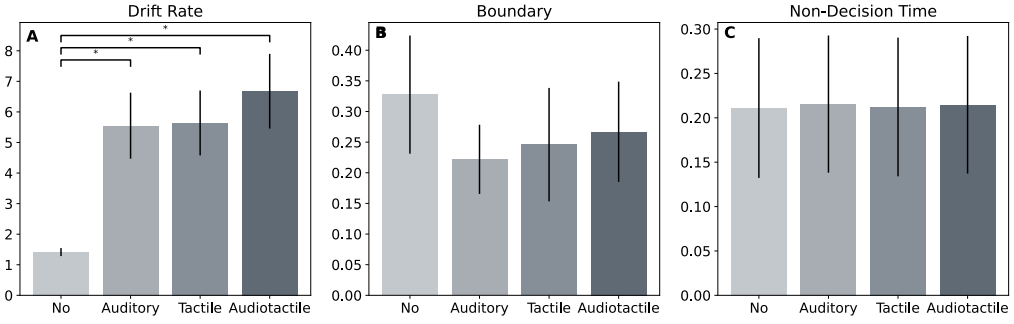

Fig. 4. Estimated parameter values of the group-level parameters extracted from the posterior distribution for the best fitting model across cues. (A) Drift rate (B) Boundary (C) Non-Decision time. Error bars depict the standard deviation of the posterior distribution. Asterisks depict significant differences.

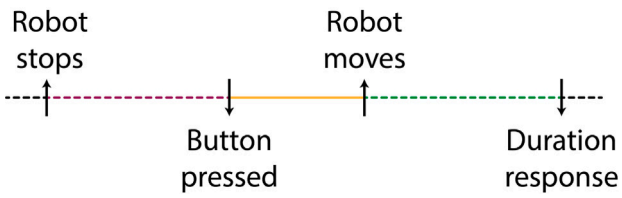

Fig. 5. Temporal bisection trial timeline. The red dashed line is the detection time of unknown duration as it depends on the participant. The yellow solid line the trial duration/delay period with a delay period of $t \in [200, 300, 400, 500, 600, 700, 800]$ ms. The green dashed line the response window in which the participant indicates whether the trial duration was perceived as "long" or "short".

interface to indicate that they had detected the robot. Following the button press, the robot would resume movement after a delay period of $t \in [200, 300, 400, 500, 600, 700, 800]$ ms, with anchor durations defined as $t_{\text{short}} = 200$ ms and $t_{\text{long}} = 800$ ms. The participants were then required to indicate whether they perceived the delay as closer to the short or long anchor duration by pressing the corresponding small button on their interface. For the better understanding of the procedure, see Fig. 5.

Overall, the experiment was structured into four phases, beginning with training. During this initial phase, participants were not required to provide input. Instead, a single robot performed the basic behavior as shown in the experiment outline and demonstrated the two anchor durations (short and long), repeating each seven times in a random sequence without any cues. Participants received auditory information through headphones, stating "short" or "long" after each trial, to help them familiarize themselves with the durations. Following this, we assessed their comfort with the task, offering the option to repeat this phase if clarification is needed.

The second phase was to measure whether the anchor durations had been learned. Similar to the training phase, it involved just one robot and the two anchor durations, each repeated seven times in a random order, without cues. Participants were asked to indicate whether the duration was long or short using the buttons. They then received auditory feedback ("false" or "correct") and visual feedback, with all robots lighting up in red or green, based on their response.

The third phase constituted the main experiment with a break for the participants after 50% of the trials. Each segment was designed to last approximately 15 minutes. In this phase, all robots were moving, and a random one would stop. Participants were instructed to press a button after they noticed the stopping and then observe the subsequent delay period. Before and after the break, every duration was presented 13 times under both non-cued and tactile-cued conditions. Participants were asked to judge if the observed delay period was more similar to the short or long anchor duration. In this round, the feedback was given for the anchor durations, with the robots illuminated in red or green and the auditory feedback ("false" or "correct") accordingly to remind the anchor durations throughout the experiment.

Throughout the experiment, we recorded the experiment state, including stopping, cue types, and button states. The data were captured with synchronized timestamps to extract the detection time of the temporal bisection response for each trial condition.

Data analysis. Logistic psychometric functions were fit to the proportion of "long" responses for each participant per condition in the timing task using the psignifit package in Python (Schütt et al., 2016). In order to assess duration judgments Bisection Points were calculated from the psychometric function as the stimulus duration with a 50% frequency of "long" responses. The Just Noticeable Difference (JND) was defined by half of the difference between the stimulus durations with 25% and 75% frequencies of "long" responses. As a temporal sensitivity measure, Weber Fractions (WF) were calculated as the ratio between JND and Bisection Point. Data from a participant were not included in the statistical analyses because of not properly adhering to task instructions.

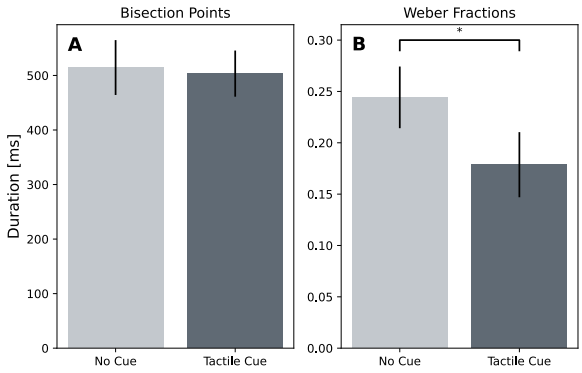

Fig. 6. Estimated parameters derived from the psychometric functions across cue conditions. (A) Bisection Points (B) Weber Fractions.Error bars depict the standard error of the mean. Asterisks depict significant differences.

4.2. Results

As expected, a one-sided paired sample t-test revealed that the RTs with tactile cue ($M = 1.08$, $SD = 0.123$) were significantly shorter than in the no cue condition ($M = 1.30$, $SD = 0.15$) $t(10) = 5.09$, $p < .001$, $d = 1.50$. To test if subjective duration judgments and temporal sensitivity were different in cue conditions, we applied one-sided t-tests between conditions to bisection points and Weber fractions, respectively. The results revealed that the tactile cue did not have a significant effect on the bisection points in the timing task $t(10) = 0.75$, $p > .46$, $d = .23$. However, as expected Weber fractions were significantly smaller in tactile cue ($M = .17$, $SD = 0.1$) than in the no cue condition ($M = .24$, $SD = 0.09$), $t(10) = 4.91$, $p < .001$, $d = .64$ showing that temporal sensitivity was better when tactile cues were present (Fig. 6).

5. Discussion

We investigated the effect of spatially uninformative signals on visual search performance and temporal judgments in a dynamic environment cluttered with robots. For this environment, we chose to work with small mobile robots that operate in a controlled but unpredictable pattern, mimicking the complexity and unpredictability of real-world scenarios. We found that auditory, tactile, and audiotactile brief signals decreased RTs in a visual search task. Furthermore, the second experiment showed that the presence of these cues increases the sensitivity of duration judgments, which might indicate a greater attentional orientation towards the target (Chica and Christie, 2009). Our findings extend the "pip-and-pop" effect (Van der Burg et al., 2008), demonstrating its applicability beyond simple visual displays to dynamic, real-world settings where robots are in motion.

The first experiment showed that auditory, tactile, and audiotactile brief cues facilitate the detection of a stopping robot amongst moving ones. Specifically, signals, that are temporally synchronized with an abrupt visual change in the environment, decrease RTs in the search task. We argue that this facilitation is essentially the "pip-and-pop" effect that has been investigated in simpler visual display settings (Van der Burg et al., 2008, 2009, 2010; Zou et al., 2012; Ngo and Spence, 2010a). This result might imply that auditory, tactile, and audiotactile cues can be incorporated into HSI and HRI interfaces to enhance the performance of the human supervisor. In dynamic and congested environments, brief and non-spatial cues can be easily used to direct the attention of the supervisor to the most crucial information that needs an urgent response. Various explanations for this

phenomenon have been proposed; oddball effect, early multisensory integration, and freezing effect. It has been argued that temporally aligned signals from another modality can turn the visual target in a segregated sequence of displays into an "oddball" that attracts attention (Ngo and Spence, 2012, 2010b). That is, it has been suggested that the auditory cue synchronized with the target display rendered that display an "oddball" which automatically attracted attention. Here, we did however not use a segregated sequence of visual displays; instead, we incorporated a continuous display that included an abrupt change at a random point in time. Therefore, there was no singular and distinct visual display that could have turned into an "oddball". Second, Van der Burg et al. (2008) proposed that the presentation of cues from another modality promotes an early multisensory integration process, which, in turn, enhances the saliency of a synchronized visual change and attracts attention. Third, Zou et al. (2012) found that the duration of fixation after the onset of the cues was lengthened, creating a "freezing effect" in which participants waited for further sample information from the visual environment. In this study, we did not investigate the validity of early multisensory integration or freezing effect accounts. However, we can confirm from both hypotheses that spatially uninformative but temporally synchronized signals increase the efficiency of the sensory processing towards the visual target by enhancing its saliency either by increased sampling or by early multisensory enhancement.

The RTs in the first experiment also showed that there were no differences between the cue conditions. That is, synchronized auditory, tactile, and audiotactile cues led to similar performance benefits, comparable to previous studies (Ngo and Spence, 2010a,b). This finding suggests that the tactile and auditory modalities can be similarly effective in realistic environments, such as robot supervision. Thus, the cue modality can be effectively exchanged in favor of the non-crowded channel depending on the situation. Moreover, a multisensory cue from audiotactile signals did not provide any further benefit compared to unisensory auditory or tactile cues. Thus, auditory and tactile cues alone were sufficient, and the addition of another sensory channel seems to be redundant. The explanations for both "multisensory integration" and "freezing effect" can provide an explanation on this result. It could be argued that already the presence of a single cue from another modality led to an early integration and boosted the saliency of the visual target. In this case, the addition of a third sensory channel can be redundant as unisensory cues were sufficient to provoke multisensory integration (Hagmann and Russo, 2016). In addition, both unisensory and bisensory cues could have led to the "freezing effect" or the "wait-at-beep" strategy. Regardless of the type of cue, participants may have extended the duration of fixation after cue onset (Zou et al., 2012).

An alternative explanation for the observed facilitation could be that the cues may have induced a general alerting effect, acting as a warning signal. Even though we did not experimentally rule out this possibility, we argue that an alerting mechanism is unlikely given the experimental design and the pattern of results. Alerting signals have been shown to facilitate reaction times while reducing accuracy (Fernandez-Duque and Posner, 1997). However, the results of the first experiment did not show any reduction in accuracy under cue conditions. Additionally, the alerting signals have been known to be most effective when they precede the event by 100 to 300 ms (Spence and Driver, 1997; McDonald et al., 2000). In contrast, the cues in our study were temporally synchronized with the visual change, which is consistent with previous findings of the "pip-and-pop" effect (Van der Burg et al., 2008, 2009). Therefore, while a general alerting mechanism cannot be completely ruled out, we argue that the facilitation is better explained by the "multisensory integration" and "freezing" mechanisms.

To investigate which sensory processes are involved in explaining this facilitation effect in visual search performance, we applied hierarchical drift–diffusion modeling. Although model comparisons showed that the best performing model was the one that kept all parameters free, further statistical analyses of the estimated parameters provided evidence that only the drift rate changes with the cue.This result can be attributed to the nature of hierarchical modeling. While the model comparison metrics evaluate the model by considering both fit and complexity, allowing all parameters to vary leads to more flexibility to capture individual-level variations. In this case, even though systematic differences of boundary and non-decision time between cue conditions were not captured in the group-level parameters, the best model, which kept all the parameters free, provided better fits, maybe because it captured individual variance better. Specifically, drift rates were higher for all of the cue conditions compared to no cue. Since higher drift rates reflect a faster and more efficient sensory acquisition, we can argue that auditory, tactile, and audiotactile cues increased the information sampling speed and efficiency (Ratcliff and McKoon, 2008). This effect is consistent with multisensory integration and subsequent increase in saliency of pip and pop, as increased saliency can facilitate target processing by attention capture (Zehetleitner and Müller, 2010; Itti and Koch, 2000). Similarly, the "freezing" effect proposed by Zou et al. (2012) could well explain the differences in the drift rates since the increased sampling period on the visual display essentially elicits faster and more efficient information gathering from the display. Consequently, the current modeling results are in favor of both accounts of the "pip-and-pop" effect.

We used a haptic vest to provide tactile cues on the human torso. Previously, Van der Burg et al. (2009) presented these cues from the wrist and Ngo and Spence (2010b) from the waist. Since we observed the "pip-and-pop" effect in this study, we can conclude that tactile cues elicit comparable effects regardless of the location they were presented from. Furthermore, the wrist and waist as well as hands provide restricted areas and can interfere with the user's movements. In contrast, the usage of the torso as a medium for cue presentation poses its advantages due to its extensive surface area and notably its non-disruptive nature during individuals' ongoing tasks or work (Botev et al., 2021).

The second experiment replicated the same effect on RTs with only tactile cues. Furthermore, the timing task after the visual search revealed that participants had a higher temporal sensitivity in the tactile cue condition compared to the no cue. However, subjective duration judgments were similar. Therefore, participants were more precise in timing the duration elapsed when a tactile cue was present. What is important to note here is that the timing task was done immediately after the participants identified the visual target. Thus, these results were sampled from a period in which attention was already deployed to the target as a result of a successful search. Here, we suggest that, as a consequence of a boosted saliency (Van der Burg et al., 2008) or increased sampling from the environment and the target (Zou et al., 2012), participants were able to direct their attention to the target more efficiently. Consequently, more resources were made available to further process the target because of the enhanced attention. In this context, previous studies have shown that, under temporally unconstrained response windows, temporal sensitivity increases with spatial attention (Chica and Christie, 2009). Similarly, in this experiment, increased attention on the visual target caused by the non-visual cues improved the temporal sensitivity.

The application side of this research highlights the potential for integrating multisensory cues in HSI to enhance performance in complex real-world tasks. By demonstrating that spatially uninformative, temporally aligned signals can significantly improve human performance in a visual search task in terms of RTs, our findings suggest that similar strategies could be employed to design more effective HSI systems. For example, in scenarios where humans and robots work together in dynamic environments, such as search and rescue operations or industrial settings, the use of auditory, tactile, or audiotactile cues could facilitate quicker and more accurate human responses to changes in robot behavior or to critical events. This approach not only leverages the natural human sensory processing capabilities by adding new perception channels but also suggests a practical pathway to improve the efficiency and safety of HSI applications.

While our results demonstrate the effectiveness of brief and spatially uninformative signals to facilitate visual search and enhance temporal sensitivity in HSI applications, we did not investigate the underlying cognitive processes such as multisensory integration or freezing effect. Future studies could employ neurophysiological and eye-tracking methods with appropriate experimental designs to explore these mechanisms. Moreover, future work should assess the temporal boundaries of this effects since in practical scenarios, delivering cues precisely at the moment of an unexpected robot behavior may not be feasible. Therefore, it is crucial to examine how effectiveness of these cues is influenced by temporal delays.

Overall, we demonstrated that spatially uninformative auditory, tactile, and audiotactile brief signals, temporally aligned with visual change onset, effectively reduce RTs in a visual search task involving moving robots. This outcome extends the understanding of the "pip-and-pop" effect observed in controlled laboratory environments to realistic environments. Additionally, our findings underscore the utility of these signals in improving temporal sensitivity, possibly due to enhanced attentional orientation to the visual target. Thus, spatially uninformative signals provide considerable benefits to both human search performance and precision in judging the time of elapsed events. This has significant implications for improving HSI, particularly in environments where quick decision-making and attention to critical events are important.

More specifically, the scenario highlights the possibility of decreasing human response time in detecting a robot signaling via a motion stop, a behavior applicable in contexts ranging from malfunction detection to signaling points of interest for operator evaluation. Based on these findings, we recommend that HSI system designers incorporate brief auditory and tactile cues that are precisely synchronized with critical visual events, such as a robot stopping or signaling, even if these cues do not provide spatial information. Our data indicate that such spatially uninformative cues effectively capture attention, as reflected in the faster RTs. Importantly, the effectiveness of these cues was evident in a cluttered, dynamic environment with multiple moving robots, demonstrating their practical utility in real-world scenarios. To maximize benefits, combining auditory and tactile cues appears to be most effective, likely due to cross-modal enhancement of saliency. Designers should also ensure cues remain brief and salient enough to attract attention without overwhelming the operator or interfering with ongoing tasks, as overly long or delayed cues could disrupt performance. Implementing these guidelines can enhance human supervisory performance by reducing detection times and improving temporal judgments, supporting quicker and more accurate decision-making in complex, time-critical HSI contexts.

CRediT authorship contribution statement

Bora Celebi: Writing – original draft, Visualization, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Julian Kaduk:** Writing – original draft, Visualization, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Müge Cavdan:** Writing – review & editing, Validation, Methodology, Investigation, Data curation. **Heiko Hamann:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Knut Drewing:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Knut Drewing reports financial support was provided by European Union. Heiko Hamann reports financial support was provided by European Union. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Table A.1
Proportion of differences of posterior distribution of fitted parameters between conditions.

Parameter	Conditions	$P_{P\|D}$
Drift Rate	No Cue < Auditory	0.99
	No Cue < Tactile	1.0
	No Cue < Autiotactile	1.0
	Auditory < Tactile	0.52
	Auditory < Audiotactile	0.76
	Tactile < Audiotactile	0.74
Boundary	No Cue < Auditory	0.16
	No Cue < Tactile	0.25
	No Cue < Autiotactile	0.30
	Auditory < Tactile	0.56
	Auditory < Audiotactile	0.67
	Tactile < Audiotactile	0.59
Non-Decision Time	No Cue < Auditory	0.51
	No Cue < Tactile	0.50
	No Cue < Autiotactile	0.51
	Auditory < Tactile	0.49
	Auditory < Audiotactile	0.50
	Tactile < Audiotactile	0.51

Acknowledgments

This research received funding from the European Union's Horizon 2020 FET Open research program under grant agreement no. 964464, for the ChronoPilot project.

Appendix. Supplementary table

See Table A.1.

Data availability

The link to the data and code repositories are included in the manuscript.

References

Allenmark, F., Müller, H.J., Shi, Z., 2018. Inter-trial effects in visual pop-out search: Factorial comparison of Bayesian updating models. PLoS Comput. Biol. 14 (7), e1006328.

Andrieu, C., de Freitas, N., Doucet, A., Jordan, M.I., 2003. An introduction to MCMC for machine learning. Mach. Learn. 50 (1), 5–43.

Arend, M.G., Benz, T.M., Mertens, A., Brandl, C., Nitsch, V., 2023. Do multimodal search cues help or hinder teleoperated search and rescue missions? Ergonomics 66 (9), 1255–1269.

Association, W.M., et al., 2013. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. Jama 310 (20), 2191–2194.

Barros, P.G.d., Lindeman, R.W., 2014. Multi-sensory urban search-and-rescue robotics: Improving the operator's omni-directional perception. Front. Robot. AI 1, 14.

Bolia, R.S., D'Angelo, W.R., McKinley, R.L., 1999. Aurally aided visual search in three-dimensional space. Hum. Factors 41 (4), 664–669.

Botev, J., Drewing, K., Hamann, H., Khaluf, Y., Simoens, P., Vatakis, A., 2021. ChronoPilot — Modulating time perception. In: 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality. AIVR, IEEE, pp. 215–218.

Van der Burg, E., Cass, J., Olivers, C.N.L., Theeuwes, J., Alais, D., 2010. Efficient visual search from synchronized auditory signals requires transient audiovisual events. PLoS One 5 (5), e10664.

Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2008. Pip and pop: nonspatial auditory signals improve spatial visual search. J. Exp. Psychol. Hum. Percept. Perform. 34 (5), 1053–1065.

Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2009. Poke and pop: Tactile–visual synchrony increases visual saliency. Neurosci. Lett. 450 (1), 60–64.

Celebi, B., Cavdan, M., Drewing, K., 2025. Spatial attention modulates time perception on the human torso. Atten. Percept. Psychophys. 87 (3), 779–793.

Cha, E., Mataric, M., Fong, T., 2016. Nonverbal signaling for non-humanoid robots during human-robot collaboration. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction. HRI, pp. 601–602.

Chan, L.K.H., Hayward, W.G., 2013. Visual search. Wiley Interdiscip. Rev. Cogn. Sci. 4 (4), 415–429.
Chica, A.B., Christie, J., 2009. Spatial attention does improve temporal discrimination. Atten. Percept. Psychophys. 71 (2), 273–280.
Correa, A., Sanabria, D., Spence, C., Tudela, P., Lupiáñez, J., 2006. Selective temporal attention enhances the temporal resolution of visual perception: Evidence from a temporal order judgment task. Brain Res. 1070 (1), 202–205.
Coull, J.T., Nobre, A.C., 1998. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. J. Neurosci. 18 (18), 7426–7435.
Doyle, M.C., Snowden, R.J., 1998. Facilitation of visual conjunctive search by auditory spatial information. Perception.
Egeth, H., 1977. Attention and preattention. In: Bower, G.H. (Ed.), Psychology of Learning and Motivation. In: The psychology of learning and motivation, vol. 11, Elsevier, San Diego, CA, pp. 277–320.
Erp, J.B.V., Veen, H.A.V., 2004. Vibrotactile in-vehicle navigation system. Transp. Res. Part F: Traffic Psychol. Behav. 7 (4–5), 247–256.
Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39 (2), 175–191.
Fernandez-Duque, D., Posner, M.I., 1997. Relating the mechanisms of orienting and alerting. Neuropsychologia 35 (4), 477–486.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall/CRC.
Gray, R., Mohebbi, R., Tan, H.Z., 2009. The spatial resolution of crossmodal attention: Implications for the design of multimodal interfaces. ACM Trans. Appl. Percept. 6 (1), 1–14.
Hagmann, C.E., Russo, N., 2016. Multisensory integration of redundant trisensory stimulation. Atten. Percept. Psychophys. 78 (8), 2558–2568.
Ho, C., Tan, H.Z., Spence, C., 2005. Using spatial vibrotactile cues to direct visual attention in driving scenes. Transp. Res. Part F Traffic Psychol. Behav. 8 (6), 397–412.
Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40 (10–12), 1489–1506.
Kassner, M., Patera, W., Bulling, A., 2014. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In: Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. In: UbiComp '14 Adjunct, ACM, New York, NY, USA, pp. 1151–1160.
Kothe, C., 2024. Lab streaming layer (LSL). Software framework for real-time synchronization, https://github.com/sccn/labstreaminglayer.
Kruschke, J.K., 2010. Bayesian data analysis. Wiley Interdiscip. Rev.: Cogn. Sci. 1 (5), 658–676.
Kumar, R., Carroll, C., Hartikainen, A., Martin, O., 2019. ArviZ a unified library for exploratory analysis of Bayesian models in python. J. Open Source Softw. 4 (33), 1143.
Lee, M., Fuss, I., Navarro, D., 2006. A Bayesian approach to diffusion models of decision-making and response time. Adv. Neural Inf. Process. Syst. 19.
Mattes, S., Ulrich, R., 1998. Directed attention prolongs the perceived duration of a brief stimulus. Percept. Psychophys. 60 (8), 1305–1317.
Matthews, W.J., Meck, W.H., 2016. Temporal cognition: Connecting subjective time to perception, attention, and memory. Psychol. Bull. 142 (8), 865–907.
McDonald, J.J., Teder-Sälejärvi, W.A., Hillyard, S.A., 2000. Involuntary orienting to sound improves visual perception. Nature 407 (6806), 906–908.
Ngo, M.K., Spence, C., 2010a. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. Atten. Percept. Psychophys. 72 (6), 1654–1665.
Ngo, M.K., Spence, C., 2010b. Crossmodal facilitation of masked visual target identification. Atten. Percept. Psychophys. 72 (7), 1938–1947.
Ngo, M.K., Spence, C., 2012. Facilitating masked visual target identification with auditory oddball stimuli. Exp. Brain Res. 221 (2), 129–136.
Oriol, A.-P., Virgile, A., Colin, C., Larry, D., J., F.C., Maxim, K., Ravin, K., Jupeng, L., C., L.C., A., M.O., Michael, O., Ricardo, V., Thomas, W., Robert, Z., 2023. PyMC: A modern and comprehensive probabilistic programming framework in python. PeerJ Comput. Sci. 9, e1516.
Pawlak, W.S., Vicente, K.J., 1996. Inducing effective operator control through ecological interface design. Int. J. Hum. Comput. Stud. 44 (5), 653–688.
Polar, 2024. Polar H10 heart rate sensor. ECG strap, https://www.polar.com/us-en/products/accessories/h10_heart_rate_sensor.
Quinlan, P.T., 2003. Visual feature integration theory: past, present, and future. Psychol. Bull. 129 (5), 643–673.
Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85 (2), 59–108.
Ratcliff, R., McKoon, G., 2008. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput. 20 (4), 873–922.
Reardon, C., Lee, K., Fink, J., 2018. Come see this! augmented reality to enable human-robot cooperative search. In: 2018 IEEE International Symposium on Safety, Security, and Rescue Robotics. SSRR, 00, pp. 1–7.
Riedo, F., Chevalier, M., Magnenat, S., Mondada, F., 2013. Thymio II, a robot that grows wiser with children * *this work was supported by the swiss national center of the competence in research "robotics". In: 2013 IEEE Workshop on Advanced Robotics and Its Social Impacts. pp. 187–193.
Schütt, H.H., Harmeling, S., Macke, J.H., Wichmann, F.A., 2016. Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vis. Res. 122, 105–123.
Seifried, T., Ulrich, R., 2011. Exogenous visual attention prolongs perceived duration. Atten. Percept. Psychophys. 73 (1), 68–85.
Shi, Z., Chen, L., Müller, H.J., 2010. Auditory temporal modulation of the visual ternus effect: the influence of time interval. Exp. Brain Res. 203 (4), 723–735.
Spence, C., Driver, J., 1997. Audiovisual links in exogenous covert spatial orienting. Percept. Psychophys. 59 (1), 1–22.
Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27 (5), 1413–1432.
Vroomen, J., de Gelder, B., 2000. Sound enhances visual perception: cross-modal effects of auditory organization on vision. J. Exp. Psychol. Hum. Percept. Perform. 26 (5), 1583–1590.
Wearden, J., Ferrara, A., 1995. Stimulus spacing effects in temporal bisection by humans. Q. J. Exp. Psychol. 48 (4), 289–310.
Wiecki, T.V., Sofer, I., Frank, M.J., 2013. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion model in python. Front. Neuroinform. 7, 14.
Wolfe, J.M., 2020. Visual search: How do we find what we are looking for? Annu. Rev. Vis. Sci. 6 (1), 539–562.
Yeshurun, Y., Marom, G., 2008. Transient spatial attention and the perceived duration of brief visual events. 16, (6), pp. 826–848.
Zehetleitner, M., Müller, H.J., 2010. Salience from the decision perspective: You know where it is before you know it is there. J. Vis. 10 (14).
Zou, H., Muller, H.J., Shi, Z., 2012. Non-spatial sounds regulate eye movements and enhance visual search. J. Vis. 12 (5), 2–2.