Autonomous is Not Enough: Designing Multisensory Mid-Air Gestures for Vehicle Interactions Among People with Visual Impairments

Paul D. S. Fink*
VEMI Lab, The University of Maine
Orono, ME, USA
paul.fink@maine.edu

Velin Dimitrov
Toyota Research Institute
Cambridge, MA, USA
velin.dimitrov@tri.global

Hiroshi Yasuda
Toyota Research Institute
Los Altos, CA, USA
hiroshi.yasuda@tri.global

Tiffany L. Chen
Toyota Research Institute
Los Altos, CA, USA
tiffany.chen@tri.global

Richard R. Corey
VEMI Lab, The University of Maine
Orono, ME, USA
richard.r.corey@maine.edu

Nicholas A. Giudice
VEMI Lab, The University of Maine
Orono, ME, USA
nicholas.giudice@maine.edu

Emily S. Sumner
Toyota Research Institute
Los Altos, CA, USA
emily.sumner@tri.global

ABSTRACT

Should fully autonomous vehicles (FAVs) be designed inclusively and accessibly, independence will be transformed for millions of people experiencing transportation-limiting disabilities worldwide. Although FAVs hold promise to improve efficient transportation without intervention, a truly accessible experience must enable user input, for all people, in many driving scenarios (e.g., to alter a route or pull over during an emergency). Therefore, this paper explores desires for control in FAVs among (n=23) people who are blind and visually impaired. Results indicate strong support for control across a battery of driving tasks, as well as the need for multimodal information. These findings inspired the design and evaluation of a novel multisensory interface leveraging mid-air gestures, audio, and haptics. All participants successfully navigated driving scenarios using our gestural-audio interface, reporting high ease-of-use. Contributions include the first inclusively designed gesture set for FAV control and insight regarding supplemental haptic and audio cues.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility design and evaluation methods**; **Empirical studies in accessibility**; *Haptic devices*; **Gestural input**.

KEYWORDS

Autonomous vehicles, Accessible design, Interfaces for blind or visually impaired individuals, Spatial audio, Gestures, Situational awareness

ACM Reference Format:
Paul D. S. Fink, Velin Dimitrov, Hiroshi Yasuda, Tiffany L. Chen, Richard R. Corey, Nicholas A. Giudice, and Emily S. Sumner. 2023. Autonomous is Not Enough: Designing Multisensory Mid-Air Gestures for Vehicle Interactions Among People with Visual Impairments . In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544548.3580762

*Author conducted this work at Toyota Research Institute.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580762

1 INTRODUCTION

Fully autonomous vehicles (FAVs) hold enormous potential to transform mobility for the roughly 300 million people who are blind and visually impaired (BVI) worldwide [2, 25]. Today, BVI people must rely on others for transportation, either through friends, family, public transportation, or rideshare. FAVs that are designed accessibly will enable independent travel among BVI people, thus resulting in more mobility and personal autonomy. However, for independence and autonomy to be fully maximized, it is argued here that people will desire to be "in the loop" of vehicle control. Whether it is personalizing the vehicle's driving style (e.g., speed or following distance), giving input on where to be dropped off, or even changing the route entirely, it stands to reason that having some sense of control over the trip is part and parcel with increased independence and autonomy. Indeed, the connective tissue between FAV control and independence was illuminated in Brewer and Kameswaren's (2018) focus group study exploring (n=15) BVI people's perceptions of autonomous vehicles. Their findings demonstrated that people desire control across the spectrum of vehicle autonomy and that new mechanisms are needed to enable actionable behavior [4].

It should be noted that although FAV control for BVI people is a promising goal, information access to the surrounding environment and driving context is a prerequisite. Termed situational awareness, recent research has found that to be inclusive, FAVs must be designed to increase understanding of the vehicle's decision making process [3, 5] as well as to give details about the surrounding environment [5, 6]. Situational awareness is essential to independence because it increases understanding of the driving environment such that actionable behavior and control are possible across the trip (e.g., for route planning). The extant research in this domain has postulated that multisensory interfaces that combine haptic (active touch) and auditory cues in FAVs represent an exemplary approach in pursuit of this goal. For instance, Brewer and Kameswaren (2018) suggest employing tactile interactions (e.g., those that mimicked the use of a white cane) in tandem with audio cues for conveying the driving environment and altering vehicle behavior, and Fink et al. (2021) suggest using vibro-audio maps for conveying route-based information [12]. The rationale for designing BVI interfaces with multimodal input-output processes include supporting distributed cognitive load across the senses and fewer demands on working memory than those that rely on a single modality [26]. However, despite being proposed for use in FAVs, few systems have been developed to convey situational awareness harnessing the benefits of multimodality, instead relying on auditory interaction [7] or vibrotactile output [30] alone.

Mid-air gestural systems are an emerging interaction technology with a number of advantages that motivated this research in terms of multisensory and accessible FAV use among people with visual impairments. First, a key advantage of mid-air gestures is that, unlike traditional touchscreen-based vehicle displays, gestures can be performed in free space without the guiding use of vision [18]. This non-visually dependent nature of gestural systems affords significant opportunity to increase natural and accessible interactions for BVI users over traditional visual-only vehicle displays. There are also significant hygienic advantages of this approach (consider that FAVs may well adopt rideshared service models and that knowing if a shared service is clean is often challenging for BVI users). Furthermore, gestures are location-independent and can be performed at a distance [35], for example throughout a vehicle cabin opposed to confined at a central display, which would afford greater flexibility for seating arrangements in future FAVs. It should be noted that handheld smartphones and dedicated accessibility devices offer some of these same advantages and will likely continue to be popular among the BVI demographic. Indeed, handheld devices are ideal for certain navigation tasks like feeling a map. However, mid-air gestural interaction presents the opportunity to offload tasks that can be performed as a natural extension of body movement, thereby enabling computational resources and interaction on existing devices that support the benefits all people will gain from driverless transportation: more time for socializing, work, and relaxation.

Recognizing these advantages, this paper explores a user-driven interface for increasing situational awareness and control in automated vehicles via mid-air gestural interaction. To do so, we first conducted a needs assessment with (n=23) BVI users to identify the types of vehicle control that are important to this demographic and the situational information necessary to be conveyed (Study 1 in Section 3). A subsequent user study session involving (n=15) participants who also completed the Study 1 survey explored the design of a mid-air gestural system to promote multisensory control (Section 4). Finally, the resulting experimental interface, which combines ultrasound-based haptic representations of the driving environment, queryable spatialized audio descriptions, and mid-air gestures to mediate between the two, was evaluated with (n=8) BVI participants from the original Study 1 group (separate from the Study 2 group). Results provide compelling evidence for increased BVI situational awareness and control potential in partnership with FAVs and identifies a first-of-its-kind gesture set for FAV control that promotes inclusion (Section 5). This system is designed to serve both BVI people who have previously operated traditional vehicles, as well as people who have never driven before, representing broad and inclusive usability across the spectrum of vision loss.

2 RELATED WORK

The research presented here was informed by the small but growing body of work exploring accessibility in FAVs for BVI users. The following reviews this work, as well as the ways in which mid-air gestural interaction has been used in the driving context and among the BVI demographic more generally.

2.1 FAVs and BVI Individuals

FAVs are predicted to have outsized impacts on underserved transportation populations, including BVI travelers, in terms of increased mobility, workforce participation, and overall quality of life [9]. A number of studies have examined the perceptions of this demographic with regard to automated driving. For instance, Brinkley et al. (2020) conducted a survey with 516 BVI respondents and subsequent focus groups (n=38), with results indicating strong support for FAVs, interest in ownership, but concerns regarding accessible design [6]. Likewise, Bennett, Vijaygopal, and Kottasz (2020) conducted a survey with 211 BVI participants and found favorable attitudes towards automated driving but, again, skepticism with regard to the accessibility of this technology [1]. Related research has also postulated how to make this technology accessible, indicating the need for new policy frameworks [12] and interfaces that enable understanding and control [4, 12]. Perhaps most important to enabling this understanding and control are findings that suggest that BVI users desire increased situational awareness in FAVs [3, 5, 6]. The logic here is that in order to adequately understand the environment such that control actions can be performed safely, users desire more information about the driving situation and context.

Although the available research suggests the importance of new human-machine interfaces (HMIs) to increase access and situational awareness in FAVs, there has been relatively little work exploring accessible FAV user interfaces. A 2021 systematic review of the literature indicates that only two HMIs have been designed or evaluated for fully autonomous use among BVI people, with only one involving a user study [11]. For instance, a text-to-speech and speech-to-text system was developed for use in FAVs with computationally efficient results, but did not involve user testing [33]. The Accessible Technology Leveraged for Autonomous vehicles System (ATLAS) was designed as a speech input and audio output system with extensive feedback from users and tested with 20 participants [7]. Although the ATLAS study results are incredibly encouraging in terms of user trust and usability, the system relies on audio as the only non-visual modality, which may present disadvantages when audio interaction is undesirable (e.g., in a loud scenario or when a fellow passenger is sleeping), nor does it harness the previously discussed intrinsic benefits of multimodality. Likewise, a 2022 study with (n=5) blind participants investigated vibrotactile feedback delivered using the Ready-Move and Ready-Ride devices on the wrist, hands, chest, and back, providing encouraging results in terms of finding the vehicle, receiving information during the trip, and arriving at the destination, but did not explore modalities beyond haptics [30]. As such, no research to date has evaluated a non-visual interface for FAV use leveraging multiple senses, as we propose to do here mediated by mid-air gestural interaction.

2.2 Mid-Air Gestural Interaction

Mid-air gestures (such as a wave hello or a thumbs up) are location independent movements performed in free space that predominately involve manipulation of the wrist, arm, and hand position [24, 34]. Unlike vehicle touchscreens, which predominantly require the use of vision (see [27, 28] for the limited examples of multisensory touchscreen usage), gestures are non-viually dependent. Recognizing this advantage for eyes-on-the-road time, mid-air gestures have begun to gain traction in the automotive domain, with several studies exploring the design and implementation of UI elements on infotainment displays using mid-air gesturing as the primary interaction modality [8, 19, 23, 32]. This body of work demonstrates that driving performance and safety can be improved by complementing gestural interaction with haptic and audio interaction. However, no work to our knowledge has leveraged the non-visual advantage of mid-air gestural systems to improve access to control in automated vehicles for people with visual impairment, as is the focus of this research.

Gestural interaction has, on the other hand, begun to gain traction in the FAV literature for manipulating driving behavior among sighted users. For instance, Qian et al. (2020) conducted a user study in a vehicle simulator to identify a set of static hand-shape gestures (held for 10 seconds) for controlling autonomous vehicles across common driving tasks (i.e., go straight, turn left, turn right, stop, slow down, back up, turn around, and pull over). Users performed gestures in three locations (steering wheel area, shifting area, and free region/open-cabin), and despite executing gestures more efficiently in the shifting area, preferred the free region condition. Questionnaire results supported the use of gestural based navigation in autonomous vehicles, particularly over short distances or as a backup form of interaction if other software failed [29]. Detjen et al. (2020) also found encouraging results of maneuver based vehicle control via gestural interaction during driving tasks similar to Qian et al.'s (2020) stimulus set, albeit with higher task load than speech and touch [10]. Although research investigating vehicle control via gestural systems has yet to include BVI people, gestural interaction has shown promising results when combined with multimodal feedback among this population more generally. For instance, Kim et al. (2016) explored use of a mid-air gesture system by BVI people to navigate a large public video display and found that audio and haptic feedback improved navigation performance compared to one modality alone [21]. Likewise, Gross et al. (2018) found positive navigation performance and low cognitive load among BVI people using a gestural system combined with audio to navigate web-based menu structure [18]. Taken together, this body of work suggests that the nonvisual advantages of mid-air gestures have the potential to increase access and control in FAVs among BVI people, particularly when combined with supplemental audio and haptic cues.

3 STUDY 1, NEEDS ASSESSMENT SURVEY

3.1 Motivation for User-driven Design

The studies presented in this work followed a principled trajectory where early results informed later design decisions, beginning with a user needs assessment via a survey delivered to BVI individuals (n=23). The needs assessment survey sought to identify the FAV driving tasks over which participants desired control, as well as the situational information that would be necessary for each driving task. Our goal was to use results from this initial phase (i.e., driving task importance and information needs) to inform the driving context and information required in the subsequent interface study (Section 5).

3.2 Methods

The Study 1 survey aimed to assess the types of information and importance of control across a range of common driving tasks. Driving tasks were adapted from Qian et al.'s (2020) stimulus set and were grouped along categories of task to reduce redundancy: *stop/start behavior*, *maneuvering behavior* (e.g., left, right, straight), *speed manipulation* (e.g., speed up, slow down), and *pulling over behavior*. Two other types of driving tasks were added to the stimulus set: *altering the route* and *adjusting the following distance*. *Altering the route* was added because of its relevance to the fully autonomous context of interest to this paper and *following distance manipulation* was added given current capabilities in consumer available driver assistance systems. The study included two types of questions. Five-point Likert scales were used to rate the importance of personal control over each type of driving task from 1 - Not important to 5 - Very important. Open-ended questions were used to identify the situational information users would need or want to issue a specific driving task command (e.g., change the following distance) when riding with FAVs.

3.2.1 Participants. Participants were recruited through a mailing list by the Carroll Center for the Blind, a facility that focuses on serving the blind and low vision community in the greater Boston, Massachusetts area. Participants (n=23), all identifying as blind, represented a broad spectrum of vision loss, onset, etiology (specific visual demographics for each participant can be found in Table 1) and age, ranging from 28-71 (M = 50.48, SD = 14.25). Of these, 13 identified as former drivers and 10 identified as having never driven before. Participants predominantly identified as white or of European descent (73.91%). 8.72% identified as black or African American and 95.65% reported as not identifying as ethnically Hispanic or Latino/x, while 4.35% of participants did. Four participants chose not to indicate racial or ethnic identify. 30.43% of participants had attained a Bachelor's degree, 26.09% a Master's degree, 21.74% some college but no degree, 4.35% an associate's degree, 4.35% High

Table 1: Vision loss etiology and extent for each participant

Etiology of Blindness	Residual Vision	Study
Retinitis pigmentosa	No usable vision	1 & 2
Unknown	Severe vision loss	1 & 2
Norrie syndrome	No usable vision	1 & 2
Cancer of the retina	No usable vision	1 & 2
Retinopathy of prematurity	No usable vision	1 & 2
Retinitis pigmentosa	Some light and shape perception	1 & 2
Diabetic retinopathy	Central vision with a 10 degree field	1 & 2
Unknown	No usable vision	1 & 2
Retinitis pigmentosa and cataracts	Some light and shape perception	1 & 2
Cortical blindness due to stroke	No usable vision	1 & 2
Congenitally low vision	No usable vision in right eye. 20/250 in left eye	1 & 2
Ushers Syndrome II	No reported vision	1 & 2
Retinopathy of prematurity	No usable vision	1 & 2
Autoimmune retinopathy and posterior sclerosis	No reported vision	1 & 2
Retinopathy of prematurity	20/400 in one eye with limited field	1 & 2
Glaucoma and corneal opacities	Able to recognize large objects at 1 foot or closer	1 & 3
Retinopathy of prematurity and glaucoma	20/7000. Some light and contrast vision. 10 degree field	1 & 3
Leber congenital amaurosis	No usable vision	1 & 3
Unknown	20/300. Steady	1 & 3
High blood pressure in eyes	No usable vision	1 & 3
Cone dystrophy	1 or 2 fingers at approximately 1 foot	1 & 3
Injury	No usable vision	1 & 3
Injury	Usable peripheral vision in both eyes	1 & 3

school or equivalent, and 4.35% a Ph.D. Two participants chose not to report educational attainment.

3.2.2 Procedure. The survey was delivered in-person and proctored by an experimenter who entered participant responses in Qualtrics. Each response was read aloud to participants and verified to be accurate prior to submission. This research was approved by the University of Maine IRB and participants were compensated for their travel and participation time ($100/study hour), in line with the Carroll Center for the Blind's recommendations.

Participants began the survey being asked to "imagine riding in a fully autonomous vehicle that can take you where you need to go safely, efficiently, and legally, without any required intervention on your part." Then participants were asked to think about and tell the experimenter what information they would want or need to decide to control the driving task (e.g., control the speed, either speed up or slow down). After this, participants were asked to rate how important being able to control that driving task would be from 1-Not Important to 5-Very Important. Both the long answer question and importance score were recorded in Qualtrics. This process repeated across the *stop/start*, *maneuvering*, *speed manipulation*, *following distance*, *altering the route*, and *pulling over* driving behaviors.

3.2.3 Hypotheses. The four Hypotheses for Study 1 were organized under two overarching research questions:

RQ1: What **types of vehicle control are important to BVI people** in autonomous vehicles?

The first hypothesis was derived both from the existing literature [3, 5] and informal input our group has received with regard to the importance of route-based control, as these behaviors have most influence on the success of the trip.

> H1: BVI people will have stronger preference for controlling and *altering the route* than other driving tasks (e.g., *following distance*, *vehicle speed*, and *starting/stopping*).

Although we predicted that route-based control would be the most important across participant, it stands to reason that former drivers might value control over the process of driving than those who have never driven before. As such, our second hypothesis stated:

> H2: People who have driven before will demonstrate stronger preference for non-route based control (e.g., *following distance*, *speed*, *turning behavior*, *starting and stopping*) than people who have never driven before.

The second research question pertained to the required information for situational awareness:

RQ2: What **types of driving information are necessary** when considering control in autonomous vehicles? Given that situational awareness includes both the vehicle's operational space and the surrounding environment, we hypothesized that:

> H3: Both behavioral information (what the vehicle is doing and will do next) and environmental information (what is in the driving environment) will be important to BVI people opposed to one category over the other.

Much like our first set of hypotheses, we also predicted that information pertaining to the route would be prioritized, as this is most relevant to the driving task of efficiently and safely reaching the destination:

> H4: *Route-based information* (e.g., time-to-destination) and *route objects* (e.g., roads/intersections, points of interest (POIs)) will be emphasized more than *non-route information/objects* (e.g., speed, following distance, pedestrians, etc.).

3.3 Results

The results of Study 1 (Figure 1) showed strong support for control across driving action. The mean importance score for each driving task category was greater than 3, with control over *altering the route* and *starting/stopping* equal to 4.9 and 4.7 respectively.

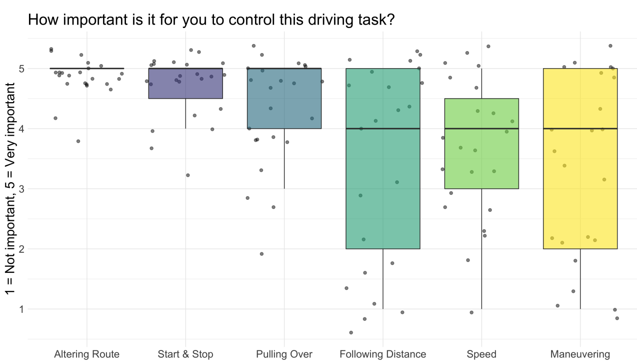


Figure 1: The perceived importance of being able to control certain driving tasks

As within-subject, non-parametric factors, we conducted a Friedman's test to analyze statistical significance of this difference. Control Type demonstrated a significant effect on subjective importance scores ($\chi^2(5) = 40.819, p < .001$). As shown in Table 2, post-hoc pairwise comparisons showed that importance is significantly different between *altering the route* and the remaining types of control (all $p < 0.05$), except *starting and stopping behavior* ($p = .417$), and *pulling over* when using Bonferroni and Holm correction ($p = .348$ and $p = .188$).

Table 2: Conover's Post Hoc Comparisons - Control Type

		T-Stat	df	W_i	W_j	p	p_{bonf}	p_{holm}
Altering Route	Start/Stop	0.816	110	109	100.5	0.417	1.000	1.000
	Pulling Over	2.303	110	109	85	0.023	0.348	0.185
	Following Distance	4.318	110	109	64	< .001	< .001	< .001
	Speed	4.174	110	109	65.5	< .001	< .001	< .001
	Maneuvering	4.797	110	109	59	< .001	< .001	< .001
Start/Stop	Pulling Over	1.487	110	100.5	85	0.140	1.000	0.699
	Following Distance	3.502	110	100.5	64	< .001	0.010	0.007
	Speed	3.358	110	100.5	65.5	0.001	0.016	0.011
	Maneuvering	3.982	110	100.5	59	< .001	0.002	0.001
Pulling Over	Following Distance	2.015	110	85	64	0.046	0.695	0.324
	Speed	1.871	110	85	65.5	0.064	0.960	0.384
	Maneuvering	2.495	110	85	59	0.014	0.211	0.127
Following Distance	Speed	0.144	110	64	65.5	0.886	1.000	1.000
	Maneuvering	0.480	110	64	59	0.632	1.000	1.000
Speed	Maneuvering	0.624	110	65	59	0.534	1.000	1.000

Taken together, these results demonstrate support for H1, which predicted that *altering the route* would be rated as more important than other types of FAV control, understanding that *starting and stopping* the vehicle and *pulling over behavior* are also important relative to other types of vehicle control.

We also analyzed the extent to which prior driving experience impacts rated importance across driving control type. Although, surprisingly, Figure 2 suggests that people who have driven before rate control importance lower than people who have never driven before across the types of control, a mixed-model non-parametric test suggests that this difference is not statistically significant $F(1, 21) = .284, p = .107$. Taken together, this analysis does not find support for H2, which predicted that people who have driven before will demonstrate stronger preference for *non-route based control* than people who have never driven.

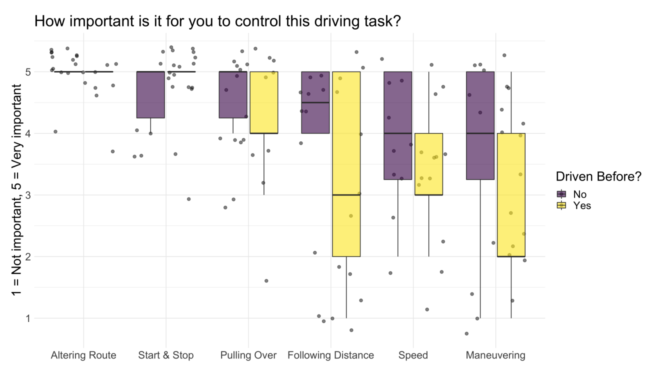


Figure 2: The perceived importance of being able to control certain driving tasks broken down by whether the participant has prior experience driving or not

The long answer questions illuminated the importance of situational information and awareness across FAV control type. In general, participants imagined wanting a significant amount of information during the trip. For instance, one participant mentioned:

> P16: "I want any relevant information [the FAV] could give me: what caused the [driving] situation to begin with and will it pose a problem if we change."

Participants also mentioned wanting the capability of control over driving tasks, even if they chose not to intervene. Another participant mentioned:

> P13: "I want control, even just for the sense of it. I might not always use it, but if I knew something was going on, like there were emergency cars ahead or a problem ahead, I want to know that I can say, 'let's take a different route, let's turn around.'"

In order to inform the design of the experimental interface and task used in Study 3, we also coded these long answer questions to determine the most frequently mentioned information or scenarios under which participants would want to undertake control of the vehicle. In support of H3, people mentioned wanting to know vehicle behavioral information like its speed or upcoming turn, as well as environmental information such as traffic. Information related to safety or that which would be important during an emergency was mentioned frequently across questions. This included

Table 3: Frequency of Long Answer Codes

Altering Route (56)		Start/Stop (63)		Pulling Over (56)	
POIs	13 (23%)	Safety and emergencies	8 (13%)	Safety and emergencies	19 (34%)
Route	9 (16%)	Route	7 (11%)	Car status	8 (14%)
Intersections and turns	7 (13%)	Obstacles	6 (10%)	Surrounding environment	5 (9%))
Traffic	7 (13%)	Surrounding Environment	5 (8%)	Obstacles	5 (9%)
Current Location	6 (11%)	Entry and exit	5 (8%)	Road status	4 (7%)
Distance	6 (11%)	Current location	5 (8%)	POIs	3 (5%)
Weather	5 (9%)	Traffic	3 (5%)	Weather	3 (5%)

Following Distance (45)		Speed (63)		Maneuvering (37)	
Distance	7 (16%)	Current speed and limit	13 (21%)	Obstacles	12 (32%)
Speed	7 (16%)	Road condition	9 (14%)	Route	10 (27%)
Other vehicle behavior	6 (13%)	Obstacles	8 (13%)	Current location	3 (8%)
Traffic	5 (11%)	Weather	7 (11%)	Traffic	3 (8%)
Weather	5 (11%)	Pedestrians	6 (10%)	Road type	2 (5%)
Obstacles	5 (11%)	Route	5 (8%)	Other vehicles	2 (5%)
Road conditions	5 (11%)	Other vehicles	5 (8%)	POIs	1 (3%)

any sort of malfunction in the vehicle, ways to enter and exit the vehicle safely, where it might be safe to stop the vehicle, and notifications regarding approaching or nearby emergency vehicles. Table 3 summarizes the frequency of codes identified from these long answer questions sorted by control type. The total number of code instances are reported per question type, along with the seven most frequent codes per question (as well as the percentage of codes within that question). Frequency reduced to five instances or fewer beyond this threshold across control type.

Germane to the later interface study, and in subjective support of H4, participants consistently noted wanting more information about POIs, such as nearby businesses, landmarks, or other places to visit. *Route-based information*, including intersections or what roads were nearby, was also mentioned frequently. These results informed the design of the experimental tasks in the subsequent studies.

4 STUDY 2, MID-AIR GESTURAL IDENTIFICATION

4.1 Motivation for Mid-Air Gestures

As discussed in 2.2, the non-visually dependent nature of mid-air gestural interaction and its applicability to the driving context would seem to suggest promise for use in FAVs, particularly among people who are visually impaired. Related research has also identified the need for new multimodal mechanisms to promote accessibility and control in FAVs among this demographic. As such, our goal in this second study session was to first identify a user-driven set of gestures for FAV control as performed by BVI participants (n=15). We also sought to understand what sensory modalities would best support gestural control and to what extent this type of navigation as desirable among BVI people. The resulting set of gestures and multimodal components were used in the subsequent interface test (Section 5).

4.2 Methods

Study 2 involved participants performing gestures for driving actions from the control type categories used in Study 1. Table 4 summarizes these driving actions.

Table 4: Study 2 Actions

(1) Start the Route	(5) Go Straight	(9) Turn Around	(11) Altering Route
(2) Stop the Route	(6) Turn Left	(10) Following Distance	(a) Locate New Route
(3) Speed Up	(7) Turn Right	(a) Closer	(b) Receive More Information
(4) Slow Down	(8) Pull Over	(b) Further Away	(c) Confirm the Route

Given our hypotheses from Study 1 and the supporting results, three subcategories for *altering the route* (locate a new route, receive more information, and confirm the new route) were included to identify gestures that could elicit the situational awareness information necessary to undertake route changing in FAVs.

4.2.1 Participants. The first 15 participants from the Study 1 group participated in Study 2. As such, these participants were also recruited by the Carroll Center for the Blind. These participants again represented a broad spectrum of vision loss, onset, etiology (specific visual demographics can be found in Table 1) and age, ranging from 28-70 (M = 55.53, SD = 13.88). Of these, eight identified as former drivers and seven identified as having never driven before. Participants predominantly identified as white or of European descent without identifying as ethnically Hispanic or Latino/x (86.67%). Two participants chose not to indicate racial or ethnic identify. 26.67% of participants had attained a Bachelor's degree, 26.67% a Master's degree, 20% some college but no degree, 6.67% High school or equivalent, and 6.67% a Ph.D. One participant chose not to report educational attainment.

4.2.2 Procedure. The experimental procedure, as with Study 1, began with participants being told, "Imagine riding in a fully autonomous vehicle that can safely, efficiently, and legally automate the trip." In this scenario, participants were told that gestures could give commands to control the vehicle. The experimenter clarified that these gestures were performed in mid-air, not on a device like a touchscreen. Participants were deliberately not given an example of what a gesture might look like, as the goal was to elicit whatever felt most intuitive (types of movement, one hand vs. two hand, etc.). After clarifying with participants that they understood their task, the experimenter started the video camera and read the first driving action. Once the participant performed the first action, the experimenter ended the recording and began the next, followed by reading the second driving action. This process repeated until all 14 driving actions were recorded.

The only modification to this procedure involved the three steps for *altering the route*. First, participants were asked to perform a gesture for locating a new route, but not knowing where it was. Then participants were asked to perform a gesture for receiving more information about the new route. And finally, participants were asked to imagine having received that information and to perform a gesture to confirm the new route. Gestures were recorded using a GoPro video camera. Video analysis was undertaken using the GoPro Player video software. Video analysis involved the gestural recordings being scored along four dimensions: movement (yes/no), type of movement (e.g., forward movement of the arm), hand position (e.g., pointed finger or open palm), and repetition (yes/no).

After performing the gestures, participants answered a brief post-test where they were asked what types of information should complement a gestural navigation system: haptic (active touch), audio (e.g., voice), or combinations of audio and haptic. Participants were also asked to what extent they agreed with the statement "I would want a hand gesture navigation system" from strongly disagree to strongly agree. The post-test survey was developed and delivered using Qualtrics.

4.2.3 Hypotheses. Prior to this experiment, we piloted a subset of gestures with 10 sighted users. From the 43 gestural videos we collected in the pilot, it was clear that people were inclined to incorporate movement and directionality in their gestures. This was also in-line with Qian et al.'s (2020) finding that people preferred dynamic opposed to static gestures. As such, our first hypothesis for this study was the following:

> H1: Gestures will prioritize the use of motion and directionality opposed to being statically performed (i.e., held in one position).

In the pilot we also observed that people tended to rely on driving metaphors for their gestures. For example, turning behavior was represented several times by the manipulation of an invisible steering wheel. As such, our second hypothesis was:

> H2: BVI people with prior driving experience will utilize gestures similar to in-vehicle elements (e.g., steering wheel or pedal manipulation) more so than people without prior driving experience.

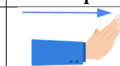
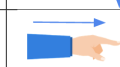
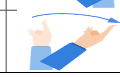
From the related research reviewed here that suggests the importance of multimodal feedback for BVI people in autonomous vehicles, our third hypothesis was:

> H3: Combinations of audio and haptic cues will be more desirable than audio or haptic alone.

4.3 Results

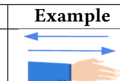
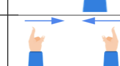
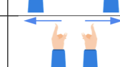
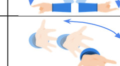
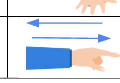
Of the 210 recorded gestures, 206 included significant hand or arm movement deemed important to the meaning of the gesture during video analysis. Three of the four gestures that were held statically were performed by a single participant, suggesting that some people may prefer motionless gestures. However, the finding that 98% of the gestures involved dynamic hand movement is strongly supportive of H1, which predicted that gestures would include motion and directionality.

Table 5: Most frequently used gestures in Study 2

Driving Action	Movement	Handshape	Repetition	Example
Start	Forward 13 (87%)	Open up 7 (47%)	No 12 (80%)	
Stop	Up 6 (40%)	Open up 10 (67%)	No 15 (100%)	
Speed Up	Rotational 5 (33%)	Finger point 5 (33%)	Yes 10 (67%)	
Slow Down	Down 6 (40%)	Open down 7 (47%)	Yes 9 (60%)	
Go Straight	Forward 14 (93%)	Finger point 6 (40%)	No 14 (93%)	
Turn Left	Left 15 (100%)	Finger point 9 (60%)	No 14 (93%)	
Turn Right	Right 15 (100%)	Finger point 8 (53%)	No 14 (93%)	

Tables 5 and 6 summarize the most frequently used types of movement and handshape for participants' gestures across the driving actions used in this study. Gestures utilized a variety of movement types (e.g., forward, up, directional left/right) and handshapes (e.g., open palm up, pointed finger). All but *following distance further* and *following distance closer* utilized one hand opposed to two. Each gesture reported in the tables is unique (note that*speed up* typically involved a participant's arm being held horizontally across the

Table 6: Most frequently used gestures in Study 2 (contd.)

Driving Action	Movement	Handshape	Repetition	Example
Pull Over	Left or Right 13 (87%)	Open side 5 (33%)	No 11 (73%)	
Turn Around	Rotational 14 (93%)	Finger point 8 (53%)	No 10 (67%)	
Closer	Together 8 (53%)	Finger point 4 (27%)	No 12 (80%)	
Further	Apart 10 (67%)	Finger point 4 (27%)	No 12 (80%)	
Locate	Arc 9 (60%)	Finger point 9 (60%)	No 13 (87%)	
Select/More	Compound 11 (73%)	Open palm 5 (33%)	No 14 (93%)	
Confirm	Left or Right 7 (47%)	Finger point 6 (40%)	No 15 (100%)	

body, whereas *turn around* involved a participant's arm being held vertically, perpendicular to the floor). The tables also provide the extent to which repetition was used in each gesture. Interestingly, *speed up* and *slow down* were the only driving actions for which the majority of participants (67% and 60% respectively) utilized continuous or repetitive gestures. This is logical given that these commands, more so than others, beg the question, "how much?".

Contrary to our expectations, only one participant related gestures to traditional driving controls (i.e., a steering wheel) and this participant did not report prior driving experience. Therefore, results from this gestural task did not support H2.

In addition to performing gestures, participants completed a post-test where they were asked what information modalities should be used in a gestural system. 14 (93%) noted wanting combinations of auditory and haptic information to complement gestural navigation, which supported H3. 11 (73%) indicated that they either agreed or strongly agreed with wanting a hand-gesture navigation system. These results, combined with Study 1 results with regard to the importance of *altering the route*, informed the design of the interface and scenario used in Study 3.

5 STUDY 3, INTERFACE TEST

5.1 Motivation for Gestural-Audio and Interface Test

Given the strong desire for gestural navigation supported by both audio and haptics from Study 2, our resulting interface combines what we refer to here as gestural-audio with haptic feedback. The gestural-audio component utilizes a ring of speakers that users can elicit through gestures to hear spatialized information. This design decision has strong support from the BVI and nonvisual navigation literature, where spatial delivery of auditory information has been shown to increase environmental learning and spatial memory by up to 50 percent [16] and significantly reduce cognitive load compared to using nonspatialized (traditional) auditory descriptions [22]. The haptic component of the interface relies on ultrasonic haptic feedback, which has gained traction in recent years for use in the automotive domain [19, 36]. Haptic exploration, like vision, presents advantages in terms of the relative ease at which information can be conveyed with spatial properties, such as lines, contours, and map elements [13]. As such, the goal of Study 3 was to test the feasibility of this experimental interface using the driving task Study 1 identified as most important: *altering the route*. Of interest was testing if users could receive adequate information from the interface to engage in a route alteration task using the inclusively designed gesture set identified by Study 2. We also sought to compare performance, measured by task completion time, between two conditions: Gestural-audio only and Gestural-audio with haptic feedback, as well as which condition users preferred.

5.2 Methods

The gestural-audio interface (Figure 3) was built using 3" Kicker motorcycle-style speakers mounted on a TrakRacer TR160 racing simulator. The TR160 is designed out of slotted extrusion rails, providing an easy way to mount devices in a modular manner around a vehicle seat. Seven channels of audio are generated through an Alcorn Mcbride RideAmp-25H Dante amplifier, and fed to the speakers arranged at the clock face positions around the user (9 o'clock, 10 o'clock, 11 o'clock, 12 o'clock, 1 o'clock, 2 o'clock, and 3 o'clock). Clock face positions were chosen throughout this design given the frequency of use in training for navigation among BVI people. The audio files were created using the AI voice generator Voicemaker.

Figure 3: The Gestural-Audio interface used in Study 3

Audio was delivered using SoundPlant, a software package that can map audio files to specific key strokes on a keyboard. The study as a whole utilized a Wizard-of-Oz methodology whereby the experimenter triggered the audio cues as opposed to being triggered via computer vision. Using results from the first two studies, we designed the audio such that it could enable route changing at an intersection. The route changing process, and its related audio was triggered via the following gestures identified from Study 2:

(1) A sweeping gesture, used to locate streets outside the car (see Locate in Table 6)
(2) A selection gesture, used to get more information about what was on a particular street (see Select/More in Table 6)
(3) A confirmation gesture, used to navigate the car in that direction (see Confirm in Table 6).

Using the sweeping/scanning gesture, participants could trigger audio clips based on the direction of their pointed finger. For example, at an intersection of 12 o'clock and 3 o'clock, participants would hear "Right turn, 3 o'clock" from the 90 degree azimuth at 3 o'clock as their finger passed the speaker at 3 o'clock. Using the selection gesture, participants could trigger audio for more information about that road. Using the previous example, participants could perform the selection gesture at the 3 o'clock speaker and hear "Main Street, there's a coffee shop nearby" from that location. Finally, participants could perform the confirmation gesture in the direction of the speaker with the coffee shop and were told by the experimenter that they successfully navigated the car in that direction.

The intersection was also conveyed using haptic representations delivered via a promising haptic modality emerging in related research: mid-air ultrasonic haptics.

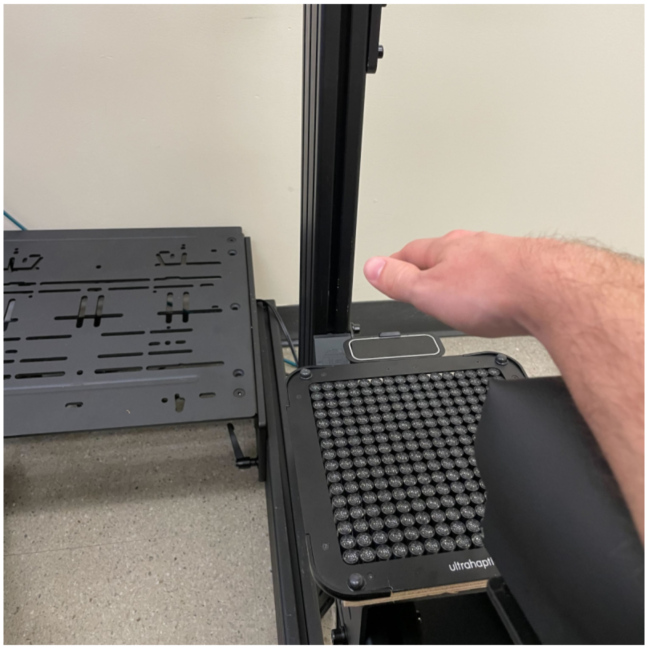


Figure 4: The UltraHaptics mid-air haptic device shown with arm rest and approximate location of the hand during the study.

The UltraHaptics (UH) is a device capable of creating complex, mid-air, haptic sensations using standing waves generated by an array of ultrasonic transducers. We used the UH in our user study as a means of generating haptic sensations to represent abstractions of a street intersection. Pilot testing revealed that hovering the hand over the device for extended periods became tiring, so a rest was designed where the user could rest their forearm during use. The UH is depicted in Figure 4.

Through significant pilot testing with one of the authors on this paper who is congenitally blind, as well as several blindfolded sighted users, we identified a haptic intersection representation that was perceptually salient. When users hover their hand over the device, it utilizes a series of pulses and "drawn" vibrating lines to represent the intersection and roads. First, a user feels pulses in the center of their palm to indicate the number of roads in the intersection. Then, a line is drawn from the palm in the direction of the first road (e.g., towards the thumb of their right hand for a road at 9'oclock). Pulses at the end of the line indicate that the line is done being drawn before the next line is drawn (again from the palm-out). This process repeats until all lines in the intersection are drawn in clockwise fashion and then repeats the sequence. The following summarizes this sequence:

(1) Pulse n-times in the center of the palm indicating the number of roads.
(2) Draw a line from the center towards the direction of the clock face position.
(3) Pulse two times at the end of the line.
(4) Repeat 1-3 until participant responds.

5.2.1 Participants. Eight participants who completed the Study 1 survey participated in Study 3. No participant from the Study 2 group participated in Study 3, as participants were recruited from the same facility in the prior studies for a one hour study (Study 2) or a two hour study (Study 3). Participants again represented a broad spectrum of vision loss, onset, etiology (specific visual demographics can be found in Table 1) and age, ranging from 31-59 (M = 41, SD = 9.78). Of these, five identified as former drivers and three identified as having never driven before. No participants reported any known tactile sensitivity loss. 50% of participants identified as white or of European descent. 25% identified as black or African American and 62.50% reported as not identifying as ethnically Hispanic or Latino/x, while 12.50% of participants did. Two participants chose not to indicate racial or ethnic identify. 37.50% of participants had attained a Bachelor's degree, 25% some college but no degree, 12.50% a Master's degree, and 12.50% an associate's degree. One participant chose not to report educational attainment.

5.2.2 Procedure. The experimental procedure began with participants being asked to imagine riding in a fully autonomous vehicle that could take them where they needed to go safely, efficiently, and legally. Given results from Study 1 in terms of important information for *altering the route* (see Table 3: POIs, intersections and turns) the scenario used in this test involved participants imagining being stopped at an intersection with the goal of changing the route to a nearby coffee shop. Participants were told that they would experience several intersections throughout the study, each with two, three, or four roads extending from their position. They were also told that they would experience the intersections through combinations of gestural-audio and haptic feedback.

Gestural-audio Examples and Practice: First, participants were exposed to two examples of representative intersections using gestural-audio. At the beginning of each intersection presentation, audio played from the speaker mounted at 12 o'clock to indicate that there were two roads, three roads, or four roads. Participants were then told that they could use a series of gestures to receive more information about where the roads were, what was along each, and to direct the vehicle's route. Only one road in each intersection had a coffee shop nearby (the others containing either a flower shop, houses, or a gas station nearby) and participants were instructed that their goal throughout the experiment was to navigate to the street with the coffee shop.

Participants were then instructed how to use the three gestures utilized throughout the experiment (provided in Section 5.2). The experimenter verified that participants understood each gesture and confirmed that each was performed accurately during this example phase.

After completing the examples, participants were exposed to two test intersections to determine that they could independently use the three gestures to interpret the intersection and navigate the vehicle to the coffee shop. All participants successfully completed the two tests without error.

Haptic Examples and Practice: Participants were then exposed to the Ultrahaptic device and told that it was able to project the number of roads and shape of the intersection onto their hand. As in the gestural-audio phase, participants were given two examples of intersections followed by two tests to determine if they could understand the intersection. In these tests, participants were instructed to say aloud how many roads there were (indicated by the pulses at the beginning of the haptic sequence) and their related clock face positions (indicated by the lines drawn from their palms). Six of the eight participants successfully passed these competency tests on their first try. The remaining two were given three tries of repeated examples of the intersections, but were unable to determine the position of the roads. Given that the experimenter verified that the device was working, and that the participants did not report any known tactile sensation loss, the reason for these failures may have been due to the learning curve associated with haptic navigation or the signal intensity from the device. Regardless of the reason, these participants were directed to the post-test and did not complete the remainder of the study.

Experimental Conditions: After the examples, practice, and competency tests with the two modalities, six participants began either the gestural-audio only condition or the gestural-audio and haptic condition. The ordering of these conditions were counterbalanced between participants to avoid any ordering or learning effects.

In both conditions, the participant's goal was to navigate to the coffee shop as quickly as possible while still being accurate. For the gestural-audio only condition, participants heard the number of roads and began scanning using the three gestures learned in the practice phase. In the gestural-audio + haptic condition, participants felt the number of roads and intersection geometry prior to scanning the intersection using the same gestures. The experimenter used a stopwatch to measure the time it took to successfully navigate to the coffee shop. In the haptic condition, the experimenter measured both the total time it took to navigate to the coffee shop and the time spent learning the haptic intersection. The stimulus set for both conditions included 6 intersections: two two-road intersections, two three-road intersections, and two four-road intersections with the ordering of these stimuli randomized between participants and condition to avoid any ordering or learning effects. The position of the coffee shop was different in each condition and was balanced across clock face positions. After completing both conditions, participants completed a brief post-test interview with the experimenter to assess ease of use, preference for interface, and to collect qualitative feedback on improvements to be made. Participants were also asked to assess to what extent they agreed with the statement: "I would want a mid-air gestural navigation system" on a 5-point Likert scale from 1-Strongly Disagree to 5-Strongly agree.

5.2.3 *Hypotheses.*

Given the related research and our results from Study 2 showing preference for combinations of audio and haptic information to support gestural navigation, our hypotheses for Study 3 were the following:

> H1: People will navigate faster in the audio + haptic condition (after experiencing the haptic representation) than in the audio only condition.

We predicted that combining haptics with the audio system would improve performance by providing redundant spatial cues as to the geometry of the intersection.

> H2: People will prefer the gestural audio + haptic condition opposed to the audio only condition

5.3 Results

Importantly, all six participants successfully navigated using the gestural-audio interface. Although the mean navigation time for the gestural-audio + haptic condition (13.16s) was slightly faster than the gestural-audio only condition (14.09s), a paired samples t-test demonstrated that this numeric difference was not statistically significant (p=.48). Additionally, a Bayesian paired samples t-test performed in JASP [20] resulted in a Bayes factor of .226, suggesting moderate evidence in favor of the null hypothesis for H1. The six participants who completed both conditions rated their preferred interface, with four preferring the gestural audio only and two preferring the gestural-audio and haptic interface. This finding did not support H2.

Figure 5 reports ease of use for both interfaces from 1-Very Difficult to 5-Very Easy. In general, participants rated both interfaces easy to use with all responses but one rated at a 3 or higher. This was particularly the case for the gestural-audio only condition, where five participants rated the interface as 5-very easy to use and one participant rated the interface as 4-easy to use.

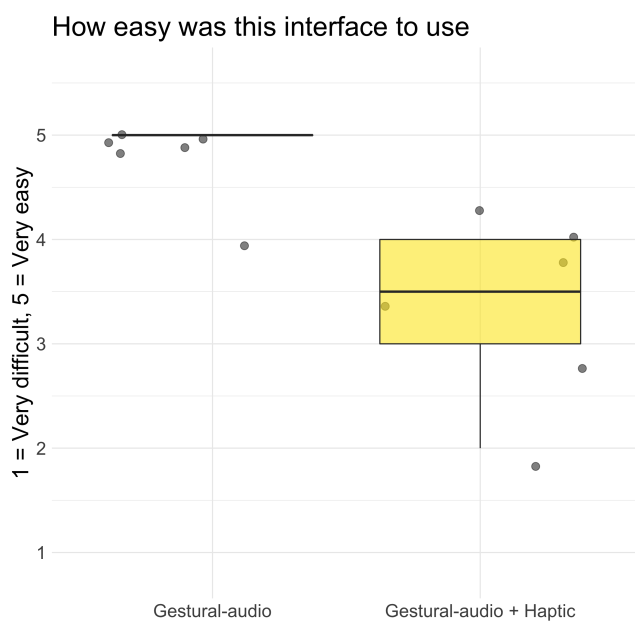


Figure 5: Ease of use

For subjective input on improving the haptic portion of the interface, every participant but one mentioned wanting stronger haptic feedback from the ultrahaptic device. Furthermore, participants noted wanting a frame of reference for where their hand should be or some cue to indicate the optimal position. This may have contributed to why some of our participants failed the competency test for the haptic condition. For improving the audio portion of the interface, participants mentioned wanting to be able to customize several features including the speech rate, volume, units of measure (degrees vs. clock face positions), and the voice itself.

In the post-test for Studies 2 and 3, participants were asked to rate the extent to which they agreed with the statement, "I would want a hand-gesture navigation system." Figure 6 displays the results for both groups, which indicate positive support for gestural navigation. As reported in the Study 2 results, eleven (73%) in the Study 2 group indicated a four (agree) or five (strongly agree), whereas 7 (88%) in the Study 3 group responded agree or strongly agree. Although the unequal sample sizes make a statistical comparison inappropriate, these descriptive results are encouraging considering that the Study 3 group experienced a more realistic scenario using gestural navigation.

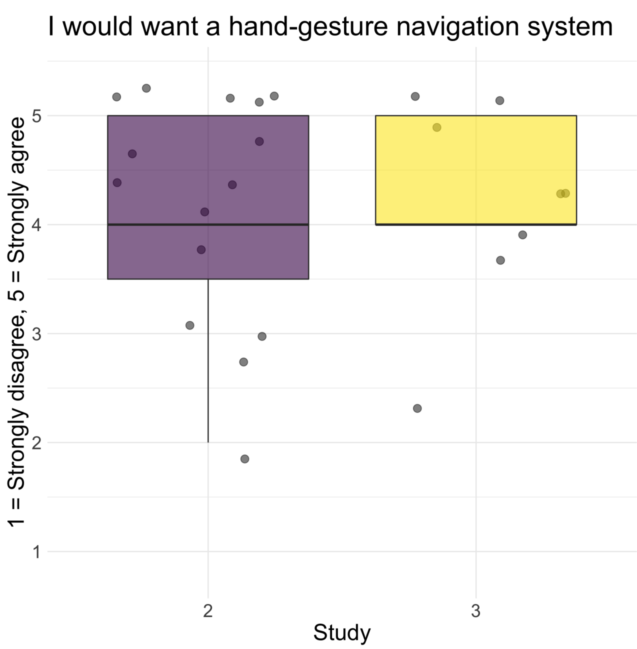


Figure 6: Participant's response to "I would want a hand-gesture navigation system" broken up by study

6 DISCUSSION

This research was motivated by the need for new accessible interfaces and interaction modalities to support autonomous transportation among people with visual impairments. Study 1 results demonstrate the importance of enabling vehicle control for this demographic, as well as the information required for human-in-the-loop control during fully autonomous vehicle (FAV) travel. In Study 2, we enumerate the first inclusively designed and accessible gesture set for FAV control, with results informing the development of a novel gestural-audio interface. The resulting experimental interface developed for and used in Study 3 promotes situational awareness and usability for altering the route, demonstrating strong support from users in terms of ease of use and desires to use. Together, this research represents a critical step towards fully accessible FAV user experiences fundamental to our inclusive transportation future.

6.1 Importance of Human-in-the-loop Control

Despite being termed "fully" autonomous vehicles, Study 1 results suggest that BVI people desire to be in the loop of FAV vehicle control across common driving tasks. At a high level, these findings are in line with existing accessibility research in this domain [4], but provide additional granularity as to the situations in which control over driving behavior itself is important (of note, *altering a route* and deciding when the vehicle *starts and stops* its journey). By providing specific information types and driving tasks for which control is desirable, the results of this research elucidate specific connections between situational awareness and actionable behavior. That is, when information is provided to convey the situational awareness necessary to undertake control, and complemented with HMIs harnessing multisensory input/output functions, as we do here, the results indicate this demographic is able to independently operate FAVs with high ease of use.

It is worth noting that this desire for control extends beyond the BVI demographic and may well be true for all users. For instance, 92% of participants in an automated vehicle demonstration ride noted wanting shared control [31]. As designers and industry stakeholders develop the next generation of full autonomy, development efforts can (and should) be cognizant of user desires for input into the vehicle's decision space.

6.2 Advantages of Gestural-Audio

Audio interfaces are a common approach in navigation systems for providing navigational cues and spatial information. From in-vehicle GPS to accessible indoor systems (see [15] for review), audio navigation is a useful and natural approach for many users. Indeed, during Study 2, many participants naturally complemented their gestures with voice directions. For example, when pointing left, participants often also said aloud "Go left." This tendency can be capitalized on in future work to compliment navigation systems, not only in terms of usability, but also precision. As such, we argue that the gestural-audio system presented and validated here affords benefits compared to interfaces relying on voice input alone.

Although audio output is particularly applicable across environment and scenario, voice as a system input is not desirable in every scenario. Voice input can be imprecise, inaccurate in noisy areas, and can cause concerns for users related to privacy or in shared spaces [13, 15, 17]. Consider, for example, not wanting to wake a fellow passenger in a vehicle. Indeed, this need for additional inputs beyond voice is a significant advantage of our resulting multisensory gestural-audio interface. Not only can gestural-audio provide an alternative during situations where voice input is undesirable, but it could also reduce imprecision when performed in tandem with voice commands, as many participants naturally did.

Another major advantage of incorporating gestures in a navigation system is the applicability to people who are hard of hearing, deaf, or blind/deaf. Although not the focus of this work, several participants in Study 3 thoughtfully mentioned how our system should be explored and extended to promote inclusion among the blind/deaf and across sensory impairments. Future incarnations of in-vehicle gestural-audio should also explore visual cues and enhanced visual cues/compensatory augmentations to further support sighted people and people with moderate visual impairment with significant residual vision. Indeed, gestural-audio could be well suited for all users to receive spatially salient information about the driving environment. By coupling in-vehicle UIs with onboard mapping and data software, future work could enable users to gesture towards any object in the environment (e.g., landmarks, other vehicles, signage) for more information and to undertake potential control actions.

One limitation of our approach is that the array of gestural-audio speakers currently encompasses 180 degrees opposed to a full cabin or 360 degree implementation. Given that user orientation in FAVs may not be fixed (consider that seat belts and typical vehicle seating arrangements may become obsolete), a truly spatialized, full cabin implementation would be most practical. Future work will involve computer vision recognition of user gestures without orientation limitations, opposed to the Wizard-of-Oz methodology used here.

6.3 Support for Haptics

Results from Study 2 demonstrate that BVI users desire combinations of audio and haptics in FAVs. Although performance did not improve using haptic cues in Study 3, we conclude this is not an indictment of the modality itself, but speaks to its specific and somewhat limited implementation in this work (see limitations section below). This is also likely an issue stemming from lack of exposure to haptic interfaces. People are generally accustomed to auditory UIs and navigation systems but have little experience with haptic UIs, despite promising results for spatial learning and behavior [14, 28]. Future work is necessary to explore how haptic cues can be successfully implemented in FAVs to leverage the intrinsic spatial advantages of this non-visual modality.

7 LIMITATIONS

The limitations of this work primarily concern the somewhat simplistic nature of Study 3. First, as mentioned, the Wizard-of-Oz approach utilized could be extended in future work to include computer vision recognition of gestures. While we argue that mid-air gestural interaction is ripe for accessible multisensory interaction coupled with audio and/or haptics, more investigation is necessary to explore the ideal sensory combinations and devices beyond gestural-audio with haptics and gestural-audio without haptics, as studied here. The task in this study was also simple in nature, with users only completing a predefined action (i.e., find the coffee shop) opposed to the vast state space of decisions that people will likely want to make in FAVs, as supported by our Study 1 results. These findings may also generalize beyond the BVI demographic. As such, future work should explore how all users, BVI or otherwise, can utilize gestural-audio across more complex driving actions and demands "in the wild" opposed to the controlled environment used in this research.

8 CONCLUSION

Fully autonomous vehicles hold enormous potential to transform the lives of people with disabilities by fundamentally increasing independence, mobility, and personal freedom. This project, consisting of three experiments with people who are blind or low vision (n = 23), explored the user-driven design of an accessible, non-visually dependent, human machine interface. Results indicate that gestural-audio holds the potential to enable people who are blind or low vision to independently operate fully autonomous vehicles. The interface test also provides compelling evidence for conveying situational awareness and increasing control across the spectrum of vision loss, with strong implications for all people during situations of reduced visibility (e.g., at night) or limited information access (e.g., in unfamiliar environments).

ACKNOWLEDGMENTS

Toyota Research Institute provided funds to support this work. We acknowledge support from NSF grant CHS-1910603 on this project and would like to thank our participants and The Carroll Center for the Blind for assistance with participant recruitment. Special thanks to Bruce Howell, Dina Rosenbaum, Nick Corbett, and Howard Sumner for their support and insight.

REFERENCES

[1] Roger Bennett, Rohini Vijaygopal, and Rita Kottasz. 2020. Willingness of people who are blind to accept autonomous vehicles: An empirical investigation. *Transportation Research Part F: Traffic Psychology and Behaviour* 69 (Feb. 2020), 13–27. https://doi.org/10.1016/j.trf.2019.12.012

[2] Rupert RA Bourne, Jaimie Adelson, Seth Flaxman, Paul Briant, Michele Bottone, Theo Vos, Kovin Naidoo, Tasanee Braithwaite, Maria Cicinelli, Jost Jonas, et al. 2020. Global Prevalence of Blindness and Distance and Near Vision Impairment in 2020: progress towards the Vision 2020 targets and what the future holds. *Investigative Ophthalmology & Visual Science* 61, 7 (2020), 2317–2317.

[3] Robin Brewer and Nicole Ellison. 2020. *Supporting People with Vision Impairments in Automated Vehicles: Challenge and Opportunities.* Technical Report. University of Michigan, Ann Arbor, Transportation Research Institute.

[4] Robin N Brewer and Vaishnav Kameswaran. 2018. Understanding the power of control in autonomous vehicles for people with vision impairment. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility.* 185–197.

[5] Julian Brinkley. 2021. Using personas with visual impairments to explore the design of an accessible self-driving vehicle human-machine interface. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 65. SAGE Publications Sage CA: Los Angeles, CA, 337–341.

[6] Julian Brinkley, Earl W Huff Jr, Briana Posadas, Julia Woodward, Shaundra B Daily, and Juan E Gilbert. 2020. Exploring the needs, preferences, and concerns of persons with visual impairments regarding autonomous vehicles. *ACM Transactions on Accessible Computing (TACCESS)* 13, 1 (2020), 1–34.

[7] Julian Brinkley, Brianna Posadas, Imani Sherman, Shaundra B Daily, and Juan E Gilbert. 2019. An open road evaluation of a self-driving vehicle human–machine interface designed for visually impaired users. *International Journal of Human–Computer Interaction* 35, 11 (2019), 1018–1032.

[8] Eddie Brown, David R. Large, Hannah Limerick, and Gary Burnett. 2020. Ultrahapticons:"Haptifying" drivers' mental models to transform automotive mid-air haptic gesture infotainment interfaces. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 54–57.

[9] Henry Claypool, Amitai Bin-Nun, and Jeffrey Gerlach. 2017. Self-Driving Cars: The Impact on People with Disabilities. *Ruderman Foundation* (2017), 35.

[10] Henrik Detjen, Stefan Geisler, and Stefan Schneegass. 2020. Maneuver-based control interventions during automated driving: Comparing touch, voice, and mid-air gestures as input modalities. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC).* IEEE, 3268–3274.

[11] Brad E. Dicianno, Sivashankar Sivakanthan, S. Andrea Sundaram, Shantanu Satpute, Hailee Kulich, Elizabeth Powers, Nikitha Deepak, Rebecca Russell, Rosemarie Cooper, and Rory A. Cooper. 2021. Systematic review: Automated vehicles and services for people with disabilities. *Neuroscience Letters* 761 (Sept. 2021), 136103. https://doi.org/10.1016/j.neulet.2021.136103

[12] Paul DS Fink, Jessica A Holz, and Nicholas A Giudice. 2021. Fully autonomous vehicles for people with visual impairment: Policy, accessibility, and future directions. *ACM Transactions on Accessible Computing (TACCESS)* 14, 3 (2021), 1–17.

[13] Nicholas A. Giudice. 2018. Navigating without vision: principles of blind spatial cognition. *Handbook of Behavioral and Cognitive Geography* (April 2018), 260–288. https://www.elgaronline.com/view/edcoll/9781784717537/9781784717537.00024.xml ISBN: 9781784717544 Publisher: Edward Elgar Publishing Section: Handbook of Behavioral and Cognitive Geography.

[14] Nicholas A. Giudice, Benjamin A. Guenther, Nicholas A. Jensen, and Kaitlyn N. Haase. 2020. Cognitive Mapping Without Vision: Comparing Wayfinding Performance After Learning From Digital Touchscreen-Based Multimodal Maps vs. Embossed Tactile Overlays. *Frontiers in Human Neuroscience* 14 (2020). https://doi.org/10.3389/fnhum.2020.00087

[15] Nicholas A. Giudice, Benjamin A. Guenther, Toni M. Kaplan, Shane M. Anderson, Robert J. Knuesel, and Joseph F. Cioffi. 2020. Use of an Indoor Navigation System by Sighted and Blind Travelers: Performance Similarities across Visual Status and Age. *ACM Trans. Access. Comput.* 13, 3, Article 11 (aug 2020), 27 pages. https://doi.org/10.1145/3407191

[16] Nicholas A. Giudice and Jerome D. Tietz. 2008. Learning with Virtual Verbal Displays: Effects of Interface Fidelity on Cognitive Map Development. In *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, Christian Freksa, Nora S. Newcombe, Peter Gärdenfors, and Stefan Wölfl (Eds.). Vol. 5248. Springer Berlin Heidelberg, Berlin, Heidelberg, 121–137. https://doi.org/10.1007/978-3-540-87601-4_11 ISSN: 0302-9743, 1611-3349 Series Title: Lecture Notes in Computer Science.

[17] Nicholas A. Giudice, William E. Whalen, Timothy H. Riehle, Shane M. Anderson, and Stacy A. Doore. 2019. Evaluation of an Accessible, Real-Time, and Infrastructure-Free Indoor Navigation System by Users Who Are Blind in the Mall of America. *Journal of Visual Impairment & Blindness* 113, 2 (March 2019), 140–155. https://doi.org/10.1177/0145482X19840918 Publisher: SAGE Publications Inc.

[18] Mikaylah Gross, Joe Dara, Christopher Meyer, and Davide Bolchini. 2018. Exploring Aural Navigation by Screenless Access. In *Proceedings of the 15th International Web for All Conference (W4A '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3192714.3192815

[19] Kyle Harrington, David R. Large, Gary Burnett, and Orestis Georgiou. 2018. Exploring the Use of Mid-Air Ultrasonic Feedback to Enhance Automotive User Interfaces. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON, Canada) *(AutomotiveUI '18)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/3239060.3239089

[20] JASP Team. 2022. JASP (Version 0.16.3)[Computer software]. https://jasp-stats.org/

[21] Kibum Kim, Xiangshi Ren, Seungmoon Choi, and Hong Z Tan. 2016. Assisting people with visual impairments in aiming at a target on a large wall-mounted display. *International Journal of Human-Computer Studies* 86 (2016), 109–120.

[22] Roberta L. Klatzky, James R. Marston, Nicholas A. Giudice, Reginald G. Golledge, and Jack M. Loomis. 2006. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of Experimental Psychology. Applied* 12, 4 (Dec. 2006), 223–232. https://doi.org/10.1037/1076-898X.12.4.223

[23] David R Large, Kyle Harrington, Gary Burnett, and Orestis Georgiou. 2019. Feel the noise: Mid-air ultrasound haptics as a novel human-vehicle interaction paradigm. *Applied ergonomics* 81 (2019), 102909.

[24] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-Air Pan-and-Zoom on Wall-Sized Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 177–186. https://doi.org/10.1145/1978942.1978969

[25] World Health Organization. 2021. *Blindness and Visual Impairment.* Technical Report. Retrieved Feb 15, 2022.

[26] Sharon Oviatt, Rachel Coulston, and Rebecca Lunsford. 2004. When do we interact multimodally? Cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces.* 129–136.

[27] Hari P. Palani, Paul D. S. Fink, and Nicholas A. Giudice. 2020. Design Guidelines for Schematizing and Rendering Haptically Perceivable Graphical Elements on Touchscreen Devices. *International Journal of Human–Computer Interaction* (April 2020), 1–22. https://doi.org/10.1080/10447318.2020.1752464

[28] Hari Prasath Palani, Paul D. S. Fink, and Nicholas A. Giudice. 2022. Comparing Map Learning between Touchscreen-Based Visual and Haptic Displays: A Behavioral Evaluation with Blind and Sighted Users. *Multimodal Technologies and Interaction* 6, 1 (Jan. 2022), 1. https://doi.org/10.3390/mti6010001 Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[29] Xiaosong Qian, Wendy Ju, and David Michael Sirkin. 2020. Aladdin's magic carpet: Navigation by in-air static hand gesture in autonomous vehicles. *International Journal of Human–Computer Interaction* 36, 20 (2020), 1912–1927.

[30] Parivash Ranjbar, Pournami Krishnan Krishnakumari, Jonas Andersson, and Maria Klingegård. 2022. Vibrotactile guidance for trips with autonomous vehicles for persons with blindness, deafblindness, and deafness. *Transportation Research Interdisciplinary Perspectives* 15 (2022), 100630.

[31] SAE. 2019. *SAE Demo Days Survey.* SAE. https://www.sae.org/news/press-room/2019/11/sae-international-survey-shows-public-enthusiasm-about-self-driving-cars

[32] Gözel Shakeri, John H Williamson, and Stephen Brewster. 2018. May the force be with you: Ultrasound haptic feedback for mid-air gesture interaction in cars. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.* 1–10.

[33] Surak Son, YiNa Jeong, and Byungkwan Lee. 2019. An Audification and Visualization System (AVS) of an Autonomous Vehicle for Blind and Deaf People Based on Deep Learning. *Sensors* 19, 22 (Jan. 2019), 5035. https://doi.org/10.3390/s19225035 Number: 22 Publisher: Multidisciplinary Digital Publishing Institute.

[34] Peng Song, Wooi Boon Goh, William Hutama, Chi-Wing Fu, and Xiaopei Liu. 2012. A Handle Bar Metaphor for Virtual Object Manipulation with Mid-Air Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1297–1306. https://doi.org/10.1145/2207676.2208585

[35] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 841–850. https://doi.org/10.1145/2470654.2470774

[36] Gareth Young, Hamish Milne, Daniel Griffiths, Elliot Padfield, Robert Blenkinsopp, and Orestis Georgiou. 2020. Designing Mid-Air Haptic Gesture Controlled User Interfaces for Cars. *Proceedings of the ACM on Human-Computer Interaction* 4, EICS (June 2020), 81:1–81:23. https://doi.org/10.1145/3397869