

Research Article

Towards a Conceptual Model of Users' Expectations of an Autonomous In-Vehicle Multimodal Experience

Ecem Berfin Ince ¹, Kyungjoo Cha ¹ and Junghyun Cho²

¹Department of Design and Engineering, Bournemouth University, Bournemouth, UK

²Holistic UX Group, Automotive R&D Division, Hyundai Motors Group, Seoul, Republic of Korea

Correspondence should be addressed to Kyungjoo Cha; kcha@bournemouth.ac.uk

Received 23 August 2023; Revised 8 February 2024; Accepted 16 February 2024; Published 14 March 2024

Academic Editor: Tze Wei Liew

Copyright © 2024 Ecem Berfin Ince et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

People are expected to have more opportunities to spend their free time inside the vehicle with advanced vehicle automation in the near future. This will enable people to turn their attention to desirable activities other than driving and to have varied in-vehicle interactions through multimodal ways of conveying and receiving information. Previous studies on in-vehicle multimodal interactions primarily have focused on making users evaluate the impacts of particular multimodal integrations on them, which do not fully provide an overall understanding of user expectations of the multimodal experience in autonomous vehicles. The research was thus designed to fill the research gap by posing the key question “What are the critical aspects that differentiate and characterise in-vehicle multimodal experiences?” To answer this question, five sessions of design fiction workshops were separately conducted with 17 people to understand the users' expectations of the multimodal experience in autonomous vehicles. Twenty-two subthemes of users' expected tasks of multimodal experience were extracted through thematic analysis. The research found that two dimensions, *attention* and *duration*, are critical aspects that impact in-vehicle multimodal interactions. With this knowledge, a conceptual model of the users' in-vehicle multimodal experience was proposed with a two-dimensional spectrum, which populates four different layers: *sustained*, *distinct*, *concurrent*, and *coherent*. The proposed conceptual model could help designers understand and approach users' expectations more clearly, allowing them to make more informed decisions from the initial stages of the design process.

1. Introduction

People have had more free time inside the vehicle with the increment of vehicle autonomy [1, 2]. In a fully autonomous vehicle (AV), occupants are expected to not be required to keep their eyes on the road all the time [3]. Thus, occupants can fully turn their attention to other activities, meaning information no longer needs to be displayed in either one device or one fixed spot. This provides opportunities for vehicle occupants to experience new forms of in-vehicle spatial orientations and to have varied interactions. This would open up new possibilities for designers and researchers to enhance the users' experience of forthcoming in-vehicle interactions.

Designing for in-vehicle interaction is multifaceted, as occupants focus on a driving task and interactions with equipped infotainment systems, brought-in devices, external services, AI assistants, and social interactions [4]. The in-vehicle interactions have evolved with advanced technology that adapts to users' changing contexts, psychological states, preferences, and habits. The interactions can be enabled through different inputs and outputs, including visual and audio, voice and haptic with knobs, buttons, touch screens, and more [5]. To provide occupants with optimal modes, it draws increased attention needed on what they expect from multimodal interaction within AVs and what characteristics of the multimodal experience should be considered when designing overall interactions. This study, therefore,

is aimed at investigating users' expectations of the in-vehicle multimodal experience in future AVs.

2. Background and Relevant Work

2.1. In-Vehicle Multimodality. As a notion, "multimodality" is linked with "mode," which means "the particular way of accomplishing something" [6], and "modality," which means "those aspects of a thing which relate to its mode, or manner or state of being, as distinct from its substance or identity" [7]. Multimodality means having multiple modalities. In a system, multimodality is an exchange between device and human where multiple inputs or outputs may be used simultaneously or sequentially depending upon context and preference [8]. Such systems process two or more combined user input modes—such as speech, pen, touch, manual gestures, gaze, and head and body movements—in a coordinated manner with multimedia system output [9]. In the context of a vehicle, it may cover various interactions such as haptic, kinetic, speech, gesture, touch, virtual touch, gaze, eye gaze switch, and facial input, plus devices such as touch screen, the pedals, buttons, face cam, microphone, and multitouch displays [1, 10–12].

Recent research shows multimodal interactions improve the in-vehicle user experience [13]. Indeed, it has been found that multimodality inside a vehicle can play a crucial role in enhancing the user experience:

- (i) As an effective means of interaction and conveying information [14, 15] by giving users ways to interact with the system through the modality that is most appropriate for the present context of use [12]
- (ii) As an enabler for emotional design and particularly positive emotions through in-vehicle systems [16]
- (iii) As a means to increase learnability, as well as accommodate a broader range of users, tasks, and contexts [17]
- (iv) As a means to offer better flexibility and reliability and to offer interaction alternatives to better meet the needs of diverse users with a range of usage patterns and preferences [18]
- (v) As a way to achieve a seamless interaction closer to human-human communication [19, 20]
- (vi) As a way to increase the trust between the machine and the passengers [21]

With a growing number of multimodal technologies diversifying the role of vehicles, understanding what people desire from a multimodal self-driving experience has become crucial to meeting people's needs.

2.2. Understanding Users' Expectations of In-Vehicle Multimodal Experience. Vehicle users' expectations informs the early design process by helping designers rethink the occupant-vehicle relationship, identify the design requirements, and design future artefacts [22–24].

Previous studies have emphasised the expected activities, use cases, and scenarios in future AVs [25–29]. Pettersson

[30] proposed a model of users' expectations of AVs. Lee et al. [23] identified a comprehensive design taxonomy and design requirements for full AVs. Cha [25] developed AV scenarios by focusing on human emotions. In these studies, user-centred values were elicited by investigating what people would want and expect from future vehicles and breaking down the user experience and expectations.

Moreover, users' expectations have also been investigated, focusing on in-vehicle multimodal interactions. Politis et al. [31] evaluated multimodal displays with users under varying degrees of situational urgency. Väänänen-Vainio-Mattila et al. [12] explored the user experience and the expectations of haptic feedback in a vehicle. Multiple studies have investigated the usefulness of the combined feedback modalities such as visualisation, light, audio, text, haptic, or vibration for conveying information to the driver that supports a feeling of trust and safety [32, 33]. These studies mainly focused on evaluating or testing a particular multimodal prototype and its real-time impact on users [13, 15, 34, 35].

Additionally, there have been several attempts to map, model, or create a framework for multimodal interaction. For example, Nigay and Coutaz [36] classified multimodal interfaces in a 2×2 table (combined or independent, sequential or parallel). Obrenovic and Starčević [37] modelled the multimodal human-computer interaction. Lalanne et al. [38] proposed a model for multimodality integration. Serrano et al. [39] developed a framework as a tool for creating multimodal input interfaces. Similarly, some studies offer a methodological perspective. Di Mitri et al. [40] provided a conceptual model for multimodal learning analytics with its objectives. Chang and Bourguet [41] proposed a usability framework for designing and evaluating multimodal interactions. Likewise, Zaidi and Kirazci [42] suggested a single-axis multimodality spectrum for Google Assistant, and Platz [8] mapped the spectrum of multimodal interaction. Most studies focus on a technical perspective that categorises multimodal devices or services.

Given that multimodal interactions depend on contextual preferences [8, 43, 44] and it entails the ability to adapt to different environments and users [45], understanding the user contexts is critical to better shape the multimodal experiences. A "user context" refers to a broad range of contextual elements in a user's experience [46, 47], the context of use or activities [26–28], or situations [48]. The most commonly mentioned elements of the in-vehicle context include *actors (occupants)*, *systems (vehicles)*, and *environment* such as a driver, vehicle, environment, performance, system, passengers, applications, inner worlds of people, ambience, surrounding, the position of people and objects, and time of the day. Therefore, a working definition of user context for this study is "the particular moments caused by the vehicle, environment or occupants that would trigger occupants to request intuitive interactions in a future AV."

With this understanding of user contexts, the study is aimed at investigating users' overall expectations of the multimodal experience in an AV. Next, a conceptual model that categorises the user's in-vehicle multimodal experience, with two dimensions and four quadrants, is proposed and discussed thoroughly. The paper goes beyond evaluating

specific modalities and instead suggests a model that provides a holistic perspective on user interactions. This broader focus contributes to a more comprehensive understanding of user expectations concerning the multimodal experience in upcoming AVs. The model is expected to help designers make more informed decisions when designing future in-vehicle multimodal experiences by considering user insights, and it can be used as a basis for design guidelines for multimodal interactions.

3. Research Design

A qualitative approach was adopted to extract users' expectations about the in-vehicle multimodal experience. It was the most appropriate approach to gain detailed knowledge about participants' future desires and expectations [49, 50]. The paper consists of two linked studies—study 1 and study 2. Study 1, as the preliminary study [22], is aimed at exploring in-vehicle user contexts in which users might desire effortless interactions in an AV. Study 2 was designed based on the study 1 findings, which will be the focus of this paper.

Study 1: exploring in-vehicle user contexts for effortless interactions in an AV

An open-ended online survey with 150 people was conducted, and the qualitative responses were thematically analysed, which resulted in a *taxonomy of in-vehicle user contexts* [22]. This taxonomy, with six contexts, related sub-contexts, and their frequency (f) with which they appeared in the entire responses, shown in Table 1, created a baseline and provided guidance for study 2.

Study 2: understanding users' expectations of the in-vehicle multimodal experience

Study 2 was designed to understand what users expect of their future vehicle multimodal experience in the six of each in-vehicle user context (Table 1). Five separate design fiction workshops with a total of 17 participants were conducted.

3.1. Design Fiction Workshop. The design fiction sessions support effective engagement and creative expression [51, 52] and promote critical engagement by encouraging people to question current beliefs. Creating an environment that allows participants to engage, collaborate, and play gives them a more dynamic role than being an informant [30]. Hence, these collaborative workshops seemed to be a helpful approach in gathering people to interact with each other to explore their expectations about a particular experience.

However, involving potential users in idea generation may also have disadvantages. While these approaches could enable us to reveal opportunities for users [53], users may not know what they want, and it is hard for the general public to engage with future artefacts and events [54, 55]. Since using creative, playful, and artistic ideation methods and approaches can offer a solution for future imagining and engagement [56, 57], this study implemented future-oriented methods to eliminate the disadvantages of the aforementioned user engagement and future imagining issues. The workshops included the following:

- (i) *Brainstorming and group discussion*, a well-known technique for collaborative idea generation, were implemented in the workshops. The workshops included individual and group brainstorming sessions, which helped achieve both creative quality and quantity [58]
- (ii) *Science-fiction (sci-fi) scenario writing* as a part of a *Design Fiction* practice was implemented as a significant part of the workshop. Fiction as an artistic form of expression helped participants to imagine the future, envision, and share their thoughts on future technologies and creatively express themselves [56, 59]
- (iii) *Roleplaying* as a part of the drama technique [60] was also implemented in the workshops as a future envisioning, engaging, and prompting method, which helped extract the future expectations of the participants

3.1.1. Design Fiction Workshop: Participant Selection. Recruiting participants was a crucial part of the study because the workshop activity required creative skills for getting into the role, engaging with the future, and writing. Conducting such a creative workshop activity with the public can be challenging due to a lack of the mentioned skills among participants. Thus, a purposeful sampling strategy [61] was used to overcome the potential challenges. Considering that speculating with fiction writers was identified as an effective options [56] and considering practicalities, the following recruitment inclusion criteria were set: (1) (sci)fiction writer, either a professional or a young or amateur writer, (2) willing to participate and engage with other participants in a session, (3) being a vehicle user, (4) aged over 18, and (5) the ability to access online tools. Screening questions including demographic information were used to ensure a gender balance and to recruit participants that met the inclusion criteria. Online channels were used for the distribution of the research adverts such as social media platforms, research sites, and online discussion forums. Research posters were also posted in a variety of “research,” “writing,” “fiction,” and “sci-fi” relevant social media communities such as research-related Facebook groups and subreddits.

Table 2 presents participants' information about the workshop sessions. Seventeen fiction and sci-fiction writers (8 F, 9 M) in total participated in the workshops. The participants' ages ranged between 18 and 65. The average age was 27. Eleven were amateur or hobbyist writers including people who published magazines, were involved in creative writing workshops or writer clubs, were sci-fi story or poet writers, and studied literature with a strong interest in the sci-fi genre. Six were professional writers including content writers, academic writers, and professional young adult fiction writers, who were already writing a sci-fi book or writing fiction for TV shows.

3.1.2. Design Fiction Workshop: Materials. As multimodal experiences should be intuitive, knowing when intuitive interaction may be critical can help to design multimodal interaction. Six contexts from the preliminary study findings

TABLE 1: A taxonomy of in-car contexts where intuitive interaction would be critical [22].

Context	In-car contexts	f	Subcontexts
1	Switching-required context	27.17%	Adjusting, changing, setting up
2	Emotion-underlain context	24.59%	Emotionally charged, perfect aura, sensorily-loaded
3	Idle-away context	18.64%	Being recommended, being served, consuming media
4	Less-controllable context	15.75%	Other road users, outer conditions, Breakdown
5	Time-sensitive context	9.8%	Personal schedule, immediate danger, waiting/impatience
6	Task-oriented context	4.02%	Consecutive tasks, clutter activities

TABLE 2: Participant information in relation to the workshop sessions.

Workshop session	Participants' age	Participants' gender	Participants' writing background	Assigned in-car contexts
1	21, 25, 29, 35	4 M	2 professional, 2 amateur	1, 2,3,4
2	25, 65	1 F, 1 M	1 professional 1, amateur	5, 6
3	19, 22, 23	1 F, 2 M	1 professional 2, amateur	1, 3, 6
4	18, 22, 24, 25, 34	3 F, 2 M	1 professional 4, amateur	1, 2, 4, 5
5	23, 25, 27	3 F	1 professional 2, amateur	2, 3, 4

[22] were employed to investigate users' expectations of the in-vehicle multimodal experience. It was intended to frame the in-vehicle interactions in a specific context to reveal rich and in-depth context-specific thoughts about their multimodality expectations.

Each context was introduced to participants with the following details: (1) the key definition, (2) examples, and (3) quotes. An example of how the context was presented to participants is shown in Figure 1.

3.1.3. Design Fiction Workshop: Procedure and Questions. Since this research is aimed at extracting perspectives from different contexts, the participants were assigned a context they would explore throughout the workshop. The in-car contexts were allocated randomly to the participants. The study was conducted through an online video conference platform with an online collaboration whiteboard application. Everyone's video camera was on, and participants could see and interact with each other as they went through the workshop activity. Each session lasted approximately 120 minutes, and five different workshops were conducted. Nine questions in total were used to extract participants' expectations of the in-vehicle multimodal experience.

In each session, the participants were first welcomed and informed of the goal and procedure of the workshops. Then, the whiteboard link was shared in the chat section, so participants could meet there. The workshop material (context) was preuploaded, and each task for individual participants was prewritten on the whiteboard application, which was separated by columns for each participant. When participants met on the whiteboard, everyone's name was located at the top of their column, and then, the context with explanation, examples, and quotes (see Figure 1) was shown just under it, which signified the context they were assigned.

Before the session started, participants were given three minutes to familiarise themselves with the platform and work out how to navigate it. Throughout the workshop, par-

ticipants' contexts and the future were emphasised to prevent them from going out of the context without realising it. They were provided with the necessary time and digital space to fuse their sense of freedom while conducting the activities. Breakout rooms were used in group discussions to encourage participants to speak up and engage within smaller groups. The following steps and questions were employed in each workshop session:

(i) Step 1: familiarisation and individual brainstorming

Each participant was assigned to a different future in-vehicle context with which to familiarise themselves. The initial session was intentionally an individual exploration because personal time was required for participants to get familiarised. They were encouraged to engage themselves with the contexts. A question was used to help participants to relate to and interpret the contexts both personally and in a broader sense—"What could this multimodal in-vehicle context mean in a broader sense? (e.g., inner, external, social, communal, political, environmental, economic, spiritual)."

(ii) Step 2: group discussion and brainstorming

After the participants had familiarised themselves with the contexts, they were encouraged to explore and extract ideas about future vehicles' multimodal usage experiences. First, they were asked to imagine a future scene in which they had a multimodal AV based on their assigned context. Then, they were asked to illustrate the details of the capabilities and tasks they expect from the vehicle in the context, which included their preferred ways of interacting with the multimodality effortlessly (e.g., showing intention and receiving actions). Interactive group cohesion was used in step 2.

(iii) Step 3: sci-fi scenario writing

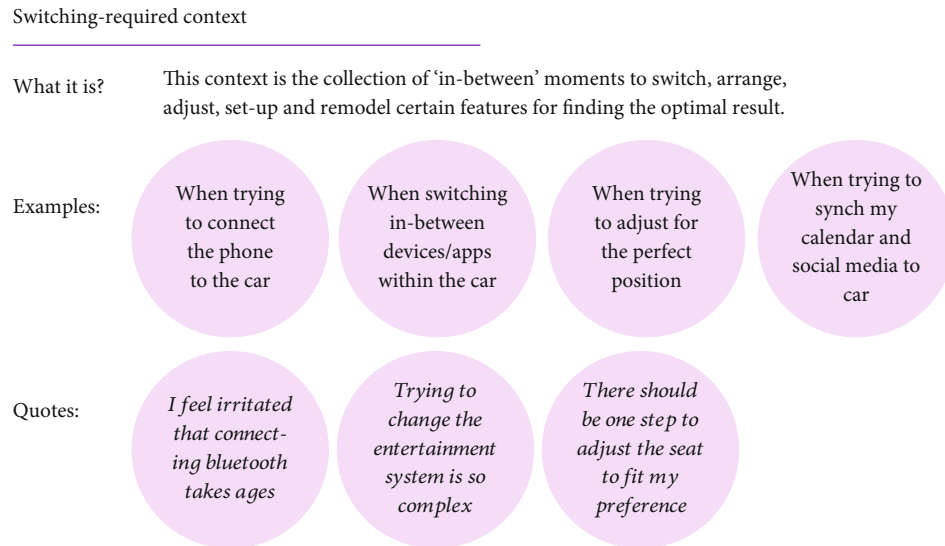


FIGURE 1: Example of how the context is presented to the participants.

Participants were prompted to imagine a future and write a fiction scenario within a future multimodal vehicle by considering “before”–“during”–“after” the journey. During their scenario writing, the questions were used to prompt elaboration of their conceptions and expectations about the in-vehicle multimodal experience—“In which ways does your multimodal vehicle penetrate your (or others around you) life and experience? Why and How?” or “How do you (or others around you) interact, collaborate and help each other with your vehicle for the optimal enhancement of your experience?”

(iv) Step 4: roleplaying

Participants were asked to empathise with the future character in their fiction writing and to roleplay their scenario. Also, they were asked to interact with other participants, exchange ideas, give recommendations to each other, explore the possibilities when two contexts are merged, and discuss the remarkable parts of their interaction and the reasons behind the interactions.

The research activities were performed in accordance with the code of practice of the university. The ethics were approved before any research activity. All of the participants were provided with a participant information form and then agreed to an ethics consent form before the study.

3.1.4. Design Fiction Workshop: Data Analysis. To analyse the results, it was necessary to inquire beyond what was immediately visible and instead search for a semantically higher meaning consistent with understanding the deeper meaning and aspects that impact users’ experiences in each context. The aspects and dimensions that greatly impact what users expect from a multimodal experience were therefore investigated in the analysis.

The general outline of the analysis was drawn from thematic analysis [62]. It was conducted following the step-by-

step thematic analysis guidelines. An online collaborative workspace application was used to organise the codes, which was helpful in organising and working with the large amount of data while increasing the flexibility of the process. The following steps were taken:

- (i) The audio data was transcribed into a written form and combined with the data from post-its and fiction scenarios that participants generated (see Table 3)
- (ii) Transcriptions were read and cross-checked by the researchers (authors) to familiarise them with the data and ensure they were represented well
- (iii) The data was thematically coded according to the research aim of understanding users’ expectations through their views, the tasks, and activities they desire to perform, their preferred input-output interactions with the car, and the narrations they created within a fictional multimodal car
- (iv) Patterns of expectations in the forms of tasks started to emerge for each context. The subthemes were extracted based on the repeating patterns of tasks. They were all colour coded separately
- (v) This part of the study was conducted separately for each context, which helped us to reveal more details
- (vi) Finally, the data was analysed as a whole to determine whether there were recurring aspects

Three reviewers (authors) completed an inter-rater reliability check independently to ensure the reliability of the results. In total, 11 themes were revealed from the analysis. Each reviewer checked two different measures of intercoder reliability: the per cent agreement and Fleiss’s kappa [63]. The checks suggested that two of the 11 themes could be

TABLE 3: An excerpt from the scenarios of participants 17 and 14.

We talk about the kids and John, until I must politely make it clear we need to get going. The two-dimensional man grabs the doorhandle, and the car door swings open. I enter and let my hand rest on the leather of my seat. It is soft, but not as soft as I would like it to be as I long for the comfort of my bed. I feel as the leather fades into cotton, just like my bedsheets. The little chalk man Tom appears on the monitor where, 100 years ago, a radio would have been. He is content to sit on a chalk chair, occasionally meeting my eyes and signalling me a joke. I am not frightened by the distraction, because I know Tom takes over the wheel the second my eyes shift away from the road. I know he will protect me. I see the change in Tom's face before I feel the car automatically slowing down. I take my hands away from the wheel, granting Tom full control, and ask "What is wrong?". Tom makes a sigh motion and responds "Well, you will not like this, but there has been an accident on the freeway."

I frown angrily "Are you kidding me? I need to get to work!" The car stops, stuck behind rows upon rows of other cars. Tom looks sad and I take a deep breath. "Sorry, I did not mean that. It's not your fault." I pause. "Can you hologram me into the meeting room, please?" (Participant 17)

With a rumbling stomach, I get into my seat and start the car. Working on New Year's Eve is quite dreadful, and I could see it on everyone's faces. If it's not their faces only, it is on their stories, it is on their likes, it is on their WhatsApp group made after a pun on our boss' name. As I press the button for a random podcast to listen to, it plays something called "Owning the Bourgeois." It's a bit cheeky, but it is nice. It feels like a friend. Similar to whenever me and the car play a game where we try to guess what object or place the other is describing. It is nice to feel someone's presence on the road, even if they are not there. The podcast skips to an episode on how we are controlled through food. How sugar is pumped into even chicken, how food deserts are a tool to hurt others, and although I still do have the urge to get the Taco-bell, the car takes the right. It stings a little bit less, with all the conspiracies. Speaking about the conspiracies, I look at my right and see one of those people with their anti-mask emblems carved onto the screenings on their cars. If it were a few years ago, those would have a hard time with a change like that on their vehicles. Now, though, you can change the looks of your car with a click, and I hear they will be blocking harmful messages that could be put on the design. (Participant 14)

considered highly reliable. The reliable themes—*attention* and *duration*—had a reliability rate of 100% (excellent agreement), while the other nine themes—*proactivity*, *inclusivity*, *control*, *individuality*, *information density*, *physicality*, *proximity*, *space usability*, and *temporality*—achieved the reliability rate of 85.185%, 82.716%, 90.123%, 87.654%, 75.309%, 55.556%, 80.247%, 85.185%, and 70.37%, respectively. Only the themes that achieved 100% were selected.

4. Research Findings

As a result of the thematic analysis, 22 subthemes were derived with 27 subconditions as the categories of users' expected tasks of multimodal experience (see Table 4). Each task (from T1 to T22) indicates the details of preferred ways to interact effortlessly with the multimodality in an AV.

As a result of data analysis, as higher categories, two dimensions—"attention" and "duration"—were identified as the most prominent aspects that crucially impact participants' expectations of in-vehicle multimodality (see Table 5). Each task was able to be categorised by the dimensions. Table 5 shows the relationship between each dimension and tasks. This section will explain the research results in detail, by justifying how the two dimensions are drawn from the user's insights. Then, in the next section, these dimensions and their interaction with each other and their implications will be discussed comprehensively.

4.1. Dimension 1: Attention

LOW ATTENTION \longleftrightarrow HIGH ATTENTION

From the participants' narratives, "attention" appeared as one of the most important dimension of their expectations of the multimodal experience. This could be interpreted as users expecting that the level of their attention will be a significant determinant in how they experience future multimodal implementations inside an AV. Here,

"attention" refers to the attention between the occupant and the potential multimodal experience involved in the vehicle.

In previous studies, users' attention and cognitive workload have been regarded as factors to be considered in multimodal interactions [20, 64, 65]. Particularly, context-aware interactions, such as interactions with a visually attentive interface, rely on a person's attention as the primary input [66]. Obrenovic and Starčević [37] also mentioned "attention: focus and context" under the category of cognitive effects when defining the effects of multimodal interaction. Moreover, interaction types are often categorised based on the level of attention [67].

When users mentioned their desires for multimodal interactions in a vehicle, the level of attention between the occupants and experiences was typically depicted in direct or indirect ways, to impact their experience. A dimensional spectrum from low to high attention appeared through the different expectations. The suggested spectrum for attention is detailed in Table 6.

4.1.1. Low Attention. Some expectations were naturally related to easy and quick desires linked with low and distracted attention, categorised as "low attention." For example, participants expressed their expectations about being able to request information easily and quickly (T3); to skip and switch the media (T2); and to start, pause, and stop actions (T1). These subthemes of tasks were observable through participants' comments:

I might just go by the name of station 1 most you know, to say like, "BBC One extra" or "Kiss." (T3)

If I don't like the music, I can just wave it with my hands. And we can go to the next genre. (T2)

I snap my fingers to get the music to turn off. (T1)

Similarly, participants mentioned their expectation of avoiding intrusiveness when getting feedback from the

TABLE 4: Retrieved subthemes (tasks) from each context.

Context	Subthemes (tasks)	Subconditions
Switching-required context	T1: quick starting/pausing/stopping actions	
	T2: skipping and switching	
	T3: quick request for info/visual info	
	T4: synchronising	
	T5: arranging interior physical space	
	T6: picking and selecting the specific information	
	T7: adjusting specific settings that require graduality	
Emotion-underlain context	T8: finding stability and calmness	T8a: when in low arousal T8b: when in high arousal
	T9: gaining control and confidence	
	T10: creating/activating	
Idle-away context	T11: consuming media	T11a: when activating the media T11b: when conducting the secondary activity when consuming media
	T12: receiving recommendations	T12a: requested recommendations T12b: unrequested recommendations
Less-controllable context	T13: getting assurance (for what I cannot control)	
	T14: getting information (for what I can still control)	T14a: when mildly hazardous T14b: when severely hazardous
	T15: reporting the error	
Time-sensitive context	T16: showing the urgency	T16a: when the vehicle is reactive T16b: when the vehicle is proactive
	T17: relieving the stress of the urgency	
	T18: getting reassurance	
Task-oriented context	T19: focusing on a task	
	T20: becoming aware of surrounding when focused	
	T21: conducting consecutive tasks	
	T22: avoiding intrusiveness	

TABLE 5: The relationship between dimensions and tasks.

Dimension	Tasks
Attention	
Low attention	T8b, T9, T10, T11a, T11b, T12a, T16a, T17, T20, T21, T22
High attention	T4, T5, T6, T7, T8a, T10, T12b, T13, T14a, T14b, T15, T16b, T18, T19, T20
Duration	
Short duration	T1, T2, T3, T5, T6, T7, T8a, T8b, T11a, T11b, T12a, T12b, T14a, T15, T16a, T19
Long duration	T4, T9, T10, T13, T14b, T16b, T17, T18, T20, T21, T22

vehicle (T22)—which is about creating a space where they will not be distracted. Hence, the interaction when the vehicle sends feedback to the occupant should require low attention, for example,

If it just popped up on the screen - So, I might kind of ignore it. (T22)

4.1.2. High Attention. Conversely, some expectations were linked to high attention in an experience. These were more focused, concentration-requiring tasks categorised under

“high attention.” For example, participants expressed their expectation of attentive interactions when focusing on a task, showing urgency to the vehicle and getting information in hazardous situations. These subthemes of tasks were observable through participants’ comments, such as the following:

Getting information in a severely hazardous situation—*T14b*

If the car needs to interrupt me for a safety reason, even if it seems minor, I’m interested in what it has to tell me.

TABLE 6: The level of attention between the occupant and the potential multimodal experience involved in the vehicle.

Dimension	Description
Low attention (distracted)	The occupant is not fully focused on a particular experience. They choose to have or there happens to be a low awareness and divided attention during the interaction.
Mixed	The experience can consist of both attentive and distracted interactions.
High attention (concentrated)	The occupant is fully concentrated on the experience. High awareness and concentration occur during the interaction.

I would like to have the options for it to grab my attention to the fullest as multi-channel.

Showing the urgency—T16b

If I'm going to have an intelligent car, I want it sort of synced and linked with all my devices... It can take my time and read it if it understands the urgency of the situation.

Focusing on a task—T19

I prefer to give attention to one thing.

I tell her that "I want to focus on rehearsing a presentation." I want it to automatically do things like dim my windows, turn the music down or off.

4.2. Dimension 2: Duration

SHORT DURATION \longleftrightarrow LONG DURATION

Another dimension of the user's expectation of a multimodal experience is the "duration." This refers to the time that passes while the occupant experiences the multimodal interaction and its effects. It appeared in the data as differences between expectations of tasks that take a short time and a long time.

In previous research, "duration" has been considered an impactful factor in users' experiences. The relevancy of time in users' experience has categorised experiences as (1) temporal, meaning the experiences are dynamic and moment-to-moment experiences, and (2) long term, meaning the time is highly related to the experience of the user where experiences remain in their memory and influence their overall evaluation [68]. There are studies on the impact of time and duration on users' perceived experience [69] and on developing positive long-term user experiences with users [68, 70]. These are equally relevant to in-vehicle experiences, as the level of focus and time users take can play a crucial role in impacting users' in-vehicle multimodal expectations.

There was a noticeable difference in the length of the time of the envisioned expectations. A short- to long-dimensional spectrum was constructed by considering the differences mentioned in the participants' narratives. The suggested spectrum for the duration is detailed in Table 7.

4.2.1. Short Duration. Some expectations referred to quick, intermittent, and short-time-required tasks and activities. These activities were categorised under short-term activities. For example, participants mentioned their expectations about receiving unrequested recommendations (T12b), getting information about mildly hazardous situations (T14a), arranging the interior physical space (T5), reporting an error (T15), and starting, pausing, and stopping actions (T1)—all of which were characterised as short-term experiences.

These subthemes of tasks were observable through participants' comments.

If it was kind of unrequested (recommendation), I would prefer it to just be a pop-up on the screen as an option to view. (T12b)

If it's snowing as well, I quickly think I'd want the car to be able to tell me on the dashboard whether or not I'm sliding or if the temperature is quite cold. I think visual icons would be perfect for that. (T14a)

If it's not understanding what I'm trying to say. And it needs a very quick activation. (T15)

It has to be something short and sweet; I'd say something like "Alex, put radio one" on whilst I'm driving" and "I like to set up like straight when I when I start driving or if I suddenly remember I want it to be quick and easy." (T1)

4.2.2. Long Duration. On the other hand, some of the user's expectations were more constant, long-term desires. These activities required an ongoing or long-term experience to be satisfied, for example, getting reassurance from the vehicle, synchronising, creating the aura bubble in the vehicle, or getting control and confidence from the vehicle. These subthemes of tasks were observable through participants' comments.

Getting reassurance—T18

If I see something visually, I would be more prone to listen or to do what I'm saying. Whereas if I listen, I'm just going to ignore it.

I want the car to be giving me updates on the blockage. And I would expect this in the form of a new kind of (display)segment. It would just give that kind of constant visual update.

Synchronising—T4

I don't want to connect my Bluetooth. I just want it to be connected already.

Creating the aura bubble—T10

It can sense wherever you're annoyed or angry, or wherever you're in a good mood or whatever. So it automatically does that for you.

The car should understand that something is wrong with me.

Gaining control and confidence—T9

So you maybe don't want the car to talk back to you. If it's just a message on the screen, it's gonna feel less invasive, and you can choose to not even read it. Whereas if it's talking to you, will be like, "Okay, I get it, just shut up now."

It will be more control and wanting less feedback from the car. So I can just be stressed.

This means that the car will be constantly aware of this and will not interrupt and will give the user more control and confidence.

TABLE 7: The duration of the interaction between the occupant and the potential multimodal experience involved in the vehicle.

Dimension	Description
Short term	The occupant is experiencing the experience in a short-term or intermittent way, and its impact is short term.
Mixed	The experience and its impacts can be short or long term.
Long term	The occupant is experiencing the experience in a long term or continuous way.

The two dimensions observed can impact how the multimodal interaction is designed and how input and output modalities are selected, designed, located, combined, and intended.

5. Discussion

5.1. A Map of Expectations of In-Vehicle Multimodal Experiences. Two dimensions, which cover two different spectrums (from low to high and from short to long), can be used to cluster all different tasks by placing “duration” on the x -axis and “attention” on the y -axis. One of the suggestions could be demonstrating the relationship between the themes and subthemes, and the tasks are summarised in Figure 2. Although the exact location of tasks cannot be determined with the data we have drawn, the tasks’ relation to quadrants can be interpreted determined to be able to create this quadrant. For example, T1 (quick starting/pausing/stopping actions) and T2 (skipping and switching) are likely to simultaneously require very short duration and low attention; that is, these interactions do not require much attention and last only for a short time. Similarly, T18 or T4 is more likely to be in effect longer term and requires more attention because one is about getting reassurance from the car, and the other is about synchronising and constantly staying connected with the car.

This two-dimensional spectrum allows for the creation of four different quadrants, on which users’ expectations of in-car multimodality can be placed, as shown in Figure 3. The spectrum enables designers and researchers to map the user expectations of multimodal experiences before implementing them into a vehicle design space. It is important to emphasise that this mapping is not intended to present any superior or inferior quadrants; each may have advantages or disadvantages.

5.1.1. Quadrant I: Sustained. The first quadrant represents the continuous expectations with a long-term impact on user experience and that require high attention and presence. The interactions in this quadrant are influenced by a sustainable, focused, and mindful relationship between the user and the vehicle. Some exemplary sustained interactions could be related to a vehicle’s self-learning capabilities. The vehicle constantly learns and adapts to the user, creates and maintains a mindful relationship with the user and their circle, and becomes more familiar with them through a growing experience.

This quadrant may be associated with the experiences that offer constant connection and connectivity in the vehicle, attentive engagement with the vehicle’s connected system, and building and maintaining a relationship with the

vehicle. Relevant studies include those that emphasise sustaining a long-term and trustworthy relationship between the human and the computer [71–73], ensuring personalisation and customisation [74], and creating active engagement and immersive experiences [75, 76] with the user.

This quadrant highlights the importance of “establishing and maintaining a continuous, familiar and engaging relationship” as a crucial part of users’ expectations of the in-vehicle multimodal experience. Considerations to take into account to address this include implementing robust software and technical capabilities with efficient machine learning, safety, personalisation, and recognition to get to know the user; creating familiarity and long-term interactive relationships; and implementing VR, AR, and XR applications to devise engaging content and an immersive user interface. Designers may employ these solutions to integrate multimodal interactions into future AV that establish and maintain long-term, trustworthy and engaging, immersive relationships with their users.

5.1.2. Quadrant II: Distinct. The second quadrant represents the expectations that are short term or intermittent and where experiences in these ranges are concentrated. It includes focused and short-term experiences inside the vehicle, such as users enjoying focused, productive time when gaming inside the vehicle or when the vehicle offers an exclusive multimodal space with a range of focused modes. These expectations relate to the possibility of transitioning from one focused event to another, so the vehicle may also require the ability to adapt to changes.

Since the user in this quadrant is engaged with the vehicle in a focused but intermittent and short-term way, this quadrant is associated with multimodal concepts such as temporality, transitioning, adaptability, productivity, and time management. It links with experiences that are focused, attentive, and reserved in their world but short term—which will require switching from one focused mood to another. Studies that relate to the idea of ensuring adaptability [77], creating focused activities [28] and time management [29] within a vehicle, correlate with this quadrant.

This quadrant highlights “creating short-term but engaged experiences and enabling efficient ways of transitioning between them as a crucial part of users’ expectations of the in-vehicle multimodal experience. It may necessitate tools and techniques that adapt in a short time. This quadrant implies the need for multimodality in “adapting experiences,” raising the following question: how can designers apply multimodal interactions in a future AV to create efficient, engaging time and adaptable transitions between focused modes?

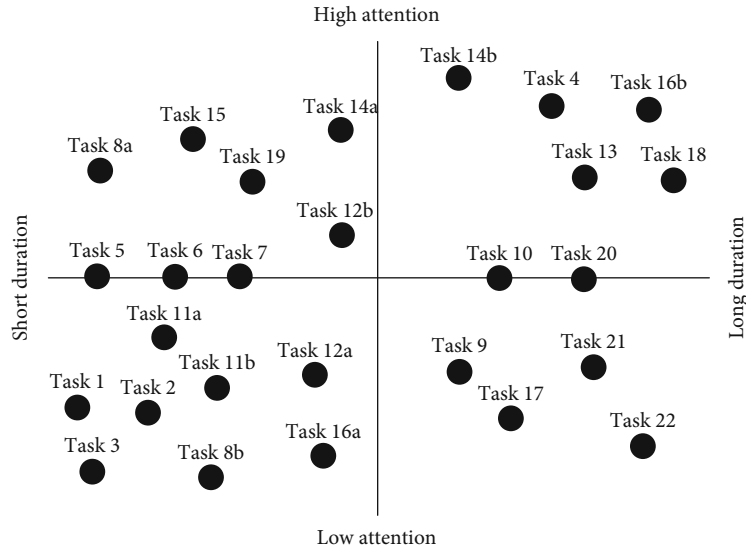


FIGURE 2: Mapping tasks by dimensions (attention and duration).

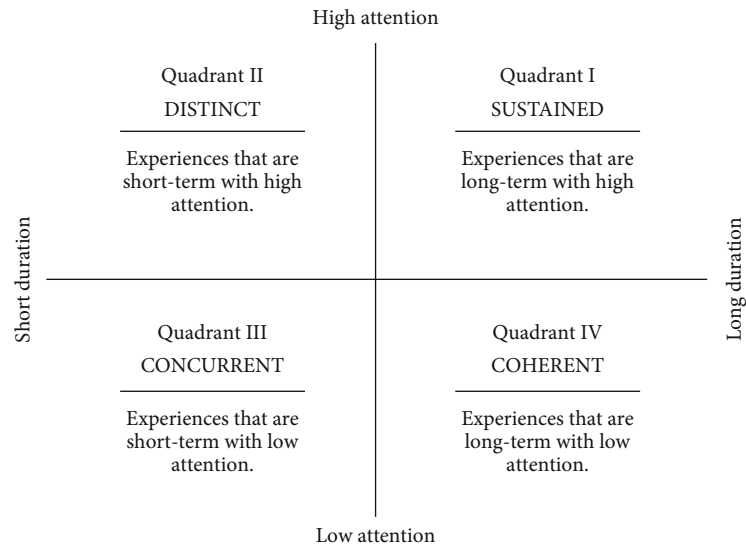


FIGURE 3: A model of users' expectations of an autonomous in-vehicle multimodal experience.

5.1.3. *Quadrant III: Concurrent.* The third quadrant embodies the expectations that involve a short-term duration and low attention. These are momentary, distracted experiences that do not necessarily require particularly focused attention. Such experiences could involve switching between different platforms, changing the music, getting quick movie recommendations, and urgently informing the vehicle. For example, when the user switches from listening to music from Spotify to listening from YouTube while reading an e-book, the vehicle may offer a range of multimodal synchronisation options.

Due to its short-term and low-attention nature, this quadrant links with multimodal concepts such as momentariness, transitivity, seamlessness, and intuitiveness. Since the experiences are quick and do not have to include particularly focused attention, this quadrant highlights “creating

rapid, practical, and absorbed experiences” as a crucial part of users’ expectations of the in-vehicle multimodal experience. Relevant studies such as Detjen et al. [78] and D’Eusano et al. [79] have explored intuitive interactions that may link with concurrent activities. Implementing technical capabilities that enable rapid, smooth, and intuitive transitions from one modality to another, one device to another, and one activity to another could be considered. The implication of this quadrant could invite designers and researchers to address how multimodal interactions can be integrated into a future AV as methods to establish and maintain concurrent transitioning that would allow intuitive experiences for users.

5.1.4. *Quadrant IV: Coherent.* The final quadrant refers to the expectations that are long term and continuous and that require only low attention. It has an emphasis on the

continuous but not discernible experiences that may create a particular undertone or mood within the vehicle. For example, the vehicle may constantly filter out the information that the user finds undesirable before exposing the remainder to them on the display. Similarly, the vehicle may constantly arrange lighting, colour, sound, and tone based on the users' moods without them noticing it.

This quadrant correlates with multimodal concepts such as intangibility, intuitiveness, and sensitivity, as the impact of the experiences on this quadrant may not be immediately perceivable. However, their impacts are complementary to the in-vehicle experience. Studies such as Löcken et al. [80] and Mahmood et al. [81] are related to this quadrant due to their exploration of user experiences that do not occupy users' attention. Since the experiences are effective in the long term but not necessarily noticeable, this quadrant highlights "creating a coherent undertone, foundation, and ambiances for the other experiences to occur." This therefore becomes an integral part of users' expectations of the in-vehicle multimodal experience. Understanding this element should inform how multimodal interactions are integrated into a future AV to establish and maintain a coherent undertone and intuitive user experience.

Various user studies have investigated in-vehicle multimodal experiences or interactions [2, 13, 15, 34, 35, 82]. These studies focused on testing and evaluating a particular in-vehicle multimodal interaction (e.g., speech, gesture, dials, and haptic) on users but do not consider the experience of an overall multimodal journey in a vehicle. Other studies have presented a classification, model, or mapping of multimodality [8, 36–38, 40, 41]. However, these are either based on general interactions with devices or on the technical aspects but do not focus on the AVs and occupants' experiences in them. Hence, the proposed conceptual model could facilitate a discussion in the early design process regarding AVs' multimodal applications by holistically understanding users' expectations with more clarity and influencing design decisions accordingly.

6. Conclusion and Future Work

This paper presented two studies, focusing on the latter study, to understand users' expectations of a multimodal experience in an AV. The attention and the duration of the multimodal interaction and their effects were determined as the two critical dimensions to map out users' expectations. The outcome was a map of expectations of in-vehicle multimodal experiences structured, which contained four quadrants: *sustained* (long term-high attention), *distinct* (short term-high attention), *concurrent* (short term-low attention), and *coherent* (long term-low attention). This research presents a novel conceptual model for understanding users' expectations in autonomous in-vehicle multimodal experiences. Distinct from conventional evaluations of specific multimodalities, it takes a significant stride in this field by presenting a model that categorises users' in-vehicle multimodal experiences thoroughly and offers a holistic perspective on users' in-vehicle expectations of a future AV. The map is poised to empower designers to make well-

informed decisions for future in-vehicle multimodal experiences by leveraging user insights. Further, it can serve as a foundational resource for crafting design guidelines for in-vehicle multimodal interactions.

The dynamics of technological development today require us to constantly understand users' altering expectations and desires to provide them with a more advanced user experience. These research findings provide a theoretical basis for researchers to develop further guidelines for in-vehicle multimodal interaction. Using the map of expectations, in which in-vehicle multimodal experiences have been broken down into layers, as a starting point, the following questions can be raised to develop this model further:

- (i) Which technical capabilities, skills, implications, and conditions may each quadrant bring or require?
- (ii) What kind of devices and design experiences can assist users in maximising their lives?
- (iii) What about experiences that move between quadrants in a single scenario, and which interactions may help navigate the transitions between quadrants?

While this study has made original contributions, it is also crucial to mention its limitations. Both the survey and the workshops were conducted remotely. Although this allowed the researchers to reach more people, it would be better to proceed with them face to face, collaborate, and roleplay on-site. Further, although participants were strongly encouraged to consider a future mobility context through contexts, fiction, scenario writing, and roleplaying, it was inevitable for them to be distracted by the present context. This is a reoccurring problem with user studies that explore the future. In future studies, researchers could enable immersion when conducting workshop activities with the help of technologies such as virtual or augmented reality.

It would also be interesting to evaluate whether particular traits and skills of users, such as gender, age, or level of driving experience, impact the attention levels. For instance, future studies could examine if attention levels increase or decrease with increasing user age and explore how user expectations may differ based on that. Moreover, in-depth qualitative research could be conducted to detail and specify the step-by-step tasks for each quadrant. Similarly, an in-depth quantitative study could determine where the tasks would be located on the map in each quadrant by exploring attention and duration levels, to further develop knowledge of the in-vehicle multimodal user experience. Neuropsychological assessments and cognitive tests could be implemented to preliminarily evaluate users' attention levels.

The outcome of this paper provides designers with a better understanding of user experiences to design multimodal interactions, as well as a theoretical basis for future researchers to develop more applicable design guidelines for in-vehicle multimodal interactions. This will contribute to the adaptation of multimodal interactions inside the vehicle, facilitate the transition from unidirectional tasks to experiences, and enable future users to enjoy improved multimodal experiences instead of fixed and nonreciprocal ones.

Data Availability

Due to the commercial nature of this research, it was agreed that the data from this research would not be shared publicly. Hence, supporting data is not available.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was fully funded by the Holistic UX Group, Automotive R&D Division, HMG, Seoul, South Korea. The grant ID is 12597. Open Access funding was enabled and organised by JISC.

References

- [1] C. Müller and G. Weinberg, "Multimodal input in the car, today and tomorrow," *IEEE Multimedia*, vol. 18, no. 1, pp. 98–103, 2011.
- [2] D. Siewiorek, A. Smailagic, and M. Hornyak, "Multimodal contextual car-driver interface," in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pp. 367–373, Pittsburgh, PA, USA, 2002.
- [3] R. Viereckl, D. Ahlemann, A. Koster, and S. Jursch, "Racing ahead with autonomous cars and digital innovation," *Auto Tech Review*, vol. 4, no. 12, pp. 18–23, 2015.
- [4] D. Gkouskos, I. Petterson, M. Karlsson, and F. Chen, "Exploring User Experience in the Wild: Facets of the Modern Car," in *Design, User Experience, and Usability: Interactive Experience Design. DUXU 2015*, A. Marcus, Ed., vol. 9188 of Lecture Notes in Computer Science(), Springer, Cham, 2015.
- [5] H. Strömberg, I. Petterson, and W. Ju, "Enacting metaphors to explore relations and interactions with automated driving systems," *Design Studies*, vol. 67, pp. 77–101, 2020.
- [6] O. A. L.'s. Dictionary, *Mode noun*, Oxford Advanced Learner's Dictionary, 2023, <https://www.oxfordlearnersdictionaries.com/definition/english/mode>.
- [7] O. A. L.'s. Dictionary, *Modality noun*, Oxford Advanced Learner's Dictionary, 2023, <https://www.oxfordlearnersdictionaries.com/definition/english/modality?q=modality>.
- [8] C. Platz, *Design beyond Devices: Creating Multimodal, Cross-Device Experiences*, Rosenfeld Media, New York, New York, USA, 2020.
- [9] S. Oviatt, R. Lunsford, and R. Coulston, "Individual Differences in Multimodal Integration Patterns: What Are They and why Do They Exist?," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 241–249, New York, 2005.
- [10] A. R. Aftab, "Multimodal Driver Interaction with Gesture, Gaze and Speech," in *2019 International Conference on Multimodal Interaction*, pp. 487–492, New York, 2019.
- [11] B. Pfleging, S. Schneegass, and A. Schmidt, "Multimodal interaction in the car: combining speech and gestures on the steering wheel," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 155–162, Association for Computing Machinery, New York, NY, USA, 2012.
- [12] K. Väänänen-Vainio-Mattila, J. Heikkinen, A. Farooq, G. Evreinov, E. Mäkinen, and R. Raisamo, "User experience and expectations of haptic feedback in in-car interaction," in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia*, pp. 248–251, New York, 2014.
- [13] T. Schneider, S. Ghellal, S. Love, and A. R. Gerlicher, "Increasing the user experience in autonomous driving through different feedback modalities," in *26th International Conference on Intelligent User Interfaces*, New York, 2021.
- [14] N. Gang, S. Sibi, R. Michon, B. Mok, C. Chafe, and W. Ju, "Don't be alarmed: sonifying autonomous vehicle perception to increase situation awareness," in *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, pp. 237–246, New York, 2018.
- [15] I. Politis, S. Brewster, and F. Pollick, "Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers," *International Journal of Mobile Human Computer Interaction (IJMHCI)*, vol. 9, no. 3, pp. 1–16, 2017.
- [16] S. Zhou, R. Lan, X. Sun, J. Bai, Y. Zhang, and X. Jiang, "Emotional design for in-vehicle infotainment systems: an exploratory co-design study," in *International Conference on Human-Computer Interaction*, pp. 326–336, Springer International Publishing, Cham, 2022.
- [17] D. Schnelle-Walka, D. R. McGee, and B. Pfleging, "Multimodal Interaction in Automotive Applications," *Journal on Multimodal User Interfaces*, vol. 13, no. 2, pp. 53–54, 2019.
- [18] M. Turk, "Multimodal interaction: a review," *Pattern Recognition Letters*, vol. 36, pp. 189–195, 2014.
- [19] P. K. Murali, M. Kaboli, and R. Dahiya, "Intelligent in-vehicle interaction technologies," *Advanced Intelligent Systems*, vol. 4, no. 2, article 2100122, 2022.
- [20] W. Wang, F. Song, and T. Zhu, "SAME: the design space for seamless automotive multimodal experience," *IT Professional*, vol. 24, no. 3, pp. 35–42, 2022.
- [21] R. Häuslschmid, M. von Buelow, B. Pfleging, and A. Butz, "SupportingTrust in autonomous driving," in *Proceedings of the 22nd international conference on intelligent user interfaces*, pp. 319–329, New York, 2017.
- [22] E. Ince, K. Cha, and J. Cho, "When do people expect effortless in-car interactions?," *Intelligent Human Systems Integration (IHSI 2023): Integrating People and Intelligent Systems*, vol. 69, no. 2023, pp. 17–26, 2023.
- [23] S. C. Lee, C. Nadri, H. Sanghavi, and M. Jeon, "Eliciting user needs and design requirements for user experience in fully automated vehicles," *International Journal of Human-Computer Interaction*, vol. 38, no. 3, pp. 227–239, 2022.
- [24] E. Luger and A. Sellen, "'Like having a really bad PA' the gulf between user expectation and experience of conversational agents," in *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 5286–5297, New York, 2016.
- [25] K. Cha, *Affective Scenarios in Automotive Design: A Human-Centred Approach towards Understanding of Emotional Experience*, Brunel University London, 2019.
- [26] T. Hecht, A. Feldhütter, K. Draeger, and K. Bengler, "What do you do? An analysis of non-driving related activities during a 60 minutes conditionally automated highway drive," in *Human Interaction and Emerging Technologies: Proceedings*

- of the 1st International Conference on Human Interaction and Emerging Technologies (IHET 2019), August 22-24, 2019, pp. 28–34, Springer International Publishing, Nice, France, 2019.
- [27] U. Montanaro, S. Dixit, S. Fallah et al., “Towards connected autonomous driving: review of use-cases,” *Vehicle System Dynamics*, vol. 57, no. 6, pp. 779–814, 2019.
- [28] B. Pflöging, M. Rang, and N. Broy, “Investigating user needs for non-driving-related activities during automated driving,” in *Proceedings of the 15th international conference on mobile and ubiquitous multimedia*, pp. 91–99, New York, 2016.
- [29] G. Stevens, P. Bossauer, S. Vonholdt, and C. Pakusch, “Using time and space efficiently in driverless cars: findings of a co-design study,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA, 2019.
- [30] I. Pettersson, “Travelling from fascination to new meanings: understanding user expectations through a case study of autonomous cars,” *International Journal of Design*, vol. 11, no. 2, p. 1, 2017.
- [31] I. Politis, S. A. Brewster, and F. Pollick, “Evaluating Multimodal Driver Displays under Varying Situational Urgency,” in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 4067–4076, New York, 2014.
- [32] F. Ekman, M. Johansson, and J. Sochor, “To see or not to see: the effect of object recognition on users’ trust in automated vehicles,” in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, New York, 2016.
- [33] S. Faltalous, M. Baumann, S. Schneegass, and L. L. Chuang, “Design guidelines for reliability communication in autonomous vehicles,” in *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, pp. 258–267, New York, 2018.
- [34] J. Kim, J. H. Ryu, and T. M. Han, “Multimodal interface based on novel HMI UI/UX for in-vehicle infotainment system,” *ETRI Journal*, vol. 37, no. 4, pp. 793–803, 2015.
- [35] B. A. Lewis, M. N. Penaranda, D. M. Roberts, and C. L. Baldwin, “Effectiveness of Bimodal Versus Unimodal Alerts for Distracted Drivers,” in *Driving Assessment Conference*, University of Iowa, 2013.
- [36] L. Nigay and J. Coutaz, “A design space for multimodal systems: concurrent processing and data fusion,” in *Proceedings of the INTERACT’93 and CHI’93 conference on Human factors in computing systems*, pp. 172–178, New York, 1993.
- [37] Z. Obrenovic and D. Starčević, “Modeling multimodal human-computer interaction,” *Computer*, vol. 37, no. 9, pp. 65–72, 2004.
- [38] D. Lalanne, L. Nigay, P. Palanque, P. Robinson, J. Vanderdonckt, and J.-F. Ladry, “Fusion engines for multimodal input: a survey,” in *Proceedings of the 2009 international conference on Multimodal interfaces*, New York, 2009.
- [39] M. Serrano, D. Juras, and L. Nigay, “A three-dimensional characterization space of software components for rapidly developing multimodal interfaces,” in *Proceedings of the 10th International Conference on Multimodal Interfaces*, pp. 149–156, Association for Computing Machinery, New York, NY, USA, 2008.
- [40] D. Di Mitri, J. Schneider, M. Specht, and H. Drachslar, “From signals to knowledge: a conceptual model for multimodal learning analytics,” *Journal of Computer Assisted Learning*, vol. 34, no. 4, pp. 338–349, 2018.
- [41] J. Chang and M.-L. Bourguet, “Usability framework for the design and evaluation of multimodal interaction,” in *Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability*, pp. 123–126, IGI Global, 2008.
- [42] S. Zaidi and U. Kirazci, “Design actions for the Google assistant beyond smart speakers,” 2018, https://www.youtube.com/watch?v=JDakZMLXpQo&ab_channel=GoogleDevelopers.
- [43] S. Rafael and V. M. Almeida, “Implementation Goals for Multimodal Interfaces in Human-Computer Interaction,” in *Human-Computer Interaction. Theory, Methods and Tools. HCII 2021*, M. Kurosu, Ed., vol. 12762 of Lecture Notes in Computer Science(), Springer, Cham, 2021.
- [44] T. Reis, M. de Sá, and L. Carriço, “Multimodal Interaction: Real Context Studies on Mobile Digital Artefacts,” in *Haptic and Audio Interaction Design. HAID 2008*, A. Pirhonen and S. Brewster, Eds., vol. 5270 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2008.
- [45] U. E. Manawadu, M. Kamezaki, M. Ishikawa, T. Kawano, and S. Sugano, “A multimodal human-machine interface enabling situation-adaptive control inputs for highly automated vehicles,” in *2017 IEEE Intelligent Vehicles Symposium (IV)*, Los Angeles, CA, USA, 2017.
- [46] A. K. Dey, “Understanding and using context,” *Personal and Ubiquitous Computing*, vol. 5, no. 1, pp. 4–7, 2001.
- [47] J. Masuhr, F. Klomp maker, C. Reimann, and K. Nebe, “Designing context-aware in-car information systems,” in *5th International ICST Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, ICST, 2008.
- [48] R. McCall, F. McGee, A. Meschtscherjakov, N. Louveton, and T. Engel, “Towards a taxonomy of autonomous vehicle hand-over situations,” in *Proceedings of the 8th International Conference on automotive user interfaces and interactive vehicular applications*, pp. 193–200, New York, 2016.
- [49] V. Braun, V. Clarke, E. Boulton, L. Davey, and C. McEvoy, “The online survey as a qualitative research tool,” *International Journal of Social Research Methodology*, vol. 24, no. 6, pp. 641–654, 2021.
- [50] V. Braun, V. Clarke, and D. Gray, “Innovations in Qualitative Methods,” in *The Palgrave handbook of critical social psychology*, pp. 243–266, Springer, 2017.
- [51] J. Bleecker, “Design fiction: a short essay on design, science, fact, and fiction,” in *Machine Learning and the City: Applications in Architecture and Urban Design*, pp. 561–578, Wiley, 2022.
- [52] C. Harrington and T. R. Dillahunt, “Eliciting tech futures among Black young adults: a case study of remote speculative co-design,” in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, 2021.
- [53] S. Landry and M. Jeon, “Participatory Design Research Methodologies: A Case Study in Dancer Sonification,” *International Conference on Auditory Display*, vol. 69, 2017.
- [54] I. Hussein, M. Mahmud, and A. Osman, “A survey of user experience practice: a point of meet between academic and industry,” in *2014 3rd International Conference on User Science and Engineering (i-USER)*, Shah Alam, Malaysia, 2014.
- [55] J. L. Sheh and X. Shen, “User requirements in mobile systems,” 2001, <https://api.semanticscholar.org/CorpusID:58423161>.
- [56] A. H. Ambe, M. Brereton, A. Soro, L. Buys, and P. Roe, “The Adventures of Older Authors: Exploring Futures through Co-Design Fictions,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–16, Association for Computing Machinery, New York, NY, USA, 2019.

- [57] N. Bozic Yams and Á. Aranda Muñoz, "Poetics of future work: blending speculative design with artistic methodology," in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, 2021.
- [58] O. Goldenberg and J. Wiley, "Individual and group brainstorming: does the question matter?," *Creativity Research Journal*, vol. 31, no. 3, pp. 261–271, 2019.
- [59] E. Cheon, S. T.-H. Sher, Š. Sabanović, and N. M. Su, "I beg to differ: soft conflicts in collaborative design using design fictions," in *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 201–214, New York, 2019.
- [60] B. Medler and B. Magerko, "The Implications of Improvisational Acting and Role-Playing on Design Methodologies," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 483–492, New York, 2010.
- [61] S. Campbell, M. Greenwood, S. Prior et al., "Purposive sampling: complex or simple? Research case examples," *Journal of Research in Nursing*, vol. 25, no. 8, pp. 652–661, 2020.
- [62] V. Braun and V. Clarke, *Thematic Analysis*, American Psychological Association, 2012.
- [63] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.
- [64] F. Chen, N. Ruiz, E. Choi et al., "Multimodal behavior and interaction as indicators of cognitive load," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 4, pp. 1–36, 2013.
- [65] S. Norris, *Analyzing Multimodal Interaction: A Methodological Framework*, Routledge, 2004.
- [66] T. Selker, "Visual attentive interfaces," *BT Technology Journal*, vol. 22, no. 4, pp. 146–150, 2004.
- [67] S. Bakker, E. van den Hoven, and B. Eggen, "Peripheral interaction: characteristics and considerations," *Personal and Ubiquitous Computing*, vol. 19, no. 1, pp. 239–254, 2015.
- [68] S. Kujala and T. Miron-Shatz, "The evolving role of expectations in long-term user experience," in *Proceedings of the 19th International Academic Mindtrek Conference*, pp. 167–174, New York, 2015.
- [69] A. K. Trapp, *Time perception in human-computer interaction: the role of duration perception and passage of time in user experience*, 2001, <https://osf.io/hcbg6/>.
- [70] T. S. Amer and T. L. Johnson, "IT progress indicators," *International Journal of Technology and Human Interaction (IJTHI)*, vol. 10, no. 3, pp. 58–71, 2014.
- [71] T. Bickmore and R. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.
- [72] L. Morra, F. Lamberti, F. G. Praticó, S. L. Rosa, and P. Montuschi, "Building trust in autonomous vehicles: role of virtual reality driving simulators in HMI design," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 9438–9450, 2019.
- [73] T. Zhang, X. Qu, X. Zhang, J. Zeng, H. Zhu, and H. Zhu, "Automated vehicle acceptance in China: social influence and initial trust are key determinants," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 220–233, 2020.
- [74] X. Sun, J. Li, P. Tang et al., "Exploring personalised autonomous vehicles to influence user trust," *Cognitive Computation*, vol. 12, no. 6, pp. 1170–1186, 2020.
- [75] V. Distler, C. Lallemand, and T. Bellet, "Acceptability and acceptance of autonomous mobility on demand: the impact of an immersive experience," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, New York, 2018.
- [76] C. Wienrich, N. Döllinger, and R. Hein, "Behavioral framework of immersive technologies (BehaveFIT): how and why virtual reality can support behavioral change processes," *Frontiers in Virtual Reality*, vol. 2, article 627194, 2021.
- [77] D. B. Kaber, Y. Liang, Y. Zhang, M. L. Rogers, and S. Gangakhedkar, "Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 15, no. 5, pp. 491–501, 2012.
- [78] H. Detjen, S. Geisler, and S. Schneegass, "Driving as side task: exploring intuitive input modalities for multitasking in automated vehicles," in *Extended abstracts of the 2021 chi conference on human factors in computing systems*, Association for Computing Machinery, New York, NY, USA, 2021.
- [79] A. D'Eusano, A. Simoni, S. Pini, G. Borghi, R. Vezzani, and R. Cucchiara, "Multimodal Hand Gesture Classification for the Human–Car Interaction," *Informatics*, vol. 7, no. 3, p. 31, 2020.
- [80] A. Löcken, W. Heuten, and S. Boll, "AutoAmbiCar: using ambient light to inform drivers about intentions of their automated cars," in *Adjunct proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications*, New York, 2016.
- [81] A. Mahmood, B. Butler, Q. Z. Sheng, W. E. Zhang, and B. Jennings, "Need of Ambient Intelligence for Next-Generation Connected and Autonomous Vehicles," in *Guide to Ambient Intelligence in the IoT Environment. Computer Communications and Networks*, Z. Mahmood, Ed., Springer, Cham, 2019.
- [82] S. Castronovo, A. Mahr, M. Pentcheva, and C. Müller, "Multimodal Dialog in the Car: Combining Speech and Turn-and-Push Dial to Control Comfort Functions," in *11th Annual Conference of the International Speech Communication Association 2010 (INTERSPEECH 2010)*, Makuhari, Chiba, Japan, 2010.