2023-01-0788 Published 11 Apr 2023

Development and Evaluation of Comfort Assessment Approaches for Passengers in Autonomous Vehicles

Haotian Su, Johnell Brooks, and Yunyi Jia Clemson University

Citation: Su, H., Brooks, J., and Jia, Y., "Development and Evaluation of Comfort Assessment Approaches for Passengers in Autonomous Vehicles," SAE Technical Paper 2023-01-0788, 2023, doi:10.4271/2023-01-0788.

Received: 25 Oct 2022 Revised: 23 Jan 2023 Accepted: 28 Jan 2023

Abstract

Passenger comfort is a critical factor in user acceptance of autonomous vehicles (AVs). Despite existing methods for passenger comfort assessment, new approaches to assessing passenger comfort in AVs may be valuable to the automotive industry. In this paper, continuous pressing-based and discrete smartphone-based approaches for comfort assessment were designed and implemented in a user study. Participants used the two approaches to evaluate their comfort levels in an experimental study based on a high-fidelity autonomous driving simulator. Performances of the two approaches in assessing comfort levels were analyzed and compared. In general, the discrete approach showed better measurement repeatability and lower measurement bias than the continuous approach. The performance gap of the continuous approach could be reduced with proper post-processing measures. Discussions on the potential uses of the approaches were also raised.

Introduction

Despite the recent advancements in autonomous vehicles (AVs), user acceptance of AVs is still low. A survey [1] revealed a significant portion of respondents with low willingness-to-pay for AVs. Another survey [2] also indicated that higher levels of autonomy would lead to lower user acceptance and poorer user experience. One topic that is important in AVs to user acceptance is perceived comfort [3]. One study showed low overall consumer confidence regarding AVs [4]. Comfort when riding in and being in traffic with AVs were the two topics with the lowest confidence among all topics investigated in the survey. Therefore, it is crucial to investigate and improve the passenger comfort in AVs to work towards increasing user acceptance of AVs.

To study passenger comfort in AVs, it is necessary to collect passengers' perceived comfort levels efficiently and effectively. Various methods have been employed to measure passenger comfort, including the use of both subjective and objective metrics. A typical approach of subjective comfort assessments is through questionnaires with structured questions measuring one's attitude through the use of Likert scales [5, 3, 6, 7]. Other researchers used a mechanical interface with seven buttons corresponding to different levels of comfort to measure the construct [8]. An objective comfort measurement typically records an individual's physiological measurements. In [9], researchers used an electromyographic system to measure muscle activities and fatigue characteristics and used video systems and electrogoniometers to measure the angles of different joints on the human body, which were all helpful for comfort evaluation.

This study focused on subjective assessment approaches. Existing subjective assessment tools either use a discrete or continuous form. A discrete assessment tool can provide an ordinal or categorical output with discrete levels, whereas a continuous assessment tool generates a numeric value as the output. The Likert scale is a widely-used discrete assessment tool. The deployment of Likert scales typically uses a questionnaire or an interview format. However, questionnaires and interviews usually require interruption during the experimental manipulation. In [8], a discrete comfort level rating interface based on a 7-point Likert scale was used with a fixed time or travel range during the simulation of a semi-automated drive. The measurement was conducted within the drive, but the process could not capture real-time changes in the participants' comfort levels. A study by Azevedo-Sa [10] focused on a similar topic where trust estimation was assessed during semi-automated driving journeys. A 5-point Likert scale was used to measure the level of trust towards the semi-automated driving system during the experiment. Participants were required to manually take over the vehicle and stop at the shoulder of the road to complete the Likert scale each time after the system warned the participant of a potential collision. The discrete assessment tools in these studies were completed with a large time interval between two measured values.

Although the methods provided online measurement, they could not achieve real-time measurement that can capture the delicate variations in the measured values within the time interval being monitored.

Besides discrete assessment tools, several studies have employed continuous assessment tools for a subjective data measurement. Koay [11] proposed a hand-held device used during the assessment of passenger comfort in human-robot interactions (HRI). The participants were required to use a hand-held device to report their perceived comfort in continuous levels during different HRI tasks during the experiment. The position of a slider on the device was recorded as the continuous comfort measurement. Hartwich [12] used a handset control for the continuous assessment of driving comfort. The position of a trigger on the device was taken as the continuous comfort measurement. Both studies employed a hand-held device for continuous subjective measurement and collected real-time comfort measurements. Due to the inaccuracy in humans' control over the device, the performance of such approaches was doubtful. However, neither in the mentioned studies nor in any further studies have the researchers explored the performance of this type of hand-held device.

Despite existing methods to assess comfort, a thorough analysis on the performance of these discrete and continuous approaches for the context of autonomous vehicles is needed. Therefore, the goal of this study is to compare and evaluate different methods for assessing passenger comfort in AVs. We developed two approaches to assess passenger comfort in this study, a continuous approach using a pressing device and a discrete approach using a smartphone app. Inspired by the passenger experience of holding onto a door handle tightly during a stressful maneuver along with the hand-held devices described above [11, 12], we developed a hand-held device for the passenger to press in order to explore the utility of using button pressure force data as a continuous measurement of one's subjective level of comfort. This study defined comfort as a single dimension, with comfort and discomfort as the two endpoints of the continuum. A simulator-based experiment was conducted using a fully autonomous vehicle context (SAE Level 5 [13]). For the continuous comfort assessment approach using the pressing device, participants were instructed to press the button to indicate their perceived comfort with a "soft" press to represent feeling comfortable and a "hard" press to represent feelings of discomfort. A smartphone app allowed participants to report comfort level using a 7-point Likert scale for the discrete approach. The study analyzed and compared the measurement performance of the continuous and discrete assessment approaches.

In brief, the contributions of this study can be summarized as:

- Designed and implemented a continuous pressing-based approach and a discrete smartphone-based approach for passengers to intuitively assess comfort in autonomous vehicles.
- Compared the performance of the discrete and continuous comfort assessment approaches through experimental studies using a high-fidelity autonomous driving simulator.

Designs of Intuitive Comfort Assessment Approaches in Autonomous Vehicles

Continuous Assessment Approach Using a Pressing Device

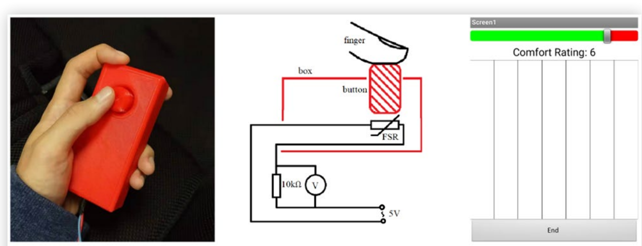
Inspired by the realistic passenger experience and the pressing devices used in [11, 12], a button box was designed and manufactured as an intuitive approach to capture the participants' pressing force as an expression of their comfort level, see Fig. 1. The button in our study allowed for the recording of the pressing force applied by the participant's thumb as the comfort level. We believed the pressing force mechanism was more intuitive to express one's level of comfort than the slider mechanism [11] and build upon the trigger mechanism [12].

A force-sensitive resistance (FSR) was placed in the box under the button to record the force applied by the participant. The resistance of the FSR would drop if a force was applied to it. A resistor was in series connected with the FSR in the measurement circuit. The voltage across the resistor was monitored so the resistance of the FSR could be calculated. By checking the calibration map of the FSR, the force applied could be obtained. A soft press represented feeling comfortable, while a hard press indicated discomfort. Since the voltage across the resistor was an analog signal, measurements from the pressing device were considered continuous measurements of comfort.

Discrete Assessment Approach Using a Smartphone App

A 7-point Likert scale was designed for participants to express their comfort level, with one representing feeling comfortable and seven indicating feeling very uncomfortable. Compared to the continuous approach, the Likert scale was a discrete

FIGURE 1 The continuous and discrete comfort assessment approaches studied in this paper. Partly reprinted from [14] with permission. ©2022 IEEE



(a) Picture of the button box. (b) Structure of the button box. (c) User interface of the app.

assessment approach where the comfort level was defined in seven progressive levels.

To collect the discrete comfort level, a smartphone app was developed. The user interface of the app is shown in Fig. 1(c). The participants could record their comfort level by operating the interface's slider. By pressing the 'End' button on the bottom of the screen, a log file containing the ratings from each time the user interacted with the slider was generated and sent to an assigned web server.

Designs of Experimental Studies and Data Collection on Comfort Assessment in Autonomous Vehicles

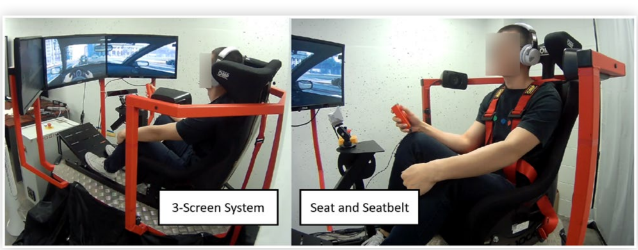
Participants

Based upon sample sizes from human factors studies [15] and the sample sizes used in relevant studies [16, 17, 8], ten participants (one female, nine males) between the ages of 22 to 52 years (M = 29.5 years, SD = 9.1 years) completed the study. The participants' years of driving experience ranged from 1 to 11 years (M = 4.4 years, SD = 3.8 years). All participants had an engineering background. The participants' experience with driver assistance systems included cruise control (N = 8) and adaptive cruise control (N = 3).

Driving Simulator

The simulation was conducted using a high-fidelity autonomous driving simulator. The participant was seated on a stationary platform during the experiment. The seat had two adjustable thigh support pads for comfort and a 4-point harness. Mounting points for a steering wheel and pedals were available, but given that the experiment was for an autonomous driving context, no steering wheel or pedals were installed. The driving simulator used three 27-inch monitors, shown in Fig. 2. The three monitors were aligned to each other with a fixed angle of 135° to compensate for the perspective distortion. Participants wore noise-canceling headphones to ensure they could hear both the road noises and communications with the experimenter.

FIGURE 2 High-fidelity autonomous driving simulator.



Driving Scenarios

This study focused on the evaluation of the two passenger comfort data collection methods. Since influential factors of passenger comfort in AVs were not the primary focus of the study, we created simple scenarios that kept participants' comfort evaluation process straight forward. All scenarios were created using a closed road course to eliminate confounds such as traffic, see Fig. 3. Simulated journeys were recorded for playback, and the length of each video was around 30 seconds.

Three sets of videos were created: *Acceleration – Deceleration Scenarios* (ADS), *Lane - Speed Keeping Scenarios* (LSKS), and *Collision Avoidance Scenarios* (CAS). In each video set, a 3×3 design was employed, in which two influential factors were considered, each with three configurations for a combination of nine videos. Table 1 shows the factors in the scenarios and the configurations of each factor. Table 2 shows how scenarios, denoted as S_1, S_2, ..., S_{27}, were formulated by combining two configurations of the factors. Detailed descriptions of the scenarios are given below.

Acceleration – Deceleration Scenarios: Maneuvers in these scenarios started by accelerating to 25 mph then maintaining the speed for up to five seconds. Then, the vehicle decelerated until coming to a complete stop. These scenarios were created by combining different acceleration levels and deceleration levels.

Lane - Speed Keeping Scenarios: Maneuvers started by accelerating to 45 mph with the medium acceleration level. Once the vehicle reached the target speed, lateral movements and speed variations occurred for five seconds. Then the vehicle decelerated to a complete stop at the medium deceleration level. LSKS were created by combining different lateral movement and speed variation levels.

Collision Avoidance Scenarios: Similar to LSKS, maneuvers started by accelerating using the medium level of acceleration to 45 mph. The vehicle maintained the speed until reaching the desired location to brake. After reaching the location, a specific brake force was applied, and the vehicle stopped with a desired clearance to the obstacle. These maneuvers were created by combining different braking forces and stopping clearances.

Definition of Passenger Comfort in AVs

In previous studies [18, 19], passenger comfort was defined in different ways. In this study, we captured passenger comfort in AVs. A different set of measures to assess passenger comfort was considered in this study compared to the studies mentioned above. A definition of passenger comfort in AVs was established to clarify the scope of investigation in this study.

- Definition: when riding in an AV, passenger comfort is a positive feeling of not being unsafe or unnatural resulting from the behaviors of the AV.

The scope of passenger comfort was limited to the behavioral factors of the vehicle in Table 1. There exist several studies suggesting that the behavioral pattern [20] or the driving style

FIGURE 3 Image of the vehicle and roadway from the virtual journey.

TABLE 1 Configurations of influential factors in the driving scenarios.

Factors	Configurations
f_1: Acceleration	Low Acceleration (LA)
	Medium Acceleration (MA)
	High Acceleration (HA)
f_2: Deceleration	Low Deceleration (LD)
	Medium Deceleration (MD)
	High Deceleration (HD)
f_3: Speed Variation	Low Speed Variation (LSV)
	Medium Speed Variation (MSV)
	High Speed Variation (HSV)
f_4: Lateral Movement	Low Lateral Movement (LLM)
	Medium Lateral Movement (MLM)
	High Lateral Movement (HLM)
f_5: Braking Force	Low Braking Force (LBF)
	Medium Braking Force (MBF)
	High Braking Force (HBF)
f_6: Clearance to Obstacle	Close Clearance to Obstacle (CCO)
	Medium Clearance to Obstacle (MCO)
	Far Clearance to Obstacle (FCO)

[21] affects passenger's perceived comfort or acceptance of AVs. These findings on passenger comfort in AVs led us to limit the scope of the passenger comfort considered in this study to focus only on comfort related to the behavioral factors of AVs.

Procedures

After providing consent, each participant completed a questionnaire to obtain demographic information. Then the participant was seated in the simulator and fastened the seatbelt. A training protocol was performed to help the participant get familiar with the continuous assessment device.

During the training process, the experimenter explained that the amount of force used to press the button should be considered on a one-to-seven scale where a one corresponds to a soft press representing feeling comfortable, while a seven, a hard press, indicated feeling uncomfortable. The experimenter explained to the participant that later during the driving scenarios experiment (DSE), the comfort level needed to be reported in both ways, using a one-to-seven number scale and the corresponding button press. The participant was also notified that only the peak pressing force value during the press after each assigned number would be recorded as the measurement of the continuous approach.

For training, the experimenter assigned a number ranging from one to seven, and the participant pressed the button with the force corresponding to that number. Each practice round started with a "one", followed by a "seven", this was to help serve as a calibration for the participants. Then the seven options were presented in random order. This process was repeated for three practice rounds for a total of 27 button presses.

The DSE stage followed. The participant was instructed to imagine themselves being in a SAE Level 5 AV during the journey. There was no driving-related task for the participant during the journey. The scenario sets were shown to the participant in the order of ADS, LSKS, then CAS. Each scenario set was viewed and rated for two consecutive rounds. Each round started with the most conservative scenario (S_1, S_{10}, S_{19}) and the most aggressive scenario (S_9, S_{18}, S_{27}) in the set. This was designed to be the calibration process where the participant experienced the most conservative and most aggressive rides within a scenario set. After the calibration, all nine videos within the scenario set were played in a random sequence. This process was repeated for each set of scenarios before moving on to the next set of scenarios. The participant provided the perceived comfort level during the video only after the video was played. The participant was asked to provide a comfort rating based on the video first with the continuous approach and then the discrete approach.

TABLE 2 The combination of configurations in each scenario.

Acceleration – Deceleration Scenarios				*Lane - Speed Keeping Scenarios*				*Collision Avoidance Scenarios*			
f_1 Scenarios f_2	**LD**	**MD**	**HD**	**f_3 Scenarios f_4**	**LLM**	**MLM**	**HLM**	**f_5 Scenarios f_6**	**FCO**	**MCO**	**CCO**
LA	S_1	S_2	S_3	LSV	S_{10}	S_{11}	S_{12}	LBF	S_{19}	S_{20}	S_{21}
MA	S_4	S_5	S_6	MSV	S_{13}	S_{14}	S_{15}	MBF	S_{22}	S_{23}	S_{24}
HA	S_7	S_8	S_9	HSV	S_{16}	S_{17}	S_{18}	HBF	S_{25}	S_{26}	S_{27}

We monitored the motion sickness of the participant with a questionnaire [22]. After each video, the experimenter would ask the participant if any sign of motion sickness was perceived. If uneasy feelings related to motion sickness occurred, the experimenter would suggest a break for the participant until the uneasy feeling disappeared. The results were used only for filtering out data affected by motion sickness. There was no experience related to motion sickness reported during the experiment.

The total length of the experiment was approximately one hour. The research protocol was approved by the Institutional Review Board at Clemson University (IRB2017-233).

Post-Processing of Data

The pressing force values collected with the continuous approach were processed with power functions to fit the corresponding Likert scale range. The function can be expressed by the equation below

$$C_c(F) = a \cdot e^{b \cdot F} + c$$
$$S.T. \begin{cases} C_c(F) = C_{max}, \text{ if } C_c(F) > C_{max} \\ C_c(F) = C_{min}, \text{ if } C_c(F) < C_{min} \end{cases} \quad (1)$$

where C_c is the continuous measurement result based on the raw pressing force value F, C_{max} and C_{min} are the maximal and minimal output values based on the Likert scale design, which in this study were 7 and 1, a, b, and c are function parameters to be determined. Each participant had a unique set of function parameters. The pressing force values and the corresponding Likert Scale values in the training process were used to fit each participant's parameters.

The function was applied to make the comparison between the continuous and discrete approaches feasible. A Likert Scale consists of uniformly distributed integers as measurements. However, participants could not apply uniformly changing pressing forces corresponding to the Likert Scale, and ranges of pressing force also varied across participants. Therefore, a function was fitted to map the pressing force data to the levels of the Likert Scale.

We also carried out different levels of down-scaling to the measurement results from the discrete approach. By performing the down-scaling process, some adjacent scales in the original 7-scale Likert scale were combined into one scale, and the original Likert scale was converted to one with fewer scales. New post-processing functions were also fitted based on the down-scaled Likert scales. The down-scaling process was carried out to examine how the measurement performance would change if the measurement resolution was decreased. Noise existed in self-reported comfort levels when the participant was occasionally confused between two adjacent levels in the Likert scale. The distance between adjacent scales enlarged with fewer scales, and we expected the participant to have a reduced chance of giving a mistaken scale. For the measurement from the continuous approach, the negative effect of the noise should also be reduced with a function fitted based on the more compact Likert scale after down-scaling.

Results and Analysis

Training Process Evaluation

The average measurements from the continuous assessment approach in the training process across participants are displayed in Fig. 4. Box plots representing the data collected during the training process for each participant are shown in Fig. 5. In Fig. 5, a total of nine boxes were included in each box plot, where each box includes the three rounds of training for the two calibration presses comfort level one ('cal1') and seven ('cal7'), followed by the seven random levels. The legends of the sub-figures in Fig. 5 are explained in the figure caption.

From visually observing Fig. 4 and 5, all participants were able to discriminate the two ends of the continuum, which were one and seven representing feeling comfortable and very uncomfortable, respectively. In addition, an upward trend was observed in the majority of the participants' plots. This suggested that the participants generally understood the continuous approach, where a harder press represented more discomfort.

Participants 1, 2, and 4 produced the cleanest plots by consistently increasing pressing force as the discomfort level increased. However, some participants occasionally pressed softer as the number increased, e.g. participant 7 and participant 9.

For most participants, the variability of the forces for the greater discomfort levels was more significant than the variability for the lower discomfort levels, and the most significant variability was observed in the calibration for comfort level seven.

Results and Analysis in Correlation

A function demonstrated in Eq. 1 was fitted based on the data collected from the training process for each participant. During the training process, the raw pressing force data was processed with the function to obtain the fitting results. Then the processed continuous measurements were compared with the corresponding discrete levels to examine the goodness of fitting. The Pearson's correlation coefficient R values were

FIGURE 4 Average continuous measurements across all participants during the training process.

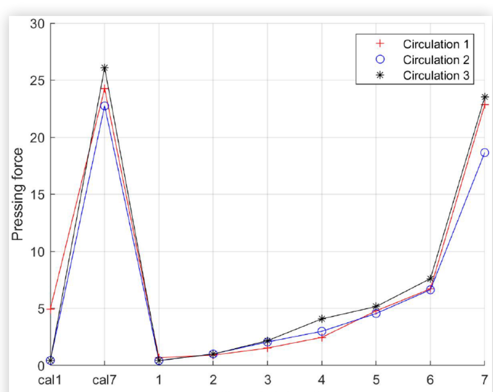
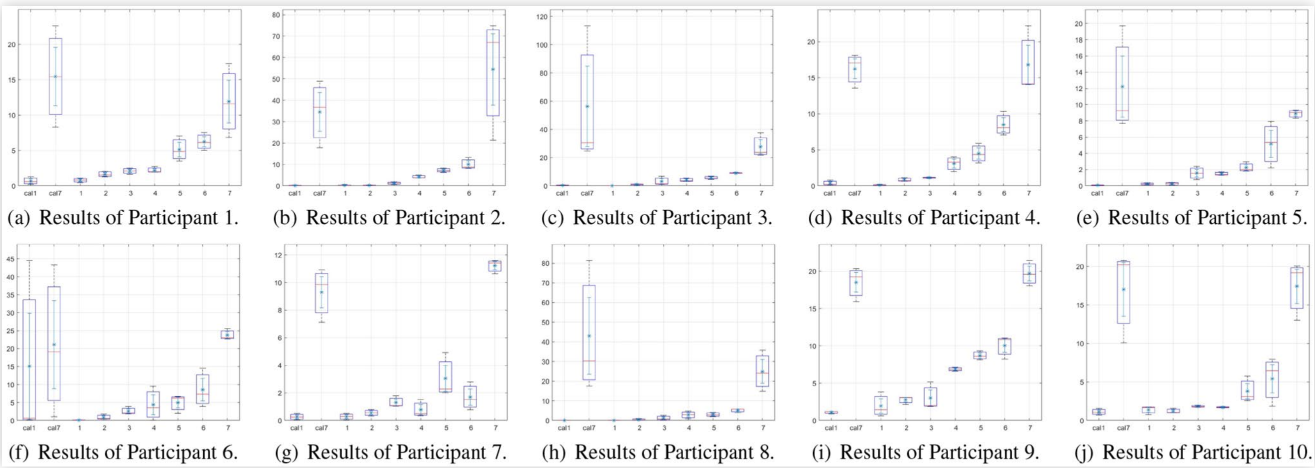


FIGURE 5 Statistics of the continuous measurement approach in the training process for the participants. The horizontal line in the box represents the median, and the asterisk is the mean value. The top and bottom edges of the box indicate the 75th and 25th percentiles. The whiskers extend out of the box mark the two extreme values, and the whiskers within the box stand for the standard error of the data.



(a) Results of Participant 1. (b) Results of Participant 2. (c) Results of Participant 3. (d) Results of Participant 4. (e) Results of Participant 5. (f) Results of Participant 6. (g) Results of Participant 7. (h) Results of Participant 8. (i) Results of Participant 9. (j) Results of Participant 10.

TABLE 3 Correlation analysis results in the training process and the driving scenarios experiment.

	Participant number									
Experimental stage	1	2	3	4	5	6	7	8	9	10
Training process	0.95	0.97	0.95	0.98	0.94	0.87	0.81	0.92	0.93	0.86
Driving scenarios experiment	0.84	0.93	0.89	0.82	0.92	0.80	0.86	0.84	0.76	0.37[a]

[a] $p = 0.007$

calculated for all participants. The evaluation results are displayed in Table 3.

The results implied a high correlation between the discrete levels and continuous measurement during the training process. All participants achieved R values of 0.8 or higher, which indicated a strong correlation between the discrete and continuous measurements. The correlation analysis results for all participants had p values below 0.001, which confirmed the statistical significance of the correlation analysis results.

Similar analysis between the continuous and discrete measurements from DSE was also carried out. Correlation coefficients were analyzed between the measurements from the two approaches. The results of the correlation analysis are displayed in Table 3.

For most participants, the measurements from the continuous and discrete approaches had a strong correlation. Nine participants achieved R values higher than 0.75 and p values of 0.001 or lower. For participant 10 with a low R value, the p value from the results was 0.007, which was still under 0.01 and indicated a significant correlation between the continuous and discrete measurements. Compared to the evaluation results from the training process, the R values for nine participants dropped to a lower value in DSE.

Results and Analysis in Repeatability and Bias

During this analyses, we performed different levels of down-scaling to the discrete approach. Three levels of down-scaling were carried out and created the 5-scale, 3-scale, and 2-scale setups for the continuous and discrete approaches. A function for the continuous approach was also fitted for each down-scaling level.

We defined repeatability as the ability to obtain repeated values when measuring the same item. In DSE, each scenario was experienced by the participant twice. Ideally, the participant should perceive the same comfort level in the two runs, and we should obtain identical measurements. Based on this understanding, we defined repeatability as the ratio of the two measurements from the same scenario, with the lower value as the numerator if a difference existed. A value closer to 1 would represent better repeatability.

The mean repeatability value across all scenarios was calculated for each participant. Results for different measurement methods and down-scaling levels were also calculated. Results are displayed in Table 4. The results showed that the discrete approach had an advantage in repeatability for all participants and all down-scaling levels over the continuous approach. As the down-scaling level increased, the mean repeatability values improved for both measurement approaches. For the continuous approach, the mean repeatability value across participants was 0.74 in the original scale setup and increased to 0.87 for the 3-scale and 2-scale setups. The performance gap between the two approaches reduced as the down-scaling level increased. The gaps between the two approaches were 0.08 for the original setup, 0.06 for the 5-scale setup, and 0.05 for the 3-scale and the 2-scale setups.

TABLE 4 Mean repeatability values across all scenarios for each participant based on the results from different measurement approaches.

Method	Participant number											
	1	2	3	4	5	6	7	8	9	10	Mean	SD
Continuous approach	0.78	0.76	0.79	0.69	0.78	0.69	0.76	0.75	0.77	0.61	0.74	0.05
Continuous approach (5-scale)	0.80	0.83	0.84	0.76	0.86	0.75	0.80	0.82	0.82	0.70	0.80	0.05
Continuous approach (3-scale)	0.86	0.90	0.90	0.84	0.91	0.83	0.89	0.89	0.90	0.89	0.87	0.04
Continuous approach (2-scale)	0.88	0.88	0.90	0.84	0.93	0.83	0.84	0.91	0.91	0.81	0.87	0.04
Discrete approach	0.83	0.84	0.86	0.76	0.81	0.77	0.78	0.90	0.90	0.76	0.82	0.05
Discrete approach (5-scale)	0.83	0.86	0.90	0.82	0.83	0.81	0.84	0.93	0.97	0.83	0.86	0.05
Discrete approach (3-scale)	0.89	0.90	0.93	0.87	0.88	0.92	0.92	0.96	0.99	0.94	0.92	0.04
Discrete approach (2-scale)	0.94	0.89	0.94	0.89	0.98	0.87	0.94	1.00	0.91	0.80	0.92	0.06

TABLE 5 Repeatability values and mean repeatability values of the continuous and discrete approaches for different participants.

Method	Low-discomfort scenarios		Mid-discomfort scenarios		Hi-discomfort scenarios	
	Mean	SD	Mean	SD	Mean	SD
Continuous approach	0.74	0.07	0.71	0.07	0.80	0.08
Continuous approach (5-scale)	0.83	0.06	0.78	0.04	0.81	0.05
Continuous approach (3-scale)	0.88	0.04	0.86	0.03	0.89	0.04
Continuous approach (2-scale)	0.91	0.04	0.85	0.02	0.87	0.04
Discrete approach	0.77	0.06	0.81	0.07	0.89	0.04
Discrete approach (5-scale)	0.83	0.06	0.86	0.04	0.90	0.02
Discrete approach (3-scale)	0.90	0.06	0.95	0.03	0.92	0.03
Discrete approach (2-scale)	0.94	0.03	0.90	0.02	0.90	0.08

To explore potential influential factors of measurement repeatability, we defined three categories of scenarios, low-, mid-, and hi-discomfort scenarios (L/M/HDS), based on the mean comfort level across participants. Each category was initially designed to contain nine scenarios. Considering the variant in the classification results based on continuous and discrete approaches, where one scenario was classified into different categories based on continuous and discrete assessment approaches, a scenario would only be classified into a category if the classification results from both approaches agreed with each other. The results of the repeatability analysis in different scenario groups are displayed in Table 5.

For both continuous and discrete approaches, an overall ascending trend in the mean repeatability value within the same scenario group could be recognized as the down-scaling level increased. For different scenario groups, it could be seen that the MDS had the lowest mean repeatability values for the results from the continuous approach. In contrast, the LDS had the lowest mean repeatability values for the discrete approach with different setups except for the 2-scale setup. Measurement bias represents the difference between the measured value and the ground-truth value. In this study, given that passenger comfort level was a subjective feeling which was challenging to obtain the ground-truth value, we defined the ground-truth comfort level as the mean value of the two measurements from the same scenario. Then the measurement bias would be half of the absolute error between the two measurements from the same scenario. Normalization was performed on the measurements from each method and down-scaling level to accommodate different measurement ranges.

The bias analysis was also carried out across all scenarios as a whole and in different groups of scenarios. The results of the mean bias value across all scenarios for each participant are displayed in Table 6. The results of the bias in different scenario groups are displayed in Table 7.

TABLE 6 Mean measurement bias values across all scenarios for each participant based on the results from different measurement approaches.

Method	Participant number											
	1	2	3	4	5	6	7	8	9	10	Mean	SD
Continuous approach	0.09	0.07	0.07	0.12	0.06	0.12	0.10	0.07	0.06	0.15	0.09	0.03
Continuous approach (5-scale)	0.08	0.06	0.06	0.10	0.05	0.10	0.08	0.06	0.05	0.12	0.08	0.03
Continuous approach (3-scale)	0.08	0.05	0.06	0.10	0.05	0.10	0.07	0.06	0.05	0.12	0.07	0.03
Continuous approach (2-scale)	0.10	0.09	0.08	0.13	0.06	0.13	0.13	0.07	0.06	0.17	0.10	0.04
Discrete approach	0.07	0.06	0.05	0.10	0.05	0.09	0.07	0.03	0.04	0.10	0.07	0.02
Discrete approach (5-scale)	0.07	0.06	0.04	0.09	0.06	0.08	0.06	0.03	0.02	0.07	0.06	0.02
Discrete approach (3-scale)	0.08	0.07	0.05	0.09	0.07	0.06	0.05	0.02	0.01	0.05	0.05	0.02
Discrete approach (2-scale)	0.06	0.11	0.06	0.11	0.02	0.13	0.06	0.00	0.10	0.21	0.09	0.03

TABLE 7 Measurement bias values and mean measurement bias values of the continuous and discrete approaches for different participants.

	Low-discomfort scenarios		Mid-discomfort scenarios		Hi-discomfort scenarios	
Method	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Continuous approach	0.07	0.03	0.10	0.02	0.09	0.03
Continuous approach (5-scale)	0.06	0.02	0.08	0.02	0.08	0.02
Continuous approach (3-scale)	0.06	0.03	0.08	0.02	0.07	0.03
Continuous approach (2-scale)	0.07	0.03	0.11	0.02	0.11	0.04
Discrete approach	0.07	0.02	0.07	0.02	0.05	0.02
Discrete approach (5-scale)	0.06	0.03	0.06	0.02	0.05	0.01
Discrete approach (3-scale)	0.06	0.04	0.03	0.02	0.07	0.02
Discrete approach (2-scale)	0.06	0.03	0.10	0.02	0.10	0.08

From Table 6, it could be observed that the discrete approach had a narrow performance advantage for most cases compared to the continuous approach. As the down-scaling level became higher, the mean bias values for both approaches dropped. However, when it reached the 2-scale setup, both approaches achieved the largest mean bias value among different down-scaling levels. The performance gap between the continuous and discrete approaches did not vary much as the down-scaling level changed.

Table 7 shows that mean bias values in LDS were the lowest among the three categories for the continuous approach at all down-scaling levels. However, no significant difference in mean bias values could be recognized across different scenario groups with the discrete approach under the original and 5-scale setups. The bias value with the 3-scale discrete approach in MDS was significantly smaller than the other two scenario groups. The bias values with the continuous and discrete approaches in the 2-scale setup grew significantly compared to the 3-scale setup in MDS and HDS.

Discussions

By directly observing the data from the training process in Fig. 5, it can be concluded that the participants can understand how to use the button pressing to rate their comfort level. However, some overlap in forces corresponding to different discrete levels were observed. This suggested that the continuous approach's overall accuracy might be poorer than the discrete approach. It can also be that the raw pressing force was not linearly correlated with the discrete levels.

Post-processing of the raw pressing force data was carried out to improve the data quality from the continuous approach. A function described in Eq. (1) was fitted for each participant based on the training process results. To examine the goodness of the fit, the Pearson's correlation coefficient R values were calculated based on the fit results. The results in Table 3 showed that continuous measurements had a strong linear correlation with discrete levels after the post-processing with the function. This proved the effectiveness of the post-processing method.

In both repeatability and bias analyses, we found that the discrete approach had an overall performance advantage compared to the continuous approach. Our observation from the training process data was within our expectations. As we performed the higher level of down-scaling, we found that the repeatability increased for both measurement approaches, the bias decreased for both approaches until reaching the 2-scale setup, and the repeatability performance gap between the two approaches decreased. This suggested that by applying the proper level of down-scaling, the repeatability performance of the continuous approach could be acceptably good compared to the discrete approach.

Despite the excellent performance in terms of repeatability and measurement bias, it should be noted that the discrete approach does not provide real-time measurement. This limitation has limited the application of the discrete approach to the collection of the holistic feelings after a session within an experiment [8, 10]. In comparison, the continuous approach showed the ability to provide real-time measurement of comfort [11, 12]. The participants of these studies provided their comfort levels in real-time during the experiment. The focus of this study was to compare the measurement performance of the two approaches in terms of measurement repeatability and bias. Therefore, the continuous approach was not implemented in real-time during the videos to compromise the limitation of the discrete approach. Considering the circumstance when real-time passenger comfort during an AV journey is desired, the continuous approach proposed in this study would be an appropriate method for the task.

Given the circumstance when the real-time comfort assessment is possible, the immediate feedback of a passenger's comfort level will help researchers and engineers identify vehicle behavioral factors that influence comfort in AVs more precisely. The information can be employed by the AV to quickly update the decision-making policy and avoid similar behaviors that made the passenger feel uncomfortable. The merits of such real-time comfort assessment are to be further explored in future studies.

Conclusion

In this study, continuous and discrete comfort assessment approaches were developed and implemented in a simulator-based user study. Participants were instructed to rate their comfort levels in simulated AV rides using both assessment approaches. The measurements from the two approaches were compared to each other to evaluate their performances. Based on the comparison results, we summarized the performance characteristics of the two approaches. The discrete approach showed good repeatability and small bias in measurement and could be used to collect holistic feelings after an experimental session. With proper post-processing measures, the continuous approach can achieve similar performance in measurement repeatability and bias as the discrete approach, and it may be used in real-time data collection of complex human feelings.

References

1. Kyriakidis, M., Happee, R., and de Winter, J.C., "Public Opinion on Automated Driving: Results of an International Questionnaire among 5000 Respondents," *Transportation Research Part F: Traffic Psychology and Behaviour* 32 (2015): 127 -140.
2. Rödel, C., Stadler, S., Meschtscherjakov, A., and Tscheligi, M., "Towards Autonomous Cars: The Effect of Autonomy Levels on Acceptance and User Experience," in *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '14, (New York, NY, USA), p. 1–8, Association for Computing Machinery, 2014.
3. Bellem, H., Klüver, M., Schrauf, M., Schöner, H.-P. et al., "Can we Study Autonomous Driving Comfort in Moving-Base Driving Simulators? A Validation Study," *Human Factors* 59, no. 3 (2017): 442-456.
4. Power, J.D., "Mobility Pipe Dreams? J.D. Power and Survey Monkey Uncover Shaky Consumer Confidence About the Future," Available at https://www.jdpower.com/business/pressreleases/2019-mobility-confidence-index-study-fueledsurveymonkey-audience/ (2019/10/23), 2019.
5. Koo, J., Kwac, J., Ju, W., Steinert, M. et al., "Why Did my Car Just Do that? Explaining Semi-Autonomous Driving Actions to Improve Driver Understanding, Trust, and Performance," *International Journal on Interactive Design and Manufacturing (IJIDeM)* 9, no. 4 (2015): 269-275.
6. Muir, B.M., "Trust between Humans and Machines, and the Design of Decision Aids," *International Journal of Man-Machine Studies* 27, no. 5-6 (1987): 527-539.
7. Hart S.G. and Staveland L.E., "Development of Nasa-Tlx (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*, vol. 52, pp. 139-183, Elsevier, 1988.
8. De Vos, A.P., Theeuwes, J., Hoekstra, W., and Coëmet, M.J., "Behavioral Aspects of Automatic Vehicle Guidance: Relationship between Headway and Driver Comfort," *Transportation Research Record* 1573, no. 1 (1997): 17-22.
9. Reynolds, H., Rayes, K., Eppler, M., Neal, D. et al., "Development of a Vehicle Laboratory to Investigate Driver Comfort from Physical Measurements," SAE Technical Paper 960480, 1996, https://doi.org/10.4271/960480.
10. Azevedo-Sa, H., Jayaraman, S.K., Esterwood, C.T., Yang, X.J. et al., "Real-Time Estimation of Drivers' Trust in Automated Driving Systems," *International Journal of Social Robotics* (2020): 1-17.
11. Koay, K.L., Walters, M.L., and Dautenhahn, K., "Methodological Issues Using a Comfort Level Device in Human-Robot Interactions," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pp. 359–364, IEEE, 2005.
12. Hartwich, F., Beggiato, M., and Krems, J.F., "Driving Comfort, Enjoyment and Acceptance of Automated Driving–Effects of Drivers' Age and Driving Style Familiarity," *Ergonomics* 61, no. 8 (2018): 1017-1032.
13. SAE International, "Taxonomy and Definitions for Terms Related to on-Road Motor Vehicle Automated Driving Systems," *SAE Standard J* 3016 (2014): 1-16.
14. Su, H. and Jia, Y., "Study of Human Comfort in Autonomous Vehicles Using Wearable Sensors," *IEEE Transactions on Intelligent Transportation Systems* 23, no. 8 (2022): 11490-11504.
15. Sauro, J. and Lewis, J.R., *Quantifying the User Experience: Practical Statistics for User Research* (Morgan Kaufmann, 2016)
16. Haspiel J., Du N., Meyerson J., Robert L.P. Jr., Tilbury D., Yang X.J., and Pradhan A.K., "Explanations and Expectations: Trust Building in Automated Vehicles," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI '18*, New York, NY, USA, p. 119-120, Association for Computing Machinery, 2018.
17. Yamakoshi, T., Yamakoshi, K., Tanaka, S., Nogawa, M., et al., "A Preliminary Study on Driver's Stress Index Using A New Method Based on Differential Skin Temperature Measurement," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 722–725, 2007.
18. Oborne, D., "Vibration and Passenger Comfort," *Applied Ergonomics* 8, no. 2 (1977): 97-101.
19. De Looze, M.P., Kuijt-Evers, L.F., and Van Dieen, J., "Sitting Comfort and Discomfort and the Relationships with Objective Measures," *Ergonomics* 46, no. 10 (2003): 985-997.
20. Elbanhawi, M., Simic, M., and Jazar, R., "In the Passenger Seat: Investigating Ride Comfort Measures in Autonomous Cars," *IEEE Intelligent Transportation Systems Magazine* 7, no. 3 (2015): 4-17.
21. Hartwich, F., Beggiato, M., and Krems, J.F., "Driving Comfort, Enjoyment and Acceptance of Automated Driving–Effects of Drivers' Age and Driving Style Familiarity," *Ergonomics* 61, no. 8 (2018): 1017-1032.
22. Gianaros, P.J., Muth, E.R., Mordkoff, J.T., Levine, M.E. et al., "A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness," *Aviation, Space, and Environmental Medicine* 72, no. 2 (2001): 115-119.

Contact Information

Yunyi Jia, Ph.D.
Mailing address: 4 Research Dr, Greenville, SC 29607
yunyij@clemson.edu

Acknowledgments

This work was partially supported by the National Science Foundation under Grants CNS-1755771 and IIS-1845779.

ISSN 0148-7191