Design of a Physiology-based Adaptive Virtual Reality Driving Platform for Individuals with ASD

DAYI BIAN, Department of Electrical Engineering and Computer Science, Vanderbilt University
JOSHUA WADE, Department of Mechanical Engineering, Vanderbilt University
AMY SWANSON, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University
AMY WEITLAUF, Department of Pediatrics, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University
ZACHARY WARREN, Department of Pediatrics, Department of Psychiatry, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University
NILANJAN SARKAR, Department of Electrical Engineering and Computer Science, Department of Mechanical Engineering, Vanderbilt University

Driving is essential for many people in developed countries to achieve independence. Individuals with Autism Spectrum Disorder (ASD), in addition to having social skill deficits, may experience difficulty in learning to drive due to deficits in attention-shifting, performing sequential tasks, integrating visual-motor responses, and coordinating motor response. Lacking confidence and feeling anxiety further exacerbates these concerns. While there is a growing body of research regarding assessment of driving behavior or comparisons of driving behaviors between individuals with and without ASD, there is a lack of driving simulator that is catered toward the needs of individuals with ASD. We present the development of a novel closed-loop adaptive Virtual Reality (VR) driving simulator for individuals with ASD that can infer one's engagement based on his/her physiological responses and adapts driving task difficulty based on engagement level in real-time. We believe that this simulator will provide opportunities for learning driving skills in a safe and individualized environment to individuals with ASD and help them with independent living. We also conducted a small user study with teenagers with ASD to demonstrate the feasibility and tolerability of such a driving simulator. Preliminary results showed that the participants found the engagement-sensitive system more engaging and more enjoyable than a purely performance-sensitive system. These findings could support future work into driving simulator technologies, which could provide opportunities to practice driving skills in cost-effective, supportive, and safe environments.

CCS Concepts: • **Human-centered computing** → **Accessibility design and evaluation methods**; *Usability testing*; • **Computing methodologies** → *Machine learning*;

This work was supported by the National Institute of Health Grant No. 1R01MH091102-01A1.
Authors' addresses: D. Bian, Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, 37212; email: dayi.bian@vanderbilt.edu; J. Wade and N. Sarkar, Department of Mechanical Engineering, Nashville, TN, 37212; emails: {joshua.w.wade, nilanjan.sarkar}@vanderbilt.edu; A. Swanson, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University, Nashville, TN, 37212; email: aamy.r.swanson@vumc.org; A. Weitlauf, Department of Pediatrics, TRIAD, Vanderbilt University, Nashville, TN, 37212; email: amy.s.weitlauf@vumc.org; Z. Warren, Department of Pediatrics, Department of Psychiatry, TRIAD, Vanderbilt University, Nashville, TN, 37212; email: zachary.e.warren@vumc.org.

1936-7228/2019/02-ART2 $15.00
https://doi.org/10.1145/3301498

Additional Key Words and Phrases: Driving, physiological signal, affective computing, dynamic difficulty adjustment (DDA), autism intervention, ASD

ACM Reference format:

Dayi Bian, Joshua Wade, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. 2019. Design of a Physiology-based Adaptive Virtual Reality Driving Platform for Individuals with ASD. *ACM Trans. Access. Comput.* 12, 1, Article 2 (February 2019), 24 pages.

https://doi.org/10.1145/3301498

1 INTRODUCTION

Autism Spectrum Disorder (ASD) has a prevalence rate of 1 in 59 among children in the U.S. (Baio et al. 2018) and is associated with high familial and societal cost (Ganz 2006; Chasson et al. 2007). ASD is characterized by social interaction deficits, verbal and non-verbal communication skill deficits, repetitive behaviors, and fixed interests (American Psychiatric Association 2013). Although there is no single agreed-upon treatment or known cure for ASD, there is growing consensus that adaptive training and educational intervention programs can improve long-term outcome for individuals with ASD and their families (Buescher et al. 2014). While substantial effort has been dedicated to improving social skill deficits (White et al. 2007), much less research has focused on other daily living tasks that allow for increased independence in adulthood. One critical skill necessary for attaining independence, securing a job, and maintaining social relationships is driving, but it is only recently that researchers have turned their attention towards the issue of driving in the ASD population.

In the U.S., the demographic of 16- to 20-year-old drivers is three times more likely to be involved in a fatal motor vehicle crash than older and presumably more experienced drivers (Centers for Disease Control and Prevention 2012). Young drivers are disproportionately represented in rates of highway fatalities; 9% of fatalities are young drivers, despite this group constituting only 5.4% of all drivers on the road (National Highway Traffic Safety Administration 2018). Over 2,000 teens die each year from motor vehicle crashes, and total annual fatalities of drivers was 37,461 in 2016 (National Highway Traffic Safety Administration 2018). Mounting evidence suggests that learning to drive is particularly difficult for many individuals with ASD due to deficits in attention-shifting, performing sequential tasks, integrating visual-motor responses, and coordinating motor response (Hill 2004; Huang et al. 2012). In addition, lack of confidence—both from the driver with ASD in themselves and the parent of a driver with ASD in their child's driving ability—further exacerbates these concerns. Parent respondents to a survey by Cox et al. (2012) reported that the characteristics associated with ASD exerted a moderately to extremely negative influence on their child's ability to drive safely. Daly et al. (2014) surveyed licensed driving adults with and without ASD about their driving histories. They found that individuals with ASD reported being older at the age of licensure, spending less time driving, feeling less confident about their driving ability, and experiencing greater number of traffic violations than neurotypical peers.

Given that some 1.4 million individuals with ASD are estimated to be driving already—with approximately 60,000 more becoming age-eligible each year—there is an urgent need to develop intelligent driving intervention tools capable of addressing both the performance and processing-level issues of drivers with ASD (Howden and Meyer 2010; Cox et al. 2012; Huang et al. 2012; National Highway Traffic Safety Administration 2016). An intelligent driving intervention tool could not only provide driving training sessions to the users, but could also adapt the training strategy (e.g., altering task difficulty) in real time based on the users' feedback (e.g., performance, engagement state). Parents of individuals with ASD suggested having driving-like experiences like driving video games before trying to drive a car is one of the most useful strategies in teaching driving skills (Cox et al. 2012). We believe such driving intervention technology provides an assistive system that can help individuals with ASD to build confidence toward on-road driving.

The primary goal of this research is to design a Virtual Reality (VR)-based driving platform for individuals with ASD that can adapt task difficulty in real-time based not only on task performance but also on participant engagement. We present a closed-loop physiology-based engagement recognition system embedded within a dynamic difficulty adjustment mechanism that controls driving task presentation in the driving platform. We also present results from a small user study of teenagers with ASD to demonstrate the feasibility and tolerability of such a driving platform. Moreover, by utilizing the participants' performance data and physiological data, we provide methods to document the participants' driving performance and engagement level, which makes designing of long-term driving sessions and tracking the users' performance and engagement level over time more accessible. However, a long-term skill learning and skill generalization study is beyond the scope of the current work.

VR technology offers promise as an intervention modality for people with ASD because it can provide a controllable, replicable, and safe learning environment (Strickland 1997). Compared to reality, VR creates a less hazardous and more forgiving method for practicing potentially dangerous skills associated with daily life, such as driving. It can also be individualized not only to the general needs of specific learners but also to their day-to-day levels of engagement. By exploiting these advantages, VR-based driving simulators could provide a variety of modes for teaching at-risk drivers who are not ready for on-road assessment.

Several previous studies have examined the generalizability of driving simulators for training real-world driving skills. In a neurotypical population, Shechtman et al. (2009) did not find significant differences in the type or frequency of driving errors when comparing on-road and simulated driving assessments. This suggests that the driving skills learned in driving simulators can be generalized or transferred to the road under the same testing conditions. Reimer et al. (2013) assessed driving behaviors in young adults with ASD and found that they displayed a nominally higher and unvaried heart rate compared to controls; they also tended to focus visual attention away from the high stimulus area of the roadway. Our previous work also discovered similar differences in visual attention (Wade et al. 2014). Other assessment studies have shown that individuals with ASD performed more poorly on selective attention, visual-motor integration (Classen et al. 2013), increased working memory tasks (Cox et al. 2016) and predicting time-to-arrival (Sheppard et al. 2016), highlighting the need for systems that can rapidly assess specific types of driver errors and adjust training goals accordingly.

VR-based skill training systems have a demonstrated value for training specific skills in individuals with ASD. However, a simple performance-based system without appropriate feedback to the user may not be optimal for enhancing and maximizing learning. Managing task difficulty to keep the user optimally challenged is beneficial for designing an effective skill training system (Hunicke 2005; Adams 2008; Liu et al. 2008). Task difficulty can affect cognitive workload and induce a variety of affective states (Giakoumis et al. 2010). A task that is beyond a user's capability could be overwhelming and may cause anxiety while a task that does not fully utilize the capacity of a user might result in boredom. When people feel anxious or bored, as opposed to engaged, they tend to lose focus on their task, learn less, be less productive and are more likely to make mistakes (Pekrun et al. 2010). Thus, the difficulty of the task should be adjusted to keep the user optimally challenged based on the ability and affective state of the user. As depicted in Figure 1, keeping the user in the "flow state" will help maintain a positive emotional state during a task by avoiding anxiety and boredom (Nacke and Lindley 2008; Fairclough et al. 2013). When users played Tetris at different levels of difficulty, Chanel et al. found that easy task difficulty was related to boredom and hard task difficulty elicited anxiety, while medium difficulty was associated with engagement

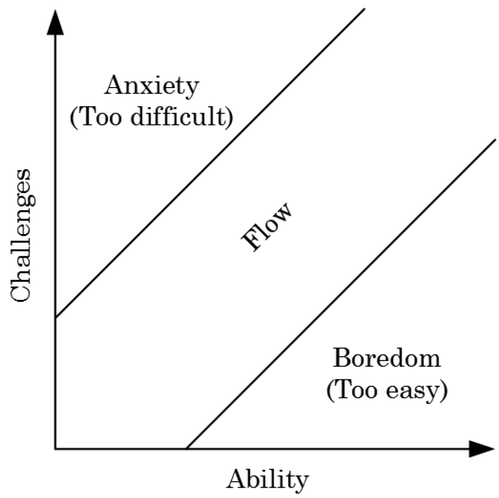

Fig. 1. Model of flow.

(Chanel et al. 2011). This suggests that adjusting task difficulty based on user boredom and anxiety might help him/her become and remain engaged in the task. When considering optimum levels of engagement, we also wanted to assess for enjoyment of participants. However engaged an individual may be, if they find a task aversive, then they may be less likely to persist or re-attempt it over time.

To exploit these benefits, task designers have attempted to keep the user in the flow state by using a mechanism called Dynamic Difficulty Adjustment (DDA), which automatically adjusts task difficulty based on a model of the user's skills (Adams 2008). For example, recent research has shown that game users demonstrate faster gains in performance and feel a greater sense of control when the difficulty adjusts according to their current skill level (Jegers 2007; Stripling et al. 2007; Orvis et al. 2008). Game users also feel more immersed in game scenarios that adjust the game difficulty according to their current performance, as opposed to those that simply increase the game difficulty over the course of the game (Stripling et al. 2007). Outside of gaming applications, DDA can also be used in applications like human machine interface design to facilitate the progression of the user from novice to expert (Bederson 2004).

Besides these performance metrics, the user's affective states can also be utilized to build a model of the user's skills. These affective states can be derived from physiological responses (Sarkar 2002). Several studies have investigated how these physiological responses impact performance and flow. Compared to strictly performance-based feedback, physiology-based affective state feedback can be more efficient in providing optimal challenge to users, keeping them in a state of flow and improving their performance (Rani et al. 2005; Lahiri et al. 2013; Zhou et al. 2015).

As can be seen from the aforementioned literature survey, most existing studies of driving intervention for ASD focus on assessment of driving performance and comparison between individuals with and without ASD. To the best of our knowledge, there is no study incorporating physiology-based DDA into a driving simulation platform for ASD intervention. The goal of this work is to bridge this gap and develop a driving skill training platform that can (1) allow real-time measurement of engagement-related physiological signals while the user takes part in the driving task; (2) predict the user's engagement level based on real-time physiological signals, and (3) alter the task difficulty based on both the user's engagement level and performance metrics. We hypothesized that this engagement-based DDA driving skill training platform (1) would be tolerated by the individuals with ASD, (2) would generate more task engagement, and (3) would result in greater levels of enjoyment for participants when compared to the participants who used only the performance-based DDA modality of the driving skill training platform. While we did not expect to observe any significant changes in driving performance in only one experimental session of the user study—as it takes time to become proficient in driving—we believe that providing tasks that are inherently enjoyable, and perhaps rewarding, could increase the likelihood that individuals, particularly those on the autism spectrum, would be willing to attempt and engage in learning activities. Because on-road driving training option for individuals with ASD is rare and expensive, a driving simulator like the one presented in the article could supplement existing services and provide more opportunities to learn driving skills in a cost-effective and safe and forgiving environment. Ultimately, we believe that such a system will have the potential to make driver training more accessible to a large number of individuals with ASD.

In terms of understanding the conceptual framework for such intervention, we have published several works (Warren et al. 2015; Zheng et al. 2015; Bekele et al. 2016; Zhang et al. 2018) based on the underlying understanding that individuals with ASD show fundamental differences in social communication and interaction as well as differences in behavior and interest that impede their ability to successfully complete and engage in numerous life tasks—from social learning to complex adaptive learning (i.e., driving). Given these core deficits—inclusive of communication challenges that may limit dynamic adjustment and feedback within traditional learning environments—systems capable of introducing physiological and attentional metrics that can be used to create richer learning experiences may be of substantial value.

This article is a substantial extension of our previous work (Bian et al. 2016). We have significantly expanded the presentation and the implementation of the closed-loop driving system as well as conducted a new set of study with completely new set of results to demonstrate the feasibility and tolerability of such a system.

The rest of the article is organized as follows. Section 2 presents the system design of the VR-based adaptive driving environment. Sections 3 and 4 present the methodology and the results of the user study, respectively. Finally, we summarize the contributions of this work in Section 5.

2 SYSTEM DESIGN

The proposed VR-based Driving Environment with Adaptive Response technology (VDEAR) is comprised of (1) a VR-based driving task module; (2) a real-time physiological data acquisition module; and (3) an individualized dynamic difficulty adjustment module that embeds a real-time physiology-based engagement prediction mechanism (Figure 2).

2.1 VR-based Driving Environment

We used "narrow field of view" version of desktop VR applications for three reasons. First, they potentially minimize cyber sickness—an important consideration for people with ASD, because of their atypical sensory perception (Bellani et al. 2011). Second, the ultimate goal of this study is to develop an affordable version of driving intervention platform that can be more accessible by individuals with ASD. "Wide field of view" versions would increase the cost significantly. Finally, although not used in the current work, a remote eye tracker informed the design of the narrow field of view because it cannot be calibrated to more than one monitor. Having access to driving simulation software is a prerequisite for developing a driving intervention program. There are a range of high-quality off-the-shelf driving simulation systems that can be used in the assessment of driving behaviors. However, these tools are not suitable for our study because they do not provide access to the source code. Access to the source code is necessary for designing a closed-loop system that utilizes information from sensors and driving performance.

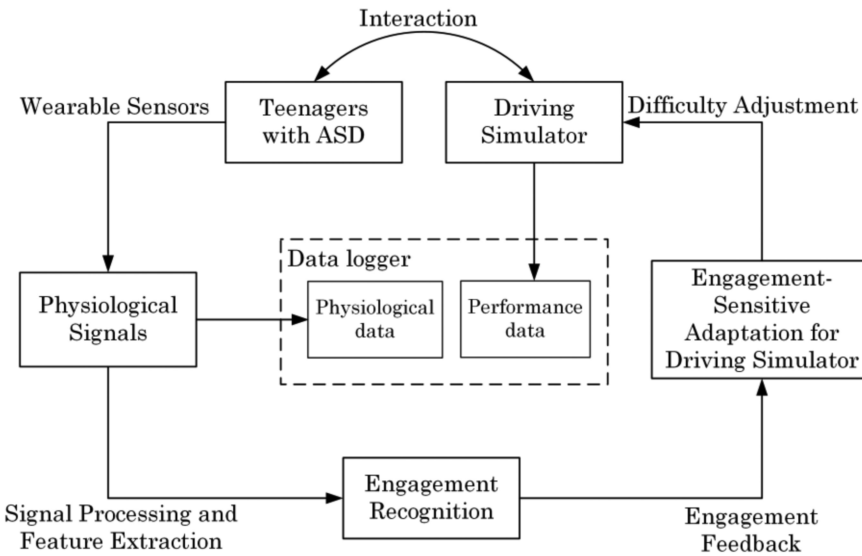

Fig. 2. VDEAR System framework overview.

Fig. 3. Driving simulator developed in our lab.

As a result, we have utilized a driving simulator that was recently designed in our laboratory (Figure 3) (Bian et al. 2013). A Logitech G27 steering wheel controller was used to control the virtual agent vehicle in the virtual driving environment. Models in the virtual driving environment, such as traffic lights, stop signs, and vehicles, were developed with the modeling tools ESRI CityEngine (www.esri.com/cityengine) and Autodesk Maya (www.autodesk.com/maya). The game development platform Unity3D (www.unity3d.com), was used to implement the system logic. A total of five difficulty levels were chosen for this adaptive driving environment. These difficulty levels were tested and validated in our previous work (Wade et al. 2014). Two sets of parameters including road conditions and environmental changes were manipulated to produce five difficulty levels. We designed the difficulty levels in such a way that they are distinguishable but not abrupt, so the participant would not feel sudden changes in the way that they manipulated the driving

Table 1. Difficulty Parameters

Group	Parameters	Domain	Description
Road conditions	Changes in responsiveness of brake pedal	[0.35, 1]	Value 1 means normal; smaller value dampens the input, indicates slippery road, which increases the difficulty
	Changes in responsiveness of the accelerator pedal	[1, 1.5]	Value 1 means normal; larger value amplifies the input, indicates downhill, which increases the difficulty
	Changes in responsiveness of the steering wheel	[1, 3.75]	Value 1 means normal; larger value amplifies the input, indicates uphill, which increases the difficulty
Environment	Intensity of light in the environment	[0.01, 0.5]	Smaller value indicates darker environment, which increases the difficulty
	Speed of agent vehicles	[0.85, 1.75]	Larger value indicates more aggressive driving of the agent vehicles, which increases the difficulty

controllers. Specifically, to provide different road conditions, we created different slopes of the road (e.g., uphill, downhill, and even road), different road textures (e.g., normal road and raining road), and different turns (e.g., normal turn and sharp turns). When the participants drive through different road conditions, they usually feel the changes of the responsiveness of the brake pedal, the accelerator pedal, and the steering wheel. The environmental changes were manipulated by changing the intensity of light in the environment and the speed of agent vehicles. The details of the difficulty parameters were listed in Table 1.

The VR system was modeled as a video game consisting of a set number of different trials. Each trial was designed to train a specific driving skill such as speed maintenance, turning, merging, and following traffic laws. At the beginning of each trial, there was a voice-based instruction, such as "Turn right," "Merge into highway," "Pass the vehicle in front," and so on, telling the participant how to progress through a given trial. A directional arrow graphic on the top right side of the screen served to provide navigational information to the participant. Accompanying audio-based prompts were also given when the drivers approached an upcoming turn (e.g., "Right turn ahead"). While the information provided by the navigation system may have served to offload some of the cognitive demands on the driver, it was an essential component of the task design as it ensured that drivers remained on the course in which predefined tasks were setup. In addition, drivers had to be on alert in order not to miss turns, for example, passing them by at high speeds. It is to be noted that as more and more auto manufacturers make navigation systems standard even for low-priced models, this feature in our system makes the VR vehicle more realistic. Each trial included within it a *critical region* within which the driving performance was monitored to provide "Failure" or "Success" feedback relevant to the content of the trial. For a "Success," the participant was awarded five points at the same time that congratulatory audio clip was played. The total points were displayed on the top left side of the screen. For a "Fail," participants were told the reason for their failure, the task was reset to the starting point of this trial's critical region, and no points were awarded for the failed trial. When the participant's vehicle was reset to the prior starting point, the system would also give detailed instructions on how to be successful on subsequent attempts, such as, "Stop before the stop sign."

Performance data, which included the turning angle of the driving wheel, the pressure on the gas and brake pedals, the speed of the vehicle, location of the vehicle, and detailed information

Table 2. Performance Metrics

Category	Good (3 points)	Fair (2 points)	Inadequate (1 point)
Number of failures (n)	n=0	n=1	n>1
Difference between driver speed and speed limit (D)	D<5	5<=D<=10	D>10
Steering wheel manipulation pattern (Kurtosis (sk) and Skewness (ss))	Small deviation from "good driving"	Medium deviation from "good driving"	Large deviation from "good driving"
Gas pedal pressing pattern (Mean (gm) and Kurtosis (gk))	Small deviation from "good driving"	Medium deviation from "good driving"	Large deviation from "good driving"

of the failure, were logged in both text (.txt) and comma-separated values (.csv) file formats. The sampling frequency of the data were the same as the frame rate of the virtual environment, which was typically 60Hz. The logged data were used in real-time performance assessment and offline data analysis.

Three-minute performance data were used to assess the user's performance in real time. As can be seen from Table 2, performance metrics included number of failures, difference between driver speed and speed limit, steering wheel manipulation pattern (kurtosis and skewness of the values) and gas pedal pressing pattern (mean and kurtosis of the values). We chose these driving features based on feature analysis of the driving data from our previous driving studies that provided the most discriminating information in assessing driving performance (Wade et al. 2014, 2015). Because the driving course was predefined, we utilized data from our previous study that were indicative of driving skill to assess the quality of driving. Value changes due to different road conditions (e.g., uphill and downhill) were normalized by using the responsiveness scale in Table 1. Each subcategory of performance metrics has three possible outcomes: Good (worth 3 points), Fair (2 points), and Inadequate (1 points). For example, a participant could score 3 points for gas pedal pressing but only 1 point for steering wheel manipulation. If a participant scored 50% of the maximum possible points within the 3min performance window, then the overall performance was considered "Good"; otherwise, this was considered as "Poor."

2.2 Physiological Data Acquisition Module

The physiological data were collected using the BIOPAC MP150 physiological data acquisition system (www.biopac.com) with a sampling rate of 1,000Hz. Using the hardware API provided by BIOPAC, we developed a customized physiological data acquisition program, in which we integrated socket-based communication with the driving program to record event information with automated time stamps (Bian et al. 2016). In this study, three physiological signals, photoplethysmogram (PPG), galvanic skin response (GSR), and respiration (RSP), were investigated. We chose these signals because they contain important information about the state of one's task engagement as shown in the literature (Healey 2000; Rainville et al. 2006; Bian et al. 2015). These signals were measured by using light-weight, non-invasive wireless sensors. PPG and GSR were measured from toes of left foot instead of fingers to reduce the interference to the driving wheel manipulation and the motion artifact from driving. PPG measures the blood volume in participant's middle toe. This measurement can also be used to compute the heart rate (HR) by identifying local maxima (i.e., heart beats), and heart rate variability (HRV). Blood pressure and HRV correlate with engagement (Kim and André 2008). GSR provides a measure of the resistance of the skin by positioning two electrodes on the distal phalanges of the index and the ring toe. This resistance decreases with an increase of perspiration, which has been shown to be associated with task engagement

(Pecchinenda 1996). Respiration was measured by using a respiration belt tied around the participant's abdomen. Slow respiration is often linked to relaxation while irregular rhythm, quick variations, and cessation of respiration relate with high level of arousal (Rainville et al. 2006).

The physiological signals acquired from BIOPAC were raw signals contaminated by motion artifacts and utility frequency. Thus, signal preprocessing was required before extracting features. First, outliers were removed from GSR, PPG, and RSP signals and were interpolated with the mean value of the adjacent data points. A Notch filter with a cutoff frequency 60Hz was used to remove the 60Hz utility frequency from these three signals. Notch filter is a band-stop filter with a narrow stopband that can cut the loss of useful frequency component in the process of filtering. Considering different properties of each signal, highpass and lowpass filters with different cutoff frequencies were used as discussed below:

Tonic and phasic components of GSR were decomposed separately from the raw signal. Tonic skin conductance is the baseline level of skin conductance and is generally referred to as skin conductance level (SCL). Phasic skin conductance is the part of the signal that changes when task-related events (e.g., task failure or success) take place. The signal was first filtered by a lowpass filter with a cutoff frequency of 0.5 Hz to remove noise. Then tonic component was acquired by using a highpass filter with a cutoff frequency of 0.05Hz. Phasic component was then calculated by deducting tonic component from preprocessed signal.

PPG is a low-frequency signal and can easily be contaminated by motion effect. A highpass filter with a cutoff frequency of 1.1Hz and a lowpass filter with a cutoff frequency of 3.6Hz were used to remove noise.

Respiration is also a low-frequency signal. A highpass filter with a cutoff frequency of 0.05Hz and lowpass filter with a cutoff frequency of 0.35 were used to remove noise.

Note that we recognized noise and motion artifacts by transforming the signals into frequency domains. By doing so, we found that some of the frequency component should not be there for certain signals (e.g., GSR typically should not contain high frequency signal components; a 60Hz frequency spike within signals usually indicates utility frequency noise). The noise and motion artifacts were removed by using the above-mentioned filters to remove the unwanted frequency bands from each signal. The cutoff frequencies of the filters were found empirically and were derived from our specific dataset (as recorded by BIOPAC MP150) and may not be applicable to other datasets, although this process could be replicated using different sample-specific parameters. The outliers in the signals were spikes generated by the recording device. To remove outliers, we ran a program to detect the outlier when the maximum absolute value of a data segment exceeded a certain value, and then removed those outliers and interpolated the removed values with mean value around that data segment. This resulted in approximately 1% data loss. After outlier removal and filtering, the data were down sampled to 100Hz. Subsampling can significantly reduce computational time of feature extraction, which is important in a real-time closed-loop system.

2.3 Online Engagement Detection Module

Physiological data and engagement ratings from our previous driving experiments (Bian et al. 2015) were used to build the engagement model. Then the model was used in the current closed-loop system for real-time engagement detection.

2.3.1 Offline Engagement Model Building. In our previous work where participants with ASD were engaged in driving tasks using the VR-based driving simulator (Bian et al. 2015), eight channels of physiological signals were examined for offline analysis to build an engagement detection model using machine learning algorithms. Subjective reports of the participants' engagement level

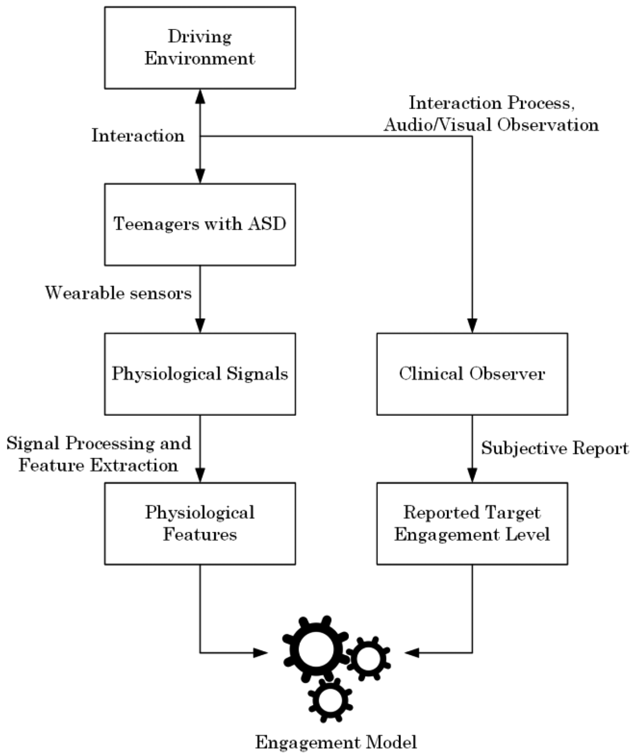

Fig. 4. The procedure for building an engagement model based on data from previous driving experiments.

from a clinical observer were used as the ground truth (Figure 4). The purpose of this model building was to build an engagement detection model that mimics a clinical observer's observations of engagement. For the purpose of developing this model and piloting this system, we utilized two observers who were highly trained in behavioral assessment and observation, especially for individuals with ASD. These observers rated engagement using a 9-point Likert scale at 3min intervals during sessions. To reduce risk of bias, approximately 20% of sessions were double-coded for reliability, and the observers resolved disagreements through consensus and discussion. The inter-rater correlation coefficient for engagement ratings was 0.9. Twenty participants' data from these previous experiments were used to build the engagement model for the current study. Note that the participants for the current study are all different from the participants for the previous study. Each participant finished six sessions in different days and each session lasted about 60min. Three-minute baseline data were collected to normalize the data. The physiological data from each session were segmented and labeled using 3min window. We chose the 3min window for two reasons: (1) physiological signals (especially the GSR signal) are relatively slow, and thus it requires a certain amount of time to reflect the change of affective states; and (2) after consulting with our psychology team, we believe that for this specific driving task, 3min is an appropriate time window for the participant to feel engaged/not engaged with the task for the purposes of using this information to alter the system. Note that this time window can be adjusted based on different tasks but must be consistent with the time window used to build the engagement model.

In the current study, to improve computational speed that is needed for the real-time closed-loop DDA mechanism, we chose only PPG, RSP, and GSR signals, the three physiological signals

Table 3. Physiological Features used in the Engagement Model

Physiological signal	Feature	Unit of measurement
Photoplethysmogram (PPG)	Mean of heart rate variability	Ms
	Mean of amplitude of the peak values	μV
Respiration (RSP)	Mean amplitude	No unit
	Standard deviation of amplitude	
	Subband spectral entropy (Spectral entropy calculated from frequency band 0.05Hz to 0.35Hz)	
	Mean of peak-valley magnitude	
	Standard deviation of peak-valley magnitude	
Galvanic Skin Response (GSR)	Rate of phasic activity	Response peaks/s
	Standard deviation of tonic activity level	No unit
	Slope of tonic activity	μS/s

Table 4. Classification Results by using Different Machine Learning Algorithms

Algorithm	Parameters	Accuracy
Multilayer Perceptron	hiddenLayers: 6; learningRate: 0.3; momentum: 0.2	81.94%
Random Forest	maxDepth: unlimited; numTrees: 100	84.72%
Support Vector Machine	kernel: radial basis function; loss: 0.1; degree: 3	80.67%
BayesNet	Estimator: SimpleEstimator; K2 search algorithm	72.95%
NaiveBayes	Numeric estimator precision values were chosen based on analysis of the training data	75.93%
J48 DecisionTree	The minimum number of instances per leaf was 2, 1- of 3-folds data were used for reduced-error pruning	80.37%

that had the most correlation with engagement to train the engagement detection model. Several features from these three signals were extracted as shown in Table 3.

After feature extraction, several well-known machine learning algorithms, which have shown success in emotion recognition tasks (Jerritta et al. 2011), were explored to build and train an engagement detection model. The engagement detection model produced binary output, which was either "High Engaged" or "Low Engaged." The Waikato Environment for Knowledge Analysis (WEKA) (Hall et al. 2009) was used to build and compare these models. The machine learning algorithms listed in Table 4 were examined and a group 10-fold cross-validation was used to assess the generalization of the models. Group 10-fold cross-validation was used here to make sure that each fold contains all the samples from one participant. In the end, Random Forest-based engagement detection model was chosen, since it yielded the greatest accuracy.

2.3.2 Online Engagement Detection. Because engagement was elicited in the same context (i.e., the driving simulator task), inter-participant variability was expected to be low (Stemmler et al. 2001). Nevertheless, to further reduce this variability, 3min of baseline physiological data were collected for each participant at the beginning of the experiment that was used to normalize the data. The driving task program and the engagement detection module communicated via a TCP socket over a local area network (LAN). When defined events (e.g., start/end of a trial, task failure/success) occurred, a JSON (http://www.json.org) string containing a time stamp and event

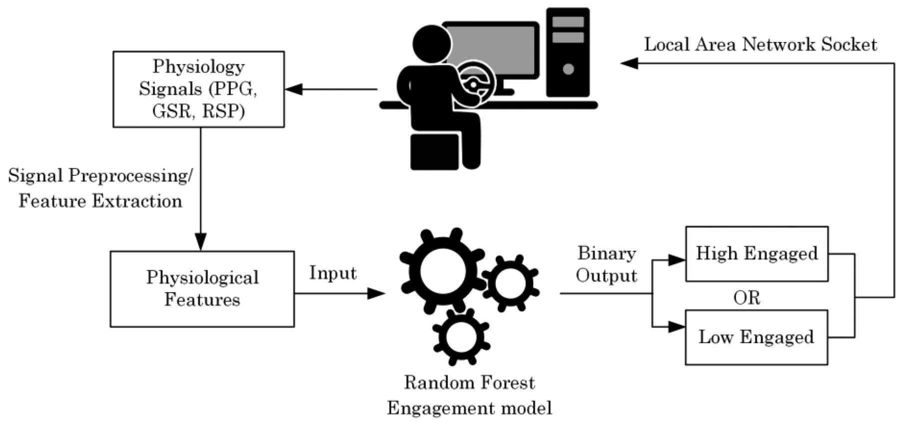

Fig. 5. Online engagement detection module diagram.

message was sent to the physiological data acquisition module. All event information and physiological data were recorded for offline analysis. Every 3min, the driving task sent an event to trigger engagement detection. The engagement detection module acquired 3min of data before this trigger. The 3min data were preprocessed to remove electrical noise and motion artifacts. After that, the 10 selected features were extracted and baseline features were subtracted from the features to offset environmental and participant differences. Subsequently, these features were fed into the previously established model to predict the engagement level. A binary label, "High Engaged" or "Low Engaged," was sent to the driving task program via the socket, which was then utilized by the difficulty adjustment module to make the decision about switching the difficulty levels (Figure 5). By measuring the execution time of the engagement detection code, we observed that the offline analysis for engagement detection took less than 100ms. The data collection resumed immediately after the engagement detection, a minimal enough delay to avoid any problems for our application. The difficulty adjustment process took place in another computer (i.e., in the main task computer) and thus did not cause any delay for the data collection.

2.4 Difficulty Adjustment Module

We developed two different difficulty adjustment strategies for the proposed driving system. One strategy used only the participant's performance metric to adjust the driving task difficulty, while the other strategy utilized both the participant's performance metric and engagement level to make the adjustment. Within the framework of this study, the independent variable was the difficulty adjustment mechanisms we described below, which was determined before each experiment.

2.4.1 Performance-Sensitive System (PS). For the PS, a task-switching mechanism adjusts the difficulty states solely based on the participant's performance. When a participant's performance is "Good" (Case 1), the task progression continues step-wise as the task difficulty level increases unless it reaches the most difficult level (Figure 6). However, if a participant's performance is "Poor" (Case 2), then the task progression continues step-wise by decreasing the task difficulty level unless it reaches the easiest level (Table 5, Algorithm 1).

2.4.2 Engagement-Sensitive System (ES). For the ES, the task difficulty manipulation is not only based on participant's performance but also on his/her engagement level. We fuse the participant's performance metric and engagement level to make the switching decision. In two cases where engagement and performance metrics agree with one another, the switching strategy is

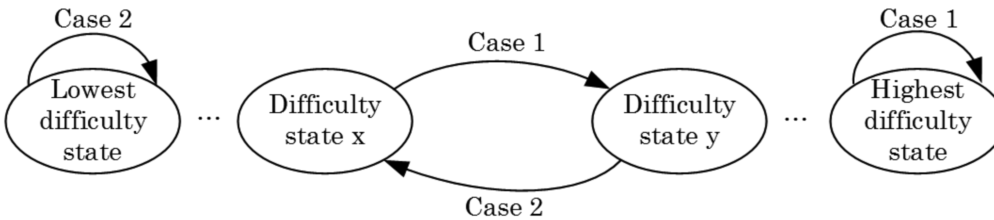

Fig. 6. Performance-Sensitive System difficulty switching FSM (Only draw four difficulty states for simplicity, assuming difficulty state y is more difficult than x.).

Table 5. Performance-Sensitive System Switching Cases

	Performance	**Action**
Case 1	Good	Increase difficulty
Case 2	Poor	Decrease difficulty

ALGORITHM 1: Performance-sensitive Difficulty Switching Action Computation

Input: Participant's performance: *perf*, current difficulty state: *diff_state*
Output: Difficulty switching action
repeat every 3min:
  if *perf == "Good"*
    if *diff_state* reaches *highest_diff_state*
      stay_the_same;
    else
      increase_difficulty;
    end
  else if *perf == "Poor"*
    if *diff_state* reaches *lowest_diff_state*
      stay_the_same;
    else
      decrease_difficulty;
    end
  end
until *task_end*;

straightforward. If a participant is "High Engaged" and his/her performance is "Good," then the system increases the difficulty level based on the finite state machine representation (Figure 7). However, if a participant is "Low Engaged" and his/her performance is "Poor," then the system decreases the difficulty level. However, in the other two cases, where engagement and the performance metric do not agree with each other, the switching strategy gives priority to the performance metric. For Case 2, in which engagement is "High" but performance is "Poor," the system recommends decreasing the difficulty level. For Case 3, when a participant is "Low Engaged" but his/her performance is "Good," the difficulty level remains the same and waits until the next trial for potential adjustment. At the next adjustment point, if the participant is still "Low Engaged" and his/her performance is "Good," then the system decreases the difficulty level (Table 6, Algorithm 2). While there are alternative interpretations for different cases (refer Section 5), we involved a clinical psychology team to make such decisions and to provide a clear path for subsequent decisions within the learning chain.

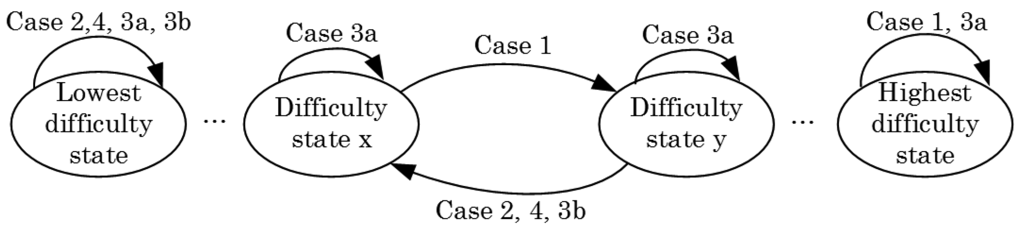

Fig. 7. Engagement-Sensitive System difficulty switching FSM (Only four difficulty states are drawn for simplicity, assuming difficulty state y is more difficult than x.).

Table 6. Engagement-Sensitive System Switching Cases

	Engagement	Performance	Action
Case 1	High	Good	Increase difficulty
Case 2	High	Poor	Decrease difficulty
Case 3a/3b	Low	Good	Same/Decrease difficulty
Case 4	Low	Poor	Decrease difficulty

3 METHODS AND PROCEDURE

3.1 Experimental Setup

VDEAR was run on a server-grade computer that could provide high-quality graphical rendering, while participants' peripheral physiological signals were acquired and processed in parallel on a separate computer. Both computers communicated over a LAN using socket-based connection. The VR driving task was presented on a 24in. LCD monitor (at a resolution of 1980 × 1080). Participants sat on a play seat and interacted with the driving environment using a Logitech G27 driving controller. The experiment was conducted in a laboratory with two rooms separated by a one-way mirror for observation. The researcher sat in the outer room to observe the experiment.

3.2 Participants

We recruited 23 teenagers (21 males and 2 females) with ASD for this phase of study. This high number of male participants is in keeping with epidemiological data indicating significantly elevated prevalence for males as compared to females (Baio et al. 2018). All participants had a clinical diagnosis of ASD from a licensed clinical psychologist as well as scores at or above clinical cutoff on the Autism Diagnostic Observation Schedule (Table 7) (Lord et al. 2000). The participants either had a learners' permit or was pursuing a permit. The Institutional Review Board (IRB) approval for conducting the experiment was sought and obtained. In the end, 20 participants' data were used in this study. Three participants (all males) were excluded from the study due to compromised data or equipment failure.

3.3 Procedure

The participants were randomly assigned to either PS or ES groups. Each participant visited the lab once and completed a 90min session. At the start of the session, physiological sensors were placed on the participant's body by an experienced researcher. Participants watched a tutorial video that discussed basic traffic rules as well as manipulation of the driving interface. After the tutorial, the participant was asked to remain calm and relaxed for 3min during which baseline physiological data were collected. To get familiar with the driving interface, participants also

ALGORITHM 2: Engagement-sensitive Difficulty Switching Action Computation

```
Input: Participant's performance: perf, participant's engagement level: engage, current difficulty state:
    diff_state
Output: Difficulty switching action
next = False;
repeat every 3min:
    if perf == "Good" and engage == "High"
        if diff_state reaches highest_diff_state
          stay_the_same;
        else
          increase_difficulty;
        end
    else if perf == "Poor" and engage == "Low"
        if diff_state reaches lowest_diff_state
          stay_the_same;
        else
          decrease_difficulty;
        end
    else if perf == "Good" and engage == "Low"
        if next == False or diff_state reaches lowest_diff_state
          stay_the_same;
          next = not(next);
        else
          decrease_difficulty;
          next = not(next);
    else if perf == "Poor" and engage == "High"
        if diff_state reaches lowest_diff_state
          stay_the_same;
        else
          decrease_difficulty;
        end
    end
until task_end;
```

received 3min of practice driving in which there were no pedestrians and no other vehicles in the VR environment. After the 3min practice drive, participants began the driving assignment at medium difficulty. Through the assignment, participants were required to follow the navigation directions and obey traffic rules. Disobeying any traffic rules (i.e., running a red light) resulted in a performance failure. In addition, the system provided instructions and feedback to the individuals regarding performance such that even if one was novice in terms of traffic laws, he/she would be able to follow the rules of the game. Task difficulty was adjusted every 3min in real time. The difficulty adjustment strategy was based on the participant's group assignment. The duration of the assignment was typically between 30 and 40min, depending on the participant's performance. A post-task survey was completed once the participant finished the assignment (Figure 8).

4 RESULTS

4.1 Tolerability of the VDEAR System

We excluded the data from three participants because (1) physiological sensors became detached midway through the experiment, resulting in diminished signal quality; (2) the equipment failed,

Table 7. The 20 Participant Profiles from PS and ES Groups

Demographic Information	PS Group Mean (SD)	ES Group Mean (SD)	*t*	*p*
Age	15.18 (1.27)	15.19 (1.89)	−0.02	0.99
Driving Permit Status	4 (Yes)/6 (No)	3 (Yes)/7 (No)	NA	NA
Sex	9M/1F	9M/1F	NA	NA
ADOS New Algorithm Total Raw Score	13.57 (6.37)	12.5 (2.83)	0.41	0.69
ADOS Severity Score	7.71 (1.89)	7.25 (1.39)	0.54	0.60
SRS-2 Total Raw score	100.63 (30.64)	89.89 (31.38)	0.71	0.49
SRS-2 T-score	76.63 (12.16)	72.00 (10.71)	−0.83	0.42
SCQ Lifetime Total Score	21.44 (8.47)	15.9 (6.54)	1.58	0.13
IQ	88.6 (23.38)	101.38 (20.21)	1.01	0.34

ADOS = Autism Diagnostic Observation Schedule, it is an instrument for diagnosing and assessing autism.
SRS = Social Responsiveness Scale (Constantino and Gruber 2007), it measures social ability of children from 4 years to 18 years old and is used primary with individuals with ASD.
SCQ = Social Communication Questionnaire (Rutter et al. 2003), it is one tool clinicians use when screening an individual for ASD and is a measure for caregivers to complete.
IQ = Intelligence Quotient.
NA = Not Applicable.

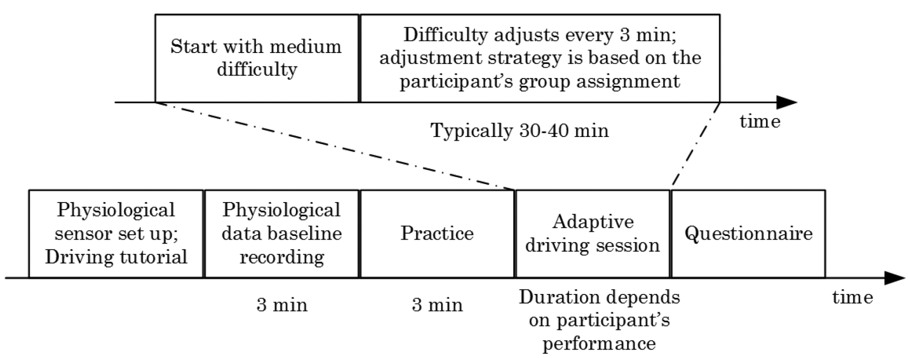

Fig. 8. Procedure of the experiment.

thus we terminated the experiment. All participants who completed the study reported that they found VDEAR to be engaging and enjoyable. None felt uncomfortable wearing the physiological sensors. The average duration of the completed driving tasks was 34min (SD=9).

4.2 Offline Analysis of Physiological Data

The physiological data (PPG, GSR, and respiration signals) were collected from all participants. For those in ES condition, physiological data were used to detect real-time engagement level to help the system determine the appropriate task difficulty level. In addition, we also computed the average engagement level for each group. To get a reliable estimate of the average engagement level for each participant, we used a 3min window size and 1min step size to segment the physiological data into a certain number of samples. As explained in Section 2.3.1, a number of features were extracted from these data and were normalized using the baseline features. These features were then used as input to the engagement model that we had built earlier. The outputs of the engagement model were the probabilities of two classes, "High Engaged" and "Low Engaged," and the sum of these two probabilities was 1. We used the probabilities of the "High Engaged" class as the engagement index of the participants. Thus, a higher engagement index reflects a higher engagement level.

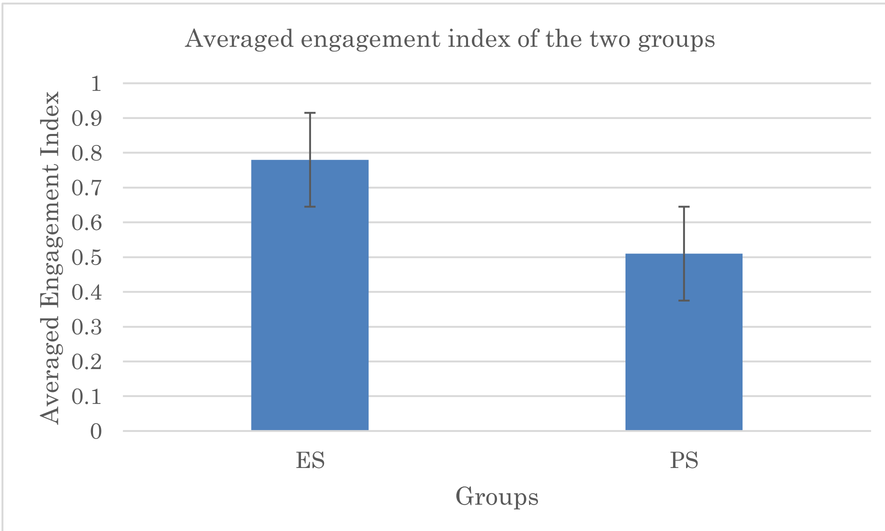

Fig. 9. Average engagement index of the two groups.

The average engagement index for each participant was computed using the following equation:

$$\overline{EI} = \frac{\sum EI}{N}, \tag{1}$$

where EI is the engagement index for each 3min sample and N is the 3min physiological data sample size for each participant.

Figure 9 shows the average engagement index for the two groups. The t test shows that the average engagement index for the ES group ($M = 0.78$, $SD = 0.18$) was statistically significantly higher than the average engagement index for the PS group ($M = 0.51$, $SD = 0.25$), ($t = -2.46$, $p = 0.03$). We also found a large effect size, Cohen's $d = 0.85$, for this difference in engagement level.

4.3 Performance Data Analysis

Although the current study was not designed to investigate whether ES-based training was more effective than PS-based training—which will require a systematic long-term user study—we did analyze participants' performance for this study to get a preliminary understanding. We could not use raw performance scores directly to assess the performance of each participant because it is not correct to conclude that two participants with the same raw performance score in different difficulty levels performed equally well. We therefore normalized the raw performance score data by assigning a weight to each difficulty level. The following equation was used to calculate the weight for each difficulty level:

$$w_i = \frac{N_i \times S_{\max,\, i}}{\sum S_{raw,\, i}}, \tag{2}$$

where w_i is the weight, $S_{raw,i}$ is the raw performance score, N_i is the sample size, and $S_{max,i}$ is the maximum possible score, in difficulty level i. The general intuition behind this method is to give greater weight to those levels with lower average raw score because they are more difficult. Then the mean weighted score for each participant was computed by using the following equation:

$$\overline{S} = \frac{\sum w_i S_{raw,\, i}}{N}, \tag{3}$$

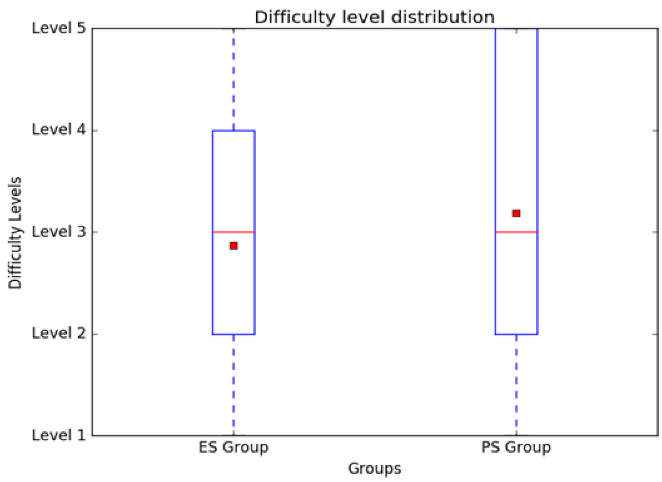

Fig. 10. Difficulty level distribution for ES and PS groups.

Table 8. Results of Questionnaire

Questions	Mean score of ES group	Mean score of PS group	p	Cohen's d
Overall difficulty of the driving task	3	3.5	0.17	0.66
*Whether noticed difficulty changes?	7 (Yes)/10	9 (Yes)/10	NA	NA
How do you like the difficulty adjustment?	4.4	3.5	**0.005**	1.45
Whether the adjustment is helpful?	4.3	4.1	0.54	0.28
Overall enjoyment	4.3	3.8	0.18	0.62
Overall frustration	1.9	2.3	0.42	0.37
Overall boredom	1.8	1.7	0.81	0.11

*This is a Yes or No question. All the other questions use the 5-Likert scale, 1 means the least and 5 means the most.

where S is the weighted score and N is the sample size for each participant. The mean weighted score for the ES group (M = 12.25, SD = 1.12) was slightly higher than that for the PS group (M = 11.92, SD = 1.24). Although the difference was not statistically significant ($t = -0.63$, p = 0.53, Cohen's $d = 0.28$), an adequately powered study might demonstrate a trend towards better training performance using the ES-based system.

In addition, we explored the distribution of the difficulty levels for both the ES and the PS group. As seen in Figure 10, participants in the PS group tended, though not significantly, to progress to a higher difficulty level as compared to those in the ES group ($t = -1.79$, $p = 0.07$, Cohen's $d = 0.23$). This result agrees with the self-report about the overall task difficulty discussed in the following section.

4.4 Questionnaire Results

A questionnaire evaluating participant responses to the system was designed using a 5-point Likert scale. The questions, the mean values from both groups, t test results and effect size (Cohen's d) are listed in the Table 8.

As seen in Table 7, most of the participants (16 of 20) from both groups noticed the difficulty adjustment in the driving task. Within those who noticed the difficulty adjustment, participants from ES group liked the difficulty adjustment more. Although other results did not show statistically significant difference, it is interesting to see that ES group reported lower overall difficulty, thought the difficulty adjustment was more helpful than the PS group, experienced less frustration and enjoyed the interaction more. Also, because the driving simulator was designed to be interesting for teenagers on the autism spectrum, and because participants attended only a single training session, the study design did not afford opportunities for variability in boredom levels. As a result, it is not surprising for us to see that both groups reported similar low boredom.

5 DISCUSSION AND CONCLUSION

This work presents the creation and pilot validation of a novel platform, VR-based Driving Environment with Adaptive Response technology (VDEAR), to provide an assistive system for the individuals with ASD to build confidence toward on-road driving. VDEAR capitalizes on the affinity of many people with ASD for technology while also providing for a closed-loop system of data capture, analysis, and system adjustment. Twenty of 23 participants completed the driving task, and performance data as well as physiological data were properly recorded for offline analysis. All participants who completed the study reported positive experiences with the system. They all tolerated the physiological sensors, an important finding for adolescents with ASD who may have sensory sensitivities to wearable devices.

During offline analysis, we fed the physiological data into the engagement model to detect the average engagement level of each participant and found that participants in the ES group had statistically significantly higher engagement than participants in the PS group. In addition, participants in the ES group subjectively reported that they liked the difficulty adjustment more than participants in the PS group. This further suggests that the physiology-based DDA mechanism may be more effective at keeping the user in the "flow state" when compared to the performance-based DDA mechanism. No differences were found in performance data, however, with participants in ES group achieving similar performance as participants in PS group. Also, participants in the PS group spent more time in higher difficulty levels in spite of the fact that they were in lower engagement levels. Although we did not expect to observe performance improvement for this one-session study, the ability to document the driving performance as well as engagement level makes designing of long-term driving sessions and assessing users' performance and engagement level over time more feasible using the VDEAR system. We believe that such a system will have the potential to make driver training more accessible to a large number of individuals with ASD.

It can be inferred from the above results that in a one-session study, the physiology-based DDA mechanism could create higher engagement in the participants and result in similar rates of success compared to a performance-only-based DDA. This could be potentially beneficial for long-term driving training programs, since the participants will likely be more willing to participate in multiple driving sessions that utilize their engagement feedback, as opposed to the one in which only performance is a factor. While we expect that an engaging and enjoyable training system will foster skill learning, it is beyond the scope of the current work to investigate skill improvement using a longitudinal user study.

This novel driving skill training platform does not rely on the user's input to adjust the system difficulty, which is potentially advantageous because individuals with ASD may not have the self-awareness and cognitive capability to provide such information. Although our system's engagement model must be developed beforehand, in the future, we can utilize real-time affective computing methods to update the model to make the system generalizable. Moreover, with the rapid development in the field of wearable and less invasive physiological sensors (e.g., smart watches), this kind of adaptive system, in which a user's affective cues will become easier to implement, will be more reliable and commonplace.

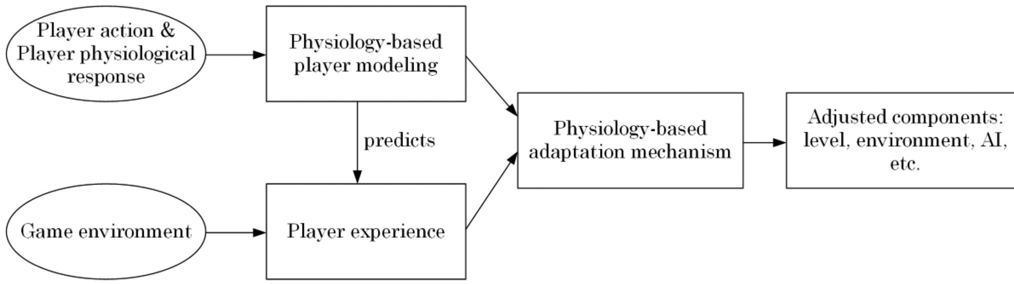

Fig. 11. Physiology-based adaptation mechanism in game design.

Although the engagement-sensitive adaptive driving task presented in this article is designed for validated among the ASD population, this driving intervention platform could also be appropriate to train other young drivers who suffer from a degree of anxiety. In addition, the presented physiology-based DDA mechanism is independent of the driving environment and thus can be integrated into other gaming or training environments. The engagement information is detected separately and sent to the main program in JSON format over a LAN using standard internet protocols. Once we defined the difficulty switching logic in the game environment, this physiology-based adaptation mechanism can be integrated seamlessly. The following figure (Figure 11) depicts the framework for designing a physiology-based adaptive game environment.

Note that the current work is limited in interpreting and modeling behaviors based on single estimates of both Engagement (E) and Performance (P). It is quite accurate to state that iterations of E and P in their specific combination may be reflective of different learning states (for example low E and Low P could be reflective of boredom or could be reflective of a task that is too challenging). In this capacity, the true value of intelligent, dynamic decision-making systems will come in the form of interpretation of chains of decision and reactions. That is, it becomes clearer to differentiate boredom and challenge if the manipulation hypothesis is not born out in subsequent modeled state. Such delineation and interpretation of sophisticated chains of learning are somewhat beyond the scope of the current work; however, we should mention that these modeling decisions were seen as the best initial hypotheses of manipulation such that an intelligent system could dynamically interpret and discern changes in learning over time. Importantly, we involved a clinical psychology team with requisite experience in Applied Behavior Analysis to inform our initial interpretations regarding potential functions of behavior and decisions that would yield clarity over time. In this context, it was determined that making initial classification decisions on Low E across adequate and low P would provide a clear path for subsequent decisions within the learning chain. Another limitation of the current work is that our engagement system utilized engagement ratings both to manipulate system characteristics and to chart some aspects of within system change. The ultimate benefit of such a system would be tied to independent and objective outcomes. In this respect, it is promising that the participant ratings noted benefits of the system. The ultimate test of such a system would be objective benefits independently documented with direct relevance to driving/learning to drive. In the future an additional measure of engagement will strengthen the work. One threat to the validity of our proposed work is the lack of quantitative measure of participants' driving-related experience before they participated in the study. Different driving-related experience might lead to different confidence levels when completing the driving task. In future work, we will have participants complete a questionnaire about their past driving-related experience. For example, we might want to know whether/how much they have played driving-related video games, whether/how much they have behind-the-wheel driving experience, and so on. Note that we did not observe any performance gains in the ES as compared to the PS in the one-session user study. Driving skill improvement requires greater exposure to training and it is unsurprising not to have observed any gains in just one session. A more systematic longitudinal study will be needed to objectively assess whether ES-based training will result in higher gains in driving skills. What is encouraging, however, is that participants in the ES were found to be significantly more engaged than those in the PS, which may indicate superior compliance among the ES participants in a long-term protocol.

APPENDIX

Driving Task Post Survey

The intelligent system will change the driving conditions at various times during the assignments. As best you can, answer the following questions about your experience playing the games.

1. Overall, how would you rate the difficulty of the driving game?

Extremely Difficult	Somewhat Difficult	Neutral	Easy	Extremely Easy
1	2	3	4	5

2. Did you notice a change or adjustment in difficulty as you played the game?

No	Yes

3. If you noticed the difficulty change, then did you like how the system changed or adjusted the difficulty for you?

Disliked it a great deal	Did not like it	Neutral	Liked it	Liked it a great deal
1	2	3	4	5

4. Overall, when the system changed or adjusted the difficulty, did you find the change to be helpful to playing the game?

No, extremely unhelpful	No, not helpful	No difference/ Stayed the same	Yes, somewhat helpful	Yes, extremely helpful
1	2	3	4	5

As best you can, rank your level of <<_____>> when completing the task.

5. Enjoyment

Did not enjoy game at all	Did not enjoy game	Neutral	Enjoyed game	Enjoyed game a great deal
1	2	3	4	5

6. Frustration

Never Frustrated 1	Frustrated Once or Twice 2	Neutral 3	Frustrated Some of the Time 4	Frustrated Most of the Time 5

7. Boredom

Never Bored 1	Bored Once or Twice 2	Neutral 3	Bored Some of the Time 4	Bored Most of the Time 5

8. If you could change the system in any way you wanted, then what would you change?

REFERENCES

E. Adams. 2008. The designer's notebook: Difficulty modes and dynamic difficulty adjustment. *Gamasutra*. Retrieved from http://www.gamasutra.com/.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Pub.

J. Baio, L. Wiggins, D. L. Christensen, et al. 2018. Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR. Surveill. Sum.* 67 (2018), 1–23. DOI: 10.15585/mmwr.ss6706a1

B. B. Bederson. 2004. Interfaces for staying in the flow. *Ubiquity* (2004), 1.

M. Bellani, L. Fornasari, L. Chittaro, and P. Brambilla. 2011. Virtual reality in autism: State of the art. *Epidemiol. Psych. Sci.* 20, 3 (2011), 235–238.

E. Bekele, et al. 2016. Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with Autism spectrum disorders (ASD). In *IEEE Virtual Reality (VR), Greenville, SC.* 121–130. DOI: 10.1109/VR.2016.7504695

D. Bian, J. Wade, A. Swanson, Z. Warren, and N. Sarkar. 2015. Physiology-based affect recognition during driving in virtual environment for autism intervention. In *Proceedings of the 2nd International Conference on Physiological Computing Systems*.

D. Bian, J. Wade, Z. Warren, and N. Sarkar. 2016. Online engagement detection and task adaptation in a virtual reality based driving simulator for autism intervention. *Proceedings of the International Conference on Universal Access in Human-Computer Interaction*. Springer.

D. Bian, J. W. Wade, L. Zhang, E. Bekele, A. Swanson, J. A. Crittendon, M. Sarkar, Z. Warren, and N. Sarkar. 2013. A novel virtual reality driving environment for autism intervention. In *Proceedings of the Universal Access in Human-Computer Interaction. User and Context Diversity*. Springer, 474–483.

A. V. Buescher, Z. Cidav, M. Knapp, and D. S. Mandell. 2014. Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr.* 168, 8 (2014), 721–728.

Centers for Disease Control and Prevention. 2012. Injury Prevention and Control: Motor Vehicle Safety—Distracted Driving. https://www.cdc.gov/motorvehiclesafety/distracted_driving/index.html.

J. Baio, L. Wiggins, D. L. Christensen, et al. Prevalence of autism spectrum disorder among children aged 8 years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. 2018. *MMWR Surveill Summ.* 67, 6 (2018), 1–23. DOI: 10.15585/mmwr.ss6706a1

G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun. 2011. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Trans. Syst. Man Cybernet. Part A: Syst. Hum.* 41, 6 (2011), 1052–1063.

G. S. Chasson, G. E. Harris, and W. J. Neely. 2007. Cost comparison of early intensive behavioral intervention and special education for children with autism. *J. Child Fam. Studies* 16, 3 (2007), 401–413.

S. Classen, M. Monahan, and Y. Wang. 2013. Driving characteristics of teens with attention deficit hyperactivity and autism spectrum disorder. *Amer. J. Occupat. Ther.* 67, 6 (2013), 664–673.

J. N. Constantino and C. P. Gruber. 2007. *Social Responsiveness Scale (SRS)*. Western Psychological Services, Los Angeles, CA.

N. Cox, R. Reeve, S. Cox, and D. Cox. 2012. Brief report: Driving and young adults with ASD: Parents' experiences. *J. Autism Dev. Disorders* 42, 10 (2012), 2257–2262.

S. M. Cox, D. J. Cox, M. J. Kofler, M. A. Moncrief, R. J. Johnson, A. E. Lambert, S. A. Cain, and R. E. Reeve. 2016. Driving simulator performance in novice drivers with autism spectrum disorder: The role of executive functions and basic motor skills. *J. Autism Dev. Disorders* 46, 4 (2016), 1379–1391.

B. P. Daly, E. G. Nicholls, K. E. Patrick, D. D. Brinckman, and M. T. Schultheis. 2014. Driving behaviors in adults with autism spectrum disorders. *J. Autism Dev. Disorders* 44, 12 (2014), 3119–3128.

S. H. Fairclough, K. Gilleade, K. C. Ewing, and J. Roberts. 2013. Capturing user engagement via psychophysiology: Measures and mechanisms for biocybernetic adaptation. *Int. J. Auton. Adapt. Commun. Syst.* 6, 1 (2013), 63–79.

M. L. Ganz. 2006. *The Costs of Autism*. CRC Press, New York, 475–502.

D. Giakoumis, A. Vogiannou, I. Kosunen, K. Moustakas, D. Tzovaras, and G. Hassapis. 2010. *Identifying Psychophysiological Correlates of Boredom and Negative Mood Induced during HCI*. B-Interface.

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. 2009. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newslett.* 11, 1 (2009), 10–18.

E. L. Hill. 2004. Evaluating the theory of executive dysfunction in autism. *Develop. Rev.* 24, 2 (2004), 189–233.

L. M. Howden and J. A. Meyer. 2010. Age and sex composition: 2010. *2010 Census Briefs*, U.S. Department of Commerce, Economics and Statistics Administration. U.S. Census Bureau.

P. Huang, T. Kao, A. E. Curry, and D. R. Durbin. 2012. Factors associated with driving in teens with autism spectrum disorders. *J. Dev. Behav. Pediatr.* 33, 1 (2012), 70–74. 10.1097/DBP.1090b1013e31823a31843b31827.

R. Hunicke. 2005. The case for dynamic difficulty adjustment in games. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*. ACM.

K. Jegers. 2007. Pervasive game flow: Understanding player enjoyment in pervasive gaming. *Comput. Entertain.* 5, 1 (2007), 9.

S. Jerritta, M. Murugappan, R. Nagarajan, and K. Wan. 2011. Physiological signals-based human emotion recognition: A review. In *Proceedings of the IEEE 7th International Colloquium on Signal Processing and its Applications (CSPA'11)*. IEEE.

J. Kim and E. André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 12 (2008), 2067–2083.

U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar. 2013. Design of a virtual reality-based adaptive response technology for children with autism. *IEEE Trans. Neural Syst. Rehab. Eng.* 21, 1 (2013), 55–64.

C. Liu, K. Conn, N. Sarkar, and W. Stone. 2008. Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Trans. Robot.* 24, 4 (2008), 883–896.

C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter. 2000. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disorders* 30, 3 (2000), 205–223.

L. Nacke and C. A. Lindley. 2008. Flow and immersion in first-person shooters: Measuring the player's gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*. ACM.

National Highway Traffic Safety Administration. 2016. 2015 motor vehicle crashes: Overview (DOT HS 812 456). *Traffic Safety Facts Research Note* (2016), 1–9. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812456.

National Highway Traffic Safety Administration. 2018. 2016 motor vehicle crashes: Overview (DOT HS 812 498). *Traffic Safety Facts Research Note* (2018). https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812498.

K. A. Orvis, D. B. Horn, and J. Belanich. 2008. The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Comput. Hum. Behav.* 24, 5 (2008), 2415–2433.

A. Pecchinenda. 1996. The affective significance of skin conductance activity during a difficult problem-solving task. *Cogn. Emot.* 10, 5 (1996), 481–504.

R. Pekrun, T. Goetz, L. M. Daniels, R. H. Stupnisky, and R. P. Perry. 2010. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *J. Edu. Psychol.* 102, 3 (2010), 531.

J. A. Healey. 2000. Wearable and automotive systems for affect recognition from physiology. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/9067.

P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio. 2006. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *Int. J. Psychophysiol.* 61, 1 (2006), 5–18.

P. Rani, N. Sarkar, and C. Liu. 2005. Maintaining optimal challenge in computer games through real-time physiological feedback. In *Proceedings of the 11th International Conference on Human Computer Interaction*.

B. Reimer, R. Fried, B. Mehler, G. Joshi, A. Bolfek, K. Godfrey, N. Zhao, R. Goldin, and J. Biederman. 2013. Brief report: Examining driving behavior in young adults with high functioning autism spectrum disorders: A pilot study using a driving simulation paradigm. *J. Autism Dev. Disorders.* 1–7.

M. Rutter, A. Bailey, and C. Lord. 2003. *The Social Communication Questionnaire: Manual.* Western Psychological Services.

N. Sarkar. 2002. Psychophysiological control architecture for human-robot coordination-concepts and initial experiments. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02).* IEEE.

O. Shechtman, S. Classen, K. Awadzi, and W. Mann. 2009. Comparison of driving errors between on-the-road and simulated driving assessment: A validation study. *Traffic Injury Prevent.* 10, 4 (2009), 379–385.

E. Sheppard, E. van Loon, G. Underwood, and D. Ropar. 2016. Difficulties predicting time-to-arrival in individuals with autism spectrum disorders. *Res. Autism Spectrum Disorders* 28 (2016), 17–23.

G. Stemmler, M. Heldmann, C. A. Pauls, and T. Scherer. 2001. Constraints for emotion specificity in fear and anger: The context counts. *Psychophysiology* 38, 2 (2001), 275–291.

D. Strickland. 1997. Virtual reality for the treatment of autism. *Studies Health Technol. Info.* 44 (1997), 81–86. DOI: 10.3233/978-1-60750-888-5-81

R. Stripling, J. T. Coyne, A. Cole, D. Afergan, R. L. Barnes, K. A. Rossi, L. M. Reeves, and D. D. Schmorrow. 2007. Automated SAF adaptation tool (ASAT). In *Proceedings of the International Conference on Foundations of Augmented Cognition.* Springer.

J. Wade, D. Bian, J. Fan, L. Zhang, A. Swanson, M. Sarkar, A. Weitlauf, Z. Warren, and N. Sarkar. 2015. A virtual reality driving environment for training safe gaze patterns: Application in individuals with ASD. In *Proceedings of the Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being.* Springer, 689–697.

J. Wade, D. Bian, L. Zhang, A. Swanson, M. Sarkar, Z. Warren, and N. Sarkar. 2014. Design of a virtual reality driving environment to assess performance of teenagers with ASD. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction.* Springer International Publishing.

S. W. White, K. Keonig, and L. Scahill. 2007. Social skills development in children with autism spectrum disorders: A review of the intervention research. *J. Autism Dev. Disorders* 37, 10 (2007), 1858–1868.

Z. E. Warren, Z. Zheng, A. R. Swanson, E. Bekele, L. Zhang, J. A. Crittendon, A. F. Weitlauf, and N. Sarkar. 2015. Can robotic interaction improve joint attention skills? *Journal of Autism and Developmental Disorders* 45, 11 (2015), 3726–3734.

L. Zhang, Z. Warren, A. Swanson, et al. 2018. *J Autism Dev. Disord.* 48 (2018), 2779. DOI: https://doi.org/10.1007/s10803-018-3544-7

Z. Zheng, et al. 2015. Design of a computer-assisted system for teaching attentional skills to toddlers with ASD. In *Universal Access in Human-Computer Interaction. Access to Learning, Health and Well-Being (UAHCI'15)*, M. Antona and C. Stephanidis (Eds.). Lecture Notes in Computer Science, Vol. 9177. Springer, Cham.

J. Zhou, J. Sun, F. Chen, Y. Wang, R. Taib, A. Khawaji, and Z. Li. 2015. Measurable decision making with GSR and pupillary analysis for intelligent user interface. *ACM Trans. Comput.-Hum. Interact.* 21, 6 (2015), 1–23.

Received December 2017; revised September 2018; accepted November 2018