Protocol for Eliciting Driver Frustration in an In-vehicle Environment

Ajjen Joshi
Affectiva
Boston, MA, USA
ajjen.joshi@affectiva.com

Youssef Attia
Affectiva
Cairo, Egypt
youssef.attia@affectiva.com

Taniya Mishra
Affectiva
Boston, MA, USA
taniya.mishra@affectiva.com

***Abstract*—A state of frustration can impair a driver's ability to make decisions that optimize the safety of the driver as well as that of those around him or her. Equipping a car with the capability of detecting signs of driver frustration and responding with appropriate interventions can be an effective method to improve driver safety. In this paper, we first describe the design and implementation of a novel protocol used to elicit true frustration of varying intensities in participants interacting with an in-car Human Machine Interface (HMI) while driving in a simulator. We detail the instrumentation that was used to capture the participants' interactions with the HMI. We provide a computational analysis of signs of frustration displayed by participants in the facial and vocal modalities and discuss trends that were observed. Finally, we present baseline machine learning methods trained on features computed from facial and vocal modalities to predict the difficulty of the completed task and whether the participant is multitasking, validating the assumptions that informed our protocol design.**

***Index Terms*—frustration, in-car sensing, emotion elicitation protocol, multimodal**

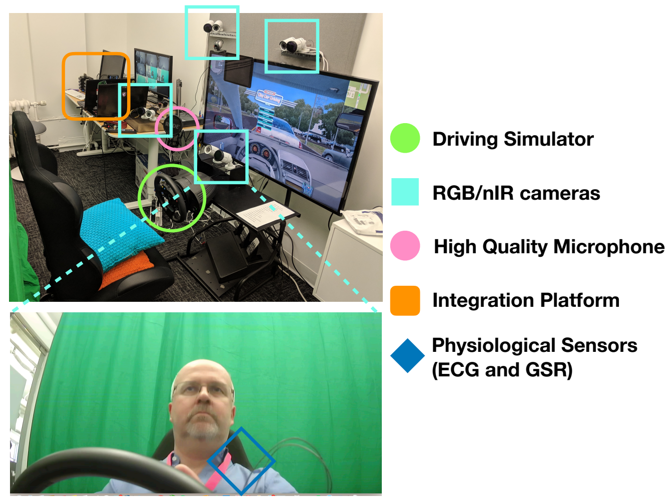

Fig. 1. The data collection lab, illustrating the setup of various sensors and instruments, used for running protocol to elicit frustration in participants driving a simulator.

I. Introduction

Interfaces that enable interaction with computers have become more diverse and commonplace in recent years. People interact with their computing devices and services using increasingly natural modes of communication, such as text, voice and gestures. As these interfaces become more ubiquitous, the reliance on these technologies to seamlessly perform tasks intended for them by the user will grow, along with the expectations regarding the robustness and efficacy of these systems. Despite the impressive autonomous abilities of current AI enabled machines, human machine interactions continue to be frustrating. A successful Human Machine Interface (HMI) of the future may therefore need to understand and cater to the unstated emotional state of users, who are intolerant of the HMI's failure to respond to their requests or demands.

Frustration induced by users' interactions with HMIs is especially a problem in the automotive use-case. Recent statistics show that every single day, 9 people die and a thousand are injured due to distracted drivers on the road [1]. People in the United States already spend an enormous amount of time in their cars, commuting further and further to get to and from their jobs. It is estimated that people spend an average of 17600 minutes per year in a car [2], equivalent to more than 12 24-hour days. This means that invariably some of that time, people will be interacting with their in-car HMI to perform tasks such as getting directions to a destination, sending messages to friends and family or changing radio stations. A failure of the in-car HMI to understand the driver or perform the task intended for it may cause frustration.

Studies have shown that a frustrated driver can be a dangerous driver [3]. Equipping vehicles with in-cabin sensors that can measure signals in relevant modalities and models that can process the signals accurately to detect if the driver is frustrated can enable the deployment of effective real-time intervention systems to de-escalate the driver from a state of frustration. For example, Hernandez et al. proposed measures such as an empathetic GPS voice, calming temperature and corrective headlights to counter stress and frustration [4].

Over the past decade, supervised deep machine learning models have enabled significant progress in a variety of tasks spanning computer vision, computational speech and natural language processing. These methods have also engendered significant progress in affective computing but most of the research has been focused in modeling basic emotions [5] and expressions [6]. More nuanced emotional states such as frustration have not had the same attention.

While there exist many datasets of basic emotions, datasets

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

of frustration are comparatively fewer and mostly limited to the education domain [7]. One of the challenges of developing frustration datasets is that the manifestations of frustration are subtler than basic emotions like anger or rage, which makes it difficult to simulate or act out, while also being time-consuming to collect naturally because it is highly variant across individuals and thus needs to be collected from a wider population. Notwithstanding these challenges, 'true' or natural frustration datasets need to be collected in order to build machine learning models that can detect frustration in humans.

In this paper, we focus primarily on the description of a data collection protocol that elicits true frustration of different intensities in participants, thereby ensuring a relatively balanced distribution of frustration intensities in the dataset. In order to study signs of frustration displayed by drivers while interacting with an in-car HMI, we designed and implemented a novel protocol involving participants interacting with the HMI in a Wizard-of-Oz setting to perform various tasks while driving in a simulator. The protocol involved capturing the participant's interactions with the HMI through an array of sensors.

The contributions of this paper are threefold. First, we propose a novel frustration data collection protocol, which we implemented to elicit varying intensities of frustration in 105 participants. Second, we provide an analysis of face and speech data that were captured during the protocol, illustrating observed trends in facial and vocal displays of frustration. Third, we present baseline machine learning methods to recognize the intended difficulty of the task and whether the user is multitasking from the displayed facial and vocal signals, validating the assumptions that informed our protocol design.

II. Related Work

Lawson defined frustration as "the occurrence of an obstacle that prevented the satisfaction of a need" [8]. Frustration can cause an "increase in arousal" [9], which can cause aggressive behavior according to the frustration-aggression hypothesis [10]. In the context of driving, frustration has been shown to occur when the goal of achieving mobility is impeded, for example, by red-light signals, slow moving vehicles, or blocked path by other vehicles or pedestrians [11]. Frustration can hinder cognitive processes essential for driving, such as judgment, decision making and attention [12]. Shinar contended that a driver's frustration is enabled by the driver's environment and is a cause of aggressive driving behavior [3] such as tailgating, headlight flashing, obscene gestures, obstructing other vehicles and verbal abuse [13].

In order to study frustration and identify strategies that can be used to measure it, researchers have experimented with various methods to elicit frustration. For example, Riseberg et al. designed software interventions to simulate the mouse failing or sticking, which hampered the scores achieved by participants while playing a mouse-based game [9]. Taib et al. induced frustration by adding constraints such as noise, speed bumps, narrow road u-turns, tight parking spots while completing driving tasks [14]. They trained a binary classifier to determine high frustration in a simulated driving task based on features extracted from pressure sensors attached to the driving seat but theirs is a preliminary study based on only 10 participants. Abdic et al. asked participants to complete several tasks, such as entering addresses into the navigation system, while driving around a preset route [15]. They trained SVMs based on data from 20 participants to distinguish "frustrated" sequences from "satisfied" sequences based on facial and audio features computed on a dataset of participants interacting with a voice interface while driving [15]. However, the frustration induced in participants in their study was dependent on existing voice agents inside the car and constrained to a single task: navigation. In contrast, we used a Wizard-of-Oz setting to emulate a more complex and expressive in-car HMI, which not only enabled us to design a number of controlled scenarios to elicit varying levels of frustration in participants but also elicit frustrations not likely to arise in today's transactional, single-turn HMIs.

A. Manifestations of Frustration

Frustration is manifested in a variety of complex ways. In the context of driving, it has been shown to be accompanied by various behaviors, such as, horn honking, purposeful tailgating, flashing high beams [16], overtaking [17] as well as in the increased intensity in pedal actuation [11]. Frustrated drivers are more likely to yell and swear at other drivers [16]. Gestural indicators of frustration while driving have also been identified: clenched fist, insulting gestures aimed at others, and banging hand against the steering wheel [18].

Facial features are also an important signal in predicting frustration. Graafsgard et al. indicated several facial action units (AU), anatomically described facial expressions developed by Ekman and Friesen as part of the Facial Action Coding System (FACS) [19], to be correlated with frustration during learning [7]. Ihme et al. found that muscles in the mouth region (nose wrinkler, chin raiser, lip pucker, lip pressor) are more frequently activated during frustration [20]. Abdic et al. found that many people often smile when frustrated [15]. However, identifying global indicators of frustration can be challenging because it may not manifest uniformly for all people. Signs of frustration are displayed differently by culture [21], personality types [18], [22], gender [3] and age [23], [24].

B. Annotating Frustration Data

The most common method of obtaining frustration labels for a sequence is through self-reports completed by participants. Several standard questionnaires have been developed for drivers to fill in studies of frustration, anxiety or stress, such as, the Driving Behavior Inventory - General (DBI-Gen) and the State Driving Behavior Checklist [16]. Gunatillake et al. developed a Traffic Frustration Index (TFI) in order to quantify the events which contribute to driver frustration [25]. Abdic et al. asked participants to complete a self-report which included a question measuring frustration in a scale of 1 to 10 [15]. Grafsgaard et al. had participants self-report pre-test and post-test measures of frustration while interacting with a computer-mediated human tutor [7].

Another method used for annotating frustration is to pre-define "frustrating" and "non-frustrating" tasks. For example, Ihme et al. designed driving tasks in a simulator where the frustrated condition contained time constraints and unfavorable traffic conditions that hindered task completion [20].

C. Modeling Frustration

Researchers have attempted to model, measure and predict frustration using signals from various modalities. Bosch et al. trained simple models to classify learning-centered affect states (frustration, boredom, confusion, delight and engagement) on facial action unit features extracted from a dataset of students interacting with an intelligent tutor [26]. Ihme et al. built classifiers trained on facial action unit features as well as cortical activations with functional near-infrared spectroscopy (fNIRS) to classify the frustration levels of participants driving in a simulator [20]. McCuaig et al. trained a neural network to detect frustration based on features collected from an eye-tracker [27]. Ang et al. investigated the use of prosody-based and semantics-based features in building models to detect frustration and annoyance [28]. Fernandez and Picard showed that electrodermal response (GSR) was indicative of human frustration [29]. Belle et al. distinguished frustrated students from calm students using ECG data [30].

Kapoor et al. detected students frustration while interacting with intelligent tutoring systems, training models on features extracted from the participants face, posture sensors in the chair, skin conductance as well as pressure sensors attached to the mouse [31]. Malta et al. constructed a Bayesian network that modeled frustration using traffic variables (e.g. traffic density, red lights), vehicle measurements (pedal actuation), behavioral measures (facial features, speech recognition errors) and physiological measures (electrodermal activity) [11]. Grafsgaard et al. trained models to detect learning-centered affective states of students interacting with an intelligent tutor on features extracted from textual dialogue with the tutor, nonverbal behavior and task action input streams [32].

III. Data Collection Protocol

The goal of the proposed data collection protocol was to elicit in a random group of human participants a frustrated emotional state to the point where they displayed perceptible facial, vocal and physiological markers of frustration. Specifically, we wanted to collect audio, video and physiological recordings corresponding to the participants frustration while interacting with an in-car HMI under two conditions:

- Performing a driving task *and* interacting with the HMI
- Performing no other task while interacting with the HMI

The driving task was performed in an in-lab simulator.

A. Participants

105 participants (55 female, 47 male, and 3 who did not specify gender) participated in this collection. The participants were health screened for use of any Beta-Blockers, flu or cold-like symptoms or recent life events. Participants were recruited to participate in a "conversational agent interaction study" with no reference to frustration, so as to prevent any preconditioning or expectation that the experience may be frustrating. Participants were simply told the purpose of the study was to understand how humans interact with conversational agents in order to accomplish different tasks. Before beginning the study, participants signed consent forms, acknowledging that their participation in the study was voluntary. They were then given 5 minutes to acclimatize to the driving simulator.

B. Methodology

We aimed to elicit true frustration at varying intensities in participants by having them interact with the Amazon Alexa voice agent [33] to accomplish tasks of varying difficulty. Participants were asked to perform a variety of interaction tasks in a timed scenario, which were designed to mimic real interactive conversations that people might have with an in-car HMI. However, participants were unaware of the fact that their interactions were performed in a Wizard-of-Oz setting, where the questions and responses from Alexa were either pre-recorded and played by a researcher running the study, or produced via imitation using Alexa's "Simon says" feature. The responses (pre-recorded or imitated) were created so as to deliberately impede the completion of the instructed task.

Each participant performed two sessions, each consisting of the same 6 voice interaction tasks: A "multitasking" session, where the participant performed the 6 voice interaction tasks while free driving on an urban route using a driving simulator (Thrustmaster TMX PRO Racing Wheel running 'City Car Driving' software [34] running on an MSI Windows gaming laptop; and a "unitasking"session, where the participant did not need to drive while performing the same voice interaction tasks. After each session, the participants reported their frustration for each task on a scale of 1 (not frustrated) to 4 (highly frustrated). The self-reports were administered using a custom questionnaire containing a subset of questions on the State-Trait Anger Expression Inventory [35]. Participants could also provide subjective feedback after the final session.

Participants were allowed between 30 and 150 seconds to perform each task, depending on task difficulty. To control for any learning effects, we randomized the order of the sessions, so that half of the population performed the tasks first without driving, and then in conjunction with driving, and the remaining half performed the tasks in reverse order. Each participant was paid \$120 for participating in the study. However, to motivate the participants to complete each task, and therefore increase the likelihood of frustrating them when they were unable to complete the timed tasks, the participants were told that they would get paid a minimum of \$80 and up to \$40 more depending on how many tasks they completed successfully. Each participant took approximately one hour to complete both sessions.

C. Tasks

The tasks that participants were asked to perform in each of their sessions are as follows:

- Add and remove items off of a shopping list

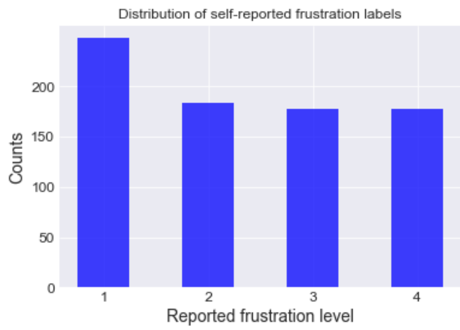

Fig. 2. Histogram of self-reported frustration intensities.

- Ask the HMI to tell a joke
- Ask the HMI to set timers
- Ask the HMI to play various radio stations
- Ask the HMI to play various media (e.g. songs, e-books, news etc.)
- Compose and send text messages to another person

These tasks were chosen as examples of interactions that people may have with a real in-car HMI agent. In order to impede the completion of the tasks and thus induce frustration, the responses were either pre-recorded or manipulated using Alexas "Simon Says" feature. Strategies for impeding task completion included deliberately misunderstanding users commands, ignoring or denying users' commands, interrupting user in the middle of making a command as well as setting off annoying and loud countdown clocks and buzzers.

D. Instrumentation and Sensors

The driving simulator consisted of a bench with a driver seat, steering wheel, gas and brake pedals, accompanied by driving simulator software (Figure 1). The lab was equipped with the following suite of sensors:

- Multi-camera setup: 4 pairs of NIR and RGB cameras (PRO-T858 3 Megapixel HD Bullet Camera for Swann Super HD 4750 Series DVRs) were used to capture multiple views of the participant.
- Audio: Along with the audio stream captured by one of the cameras, a RODE NT1-A 1" cardioid condenser microphone was placed to capture high-quality audio of the participant interaction.
- Baby monitor: A baby monitor was used by the researcher conducting the study to observe the participant so that the pre-recorded Alexa responses could be played at appropriate times.
- Electrocardiogram (ECG): Participants were asked to place 4 ECG Shimmer [36] sensors on their body that enabled measurement of heart rate.
- Galvanic Skin Response (GSR): Participants were asked to wear a Shimmer [36] skin conductance sensor on their index finger. Both the ECG and GSR sensors were synced to the iMotions software suite [37].

All signals were synchronized with the help of audio cues.

IV. Exploratory Data Analysis

For preliminary data analysis, we selected a subset of 74 participants, each of whom completed 12 tasks. The remaining participants were ignored in this analysis as the high quality audio for these participants were corrupted in some segments. For these participants, we analyzed frustration intensities as reported by participants themselves for each task, and the facial expressions and vocal characteristics they displayed when engaged in performing the tasks in the collection protocol.

A. Distribution of Frustration Intensities

One reason for designing a frustration elicitation protocol was to engineer interactions that could induce varying states of frustration in participants. In Figure 2, we plot the counts of the various intensities of frustration as reported by the participants for every task. As can be observed, the distribution of self-reported intensities is fairly uniform, with each intensity of frustration being reported more than 150 times.

B. Multitasking Induces Observable Signs of Frustration

We analyzed the average activations of facial action units and emotions for all tasks using Affectiva's SDK [38]. Participants displayed more brow furrows, chin raises, inner brow raises, lid tightens and lip stretches for tasks completed while driving compared to tasks where they focused solely on interacting with the HMI. Interestingly, participants displayed the reverse trend for brow raise (Figure 3 a). One hypothesis explaining this observation is that brow raises are often inhibited by brow furrows, which occur when a person is frowning.

We also computed acoustic features for segments of each task where the participant was speaking. Prior to analysis, the audio was denoised using logmmse [39]. Spoken segments were identified manually using human annotators. For each spoken segment, PyAudioAnalysis [40] was used to extract the acoustic features, including energy features, spectral features, mel-frequency cepstral features and chroma features; chroma represents the distribution of an audio signals energy across a predefined set of pitch classes. We also computed the total amount of speech produced compared to the overall length of the task. A subset of acoustic features that showed high variance across the tasks is shown in Figure 3 c. From this, we observe that for matched tasks participants spoke less when they were driving. The other features had an approximately monotonic trend (increasing or decreasing) in relation to task difficulty, and an observable difference when comparing speech produced for tasks completed while driving to speech produced for the same tasks where the participants were not.

We also computed the average classifier activations for pre-trained vision and speech anger models in the Affectiva SDK across all tasks. This also illustrates a similar trend: averaged across all frames, across all tasks for both visual and audio signals, pre-trained anger models are activated more for tasks completed in conjunction with driving compared to tasks where the participants focus solely on interacting with the in-car HMI (Figure 3 b and d).

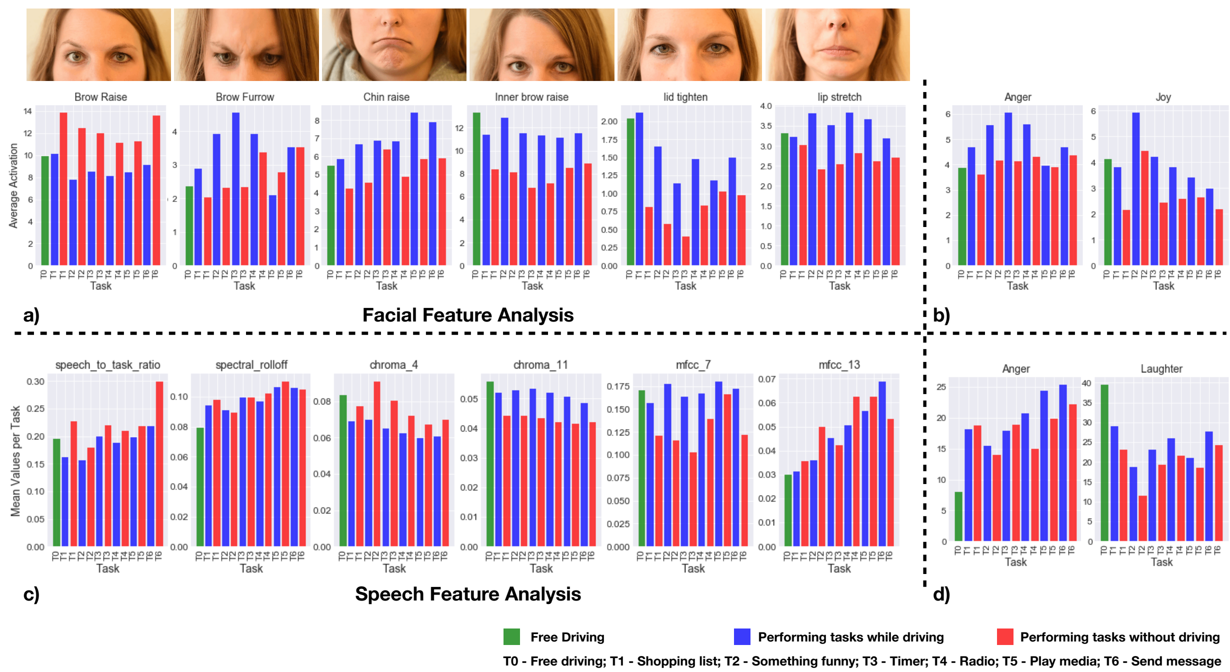

Fig. 3. a) For each task, we plotted the the average activations of facial action unit classification models. b) For each task, we plotted the average activations of canonical anger and joy classification models trained on facial displays of anger and joy respectively. c) For each task, we plotted the mean values of various speech features. d) For each task, we plotted the average activations of anger and laughter classification models trained on vocal displays of anger and laughter respectively. *Images displaying Facial Action Units were obtained from https://imotions.com/blog/facial-action-coding-system/*

Another observation from the anger plots are the relative degree of estimated anger relative to the task of free driving (where participants are driving without having to perform any task). For the task designed and self-reported to be the most frustrating (Send message) while driving, the vision and speech anger models were activated, on average, 1.2 and 3.2 times more respectively compared to the task of free driving. These observations quantify the assumption that multi-tasking while driving can be frustrating and hence distracting.

C. Co-occurrence of Joy and Laughter

We also observed that participants have interesting, non-intuitive manifestations when responding to a state of frustration. For example, participants smiled, laughed or showed facial expressions of joy when frustrated (Figs. 3 b and d). This can be seen in the classifier activations of a face-based "Joy" classifier as well as those of a voice-based "Laughter" classifier. On average, activations for these positive valence classifiers scored higher for all tasks when participants were multitasking compared to when they were not. This is consistent with prior art ([15], [41]) where researchers found that people may display what are otherwise considered signs of positive affect despite being in a state of frustration.

D. Subjective Comments and Self-reports

Participants also wrote their opinions in a final survey after completing both sessions. Participant 39 stated, "I found this confusing at times and not having driven a car in years stressful especially when attempting to multitask." Participant 24 stated, "Me and Alexa are in a fight. I think she knows I prefer Google assistant." Participant 14 stated "Alexa doesnt acknowledge that she hears me. Its polite to respond when someone says your name." These opinions are indicative of the frustration induced when the driver had to perform challenging voice interactions while driving.

The average reported frustration, difficulty and stress intensities (on a scale of 1-4) for each task is shown in Figure 4 a. We observed that for all tasks, the mean of the self-reported frustration, difficulty and stress intensities for completing that task while also driving is higher, than for completing the corresponding task while not driving. This suggests that difficult and unexpected interactions with an in-car HMI induces more frustration in people while they are in the act of driving, compared to when they focus their attention solely on communicating with the agent without having to attend to the task of safely driving.

V. Predicting Task Difficulty and Multitasking

The assumptions behind our protocol design were: increasing task difficulty can cause perceptible increase in displays of frustration and the higher cognitive load placed on the participant while multitasking can trigger a heightened sense of frustration. To evaluate the assumptions explicitly, we trained preliminary speech and vision models to predict: a) the task difficulty level, and b) the multitasking label, which represents whether the participant was driving while performing the interaction or was focused solely on the interaction. We

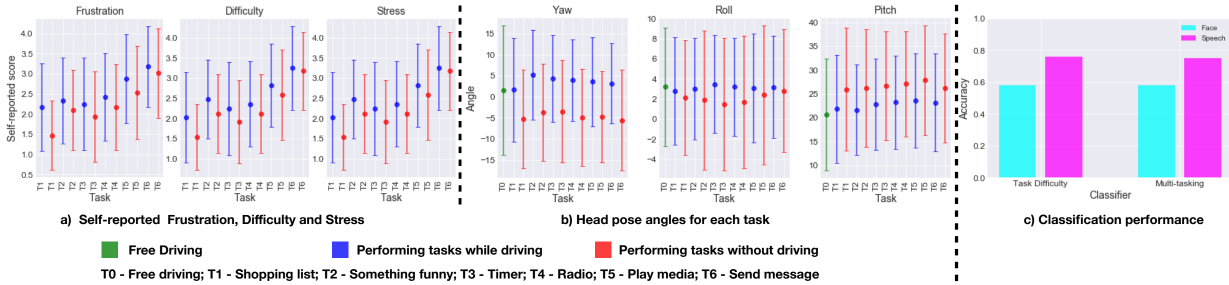

Fig. 4. a) For each task, we plot the average self-reported frustration, difficulty and stress level (on a scale 1-4). b) For each task, we plot the average head pose angle. c) We plot the classification performance of models trained to predict task difficulty and multitasking.

binarized all labels: the tasks “Play radio”, “Play media” and “Send messages” were assigned the “difficult” label as they were designed to be more difficult than the rest. Multitasking binary labels were assigned based on whether the participant was driving while interacting with the HMI.

In order to train our baseline vision models, we used Affectiva’s SDK to first extract classifier activations for a set of facial action units (Brow Furrow, Brow Raise, Eye Closure, Mouth Open, Nose Wrinkle, Upper Lip Raise, Yawn, Cheek Raise, Chin Raise, Inner Brow Raise, Jaw Drop, Lid Tighten, Lip Corner Depressor, Lip Press, Lip Pucker, Lip Stretch, Smile), and facial emotions (Anger, Joy, Surprise). We did not use head pose (Pitch, Roll, Yawn) to train our models because we observed that people mostly looked ahead in sessions where they are driving but exhibited higher variance in head pose in sessions where they are not driving, as illustrated in Figure 4 b. Each task was represented by an input vector of 20 facial features over a number of frames. As an aggregate feature representation, we computed the mean, standard deviation, min and max for each aforementioned feature, resulting in a 80-dimension feature vector.

To train our baseline speech models, we extracted acoustic features using the PyAudioAnalysis library, including energy features, spectral and ceptstral features and chroma features from audio samples corresponding to segments where the participant was speaking while performing the task. For each task, we computed the same set of statistics of 35 different features in addition to the speech-to-task ratio resulting in a 141-dimensional feature vector.

We trained and tested our models using 4-fold subject-independent splits of the data. We trained a random forest model with a 100 trees, each to a depth of 10. The results shown in Figure 4 c indicate that speech features were more predictive of both difficulty level (Mean accuracy of 0.76 for speech, 0.59 for vision) and multitasking (Mean accuracy of 0.78 for speech, 0.58 for vision). The ability to distinguish task difficulty and multitasking based on simple classifiers trained on facial and vocal features validates the protocol design, which aimed to elicit frustration display perceptible in face and voice. This difference in performance between the speech and vision models points towards the hypothesis that frustration may not be manifested equally in all modalities.

VI. Conclusions and Future Work

Here, we have proposed a novel frustration elicitation protocol that we designed and implemented to elicit varying intensities of true frustration in a realistic driving scenario. In this protocol, participants interacted with an HMI to perform tasks of varying difficulty, while driving in a simulator. The HMI interactions were manipulated by a human in the background to deliberately impede task completion, and thus induce subject frustration. During administration of this protocol, all participant interactions with the HMI were captured through an array of audio, video and physiological sensors. This paradigm allowed us more control in the elicitation of subject frustration, and can be expanded to better study interaction design patterns with smarter conversational agents of the future.

We provided a detailed exploratory analysis of the audio and video data captured from 74 study participants, illustrating several observable signs of frustration in facial and vocal display. Finally, we presented baseline machine learning models to recognize task difficulty labels as well as the presence of multitasking from the underlying audio and video signals, to validate assumptions that informed our protocol design.

There are several avenues for future work. First, we plan to use human annotators to find correlations between participant self-reports of frustration and human observations of frustration in face and voice. We will also obtain dense temporal frustration labels, which will enable us to train frustration detection models, not at the task-level but at the much-denser elicitation level. More labeled data will allow us to experiment with more sophisticated but data hungry deep architectures. Furthermore, we will leverage the multimodal nature of frustration manifestation by using signals extracted from the physiological sensors (which we haven’t yet analyzed), in addition to video and audio data to build richer models of representation and classification. A further improvement would be to personalize models to individual nuances and idiosyncrasies regarding how signs of frustration are displayed.

Acknowledgments

The authors are grateful to the reviewers for their constructive feedback. The authors would also like to thank Mohammad Mavadati, Greg Pettigrew, Jessica Wilson and Roxanna Salim for their input and feedback.

REFERENCES

[1] N. H. T. S. Administration, "Distracted driving 2016," 2018.

[2] A. A. Association, "Americans spend an average of 17,600 minutes driving each year," 2016.

[3] D. Shinar, "Aggressive driving: the contribution of the drivers and the situation," *Transportation Research Part F: traffic psychology and behaviour*, vol. 1, no. 2, pp. 137–160, 1998.

[4] J. Hernandez, D. McDuff, X. Benavides, J. Amores, P. Maes, and R. Picard, "Autoemotive: bringing empathy to the driving experience to manage stress," in *Proceedings of the 2014 companion publication on Designing interactive systems*, pp. 53–56, ACM, 2014.

[5] G. Levi and T. Hassner, "Emotion recognition in the wild via convolutional neural networks and mapped binary patterns," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pp. 503–510, ACM, 2015.

[6] A. Gudi, H. E. Tasli, T. M. Den Uyl, and A. Maroulis, "Deep learning based facs action unit occurrence and intensity estimation," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 6, pp. 1–5, IEEE, 2015.

[7] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Automatically recognizing facial indicators of frustration: a learning-centric analysis," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 159–165, IEEE, 2013.

[8] R. Lawson, *Frustration; the development of a scientific concept*. Macmillan, 1965.

[9] J. Riseberg, J. Klein, R. Fernandez, R. W. Picard, *et al.*, "Frustrating the user on purpose: Using biosignals in a pilot study to detect the user's emotional state," in *Conference on Human Factors in Computing Systems: CHI 98 conference summary on Human factors in computing systems*, vol. 18, pp. 227–228, 1998.

[10] J. Dollard, N. E. Miller, L. W. Doob, O. H. Mowrer, and R. R. Sears, "Frustration and aggression.," 1939.

[11] L. Malta, C. Miyajima, N. Kitaoka, and K. Takeda, "Analysis of real-world driver's frustration," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 109–118, 2011.

[12] M. Jeon, "Towards affect-integrated driving behaviour research," *Theoretical Issues in Ergonomics Science*, vol. 16, no. 6, pp. 553–585, 2015.

[13] M. Joint, "Road rage," 1995.

[14] R. Taib, J. Tederry, and B. Itzstein, "Quantifying driver frustration to improve road safety," in *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems*, pp. 1777–1782, ACM, 2014.

[15] I. Abdic, L. Fridman, D. McDuff, E. Marchi, B. Reimer, and B. Schuller, "Driver frustration detection from audio and video in the wild," in *KI 2016: Advances in Artificial Intelligence: 39th Annual German Conference on AI, Klagenfurt, Austria, September 26-30, 2016, Proceedings*, vol. 9904, p. 237, Springer, 2016.

[16] D. A. Hennessy and D. L. Wiesenthal, "Traffic congestion, driver stress, and driver aggression," *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, vol. 25, no. 6, pp. 409–423, 1999.

[17] N. Kinnear, S. Helman, C. Wallbank, and G. Grayson, "An experimental study of factors associated with driver frustration and overtaking intentions," *Accident Analysis & Prevention*, vol. 79, pp. 221–230, 2015.

[18] D. E. Miles and G. L. Johnson, "Aggressive driving behaviors: are there psychological and attitudinal predictors?," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 6, no. 2, pp. 147–161, 2003.

[19] P. Ekman, "Facial action coding system," 1977.

[20] K. Ihme, A. Unni, M. Zhang, J. W. Rieger, and M. Jipp, "Recognizing frustration of drivers from face video recordings and brain activation measurements with functional near-infrared spectroscopy," *Frontiers in human neuroscience*, vol. 12, 2018.

[21] G. Bateson, "Iv. the frustration-aggression hypothesis and culture.," *Psychological Review*, vol. 48, no. 4, p. 350, 1941.

[22] E. R. Dahlen, B. D. Edwards, T. Tubré, M. J. Zyphur, and C. R. Warren, "Taking a look behind the wheel: An investigation into the personality predictors of aggressive driving," *Accident Analysis & Prevention*, vol. 45, pp. 1–9, 2012.

[23] S. D. Lawson, "Red-light running: accidents and surveillance cameras," 1991.

[24] A. R. Hauber, "The social psychology of driving behaviour and the traffic environment: Research on aggressive behaviour in traffic," tech. rep., 1983.

[25] T. Gunatillake, P. Cairney, and R. Akcelik, "Traffic management performance: Development of a traffic frustration index," 2000.

[26] N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura, L. Wang, and W. Zhao, "Automatic detection of learning-centered affective states in the wild," in *Proceedings of the 20th international conference on intelligent user interfaces*, pp. 379–388, ACM, 2015.

[27] J. McCuaig, M. Pearlstein, and A. Judd, "Detecting learner frustration: towards mainstream use cases," in *International Conference on Intelligent Tutoring Systems*, pp. 21–30, Springer, 2010.

[28] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke, "Prosody-based automatic detection of annoyance and frustration in human-computer dialog," in *Seventh International Conference on Spoken Language Processing*, 2002.

[29] R. Fernandez and R. W. Picard, "Signal processing for recognition of human frustration," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, vol. 6, pp. 3773–3776, IEEE, 1998.

[30] A. Belle, S.-Y. Ji, S. Ansari, R. Hakimzadeh, K. Ward, and K. Najarian, "Frustration detection with electrocardiograph signal using wavelet transform," in *2010 International Conference on Biosciences*, pp. 91–94, IEEE, 2010.

[31] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," *International journal of human-computer studies*, vol. 65, no. 8, pp. 724–736, 2007.

[32] J. F. Grafsgaard, J. B. Wiggins, A. K. Vail, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "The additive value of multimodal features for predicting engagement, frustration, and learning during tutoring," in *Proceedings of the 16th International Conference on Multimodal Interaction*, pp. 42–49, ACM, 2014.

[33] M. Alexa, *Amazon Alexa: 2017 User Guide + 200 Ester Eggs*. Independently published, 2017.

[34] "City car driving - car driving simulator, pc game." https://citycardriving.com/. Accessed: 2019-07-17.

[35] C. Spielberger, "Manual for the state-trait anger expression scale (staxi)," *Odessa, FL: Psychological Assessment Resources*, 1988.

[36] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "Shimmer–a wireless sensor platform for noninvasive biomedical research," *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1527–1534, 2010.

[37] "imotions." https://imotions.com/. Accessed: 2019-07-17.

[38] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*, pp. 3723–3726, ACM, 2016.

[39] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 33, pp. 443 – 445, 05 1985.

[40] T. Giannakopoulos, "pyaudioanalysis: An open-source python library for audio signal analysis," *PloS one*, vol. 10, no. 12, 2015.

[41] M. E. Hoque, D. J. McDuff, and R. W. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 323–334, 2012.