Nass, C., Jonsson, I., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. Extended Abstracts of Human Factors in Computing Systems: CHI 2005, USA, 1973-1976.

Improving Automotive Safety by Pairing Driver Emotion and Car Voice Emotion

Clifford Nass[1] Ing-Marie Jonsson[2] Helen Harris[2] Ben Reaves[2] Jack Endo[2] Scott Brave[1] Leila Takayama[1]

(1) Department of Communication
Stanford University
Stanford, CA 94305-2050
{nass,brave,takayama}@stanford.edu

(2) Toyota Information Technology Center
4009 Miranda Avenue
Palo Alto, CA 94025
{ingmarie, hharris, ben, endojack}@us.toyota-itc.com

ABSTRACT

This study examines whether characteristics of a car voice can affect driver performance and affect. In a 2 (driver emotion: happy or upset) x 2 (car voice emotion: energetic vs. subdued) experimental study, participants (N=40) had emotion induced through watching one of two sets of 5-minute video clips. Participants then spent 20 minutes in a driving simulator where a voice in the car spoke 36 questions (e.g., "How do you think that the car is performing?") and comments ("My favorite part of this drive is the lighthouse.") in either an energetic or subdued voice. Participants were invited to interact with the car voice. When user emotion matched car voice emotion (happy/energetic and upset/subdued), drivers had fewer accidents, attended more to the road (actual and perceived), and spoke more to the car. Implications for car design and voice user interface design are discussed.

Categories and Subject Descriptors: H.5.2 User Interfaces; H.1.2 User/Machine Systems

Keywords: Safety; voice user interfaces; affective computing; automobile interfaces; user behavior.

INTRODUCTION

The traditional approach to classifying users is by *traits,* such as gender and personality, which are stable characteristics of an individual. Even user expertise can be thought of as a trait, unchanging within the context of a session. However, user behaviors, cognitions, and attitudes are also influenced by *states,* the moment-to-moment feelings, knowledge, and physical situation of the person. States can change multiple times during a session, sometimes even in seconds, in response to and influencing reactions to the interface [8].

Predicting a user's behavior at any given moment in time, then, requires attention to their states as well as their traits. For example, while extroverts are generally talkative, they might be as silent as introverts in a library or even quieter when they bump into their secret crush.

The most important user states are *emotions*. Although emotion was one of the primary foci of the early field of psychology, the study of emotion has lain dormant for a long time [5]. However, it is now understood that rich emotions are a fundamental component of being human, and emotions powerfully predict how a person will behave [1]. Affective states—whether short-lived emotions or longer-term moods—color almost everything people do and experience. Emotion is not limited to the occasional outburst of fury when being insulted, or frustration when trapped in a traffic jam. Indeed, many psychologists now argue that it is impossible for a person to have a thought or perform an action without engaging, at least unconsciously, his or her emotional systems.

Just as users can have emotions, interfaces can manifest emotion [9]. Textual interfaces can exhibit emotion through word choice, pictorial interfaces can smile or frown, and voice interfaces can exhibit emotion through tone of voice. Unfortunately, virtually no research has been done on how user emotion and interface emotion interact [8]. This paper addresses the question, "how do user emotion and voice interface emotion interact to influence drivers' performance and attitudes?"

SPEECH AND EMOTIONAL CUES

Speech is a powerful carrier of emotional information. It has been shown that most basic emotions are associated with acoustic properties in a voice, such as loudness, fundamental frequency, frequency range, and speech-rate. For example, sadness is generally conveyed by slow and low-pitched speech, while happiness is associated with fast and louder speech [8].

Emotional cues in speech interfaces have been studied in contexts other than the car [8]. Findings show that emotions and moods affect performance, with positive affective states favorably affecting problem solving and decision making [6]. Emotions are also contagious: people often catch each other's emotions. This has been confirmed for textual interfaces, where excitement and positive effects conveyed by word choice are transferred to the user [1].

CHI 2005, April 2–7, 2004, Portland, Oregon, USA.

ACM 1-59593-002-7/05/0004.

VOICE INTERFACE IN THE CAR

Driving is an activity that presents a context in which emotion can have enormous consequences. Attention, performance, and judgment are of paramount importance in automobile operation, with even the smallest disturbance potentially having grave repercussions. The road-rage phenomenon [3] provides one undeniable example of the impact that emotion can have on the safety of the roadways. In general, negative emotions, as well as over-relaxation, lead to poorer driving performance.

The issue of voices in cars primarily has been discussed in terms of cell phone use (or misuse). Car manufacturers are increasingly turning to voice as a promising strategy for safe and engaging interactions with drivers. There has been dramatic growth in a wide range of voice services in the car. Navigation, warning messages, restaurant recommendations, advertising, etc., are all voice-based, because drivers must have "eyes free" and "hands free" access to a wide range of information, and increasingly are grounded in two-way speech communication.

This trend makes it critical to know how emotion expressed by an in-car voice interface interacts with a driver's emotion in affecting attention, performance, and judgment. The consequences of using speech-based interfaces for in-car information systems are not well-understood. For example, might the emotional characteristics of the voice have as much impact on attention, performance, and judgment as the emotion of the driver? More specifically, what happens when the emotion of the voice and the emotion of the driver are mismatched, e.g., an upset driver encountering an upbeat voice? The most important possible effect of emotion matching is safety. However, because cars are fundamentally consumer products, designers cannot simply focus on safety: user feelings about the car are also critical for any voice interface to be successful.

EXPERIMENT DESIGN

Method

To investigate these questions we used a driving simulator that consisted of a PlayStation2 running the game, "Hot Pursuit." The game was configured to run a preset course, and all participants experienced the same properties for both driving conditions and car. The simulator was rear projected on a rear projection screen whose size was six foot diagonally. The participants drove a virtual car down a simulated country road (complete with other vehicles) using a gas pedal, a brake pedal, and a force-feedback steering wheel; there were no "game-like" elements in the interaction. All participants drove for approximately fifteen minutes. The participants' driving sessions were videotaped.

The experiment was a 2 (emotion of driver: happy or upset) by 2 (emotion of car voice: Energetic or Subdued) between-participants design, with random assignment to condition. Gender was balanced across conditions.

To address the question of the effects of user emotion on driving performance, at the beginning of the experiment, half of the participants were induced to be happy and the other half induced to be sad. This was accomplished by showing the participants a 5-minute video consisting of 37 six-second film and television clips derived from [2]. For happy participants, the videos reflected happy themes; for sad participants, the videos reflected sad or disturbing themes. A questionnaire served as the manipulation check.

All participants interacted with a "Virtual Passenger," represented by a professional actress that made light conversation with the driver. The Virtual Passenger introduced "herself" by saying, "*Hi. My name is Chris and I will be your Virtual Passenger for today. We are going to be driving today on a coastal road, one that I've traveled many times before, but one that may be new to you. The trip shouldn't take too long: ten or fifteen minutes. Let's get going.*"

At thirty-six separate points along the course, the Virtual Passenger made a different remark, for example, "*How do you think that the car is performing*?", "*Don't you think that these lanes are a little too narrow?*", and "*What kinds of things do you think about when you're driving?*" Half of the happy and half of the upset participants (randomly selected) drove with a voice that the actress was told to make "energetic"; for the other half, the actress was told to make the voice sound subdued. While formal speech patterns were not manipulated (as one can with a synthesized voice), the energetic voice had greater pitch range, amplitude range, and speed than the subdued voice; volume and pitch were identical.

While attention to the road is critical to driving, designers of car interfaces also want people to feel that interacting with the voice is an important part of the driving experience; otherwise, the Virtual Passenger would be no more than a radio. To assess drivers' engagement with the voice, participants were invited to speak to the Virtual Passenger as much or as little as they wished.

After completing the course, participants filled out a post-test questionnaire assessing their own emotion, the emotion of the voice and their perceived attentiveness.

Participants

40 adults, 20 female and 20 male, native English speakers with drivers' licenses were recruited from a temporary agency to participate in the study. Their ages ranged from 25 to 50. All participants were paid for their participation.

Measures

The effectiveness of the emotion manipulation was checked by a pre-test questionnaire, in which participants where asked to assess, using a variant of the Differential Emotion Scale [7], the positivity or negativity of their emotional state. The index was based on the question, "How well do each of the following adjectives describe how you feel?" followed by a list of adjectives based on five-point Likert scales (1=Describes Very Poorly to 5=Describes Very Well). The index was comprised of twelve adjectives: Ashamed, Fear, Sad, Scared, Shocked, Afraid, Dislike, Unhappy, Upset, Frightened, Guilt, and Revulsion. The index was very reliable (Cronbach's α=.95).

To assess the effects of the link between driver emotion and voice emotion on drivers' performance, we had two measures:

Number of accidents was manually coded from videotapes of the driving session.

Driver's attention was assessed by determining the driver's reaction time to a task-relevant stimulus. Drivers were instructed to honk their horn as soon as they heard a horn honk (18 honks were placed at particular points along the videotape). Greater speed is associated with greater attention because horn-honking is relevant to the driving task. This is in contrast to secondary-task reaction time experiments, in which greater response time is associated with *less* attention to the primary task.

Driver's perceived attention was based on a post-test questionnaire asking how well each of the following adjectives describes the participants' feelings while driving, each based on a 10-point Likert scale (1=Describes Very Poorly to 10=Describes Very Well). The index was comprised of four items: alert, careful, safe, and confident (α=.83).

Driver's engagement with the system was measured by the amount of time drivers spent talking back to the Virtual Passenger while driving down the simulated road.

RESULTS

The effects of the driver's emotion and the emotional coloring of the Virtual Passenger's voice were measured by a two (driver emotion) by two (voice emotion) between-participants ANOVA.

As expected, participants who saw the upsetting videotape were much more upset, M=2.88, SD=1.0, than were participants who saw the pleasant videotape, M=1.37, SD=0.29, based on a two-tailed t-test, $t(58)=7.9$, $p<.001$. Consistent with the voice manipulation, the energetic voice was perceived to be much more energetic, M=7.3, SD=2.1, than was the subdued voice, M=3.4, SD=*1.2*, based on a two-tailed t-test, $t(38)=5.1$, $p<.001$.

Matching the voice of the car to the drivers' emotions had important consequences (see Table 1). Drivers who interacted with voices that matched their own emotional state (energetic voice for happy drivers and subdued voice for upset drivers) had less than half as many accidents on average as drivers who interacted with mismatched voices (M=3.39 vs. M=8.95, $F(1,36)=5.54$, $p<.03$). This magnitude of reduction is far greater than the effects of virtually any technological change in a car at dramatically less expense; influencing the driver is more effective than reengineering the car. The effect of voice limited the usually significant difference in accident rate between both gender and happy and upset drivers, although female drivers and happy drivers tended to do better, $F(1,36)=2.1$, $p<.15$ and $F(1,36)=1.6$, $p<.18$, respectively. There was no main effect for voice emotion.

Matched groups (Happy Driver-Energetic voice and Upset Driver-Subdued voice) also communicated much more with the voice, even though the voice said exactly the same thing in all conditions, $F(1,36)=6.45$, $p<.02$. This is an important and in some sense surprising result. Although matched drivers spoke more with the Virtual Passenger and presumably paid more attention to the voice, this did not affect their driving performance; indeed, they drove better while speaking more. There were no main effects for driver emotion or for voice emotion.

Perceived attention to the road was assessed by a post-test questionnaire. Matching drivers (M=4.15, SD=1.92) perceived themselves paying much more attention to the road than did mismatched drivers, (M=2.82, SD=1.18), based on a two-tailed t-test, $t(38) = 2.4$, $p<.02$.

This result for perceived attention is consistent with the results for actual attention to the road. Matched participants tended to respond more quickly than mismatched participants, as indicated by the speed of horn honks, $F(2,36) = 3.18$, $p<.08$.

Table 1. Results

Variable	Happy Drivers		Upset Drivers	
	Energetic Voice	Subdued Voice	Energetic Voice	Subdued Voice
Number of Accidents	2.0 (2.8)	8.3 (9.2)	9.6 (10.6)	4.8 (5.2)
Amount of Speaking	5.3 (2.2)	4.1 (1.5)	4.0 (1.0)	5.3 (1.5)
Drivers' Attention (speed of horn honk)	3.3 (1.9)	6.5 (6.3)	6.9 (6.8)	4.4 (3.2)

VOICE EMOTION AND DRIVING PERFORMANCE

Pairing the voice of the car to the drivers' emotions had a strong effect both on drivers' performance and attitude. Emotion-inconsistent voices are arguably more difficult to process and attend to than emotion-consistent ones, leading to driver distraction [4]. Although there was a slight tendency for female drivers and happy drivers to be better drivers, even this effect was minimal compared to the effects of pairing.

The current research demonstrates that a very simple, inexpensive and fully controllable aspect of a car interface can have a dramatic influence on driver safety. Changing the paralinguistic characteristics of a voice is sufficient to significantly improve driving performance. Even with the same words spoken at the same times by the same voice under the same road conditions, driver performance can be strongly altered by simply changing the voice from energetic and upbeat to subdued. Designers can powerfully influence the number of accidents, the drivers' perceived attention to the road, and the driver's engagement with the car simply by changing the tone of voice.

A key finding here is that the same voice will not be effective for all drivers. For both actual and perceived performance, upset drivers clearly benefited from a subdued voice, while happy drivers clearly benefited from an energetic voice. This suggests that voices in cars must adapt to their users and raises two important questions: 1) How can an interface detect driver emotion? and 2) How can that information be used most effectively?

First, how should one assess the emotion of drivers? Future research should explore a number of possibilities including cameras to detect facial expression, sensors attached to the steering wheel, and voice analysis (made more effective when the car voice encourages conversation). Of course, changes in driving emotion during the session are also critical.

Second, designers must determine what should be varied when a particular emotion is detected. In the current study, we only varied paralinguistic cues that marked two specific emotions. Should the content that the car voice presents change? Should the amount of conversation by the car voice change? Important to address here is also the speed of change. If a normally happy driver enters the car in an upset state, should the car voice immediately change its voice characteristics? How inertial or responsive should the voice of the car voice be to changes in the driver's emotion? Should the car voice explicitly reveal its conclusions about the emotions of the driver? Will the driver deduce these conclusions from the behavior of the car voice? These are all critical questions.

The current study only examined two basic driver emotions—happy versus sad—and two basic voice emotions—aroused or subdued. Clearly, there are many other dimensions of emotion that should be explored. Beyond emotion, it is important to consider other aspects of driver characteristics and how they might interact with the car voice, including personality, gender, ethnicity, etc.

Whether the present results would occur during actual driving experiences is not clear; this research should be replicated with more realistic driving situations (e.g., fewer accidents). The present study relied on an actress's conscious manipulation of voice; it would be useful to verify (beyond a simple manipulation check) or manipulate (with synthetic voice) acoustic features associated with "energetic", "subdued", or other emotions to test the generality of the present findings. Finally, it is important to compare these results to those that would obtain from having no voice at all.

ACKNOWLEDGMENTS

This research was supported in part by Toyota Information Technology Center.

REFERENCES

1. Brave, S. and Nass, C. Emotion in human-computer interaction. In Jacko, J. and Sears, A. eds. *Handbook of human-computer interaction*, Lawrence Erlbaum Associates, (2002), 251-271.
2. Detenber, B.H. and Reeves, B. A bio-informational theory of emotion: Motion and image size effects on viewers. *Journal of Communication*, *46*, 3 (1996), 66-84.
3. Galovski, T.E. and Blanchard, E.B. Road rage: A domain for psychological intervention? *Aggression and Violent Behavior*, *9*, 2 (2004), 105-127.
4. Gong, L. *The psychology of consistency in human-computer interaction*. Doctoral dissertation. Communication, Stanford University, Stanford, CA. 2000.
5. Gross, J.J. Emotion and emotion regulation. In Pervin, L.A. and John, O.P. eds. *Handbook of personality: Theory and research*, Guildford, (1999), 525-552.
6. Isen, A.M. Positive affect and decision making. In Lewis, M. and Haviland-Jones, J.M. eds. *Handbook of Emotions*, The Guilford Press, (2000), 417-435.
7. Izard, C.E. *Patterns of emotions*. Academic Press, New York, 1972.
8. Nass, C. and Brave, S.B. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT Press, Cambridge, MA, 2005.
9. Picard, R.W. *Affective computing*. MIT Press, Cambridge, MA, 1997.