Masthead Logo
University of Iowa
Iowa Research Online

Driving Assessment Conference
2005 Driving Assessment Conference

Jun 28th, 12:00 AM

Matching In-Car Voice with Driver State : Impact on Attitude and Driving Performance

Ing-Marie Johnsson
Toyota Information Technology Center

Clifford Nass
Stanford University, Stanford

Helen Harris
Toyota Information Technology Center

Leila Takayama
Stanford University, Stanford

Johnsson, Ing-Marie; Nass, Clifford; Harris, Helen; and Takayama, Leila. Matching In-Car Voice with Driver State : Impact on Attitude and Driving Performance. In: Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, June 27-30, 2005, Rockport, Maine. Iowa City, IA: Public Policy Center, University of Iowa, 2005: 173-180. https://doi.org/10.17077/drivingassessment.1158

MATCHING IN-CAR VOICE WITH DRIVER STATE: IMPACT ON ATTITUDE AND DRIVING PERFORMANCE

Ing-Marie Jonsson[1], Clifford Nass[2], Helen Harris[1], Leila Takayama[2]
[1] Toyota Information Technology Center
Palo Alto, California, USA
[2] Stanford University
Stanford, California, USA
E-mail: {ingmarie, helenh}@us.toyota-itc.com
E-mail: {nass, takayama}@stanford.edu

Summary: This study examines whether characteristics of a car voice can influence driver behavior and attitudes, and furthermore, if the driver's emotional state will influence the perception of the car voice. Participants in a 2 (driver emotion: happy or upset) x 2 (car voice emotion: energetic or subdued) experimental study, had emotion induced by video clips and then spent 20 minutes using a driving simulator. While they were driving, a voice in the car spoke 36 questions (e.g., "Have we passed the sign for Lucia yet?") 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. Matching the car voice to the drivers' emotions had enormous consequences. Drivers who interacted with voices that matched their own emotional state had less than half as many accidents on average as drivers who interacted with mismatched voices! Drivers paired with matched voices also communicated much more with the voice (the voice said exactly the same thing in all conditions). The effects of matching emotion versus mismatching emotion were so powerful that neither driver emotion nor gender had an expected effect on driving performance. There was a slight tendency for happy drivers and female drivers to be better drivers, even though this effect was minimal compared to the effects of matching. In other words, finding the appropriate in-car voice for the driver's emotion stood out as the most critical factor in enabling a safe driving experience.

INTRODUCTION

Interactive information systems, both graphics-based and speech-based, are rapidly finding their way into the car. Current research and attention theory both suggest that speech-based interactions would enable drivers to focus more in the driving task than would interactions with a visual display. Results from speech-based interactions with in-car systems, however, show that drivers tend to take risks while talking, have slower reaction times, tend to slow down, and to increase their distance to other cars (Lee at al., 2001). Thus, speech-based interactions with in-car systems share many of the characteristics of mobile phone conversation, and there is the danger that poorly designed speech interfaces may show the same effect on driving performance.

Linguistic and paralinguistic cues such as choice of words, intonations and tone of voice all play a critical role in human-human and human-computer interactions. Voices, both computer-generated and recorded, manifest personality, gender, and accents (Nass & Brave, 2005). While these characteristics give us a broad sense of what a person will think and do, predicting a person's behavior at any given moment in time also requires attention to the user's particular feelings, knowledge, and the physical situation of the person.

Attitude, behavior, and cognition also can be influenced by momentary states such as feelings and situation. The most powerful momentary state is emotion. Emotion is not limited to the occasional outburst of fury when a person insults you, excitement when one wins the lottery, or frustration when trapped in a traffic jam. It is now understood that a wide range of emotions play a critical role in activities ranging from asking for directions to asking someone on a date.

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 (Gross, 1999). This avoidance of the study of emotions has extended to the realm of voice interfaces. Just as users have emotions, interfaces can manifest emotions (Picard, 1997). Textual interfaces can exhibit emotion through word choice, pictorial interfaces can smile or frown, and voice interfaces can exhibit emotion through tone of voice. Cues such as loudness, fundamental frequency, frequency range, and speech-rate distinguish dominant from submissive individuals and affect people interacting with systems using computer-generated speech (Nass & Gong, 2000). It has also been shown that most basic emotions or modes are associated with acoustic properties in a voice such as pitch range, rhythm, and amplitude or duration changes. For example, sadness is generally conveyed by slow and low-pitched speech, while happiness is associated with fast and louder speech.

VOICE INTERFACE IN THE CAR

Introducing speech-based interaction and conversation into the car highlights the influence of linguistic and paralinguistic cues. Previous studies with in-car information studies showed that the linguistic cues had a significant effect. The words used to inform the driver of bad driving performance have a clear positive effect only when they ascribed the bad performance to the driving environment rather than to the drivers themselves (Jonsson et. al, 2004). Emotional cues in speech interfaces have been studied in contexts other than the car, and findings show that emotions or mood are detected and affect performance. Positive affective states favorably affect problem solving and decision making (Hirt et al., 1996; Isen, 2000). 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 the interface are transferred over to the user (Morkes et al., 2000). An individual's current emotional state can also affect the experience of subsequent affective interactions. Residues from a previous affective reaction will combine with effects of emotions produced by subsequent affective stimulations. This can cause an overly intense affective reaction so that residual arousal from happiness may intensify their sense of well being, and residual unhappiness from sadness may intensify discomfort**.**

Speech is nearly as innate to humans as breathing. We speak to our cats, dogs and even to our plants, so why not to our cars or some virtual presence in the car? Driving 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 (Galovski & Blanchard, 2004) provides one undeniable example of the impact that emotion can have on the safety of the roadways. Considering the above discussion of the effects of emotion—in particular, that positive affect leads to better performance and less risk-taking—it is not surprising that research and experience demonstrate that happy drivers are better drivers (Groeger, 2000). Now that car manufacturers are increasingly turning to voice as a promising strategy for interactions with everything from in-car navigation systems and environmental controls to road-aware, chit-chatting co-pilots, it is critical to know how emotion expressed by an in-car voice interface interacts with a driver's emotion in affecting attention, performance, and judgment. 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 energetic and happy voice?

The present experiment was designed to investigate the effects of matching and mismatching a driver's mood with emotional cues used by the voice in the car. It was expected that happy drivers would exhibit better driving performance than upset drivers. This was based on results that positive affect helps performance (Isen, 2000). There was also an expectation that the energetic car voice would have a positive influence on the perception of the in-car system and the car. It was also expected that this effect would be intensified for happy drivers.

EXPERIMENT DESIGN

The overall aim of this research is to identify linguistic and paralinguistic properties of a car voice that is perceived favorably by the driver and improve driving performance. Particularly we wanted to answer three questions:

1. *Does the emotional coloring of the Virtual Passenger's voice influence the driver's emotion, performance, and perception of the car?*
2. *Does the driver's emotion influence driving performance and the perception of the car?*
3. *Will the emotion of the driver and the emotional coloring of the Virtual Passenger's voice interact to influence the driving performance and the perception of the car?*

To investigate these question we employed 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 predefined properties for both driving conditions and car. The simulator was projected on a six-foot diagonal projection screen. 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. All participants drove on the same simulated course, with highways, country roads, and cities for approximately fifteen minutes. The experiment was a 2 (emotion of driver) by 2 (emotion of car voice), between-participants design, gender balanced with random assignment to condition.

To address the question of the effects of user emotion on driving performance, it was necessary to have half of the participants be happy at the time of the experiment and the other half be sad and upset. The difficulty of ensuring the scheduling of equal numbers of equally happy and equally sad and upset participants led us to *create* emotion rather than rely on recruitment. Therefore, either a happy or sad and upset state was induced in each participant at the beginning of the experiment. This was accomplished by showing the participants 37 six-second film and television clips derived from Detenber (1996) and lasting for approximately five minutes. For happy participants, all of the videos reflected happy themes; for sad and upset participants, all of the videos reflected sad or disturbing themes. A pre-test questionnaire served as the manipulation check.

While driving the course, participants interacted with a "Virtual Passenger," represented by a recorded female voice 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?"; "Do you generally like to drive at, below, or above the speed limit?"; "Don't you think that these lanes are a little too narrow?"; "One of my favorite parts of this drive is the lighthouse…be sure to tell me what you think of it once you see it"; and "What kinds of things do you think about when you're driving?" For half of the happy and half of the sad and upset participants (randomly selected), the voice was energetic and happy while, for the other half, the voice was subdued and sad.

While everyone wants drivers to pay attention to the road, designers of car interfaces also want people to feel that the voice is an important part of the driving experience. To assess this, 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 the emotion of the voice and their perceived attentiveness.

Participants

Forty adults were recruited from a temporary agency to participate in the study. All participants had had driver's licenses for over five years and were native English speakers.

Measures

As an emotion-manipulation check, participants were asked to self-evaluate, on the pre-test questionnaire, how upset they were using a variant of the Differential Emotion Scale (Izard, 1972). 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, all measures on five-point scales. The index was very reliable (Cronbach's α=.95).

In the case of driving, safety is of utmost importance. To assess the effects of the link between driver emotion and voice emotion on drivers' performance, we had two measures:

- *Number of accidents* was coded manually by watching a video recording of each driver's simulator screen.

- *Driver's attention* was assessed by determining the driver's reaction time to a task-relevant stimulus. Specifically, drivers were instructed to honk their horn as soon as they heard a horn honk. There were 12 randomly-placed honks during the driving session. Because horn-honking is relevant to the driving situation, greater speed is associated with greater attention. 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 engagement with the system was measured by the amount of time drivers spent talking back to the Virtual Passenger while driving.

Perceived attentiveness was an index comprised of four items: alert, careful, safe, and confident (α=.83).

The items were based on the question, "How well do each of the following adjectives describe your feelings [about the Virtual Passenger/while driving]?" The questions were followed by a list of adjectives based on ten-point Likert scales ranging from Describes Very Poorly (=1) to Describes Very Well (=10).

RESULTS

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

Manipulation Checks

Participants who saw the sad and upsetting videos were much more upset, M=3.40, SD=1.3, than were participants who saw the happy and pleasant videos, M=1.64, SD=3.20, based on a two-tailed t-test, $t(38)=6.11$, $p<.001$. Likewise, the energetic/happy voice was perceived to be much more energetic, M=7.3, SD=2.1, than was the subdued/sad voice, M=3.4, SD=*1.2*, based on a two-tailed t-test, $t(38)=5.1$, $p<.001$.

Driving Performance

Matched participants (Happy driver-Energetic voice and Upset driver-Subdued voice) had significantly fewer accidents while driving than mismatched participants, as indicated by the cross-over interaction, $F(1,36) = 4.10$, $p<.05$ (see Figure 1). Happy drivers tended to have less accidents than upset drivers, but the result was not significant, $F(1,36) = 2.23$, $p<.14$.

Although not significant, there was also a cross-over interaction with respect to the response time to the horn honks, with matched participants tending to respond more quickly than mismatched participants, $F(1,36) = 2.49$, $p<.12$ (see Figure 2). There were no main effects.

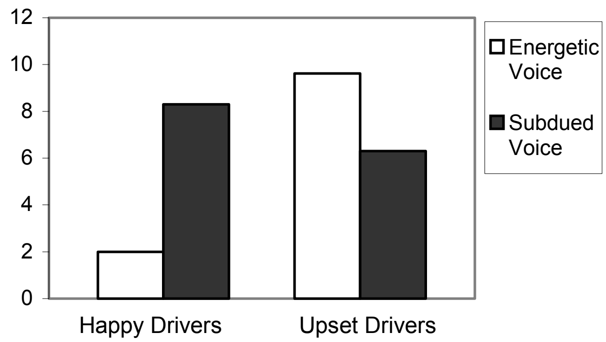

Figure 1: Number of Accidents

Figure 2: Accumulated Response Times to Honks

Matched groups (Happy Driver-Energetic voice and Upset Driver-Subdued voice) communicated more with the Virtual Passenger than mismatched groups, $F(1,36) = 4.50$, $p<.04$. There were no main effects for driver emotion or for voice emotion (see Figure 3).

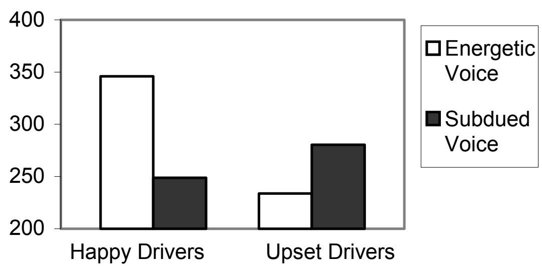

Figure 3: Amount of Communication with Virtual Passenger

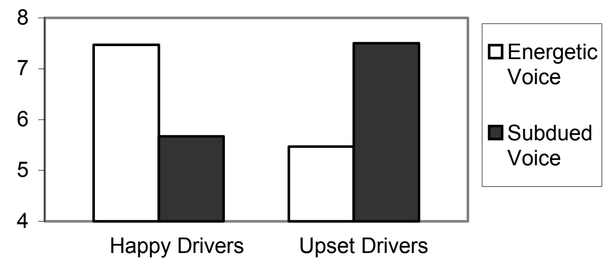

Figure 4: Attentiveness

There was a significant interaction for the driver's alertness, such that the matched drivers felt themselves to be more alert than did the mismatched drivers, $F(1,36) = 4.29$, $p<.05$ (see Figure 4). There were no main effects.

DISCUSSION

Matching the voice of the car to the drivers' emotions had enormous consequences. Drivers who interacted with voices that matched their own emotional state had *less than half* as many accidents on average as drivers who interacted with mismatched voices! Drivers paired with matched voices also communicated much more with the voice, even though the voice said exactly the same thing in all conditions. Note that although drivers who heard emotion-matched voices interacted much more, they were nonetheless better able to avoid accidents. The effects of matching emotion versus mismatching emotion were so powerful that neither driver emotion nor voice emotion by itself had a consistent effect on drivers. There was a slight tendency for happy drivers and female drivers to be better drivers, but this effect was minimal compared to the effects of matching. In other words, finding the appropriate in-car voice to match the driver's emotion stood out as the most critical factor in enabling a safe and engaging driving experience. People find it easier and more natural to attend to voice emotions that are consistent with their own, leading to better driving performance, more attention to the road, and a more enjoyable driving experience.

A key finding here is that the same voice cannot be effective for all drivers. This suggests that voices in cars must be able to both detect and adapt to a driver's emotion. Once the car makes a determination of the driver's emotion, the system must decide how to respond to that emotion. One strategy is to exhibit empathy by changing the emotion of the voice in step with the driver. Empathy greatly fosters relationship development, as it communicates support, caring, and concern for the welfare of another (Brave, 2003). A voice that matches a driver's emotions would strongly increase the connection between the user and the voice. Although rapid response to emotion of the user can be effective, there are dangers in this approach. Voices that respond and adapt to emotions at a too fast rate would increase cognitive load and be perceived as psychotic. While this can be entertaining when performed by manic comedians like Robin Williams or Jim Carrey, it is exhausting and disturbing when encountered in daily life. A rapid changing car voice would quickly be marked as manic-depressive instead of empathetic! Thus, emotion in technology-based voices must balance responsiveness and inertia by orienting to both fleeting emotions and longer term moods.

REFERENCES

Brave, S. (2003). *Agents that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent*. Doctoral dissertation. Communication, Stanford University, Stanford, CA.

Detenber, B.H. and Reeves, B. (1996). A bio-informational theory of emotion: Motion and image size effects on viewers. *Journal of Communication*, *46*, 3, 66-84.

Galovski, T.E. and Blanchard, E.B. (2004). Road rage: A domain for psychological intervention? *Aggression and Violent Behavior*, *9*, 2, 105-127.

Groeger, J.A. (2000). *Understanding driving: Applying cognitive psychology to a complex everyday task*. Philadelphia, PA: Psychology Press.

Gross, J.J. (1999). Emotion and emotion regulation. In Pervin, L.A. and John, O.P. (Eds.) *Handbook of personality: Theory and research*. Guildford, 525-552.

Hirt, E.R., Melton, R.J., McDonald, H.E. and Harackiewicz, J.M. (1996). Processing goals, task interest, and the mood-performance relationship: A mediational analysis. *Journal of Personality and Social Psychology*, *71*, 245-261.

Isen, A.M., Rosenzweig, A.S. and Young, M.J. (1991). The influence of positive affect on clinical problem solving. *Medical Decision Making*, *11*, 3, 221-227.

Izard, C.E. (1972). *Patterns of emotions*. New York: Academic Press.

Jonsson, I-M., Nass, C., Endo, J., Reaves B., Harris, H., Le Ta, J., Chan, N., Knapp, S., (2004). Don't Blame me I am Only the Driver: Impact of Blame Attribution on Attitudes and Attention to Driving Task, Proceedings of CHI 2004.

Lee, J., Caven, D., Haake, S., and Brown, T. (2001). Speech-based Interactions with In-Vehicle Computers: The Effect of Speech-based Email and Drivers' Attention to the Roadway. *Human Factors*, 43, pp. 631-640.

Morkes, J., Kernal, H., and Nass, C. (2000). Effects of Humor in Task-oriented Human-computer interaction and Computer-mediated interaction: A direct test of SRCT theory. *Human-Computer Interaction, 14(4),* 395-435.

Nass, C. and Brave, S.B. (2005). *Wired for speech: How voice activates and enhances the human-computer relationship*. Cambridge, MA: MIT Press.

Nass, C. & Gong, L. (2000). Social aspects of speech interfaces from an evolutionary perspective: Experimental research and design implications. *Communications of the ACM*, *43*(9), 36-43.

Picard, R.W. (1997). *Affective computing*. Cambridge, MA: MIT Press.