Detecting Emotions in Conversations Between Driver and In-Car Information Systems

Christian Martyn Jones[1] and Ing-Marie Jonsson[2]

[1] School of Mathematical and Computer Sciences,
Heriot-Watt University, Edinburgh, EH14 4AS, UK
cmj@macs.hw.ac.uk
[2] Department of Communication,
Stanford University, California 94305, USA
ingmarie@csli.stanford.edu

Abstract. Speech interaction with in-car controls is becoming more commonplace as the interaction is considered to be less distracting to the driver. Cars of today are equipped with speech recognition system to dial phone numbers and to control the cockpit environment. Furthermore satellite navigation systems provide the driver with verbal directions to their destination. The paper extends the speech interaction between driver and car to consider automatic recognition of the emotional state of the driver and appropriate responses by the car to improve the driver mood. The emotion of the driver has been found to influence driving performance and by actively responding to the emotional of the driver the car could improve their driving.

1 Introduction

Todays cars are fitting with interactive information systems including high quality audio/video systems, pin-point satellite navigation systems, hands free telephony and control over climate and ride conditions. Current research and attention theory suggests that speech-based interactions would be less detrimental to the driver than would interactions with a visual display [1]. With potentially more devices and more complex devices for the driver to control using speech as an interface is no longer a gimick but a necessity. Introducing speech-based interaction and conversation into the car highlights the potential influence of linguistic and paralinguistic cues. These cues play a critical role in human—human interactions, manifesting among other things, personality and emotion [2]. Previous studies show that alerting drivers to hazards in the road results in a more cautious and safer driving [3]. An in-car voice system was used to give the drivers relevant and timely road information, and thereby providing extra time and distance for them to evaluate the driving situation. Emotion of the car-voice was found to impact on driving performance. Results from a study matching and miss-matching the emotion of the car-voice and the emotion of the driver showed that matched emotions positively impacted driving performance [3]. These results make it interesting to investigate the feasibility of designing the emotionally responsive car.

J. Tao, T. Tan, and R.W. Picard (Eds.): ACII 2005, LNCS 3784, pp. 780 – 787, 2005.

2 An Emotionally Responsive Car

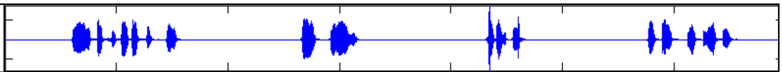
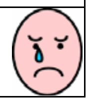
The development of an emotionally responsive car involves a number of technically demanding stages. In practise, the driver and car will converse two-way, where each will listen and respond to the others request for information and their emotional wellbeing. Systems to understand the requests from the driver and to retrieve and respond with appropriate information are not considered in this paper. Instead we concentrate on recognising the emotional state of the driver so that the car can modify its response both in the words it uses but also the presentation of the message by stressing particular words in the message and speaking in an appropriate emotional state. By the car altering its 'standard' voice response it will be able to empathise with the driver and ultimately improve the wellbeing and driving performance. This paper reports on our continued work to develop the acoustic emotion recognition part of the project and to test the technology within the car environment.

3 Development of Acoustic Emotion Recognition

There is considerable research interest in detecting and recognising human emotions automatically both academically [4],[5] and commercially [6]. Emotional information can be obtained by tracking facial motion, gestures and body language using image capture and processing; monitoring physiological changes using biometric measurements taken from the steering wheel and seat/seat-belt; and also analysing the acoustic cues contained in speech. Currently, video cameras and biometric sensors are not fitted as standard in cars, however speech controlled systems are already commonplace. Voice-controlled satellite navigation, voice-dial mobile phones and voice-controlled multimedia systems exist and drivers are educated and comfortable with their use. Therefore the project can incorporate voice-based emotion recognition without any requirement of additional hardware or changes to the driver's environment.

The acoustic emotion recognition system used in this project has been trained to recognise a range of emotions including boredom, sadness, grief, frustration, extreme anger, happiness and surprise. The system is trained on emotional speech obtained from United Kingdom and North American English speaking drama students at the Royal Scottish Academy of Music and Drama (RSAMD) using personalised and strongly emotive scenarios and constrained and free speech. All examples were validated in a blind listening study using human listeners before inclusion in our emotive speech corpus. The system uses 10 acoustic features including pitch, volume, rate of speech and other spectral coefficients and maps these features to emotions using statistical and neutral network classifiers. Using a test emotive speech data set the overall performance of the emotion recognition system is greater than 70% for 5 emotional groups of boredom, sadness/grief, frustration/extreme anger, happiness and surprise. The emotion recognition can track changes in emotional state over time and present its emotion decisions as a numerical indicator of the degree of emotional cues present in the speech. In addition, and to aid visualisation, the emotional state of the speaker can be displayed as an emotional face image, Table 1.

Table 1. Tracking changes in emotional state over time using an acoustic emotion recognition system. Plot shows speech waveform (top), numerical classification of the emotion state for happiness, surprise, anger/frustration, sadness/grief, the overal decision of the emotion recognition system (2nd bottom) and a graphical representation of the emotion for each second of time.

Speech waveform									
Boredom	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
Happiness	0.00	0.00	0.90	1.00	0.00	0.00	0.00	0.00	0.00
Surprise	0.00	0.00	0.15	1.00	1.00	0.00	0.00	0.00	0.00
Anger	0.06	0.00	0.00	0.00	0.00	1.00	1.00	0.00	0.00
Sadness	0.31	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.99
Decision	Bored	Bored	Happy	Not sure	Sur- prise	Anger	Anger	None	Sad
Emotional face images									

4 The Emotive Driver Project

The project builds on our initial research to assess the feasibility of automatically detecting driver emotions using speech [7]. The experiment sets out to record conversations between the car and the driver and to test the accuracy and validity of the automatic acoustic emotion recognition system in detecting, recognising and tracking emotions in the speech. The experiment consisted of an 8 day study at Oxford Brookes University, UK using 60 participants, 29 male and 31 female. Participants operated the STISIM driving simulator [8], Fig. 1, using an accelerator pedal, a brake pedal, and a force-feedback steering wheel, Fig. 2. All participants experienced the same pre-defined route and properties for both driving conditions and the car. The drive lasted approximately 20 minutes for each participant.

Fig. 1. STISIM simulator driving software showing front and rear views together with speedometer and rev counter

Fig. 2. Simulator driving hardware including pedals for an automatic car, steering wheel and car seat

Engine noise, brake screech, indicators, sirens etc together with verbal information from the car was played through stereo speakers. The information presented included instructions such as:

- This road is very busy during rush hour and traffic is slow
- This road too narrow and windy for cyclists
- The police often use radar here so make sure to keep to the speed limit
- This is a really windy stretch
- Pedestrians cross the road without looking in this school zone
- Some drivers need to pay more attention to the driving
- Parts of this road are under construction
- There is a traffic jam ahead if you turn left you might avoid it
- This road has many construction zones
- There is an accident ahead if you turn left here you might avoid it
- This stretch of road is always slippery
- This road is narrow and windy for slowly vehicles
- This stretch of road often has a problem with the fog

In addition the car attempts to engage the driver in conversation by saying:

- I get stress in traffic almost every day, how often do you get stressed with traffic problems?
- What do think about traffic?
- What do you think about the driving conditions?
- Do you generally like to drive at, above or below the speech limit?
- How do you like the car's performance?
- What types of situations makes you feel stressed whilst driving?
- I like to drive with people who talk to me, what is your favorite person to driver with?
- How do think you are driving?
- I like driving on mountain roads, what's your favorite road to drive on?

- How are you feeling right now?
- What kinds of things do you think about whilst driving?
- Do you find it stressful talking to people whilst driving?
- What do you think about this car?
- I'd love to ride in one of the new minis what is your favorite car to drive?
- I find that driving on narrow roads makes me feel stressed what kinds of roads makes you feel stressed?
- This is miserable, what's your strategy for coping with rain and fog whilst driving?

Speech from the participants was recorded using an Andrea directional bean with 4 microphones placed in front and about 1.5 meters away from the driver and recorded on a Sony Handicam DVD201. This microphone is typical of those used in the cars of today and provided a clean acoustic recording without overly sampling the car noise. The driving sessions were also videotaped from the front left of the driver to show driver hands, arms, upper body, head and eye motion, and facial expressions. Although this study does not consider image processing as a means to recognise driver emotions the video is used to correlate results from the acoustic emotion recognition with the emotions displayed in the faces of the drivers.

5 Results from Listener and Automatic Emotion Recognition Systems

The participants exhibit a range of emotions including boredom, sadness, anger, happiness and surprise, however for most of the drive the participants have a neutral/natural or a downbeat, bored/sad emotional state. When challenged in the drive by obstacles in the road, other drivers, difficult road conditions and pedestrians, we observe strong emotions from the drivers. During these emotional events the acoustic emotion recognition system must detect and recognise the emotional state of the driver.

By listening to the speech recording, a transcript of the drive has been created which includes not only the words of the conversation but also the emotional state of the driver. The same speech recording was processed by the acoustic emotion recognition system and its output classification represented as emotive faces for each second of the drive. The performance of the automatic emotion recognition system was determined by comparing the human emotion transcript against the output from the recognition system. The project is ongoing and we continue to process the speech data from drivers. An example speech and emotion transcript created by the human listener is included as Fig. 3, together with the emotional classification obtained automatically from the emotion recognition system Fig. 4.

There is a correlation between the emotional transcript created by the human listener and the emotion output returned automatically by the acoustic emotion recognition system. However there are occasions where the speech is masked by car noise (such as engine, brakes etc) and times when the driver could be one of two emotions such bored or sad, happiness or surprised. Classifying speech into emotions is challenging and there are occasions when the human listener is confused between sadness and boredom and happiness and surprise, rather like the automatic emotion recognition system. Grouping boredom and sadness together and calling this ‘downbeat’ may overcome much of these confusions. Similarly grouping happiness and surprise together as ‘upbeat’ also reduces confusions.

4.40 [car] What do you think about the driving conditions?

4.50 They are very good , dry, dry road  no rain, generally good I think 

5.20 Nice day 

5.30 [car] The police often use radar here so make sure to keep to the speed limit

5.35 Thank you 

6.40 [car] Do you generally like to drive at, above or below the speech limit?

6.47 Erm I generally drive just a little bit above  the speed limit, which probably explains  why I'm speeding a little bit now, but I generally also try and keep to it especially in 30 zones , erm and where the road is you know dangerous

7.15 Not like that at all 

7.30 [car] How do like the cars performance?

7.35 Erm  not exactly a Ferrari but it's okay, it's got a bit of a kick on it , handlings not great bit of under steer , can't really see myself as the next Jeremy Clarkson can you 

8.00 [car] This is a really windy stretch

8.05 Bloody ‘ell  yeh

8.20 [car] What types of situations makes you feel stressed whilst driving?

8.25 Erm generally er  I think traffic certainly has an affect, erm I think people blocking the road like this guy is now 

Fig. 3. Transcript of conversation between car and driver including the classification of emotional state of the driver as considered acoustically by the human listener

We have considered the correlation between the emotional transcript of the human listener and the automated emotion recognition system and can report that on average there is 60-75% correlation for the 5 emotional groups. The current range of emotions detected may not be the optimal range of emotions required for the emotionally responsive car and the performance of the emotion recognition can be improved by using the ‘downbeat’ and ‘upbeat’ groupings.

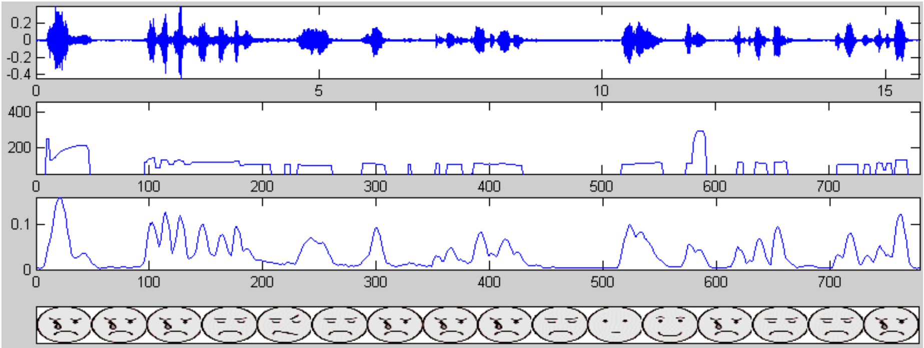


Fig. 4. Automatically generated transcript of the driver's speech using the emotion recognition system. Plots shown are waveform (top), pitch track (2nd), volume track (3rd) and the automatic classification of emotion (bottom) in the voice for a section of speech from 24.13-24.30 minutes. The driver has been asked "What is your favorite car to drive?" and responds by saying "Erm I'd go for an Austin Marin Vanquish and the whole prestige of the car, probably pretend to be James Bond and it's a lovely car". The emotion track shows that the driver is talking in a downbeat, matter-of-fact manner but gets enthusiastic and upbeat when talking about James Bond.

6 Extensions to the Current Research

The research is ongoing and we continue to analyse speech from the drivers and consider improvements for future work [7]. Although the in-car system asks questions of the driver many of the participants did not respond and engage in conversation. 25% of the participants did not converse with the car and thus we were unable to ascertain their emotional state acoustically. Future work will consider why the drivers did not talk with the car. Are they too focused on the task of driving? Do they not feel comfortable talking with the car? Do they not feel the car is listening to them and responding appropriately? We will adapt the conversational interface to help encourage conversation and enable the acoustic emotion recognition to gain insight into the mood of the driver.

Once the car knows the emotional state of the driver how should it adapt? Previous studies have considered varying the paralinguistic cues only [9], however should the content of the response also change, and how? Should the car become less or more talkative depending on the mood of the driver? Should the car alter the telematics, climate, music in the car in response the mood of the driver?

Further research should consider the affect of altering the car response and car environment to driver emotion. One strategy is to exhibit empathy by changing the emotion of the car-voice to match the user. Empathy fosters relationship development, as it communicates support, caring, and concern for the welfare of another. A voice which expresses happiness in situations where the user is happy and sounds subdued or sad in situations where the user is upset would strongly increase the connection between the user and the voice [10].

How fast should the emotion of voice change? Although rapid response to predicted emotion of the user can be effective, there are a number of dangers in this approach. Emotions can change in seconds in the human brain and body [11]. A sad person may momentarily be happy if someone tells a joke, but will fall back into their sad state relatively quickly. Conversely, happy drivers may become frustrated as they must slam on the brakes for a yellow light, but their emotion may quickly switch back to feeling positively. If the voice in the car immediately adapted to the user's emotions, drivers would experience occurrences such as the car-voice changing its emotion in mid-sentence. This would dramatically increase cognitive load constantly activate new emotions in the driver and be perceived as psychotic.

Mood must be taken into account to make the car-voice an effective interaction partner. Moods tend to bias feelings and cognition over longer terms, and while moods can be influenced by emotions, they are more stable and effectively filter events. A person in a good mood tends to view everything in a positive light, while a person in a bad mood does the opposite. Drivers that are in a good mood when entering a car are more likely to experience positive emotion during an interaction with a car-voice than drivers in a bad mood. Therefore it seems that emotion in technology-based voices must balance responsiveness and inertia by orienting to both emotion and mood.

Humans are what they feel: Performance, knowledge, beliefs, and feelings are to a large extent determined by emotions. People are also influences by voice interactions with people and interfaces. This makes it important for designers of speech based systems to work with linguistic and para-linguistic cues, especially emotional cues, to create the desired effect when people interact with the system.

References

1. Lunenfeld, H.: Human Factor Considerations of Motorist Navigation and Information Systems, Proceedings of Vehicle Navigation and Information Systems (1989) 35–42
2. Strayer, D., and Johnston, W.: Driven to Distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone", Psychological Science 12 (2001) 462–466
3. Jonsson et al.: Increasing Safety in Cars by Matching Driver Emotion and Car Voice Emotion, Proceedings of CHI 2005 (2005)
4. Cowie R. et al.: Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine (Jan 2001) 32-80
5. The Humaine Portal.: Research on Emotion and Human-Machine Interaction, http://www.emotion-research.net/ (2004)
6. Jones C.: Project to Develop Voice-Driven Emotive Technologies, Internal Report for Scottish Executive, Enterprise Transport and Lifelong Learning Department, UK (2004)
7. Jones C. and Jonsson I-M.: Speech Patterns for Older Adults While Driving, Proceedings of HCI International 2005 Las Vegas USA (2005)
8. STISIM Drive System, Systems Technology, Inc. California, http://www.systemstech.com/
9. 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
10. Brave, S.: Agents That Care: Investigating the Effects of Orientation of Emotion Exhibited by an Embodied Computer Agent, Doctoral Dissertation. Stanford University, CA (2003)
11. Picard, R.W.: Affective Computing. MIT Press, Cambridge, MA (1997)