Baselining Behaviour: Driving Towards More Realistic Simulations?

Alex W Stedmon
Chris Carter
Steven H Bayer

HUSAT Research Institute, Loughborough University, UK

Abstract

Automatic Speech Recognition (ASR) allows systems to be operated by speech input and may potentially improve the usability and safety of in-car systems. SPEECH-IDEAS is a LINK IST project investigating the use of ASR interfaces for in-vehicle systems. The success/application of in-car ASR relies on designing interfaces to match the expectations, preferences and abilities of various user groups. Driver workload (underload or overload) is a primary factor affecting the integration of in-car systems. Using multiple measures of workload, it is possible to assess relationships between actual task difficulty (objective measures), perceptions of task difficulty (subjective measures) and how individuals react to their perception of task difficulty (psychophysiological measures). This experiment sought to define a range of baseline driver workload factors in a simulator. Traffic behaviour (density, flow, speed changes, etc) and road layout/conditions (geometry, speed restrictions, fog, etc) were manipulated to assess the validity of different workload levels. Overall, the findings illustrated that increased workload affected driver performance and provide suitable baseline information for proceeding to assess the impact of in-car ASR on driving behaviour.

Introduction

Automatic Speech Recognition (ASR) allows systems to be operated by speech input. Whilst little attention has been given to ASR in the driving domain (Graham, Aldridge, Carter, & Lansdown, 1998), it may potentially improve the usability and safety of in-car systems including voice-dialling of mobile phones, operating entertainment systems and Intelligent Transportation Systems (ITS) such as route guidance or travel/traffic information services.

As in-car systems generally require drivers to operate visual displays and manual controls whilst driving (leading to long eyes-off-the-road times and compromising safety), a potential advantage of ASR is that it may allow for 'eyes-free' and/or 'hands-free' system use. Driving behaviour may benefit, therefore, from a transfer of loading from the over-burdened visual/manual modality to the auditory modality (Graham & Carter, 2000). However, the underlying assumption that speech exists as an untapped resource is a contentious issue (Stedmon, 1997) for speech may already be an active or semi-active mechanism (Linde & Shively, 1988).

ASR & Driving Behaviour

SPEECH-IDEAS is investigating the use of ASR interfaces for in-vehicle systems. In-car use of ASR is fundamentally different from a number of other domain applications. In particular, the task of using ASR is secondary to the primary task of safe driving and, therefore, the driver's attention to the ASR process (commands, menus, vocabulary, etc) may be limited. The in-car environment is typical of a 'hostile environment' (Baber & Noyes, 1996), characterised by high levels of noise, workload and stress that may affect the speech produced and subsequent recognition process.

The success/application of ASR in the driving domain relies on the careful design of the interface to match the expectations, preferences and abilities of various user groups. What might be a potential aid could just as easily prove to be hazardous by distracting drivers from the control of their vehicles (Stein, Parseghian, & Wade Allen, 1987).

Arousal, Performance & Workload

Ideally people operate at the peak of their arousal/performance (Weiner, Curry, & Faustina, 1984). Underlying this hypothesis is a concept of arousal expressed as an inverted U-curve, that assumes there is an optimal level of arousal that yields an optimal level of performance.

Driver workload (underload or overload) is a primary factor affecting the integration of in-car systems. Whilst driver underload will result in a

deviation from the top of the inverted U-curve back towards the 'low arousal - low performance' end of the scale; driver overload will cause a deviation across the scale towards 'high arousal - low performance'.

Psychophysiological, Subjective & Objective Workload

Psychophysiological measures of workload (heart rate, respiration, eye blink, brain activity, hormones, etc) provide information about an individual's reaction to their environment. Two main advantages in their use are: a lack of intrusion or need for overt participant responses, especially in a multi-task environment; and, continuous measurement throughout long/monotonous tasks, where subjective ratings may fail to detect rapid changes or transient peaks in workload (Wilson & Eggemeier, 1991).

A number of studies have sought to correlate different measures of workload, suggesting that heart rate and heart rate variability serve as an index of mental workload and time-on-task (Aasman, Mulder, & Mulder, 1987). Wilson & Eggemeier (1991) observed that heart rate variability also showed a negative correlation with subjective ratings of workload.

Objective workload may be measured as a function of the task difficulty/environment. In the driving domain workload may be manipulated through driving conditions (traffic, weather, road surface conditions, time of day etc). Measures for assessing driving performance may include: accidents, speed, lane deviation, speeding tickets, and traffic light violations (Stein, *et al*, 1987). Accidents are a clear measure of traffic safety and may be caused by lapses in attention, excess speed, poor speed control or poor lane keeping. Excess speed may cause a driver to lose control of their vehicle (due to road geometry or hitting obstacles) whilst driving faster or slower than other road users (poor speed control) may increase the likelihood of an accident. Lane deviation is another indicator of driving performance. If a driver's ability to maintain lane position is impaired, then the probability of exceeding the lane boundaries and hitting another object/vehicle also increases (Stein, *et al*, 1987). Speeding tickets and traffic light violations may be taken as indicators of driver Situational Awareness (SA). These parameters provide some indication of driver vigilance over the speedometer (inside the vehicle), and road signs, traffic lights and the behaviour of other traffic (outside the vehicle).

It is the particular sensitivies of these different measures that make their multiple use advantageous in providing a better understanding of the dynamics of workload. By their very nature multiple-task situations place various demands on human processing capabilities and no single measure is adequate in providing an overall appreciation of workload. By using multiple measures of workload, it is possible to assess relationships between actual task difficulty (objective measures), perceptions of task

difficulty (subjective measures) and how individuals react to their perception of task difficulty (psychophysiological measures).

Baselining Behaviour

This experiment sought to define a range of baseline driver workload factors, within a driving simulator, so that future comparisons could be made for workload effects on the use of in-car ASR. Within a series of scenarios traffic behaviour (density, flow, speed changes, etc) and road layout/conditions (geometry, speed restrictions, fog, etc) were manipulated to assess the validity of different workload levels. The simulator allowed for strict experimental control of the workload variables between participants, whilst ensuring driver safety and an ease of data collection.

Method

Participants 22 participants, recruited via HUSAT's participant database, consisted of 14 men (22-64 years, mean 38.1 years), and 8 women (21-45 years, mean 32.5 years). All participants had normal, or corrected to normal, vision, did not wear pacemakers and were not taking any medication. All participants held full UK driving licences, drove at least 2-3 times per week and 6,000 miles a year.

Apparatus Driving scenarios were generated and displayed using the full-size, interactive, HUSAT Driving Simulator running STI-Sim experimental software. Heart rate data were collected using ADI-Instruments MacLab/8 & Bio-Amp hardware and Chart v3.5 software. NASA-TLX and Bedford-Harper subjective workload questionnaires were administered.

Design A repeated measures, within-subjects, design was used. The independent variable (workload) was manipulated across 2x2x2 factors that produced eight conditions as outlined in Table 1 below:

Low Time Pressure	Control	Fog	Traffic	Fog & Traffic
High Time Pressure	**Control**	**Fog**	**Traffic**	**Fog & Traffic**

Table 1: Workload Variables in the Driving Scenarios

To minimise any order or carry-over effects, the conditions were counterbalanced across subjects using a Latin Square. The driving task was designed so that it represented, as far as possible, a real route. Changes were made to the layout so that aspects of road geometry (junctions, traffic lights, curves, hills, etc) were matched in all cases. Two

variations of the road layout were designed to minimise learning or fatigue effects. These were counterbalanced so that the driving task content remained homogenous, across subjects and conditions. This strengthened the analysis between scenarios and minimised any error variance from potential differences in the scenarios.

Dependent variable measures were collected for psychophysiological workload (heart rate & heart rate variability), subjective workload (NASA-TLX & Bedford-Harper scores), and objective workload (accidents, lane deviation, speed, speeding & traffic light violations).

Procedure After signing a consent form, participants received written and verbal instruction on the experimental procedure. Electrodes were connected for the heart rate analysis and a baseline measurement was taken. Participants underwent familiarisation and training in the simulator before running the experimental conditions. Prior to each condition participants received further instructions and immediately after each condition subjective workload scores were collected. Participants were given a short break half way through the experiment and upon completion were de-briefed and paid for their time.

Results

Mean data scores for the psychophysiological, subjective and objective workload were obtained for each scenario and analysed using a 2x2x2 (workload) within-subjects, and 2-way (sex) between-subjects ANOVA.

Heart Rate (HR) No significant main effects were observed for time pressure, traffic density or fog ($p>0.05$) illustrating that HR did not alter as a function of workload. In addition, no significant interactions were observed ($p>0.05$).

Heart Rate Variability (HR-V) Raw data were analysed using derivative signals between $0.02 - 0.13$ Hz. No significant main effects were observed for time pressure, traffic density or fog ($p>0.05$) illustrating that HR-V did not appear to alter as a function of workload. A significant interaction was observed for time pressure x fog [$F_{(1,20)} = 6.512$; $p< 0.05$], illustrating that when fog was present and time pressure increased, HR-V decreased in response to higher workload. No other significant interactions were observed ($p>0.05$).

NASA-RTLX Scores Significant main effects were observed for time pressure [$F_{(1,20)} = 39.44$; $p< 0.001$]; traffic density [$F_{(1,20)} = 61.93$; $p< 0.001$], and fog [$F_{(1,20)} = 4.724$; $p<0.05$], illustrating that as workload increased, drivers perceived the task to be more demanding across the six

scales of the NASA-TLX workload profile. No significant interactions were observed ($p>0.05$).

Bedford-Harper Scores Significant main effects were observed for time pressure [$F_{(1,20)} = 23.08$; $p< 0.001$], and traffic density [$F_{(1,20)} = 34.76$; $p< 0.001$], illustrating that as workload increased, drivers perceived the task to be more demanding. No main effect was observed for fog ($p>0.05$). A significant interaction was also observed for traffic density x fog [$F_{(1,20)} = 14.24$; $p< 0.01$], illustrating that when fog was present and traffic density increased, the driving task was perceived as being more difficult than when either workload variable was manipulated by itself. No other significant interactions were observed ($p>0.05$).

Road Traffic Accidents (RTAs) Significant main effects were observed for time pressure [$F_{(1,20)} = 8.46$; $p<0.01$] and traffic density [$F_{(1,20)} = 32.82$; $p<0.001$], illustrating that RTAs increased as a function of increased time pressure and traffic density. No significant main effect was observed for Fog ($p>0.05$). A significant interaction observed for time pressure x traffic density [$F_{(1,20)} = 9.64$; $p<0.01$] illustrated that as drivers came under more time pressure with increased traffic density, they were more likely to have an RTA.

Lane Deviation (LD) No significant main effects were observed for time pressure, traffic density or fog ($p>0.05$) illustrating that LD did not alter as a function of driver workload. Significant interactions were observed, however, for traffic density x sex [$F_{(1,20)} = 5.865$; $p< 0.05$], and traffic density x fog [$F_{(1,20)} = 4.633$; $p<0.05$], illustrating that women deviated more than men in low traffic, and that when fog was present and traffic density increased LD decreased significantly. No other significant interactions were observed ($p>0.05$).

Vehicle Speed A significant main effect was observed for time pressure [$F_{(1,20)} = 6.74$; $p<0.05$] illustrating that speed increased as time pressure increased. No other main effects were observed and no significant interactions were observed ($p >0.05$).

Speeding Offences A significant main effect was observed for traffic density [$F_{(1,20)} = 16.89$; $p<0.01$] illustrating that speeding offences increased as a function of traffic density. No other significant main effects were observed ($p>0.05$). A significant interaction was observed for traffic density x fog [$F_{(1,20)} = 4.65$; $p<0.05$] illustrating that when fog was present and traffic density increased, speeding offences increased more than for these variables in isolation.

Traffic Light Violations (TLVs) Significant main effects were observed for time pressure [$F_{(1,20)} = 16.47$; $p<0.01$], traffic density [$F_{(1,20)} = 6.92$;

p<0.05] and fog [F (1,20) = 11.93; p<0.01] illustrating that TLVs increased as a function of both time pressure and fog <u>but</u> decreased as a function of traffic density. A significant interaction was observed for traffic density x fog [F (1,20) = 4.89; p<0.05], illustrating when fog was present and traffic density increased TLVs decreased significantly. No other significant interactions were observed (p>0.05).

Discussion

The lack of significant data for heart rate and heart rate variability appear to undermine the subjective measures of workload that consistently rated the higher workload scenarios as being more demanding. It would appear, therefore, that even though drivers perceived the driving task to be more demanding, the psychophysiological basis of their behaviour did not alter.

This finding supports a study by Wilson & Eggemeier (1991), that compared real and simulated flying exercises and found no reliable heart rate variations in the simulator even thought they existed for real flight. They concluded that 'a subject's physiological responses in a simulator task could be different from those during actual flight, due to differences in the responsibilities and ... mental workload' (Wilson & Eggemeier, 1991).

Although participants drove faster and had more accidents under increased time pressure and traffic density, the frequency of these events was still very low. In a similar study (Stein, *et al*, 1987) found that, "despite high workload, accidents and speeding offences were infrequent events". This not only supports the findings of this study but also the realism of the simulations with participants appearing to drive much as they would on a real road.

It might have been expected that more speeding offences would have occurred under the increased time pressure. That the only significant effect was observed for traffic would seem to indicate that drivers were speeding up and slowing down with the traffic flow, perhaps intermittently breaking the limit. Speeding offences were only logged on the number of times a driver broke a particular speed limit, rather than the length of time spent speeding. As such, a driver who constantly slowed down and speeded up could invoke more penalties than someone who broke a speed limit and remained at that speed. This might explain why the time pressure condition showed an effect for vehicle speed but not for speeding violations, and why traffic density showed an effect for speeding violations but not for vehicle speed.

From the lane deviation data it appeared that participants drove consistently throughout the different conditions. They might have been expected to deviate more in the fog (without all the visual cues of other conditions) or under increased time pressure (when they might have been tempted to overtake traffic). The finding that women deviated more than

men in low traffic density is interesting but, with the ratio of male to female participants, has to be taken with some caution.

The results for traffic light violations show that TLVs increased as a function of both Time and Fog <u>but</u> decreased as a function of Traffic. This would seem to indicate that under increased time pressure drivers either could not slow down in time for the lights or consciously decided to drive through them. With fog, visibility was impaired and drivers may not have seen the road signs or traffic lights in time to stop. Under these conditions of increased workload attention may have been diverted (for different reasons) from looking at the speedometer (inside the car) or looking at road signs (outside the car) with subsequent effects on driver SA. The significant finding for traffic density would appear to indicate that other road users provided cues/reference points for stopping at traffic lights.

The interactions that were observed indicate, in general, how workload factors may compound their affects on task performance over simple main effects. Traffic and fog would appear to have the most impact across a number of measures illustrating that when fog was present and traffic density increased, the driving task was perceived as being more difficult than when either workload variable was manipulated in isolation.

These results support the notion that workload is highly problematic to define (Finch & Stedmon, 1998) and that, as Baber & Noyes (1996) state, "given the range of demands which can have a bearing upon workload, it would be difficult to provide a unified definition of the term". Workload is, therefore, more than merely *doing* a task, it encompasses an individual's *perception* of its complexity in relation to their *ability* to perform it.

The findings support the hypothesis that workload affects driver performance in a simulator and provides valuable baseline information for proceeding to assess the impact of in-car ASR on driving behaviour. Furthermore, these findings provide a more general resource for other trials where baseline behaviour is needed to compare performance/realism of a driving task in a simulator.

Acknowledgements

References

Aasman, J, Mulder, G., & Mulder, L.J.M. (1987), "Operator Effort and the Measurement of Heart Rate Variability". *Human Factors* 29, 161-170.

Baber, C., and Noyes, J. (1996), "Automatic Speech Recognition in Adverse Environments". *Human Factors* 38(1), 142-155.

Finch, M.I., & Stedmon, A.W. (1998), "The Complexities of Stress in the Operational Military Environment". In, M. Hanson (ed). *Contemporary Ergonomics 1998. Proceedings of The Ergonomics Society Annual Conference, Cirencester. 1998.* Taylor & Francis Ltd. London.

Graham, R., Aldridge, L., Carter, C., & Lansdown, T. (1998), "The Design of In-Car Speech Recognition Interfaces for Usability and User Acceptance". In, D. Harris (ed). *Engineering Psychology & Cognitive Ergonomics - Vol.4.* Ashgate.

Graham, R., & Carter, C. (2000), "Comparison of Speech Input and Manual Control of In-Car Devices whilst on the Move". *Personal Technologies* 4, 155-164.

Linde, C., & Shively, R. (1988), "Field Study of Communication and Workload in Police Helicopters: Implications for Cockpit Design". In, *Proceedings of the Human Factors Society 32nd Annual Meeting.* Human Factors Society, Santa Monica, CA. 237-241.

Stedmon, A.W. (1997), "Stressors, Speech Production, and Automatic Speech Recognition (ASR)". In, J. Isaac and R. Dangwal (eds). *Proceedings of the 3rd International Conference on Cognitive Subsystems.* New Delhi, India, 13-15 December, 1997.

Stein, A.C., Parseghian, Z., & Wade Allen, R. (1987), "A Simulator Study of the Safety Implications of Cellular Mobile Phone Use". *Proceedings of 31st Annual Meeting of the American Association for Automotive Medicine.* New Orleans, USA, 28-30 September, 1987.

Weiner, E.L., Curry, R.E., & Faustina, M.L. (1984), "Vigilance and Task Load: In Search of the Inverted U". *Human Factors* 26, 215-222.

Wilson, G.F., & Eggemeier, F.T. (1991), "Psychophysiological Assessment of Workload in Multi-Task Environments". In, D.L.Damos (ed). *Multiple Task Performance.* London: Taylor & Francis.