Comparison of Speech Input and Manual Control of In-Car Devices while on-the-move

Robert Graham & Chris Carter

HUSAT Research Institute
Loughborough University
The Elms, Elms Grove
Loughborough
Leics., LE11 1RG
UK.

tel: 01509-611088
fax: 01509-234651
email: r.graham@Lboro.ac.uk

Introduction

Devices which are used in the car, such as mobile phones, entertainment systems and intelligent transportation systems, have many of the same interface restrictions as other mobile and hand-held applications (PDAs,laptops, etc.) The 'dashboard real estate' available for displays and controls is very limited. This leads to difficulties providing complex information to the user, and feeding back the results of control operations, in standard visual display formats. Similarly, traditional controls (keys, buttons, etc.) must be used sparingly, and are not easily operated while on the move, due to limb vibration, etc. Of course the main problem for these applications is that the operation of the in-vehicle device is not the primary task; rather, the user must always concentrate on safe driving. Manual controls and visual displays require the eyes to be taken off the road and the hands to be taken off the steering wheel. Operating complex visual-manual interfaces while on the move has been shown, in real and simulated driving environments, to result in significant decrements in safety-related behaviours such as lane-keeping, headway judgement and hazard detection. For example, the safety issues associated with using a mobile phone while driving have been widely reported in the media and in the research literature (e.g. Brookhuis, De Vries, & De Waard, 1991; Alm & Nilsson, 1995).

Speech interfaces, involving automatic speech recognition (ASR) and possibly speech output (synthesised or recorded speech), have a number of potential advantages in this context. Primarily, the speech modality is both hand-free and eyes-free, meaning that the hands can be kept on the wheel and the eyes on the road. Speech is a natural, everyday activity and therefore may be more easily learnt and better accepted by users (at least, if the speech interface is designed in a user-centred way). Speech control may allow system functions and sub-functions to be more easily accessed than via some menu structure within a visual-manual interface. Speech input/ output can also be used by the large proportion of the user population who have poor eyesight and/or poor manual co-ordination.

Despite the potential advantages of speech technology for in-car and other mobile systems, the area is surprisingly under-researched and under-developed. Most of the larger telecommunications companies have been investigating speech input for phone functions for some time, but speech control technology is currently only available on the market for a limited range of applications (e.g. voice dialling by keyword) on a limited number of handsets (e.g. Philips Genie, Ericsson T18). In the automotive industry, the recently-launched Jaguar S-Type is the first car which allows speech input for more than a single system, supporting the control of phone, entertainment system and climate control functions. Although speech recognition is now widely available for PC dictation functions, the successful applications tend to be in quite specialised domains where standard keyboard input is impractical.

The current research programme aims to address this unfulfilled potential by investigating the use of speech input for in-car systems while on the move. The fundamental question being asked is: in which situations (i.e. for which users, doing which tasks, in which environments) does speech control provide advantages over traditional manual control? The relative performance of the speech and manual options for various functions is assessed by measuring (a) task performance, (b) concurrent driving performance, and (c) subjective responses. The research also considers the optimal design of the speech dialogue, in particular the nature and modality of the feedback provided to the user of the results of the speech recognition process.

Method

Two laboratory studies are reported in this paper. The first considers the mobile phone application, and the second, the in-car entertainment (ICE) application.

In the first study, forty eight participants carried out a driving-related laboratory task while simultaneously dialling phone numbers. The driving task consisted of a tracking component (using a steering wheel to keep a moving block on a PC screen within another block) and a target detection component (hitting the brake pedal in response to a peripheral stimulus). Three phone interface modalities were tested: a standard manual phone and a speech-controlled phone with either auditory-only or auditory-plus-visual feedback. The study used a Nokia Orange manual handset and an AURIX speech recognition unit, produced by the Speech Research Unit at DERA Malvern. To initiate dialling in the speech conditions, the user pressed a steering-wheel-mounted press-to-talk (PTT) button. They issued the command word 'phone', followed by the digits in chunks of any size, followed by the word 'dial'. Feedback was given each time the PTT was released; this consisted of a female voice repeating the recognised words, with or without a dashboard text display. A within-subjects design was used, with each participant experiencing each of six blocks (three interface modalities, either alone or with concurrent driving). Dialling of familiar numbers was prompted by the experimenter (e.g. "now call your mum") and the timing of numbers was randomly distributed across each 8-minute block to avoid anticipation. In all interface conditions, users were instructed to correct any errors in the phone number before they issued the dial command.

The second study tested thirty naïve participants with the ICE application, in a similar laboratory environment. Four alternative control interfaces were examined: standard manual dashboard buttons, steering wheel buttons, and speech operation either with or without explicit auditory-plus-visual feedback. All of the interfaces were provided by Jaguar Cars, and were almost identical to those incorporated into their S-Type vehicle. Various operations were tested, including selecting a track and/or disc number on the CD player (e.g. "CD play disc 8 track 8"), tuning the radio to a pre-set channel (e.g. "radio pre-set 2") and playing the tape ("tape play"). Feedback in the speech condition consisted of a male voice repeating the whole recognised command, and a visual message display mounted just below the speedometer. The speech-without-feedback option was simulated by simply turning the voice output off and covering the visual display. Again, the design was within-subjects and split across nine blocks (four interface modalities, either alone or with concurrent driving, plus a last control condition of driving without the ICE tasks).

Both studies collected a similar set of data, including driving performance measures (tracking error, reaction time to peripheral targets), task performance measures (transaction time, number of input errors) and subjective measures (perceived mental workload, attitude questionnaire ratings).

Results

At the time of writing (May 1999), the analysis of results from the second study is not complete. Therefore, this section will concentrate on the results of the first study only.

It was found that concurrent driving performance was significantly poorer in the manual control condition than the speech input conditions, both in terms of tracking error ($F(2,90)=43.7$, $p<0.0001$) and target reaction time ($F(2,94)=3.79$, $p=0.026$). Supporting this finding, participants' ratings of mental workload on the NASA-RTLX scale indicated that they found using manual phone controls while driving significantly more demanding than using either of the speech input interfaces ($F(2,90)=4.04$, $p=0.020$).

However, task performance in the speech input conditions was significantly worse than the manual control condition. Transaction time, measured as the time between the instruction from the experimenter and the initiation of dialling, was longer in the speech input conditions (M=33 seconds for the 'phoning while driving' block) than the manual control condition (M=22 seconds) ($F(2,80)=125$, $p<0.0001$). Also, numbers in the manual control condition were dialled more accurately than those in the speech input conditions ($F(2,80)=6.3$, $p=0.002$). On average, only 89% of phone calls would have got through to the intended recipient when using the speech input modality while driving, compared with 95% when using the manual controls.

The feedback modality in the speech input conditions had a small but significant effect on performance. When visual as well as auditory feedback was provided, this distracted from the tracking task, even though participants reported that they did not consciously pay much attention to the visual display ($t(47)=3.0$, $p=0.004$, comparing the combined feedback with auditory-only).

Despite the fact that the speech recogniser made a significant number of recognition errors (word recognition rate was only 89% across the whole trial), and that task performance was poorer for the speech conditions, users' attitudes to the speech interface were very positive. They rated the speech interface easy to learn, logical and useful, and most stated that they would prefer to have a speech-operated phone in their car than a manual phone. The particular phone dialling interface used in the trial was also very well received.

Discussion

This study confirmed the results of previous research in finding that using manually-operated in-car systems while on the move has a detrimental effect on both driving performance and task performance. The effects on driving can be significantly reduced, but not eliminated, through the use of speech input. This finding is clearly of relevance to the automotive industry and their suppliers, who are justifiably concerned at the potential for their systems to contribute to accident causation. It also has implications for legislation (it should be noted than in the UK, using a mobile phone or other in-car devices while on the move is 'advised against' in the Highway Code, but not illegal).

However, the finding that speech input was slower and less accurate than manual control (at least for the phone dialling function) may be of more general interest. The speech recogniser in the experiment made a large number of recognition errors, and the subsequent error correction process clearly added to the transaction times. We would expect recognition rates in a production system to be significantly better than those obtained in the experiment, and, therefore, this effect to be reduced. The reasons for the reduced dialling accuracy are less clear cut. Dialling accuracy was measured as the proportion of numbers incorrectly dialled, after speech recognition errors and their correction had been taken into account. Therefore, poor dialling accuracy must be due to poor monitoring of the recognition feedback; that is, when the recogniser made a mistake, users did not always realise that it had made a mistake. This was apparent both when using the phone while driving and when using it alone, and for both the . auditory-plus-visual and auditory-only feedback conditions. It implies that, if input accuracy is important, speech may not be the best control modality for in-car and other mobile devices.

Where speech control is incorporated into the user interface, it is vital to design the vocabulary and dialogue according to good human factors principles. The phone dialling dialogue used in the current study was quite unconstrained - users could input digits in chunks of any length, they could cancel or correct inputs at any time, they could enter numbers in different ways (e.g. "double three" instead of "three three") - and this was both efficient and well liked. It would be instructive for future work to compare this dialogue with some of the other alternatives being incorporated into real and prototype phone applications. Feedback and error correction are perhaps the most important dialogue issues. The study confirmed that auditory-only feedback is probably the best option for a car environment, to avoid the costs (in terms of distraction from the road) of visual feedback. For other domains, where visual feedback does not have any significant associated cost, a redundant combination of auditory and visual feedback would usually be preferable (ensuring that the content and timing of fed-back information is consistent across the two modalities).

The subjective responses from the study were also relevant to the applicability of speech control for mobile systems. They showed that, even when recognition accuracy is poor, transaction times are slow, and task error rates are high, speech input can be perceived as preferable to manual control. Therefore, although the designer may struggle to maximise these performance measures, the success of speech recognition will ultimately depend on whether or not users perceive a significant advantage for speech over its manual alternatives. The current

study showed that such an advantage is seen for using a mobile phone in a car, but further work is necessary to determine users' perceptions for other applications. Within the current research programme, a tool is currently under development to determine which factors are important in users' overall attitudes towards speech interfaces.