Audio and Visual Cues in a Two-Talker Divided Attention Speech-Monitoring Task

Douglas S. Brungart, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, **Alexander J. Kordik,** Sytronics, Inc., Dayton, Ohio, and **Brian D. Simpson,** Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio

Although audiovisual (AV) cues are known to improve speech intelligibility in difficult listening environments, little is known about their role in divided attention tasks that require listeners to monitor multiple talkers at the same time. In this experiment, a call-sign-based multitalker listening test was used to evaluate performance in two-talker AV configurations that combined zero, one, or two channels of visual information (neither, one, or both talkers visible) with zero, one, or two channels of audio information (no audio, both talkers played from the same loudspeaker, and both talkers played through different, spatially separated loudspeakers). The results were analyzed to determine the relative performance levels that would occur with each AV configuration with target information that was equally likely to originate from either of the two talkers in the stimulus. The results indicate that spatial separation of the audio signals has the greatest impact on performance in multichannel AV speech displays and that caution should be used when presenting a visual representation of only a single talker unless that talker is known to be the highest priority talker in the combined AV stimulus. Potential applications of this research include the design of improved audiovisual speech displays for multichannel communications systems.

INTRODUCTION

In recent years, improvements in data transmission technology have dramatically reduced the cost of telecommunications bandwidth. To this point, however, relatively little effort has been made to exploit this low-cost bandwidth in improved speech communications systems. In part, this apparent oversight reflects the fact that standard telephone-grade audio speech (with a bandwidth of roughly 3500 Hz) already produces near 100% intelligibility for typical telephone conversations involving a single talker in a quiet listening environment. There is, however, ample opportunity for higher-bandwidth speech communication systems to improve performance in complex listening tasks that involve more than one simultaneous talker. High-bandwidth multichannel speech communication systems could have a wide variety of applications, ranging from simple three-way conference calling to sophisticated command and control tasks that require listeners to monitor and respond to time-critical information that could be present in any one of a number of simultaneously presented competing speech messages.

A question of practical interest, therefore, is how additional bandwidth could best be allocated to improve the effectiveness of multichannel speech communications systems. The most obvious approaches to this problem involve the restoration of the audio and visual cues that listeners rely on to segregate speech signals in real-world multitalker listening environments, such as crowded restaurants and cocktail parties. For example, listeners in the real world rely on interaural differences between the audio signals reaching their left and right ears to help them segregate the voices of spatially separated talkers (see Bronkhorst, 2000, for a recent review of

Address correspondence to Douglas S. Brungart, AFRL/HECB, 2610 Seventh St., WPAFB, OH 45433-7901; douglas.brungart@wpafb.af.mil. **HUMAN FACTORS,** Vol. 47, No. 3, Fall 2005, pp. 562–573.

this phenomenon). When these binaural cues are restored to a speech communication signal by adding a second independent audio channel to the system, multitalker listening performance improves dramatically (Abouchacra, Tran, Besing, & Koehnke, 1997; Crispien & Ehrenberg, 1995; Ericson & McKinley, 1997; Nelson, Bolia, Ericson, & McKinley, 1999).

Additional bandwidth could also be used to restore the visual speech cues that are normally available in face-to-face conversations. These cues make it possible to extract some information from visual-only speech stimuli (a process commonly known as speechreading; Summerfield, 1987), and they contribute substantially to audiovisual (AV) speech perception when the audio signal is distorted by the presence of noise (Sumby & Pollack, 1954) or interfering speech (Rudmann, McCarley, & Kramer, 2003).

From earlier experiments, it is clear that multitalker listening performance can be improved both by the addition of binaural spatial audio cues and by the addition of visual speech information. However, relatively little is known about how audio and visual information might interact in high-bandwidth multichannel AV speech displays. Important research issues related to this topic include the following:

Divided attention versus selective attention in AV speech perception. An essential underlying assumption in the design of a multitalker speech display is that neither the system nor the operator will have reliable a priori knowledge about which talker will provide the most important information at any given time. (Otherwise, either the system or the operator would simply turn off the uninformative talkers.) Consequently, it is important to know how well listeners are able to divide their attention across the different talkers in an AV speech stimulus in order to extract important information that might originate from any one of the competing speech signals. However, virtually all experiments that have examined AV speech perception with more than one simultaneous audio speech signal (Driver, 1996; Driver & Spence, 1994; Reisberg, 1978; Rudmann et al., 2003; Spence, Ranson, & Driver, 2000) have examined performance in a selective attention paradigm in which the participants were provided with a priori information about which talker to attend to and which talker to ignore prior to the presentation of each stimulus. This makes it difficult to determine how visual cues might influence performance in situations in which listeners must rely on the content of the competing speech messages to determine where the most important information resides.

AV speech perception with multiple visible talkers. Although a number of studies have examined AV speech perception with multiple simultaneous talkers, most have been limited to cases in which only a single talker was visible at any given time (Driver, 1996; Driver & Spence, 1994; Reisberg, 1978; Rudmann et al., 2003; Spence et al., 2000). Thus it is not clear how well listeners might be able to divide their visual attention across two visible faces in a multitalker AV speech stimulus.

Semantic AV incongruencies. When visual speech stimuli are presented in conjunction with mismatched audio speech stimuli, cross-modal interactions can substantially distort the overall perception of the multimodal stimulus. One classic example of this is the McGurk effect, which causes listeners who see one word spoken and hear another word spoken to report the perception of a third word that was not presented in the stimulus (e.g., they report hearing an "ada" sound when they see a talker saying "aga" and hear a talker saying "aba"; McGurk & McDonald, 1976). Although it is unlikely that AV speech display would intentionally present mismatched audio and visual signals, it is possible that AV interactions such as the McGurk effect could cause visual cues produced by one talker to interfere with the perception of the audio speech signals produced by one of the other competing talkers in the AV stimulus.

Spatial AV incongruencies. Another cross-modal interaction that has been reported in AV speech perception is the so-called ventriloquist effect, in which listeners hear speech sounds at the location of a visual representation of the talker's face even when that face is spatially displaced from the origin of the audio signal (Bertelson & Radeau, 1976). The ventriloquist effect can influence AV speech perception in listening configurations in which the audio and visual portions of the target speech stimulus are spatially separated.

Spence and his colleagues have shown that

AV speech perception is better when the audio and visual signals are colocated than when they are spatially separated (Driver & Spence, 1994) and that interfering speech signals are more difficult to ignore when they are located at the point of visual fixation (Spence et al., 2000). Similarly, Driver (1996) has shown that performance in an AV speech perception task that combines a visual representation of the target talker's face with an audio mixture containing both the target and masking talkers' voices improves when the visual representation of the face is spatially displaced from the audio mixture, presumably because ventriloquism causes the apparent location of the target voice to shift toward the location of the face and away from the location of the masking voice. These results are important in the design of AV multitalker speech displays because they suggest that the relative spatial locations of the audio and visual portions of a multimodal speech stimulus can substantially influence the overall perception of that stimulus.

In this paper, we present the results of an experiment that attempted to address some of these research issues by examining performance in a two-talker divided attention AV listening task with all possible combinations of three levels of audio fidelity (no audio, one-channel audio, and two-channel audio) and three levels of video fidelity (no video, one visible talker, and two visible talkers). When applicable, performance was also examined in configurations in which the locations of the audio and visual portions of the AV stimuli were presented at the same locations (spatially congruent) and in which the audio and visual portions were presented at different locations (spatially incongruent). In all cases, performance was analyzed under the assumption that the target information was equally likely to originate from either of the two talkers in the stimulus. The results are discussed in terms of their implications for the design of multitalker AV speech displays.

METHOD

Participants

Eight volunteers (5 men and 3 women) were paid to participate in the experiment. All had normal hearing (<15 dB HL from 500 Hz to 8 kHz), and their ages ranged from 21 to 55 years. All of the participants had taken part in previous experiments that utilized speech materials similar to those used in this experiment.

Audio and Visual Speech Materials

The experiment was conducted with an audiovisual speech corpus based on the Coordinate Response Measure (CRM), a call-sign-based color and number identification task developed by the Air Force Research Laboratory for use in multitalker communications research. The CRM, which has been used in a number of studies of multitalker speech perception with audio-only stimuli (Brungart, Simpson, Ericson, & Scott, 2001), is ideally suited to testing the ability of listeners to divide their attention across two simultaneous stimuli because it requires them to identify the target phrase by listening for a preassigned call sign contained within that phrase.

The CRM speech materials used in this experiment consisted of phrases of the form "Ready (call sign) go to (color) (number) now" spoken with all possible combinations of two call signs (Ringo and Baron), four colors (blue, green, red, white), and eight numbers (1–8). Thus a typical example of a CRM sentence would be the phrase "Ready Baron go to blue five now."

The CRM phrases were recorded in the corner of a large anechoic chamber with a digital video camera (Sony Digital Handycam) located roughly 1.5 m in front of the talker. The talkers (1 man and 1 woman) stood in front of a black, acoustically transparent background covering the wedges on the wall of the anechoic chamber and were instructed to repeat the CRM phrases at a monotone level while keeping their heads as still as possible. Breaks were inserted between the CRM phrases to avoid any effects of coarticulation between consecutive recordings. The resulting videotapes were downloaded onto a PC, where they were partitioned into different individual Audio Video Interleaved (AVI) files for each of the 128 recorded phrases. Then a commercially available video editor (VirtualDub) was used to crop the frames of the AVI files around the locations of the talkers' heads, convert them from color to gray scale, and compress them with the Indeo 5.1 codec. Finally, the AVIs were time aligned by padding enough silent repetitions of the first video frame in each file to set the onset of the word "ready" to the same temporal position within each AVI file. The audio signals associated with each AVI file were also normalized to have the same overall root mean square level. Figure 1 shows example frames from the CRM AVIs for the male and female talkers used in this experiment.

Apparatus

The experiment was conducted in a 5.5- × 5-m sound-treated conference room with the configuration shown in Figure 2. The participant was seated at the center of the long edge of a 3.3- × 1.2-m conference table located along the midline of the room. Directly across the table from the participant (roughly 1.37 m from the location of the head), a full-sized white cotton sheet was hung from the ceiling perpendicular to his or her line of sight. This sheet, which was held taut by a weighted wooden rod along its bottom edge, served as a projection screen for the visual stimuli used in the experiment. The visual stimuli were projected onto the sheet at eye level by a Boxlight CD-40m projector located on a shelf above and behind the participant's chair. The projector was connected to the SVGA output of a PC-based control computer, which used ActiveX calls to the Microsoft Windows Media Player to present the left and right visual stimuli (AVI files) at the left and right sides of a single MATLAB figure window. The system was run in an 800 × 600 video output mode, and the video files were each 150 × 150 pixels, with a separation of 118 pixels between the two faces. This resulted in an image size of 0.40 × 0.60 m for each of the two AVI frames and a separation of 0.72 m between the midpoints of the two faces on the screen. Note that this orientation results in face locations at +15° azimuth and –15° azimuth relative to the location of the participant (identical to the configuration used by Driver, 1996).

Figure 1. Example frames from the CRM AVIs recorded for the male and female talkers used in the experiment.

The loudspeakers used to generate the audio portions of the stimuli (Bose JewelCube) were mounted on stands behind the acoustically transparent screen. These speakers were adjusted to place them directly behind the mouths of the two faces projected on the front of the screen. The speakers were driven by a stereo amplifier connected to the sound card of the control PC.

Audiovisual Display Configurations

The experiment was designed to examine all possible combinations of three audio conditions and three visual conditions that could reasonably occur in a two-talker AV speech display. The audio conditions were a *no-audio* condition, in which neither talker's voice was presented, a *one-channel audio* condition, in which both of the talkers' voices were mixed together and presented from the same loudspeaker location, and a *two-channel audio* condition, in which each talker's voice was presented from a different loudspeaker location. The video conditions were a *no-video* condition, in which neither talker's face was visible, a *one-channel video* condition, in which only one of the talkers' faces was visible, and a *two-channel video* condition, in which both talkers' faces were visible. Figure 3 shows the eight AV conditions that result from different combinations of these audio and visual conditions. The third column of the figure shows the basic template of each AV condition for two arbitrary talkers, labeled A and B; Talker A was always located to the left of the participant. These templates illustrate the AV conditions pictorially by a small box with two circled positions at the top, representing the talkers shown visually at the left and right sides of the screen, and two uncircled positions at the bottom representing the talkers presented auditorily from the left and right loudspeaker locations (as indicated by the legend in the upper-right of the figure).

In a normally configured ("spatially congruent") AV display, in which the audio signals for

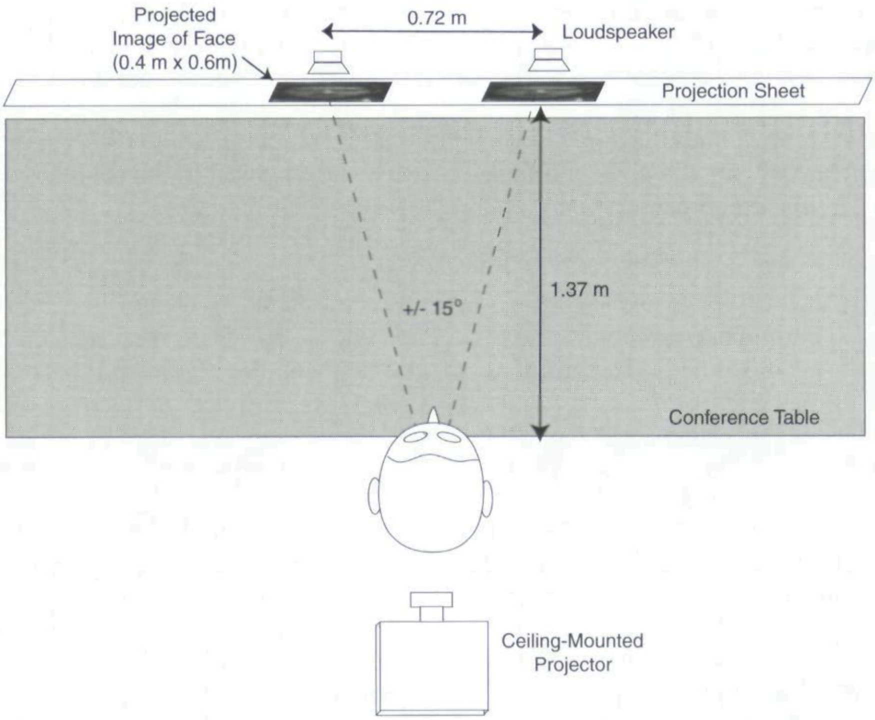

Figure 2. Configuration of listening room used in the experiment.

each talker originate from the same locations as the visual representations of those talkers, each of these basic AV templates corresponds to either two or four different spatial configurations, depending on whether the target talker is designated by A or B in the template and whether Talker A in the template is located on the participant's left or right side. The fourth column of Figure 3 illustrates the 20 possible congruent spatial configurations that could originate from the eight basic AV templates. In configurations with at least one audio channel and at least one video channel, it is also possible to construct "spatially incongruent" variations for each template, in which the target and/or masker voice originates from a location different from that of the corresponding face. The rightmost column of Figure 3 shows the 12 possible incongruent spatial configurations corresponding to the templates with one or more audio channel and one or more video channel. Thus the last two columns of Figure 3 show a total of 32 different spatial configurations corresponding to the eight basic AV conditions shown in the figure, with spatial configurations that were mirror images of each other labeled as "a" and "b" versions of each of 16 unique configuration conditions (labeled Conf 1–16a and Conf 1–16b). All 32 of these possible spatial configurations were tested in the experiment.

Procedure

The experiment was conducted with the participant seated in front of the video screen in the darkened conference room. In each trial, the participant was presented with a stimulus containing the audio and/or video portions of two CRM phrases spoken simultaneously by the same talker: a target phrase, selected randomly from the 32 phrases containing the call sign "Baron," and a masking phrase, selected randomly from the 21 phrases containing the call sign "Ringo" and a color and number different from that of the target phrase. These two phrases were presented auditorily at the left or right loudspeaker locations and/or visually at the left or right talker locations according to one of the 32 spatial configurations shown in Figure 3. The participants were instructed to attend the target phrase containing the call sign "Baron" and to use the mouse

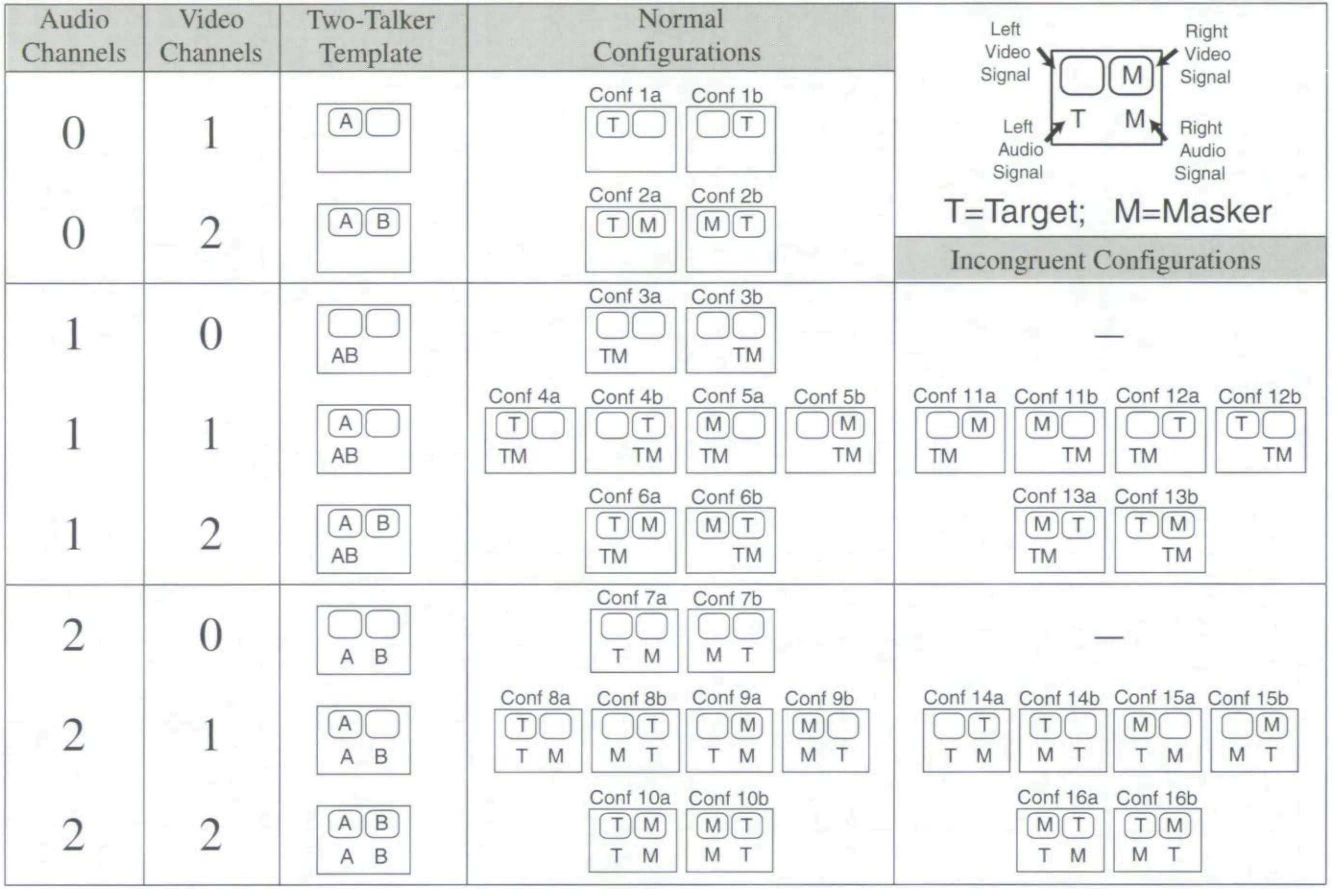

Figure 3. Audiovisual display configurations tested in the experiment. The configurations are illustrated pictorially, with the circled letters at the top of each icon representing the locations of the two possible visible faces and the uncircled letters at the bottom of each icon representing the locations of the two possible audible voices.

to select the color and number contained in the target phrase from an array of colored digits that was displayed on the projection screen after the end of each stimulus.

The data were collected in 60-trial blocks. Each block was divided into 12 sequences of five consecutive trials with the same target talker location in one of the 32 different spatial configurations shown in Figure 3. Thus, within each block of trials, the locations of the target and masker voices and faces would remain fixed for five consecutive trials, randomly change to a new spatial configuration, remain fixed for five more consecutive trials, randomly change again, and so on until five trials had been collected in each of 12 different randomly selected spatial configurations. Over the course of the experiment, the data collection was balanced to collect an equal number of trials with each of the two target talkers (male or female) with each of the 32 spatial configurations shown in Figure 3 (i.e., an equal number of trials in the "a" and "b" versions of each of the 16 configuration conditions in the figure). Each participant participated in 34 60-trial blocks, plus one additional 40-trial block that was required to equalize the number of trials collected in each of the 32 spatial configurations. Thus each participant participated in a total of 13 five-trial sequences in each of the 32 possible AV configuration conditions of the experiment, for a total of 2080 trials per participant.

RESULTS

AV Display Condition

Figure 4 shows overall color and number identification performance in each of the eight spatially congruent AV display conditions tested in the experiment. These data were generated by averaging performance across all of the normal spatial configurations associated with each basic AV template in Figure 3. For example, performance in the AV condition with one audio channel and one video channel (speckled bar in the middle of Figure 4) represents mean performance across Spatial Configurations 4a, 4b, 5a, and 5b. This mean value represents the average overall level of performance that would occur if that AV display condition were used in a two-talker listening scenario in which the relevant information was equally likely to originate from either of those two talkers. In the AV conditions in which only one of the two talkers was visible (speckled bars), the black triangles show performance in the spatial configurations in which only the target talker was visible and the white triangle shows performance in the spatial configurations in which only the masking talker was visible.

The leftmost group of bars shows performance in the visual-only AV conditions, in which no audio signal was present in the stimulus. The speckled bar represents performance in the condition with only one visible talker. The black triangle illustrates performance in the spatial configurations in which the only visible talker happened to be the target talker. The participants' responses were 67% correct in this condition, indicating that they were quite good at determining both the colors and numbers contained in the CRM phrases from just the visual cues available in the stimuli.

The relatively high speechreading scores obtained in this experiment reflect the highly redundant nature of the CRM speech stimuli: With only four color and eight number alternatives, the participants were able to make reasonably accurate guesses about the contents of the CRM phrases, even without the benefit of any audio information. Of course, because only one talker was present in this AV condition, no information could be obtained from the display when the visible talker happened to be the masking talker rather than the target talker. Thus in Figure 4 the white triangle in this condition is placed at the chance level of performance (shown by the horizontal dashed line). The level of the speckled bar represents the average of the target-visible and masker-visible spatial configurations; in other words, it represents the mean performance level that would occur in a speech display that showed the face of only a single talker, who was 50% likely to be the target talker at any given time. Note that this mean performance level was substantially worse than the mean performance level that occurred in the no-audio condition in which both talkers were visible at the same time (shown by the adjacent shaded bar in the figure). This suggests that the participants in the AV condition with two visible faces were able to successfully divide their attention

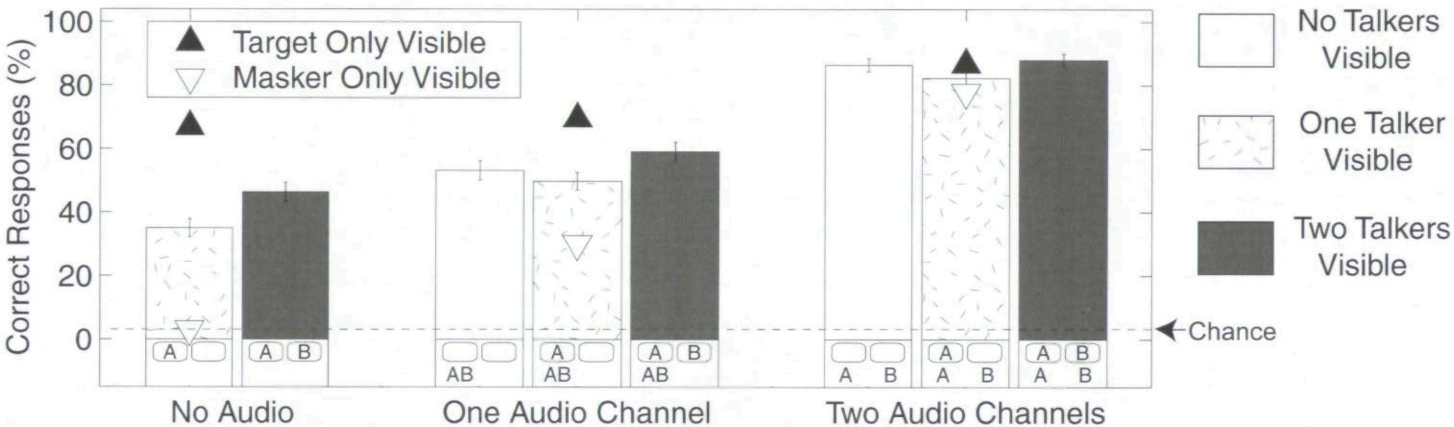

Figure 4. Overall percentage of the responses that matched both the color and the number contained in the target phrase in each spatially congruent AV condition of the experiment. In the AV conditions with only one talker visible, the center points of the black triangles show performance in the spatial configurations in which the target was visible, and the center points of the white triangles show performance in the spatial configurations in which the masker was visible. The dashed line shows chance performance (1/32 correct). The error bars show 95% confidence intervals calculated from the raw data for each data point.

across the two faces and use the embedded call signs in the CRM phrases to help them selectively focus their attention on the face of the target talker.

The right two groups of bars in Figure 4 show performance in the AV display conditions that contained at least one channel of audio. The arcsine-transformed data from each participant in these six conditions were subjected to a two-factor, within-subjects repeated measures analysis of variance (ANOVA) on the factors of audio channels (one or two) and video channels (zero, one, or two). This ANOVA showed that the main effects of audio channel, $F(1, 7) = 96.19$, and video channel, $F(2, 14) = 25.35$, were both significant at the $p < .0001$ level but that the interaction between the two was not significant. From these data, it is apparent that spatially separating the two audio channels had the largest effect on performance, improving overall identifications by roughly 32 percentage points. Also, from the mean performance values illustrated by the bars, it is apparent that overall performance was roughly 4 percentage points better than in the audio-only condition when both talkers in the stimulus were visible but that it was actually 4 percentage points *worse* than in the audio-only condition when only a single visible talker was present in the stimulus. Although these differences were small in absolute terms, a post hoc test (Tukey's honestly significant difference) confirmed that both were statistically significant at the $p < .001$ level.

The reason for the drop in mean performance in the AV conditions with only a single video channel is illustrated by the white and black triangles in Figure 4. When a visible representation of just the target talker was added to the audio-only stimulus (black triangles), performance increased slightly in the condition with one audio channel and stayed roughly constant in the condition with two audio channels. However, when the visible representation of just the masking talker was added to the stimulus (white triangles), performance decreased in both audio conditions. This performance penalty might have been the result of a McGurk-like fusion between the visible masking face and the audible target voice, or it might simply have been the result of a strong bias on the part of the participants to incorrectly assume that the visible talker was always the target talker. However, it is worth noting that the performance penalty of showing just the masking talker was much smaller in the condition with two audio channels, in which the bias to assume that the visible talker was the target should have been just as strong. Thus it seems that spatially separating the two talkers' voices produced a substantial increase in overall performance and had the additional benefit of making the participants much more resistant to confusion from the presence of a visible representation of only the masking talker.

Spatial Incongruency

Figure 5 compares performance in the congruent and incongruent spatial configurations of the AV display conditions that included one or two audio channels and one or two visual channels. The arcsine-transformed performance scores for each participant were also subjected to a three-factor within-subjects ANOVA on the factors of audio channels (one or two), video channels (one or two), and spatial congruency (yes or no). This analysis revealed that all three of these main effects were statistically significant at the $p < .05$ level but that none of their interactions were significant. Thus it seems that performance was slightly but significantly worse in the spatially incongruent configurations than in the spatially congruent configurations.

In the AV conditions with one audio channel, performance was virtually identical in the spatially congruent and incongruent configurations of the experiment. Thus there does not appear to be any evidence of the improvement in performance reported by Driver (1996) when the visual representation of the target talker was spatially displaced from the location of the audio signal containing both the target and masking speech. In a recent experiment examining bimodal speech perception, Rudmann et al. (2003) also failed to find any evidence of the improvement in performance that Driver (1996) attributed to the ventriloquism effect. (Specifically, Driver, 1996, hypothesized that the spatially displaced visual image changed the apparent location of the corresponding audio signal and thus introduced an apparent spatial separation between the target and masking speech signals.) In light of our results and those of Rudmann et al. (2003), it does not appear that the ventriloquism-based

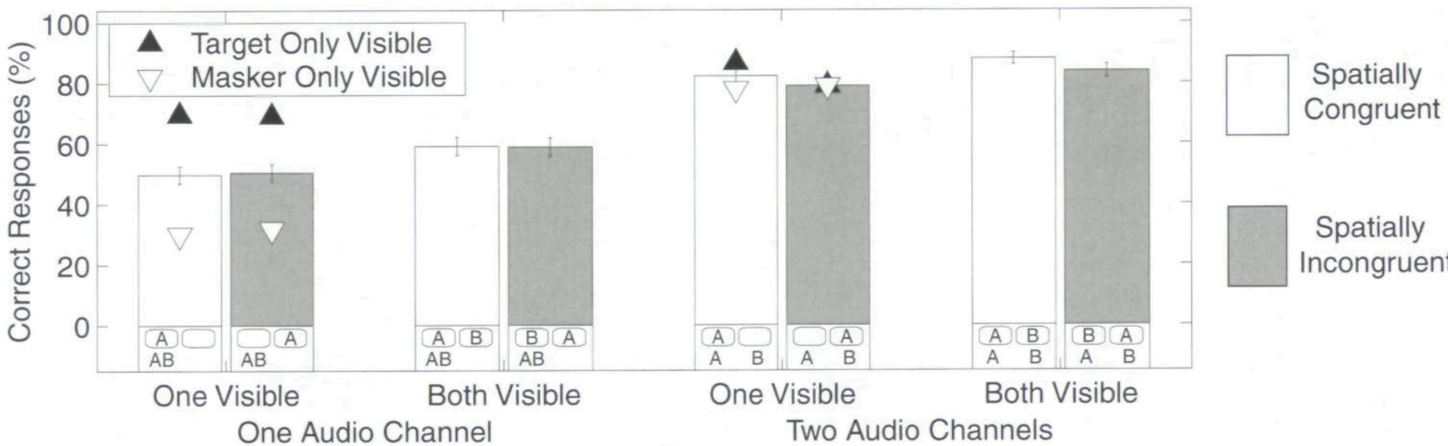

Figure 5. Percentages of correct color and number identifications in the spatially congruent (white) and spatially incongruent (shaded) configurations of the experiment. Details are as in Figure 4.

spatial separation effect reported by Driver (1996) is particularly robust.

In the AV conditions with two audio channels, there does appear to be a slight decrease in performance in the spatially incongruent configurations. This is consistent with previous results that have shown that AV speech perception is better when the audio and visual signals are colocated than when they are spatially separated (Driver & Spence, 1994). However, this also seems to be a relatively small effect.

SUMMARY AND CONCLUSIONS

In terms of display design, the important results of this experiment can be summarized by the following three major points:

1. Spatial separation of the audio signals should be the first priority of any system designed to efficiently present more than one simultaneous speech signal. In this experiment, spatially separating the audio signals from the two talkers improved overall performance by 32 percentage points, whereas adding visual representations of both talkers improved performance only 4 percentage points. Spatially separating the audio signals also substantially reduced the performance penalty that occurred when only the masking talker was visible in the stimulus.

2. Caution should be used when adding only a single visible face to a multichannel speech display when there is no guarantee that the visible talker will be the most important talker in the combined AV stimulus. In this experiment, the performance gains that occurred when the visible talker happened to be the target talker were more than offset by the performance penalties that occurred when the visible talker was the masking talker.

3. Whenever possible, the audio and visual portions of a multitalker speech signal should be spatially congruent. In this experiment, there was no evidence that the ventriloquism effect improved performance when the visible representation of the target talker was spatially separated from a single-channel audio signal containing both the target and masking voices, and there was a significant decrease in performance when the visible target talker was presented with a spatially incongruent two-channel audio signal.

The results of this experiment provide a preliminary assessment of the roles that audio and visual cues play in speech-monitoring tasks, but a great deal of additional research is needed to obtain a clear picture of the full impact that visual cues have on multitalker speech perception. One possible area for future research is a more detailed analysis of the influence that different speech materials have on audiovisual interaction in monitoring tasks. The limited vocabulary of the CRM corpus used in this experiment (two call signs, four colors, and eight numbers) introduced substantial redundancy into the audio and visual signals associated with the different conditions of this experiment. This redundancy may have caused performance to be dominated by the most salient modality. If a larger vocabulary of phonetically balanced words were used to generate the stimuli, some of this redundancy would be eliminated and the participants would probably exhibit more signs of audiovisual integration than they did with the CRM phrases.

Another important area for future research is an extension of this experiment to multitalker listening configurations with more than two competing talkers. In the monitoring task used in this experiment, participants who were able to extract the masker call sign from one of the two competing speech signals were able to use elimination to determine that the other talker had to be the target talker. If a third talker were added to the stimulus, an elimination-based monitoring strategy would be less effective and overall performance in the task would be more dependent on the ability to divide attention across more than one simultaneous speech signal. This would provide a more thorough test of the role that audiovisual cues play in multitalker speech perception and would allow an extension of the results of this experiment to complex communications situations in which listeners are more likely to require the assistance of sophisticated speech display systems to successfully complete their assigned tasks.

ACKNOWLEDGMENTS

The authors would like to thank Bob Bolia and Jessica Young for their help with the collection of the AV corpus. Portions of this research were funded by the Air Force Office of Scientific Research.

REFERENCES

Abouchacra, K., Tran, T., Besing, J., & Koehnke, J. (1997, February). *Performance on a selective attention task as a function of stimulus presentation mode.* Presented at the Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, Florida.

Bertelson, P., & Radeau, M. (1976). Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. *Perception and Psychophysics, 29,* 578–585.

Bronkhorst, A. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica, 86,* 117–128.

Brungart, D., Simpson, B., Ericson, M., & Scott, K. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America, 110,* 2527–2538.

Crispien, K., & Ehrenberg, T. (1995). Evaluation of the "cocktail party effect" for multiple speech stimuli within a spatial audio display. *Journal of the Audio Engineering Society, 43,* 932–940.

Driver, J. (1996). Enhancement of selective attention by illusory mislocation of speech sounds due to lip-reading. *Nature, 384,* 66–68.

Driver, J., & Spence, C. (1994). Spatial synergies between auditory and visual attention. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 311–331). Cambridge, MA: MIT Press.

Ericson, M., & McKinley, R. (1997). The intelligibility of multiple talkers spatially separated in noise. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 701–724). Hillsdale, NJ: Erlbaum.

McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.

Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1999). Spatial audio displays for speech communication: A comparison of free-field and virtual sources. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 1202–1205). Santa Monica, CA: Human Factors and Ergonomics Society.

Reisberg, D. (1978). Looking where you listen: Visual cues and auditory attention. *Acta Psychologica, 42,* 331–341.

Rudmann, D., McCarley, J., & Kramer, A. (2003). Bimodal displays improve speech comprehension in environments with multiple speakers. *Human Factors, 45,* 329–336.

Spence, C., Ranson, J., & Driver, J. (2000). Cross-modal selective attention: On the difficulty of ignoring sounds at the locus of visual attention. *Perception and Psychophysics, 62,* 410–424.

Sumby, W., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26,* 212–215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audiovisual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lipreading* (pp. 3–51). New York: Erlbaum.

Douglas S. Brungart is a senior computer engineer in the Air Force Research Laboratory at Wright-Patterson Air Force Base, Ohio. He received a Ph.D. in electrical engineering from the Massachusetts Institute of Technology in 1998.

Alexander J. Kordik is a human factors psychologist with Sytronics, Inc., in Dayton, Ohio. He received a B.S. in psychology from Wright State University in 2002.

Brian D. Simpson is an engineering research psychologist in the Air Force Research Laboratory at Wright-Patterson Air Force Base, Ohio. He received a M.S. in human factors psychology from Wright State University in 2002.

Date received: March 7, 2003
Date accepted: August 26, 2004

Sharing Control Between Humans and Automation Using Haptic Interface: Primary and Secondary Task Performance Benefits

Paul G. Griffiths and **R. Brent Gillespie,** University of Michigan, Ann Arbor, Michigan

This paper describes a paradigm for human/automation control sharing in which the automation acts through a motor coupled to a machine's manual control interface. The manual interface becomes a haptic display, continually informing the human about automation actions. While monitoring by feel, users may choose either to conform to the automation or override it and express their own control intentions. This paper's objective is to demonstrate that adding automation through haptic display can be used not only to improve performance on a primary task but also to reduce perceptual demands or free attention for a secondary task. Results are presented from three experiments in which 11 participants completed a lane-following task using a motorized steering wheel on a fixed-base driving simulator. The automation behaved like a copilot, assisting with lane following by applying torques to the steering wheel. Results indicate that haptic assist improves lane following by least 30%, $p < .0001$, while reducing visual demand by 29%, $p < .0001$, or improving reaction time in a secondary tone localization task by 18 ms, $p = .0009$. Potential applications of this research include the design of automation interfaces based on haptics that support human/automation control sharing better than traditional push-button automation interfaces.

INTRODUCTION

Although certain costs of adding automation to a machine or process are easy enough to anticipate (e.g., additional operator training), other costs have turned out to be much more subtle. Especially when operators are called upon to share control with an automation system, unexpected problems arise. These include bumpy transitions between automated and manual control (Meyer, Rubinstein, & Evans, 2001; Mosier, 2002) and automation surprises, which occur when the operator has a poor mental image of automation behavior or misses automation mode changes (Sarter, Woods, & Billings, 1997). The consequences arising from these unexpected problems can be dire, at times offsetting the improved precision and performance or reduced operator workload for which automation was added in the first place. Unraveling these costs and providing solutions has been the focus of much research in recent years.

A portion of the unexpected problems associated with sharing control between a human operator and an automation system can be attributed to the need for the operator to master two control interfaces and to learn two or more distinct operating modes. Whereas many machines present a manual control interface requiring continuous, direct control and certain manual skills, the typical automation interface presents a set of indicators, knobs, and buttons requiring intermittent input and certain analytical and decision-making skills (Mosier, 2002). When complex or unanticipated conditions arise, the traditional approach to "cooperation" is for the operator to interrupt the automation and take over full control through the manual interface (Christoffersen & Woods, 2002). When control is wrested away from the automation system, the advantages of

Address correspondence to Paul G. Griffiths, Department of Mechanical Engineering, University of Michigan, 2350 Hayward St., Ann Arbor, MI 48109; paulgrif@umich.edu. **HUMAN FACTORS,** Vol. 47, No. 3, Fall 2005, pp. 574–590.

automation (precision, computational speed, and other functions) are lost (Christoffersen & Woods, 2002). Bainbridge (1983) has noted that ironically, the manual skills of an operator who regularly gives up control to an automation system may begin to degrade from lack of practice, leaving him or her ill prepared to meet the heightened challenges that typically arise during automation failures.

What is needed is an intermediate, more collaborative mode of interaction. Ideally, the operator would be able to negotiate with and redirect the automation system without first disabling it and then restarting (Christoffersen & Woods, 2002). There would be a natural kind of "give and take" between the operator and the automation. In this paper, we propose an automation interface and associated control-sharing paradigm that does not include mode switching and therefore avoids the associated pitfalls. The operator needs to learn only one interface and one set of rules.

Another portion of the unexpected problems in automation can be attributed to the poor rate and quality of information transmission supported by most automation interfaces. In addition to the communication of current mode and action, efficient cooperation requires the communication of goal and intent. One approach to improving communication is the use of multimodal displays. Sarter (2002) has estimated the value of various modalities for guiding the operator's attention and effectively managing interruptions. To visual and auditory displays, Sarter (2002) has added haptic (kinesthetic and tactile) display, including vibrotactile cues on the manual control interface.

In automobile interfaces, both vibrotactile and pulse torque cues applied through a motorized steering wheel have been tested as a means of informing or warning automobile drivers (Driver & Spence, 2004; Schumann, Godthelp, & Hoekstra, 1992; Suzuki & Jansson, 2003). Beyond the use of haptics for discrete cues, haptic display has been used to automate vehicle steering in such a way that the driver can continuously monitor the automation actions (Switkes, Rossetter, Coe, & Gerdes, 2004). In this work, automation took the form of a virtual potential field to aid the driver in lane keeping. Any deviation from the center of a lane produced continuous haptic information about the automation's tendency to return the vehicle to the lane center.

In this paper, we also employ haptic display with the aim of enabling more effective negotiation and coordination of intent and actions. The promise of haptic display follows in part from its lack of overlap with the visual and auditory modalities, but further, haptic signals can be information rich in ways that are particular to the communication of intent, especially intent involving direction and magnitude. Further, we believe that signals encoded in force and motion are especially suited for informing the operator about automation intent because they can simultaneously convey intent and automation confidence in that intent. Mechanically mediated communication can even support negotiation of authority. For example, a particular motion imposed by the automation can be accompanied by larger reaction forces that resist motion inputs from the operator. Likewise, the operator could express his or her desire for increased authority by using high impedance or less "give," possibly by cocontracting the muscles in his or her arm. This is the way two human operators would communicate if they both grasped the same manual control interface: Each operator would apply muscle action to extract his or her own desired response with a certain authority while simultaneously perceiving the other's intent and desire for authority by feel.

We propose, then, to combine the machine and automation interfaces into a single interface modeled after the traditional manual control interface. The human operator is presented only the standard manual interface, over which the automatic control system is also given authority. The automation system imposes its control effort through a motor coupled directly to the interface. The motorized manual interface becomes a haptic display, relaying information about the actions of the automation system to the human operator's haptic senses. In effect, we propose a return to the "contact" or "direct" mode of interaction, in which visual/kinesthetic perception and motor response are relied upon rather than the analytical and decision-making skills usually required by the automation system. In contrast to the use of vibrotactile cues for alert and display of discrete information, we use haptics to relay continuous signals to the operator. In our shared control scheme, the automation is placed in mechanical parallel with the operator and takes the form of an assist that actually intervenes in the control loop. The operator may choose either to yield to the assist while observing its action or to override it by exerting slightly more effort.

Through the motor on the interface, the automation may apply torques according to its own rules or control law, as a function of sensed machine state. For example, a steering wheel can be given a "home" position that is itself animated according to sensed vehicle position within a lane. The automatic controller can create virtual springs that attach the steering wheel to a moving home angular position that corresponds to the vehicle direction recommended by the automation. By feel, the operator can form a mental image of the springs attached to the moving home position, especially by haptically exploring the invariants of the reaction torque to his or her own input motions. To ensure that the automation can be overridden, it uses a limited mechanical impedance (essentially a limited stiffness).

The introduction of assist through a motor on a manual control interface has been studied extensively in the applications of haptic interface to teleoperated and virtual environments (Gillespie, 2004; Hayward, Astley, Cruz-Hernandez, Grant, & Robles-De-La-Torre, 2004). In these applications, assist is offered in the form of *virtual fixtures* that may be used by the operator as mechanical guides for controlling force or motion direction. Virtual fixtures have been shown to improve performance in targeting tasks (Dennerlein & Yang, 2001; Hasser, Goldenberg, Martin, & Rosenberg, 1998), peg-in-hole tasks (Payandeh & Stanisic, 2002; Rosenberg, 1993; Sayers & Paul, 1994), and surgical interventions (Park, Howe, & Torchiana, 2001). Virtual fixtures are usually fixed in the shared workspace; however, virtual fixtures composed of functions of time or recognized operator motions were studied by Li and Okamura (2003). In this work we also employ virtual fixtures, created by the automation in the workspace shared by the automation and operator. Our fixtures, however, are animated by the automation system. By and large, the focus in the field of haptic interface has been improved human/machine performance. The possible secondary benefits, such as reduced operator workload, have been overlooked in the literature.

The setting for our investigation of mechanically mediated control sharing is a driving simulator with a motorized steering wheel. Although our particular implementation is far removed from actual driving – our former setups were in fact closer (Steele & Gillespie, 2001) – we hope that our results may nevertheless contribute to ongoing work in the design of automation interfaces for driving. The primary goal in this paper is to quantify the primary and secondary benefits of using a motorized interface to institute control sharing between an operator and an automatic controller.

We present three experiments designed to demonstrate our conception of shared control using a motorized manual control interface. Naturally, an expected outcome is improved performance on the semiautomated task, which is quantified in Experiment 1. However, there are other expected benefits. Experiment 2 was aimed at quantifying the benefits in reduced visual demand associated with the primary task. In Experiment 3, we investigated the hypothesized freeing of attention (as reflected by improved performance on a secondary task). The driving task in all experiments also includes a challenge not addressed by the automatic controller, which becomes a means for prompting negotiations between the human and automation and a basis for requiring and measuring maintained attention to the primary driving task. After the descriptions of the experiments, we present a general discussion comparing the results from all three experiments and discuss the effects of haptic assist on primary and secondary tasks, followed by a summary of our results.

METHODS

To carry out our experiments, we developed a fixed-base driving simulator that featured a computer monitor and a motorized steering wheel. The computer monitor presented participants with a view of the simulated vehicle hood and roadway, and the motorized steering wheel provided steering control and force feedback from the simulated tire-road interactions. In addition, the motor on the steering wheel was used to apply torques from the automation system when it was enabled. These torques were designed to assist the driver in holding to the center of the simulated roadway (i.e., to assist in lane keeping). The automation-produced torques could be felt by the participant's hand on the steering wheel, so we refer to them as *haptic assist.* To prompt the participants to maintain some amount of authority while being assisted by the automation, obstacles were placed in the center of the roadway. The automation system was not able to sense and avoid these obstacles. The participants were instructed that they would be solely responsible for steering around the obstacles. In this section we introduce each of the major components of the driving simulator in turn: the vehicle, roadway, and obstacle models; the automation system (including its path-planning algorithm and feedback control law), and the simulator hardware. The experimental procedures pertaining to each of the three experiments and the associated performance metrics will be described in the sections to follow.

Vehicle, Roadway, and Obstacle Models

Participants drove a vehicle model, which rolled on flat ground, and they turned right or left by steering the front wheels, as in a typical car. The speed of the vehicle's front wheel was fixed at 15 miles/hr (mph; 24 km/hr); thus interactions with brake and accelerator pedals were not included. The kinematics of the vehicle model were computed according to the bicycle model, assuming no slip between the tires and the roadway (T. D. Gillespie, 1992). Force feedback on the motorized steering wheel reflected the vehicle model's self-aligning torque. This torque acts on the steering linkage and tends to turn the front wheels into the direction of travel of the vehicle, causing the vehicle to steer straight. Given the fixed speed of the vehicle, the self-aligning torque was proportional to the steering angle of the front wheels relative to the vehicle centerline.

A roadway was defined as a sequence of 16 straight and 15 left- and right-curved road segments of varying length totaling 1993 m. An overhead view of the roadway is shown in Figure 1. All the curved segments had a curvature of 0.025 m^{-1}. To smooth transitions, segments were joined with clothoid curves 5.6 m long. Given the fixed front wheel speed, each trial lasted nearly 5 min; some variation in the time per trial arose because of the difference in length of the actual vehicle path compared with the length of the road centerline. Visually, the road segments had a gray, concrete texture with a yellow stripe along the center and green embankments on either side.

Participants were instructed to follow the yellow centerline of the road, except when they encountered orange cylindrical objects that had been placed on the road centerline at irregular intervals. All obstacles were 1.4 m in diameter and 2.0 m tall, and the spacing between obstacles varied from 20 to 80 m in a uniform random distribution. Obstacles were located on both straight and curved segments, as indicated in Figure 1. If the vehicle perimeter contacted an obstacle, a brief orange flash indicated the collision and the destruction of the obstacle.

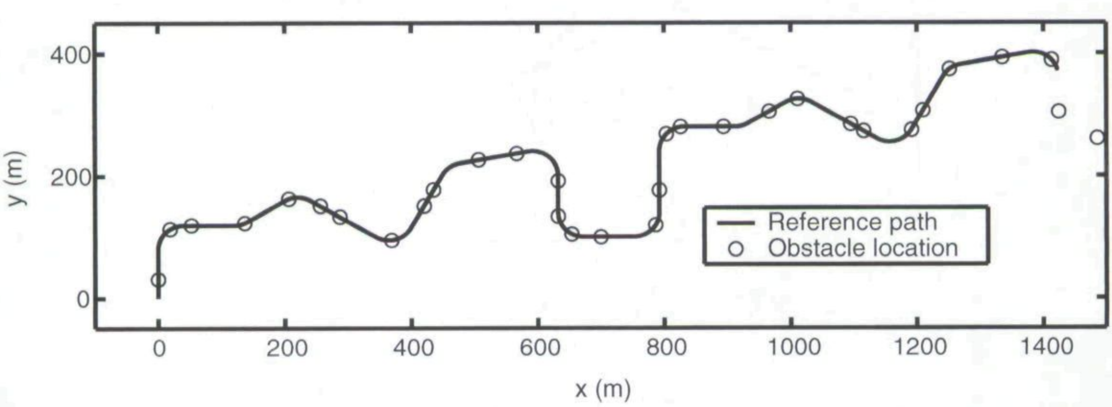

Figure 1. A top-down view of the driving course that participants traversed during each trial. The location of each obstacle along the road centerline is indicated with a circle.

Automation System

In half the trials, an automation system assisted participants in lane keeping. Conceptually, the automation divided the steering task into two problems: generating a desired path that would return a stray vehicle to the road centerline and turning the steering wheel to follow that desired path.

The path planning employs a geometric approach based on knowledge of the vehicle's position and orientation relative to the nearby road geometry. This approach follows the predictive driver model of Hess and Modjtahedzadeh (1990), using the notion of an "aim point" ahead of the vehicle on the centerline of the road. This aim point is located by finding the closest point to the vehicle on the road and then looking forward 10 m along the road. The geometry of the aim-point construction is illustrated in Figure 2. If the vehicle's front wheels are always headed toward the aim point, they follow a path that leads back to the center of the road. Given the desired front wheel heading and the current vehicle heading, the desired steering wheel angle is determined. Because our path planning was based on a model of human driver behavior, we surmised that the automation would, in some sense, not fight the human driver but rather mimic the driver's behavior. This path-planning technique was also advantageous because the desired steering angle was relatively simple to calculate from the geometry; the only challenge was calculating the closest point in real time. We addressed that problem by using a feedback-stabilized closest-point algorithm, which features computational efficiency for real-time applications (Patoglu & Gillepsie, in press).

After calculating the desired steering angle, the automation's remaining task was to exert an appropriate torque on the steering wheel. A virtual torsional spring was used to oppose motion of the steering wheel away from the desired steering angle – that is, a restoring torque proportional to the difference between the desired steering wheel angle and the current measured angle was applied to the motorized steering wheel. A torsional stiffness of 1.2 Nm/rad set the level of control authority exerted by the automation. As a further limit on the automation's authority, the maximum magnitude of the torque

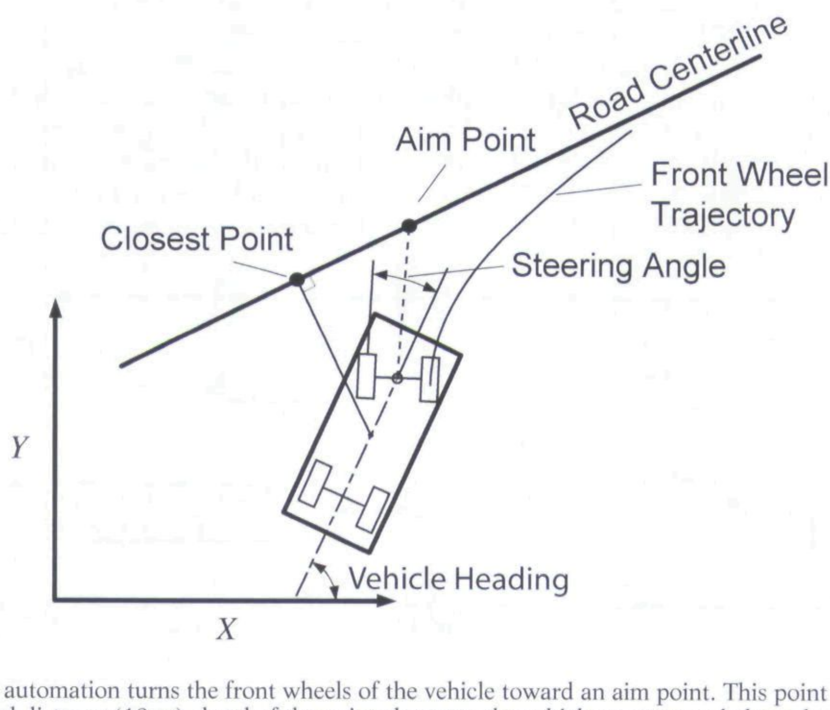

Figure 2. The automation turns the front wheels of the vehicle toward an aim point. This point remains a fixed distance (10 m) ahead of the point closest to the vehicle as measured along the road centerline.

was set to 0.82 Nm. Thus if the participant turned the wheel far away from the desired steering angle, the resistance imposed by the steering wheel motor would saturate well below the limits of human strength.

Simulator Hardware

Participants drove the simulator while seated in front of a computer monitor that displayed the roadway, with their right hand grasping the motorized steering wheel as depicted in Figure 3. The motorized wheel is described in Gillespie, Hoffman, and Freudenberg (2003). The computational hardware supporting the driving simulator included two computers: a PC for graphical display and data logging and a Motorola MPC555 microcontroller for real-time simulation of the vehicle model and the automation system. An OpenGL graphics application running on the PC rendered a 3-D animation of the hood of the car and the road. A screen shot from the animation appears in Figure 4, showing embankments on either side of the road and two obstacles on the centerline. The graphics software received the vehicle state information every 8 ms through a serial communication link to the MPC555 microcontroller. For Experiment 2, the graphics program was used to occlude the participant's view of the road, except for 1-s glimpses when requested through a key press.

The driving simulator and shared controller described in this section were used in three experiments. Experiment 1 defined a baseline measure of driving performance for participants with and without the haptic assist. In Experiment 2, the participants' visual demand while driving was measured using the visual occlusion method with and without haptic assist. In Experiment 3, participants were asked to perform a secondary task while driving. Again, the experiment was run for the two conditions (with and without haptic assist), and the effect of haptic assist on secondary task performance was measured by accuracy and reaction time.

EXPERIMENT 1: BASELINE

The objective of our first experiment was to quantify the improvement in driving performance afforded by the haptic assist controller. Driving performance was defined in terms of two performance variables: lateral error in tracking the road centerline (a path-following task) and the number of obstacles hit. The haptic assist, when present, could be entrusted to help only with path following. It was solely the driver's responsibility to avoid obstacles in the middle

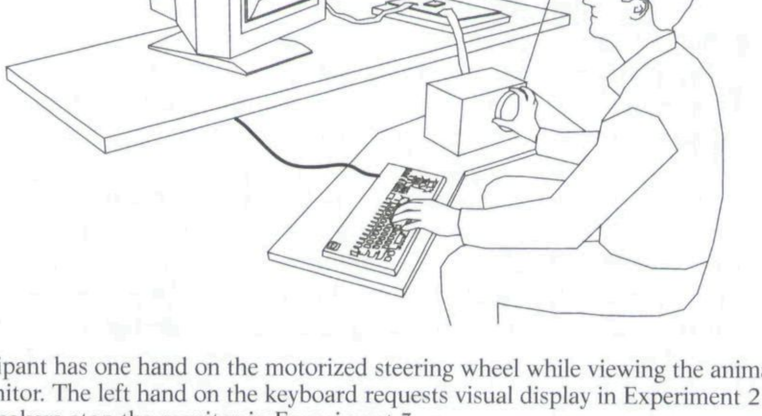

Figure 3. A participant has one hand on the motorized steering wheel while viewing the animated roadway on the computer monitor. The left hand on the keyboard requests visual display in Experiment 2 and responds to tones from the speakers atop the monitor in Experiment 3.

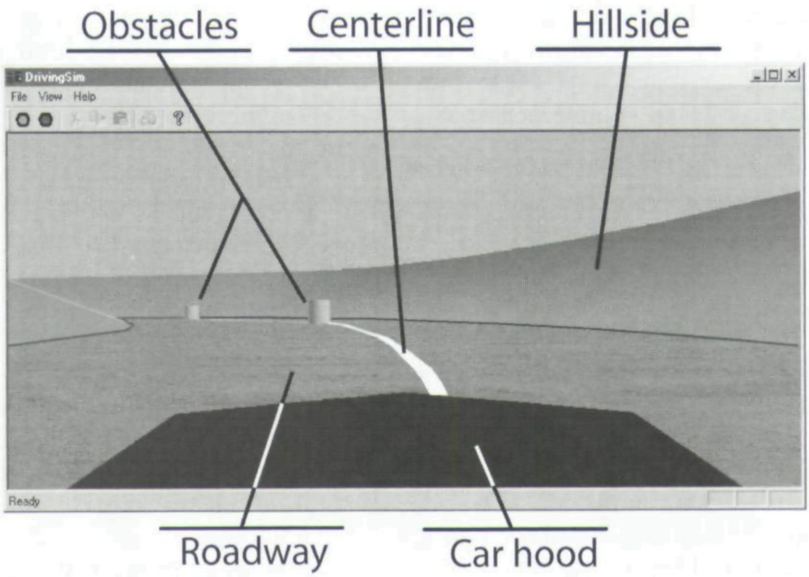

Figure 4. An OpenGL animation of the roadway visible over the hood provided participants with visual feedback (labels added).

of the road; the automation had no information about the obstacles. Thus the obstacles provided the motivation for keeping the human in the loop. The experimental condition under investigation in Experiment 1 was simply the presence of haptic assist. Experiment 1 provided the baseline performance, against which the results from Experiments 2 and 3 were compared. Experiments 2 and 3 imposed additional experimental conditions under which the effects of haptic assist were measured.

Procedure

Participants. Eleven participants (9 men and 2 women) between the ages of 20 and 63 years (M = 30, SD = 11.9 years) volunteered for the study. All participants had normal or corrected-to-normal vision. Each participant provided informed consent in accordance with University of Michigan human participant protection policies. Individuals were not paid for their participation. Each participated in one experimental session lasting approximately 1 hr. Participants, whether left or right handed, were asked to use their right hand on the motorized steering wheel to steer the simulated vehicle along the centerline of the roadway as closely as possible without colliding with any obstacles. Each participant spent one 5-min trial familiarizing him- or herself with the path-following task under the experimental conditions of all three experiments.

Design. After completing the practice trial, each participant completed six 5-min trials, each representing a unique set of experimental conditions. Two of these six trials were designed to assess baseline path-following performance: one trial without haptic assist and the other with haptic assist. Each participant was randomly assigned a sequence of trials chosen from a set designed to counterbalance the ordering of the haptic assist condition. Participants did not receive performance feedback at the end of trials; however, the roadway and excursions of the car from the centerline were visible on the graphic display, and obstacle collisions, if they occurred, were accompanied by a brief flash of orange on the entire screen. The simulator logged the time, vehicle position and orientation, obstacle collisions, the closest point on the centerline (computed in real time), and the torque displayed on the motorized steering wheel.

Dependent variables and data analysis. To assess driving performance, we defined two performance measures: one for path following and the other for obstacle avoidance. The shortest distance between the center of the car and the center of the road was computed at each sample time (every 8 ms) and defined as the lateral error (LE). The root mean square (RMS) of LE, denoted RMS[LE], was used to assess path-following performance. However, to facilitate analysis of path-following performance and its dependence on assist condition independent of the obstacle avoidance maneuvers, we partitioned the data into segments *between* obstacles and segments *near* obstacles. Partitioning was defined in time, with *near* indicating 2 s before and 1 s after the instant at which the closest point on the centerline passed the obstacle. The near data segments were then discarded from the measure of path-following performance. Obstacle avoidance performance was defined as the fraction of obstacles hit from the 30 presented, reported as a percentage and denoted %Hit.

Quantile plots were used to verify that the data fit normal distributions for both performance metrics when considered across the 11 participants. Using $\alpha = .05$ to establish statistical significance of results, multiple-factor analysis of variance (ANOVA) was performed for the two dependent performance measures: RMS[LE] and %Hit.

Results

Figure 5 shows the tracking performance of a typical participant in a generic section of the roadway with and without steering assist. The section of roadway shown is 220 m in length and took about 32 s to traverse at the constant 15 mph (24 km/hr) vehicle speed – a moderate pace given the curvature of the road. The top trace shows the curvature of the road, indicating that a left turn and two straight segments are represented in this section of roadway. Deviation from the centerline is graphed versus time in the lower two traces, in which Trace A was recorded without haptic assist and Trace B was recorded with haptic assist. The obstacles are outlined as circles of radius 1.6 m, but they appear as ellipses because of the nonunity aspect ratio. The 1.6-m radius accounts for the 0.7-m obstacle radius and 0.9-m car half width – that is, a collision occurs if the vehicle center comes within 1.6 m of an obstacle center. Obstacle avoidance maneuvers produced by the participant are apparent in both traces, and those maneuvers are not appreciably different by condition. However, differences across condition are apparent in tracking performance in the sections of roadway between obstacles. Improvement can be observed in Trace B, where haptic steering assist was provided.

As previously described in the subsection discussing dependent variables and data analysis, the data were partitioned into segments either near or between obstacles, and only values of LE sampled when the vehicle was considered between obstacles were used to calculate RMS [LE]. The shaded areas in Figure 5 indicate data within the 3-s windows around each obstacle, which were considered near.

There was some variation in driving behavior among participants, particularly in the obstacle avoidance maneuvers. Some drivers chose more "aggressive" driving styles; they would wait longer before avoiding an obstacle and would turn the steering wheel faster and with more effort. The variation across all participants is shown as pointwise percentiles in Figure 6 over

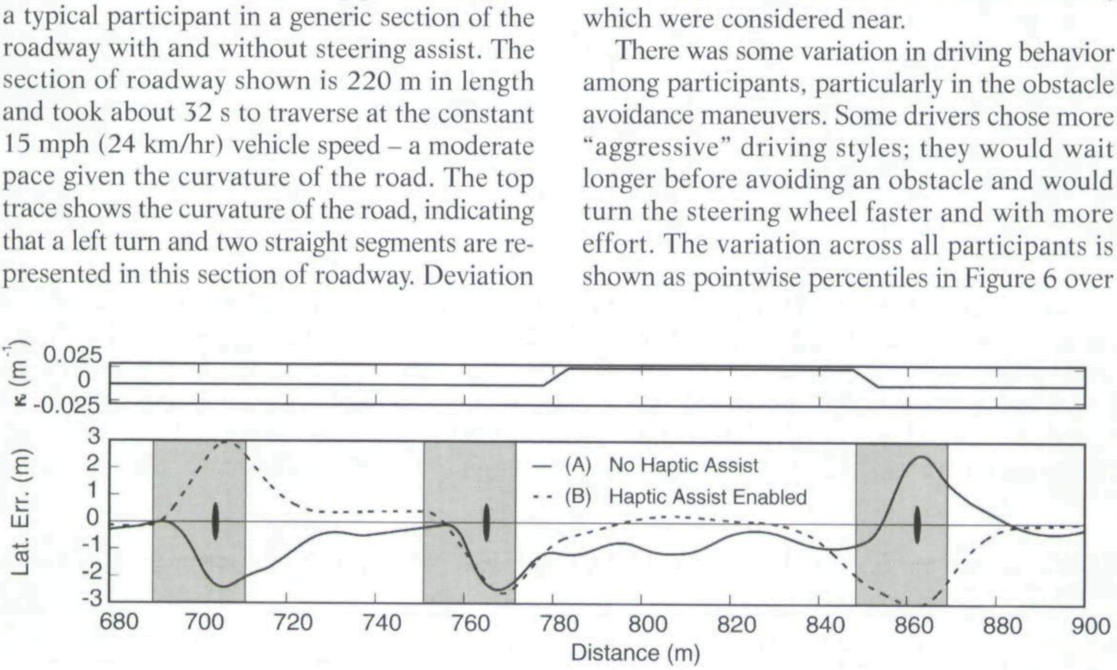

Figure 5. Plot of one participant's lateral deviation over a 220-m section of roadway under conditions without steering assist (A) and with steering assist (B). The top trace indicates roadway curvature, and the ellipses in the lower plot show the size and location of obstacles. The shaded regions indicate segments omitted from the path-following performance metric, RMS[LE].

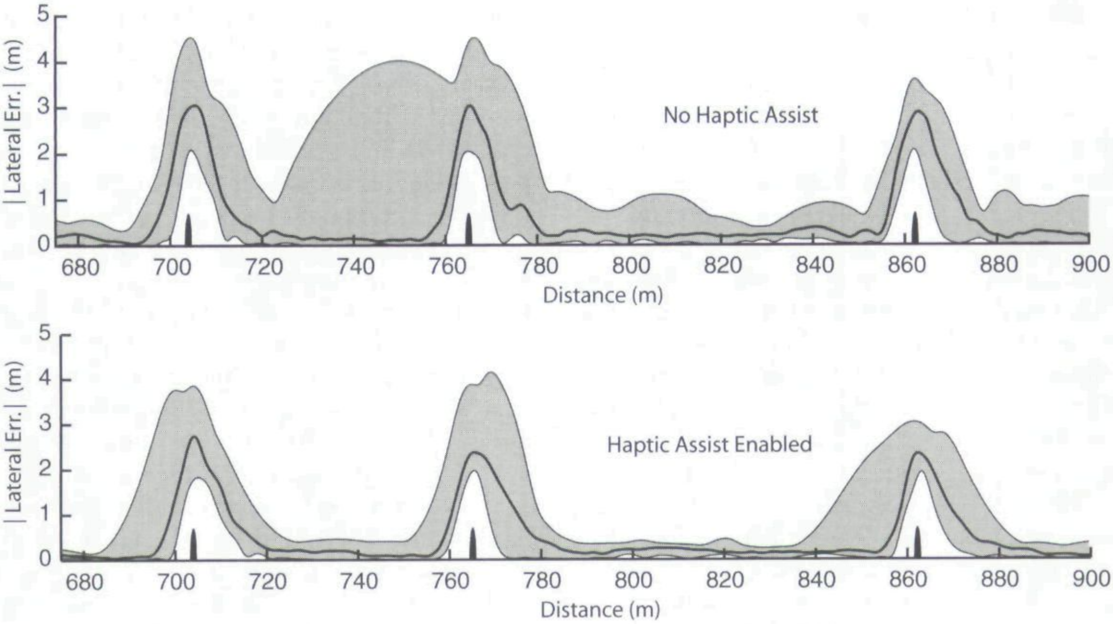

Figure 6. Two plots shade the 5th to 95th percentile intervals for the absolute value of the lateral error data without haptic assist and with haptic assist computed across all 11 participants. The line running through the shaded area indicates the 50th percentile. The portion of the course shown includes three obstacles shown as half ellipses.

the same section of road shown in Figure 5. After computing pointwise percentiles, we low-pass filtered the data with a noncausal second-order Butterworth filter with a spatial cutoff frequency of 3.7 m^{-1}. The 5th and 95th percentile of |LE| form the bottom and top edges of the shaded region in the plot. The dark line drawn through the shaded region is the pointwise median of |LE| for all participants. Qualitatively, the plot shows that the obstacle avoidance behavior is not significantly altered by the haptic assist, but the tracking error is reduced in the sections between obstacles.

Multiple-factor ANOVA applied to RMS[LE] revealed significant main effects attributable to the assist condition, $F(1, 21) = 4.9$, $p = .05$, $MSE = 0.15$, and subject, $F(10, 21) = 3.5$, $p = .03$, $MSE = 0.11$. For %Hit, again a significant main effect attributable to assist condition was found, $F(1, 21) = 9.4$, $p = .01$, $MSE = 0.0032$; however, no significant main effect was found for subject, $F(10, 21) = 2.4$, $p = .09$, $MSE = 0.0083$.

Thereafter, a paired *t*-test analysis was applied to the data. Table 1 presents the sample mean, $\bar{x}$, and standard deviation, *s*, of RMS[LE] and %Hit along with the paired *t*-test results. For both performance measures, the difference between the no-assist and the with-assist conditions was calculated per participant, and the difference of means, $\Delta\bar{x}$, and the standard deviation, $s\Delta$, of these differences along with the

TABLE 1: Mean and Standard Deviation of RMS[LE] and %Hit for Baseline Trials (Experiment 1) With and Without Assist

Performance Metric	No Assist		With Assist		Paired *t* Test			
	$\bar{x}$	*s*	$\bar{x}$	*s*	$\Delta\bar{x}$	$s\Delta$	*p*	DOF
RMS[LE] (m)	0.680	0.391	0.473	0.234	–0.207	0.264	.0134	10
%Hit	0.61	1.26	3.03	3.15	+2.42	2.62	.00594	10

p value and degrees of freedom (DOF) of the t tests are presented in Table 1. There was a 30% reduction in RMS[LE], $t(10)$, $p = .013$. However, for %Hit, there was a statistically significant increase from 0.57% to 2.8%, $t(10)$, $p = .0059$, when haptic assist was added. This result represents a cost of adding haptic assist rather than a benefit. Note, however, that the assist is always trying to drive the vehicle back to the center of the road, exactly where the obstacles are placed. Without driver intervention, the car would drive through every obstacle on the course.

EXPERIMENT 2: VISUAL DEMAND

Our second experiment aimed to quantify the ability of the haptic assist controller to reduce the demand for visual cues while aiding participants in the path-following task. Again, because of the presence of obstacles on the road centerline, the path-following task demanded a certain amount of attention from the participants, whether or not the haptic assist was present. As in Experiment 1, the independent variable is the presence of haptic assist. In Experiment 2, however, the effects of haptic assist were measured with reduced visual feedback. The graphical display was blank except when a visual refresh was requested by the participant. The requests for visual feedback provided a measure of the participants' instantaneous and average demand for visual cues.

Procedure

Participants. The same 11 individuals who participated in Experiment 1 also participated in Experiment 2.

Design. Experiment 2 concerned two of the six trials. These two trials, one with and one without haptic assist, were designed to assess driving performance and visual demand while the visual feedback was metered according to the visual occlusion method (Tsimhoni & Green, 2004). To enable us to measure the participants' demand for visual cues, the graphical display of the driving environment and roadway was blank except for 1-s glimpses provided each time participants pressed a key with their left hand on the computer keyboard. Participants were instructed to request the display whenever they felt that additional visual feedback was necessary to follow the roadway and avoid obstacles. A measure of visual demand throughout the trial was measured by the fraction of time that the visual feedback was not occluded. In addition to the data logged in Experiment 1, the simulator logged key presses on the keyboard.

Dependent variables and data analysis. To assess driving performance, the same two performance metrics already defined for Experiment 1 were used: RMS[LE] (defined in regions away from obstacles) and %Hit. In addition, an instantaneous measure of visual demand was computed using

$$VisD = \frac{1.0}{t_i - t_{i-1}},$$

in which t_i is the time of the ith key press and the numerator is the period of the time that the display was not occluded per request (Tsimhoni & Green, 2004). A measure of average visual demand over the entire trial, denoted Avg[VisD], was computed by the number of key presses in a given trial divided by the duration of the trial (~300 s).

Data analysis included quantile plots, multiple-factor ANOVA, and t tests. The value $\alpha = .05$ was used to establish statistical significance.

Results

The three performance metrics for Experiment 2 (visual demand) were RMS[LE], %Hit, and Avg[VisD]. Multiple-factor ANOVA was performed for all three dependent performance measures. Analysis of RMS[LE] revealed significant main effects attributable to assist condition, $F(1, 21) = 12$, $p = .005$, $MSE = 0.65$, and subject, $F(10, 21) = 4.5$, $p = 0.01$, $MSE = 0.23$. For %Hit, neither assist condition nor subject were significant main effects, although assist condition approached significance, $F(1, 21) = 3.5$, $p = .09$, $MSE = 0.011$. The new performance measure in this experiment was Avg[VisD], and the ANOVA results indicated significant main effects for assist condition, $F(1, 21) = 96$, $p = .0001$, $MSE = 1370$, and for subject, $F(10, 21) = 8.6$, $p = .001$, $MSE = 1220$.

As in Experiment 1, the presence of haptic assist produced a reduction in RMS[LE] in regions away from obstacles. The data presented in Figures 5 and 6 from Experiment 1 are also representative of the effects of the assist condition on path following for Experiment 2. Similarly, there was an increase in %Hit; however, in Experiment 2, the increase was not a statistically significant result. Visual demand was significantly reduced with the addition of haptic assist.

When Avg[VisD] is plotted against path length, there is no obvious, qualitative correlation with the proximity to obstacles or curves. In fact, visual demand remains relatively constant over the entire course. We conjectured that participants might try to schedule their request for a visual refresh during critical periods (e.g., a short time before an obstacle), in which case the key press frequency would be a poor measure of visual demand. However, a histogram of the number of key presses in relation to time before and after passing an obstacle revealed a flat distribution.

To further investigate the dependence of RMS [LE], %Hit, and Avg[VisD] on assist condition, we applied paired t tests. Table 2 presents the sample mean, $\bar{x}$, and standard deviation, s, of the RMS[LE], %Hit, and Avg[VisD], along with the mean differences, $\Delta\bar{x}$, by condition and associated standard deviation, $s\Delta$, and p values resulting from the paired t tests. There was a significant reduction in RMS[LE] of 41%, $t(10)$, $p = .002$, and a significant reduction in Avg[VisD] of 29%, $t(10)$, $p = .0001$, when the haptic assist was available as compared with the no-assist condition. With assist added, %Hit increased from 1.8% to 6.4%, and the increase in %Hit across the haptic assist condition is again significant, $t(10)$, $p = .045$. Figure 7 shows a box plot of the average visual demand, Avg[VisD], by assist condition, in which each box represents $n = 11$ participants. As evident in the figure, the median visual demand was lower when the haptic assist was turned on.

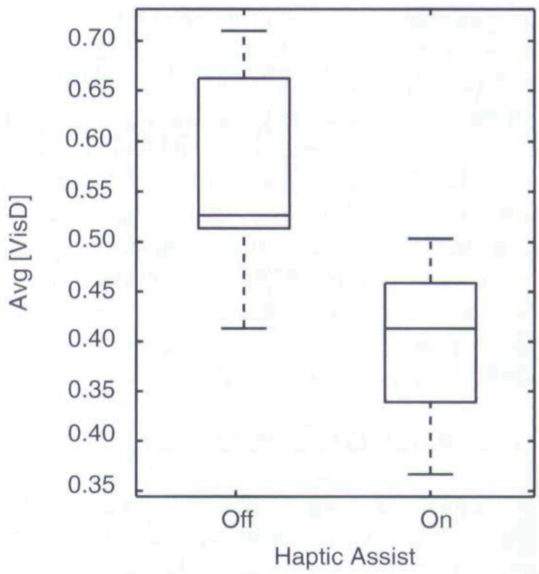

Figure 7. Box plot of Avg[VisD] by assist condition.

EXPERIMENT 3: SECONDARY TASK

The third experiment was aimed at quantifying the ability of the haptic assist steering wheel to aid the participant in a path-following task while reducing mental load in spatial processing. Mental-processing load was estimated by measuring performance on a secondary task that required the participant to localize tones emitted from three speakers. Tone localization was chosen as a secondary task on the assumption that it would interfere with the path-following task or would compete for the same spatial-processing code (Wickens & Liu, 1988). Also, the ability to localize sounds, including reaction times for localization from within a vehicle, can be critical to safety and overall driving performance (Wallace & Fisher, 1998).

TABLE 2: Mean and Standard Deviation of RMS[LE], %Hit, and Avg[VisD] For Experiment 2

Performance Metric	No Assist		With Assist		Paired t Test			
	$\bar{x}$	s	$\bar{x}$	s	$\Delta\bar{x}$	$s\Delta$	p	DOF
RMS[LE] (m)	1.412	0.727	0.8282	0.312	−0.584	0.513	.0018	10
%Hit	1.82	3.11	6.36	8.75	+4.54	8.07	.0456	10
Avg[VisD] (–)	0.570	0.0966	0.404	0.0765	−0.166	0.0563	<.0001	10

Procedure

Participants. The same 11 individuals who participated in Experiments 1 and 2 also participated in Experiment 3. Participants were screened for normal, balanced hearing by self-assessment.

Design. Two of the six trials performed by each participant pertained to Experiment 3, one each for the two haptic assist conditions. These two trials were designed to simultaneously assess driving performance and performance on a secondary task involving auditory localization of tones. The primary task was the same as in Experiments 1 and 2: to follow the center of the road as closely as possible while avoiding obstacles. Participants were not told which task was more important or how their performance on either task would be measured.

Secondary task: Tone location. Three computer speakers were placed approximately 1 m in front of the participant's head with an 18-cm center-to-center spacing on top of the computer monitor that displayed the simulated roadway. These speakers played 0.5-s square-wave tones with a fundamental frequency of middle C. The sound level reading at the participant's head location was measured to be 81 dBA. The time between tones was randomly selected, with a uniform distribution between 2 and 6 s. Participants were asked to identify which of the three speakers played the tone and to press a corresponding key on the computer keyboard. The *j* key was used to indicate that the left speaker had played, the *k* key to indicate the center speaker, and the *l* key to indicate the right speaker. Participants were not told that the speed of their response would be measured.

Performance by each participant on both the primary and secondary tasks was recorded for two 5-min trials: one trial without haptic assist and the other with haptic assist. The simulator logged the time, the vehicle position and orientation, the closest point on the centerline, the number of obstacles hit, the torque displayed on the motorized steering wheel, the tones sounded by the speakers, and the key presses registered by the keyboard.

Dependent variables and data analysis. To assess driving performance, we used the same two performance metrics defined for Experiment 1: RMS[LE] (in regions away from obstacles) and %Hit. These performance metrics were analyzed as described previously for Experiment 1.

Two additional performance metrics were defined for the secondary tone location task: accuracy and reaction time. Accuracy, denoted ToneAcc, was defined as the percentage of tones that were correctly identified. The reaction time, denoted RT, was the time in milliseconds between the tone onset and the registration of the key press by the personal computer. Because of technical limitations, RT data were quantized to 8-ms levels. The precision of the timing, however, was better than 1 μs, and timing jitter (standard deviation of sample time) was measured at 0.53 ms. Quantization and timing jitter can be considered noise in the measurements of response time. Data analysis for the various performance metrics included quantile plots, multiple-factor ANOVA, and *t* tests. The value $\alpha = .05$ was used to establish statistical significance.

ToneAcc performance was first determined independently for each speaker by computing the percentage of correct responses for a particular speaker (left, center, or right). For example, the location accuracy for the left speaker is the number of times the *j* key was hit in response to a tone from the left speaker as a percentage of the total number of tones from the left speaker during the 5-min trial (about 37 tones). Quantile plots of the tone location accuracy data showed that these data were approximately normally distributed. ANOVA showed that the speaker (right, center, left) was not a significant main effect. Thus the data by speaker were combined to define a single ToneAcc performance metric for each participant and assist condition.

Quantile plots were also used to check for normality of the RT data. The RT data were, as expected, not normally distributed, so the use of transformations and data set truncations was investigated. The inverse (1/RT) transformation, the logarithmic (log[RT]) transformation, and truncations of the RT data to 0.75, 1.0, 1.25, 1.5 and 2.0 s were applied to the data; *truncation* refers to exclusion of all data beyond the range specified. The influence of these various transformations and truncations on the

ANOVA-reported *p* values were compared with the influence of the same operations on the *p* values generated by synthesized data, as reported in Ratcliff (1993). Data were synthesized in Ratcliff (1993) by specifying a difference in means or tail sizes between two ex-Gaussian distributions. A comparison of *p* value trends produced by the operations on our data versus the operations on the synthesized data suggested that a 1.25-s cutoff produced the most statistical power and suggested further that our data showed a difference in tail sizes. The rationale behind truncation is that the longer RT data are spurious in the sense that they are strongly influenced by processes other than the condition being tested, such as distractions or the intrusion of cognitive processes not relevant to the experiment (Ulrich & Miller, 1994). Despite quantization of timing data, the high precision of the timing data and low software jitter allows the mean difference to be extracted with precision similar to the jitter (±0.5 ms). Note that each of the 11 participants reacted to 112 tones in each trial.

Results

As in Experiments 1 and 2, the presence of haptic assist produced a reduction in RMS[LE] in regions away from obstacles. The data presented in Figures 5 and 6 for Experiment 1 are also representative of the effects of the assist condition on path-following performance for Experiment 3.

Multiple-factor ANOVA applied to RMS[LE] revealed significant main effects attributable to assist condition, $F(1, 21) = 8.78$, $p = .014$, $MSE = 0.347$, and subject, $F(10, 21) = 3.4$, $p = .033$, $MSE = 0.134$. For %Hit, assist condition was not a significant main effect, but subject was a main effect, $F(1, 21) = 5.07$, $p = .0085$, $MSE = 0.00682$.

The sample mean, $\bar{x}$, and standard deviation, *s* (collapsed across participants), of RMS[LE], %Hit, and ToneAcc are presented in Table 3, along with the *p* values of paired *t* tests applied to the difference in the means, $\Delta\bar{x}$. There was a significant 38% reduction in RMS[LE], $t(10)$, $p < .0093$. That is, path following was significantly improved in the regions between obstacles with the addition of haptic assist. The percentage of obstacles hit increased from 3.4% to 4.3%, but the difference was not statistically significant. Of the two new performance metrics that measured performance on the secondary (tone localization) task, ToneAcc did not change appreciably. The difference in mean percentage of correct identifications rose slightly with the addition of haptic assist but without any statistical significance, $t(10)$, $p = .604$.

RT was the other performance metric for the secondary task. After a multiple-factor ANOVA revealed that the effect of path curvature on the RT data was not a significant main effect, $F(1, 2242) = 0.21$, a multiple-factor ANOVA was performed considering assist condition and proximity to obstacles (near/between) as experimental factors. The ANOVA reported significant main effects in haptic assist condition, $F(1, 2286) = 8.5$, $p = .003$, $MSE = 0.16$, proximity to obstacles, $F(1, 2286) = 35$, $p = .0001$, $MSE = 0.66$, and subject, $F(10, 2286) = 52$, $p = .0001$, $MSE = 0.97$, with significant interactions in assist by subject, $F(10, 2286) = 0.12$, $p = .0001$, $MSE = 0.12$, and in proximity to obstacles by subject, $F(10, 2286) = 3.2$, $p = .0005$, $MSE = 0.059$.

A paired *t* test was performed for the RT data in which the participants' mean reaction time was subtracted from their respective RT data and the population means with and without

TABLE 3: Mean and Standard Deviation of RMS[LE], %Hit, ToneAcc, and RT for Experiment 3

Performance Metric	No Assist		With Assist		Paired *t* Test			
	$\bar{x}$	*s*	$\bar{x}$	*s*	$\Delta\bar{x}$	$s\Delta$	*p*	DOF
RMS[LE] (m)	0.817	0.520	0.503	0.221	–0.314	0.264	.0093	10
%Hit	3.64	6.90	4.55	5.83	+0.91	5.18	.287	10
ToneAcc (%)	94.9	4.41	95.3	3.92	+0.35	4.29	.604	10
RT (ms)	564	147	545	132	–18.2	143	.0009	2318

haptic assist were compared. A statistically significant 18-ms decrease in RT was found with haptic assist as compared with no assist, $t(2328)$, $p = .0009$, regardless of proximity to obstacles. When only RT data between obstacles were considered, the effect of adding haptic assist was a 21-ms decrease in RT, $t(1612)$, $p = .0005$, and when only RT data near obstacles were considered, the effect of adding haptic assist was a 10-ms increase in RT, but without statistical significance, $t(714)$, $p = .172$. Because proximity to obstacles was a main effect, we determined the effect of proximity to obstacles without regard to assist condition, and the mean reaction time was found to increase by 37 ms, $t(2328)$, $p < .0001$, when participants were near obstacles.

GENERAL DISCUSSION

The data from all three experiments may be examined to compare the effects of haptic assist with the effects of imposing the conditions of the visual occlusion method and of adding a secondary task. Figure 8 shows six box plots defined by the medians and upper and lower quartiles for the two assist conditions (on/off) and by the three experiments (1 = baseline, 2 = visual demand, and 3 = secondary task). The improvement in path-following performance afforded by the haptic assist is evident in each experiment. The cost in path-following performance incurred by the conditions of the visual occlusion method and the addition of a secondary task are also evident when comparing across experiments. Although the availability of haptic assist does not restore path-following performance to baseline levels under visual occlusion, haptic assist does restore path-following performance to baseline performance with haptic assist under the secondary task. These data show that haptic assist can improve path-following performance to a degree that does not diminish even when a secondary task is added.

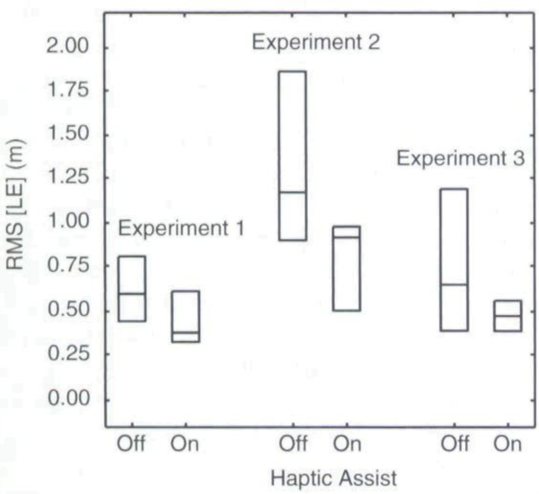

Figure 8. Box plots of RMS[LE], showing medians and quartiles for the two assist conditions and the three experiments.

To achieve path following, the stiffness of the shared controller must be tuned to balance two conflicting goals. The automation must resist deviations from its desired steering angle with enough authority to reject disturbances such as wind gusts and road crown. The automation effectively resists undesired movement of the steering wheel by constructing a virtual spring to hold the wheel. A greater spring stiffness achieves the tracking goals better. However, in sharing control, the driver must overcome the stiffness presented by the virtual spring if he or she wishes to steer with an angle other than that determined by the automation. If the virtual spring is too stiff, the driver may find it difficult to overpower the controller's actions, but if the spring is very weak, disturbances would cause excessive error in lane keeping with the controller acting alone.

As discussed in the Results sections for the three experiments, the improvement in path-following performance was usually accompanied by reduced obstacle avoidance performance (increased %Hit). Figure 9 shows the number of obstacles hit for the two assist conditions (on/off) for each of the three experiments. A marker for each participant indicates the number of obstacles hit in each of the six trials, arranged by assist condition and experiment. The dependence of the number of obstacles hit on the participant is evident. The reduction in obstacle avoidance performance with the addition of assist is clear, especially in Experiments 1 and 2. Note, however, that a reduction in performance is a natural consequence of an assist system that helps maintain path following but that is not aware of obstacles that lie on the path. Note also that the decrement in performance incurred by the haptic assist is about the same magnitude

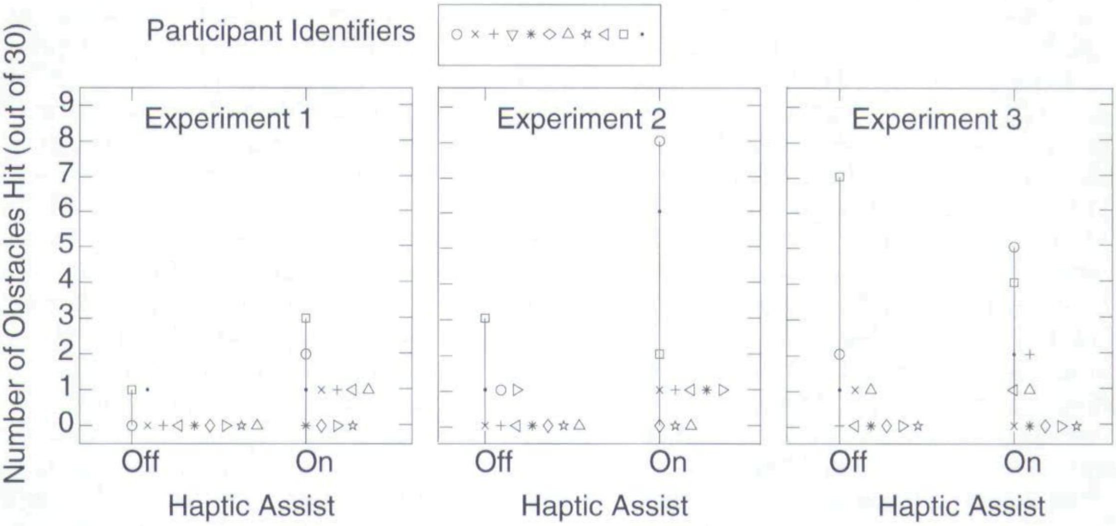

Figure 9. Plots of the number of obstacles hit for each participant are shown for the two assist conditions and the three experiments.

as the decrement incurred by the addition of a secondary task. Also, in the presence of the secondary task, the addition of haptic assist does not incur a statistically significant further performance cost in obstacle avoidance. These results, however, are somewhat inconclusive. They merit further exploration in future studies.

Certainly the reduction in obstacle avoidance performance incurred by haptic assist motivates the development of sensors in the automobile that can serve in collision warning or collision avoidance systems. The assist could be turned off and a warning (audio, visual, or haptic) sounded when an obstacle is detected by the automatic controller. A more proactive system would assess the traffic situation and help the driver make an evasive maneuver by planning a path around the obstacle.

Accuracy and reaction time were the metrics of participants' performance in the tone location experiment. The hypothesis of Experiment 3 was that the performance of the secondary task would improve if the driver was provided haptic assist. Indeed, there was an improvement (reduction) of the reaction time by 18 ms, $t(2328)$, $p = .0009$, but the improvement in the localization accuracy was small and not statistically significant. The reduction in reaction time is a desirable result for two reasons. Taken by itself, a faster reaction means that the driver can react sooner to dangerous situations – for example, the honk of a horn from another vehicle. The faster reaction time to the auditory probe also implies that with shared control, the driver experiences a reduced cognitive load in the driving task (Raney, 1993). The validity of using reaction time as an indicator of cognitive load was affirmed by the 37-ms increase in reaction time when participants were near obstacles (actively engaged in obstacle avoidance) versus when they were far from obstacles, $t(2328)$, $p < .0001$. So, in addition to the improved performance in the primary control task, the haptic assist evidently increases the availability of cognitive processing capacity for performance of a secondary task.

The presence of the secondary task decreased the path-following performance by 18%, when compared with the baseline experiment with no assist, $t(10)$, $p = .033$. This statistic is further evidence that the spatial reasoning task selected for the secondary task was competing for some of the same cognitive processing capacity required for driving (the primary task). When haptic assist was enabled and path following in the baseline condition was compared with path following in the presence of the secondary task, there was only a 6% degradation in performance, which was not statistically significant, $t(10)$, $p = .64$. This result suggests that the haptic assist can allow a driver to perform a secondary task with negligible degradation in tracking performance.

SUMMARY

We have investigated the use of haptic interface to realize and test the idea of a human driver sharing control of vehicle heading with an automatic controller. The human and controller share the same control interface (e.g., steering wheel) and are mechanically interconnected such that they may exchange information and share control authority with each other. Haptic display becomes the means to place the automatic controller in the haptic perceptual space of the human. The human is free not only to observe the actions of the controller but may override them at any time he or she sees fit, based on his or her perception of additional factors in the task environment. Shared control extends the notion of a virtual fixture to a virtual agent or copilot. Like the virtual fixture, the human is aware of the virtual agent by feel and can use the agent to negotiate a task more efficiently.

Although certain recommendations about haptic versus other types of assist remain unexplored in our present experiments, our findings hold some important implications for the design of automation systems. We demonstrated a significant reduction in visual demand and a freeing of attention by inserting automation into the control loop through a motor on the manual interface. Both of these positive effects were achieved while significantly improving performance on the primary driving task, which was shared by the human-machine team. Through the use of haptic assist, the human remains in the loop, and so we believe that shared control through a haptic interface incurs minimal loss of obstacle avoidance performance to surprise events in the primary task.

We hypothesize that the mechanical coupling between hand and manual interface and the colocated sensing and actuation functions of the hand keep the operator in the loop. Thus we expected to measure maintained obstacle avoidance, but this expectation was not fully borne out in our data. There was a statistically significant reduction in obstacle avoidance with the addition of haptic assist. However, this should be considered in light of the fact that our measure of obstacle avoidance was strongly confounded with the haptic assist condition. Specifically, the haptic assist favored the center of the path, where the obstacles were located. This condition was most apparent in the baseline trial, without the conditions of visual occlusion or the presence of a secondary task. Notably, under the visual occlusion and secondary task conditions, the diminished obstacle avoidance performance incurred by the automation disappeared.

The methods and experiments described in this paper both draw from and contribute to the fields of human factors and haptic interface. In the field of human factors, substantial knowledge already exists about the effects of adding automation on the performance of human-machine teams. However, the use of haptic feedback and its relative merit as compared with visual or auditory feedback has not been evaluated for control sharing. In particular, haptic feedback can be used to provide information regarding continuous control action taken by an automation system that does not further load the visual or auditory systems. In this paper we have demonstrated improved performance with the addition of automation while actually reducing visual demand.

In the haptic interface field, the ability of a manual interface to simultaneously display information and function as a control input device is well understood and used to advantage in numerous applications. However, only the primary task performance benefits of adding automation with haptic feedback have received attention in the field of haptic interface. The auxiliary performance benefits of haptic assist, either in the reduction of a particular mental workload (e.g., spatial processing) or the reduction in demand on other perceptual systems, are often sought in applications, and they are sometimes claimed but seldom quantified. Our work combines the concepts of colocated action and sensing inherent to haptic interfaces and examines not only the human-machine performance in the primary task but also human performance in a secondary task.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, 775–779.

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research: Automation* (Vol. 2, pp. 1–12). New York: Elsevier Science.

Dennerlein, J. T., & Yang, M. C. (2001). Haptic force-feedback devices for the office computer: Performance and musculoskeletal loading issues. *Human Factors, 43,* 278–286.

Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 179–220. Oxford, UK: Oxford University Press.

Gillespie, R. B. (2004). Haptic interface to virtual environments. In T. Kurfess (Ed.), *CRC handbook on robotics and automation* (pp. 23.1–23.23). Boca Raton, FL: CRC Press.

Gillespie, R. B., Hoffman, M., & Freudenberg, J. (2003). Haptic interface for hands-on instruction in system dynamics and embedded control. In *11th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, IEEE Virtual Reality* (pp. 125–131). Los Alamitos, CA: IEEE Computer Society.

Gillespie, T. D. (1992). *Fundamentals of vehicle dynamics.* Warrendale, PA: Society of Automotive Engineers.

Hasser, C. J., Goldenberg, A. S., Martin, K. M., & Rosenberg, L. B. (1998). User performance in a GUI pointing task with a low-cost force-feedback computer mouse. In *Proceedings of the ASME International Mechanical Engineering Congress and Exposition* (Vol. 64, pp. 151–155). New York: American Society of Mechanical Engineers, Dynamic Systems and Control Division.

Hayward, V., Astley, O. R., Cruz-Hernandez, M., Grant, D., & Robles-De-La-Torre, G. (2004). Haptic interfaces and devices. *Sensor Review, 24,* 16–21.

Hess, R. A., & Modjtahedzadeh, A. (1990). A control theoretic model of driver steering behavior. *IEEE Control Systems Magazine, 10*(5), 3–8.

Li, M., & Okamura, A. M. (2003). Recognition of operator motions for real-time assistance using virtual fixtures. In *11th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, IEEE Virtual Reality* (pp. 125–131). Los Alamitos, CA: IEEE Computer Society.

Meyer, J. S., Rubinstein, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 763–797.

Mosier, K. L. (2002). Automation and cognition: Maintaining coherence in the electronic cockpit. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research: Automation* (Vol. 2, pp. 93–122). New York: Elsevier Science.

Park, S. S., Howe, R. D., & Torchiana, D. F. (2001). Virtual fixtures for robot-assisted minimally invasive cardiac surgery. In *Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 1419–1420). Berlin: Springer-Verlag.

Patoglu, V., & Gillespie, R. (in press). Feedback stabilized minimum distance maintenance for convex parametric surfaces. *IEEE Transactions on Robotics.*

Payandeh, S., & Stanisic, Z. (2002). On application of virtual fixtures as an aid for telemanipulation and training. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 18–23). Los Alamitos, CA: IEEE Computer Society.

Raney, G. E. (1993). Monitoring changes in cognitive load during reading: An event-related brain potential and reaction time analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 51–69.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

Rosenberg, L. B. (1993). Use of virtual fixtures to enhance telemanipulation with time delay. In *Proceedings of the ASME Winter Annual Meeting* (Vol. 49, pp. 29–36). New York: American Society of Mechanical Engineers, Dynamic Systems and Control Division.

Sarter, N. B. (2002). Multimodal information presentation in support of human-automation communication and coordination. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research: Automation* (Vol. 2, pp. 13–36). New York: Elsevier Science.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1926–1943). New York: Wiley.

Sayers, C., & Paul, R. (1994). An operator interface for teleprogramming employing synthetic fixtures. *Presence, 3,* 309–320.

Schumann, J., Godthelp, J., & Hoekstra, W. (1992). *An exploratory simulator study on the use of active control devices in car driving* (Tech. Rep. No. IZF 1992 B-2). Soesterberg, Netherlands: TNO Institute for Perception.

Steele, M., & Gillespie, R. B. (2001). Shared control between human and machine: Using a haptic interface to aid in land vehicle guidance. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1671–1675). Santa Monica, CA: Human Factors and Ergonomics Society.

Suzuki, K., & Jansson, H. (2003). An analysis of driver's steering behaviour during auditory or haptic warnings for the designing of lane departure warning system. *Society of Automotive Engineers of Japan Review, 24,* 65–70.

Switkes, J. P., Rossetter, E. J., Coe, I. A., & Gerdes, J. C. (2004). Handwheel force feedback for lanekeeping assistance: Combined dynamics and stability. In *Proceedings of AVEC '04, 7th International Symposium on Advanced Vehicle Control* (pp. 47–52). Tokyo: Society of Automotive Engineers of Japan.

Tsimhoni, O., & Green, P. A. (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1586–1590). Santa Monica, CA: Human Factors and Ergonomics Society.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General, 123,* 34–80.

Wallace, J. S., & Fisher, D. L. (1998). Sound localization: Information theory analysis. *Human Factors, 40,* 50–68.

Wickens, C. D., & Liu, L. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30,* 599–616.

Paul G. Griffiths is a Ph.D. student of mechanical engineering at the University of Michigan. He received his M.S. in mechanical engineering from the University of California, Berkeley, in 2002.

R. Brent Gillespie is an assistant professor of mechanical engineering at the University of Michigan. He received his Ph.D. in mechanical engineering from Stanford University in 1996.

Date received: February 28, 2002
Date accepted: April 20, 2005