

ASSESSING COGNITIVE DISTRACTION USING EVENT RELATED POTENTIALS

James R. Coleman¹, Jonna Turrill¹, Rachel J. Hopman¹,
Joel M. Cooper^{1,2}, and David L. Strayer¹
University of Utah¹ and Precision Driving Research²

Summary: This report examines the utility of using Event-Related Brain Potentials (ERPs) to evaluate cognitive distraction in the context of driving an automobile. Across two studies, ERPs (both P300 latency and P300 amplitude) were found to be effective in quantifying the cognitive workload experienced by drivers when they interact with in-vehicle voice-command systems.

OBJECTIVES

The current study provides an assessment of the utility of event related brain potentials (ERPs) to evaluate cognitive distraction as drivers interact with voice-based systems in the vehicle. ERPs were collected from participants in two separate experiments. The first provided a baseline control for the distraction potential of several in-vehicle voice-command interactions. The second assessed the effect these same voice-based interactions on simulated driving. In both experiments, the ERPs were elicited by a detection-response task (DRT) using a go-no-go oddball variant of the task. When applied to driving, ERPs can provide a quantifiable and objective measurement of dual-task workload. However, as the driving environment becomes more complex, the signal-to-noise ratio in the ERP diminishes.

EXPERIMENT 1

Methods

Participants. Ten participants enrolled in Experiment 1 (6 male) and 10 participants enrolled in Experiment 2 (7 male). All were students at the University of Utah and ranged in age from 20 to 34 years, with an average age of 24.7 years. Participants reported having normal neurological functioning, normal or corrected-to-normal visual acuity, normal color vision, a valid driver's license, and were fluent in English with an average of 8.7 (4 min, 18 max) years of driving experience. All of the participants owned a cellular phone and 95% reported that they used their phone regularly while driving.

Equipment. Microsoft PowerPoint 2013 was used to coordinate an interactive messaging service with text to speech features. Participants were given a short list of commands (i.e., *Repeat*, *Reply*, *Delete*, *Next Message*, and *Send*) that were used to control the messaging program. The experimenter, who reacted to the participants' verbal commands, mimicked a speech detection system with perfect fidelity.

The cellular phone was manufactured by Apple (Model iPhone 5) running iOS 6 or iOS 7. An Olympus ME-15 Mono Lapel microphone was clipped to the participant's collar for a voice-controlled Siri messaging system. An iPhone headset microphone adapter was used to allow output from and input to the iPhone 5 when participants used vocal commands to interact with

Siri. TEAC CD-X70i Micro Hi-Fi system speakers were used for the presentation of the audio for each of the conditions.

NeuroScan 4.5 software was used to collect continuous EEG that was recorded using a NeuroScan 32-electrode NuAmp amplifier. The EEG was filtered online with a DC notch filter (60 Hz) with a sample A/D rate of 250 Hz. The DRT software communicated with the NeuroScan system via a parallel port connection to create event markers associated with the continuous EEG. These event markers allowed for offline stimulus-locked analysis of the EEG recordings. The influence of blinks on the EEG was corrected using ocular artifact rejection techniques (Semlitsch, Anderer, Schuster, & Presslich, 1986) and the data was epoched 200ms before to 1200ms after the onset of the DRT target stimuli. Epochs were then filtered with a band pass, zero phase shift filter of 0.01 to 12 Hz. Finally, events that exceeded an artifact rejection criterion of 100 μ V were rejected and the remaining events were averaged to obtain one subject's averaged waveform for each condition in the experiment.

The Detection Response Task (DRT) hardware and software were designed by Precision Driving Research, Inc. following ISO preliminary standards (ISO, 2012). Adopting the protocol used by Strayer et al., (2013), a red/green LED light was mounted on the participant's head via a headband with an LED light attached to the end of an adjustable arm. The light was adjusted to an average 15° to the left and 7.5° above the participant's left eye. Red/green lights were presented in an 80/20-oddball sequence. ERPs were time-locked to the onset of green (rare) lights. Response reaction time was recorded with millisecond accuracy via a microswitch attached to participants' left thumb that was depressed in response to the green light.

Procedure. In Experiment 1, participants were seated 65 cm in front of a computer displaying a static fixation cross and asked to look forward and avoid making excessive head and eye movements. In Experiment 2, participants were seated in a fixed-base driving simulator with a 180-degree field of view (manufactured by L-3 Communications) and asked to drive on a multilane freeway with moderate traffic (approximately 1500 vehicles/lane/hour). Participants were asked to maintain a 2-second following distance behind a pace car and they were given 5 minutes of practice to familiarize themselves with the driving simulator.

Table 1 presents the nine different 9-minute conditions that were counterbalanced across participants using a Latin Square design. Before each condition began, participants were familiarized with the procedures for interacting with the system and they were required to demonstrate proficiency before data collection for that condition commenced. Message type and duration were equated in conditions 3-6. In Experiment 2, participants were given a 5-minute practice session to familiarize themselves with the driving simulator, scenario design, and prescribed following distance.

Table 1. The nine conditions used in Experiments 1 and 2

1) Single Task	The single task condition provided a baseline measure of cognitive workload without a concurrent secondary task.
2) Car Commands	Every 30-45 seconds, a short audio cue was played (e.g., “ <i>You are getting hot</i> ” or “ <i>You want to change the radio station</i> ”). Participants interpreted the cue and then stated a verbal command in response to the cue (e.g., “ <i>Turn AC on low</i> ” or “ <i>Tune radio to 88.3</i> ”).
3) Natural Listen	Participants interacted with a simulated email/text messaging service. The system was fully automated with perfect speech recognition capability implemented using a “Wizard-of-Oz” technique (Kelley, 1983; Lee, Caven, Haake, & Brown, 2001; Strayer et al., 2013). The email and text messages and the system confirmations were pre-recorded using a female voice (author J.T.). Participants were asked to <i>listen</i> to the messages, but they were not allowed to compose or send messages in reply.
4) Synthetic Listen	Participants interacted with a simulated email/text messaging service. The system was fully automated with perfect speech recognition; however, the messages and system confirmations were pre-recorded using a synthetic, computerized female voice, “Kate,” from NeoSpeech (NeoSpeech, 2012). Participants were asked to <i>listen</i> to the messages, but were not allowed to compose or send messages in reply.
5) Natural Compose	Participants interacted with a simulated email/text messaging service. The system was fully automated with perfect speech recognition. The messages and system confirmations were pre-recorded using the same female voice (author J.T.) used in the third condition. Participants were asked to <i>listen</i> and then <i>compose</i> a response to messages that required a response.
6) Synthetic Compose	Participants interacted with a simulated email/text messaging service. The system was fully automated with perfect speech recognition. The messages and system confirmations were pre-recorded using NeoSpeech. Participants were asked to <i>listen</i> and then <i>compose</i> a response to messages that required a response.
7) Menu High Reliability	Participants interacted with a simulated, infotainment/navigation system. They were instructed to <i>navigate</i> through an auditory menu system to <i>select</i> a grocery store, coffee shop, gas station, bank, or a restaurant location.
8) Menu Low Reliability	Participants interacted with a simulated, infotainment/navigation system. They were instructed to <i>navigate</i> through an auditory menu system to select a grocery store, coffee shop, gas station, bank, or a restaurant location. However, the system randomly introduced errors that occurred on average 7.8 (sd = 2.15) times during the 9-minute condition.
9) Eyes/hands-free Siri	Participants interacted with Apple’s Siri to perform three tasks in a self-paced order: Listening to and sending text messages, updating Facebook or Twitter status, and modifying and reviewing calendar appointments. Participants were instructed to activate an eyes-free version of Siri by saying, “ <i>Hello/Hi Siri.</i> ” The researcher would then manually activate Siri. The participant neither looked at nor made physical contact with the iPhone.

Results

Figure 1 presents the grand average ERP waveforms obtained in Experiment 1 at the midline Parietal electrode site (Pz) that were time-locked to the onset of green lights in the DRT task. In the figure, the amplitude in microvolts is cross-plotted with time in msec. Inspection of the figure reveals a well-defined P2-N2-P300 ERP component structure. We focused on the P300 component of the ERP because of its sensitivity to cognitive workload and we measured both its peak latency and amplitude (Kutas, McCarthy, & Donchin, 1997; Fowler, 1994).

In Figure 2, P300 peak latency, measured as the point in time of maximum positivity in a window between 350 and 800 msec, is plotted for each of the conditions in the experiment. A repeated measures ANOVA found no significant main effect of condition on P300 latency, $F(8, 72) = 1.57, p = ns, \text{partial } \eta^2 = .15$. The P300 amplitude was quantified by computing the average area under the curve between 350 and 800 msec. Figure 3 plots P300 amplitude as a function of condition. A repeated measures ANOVA found a main effect of condition, $F(8, 72) = 2.39, p < .05, \text{partial } \eta^2 = .21$.

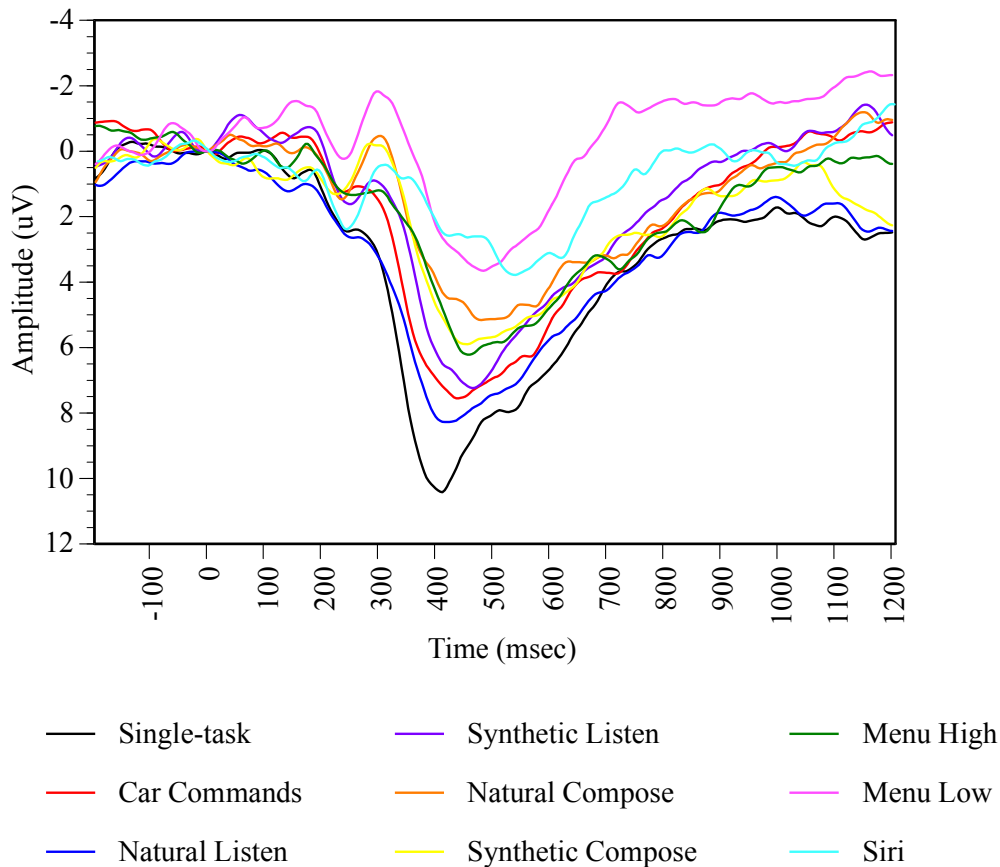


Figure 1. Experiment 1 ERPs at Pz

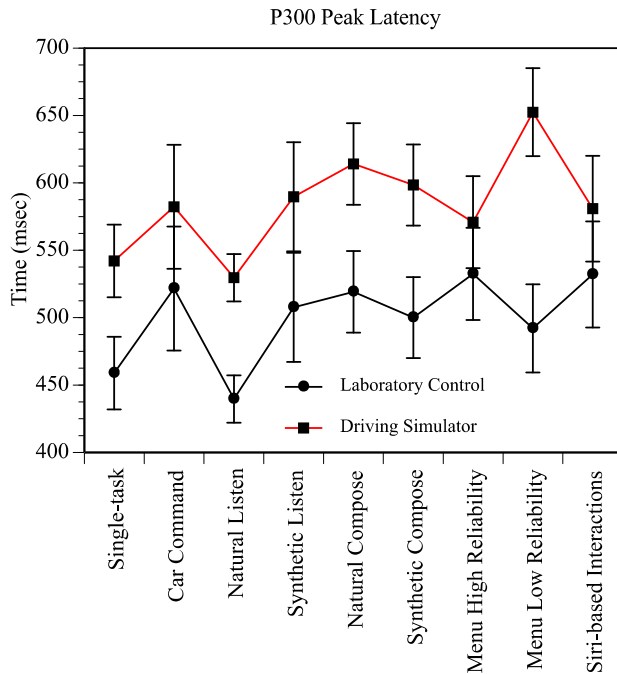


Figure 2. P300 Latency (Experiments 1-2)

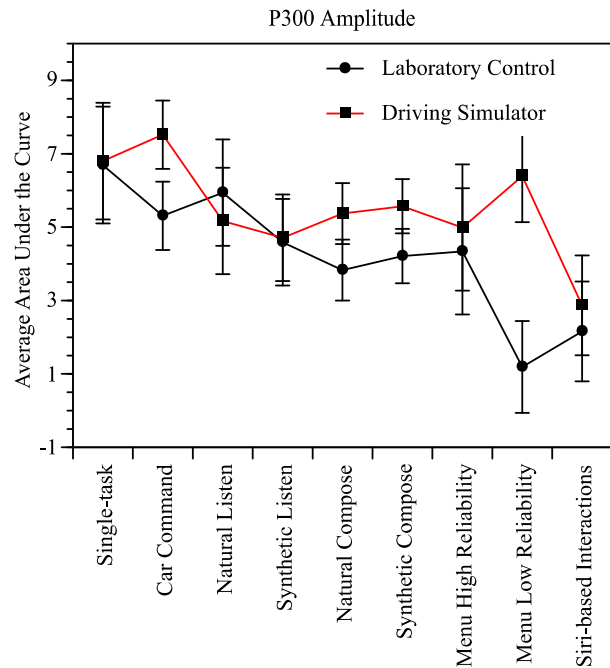


Figure 3. P300 Amplitude (Experiments 1-2)

EXPERIMENT 2

Results

EEG was recorded and analyzed in Experiment 2 using the same protocol as that of Experiment 1. Figure 4 presents the grand average ERP waveforms obtained in Experiment 2 at the midline Parietal electrode site (Pz) that were time-locked to the onset of green lights in the DRT task. Figure 2 presents the P300 latency means and standard error as a function of condition. A repeated measures ANOVA found no significant main effect of condition of P300 latency $F(8, 72) = 1.49, p = ns, \text{partial } \eta^2 = .14$. As in Experiment 1, the P300, presented in Figure 3 was quantified by computing the average area under the curve between 350 and 800 msec. Means and standard error are presented Figure 3. A repeated measures ANOVA of the P300 area under the curve found no effect of condition, $F(8, 72) = 1.05, p = ns, \text{partial } \eta^2 = .10$.

A MANOVA combining Experiments 1 and 2 was performed on P300 latency and amplitude. The P300 latency MANOVA revealed a main effect of condition, $F(8,144) = 2.25, p < .05, \text{partial } \eta^2 = .11$ and an effect of experiment, $F(1,18) = 9.65, p < .01, \text{partial } \eta^2 = .35$; however, the condition X experiment interaction was not significant ($p > .56$). A subsidiary trend analysis found that P300 latency *increased* as a linear function of condition, $F(1,18)=7.04, p < .05, \text{partial } \eta^2 = .28$. The P300 amplitude MANOVA revealed a main effect of condition, $F(8,144) = 2.22, p < .05, \text{partial } \eta^2 = .11$; however, neither the effect of experiment nor the condition X experiment interaction was significant (all p 's $> .38$). A subsidiary trend analysis found that P300 amplitude *decreased* as a linear function of condition, $F(1,18)=8.5, p < .01, \text{partial } \eta^2 = .32$.

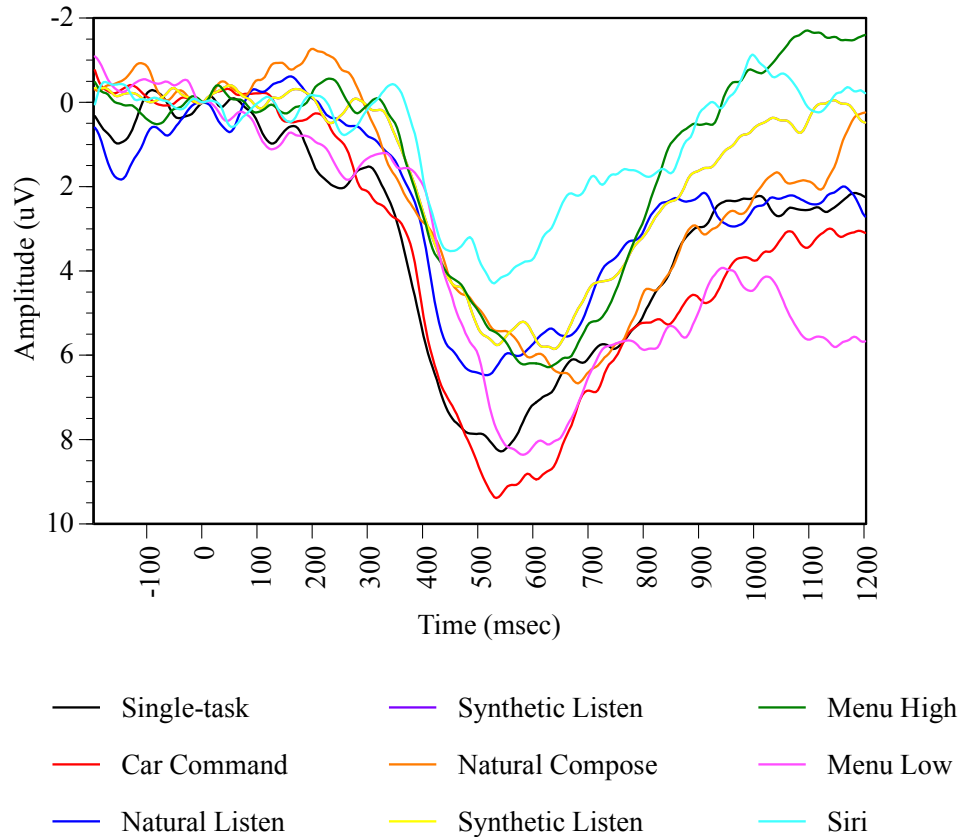


Figure 4. Experiment 2 ERPs at Pz

GENERAL CONCLUSIONS

ERPs were found to be an effective way of quantifying cognitive workload while driving in simulated environments, thereby providing a clear and objective assessment of attentional allocation in the driving environment. Both speech generation and the fidelity of the in-vehicle voice-command system were the largest factors affecting the cognitive workload of the driver. In tasks where participants had to verbally respond to the system and/or compose messages, a marked change in the P300 was observed compared to single task conditions (i.e., P300 latency increased and P300 amplitude decreased). An even greater change was observed when the systems were error-prone in their interactions with participants (i.e. Siri and Menu Low conditions).

The ERPs became degraded as we moved from the laboratory to the driving simulator due to the increased biological noise from eye/head/body movements and electronic noise from the driving simulator. This degradation of signal quality may prove problematic for on-road driving studies (with greater ambient noise), as the authors have found in prior research (Strayer et. al, 2013). Nevertheless, in controlled laboratory settings, ERP measures provide a sensitive and objective metric of the cognitive workload experienced by drivers as they interact with different in-vehicle systems.

ACKNOWLEDGMENTS

We acknowledge the support of the AAA Foundation for Traffic Safety.

REFERENCES

- Fowler, B. (1994). P300 as a measure of workload during a simulated aircraft landing task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 36(4), 670-683.
- ISO. (2012). Road vehicles -- Transport information and control systems -- Detection-Response Task (DRT) for assessing selective attention in driving. ISO TC 22 SC 13 N17488 (Working Draft). *Under development by Working Group 8 of ISO TC22, SC 13*.
- Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. *Proceedings of ACM SIG-CHI '83 Human Factors in Computing Systems* (pp. 193-196). Boston: New York, ACM.
- Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interactions with in-vehicle computers; The effect of speech-based e-mail on drivers' attention and roadway. *Human Factors*, 43, 631-640.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, 197(4305), 792-795.
- NeoSpeech (2012). NeoSpeech Text-to-Speech voices [computer software]. Santa Clara, CA. Available from <http://www.neospeech.com/>.
- Scheer, H. J., Sander, T., & Trahms, L. (2006). The influence of amplifier, interface and biological noise on signal quality in high-resolution EEG recordings. *Physiological measurement*, 27(2), 109.
- Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, 23, 695-703.
- Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile. *AAA Foundation for Traffic Safety*.