Using Simplified Grammar for Voice Commands to Decrease Driver Distraction

Ashraf Gaffar¹, Shokoufeh Monjezi Kouchak²

¹Ira.A Fulton School of Engineering, Arizona State University, Arizona, USA, agaffar@asu.edu
²Ira.A Fulton School of Engineering, Arizona State University, Arizona, USA, smonjezi@asu.edu

Abstract - Today’s car is considered a Cyber Physical System implying a shift of control and management of the car’s features to complex onboard software. Since the computer has taken control over most of the in-car driving and entertainment (infotainment) activities, driver distraction has become one of the primary reasons for car-related fatal accidents. As a result, we need to decrease driver distraction despite the growing presence of cockpit interactive digital technologies. Using voice-based commands is a potential solution to decrease visual distraction but it causes significant cognitive distraction. In this paper, we present a new method for voice-based commands to decreasing the cognitive distraction while using command based voice inputs.

Keywords: distraction; human car distraction; cognitive load, cyber-physical systems

1 Introduction

Driver distraction is one of the main reasons of car-related fatal crashes. NHTSA estimates that around 25% of traffic crashes in United States of America are caused by driver distraction [1]. Driver distraction is generally agreed upon to be caused by every person, event or object that persuades the driver to shift her attention from the fundamental task of driving [2]. Although distraction is not directly or readily observable, it is clearly manifested in why -under some conditions- the performance of a task is degraded [3]. These conditions are typically considered related to a distractor.

As mentioned earlier, today’s car system is a Cyber Physical System [4]. The car lifecycle right from its early stage of design and build phase to its running phase is controlled by software. Automotive industry is a software-heavy domain with software dominating the design, manufacturing, and maintenance of a car. Besides, the car itself is loaded with complex hardware and software components. Considering this, we focus our work on the interactive relationship between a driver and the digital environment of the cockpit, mostly in the infotainment system built in in the mid console, also known as the center stack. The modern infotainment system has three main groups of interactive software:

1. Built-In Features: Standard car information and local entertainment group originally loaded with the car or downloaded by the driver.

2. Bluetooth Connectivity: Providing additional features accessible via externally connected smart devices. These typically have hundreds of additional apps, contact lists, personal calendars, and other features. While originally designed for non-driving smart device user, they can often be synchronized, displayed and managed from the infotainment screen.

3. Web Connectivity: Onboard access to the World Wide Web, which can add a virtually unlimited number of interactive online features with a potential for serious distraction.

The main impact on the negative side of advanced Infotainment system is growing driver distraction by allowing drivers’ access to such apps and features originally not intended for driving environment. These can add to the large number of other distracting tasks like mobile phone use or talking to other passengers [5].

One challenge of measuring driver distraction is the fact that there are mixed types of distraction, and they can affect driver in different ways. The research community identifies four main types of driver distraction that can be present in isolation or together: Visual, auditory, manual and cognitive distraction [2]. Visual distraction occurs when driver focuses on entertainment system instead of keeping eyes on the road. Auditory distraction occurs when the sound in the car such as passenger conversation, traffic noise or music masks the warning noise that is necessary for self-monitored driving performance. Manual distraction happens when the driver

takes hands off the wheel to do non-driving related activity like eating or drinking. Cognitive distraction happens when the driver is absent-minded (looking but not seeing) or when the driver is busy with deep thinking about something [2][6].

A major research challenge is that depending on the method of study, evaluation, and interpretation, similar activities can be considered highly positive and highly negative in different scientific studies. For example, listening to music has been considered as harmless activity during driving in some papers [7] but was considered a distraction in some other papers [1]. Few of the previous research works say that listening to music during driving has effect on driving performance in two ways: distraction and mood effect.

Listening to music causes audio, visual and biomechanical distraction. According to a research in the United States of America 91% of drivers manipulate their cars’ audio system control and it takes 1% of driving time. They adjust their audio systems 8 times every hour and each interaction took 5.5 seconds.

Analysis of police accident report provided an evidence of a connection between road accidents and using in-car infotainment system, such as audio system. This analysis showed that in-car infotainment system was the first as well as the second reason for in-car distraction. [1]

Texting by mobile phone or using iPad can cause all 4 types of distractions. So using them during driving increases the driving errors significantly [8] and [9]. For example, using cell phone during driving has negative effect on lane keeping and event detection [10]. People who use phone during driving are four times more likely to meet with an accident that will result in hospital attendance; Age group and sex don’t influence the risk of accident when driver uses mobile phone during driving. Even using the hands-free feature is not helpful and it doesn’t decrease the chance of accident [11].

Although the tasks that don’t need visual attention or manual interaction cause less distraction, the potential for cognitive distraction has to be considered [12]. It is hard to determine attention status of driver so it is hard to solve cognitive distraction problem, because there isn’t any one-to-one relation between car and brain activities, so some ambiguity will be in the consequence of estimate [13].

The number of applications on the infotainment system in modern cars is expected to increase significantly during the next few years [14], so finding new methods to reduce the driver distraction that is caused by infotainment system will reduce the fatal crashes significantly. In this paper, we explain a new method to decrease the distraction that voice based commands of infotainment systems causes.

Earlier researches have proven that using the voice-based commands in the car user Interface (UI) reduces driver distraction for two reasons:

1. It reduces the visual distraction because the driver doesn’t need to take his eyes off the road and search among lots of apps to find his favorite app. Using touch screen instead of voice commands increases visual distraction significantly since the driver needs to several steps navigation for finding each application.

2. It reduces manual distraction since the drivers doesn’t need to take his hands off the steering wheel when he is using the voice commands.

Although using voice commands reduces both visual and manual distraction, it causes cognitive distraction and audio distraction.

Our hypothesis is that much of the cognitive load associated with voice command is related to the system not being able to understand commands all the time, forcing users to repeat or change the way they issue their command, taking a significant portion of their attention off the road trying to perceive the system ill-response and trying to figure out a different way to communicate their desire. We have used simplified voice based commands instead of all-natural voice based commands with complicated grammar. It was found that using simplified grammar significantly decreases distraction. In the following sections, we explain our hypothesis followed by the description and type of participants who took part in the experiment and the equipment that we used in the experiment. The results of the experiment are given in the form of statistical analysis.

2 Voice commands experiment

2.1 Using simplified grammar

Although voice based commands decreases the visual and manual distraction, cognitive distraction and audio distraction can significantly increase and even reach a dangerous level. Using complicated grammar increases the cognitive distraction since the drivers makes more mistakes such as mispronunciation and misunderstanding, the chance of misdiagnosed voice command increases so the driver needs to repeat each voice commands several times as a result the cognitive distraction increases. When the driver uses voice commands with simple grammar the number of his mistakes decreases so the number of misdiagnosed voice commands decreases. Less misdiagnosed voice commands means less cognitive distraction. Listening to sentences during driving causes distraction so reducing the number of misdiagnosed voice commands by using voice commands with simple grammar reduces the audio distraction too.
We designed an experiment to find out the different between response time and the number of errors when the driver uses voice commands with simple grammar and when the driver uses the voice commands with complicated grammar. Response time has direct correlation with cognitive distraction so reducing the response time would decrease the cognitive distraction. Less response time for a command shows that this command causes less cognitive load so it reduces the driver distraction.

Does using simple grammar in speech mode communication instead of sophisticated voice commands reduce response time? Our experiment’s results show that using simplified grammar for the voice commands in speech based interaction reduces response time and distraction level compared to the voice commands with complicated grammar.

2.2 Equipment and participants in experiment

The experiment was conducted in a Drive Safety Research simulator DS-600(Figure 1). The DS-600 is fully integrated, high performance, high fidelity driving simulation system which includes multi-channel audio/visual system, a minimum 180° wraparound display, full width automobile cab (Ford Focus) including windshield, driver and passenger seats, center console, dash and instrumentation and real time vehicle motion simulation.

An Android application was developed to display the user interface with numbers ranging from 1 to 9. The application was hosted on the Android v4.4.2 based Samsung Galaxy Tab 4 8.0 that was connected to the Hyper Drive Simulator.

We used 45 volunteers in our experiments each belonging to different sector participle in our experiment. We asked them to drive on a previously programmed route. During driving they were asked to interact with both voice based system with simplified grammar and normal grammar.

In this experiment, normal grammar User Interface (UI) referred to usual commands present in cars that have speech based communication models to interact with the center stack and simplistic Grammar user interface, a user interface that has voice commands with simple grammar, used in order to perform a set of operations during driving. Figure 2 and Figure 3 show an example of 2 categories of commands.

We measured two variables in this experiment: Response Time (in milliseconds) and Errors.

Response time: Response time is the total time taken by a person to interact with the system using voice command. Resulting in decreased response time causing less distraction as the driver comparatively spends less time on the secondary task.

Number of errors: The count of errors refers to the number of times that the driver changed the lanes incorrectly and hit other cars or buildings while using the voice commands. More errors imply more distraction.

We used two statistical methods to analysis our experiment. The first statistical method that was used to analysis this experiment is “comparing two alternatives”. Comparing between two alternatives used to compute the mean of differences of the paired measurements. The confidence interval C1 and C2 are determined with a confidence level, say 95%. This statistical method helped us
to determine if the results of our experiment are statistically significant.

\[
c_1 = \bar{d} - z_{1-\alpha/2} \frac{s_d}{\sqrt{n}}
\]

\[
c_2 = \bar{d} + z_{1-\alpha/2} \frac{s_d}{\sqrt{n}}
\]

If \(c_1 < 0\) and \(c_2 > 0\), then the measurements are statistically insignificant. If \(c_1\) and \(c_2 > 0\) or \(c_1\) and \(c_2 < 0\), then we can conclude the measurements are statistically significant.

The second statistically method that was used to analysis this experiment is analysis of variance (ANOVA). Analysis of variance is a technique to divide the total variation observed in an experiment into different meaningful components. This technique assumes that the errors in the measurements for different setting are independent with normal distribution.

By using ANOVA we can compute the variation that observed because of changing the alternatives and the variation that observed because of the error in measurement. The measurements are significant when the variation that observed because of changing alternatives is more than variation that observed because of the errors in measurement. The variation due to alternatives SSA and variation due to errors SSE are calculated using the below equations. Where \(k\) being the number of alternatives. In our case \(k = 2\), and \(n\) is the number users in the experiment. \(n = 20\).

\[
SST = \frac{\sum_{i=1}^{k}(\sum_{j=1}^{n}x_{ij}^2 - \bar{y}_{ij}^2)k}{kn} - \frac{(\sum_{i=1}^{k}(\sum_{j=1}^{n}y_{ij})^2)}{kn} \tag{2}
\]

\[
SSA = \frac{\sum_{i=1}^{k}(\sum_{j=1}^{n}x_{ij}^2 - \bar{y}_{ij}^2)}{n} - \frac{(\sum_{i=1}^{k}(\sum_{j=1}^{n}y_{ij})^2)}{kn} \tag{3}
\]

\[
SSE = SSA - SST \tag{4}
\]

The estimate of variance of alternatives and errors are calculated using:

\[
S2a = \frac{SSA}{k-1} \quad S2e = \frac{SSE}{n(k-1)} \tag{5}
\]

F value is computed using:

\[
F = \frac{S2a}{S2e} \tag{6}
\]

This calculated F value is compared with the value \(F_{1-\alpha; (k-1),(k(n-1))}\) obtained from the table of critical F values. Then we can say the variation due to alternatives is greater than variation due to errors with a confidence level of \((1-\alpha)\).

We will use this statistical analysis to prove that the variation of response time is significantly due to the change in alternative (changing the UI setting of Center stack) rather than the variation due the different users. This also gives a statistical estimate of the effect on the response time due to change in alternative and change in user.

### 2.3 Procedure

Before the start of the experiment, each volunteer was given few minutes to get acquainted with the simulator. After that, they were asked to drive two times. First time, they were asked to do secondary tasks like operating the center stack by giving voice commands with complicated grammar. The response time and errors of all volunteers were recorded.

In the second time, they were asked to interact with center stack, during driving, by giving different set of simple commands (commands with simple grammar) and response time and errors were recorded again. This operation was repeated at three locations:

1. Very fast left when the cars would be coming from different direction.
2. When the cars suddenly change on the freeway.
3. On the left turn of the road when they encounter pedestrians crossing.

Figure 4 shows the response time for each volunteer in milliseconds when they operated on commands with simple and complex grammar.

Figure 5 shows the number of errors for each volunteer when they operated on commands with simple and complex grammar.
Table 1 shows the average and standard deviations in the response time of our experiment. Response time 1 is the response time for commands with complex grammar and Response time 2 is the response time for commands with simple grammar:

<table>
<thead>
<tr>
<th></th>
<th>Response time 1</th>
<th>Response time 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>9223.76</td>
<td>4229.6</td>
<td>4994.16</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>3372.03</td>
<td>1416.63</td>
<td>3428.51</td>
</tr>
</tbody>
</table>

### 2.4 Results

CI (Confidence Interval) is a combination of sample size and variability. In order to interpret the CI for population mean, we have to assume that samples were chosen randomly and independently from a population and that the values were distributed according to Gaussian distribution. When we accept this assumption, there is 95% chance that 95% of CI contains the true population.

We found the C1 and C2 for both response time and error. You can see the results in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response time</strong></td>
<td>4276.06</td>
<td>5712.25</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>0.254381577</td>
<td>0.67851756</td>
</tr>
</tbody>
</table>

Since C1 and C2 for response time are positive, we can say that the measurements obtained are statistically significant and the response time of alternative 2 (voice command with simple grammar) is less than alternative 1 (voice command with complex grammar), so with 95% of confidence we can say that using voice commands with simple grammar is better than using voice commands with complex grammar because they reduce the number of errors.

In this experiment we used ANOVA in order to analyse the collected data statistically. Table 3 shows the results of using ANOVA to statistical analysis of volunteer drivers’ response time in this experiment. According to this table SSA/SST=0.865 thus we can say 86.5% of total variation is due to difference between two alternatives options. And SSE/SST=0.134 so we can say that 13.5% of total variation is due to noise in the measurements of different users.

In this table F-computed is greater than F-tabulated so at 95% confidence we can say that the differences seen in the response time by changing the alternatives is statistically significant.

### Table 3

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SST</strong></td>
<td>11665942960</td>
</tr>
<tr>
<td><strong>SSA</strong></td>
<td>10094439615</td>
</tr>
<tr>
<td><strong>SSE</strong></td>
<td>1571503346</td>
</tr>
<tr>
<td><strong>SSA/SST</strong></td>
<td>0.865291357</td>
</tr>
<tr>
<td><strong>SSE/SST</strong></td>
<td>0.134708643</td>
</tr>
<tr>
<td><strong>F-computed</strong></td>
<td>1143.3703</td>
</tr>
<tr>
<td><strong>F-tabulated</strong></td>
<td>3.087</td>
</tr>
</tbody>
</table>

Table 4 shows the results of using ANOVA to statistical analysis of number of errors of drivers in this experiment. According to this table SSA/SST=0.397 thus we can say 39.7% of total variation is due to difference between two alternatives options. And SSE/SST=0.602 so we can say that 60.02% of total variation is due to noise in the measurements of different users. Since F-computed is greater than F-tabulated, at 95% confidence we can say that the differences seen in the response time by changing the alternatives is statistically significant.

According these statistical analyses we can say that using voice commands with simple grammar decreases the cognitive distraction since it reduce the number of errors also it reduces the response time. Commands with less response time cause less cognitive distraction because the driver spends less time for secondary task during driving. In addition in this experiment the number of errors of drivers reduces when they use voice commands with simple grammar. It shows that they could focus better on their primary task (driving) when they use voice commands with simple grammar so they made fewer errors.

Since C1 and C2 for error are positive, we can say that the measurements obtained are statistically significant and the response time of alternative 2 (voice command with simple grammar) is less than alternative 1 (voice command with complex grammar), so with 95% of confidence we can say that using voice commands with simple grammar is better than using voice commands with complex grammar because they reduce the number of errors.

Similarly For the errors that was measured in this experiment according Table 2 since c1 and c2 for errors are positive, we can say that the measurements obtained are statistically significant and the number of errors with alternative 2 (voice commands with simple grammar) is less than alternative 1 (voice commands with complex grammar), so with 95% of confidence we can say that using voice commands with simple grammar is better than using voice commands with complex grammar because they reduce the number of errors.
Table 4

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SST</td>
<td>456.994506</td>
</tr>
<tr>
<td>SSA</td>
<td>181.4722531</td>
</tr>
<tr>
<td>SSE</td>
<td>275.5222529</td>
</tr>
<tr>
<td>SSA/SST</td>
<td>0.39709942</td>
</tr>
<tr>
<td>SSE/SST</td>
<td>0.60290058</td>
</tr>
<tr>
<td>F-computed</td>
<td>117.2393907</td>
</tr>
<tr>
<td>F-tabulated</td>
<td>3.087</td>
</tr>
</tbody>
</table>

4 References

[14] www.speechtechmag.com, March 5th

3 Conclusions

Using voice commands in car user interface can reduce visual and manual distraction but using this kind of commands causes other kinds of distraction. It causes cognitive distraction and audio distraction. Misdiagnosed voice commands cause cognitive distraction because the user should repeat a voice command several times and sometimes the driver should divert to another kinds of command. Sometimes the noise of environment causes misdiagnosed. By considering this fact that speech recognition systems are not ideal even with the best technology in order to improve the quality of voice based commands and decreasing the response time of driver and decreasing the number of driver’s error we should use voice commands with simple grammar because this kinds of commands are easy to remember and easy to pronounce so the response time decreases if we use the voice commands with simple grammar. In this paper we talk about the results of our experiment that compared the response time and number of errors in two scenarios. In the first scenario we used commands with complex grammar and in the second scenario we used commands with simple grammar. We observed that the response time reduced by half with simplified grammar. Also the errors reduced by 1/3rd when we used simple grammar for voice based commands. Driver was taking more attempts to get the command right when using complex grammar and he was getting the command right at the first attempt when commands had simple grammar. So we can say that using commands with simplified grammar decreases driver distraction when interacting through speech commands in car.