

Improving Driver Safety through Naturalistic Data Collection and Analysis Methods

2010-01-2333

Published
10/19/2010

Zachary Doerzaph, Thomas A. Dingus and Jon Hankey
Virginia Tech.

Copyright © 2010 SAE International

ABSTRACT

The design of a safe transportation system requires numerous design decisions that should be based on data acquired by rigorous scientific method. Naturalistic data collection and analysis methods are a relatively new addition to the engineer's toolbox. The naturalistic method is based on unobtrusively monitoring driver and vehicle performance under normal, everyday, driving conditions; generally for extended collection periods. The method generates a wealth of data that is particularly well-suited for identifying the underlying causes of safety deficiencies. Furthermore, the method also provides robust data for the design and evaluation of safety enhancement systems through field studies. Recently the instrumentation required to do this type of study has become much more cost effective allowing larger numbers of vehicles to be instrumented at a fraction of the cost.

This paper will first provide an overview of the naturalistic method including comparisons to other available methods. The focus of the paper then shifts to review the evolution of the data acquisition systems (DAS) and methods that have enabled naturalistic data collection. The goal is to provide readers with an understanding of how technology and unique partnerships has allowed the naturalistic data collection method to mature.

INTRODUCTION

Although driving becomes nearly second nature to the seasoned individual, in reality it is the most complex task most people perform on a daily basis. Drivers execute a continuous stream of decisions to navigate and control their vehicle based on sensory inputs provided by their surroundings. This interaction between the driver, their vehicle, roadway, adjacent vehicles, and the environment

results in an incessant flow of potential hazards. Although drivers typically recognize and avoid these hazards, crashes are unacceptably prevalent; evidenced by the approximately 2.4 million injuries and 37 thousand fatalities reported each year [1]. The research community is focused on improving driver safety through investigation of the root causes of crashes and the development of strategies to reduce crash prevalence and mitigate the resulting injuries.

Naturalistic data collection has emerged over the last decade as a highly effective technique for gathering the information necessary to improve driver safety. The technique is based on unobtrusively monitoring drivers under normal, everyday, driving conditions. Typically, a driver's personal vehicle is instrumented with a data acquisition system (DAS) and no special instructions are provided. Drivers simply go about their daily routine while the DAS silently collects information about the driver's interaction with the vehicle, roadway, adjacent vehicles, and environment. The collected information is subsequently analyzed to identify the root causes of crashes and near crashes as well as a number of other uses such as virtual testing of countermeasures.

Although instrumented vehicle studies and field operational tests on live roadways have been conducted for many years, the first known application of naturalistic data collection methods in surface transportation was the "100 Car study" performed by Dingus et al., in 2002 [2]. This study collected 42,000 hours of continuous video and electronic data for 241 drivers over a 13 month period. This study was the first instrumented vehicle study to: 1) capture a meaningful number of crash, near crash and minor collision events, and 2) it was the first study that provided the opportunity to "mine data" to answer a large number of safety-related questions related to driver behavior and performance. Naturalistic data collection relies on technology to replace the in-vehicle experimenter. Although this was initially difficult

and expensive, particularly for larger-scale studies, modern digital embedded technologies are revolutionizing the method by providing higher performance data collection at a fraction of the cost.

Today, naturalistic data collection is a mainstream technique with ongoing data collection efforts spanning the globe [3]. Within the United States, the Academy of Sciences in conjunction with the Transportation Research Board is embarking on the largest naturalistic data collection effort to date in the Strategic Highway Research Program 2 (SHARP2) Naturalistic Driving Field Study (SHARP2 NDFS, a.k.a “2000 car study”) [4]. This ambitious project relies on partnerships with research organizations such as the Virginia Tech Transportation Institute who is developing the data acquisition systems and coordinating the data collection effort. Other partners will manage individual data collection sites located from Seattle to North Carolina. Functioning as a team, this group is creating a national laboratory housing an unparalleled volume of driving performance data. Analysis of this data will allow researchers to perform a virtually limitless number of investigations to understand crash causation and driver behavior as it occurs on live roads with all the physical and cognitive factors present (i.e. Naturalistic).

In addition to pure naturalistic study, modern field studies often employ a pseudo-naturalistic strategy for collecting data. These field studies employ the naturalistic method to an extent; with the exception of a treatment which is being assessed and possibly additional controls such as asking a participant to drive in a particular area. For example, a pilot field study was recently performed for a cooperative intersection collision avoidance system for violations (CICAS-V) [5]. During this study participants were asked to drive a vehicle that was equipped with intersection collision avoidance technologies; however, there was no experimenter present in the vehicle. Although the driver's interaction with the technology was the subject of interest, the basic techniques employed during the data collection mirror that of a true naturalistic study.

As the prevalence of naturalistic data collection and the availability of naturalistic databases continue to grow, it is important for researchers and practitioners to be aware of naturalistic methods. The purpose of this paper is to provide an overview of naturalistic data collection emphasizing the technology advancements that have brought the collection method into mainstream science.

DATA COLLECTION METHODS COMPARED

Naturalistic data collection lies in the middle of the continuum of data collection methods (Figure 1). At one end of the spectrum is epidemiological data collection which is

based on mining crash databases such as the Fatality Accident Reporting System (FARS) and the General Estimates System (GES). At the other end of the spectrum are empirical data collection methods such as simulator and test track studies. These three complementary data collection methods each have advantages; researchers should be careful to select the appropriate method(s).

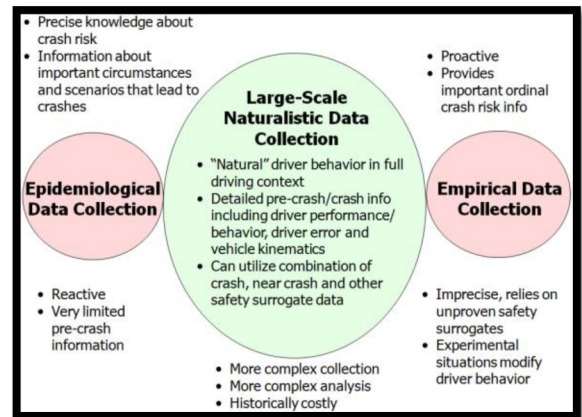


Figure 1. The primary methods for collecting driver performance data

Epidemiological data collection relies on actuarial data which provides very precise knowledge about crash risk. This precision is a direct result of measuring actual crashes after they have occurred rather than relying on a crash surrogates. Unfortunately, epidemiological data is reactive since detection of a safety deficiency is not possible until after a statistically significant number of crashes have occurred; potentially years after the deficiency manifests. Furthermore, crash databases typically have limited pre-crash information.

Most databases rely on data gathered from scene investigations (typically police reports) and/or self-report. The primary role of a policeman arriving at a scene is not filling out a police report. Rather, the police ensure public safety by providing traffic control and clearing the scene. The scene has likely changed and vehicles are frequently moved. It is difficult to obtain accurate assessments of pre-crash circumstances as drivers and passengers may be deceased, severely injured, or likely dazed. Furthermore, pre-crash events frequently occur so rapidly that key elements are forgotten or left out by witnesses. Finally, drivers and passengers may withhold information to avoid prosecution or embarrassment. These factors lead to errors in the database which can lead to false conclusions.

At the other side of the spectrum lies empirical data collection. Empirical data collection allows researchers to tightly control the research environment. They can hold a number of potentially confounding factors static while an independent measure is manipulated. This provides ordinal

crash risk information for the factors that were manipulated. It also allows proactive investigation of the treatment options such that risk can be assessed before the treatment is deployed into the real-world.

When performing an empirical study one must select a set of safety surrogate measures. Example measures may include lane keeping performance and average glance time. These surrogates are not a direct measure of crash risk which limits the ability to make conclusions regarding the safety impact in the population. Furthermore, the experimental scenario can alter the driver's behavior. For example, the driver in a simulator or test-track study will not have the same underlying motivation for driving and may purposefully modify their behavior to meet the perceived expectations of the experimenters.

Nestled in the middle, naturalistic data collection has the advantage of measuring 'natural' behaviors in the full driving context. With synchronized video, the information density obtained far exceeds that of any other data collection method. The time-history data provides the ability to observe numerous pre-crash factors that cannot be captured in a crash database. The resolution of collected data also permits thorough examination of the seconds leading up to a crash such that detailed reconstructions are possible. This allows researchers to fully characterize factors leading to a crash, including those that were not identified a priori.

The collection of naturalistic data can be somewhat more complex than empirical data collection. The DAS must be able to operate for an extended duration with minimal maintenance and participant logistics must be considered throughout the study duration. When performing analysis, the researcher must be careful to account for potential confounds that could be been controlled in an empirical study. Finally, large naturalistic studies have traditionally been costly. This limitation is alleviated by the latest generation of low-cost embedded DAS coupled with improved collection methods and standardized procedures.

It should be noted that many efforts will not fit cleanly into one category. Field studies are the most common example in which most of the naturalistic methods are used with the exception of a treatment that is introduced. Field studies are not purely naturalistic, but rather blend the empirical and naturalistic methods. Such data collection efforts provide excellent data for determining the merits of a safety improvement strategy with all the complexities of the open-roadway prior to deployment.

THE EVOLUTION OF NATURALISTIC DATA COLLECTION EQUIPMENT

Naturalistic data collection relies on technology to enable unobtrusive monitoring of the driver and environment. Only

basic naturalistic studies are possible without technology (e.g. counting seatbelt usage from outside the vehicle). In the 80s, it became possible to instrument vehicles using digital computers and video cameras (Figure 1). However, these vehicles were very expensive, time consuming to instrument, and they were far from unobtrusive. The figure is from the first instrumented vehicle evaluation of a moving map navigation system in 1984 [6]. The instrumented vehicle in this study employed two full-size video cameras mounted outside of the vehicle which captured the drivers face and the forward scene. A standard IBM PC and custom analog to digital conversion box was used to collect a variety of performance data including steering, accelerator and brake position.



Figure 2. Image of the first vehicle involved in an early data collection project

In the late 90's technologies for collecting naturalistic data were refined. An example study is the naturalistic lane change study which provided participants with a specially equipped vehicle to drive as their personal vehicle for an two week duration [7]. The technologies put into these vehicles reduced the camera size significantly by using remote video cassette recorders (VCR) to capture the images. For the first time, video collection was choreographed by a computer that automatically started when the ignition was switched on and continuously collected data until the vehicle was shut off. Furthermore, this video was also loosely synchronized with a digital recording of sensor data including measures such as turn signal activation and lateral acceleration. This data collection system was considered very compact for the time despite filling a significant volume of the vehicle trunk (Figure 3).

Similar hardware configurations employing a combination of VCRs, laptops, and sensor arrays were used for a number of years. One of the most advanced systems of this early generation DAS was constructed for a study on drivers of long-haul trucks [8]. These trucks had to be capable of collecting data for approximately one week of data during cross-country trips without any maintenance. To enable a data collection of this duration, first data was only retained for an interval of time surrounding a triggered event (e.g.