Belt Up: Investigating the Impact of In-Vehicular Conversation on Driving Performance

Amardeep Sathyanarayana, Navid Shokouhi, Seyed Omid Sadjadi, John H. L. Hansen

Abstract—In-vehicle conversations are typical when there is more than one person in the car. Although many conversations are beneficial in keeping the driver alert and active, there are also instances where a competitive conversation may adversely influence driving performance. Identifying such scenarios can improve vehicle safety systems by fusing the knowledge obtained from conversational speech analysis and vehicle dynamic signals. In this study we incorporate the use of smart portable devices to create a unified platform for recording in-vehicular conversations as well as the vehicle dynamic signals required to evaluate the driving performance. Results show that turn taking rate and overlapping speech segments under certain conditions correlate with deviations from normal driving patterns. The conversational speech analysis can thus be utilized as a component in driver assistance systems such that the impact of in-vehicle speech activity on driving performance is controlled or minimized.

I. INTRODUCTION

Driving is inherently a complicated task that requires dedicated attention towards many details both inside and outside the vehicular environment. With the rapid growth of in-vehicle infotainment systems, drivers spend more time performing various secondary non-driving related tasks such as messaging, checking emails, speaking over phone, listening to music, etc. Getting involved with these secondary tasks requires drivers to share their physical, auditory, visual, and cognitive abilities while driving. This in turn can lead to lack of sufficient attention towards driving and the road which may results in accidents/crashes of varying intensity. The 100-car Naturalistic Study reported that over 75% of crashes and 65% of near crashes were caused due to driver inattention [2]. In order to mitigate the distractions within the vehicular environment, new laws and guidelines are being enforced, particularly aiming at visual distractions [3].

Studying auditory based distractions, the work in [4] showed that although in-vehicle speech activity by itself might not influence driving performance, it imposes extra cognitive workload on drivers while performing secondary tasks. In this study, we further investigate the possible consequences of drivers’ involvement in competitive conversations while driving. For a complete definition of a competitive conversation see [5]. It is expected that active engagement in in-vehicular speech should result in deviations from usual driving patterns which could be identified through the analysis of different driving maneuvers.

There are many parameters such as road, weather, and traffic conditions that impact the driver’s decision making process while performing different maneuvers [6]. There are also evasive maneuvers which the driver performs during critical scenarios. Hence, understanding how different maneuvers are performed becomes essential in evaluating driving performance. Research in maneuver recognition has employed CAN-bus signals [7], [8] and obtained over 90% accuracy in the maneuver identification task. In addition, it has been shown that sensor information from portable devices outperform CAN-bus signals in maneuver recognition [9]. Similarly, in this study, we evaluate driving performance using sensor information from a smart portable device by analyzing maneuvers and computing deviations from the general trend [10].

In order to investigate the cause-and-effect relationship between in-vehicle conversations and driving performance, a series of experiments are conducted. In addition to detecting active speech regions, it is important to determine the amount of “focus” or “effort” the driver invests in participating in a conversation. The amount of competitiveness can be a measure to detect instances in which the driver is excessively engaged in a conversation. As an immediate result of studies in the field of Conversation Analysis, [5], we compute the occurrence and amount of overlapping speech segments and the frequency of turn takings in conversations as a measure to flag competitiveness.

The experiments in this study are designed with the goal to increase the drivers participation in conversations taken place in the vehicle. We show that speech segments with a high rate of overlap occur along “competitive” segments of a conversation in the sense that these segments of speech are shown to have a higher impact on the drivers performance compared to other speech active regions. We extend our previous work in [4] by adding a conversation analysis component in the audio analysis stage of the proposed driver assistance system. This should help improve the effectiveness of the active safety system to detect more accurately in-vehicular conversation and reduce its impact on driving performance.

II. SYSTEM DESCRIPTION

Previously [4], it was established that although in-vehicle speech activity by itself might not impact driving performance, it can add to the cognitive workload of the driver.
Performing certain secondary tasks such as controlling the navigation system, radio unit or even the AC in the car seem inevitable, there are, however, other secondary tasks such as engaging in a conversation with other passengers or over cellphone which could be easily controlled if they are determined to compromise driving performance.

This paper proposes a feedback system which not only evaluates variations in driving performance but also helps mitigate the influence of in-vehicle speech if it is found to be adversely influencing the driving performance (see Figure 1). There are two separate subsystems in the proposed framework. The first subsystem evaluates the driving performance by identifying maneuvers and then comparing them with their regular patterns. If a maneuver is recognized as not being performed “normally”, then the driver's in-vehicle speech involvement is monitored to see whether that is the cause for the unusual driving behavior. An improvisation to the work in [4] is that rather than using the proprietary CAN-bus and driver related signals, vehicle related signals are extracted and used from a portable device to analyze different maneuvers being executed by the driver.

The second subsystem evaluates the driver's involvement in in-vehicle conversations. An addition to the work in [4] is that rather than considering any in-vehicle speech activity, the emphasis is on the driver's involvement in the conversation. This involvement is analyzed by measuring the amount of overlapping speech segments as well as the turn-taking rate in a conversation. One of our objectives is to demonstrate that these two metrics are reliable indicators of the driver's focus on speaking as opposed to the primary task of driving.

If in-vehicle speech activity is identified to be adversely influencing driving performance, it can be controlled passively by providing an initial warning feedback to the driver and/or co-passengers to halt the conversation until the vehicle state returns back to normal. The feedback should be complementary to the modality which is causing the distraction. The extent to which this feedback may cause distraction deserves separate investigations. Monitoring the driving performance continues and any secondary task performed by the driver, other than those essential to driving, could be sequentially cut off. Although the current system focuses on in-vehicle speech, it can be extended to other modalities.

The uniqueness of this system is the ability to identify driving performance variations and isolate the source of distraction. In our case, if the driver's performance is not influenced by in-vehicle speech then other active/passive safety systems could be triggered until the driving performance is back to “normal”.

A. Driving Performance Evaluation

Maneuvers form the basic building blocks of a route, hence analyzing them can be employed as a key component in understanding driving performance. Variations in driving performance can be recorded by observing how each maneuver is executed and comparing the characteristics of maneuvers that fall under the same category.

In this study, 8 most commonly performed maneuvers - Right Turns (RTR), Left Turns (LTR), Right Lane Changes (RLC), Left Lane Changes (LLC), Right Road Curves (RRC), Left Road Curves (LRC), Straight driving (STR) and Stops (STP) - are considered. The execution time for each maneuver depends on the traffic, road, and weather conditions. Turns (left or right) generally take 5-9 seconds and are characterized by a decrease in speed, a significant increase in steering wheel movement which are eventually followed by an increase in speed. Turns are generally around intersections, but the traversed path does not necessarily form a 90 degree angle. Lane changes take around 3-5 seconds and are not restricted to a single shift to the adjacent lane. Stops generally take 5-15 seconds and depend mainly on the driver’s comfort and the traffic (for instance at intersections). Road curves and straight segments are road dependent and can stretch to long time segments starting from 15-30 seconds. These durations, however, serve only as typical measures and may vary drastically under different circumstances. The completion of any route consists of a combination of the aforementioned maneuvers. It is worth mentioning that evasive maneuvers are out of the scope of this study.

Maneuver recognition is performed by considering short window segments of 1 to 2 seconds and deciding whether the segment belongs to one of the 8 classes noted above, [10]. Statistical methods are adopted in maneuver recognition by clustering discriminative features from vehicle dynamics data such as vehicle speed and steering wheel angle. This approach has been previously adopted and proven to be effective for recognizing maneuvers with a high accuracy [9], [10]. The dominant features for different maneuvers are clustered in distinct regions in the feature space. This, between-class separation helps obtain a high accuracy in maneuver recognition. Once a maneuver or a small maneuver segment has been recognized to belong to a particular class (maneuver), further analysis is done to compute the within-class variability and evaluate the deviation. Detecting the outliers in each class is motivated by the variability observed.

![Fig. 1. System description](image-url)
in each class which is dictated by different levels of driving performance for each driver. The analysis of within-class variability is illustrated in Figure 2. Let us consider the 2 to be a representation of the feature space of an already recognized maneuver class X. It can be seen that all the green squares (majority of the samples in the class) are tightly clustered depicting the nominal pattern of maneuver execution for that class and driver. If the same maneuver X is performed differently, there will be a deviation in the features causing the points to move away from the cluster centroid. This is marked as the yellow and red squares. The stray outliers (red squares) are those maneuvers which are executed very differently, suggesting that the driver was inattentive or “distracted”. We evaluate the driving performance via identifying these outliers. This approach has provision not only in evaluating driving performance but also in identifying evolving driving trends. If there are many moderate variations (yellow squares) it might indicate that the maneuver cluster is growing or shifting slowly, hence the decision thresholds can be appropriately updated to include changing driving trend of the driver. In order to detect the outliers in each maneuver class, we use the Mahalanobis distance of each feature from the mean. Let \( \mathbf{F} = [f_0 f_1 \cdots f_{N-1}]^T \) represent a multivariate N-dimensional dominant feature vector from class X. Then, the Mahalanobis distance is calculated as,

\[
D_M(\mathbf{F}) = \sqrt{\left( \mathbf{F} - \mu_X \right)^T \Sigma_X^{-1} \left( \mathbf{F} - \mu_X \right)}
\]  

(1)

where \( \mu_X \) is the N-dimensional mean vector for class X, and \( \Sigma_X \) is the \( [N \times N] \) covariance matrix. The greater the Mahalanobis distance, the farther the feature vector is from the class centroid, which is an indication of deviation from “normal” driving pattern of the driver. Different threshold values can be set on the distance to detect the outliers. The per class distance in (1) represents the variation in driving performance as observed through the driver’s previous maneuver executions.

B. Conversation Analysis

As mentioned in Section II, the purpose of utilizing audio data is to analyze the conversation taken place in the vehicle and determine whether it is causing driver distraction. We speculate that the level of competitiveness in a conversation should correlate with the driving performance. In order to measure competitiveness in a conversation two features are utilized. The first is the turn taking rate in the conversation.

Fig. 2. Example of feature space for maneuver X

An increase in the number of turn takings in a given time window can imply focus of speakers on the argument. Additionally, the amount of overlapping speech in the same time window is considered as a potential feature in this study [5]. The conversation analysis setup is developed by incorporating the work accomplished in our previous studies [11], [4].

1) Turn Taking rate: A speech activity detection (SAD) system is used to detect start and end-points of active speech regions in a recorded (or buffered) conversation. The start-points are labeled as the triggering onset of a turn in the conversation. Since this definition for turn taking may not always imply that both speakers are involved in a conversation (e.g., instances when one of the speakers pauses in between sentences or words), the amount of overlapping speech in regions with a high turn taking rate is also employed to ensure that both speakers are involved.

2) Overlapping Speech rate: The overlapped-speech detection system, in addition to the baseline cepstral features (e.g., mel-frequency cepstral coefficients -MFCC ), incorporates features that emphasize on the harmonicity of speech. These features include spectral flatness measure (SFM), average Aperiodicity measure [12], and Kurtosis [13]. Overlapped-speech is referred to segments in the audio file during which more than one person is speaking. The task of designing an overlapped-speech detection system for in-vehicular environments has been investigated by the authors in a recent study [11]. The system is developed by training background models for single-speaker and double-speaker speech and adapting the models to the data collected in the vehicular environment where several noise sources are active at the same time (e.g., engine noise, turn signals, AC). The double-speaker model is trained on artificially created data by summing individual speech segments from different speakers with 0 dB average target-to-interference ratios. The maximum \textit{a posteriori} (MAP) adaptation technique is used to
adapt the mixture means in both models to the test data [14].
The adaptation data used for the single speaker is extracted
from a small portion of the UTDrive test data using the SAD
system described in [15]. On the other hand, the adaptation
data for the double-speaker model is artificially created by
adding the active speech segments extracted from separate
segments in the audio file. Prior to these steps the noisy
data is enhanced using the optimally modified log-spectral
amplitude (OM-LSA) algorithm [16]. Figure 3 illustrates
the block diagram of the overlapped-speech detection system.

It is worth remarking that the presence of overlapping
speech by itself may not always imply a highly competitive
conversation. For instance, while back-channels (e.g. uh-huh)
are common overlapping speech occurrences, they do not
necessarily imply competitiveness. Hence, the combination
of a high turn taking rate and overlapping speech is used as
the measure of competitive behavior in a conversation.

III. DATA DESCRIPTION

There are various factors that affect the visual, auditory,
physical, and cognitive abilities of a driver which may in turn
adversely influence his/her driving performance. Since this
study focuses on understanding the influence of in-vehicle
speech on the driver, effort has been made to minimize
the impact from other modalities on the driver. Experiments
are conducted with drivers operating the vehicle under real-
traffic conditions, and the data is collected under similar
weather and traffic conditions for all drivers. The route
consists of residential areas as well as highways and takes
on average twenty minutes per session to complete.

Recently, there has been a keen interest in utilizing
portable devices (sensor loaded smartphones and Tablets) to
instrument a vehicle and use it as a pseudo-data collection
platform [9], [10]. It has also been verified that employing
sensor information from portable devices can provide maneu-
ver recognition performance comparable with that obtained
using CAN-bus signals from expensive instrumented vehicles
[9], [10]. In this study, the data collection is performed using
the UTDrive instrumented vehicle (UTDrive) along with the
portable device mounted on the windshield [9] (see Figure 4).
However, only the sensor information from the portable
device are extracted and utilized in our analysis. Using
Samsung Galaxy Tablet 10.1” with as the portable device
platform, an android application (app) has been developed
to collect all the available sensor information on the device
synchronously. The available sensors and derived information
include camera, microphone, accelerometer, gyroscope,
magnetometer, orientation, compass, and GPS signals. A
detailed description of the UTDrive app for portable device
and the sensor information can be found in [9].

The data collected for the purpose of this study is designed
with the intention to induce a competitive behavior in the
driver towards the conversations. The tasks which the drivers
have been asked to perform are expected to activate their
competitiveness and involvement in order to increase the
amount of overlapped-speech segments in the conversations
[5]. However, the fact that the purpose of the study is to
collect competitive conversations is kept hidden from the
drivers to avoid self-consciousness. The driving route is
divided into four segments that are repeated in two phases.
In the first phase the driver drives through the complete route
without performing any secondary task to become familiar
with the route and the vehicle. In the second phase the driver
is asked to perform a different task in each segment of the
route. Four different tasks are chosen to include as many
variations in conversational speech as possible. The tasks
are described below:

- Segment 1: At the beginning, in order to “break the
  ice” the passengers initiate a simple conversation by
  asking the driver questions about casual topics such as
  the weather.
- Segment 2: In this segment one of the passengers
  chooses an object (in his mind) and the driver and the
  other passenger are asked to guess what the object is
  using hints provided to them. This game is chosen to
  increase turn-taking and overlapping speech segments.
- Segment 3: A set of TIMIT sentences are played

Fig. 4. UTDrive portable android application user interface.

Fig. 5. Driving route with different conversational task segments.
through a portable tablet and the driver is required to repeat each sentence before the next sentence is played.

- Segment 4: An argument is initiated by one of the passengers. This second conversation requires more involvement and attention compared to that in Segment 1. The difference here is that the driver’s opinion on a debatable topic is asked and based on his/her response, the passengers take the opposite side and attempt to argue on the subject.

Each of the tasks described above takes place on one of the legs in the route as labeled on the map in Figure 5.

IV. EXPERIMENTAL RESULTS

In [4], CAN-bus signals such as vehicle speed, steering wheel angle, gas and brake pedal pressure were used in developing maneuver models. Using these models, the effect of in-vehicle speech on the driver’s performance was investigated. Moreover, in a previous study [9, 10] it was shown that the use of sensor information extracted from low cost portable devices (compared to the available CAN-bus signals) could yield more accurate maneuver recognition. Therefore, in this study, the analysis of sensor data from portable device is extended to compute a driving performance measure and perform an in-depth analysis of in-vehicle speech to investigate the influence of overlapped speech and turn-taking on driving performance. A major advantage of this experimental setup is that both the in-vehicle speech data and the data required for maneuver recognition are recorded by the portable device which results in a more concise and cost effective data acquisition platform.

Statistical information from the inertial measurement sensors such as accelerometer and gyroscope of the portable device are extracted on a per frame basis (once per second). Statistical information such as maximum lateral acceleration, mean of the vehicle speed, variance of yaw-gyroscope (refer to [9] for a detailed list) form a dominant feature set which is used in training the maneuver specific models. Using support vector machines (SVM), the maneuver segments are classified with a high average accuracy of over 90%.

Once classified, the maneuver’s within-class distance from the class centroid is measured using the Mahalanobis metric as discussed earlier. Next, the thresholds are appropriately set to identify any abnormal or risky driving patterns.

Examples of the driving performance evaluation is shown in Figure 6. This figure represents the same segment of the route driven by the same driver in two phases of the data collection. The regions marked as “green” denote locations where the driving maneuver is similar to its usual pattern, whereas regions marked as “yellow” show segments where the driver exhibits relatively higher variations in the driving pattern. The “red” regions mark locations where the driver performs an abnormal or risky maneuver. All these regions concur with the visual and perceptual verification obtained by looking at the video from the data recordings.

As Figure 6 suggests, there are some instances in the route where no direct relationship is observed between the occurrence of overlapping speech and drop in driving performance (see the area marked by a star in the figure). This aligns well with our intuition since there are always other sources that can cause irregularities in the driver’s maneuver execution patterns. Hence, speech related features should be analyzed in more detail to confirm that the conversation is the source of distraction. With this in mind, the patterns of turn-takings and overlapping speech segments are jointly investigated over time. We define turn-taking rates and overlapped-speech rates as follows:

\[
\text{ovlrate} = \frac{\text{Number of overlapped samples in window}}{\text{window length}}
\]

\[
\text{tt\text{rate}} = \frac{\text{Number of turn-takings in window}}{\text{window length}}
\]

Figure 7 depicts the average turn-taking rate and overlapped-speech rate in the conversations before the first instance of observing a major drop in driving performance for four different occasions. Each plot belongs to a different scenario. If both the turn-taking rate and overlapped-speech rate are increasing before the performance-drop moment, the conversation can be a potential source of distraction. Plot
performance-drop. This is expected since overlapped-speech likelihood of the conversation being the source of distraction. Panels (a) and (b) in the figure represent the scenario where both the features are increasing with time. A similar pattern to that seen in panel (c), however, is also observed in our experiments where the overlapped-speech rate increases and eventually rolls off towards zero before the onset of performance-drop. This is expected since overlapped-speech generally occurs less than turn-takings.

V. CONCLUSIONS

This study investigated the impact of in-vehicle conversations on driving performance. A data acquisition platform employing an off-the-shelf portable tablet was developed and used to record audio/video as well as the vehicle dynamic signals. We incorporated certain features such as the number of turn-takings and overlapping speech segments to analyze the competitiveness of conversations. A good correlation was found between the turn-taking rate as well as overlapped speech and the drop in driving performance which was detected through maneuver analysis. The proposed component can thus enable active safety systems to isolate the competitive conversation and alert the driver and passengers to “belt up”.

REFERENCES