Applied Soft Computing xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# Automatic linguistic reporting in driving simulation environments

Luka Eciolaza [a,*], Martín Pereira-Fariña [b], Gracian Trivino [a]

[a] European Centre for Soft Computing, Gonzalo Gutierrez Quiros s/n, 33600 Mieres, Asturias, Spain
[b] Centro de Investigación en Tecnoloxías da Información (CITIUS), Universidade de Santiago de Compostela, Galicia, Spain

ARTICLE INFO

*Article history:*
Received 15 December 2011
Received in revised form 5 July 2012
Accepted 3 September 2012
Available online xxx

*Keywords:*
Computational Theory of Perceptions
Linguistic summarization of data
Driving simulators

ABSTRACT

Linguistic data summarization targets the description of patterns emerging in data by means of linguistic expressions. Just as human beings do, computers can use natural language to represent and fuse heterogeneous data in a multi criteria decision making environment. Linguistic data description is particularly well suited for applications in which there is a necessity of understanding data at different levels of expertise or human–computer interaction is involved. In this paper, an application for the linguistic descriptions of driving activity in a simulation environment has been developed. In order to ensure safe driving practices, all new onboard devices in transportation systems need to be evaluated. Work performed in this application paper will be used for the automatic evaluation of onboard devices. Based on Fuzzy Logic, and as a contribution to Computational Theory of Perceptions, the proposed solution is part of our research on granular linguistic models of phenomena. The application generates a set of valid sentences describing the quality of driving. Then a relevancy analysis is performed in order to compile the most representative and suitable statements in a final report. Real time-series data from a vehicle simulator have been used to evaluate the performance of the presented application in the framework of a real project.


## 1. Introduction

Linguistic summarization represents an attempt to describe by means of linguistic expressions patterns emerging in data. It is intended in general for applications in which there is a strong human–machine interaction involving accessing and understanding data, like supervision and control processes. Linguistic summarization is a data mining or knowledge discovery approach for providing efficient and human friendly framework for the analysis of data within a decision support environment.

There is a growing necessity for computational systems capable of providing linguistic descriptions of phenomena. New technologies allow acquiring and archiving vast volumes of data about time-evolving phenomena and the amount of information collected in different fields is overwhelming and ever growing. However, there is a lack of tools and means for processing and interpreting all this information using computers. In order to be useful, this information must be explained in an understandable way, including facts that may be derived from the data and the background knowledge available about the phenomenon under study. Human specialists describe their perceptions using natural language (NL). NL allows experts to make imprecise representations that summarize their perceptions of complex phenomena choosing the most adequate degree of granularity in each circumstance to remark the relevant aspects and to hide the irrelevant ones.

In our opinion, data should not be simply made accessible or summarized as graphics, tables and simple linguistic variables, but the needed interpretation arguments and conclusions should be provided and explained using NL. Nowadays, any organization of data as provided by a computer, either in a numerical, categorical, and/or graphical form, is just a tool that can be employed by human experts to produce an explanation in NL. Understandable linguistic descriptions of phenomena are provided by human experts, computers just being a tool for storing and accessing data in a very flexible way. This is clearly a problem as the ratio data/human experts is growing dramatically, since it is becoming easier and easier to collect data, but providing a human being with expertise on a certain area remains difficult and expensive. On the other hand, assessment by observation does not meet the validity and reliability criteria necessary for any objective evaluation. Generated expert reports should be based on unified criteria and independent to the subjective variability of different experts [1,2].

In summary, the formulation of data summaries can be seen as a very complex and non-trivial data mining task, and there is a clear need for computational systems able to produce automatic linguistic descriptions of data about phenomena.

In this paper, we present our approach to implement a computational system capable of providing linguistic descriptions of driving activity in simulation environments. The growth of vehicle onboard

* Corresponding author. Tel.: +34 985456545; fax: +34 985456699.
*E-mail addresses:* luka.eciolaza@softcomputing.es (L. Eciolaza), gracian.trivino@softcomputing.es (G. Trivino).

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2012.09.007

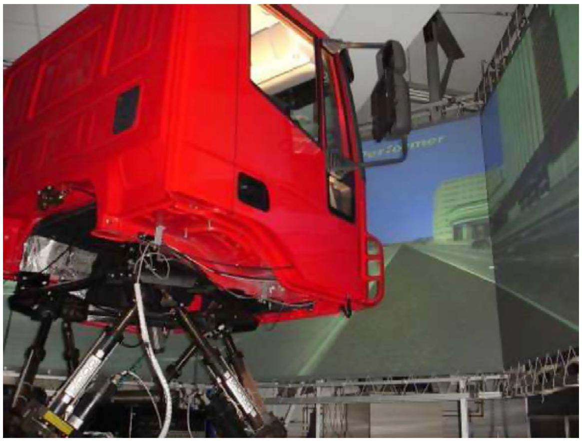

**Fig. 1.** CABINTEC: the simulator for intelligent transportation systems.

devices has increased dramatically driver's attention to secondary tasks other than driving. There are many studies [3–5] analyzing this problem that demonstrate how the integration of onboard systems can drastically increase the risk of committing mistakes on driving, potentially resulting on traffic accidents. Therefore, there is a clear need to evaluate the onboard systems in order to ensure safe driving practices.

Driving simulators (see Fig. 1) are interesting tools to perform this evaluation. Simulators can provide analysts with various information related to the driver–device interaction in a variety of situations. Human factor experts are able to assess the performance and effectiveness of the onboard devices through a variety of predefined experiments and obtain a vast amount of data. Unluckily, handling this huge amount of information is not easy and to perform a series of repetitive experiments can be very tedious.

Our research line belongs to the field of Fuzzy Logic, where since several years ago the scientific community deals with the problem of obtaining automatic linguistic summaries of all sorts of data [6–9].

Our approach is based on Computational Theory of Perceptions (CTP) [10,11]. CTP provides a framework to develop computational systems with the capacity of computing with the meaning of NL expressions, i.e., with the capacity of computing with imprecise descriptions of phenomena in a similar way humans do it. In this paper we develop on the concept of Granular Linguistic Model of a Phenomenon (GLMP) [12,13]. Our long term project deals with generating more complex and useful linguistic descriptions than the obtained using the existing technology.

State-of-the-art automatic linguistic description techniques in modern training systems based on Fuzzy Logic [2], consider expert knowledge recorded as simple linguistic variables. In our approach, we generate complete linguistic summary reports through the aggregation of all the information related to the analyzed phenomena. On the other hand, the works in the area of Natural Language Generation systems [14] contemplate the generation of complete summary reports, however, they do not consider the quantification of validity degrees for generated sentences and therefore it is difficult in this context to deal with sensor input data and uncertain information. Moreover, at the current state of development within the field of linguistic summarization of data, there is an important lack of methods for quality assessment.

The work presented in this article builds upon previous works presented in [12,13]. This work represents the last step of a series of prototypes of increasing complexity. Each prototype followed a logic step where different problematic have been addressed in order to get to the final goal of automating the onboard device evaluation process. Thus, in [12] the quality of driving during selected simulation segments was summarized into a set of fixed sentences that were later on compared with questionnaires filled by experts. In the prototype presented in [13] the generated reports were sequential descriptions of events within dynamic or time-evolving data, and there was an initial stage of validity analysis performed by the GLMP followed by a relevancy analysis to select the appropriate sentences for the final report.

The application presented here is characterized as follows; (a) it is focused on the detection of risk events during the driving activity; (b) it compares intervals of interaction with no-interaction between driver and onboard systems. Finally (c) it generates a linguistic report describing the overall quality of driving and it determines the distraction potential of analyzed onboard systems. The report includes hypothesis about the relation between the detected risk events and the distraction levels, all based on the input/output signals of the simulator.

Once calibrated, the automatically generated document should complement or replace the evaluation reports generated by human experts.

In this paper, we present the complete linguistic model and report template that allow describing the interaction between the driver and the onboard devices in a vehicle. We measure the potential distraction levels following the requirements defined within the HITO project. For this, we introduce the concept of the *Index of Potential Distraction*. For the quality assessment of the generated text, we explore a new approach that seeks to evaluate reports' usability from the point of view of human consumption, identifying different indicators that define the quality of a linguistic report.

Section 2 provides a brief review of the state of the art in the different subjects involved. Section 3 introduces the concept of Granular Linguistic Model of a Phenomenon. The definitions related to the various components are included, and the process of sentences generation with their respective validity degree are described. Section 4 describes the report generation task through the compilation of the most relevant sentences. Section 5 explains in detail the different parameters involved in the application, it shows the practical implementation of the GLMP for the report generation and a definition for the computation of the Index of Potential Distraction is proposed. Sections 6 and 7 describe the experimentation and validation of the obtained reports respectively. Finally, Section 8 contains the conclusions.

## 2. Background information

This section presents a small review of the various domains involved in this work. Information about the state of the art practices within these domains will help the reader understand better the extent and potential of the paper.

### 2.1. Intelligent transportation systems and HITO

Road transportation is changing rapidly due to technological evolution. The field of intelligent transportation systems (ITS) is focused on the minimization of traffic congestions, optimization of traffic management, and the road security improvement in general.

The increasing capability to acquire, archive and share huge amounts of heterogeneous information allow the development of a big number of vehicle onboard systems focused to improve the overall driving experience with applications for driving assistance, entertainment or information systems. Apart from traffic management and road security improvement applications, in the last decade, technologies to monitor and control drivers conditions, such as fatigue or stress showed an increased interest and there are numerous examples of these devices in the market [15].

Please cite this article in press as: L. Eciolaza, et al., Automatic linguistic reporting in driving simulation environments, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.09.007

As stated above, the growth of vehicle onboard devices implies the increased risk of distractions potentially resulting on mistakes while driving. Vehicle onboard systems must be designed with the priority of ensuring and meeting safety requirements, focusing on clear and intuitive interfaces. Therefore, there is a clear need to evaluate the onboard systems in order to ensure safe driving practices.

Driver distraction is an important issue of study both for research in investigating human multitasking abilities and for practical purposes in developing and constraining new onboard devices. See, for example, the research performed in [16,17] focused on modeling and prediction of driver distraction. The analysis of driving performance data like vehicle deviation from lane or speed control with respect to an accelerating and braking lead vehicle can be used to accurately model driver's distraction. They quantify how the vehicle's deviation from the lane center increases during period of inattention and how the vehicle returns to lane center during periods of active steering.

This paper is part of a multidisciplinary long term project with the aim of developing an "*Intelligent cabin for road transportation of goods and passengers*". Representative parameters of vehicle, environment and driver behavior are collected and evaluated in order to know, depending on the nature and criticality of the information, WHAT, HOW, WHEN and WHERE the onboard systems' information should be placed in the driving desk within the vehicles. The project aims to design a safe, usability oriented and scalable framework for the integration of new onboard devices.

In particular, HITO (Human Interface by Technology Observation) [18] aims to develop an effective methodology and framework to evaluate onboard devices used in road transport in order to optimize reliability levels and to guarantee an ergonomic and safe workplace design. For this, HITO proposed the development of a methodology to measure *potential distraction levels* caused by onboard systems. This methodology consists of a series of experiments or exercises on a simulation environment designed by experts. Initially, the exercises will be analyzed and reported manually by a team of human factor researchers for the evaluation of each device. Finally, the project will also address the essential task of automatization of the developed methodology through a software application.

Modern training systems in different areas are now incorporating automated evaluation functionality. Evaluations are based on expert knowledge and the used metrics are typically recorded as linguistic variables that can take values such as low, medium, high or other comparable terms. The work presented by Riojas et al. [2] is an interesting and updated reference of applying Fuzzy Logic in this field. In our approach, we are aiming to generate complete linguistic summary reports and not only simple linguistic variables describing the simulation activity.

### 2.2. Computational Theory of Perceptions

CTP was introduced in the Zadeh's seminal paper "From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions" [10] and further developed in subsequent papers [11,19,20]. CTP provides a framework to implement computational systems with the capacity of computing with the meaning of NL expressions, i.e. with the capacity of computing with imprecise descriptions of the world in a similar way that humans do it. According to CTP, our perception of world is granular. A granule underlies the concept of a linguistic variable [19]. A linguistic variable is a variable whose values are linguistic labels, i.e., words or sentences in NL [21]. In this approach, a fuzzy linguistic label can be viewed as a linguistic summary of numerical data, e.g., a set of temperature values are labeled as Medium. The definition of the linguistic label includes the concept of "*degree of validity*" to describe each element in the set [21].

### 2.3. Linguistic summarization of data

In our approach, linguistic summarization [6,8,22] is closely related to CTP and hence to *Granular Computing* paradigm. According to [23], information granules are conceptual entities which emerge from the needs of humans in a continuous quest to abstraction and summarization of information. Linguistic summarization seeks to process complex information and describe emerging patterns through linguistic expressions and the manipulation of information granules in the form of words. Information granulation in the form of NL is used to describe and understand complex phenomena at different levels of resolution or scales. Similar to human reasoning, NL representation is used for multi modal data fusion.

The idea of linguistic fuzzy quantifiers was introduced by Zadeh in [22]. The concept of linguistic fuzzy summary was introduced in [6] and further developed in [8]. A fuzzy linguistic summary is a set of sentences which express knowledge about a situation through the use of fuzzy linguistic summarizers and fuzzy linguistic quantifiers.

The basic concept of fuzzy linguistic summary has the general form of a quantified fuzzy proposition [6,22]: ($w$, "Q objects in database are S"); where Q is called the quantifier, S is the qualifier, also called summarizer, and $w$ is the degree of validity of the linguistic clause for representing the meaning in the specific context. For example: (0.7, "sometimes the driving quality has been low"). In recent years, this basic concept has been developed in different ways [24] and used for different applications, e.g., data mining [7], database query [8], and for describing temporal series [25]. See in [26] a review on the state of the research on this field.

During the last years, researchers in the field of computing with words and perceptions have developed an important set of resources to represent the meaning of perceptions for making decisions in specific applications [27,28]. The basic fuzzy linguistic summary covers a very small part of the possibilities of meaning in NL [29–31]. Each application will present particular challenges for the computation of the validity degrees of the different types of linguistic expressions.

## 3. Granular Linguistic Model of a Phenomenon (GLMP)

In this section, we introduce the components of the GLMP, our approach based on CTP for developing computational systems able to generate linguistic descriptions of data. In our research line, one of the contributions of this paper consists of identifying three types of computational perceptions (CP), namely, *assertive*, *derivative* and *integrative* that here, they are extensively used to linguistically model the evolution in time of phenomena.

### 3.1. Computational perception

A *CP* is the computational model of a unit of information acquired by the designer about the phenomenon to be modeled. In general, *CP*s correspond with specific parts of the phenomenon at certain degrees of granularity. A *CP* is a couple ($A$, $W$) where:

$A = (a_1, a_2, \ldots, a_n)$ is a vector of linguistic expressions (words or sentences in NL) that represents the whole linguistic domain of the *CP*. Each $a_i$ describes the value of the *CP* in each situation with specific degree of granularity. These sentences can be either simple, e.g., $a_i$ = *"The vehicle speed is high"* or more complex, e.g., $a_i$ = *"During interaction sometimes the manoeuvre execution has been bad."*.

Please cite this article in press as: L. Eciolaza, et al., Automatic linguistic reporting in driving simulation environments, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.09.007

$W=(w_1, w_2, \ldots, w_n)$ is a vector of validity degrees $w_i \in [0, 1]$ assigned to each $a_i$ in the specific context. $w_i$ is the degree in which $a_i$ is valid to describe a situation.

### 3.2. Perception mapping (PM)

We use *PMs* to create and aggregate *CPs*. There are many types of *PMs* and this paper explores several of them. A *PM* is a tuple ($U$, $y$, $g$, $T$) where:

$U$ is a vector of input *CPs*, $U=(u_1, u_2, \ldots, u_n)$, where $u_i = (A_{u_i}, W_{u_i})$. In the special case of first order Perception Mappings (1*PMs*), these are the inputs to the GLMP and they are values $z \in \mathbb{R}$ either provided by a physical sensor or obtained from a database.

$y$ is the output *CP*, $y=(A_y, W_y) = \{(a_1, w_1), (a_2, w_2), \ldots, (a_{n_y}, w_{n_y})\}$.

$g$ is an aggregation function employed to calculate the vector of validity degrees assigned to each element in $y$, $W_y = (w_1, w_2, \ldots, w_{n_y})$. It implements the aggregation of input vectors, $W_y = g(W_{u_1}, W_{u_2}, \ldots, W_{u_n})$, where $W_{u_i}$ are the degrees of validity of the input perceptions. In Fuzzy Logic many different types of aggregation functions have been developed. For example $g$ could be implemented using a set of fuzzy rules. In the case of 1*PMs*, $g$ is built using a set of membership functions as follows: $W_y = (\mu_{a_1}(z), \mu_{a_2}(z), \ldots, \mu_{a_{n_y}}(z)) = (w_1, w_2, \ldots, w_{n_y})$ where $W_y$ is the vector of degrees of validity assigned to each $a_i$, and $z$ is the input data.

$T$ is a text generation algorithm that allows generating the sentences in $A_y$. In simple cases, $T$ is a linguistic template, e.g., "*The temperature in the room is* {*high* | *medium* | *low*}".

### 3.3. Structure of the GLMP

The *GLMP* consists of a network of *PMs*. Each *PM* receives a set of input *CPs* and transmits upwards a *CP*. We say that each output *CP* is *explained* by the *PM* using a set of input *CPs*. In the network, each *CP* covers specific aspects of the phenomenon with certain degree of granularity. Fig. 2 shows an example of GLMP. In this example, at every point in time, the phenomenon can be described at a very basic level in terms of three variables $z_1$, $z_2$, and $z_3$. These variables are verbalized in through 1*PMs* $\{p_1^1, p_2^1, p_3^1\}$.

Using different aggregation functions and linguistic expressions, the paradigm *GLMP* allows the designer to model computationally his/her perceptions. In the case of Fig. 2, from the outputs of 1*PMs*, other two higher-level descriptions of the phenomenon are derived. These descriptions are given in the form of computational perceptions $CP_4$ and $CP_5$, which are explained by 2*PMs* $\{p_4^2, p_5^2\}$ in terms of $CP_1$, $CP_2$, and $CP_3$. The validity of each item in $CP_4$ and $CP_5$ is explained by those items of $CP_1$, $CP_2$ and $CP_3$. Finally, the top-order description of the phenomenon is provided, at the highest level of abstraction, by $CP_6$, explained by the 2*PM* $\{p_6^2\}$ in terms of $CP_4$ and $CP_5$. Notice that, by using this structure, one can provide a linguistic description of the phenomenon at different levels, from the very basic level to the highest or most general level of granularity.

### 3.4. Types of CP

For this application, as initially introduced in [13] and inspired in the classical Control Theory [32], we focus on the perception of three important characteristics of phenomena evolution, namely, the perception of the current state (*assertive CP*), the perception of the trend to evolve (*derivative CP*) and the summary of accumulated perceptions (*integrative CP*).

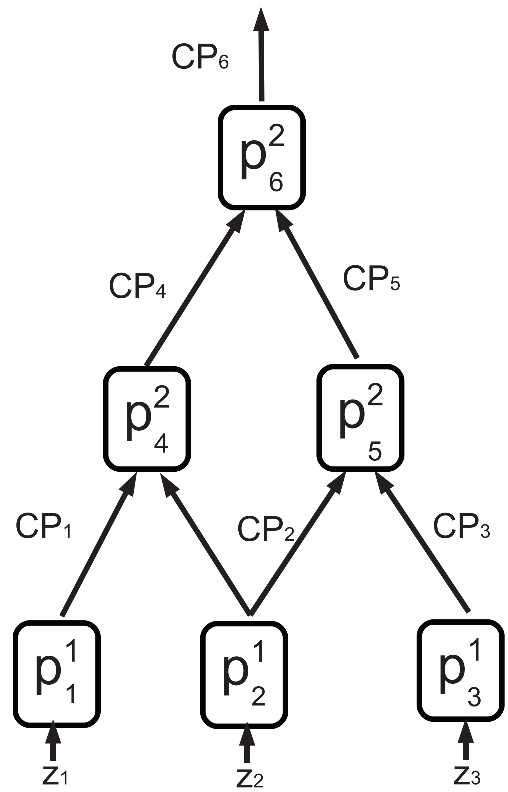


**Fig. 2.** Example of GLMP.

#### 3.4.1. Assertive CP

It is associated with a linguistic expression of type "*Y is A*". It represents the linguistic fuzzy model of the current state of a characteristic of the phenomenon, e.g., "*The Distance to the Vehicle in front is **High***".

#### 3.4.2. Derivative CP

They correspond with trend analysis information and they give insight into how the phenomenon is evolving in time. It helps contextualizing the information and it may be important for decision making.

The following example sentences clearly show the importance of the derivative information, and how it can completely change the context in which certain decision must be taken.

- "*The Distance to the Vehicle in front is* **Medium**".
- "*The Distance to the Vehicle in front is* **Medium** and **Increasing**".
- "*The Distance to the Vehicle in front is* **Medium** and **Rapidly Decreasing**".

In this work, we only use derivative 1*CPs*, i.e., they are directly obtained from the input signals of the different sensors.

In the derivative *PM*, $U$ is a time-series signal ($z = \{z(k-l+1), \ldots z(k-1), z(k)\}$) of length $l$, obtained directly from sensor input data, where $k$ represents the current sample and $l$ is defined by the designer, e.g., to filter the noise of the input signal. $z_d$ is the relative change between samples and it is calculated as follows:

$$z_d = 100 \times \frac{z(k) - z(k-l+1)}{\bar{z}}$$

where the average of the $z$ is

$$\bar{z} = \frac{1}{l}\sum_{i=1}^{l} z(k-i+1)$$

The relative change $z_d$ is directly used to describe the trend of the perceptions linguistically. Here, $T$ is defined by the template: "*The value of the attribute is* {*Rapidly Decreasing* | *Decreasing* | *Steady* | *Increasing* | *Rapidly Increasing*}".

*3.4.3. Integrative CP*

The *integrative CP* represents the accumulated perception of the phenomenon over a period of time. The text associated with these perceptions consists of summary sentences of historical event occurrences, and answers the question of "Which is usually the state of the Parameter", i.e., "Q of Ys are A".

The typical template for the answer could be:

{*Never* | *A few times* | *Sometimes* | *Many times* | *Most of the time* | *Always*}, *the parameter was* {*Low* | *Medium* | *High*}.

The accumulated perception may be very important for decision making. The following example sentences with assertive and integrative sentences show how the integrative information can completely change the context in which the decision must be taken.

- "*The Vehicle Linearity is* **Medium**".
- "*The Vehicle Linearity is* **Medium** and **Most of the time** it has been **High**".
- "*The Vehicle Linearity is* **Medium** and already **Sometimes** it has been **Medium**".

In case of driving quality assessment the last example sentence could show a distraction problem as it seems that *medium vehicle linearity* events are common and so there have been many slight distractions.

The definition of an *integrative PM* corresponds to the tuple ($U$, $y$, $g$, $T$). In this case, $U$ is an input *CP* over a time window. The designer sets the parameter $l$ defining the length of the time window from which the temporal series of $l$ samples is obtained. $U = \{u(k-l) \ldots u(k)\}$, where $k$ represents the current sample.

In the case of an attribute defined by three linguistic labels, e.g., {*Low*, *Medium*, *High*}. The output *CP* $y = (A_y, W_y)$ for the *integrative PM* will be expressed by eighteen possible sentences $\{(a_1, w_1), \ldots, (a_{18}, w_{18})\}$, combination of the six linguistic quantifiers $Q$ = {*Never, A few times, Sometimes, Several times, Most of the time, Always*} and input linguistic labels $A_i$ = {*Low*, *Medium*, *High*}.

$g$ is an aggregation function computed as a quantified sentence. There are many different approaches for evaluating quantified sentences. In this work we have used the $\alpha$-cut based method called GD introduced in [33] instead of the basic approach of quantified fuzzy propositions [6,22] where the weights $w$ associated with the linguistic expressions are computed as fuzzy cardinalities. The *GD* method has been used due to its efficiency and non-strict character. The method also fulfills some interesting properties related to relative quantifiers defined in [33].

## 4. Report generation

Using the *GLMP* defined in the previous section, we can generate a set of valid sentences describing the phenomenon in different levels of granularity or detail. The *GLMP* is built during a design stage where a corpus of NL expressions that are typically used in the application domain is collected. These expressions describe the relevant features of analyzed phenomena. The sentences describe each perception in every temporal sample and through quantified sentences the overall states of the perceptions will be described.

A medium size *GLMP* could generate a huge number of sentences describing a particular phenomenon. In case of a session of driving simulation analyzed in this work, the number of generated sentences can amount to hundreds of thousands for a normal simulation exercise. It is critical to do a relevancy analysis in order to select and compile the relevant sentences into one document highlighting the interesting characteristics of a simulation.

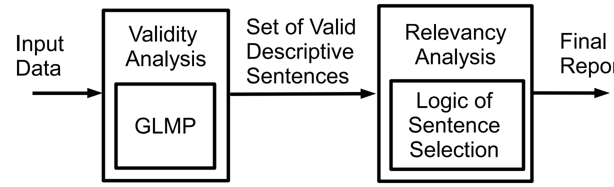


**Fig. 3.** Report generation diagram.

The report describing the temporal evolution of a phenomenon is obtained from the instantiation of the input data following a customized report template. The template of the reports is defined considering the particular needs of the users in order to highlight relevant aspects. For the application presented in this paper, a report template have been created in collaboration with human factors experts and it will be explained more in detail in Section 5.4.

Fig. 3 shows the report generation diagram followed for the automatic report generation. Initially, within the validity analysis, the full set of valid sentences describing the analyzed phenomenon are created. In a second stage, a logic of relevant sentence selection is implemented based on the customized template for a final report. The computational system selects among the available possibilities the most suitable linguistic expressions to describe the input data.

In this paper, in order to explain the possible causes of detected incidents, the report template includes de diagnosis of happening events which implies the need to solve the inverse problem concerned with fuzzy relations [34]. We have implemented a linguistic approach to present the solutions where the variables inferred in the diagnosis are explained based on NL sentences generated in previous section (see the example of application in Section 5.4).

## 5. Application: linguistic description of potential distraction levels of onboard devices

In the HITO project, driving simulators have been used in order to set a methodology for the evaluation of new vehicle onboard devices. This methodology analyzes driver–device interaction in a variety of situations in order to asses the *Index of Potential Distraction* (IPD) of the new device. The designed exercises require a sufficient degree of concentration from the driver and at the same time demand the interaction with different onboard devices. The amount of data generated in the experiments, with a number of drivers in various different situations, is huge and so the manual assessment process require many human resources.

In this work, linguistic descriptions of driving simulation exercises are generated automatically. The automated onboard device evaluation process will save time and resources and it will generate reports based on an unified expert criteria. The generated descriptions will either replace or complement manually generated expert reports and they will be mainly focused on the following aspects:

- Detection of distraction events during the exercise.
- Comparison of intervals with and without interaction between the driver and onboard devices.
- Generation of a linguistic report describing the overall quality of driving, providing an *IPD* of the analyzed onboard devices.

*5.1. Driving simulator: monitored parameters*

Vehicle simulator provides different parameters that are used to determine the driving quality and the *IPD* of the analyzed devices over the simulation exercises. These input parameters are related to the conduction, the controls of the vehicle and the simulation

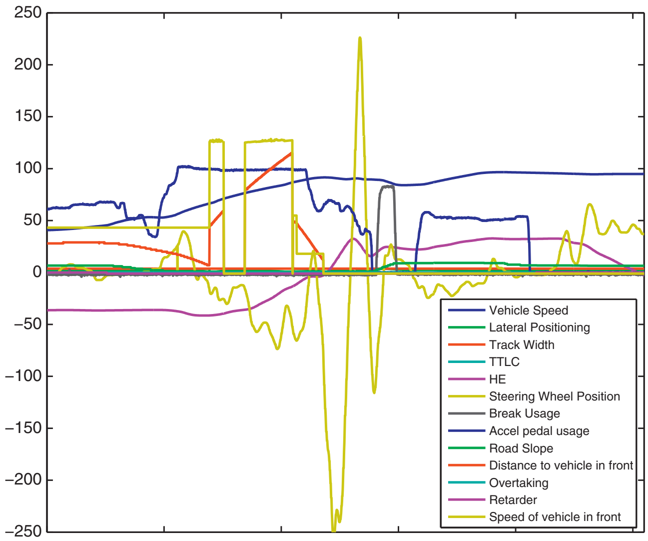


**Fig. 4.** Examples of signals obtained from the simulator CABINTEC.

environment (see Fig. 4). They are numerical values $z \in \mathbb{R}$ described as follows:

- $z_1$ *Vehicle Speed*: principal vehicle speed (km/h).
- $z_2$ *Lateral Position*: distance from the central point of the vehicle to the right extreme of the track where it is (m).
- $z_3$ *Track Width*: width of the track of circulation (cm).
- $z_4$ *TTLC*: time to cross the line on the edge of the road (in s) (*Lateral Position*/*Lateral Speed*).
- $z_5$ *HE*: angle between the tangents of the vehicle position and the road (°).
- $z_6$ *Steering Wheel Position*: measured steering wheel turning angle (°).
- $z_7$ *Percentage of brake usage*: percentage of actuation over the brake pedal (%).
- $z_8$ *Percentage of accelerator usage*: percentage of actuation over the acceleration pedal (%).
- $z_9$ *Road Slope*: slope or inclination of the road (%).
- $z_{10}$ *Distance to the Vehicle in front*: (m).
- $z_{11}$ *Vehicle Overtaking*: overtaking situation of the vehicle (boolean).
- $z_{12}$ *Retarder*: use of retarder. Hydraulic vehicle braking system (%).
- $z_{13}$ *Speed of the Vehicle in front*: (km/h).

Obviously, it is possible to extract a lot of useful information from these data. The detection of risk events and description of driving activity through the linguistic summarization of the above data inputs is only a first approach of the final automated application. For the comparison of interaction and non-interaction intervals a correlation between the appearance of risks events and the manipulation of the onboard devices is done. The information provided by the onboard devices may vary depending on the device but we are always able to determine when the interaction starts and ends within the simulation exercise. Therefore, we created a time-series signal $z_{int}$ representing the interaction intervals, as follows:

$$z_{int} = \begin{cases} 1 & \text{if interaction is active} \\ 0 & \text{if interaction is inactive} \end{cases} \qquad (1)$$

### 5.2. GLMP: Index of Potential Distraction of an onboard device

The *GLMP* in this application is designed to answer the general question of which is the *IPD* of a given onboard device (see Fig. 9). This linguistic model is an enhancement over the one presented in [13] where the description of the *Driving Quality* was performed. In this case the *Driving Quality* information is combined with the information of the analyzed onboard device in order to describe and evaluate the *IPD*.

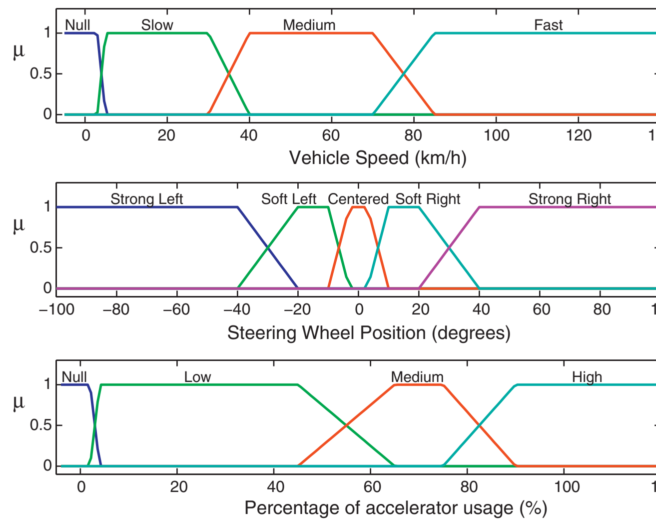


**Fig. 5.** Examples of trapezoidal linguistic labels used for fuzzification of 1*CPs*.

#### 5.2.1. 1*CPs*

In total, there are 14 input variables for the *GLMP*, 13 signals from the simulator and one signal defining the interaction with the device. Therefore, there are 14 1*CPs*, with 1*PMs* $\{p_1^1, p_2^1 \ldots p_{14}^1\}$. Each 1*PM* will be defined by the tuple $(U, y, g, T)$. As an example, the 1*PM* of *Vehicle Speed* is developed where:

- $U$ is the input value $z_1$ provided by the speed sensor.
- $y$ is a variable of type 1*CP* describing the *Vehicle Speed*. Its value is expressed by linguistic sentences and their corresponding weights of validity as follows: $(Null, w_{Null})$, $(Slow, w_{Slow})$, $(Medium, w_{Medium})$, $(Fast, w_{Fast})$. Here, e.g., *Slow* stands for complete linguistic expression *"The vehicle speed is Slow"*.
- $g$ is a function: $W_y = (\mu_{Null}(z), \mu_{Slow}(z), \mu_{Medium}(z), \mu_{Fast}(z))$ where $\mu_i$ are membership functions relating linguistic labels with the sensor's numerical value $z$.
- $T$ "*The vehicle speed is* {*Null* | *Slow* | *Medium* | *Fast*}".

Trapezoidal membership functions have been used to cover the domain of values of the different inputs parameters $z_i$. The definition of the membership functions have been made using expert knowledge. Fig. 5 shows several examples of the used trapezoidal membership functions.

#### 5.2.2. 2*CP*

According to the definition provided by the team of human factors experts involved in the HITO project, we describe and evaluate the quality of driving using three parameters, namely, *Steering Wheel Control*, *Vehicle Linearity* and *Security Distance*. These three variables can give insight on the risk events and the distraction levels while driving. They are 2*CPs* that are derived directly from the 1*CPs* described above. We used sets of fuzzy IF–THEN rules by each corresponding 2*PM* (named $p_1^2, p_2^2, p_3^2$ in Fig. 9).

$y_1^2$ The *Steering Wheel Control* is the output of $p_1^2$, function of the 1*CPs* $(y_1 \ldots y_{11})$ and the derivatives $(y_{d1} \ldots y_{d11})$. The *loss of steering wheel control* is defined as an abrupt manoeuvre, unusual in the vehicle direction control, and causing safety risk. Sudden track changes, big oscillations on the vehicle directions,

**Table 1**
Table of 1*CPs* with the corresponding linguistic variables and labels.

| *CP* ($y$) | Linguistic variables | Linguistic labels ($A_y$) |
|---|---|---|
| $y_1^1$ | *Vehicle Speed* | {Null, Slow, Medium, Fast} |
| $y_2^1$ | *Lateral Position* | {Short, Medium, Long} |
| $y_3^1$ | *Track Width* | {Narrow, Medium, Wide} |
| $y_4^1$ | *TTLC* | {Big Left, Small, Big Right} |
| $y_5^1$ | *HE* | {High Neg, Medium Neg, Low, Medium, High} |
| $y_6^1$ | *Steering Wheel Position* | {Strong Left, Soft Left, Centered, Soft Right, Strong Right} |
| $y_7^1$ | *Percentage of brake usage* | {Null, Low, Medium, High} |
| $y_8^1$ | *Percentage of accelerator usage* | {Null, Low, Medium, High} |
| $y_9^1$ | *Road Slope* | {Strong Descendent, Descendent, Null, Ascendent, Strong Ascendent} |
| $y_{10}^1$ | *Distance to the Vehicle in front* | {Not Measurable, Very Small, Medium, Big} |
| $y_{11}^1$ | *Vehicle Overtaking* | {Not, Yes} |
| $y_{12}^1$ | *Retarder* | {0, 1, 2, 3, 4} |
| $y_{13}^1$ | *Speed of Vehicle in front* | {Not Measurable, Null, Low, Medium, High} |
| $y_d$ | *Derivative* | {Rapidly Decreasing, Decreasing, Steady, Increasing, Rapidly Increasing} |

**Table 2**
Table of 2*CPs* with the corresponding linguistic variables and labels.

| | | |
|---|---|---|
| $y_1^2$ | *Steering wheel control* | {Low, Medium, High} |
| $y_2^2$ | *Vehicle Linearity* | {Low, Medium, High} |
| $y_3^2$ | *Security Distance* | {Low, Medium, High} |
| $y_4^2$ | *Driving Quality* | {Low, Medium, High} |
| $y_5^2$ | *Driving Quality during interaction* | {Low, Medium, High} |
| $y_6^2$ | *Driving Quality during non-interaction* | {Low, Medium, High} |
| $y^{Top}$ | *IPD* | {Low, Medium, High} |
| $y_i$ | *Integration* | {Never, A few times, Sometimes Many times, Most of the time, Always} |

out of track circulation, all could be indicators of loss in steering wheel control.

$y_2^2$ The *Vehicle Linearity*, output of $p_2^2$, is also function of 1*CPs* ($y_1 \ldots y_{11}$) and their derivatives ($y_{d1} \ldots y_{d11}$). It refers to the uniformity of the vehicle trajectory.

$y_3^2$ *Security Distance*, output of $p_3^2$, refers to the distance with respect to the vehicle in front. It is defined depending on Highway Code, road conditions and environmental conditions. This 2*CP* is function of 1*CPs* ($y_7 \ldots y_{13}$) and their derivatives ($y_{d7} \ldots y_{d13}$).

On the other hand, *Driving Quality* is also a 2*CP* in the *GLMP*. Any deviations on above mentioned three parameters suggest a distraction problem and a degraded quality of driving.

$y_4^2$ *Driving Quality* is the output of $p_4^2$. It is defined using the *CPs* ($y_1^2, y_2^2, y_3^2$). It determines the quality of driving at every instant during the simulation.

Each 2*PM* will be defined by the tuple ($U$, $y$, $g$, $T$). As an example, the 2*PM* of *Security Distance* ($y_3^2$) is developed where:

$U$ is a set of input *CPs* $U = (u_1, \ldots, u_n)$, where $u_i = (A_i, W_i)$ are the output *CPs* of the 1*PM*'s $\{p_7^1, p_8^1, \ldots, p_{13}^1\}$ and their derivatives. $U = \{y_7, \ldots, y_{13}, y_{d7}, \ldots, y_{d13}\}$.

$y$ is a variable of type 2*CP* describing the *Security Distance*. Its possible value is expressed by linguistic sentences and their corresponding weights of validity as follows: ($Low$, $w_{Low}$), ($Medium$, $w_{Medium}$), ($High$, $w_{High}$). Here, e.g., *Low* stands for *"The Security Distance is Low"*.

$g$ is the aggregation function $W_y = g(W_{y_7}, \ldots, W_{y_{13}}, W_{y_{d7}}, \ldots, W_{y_{d13}})$, where $W_y$ is a vector ($w_{Low}$, $w_{Medium}$, $w_{High}$) of validity degrees of the perception's linguistic labels in $y_3^2$. $W_{y_i}$ are the degrees of validity of the input computational perceptions.

$T$ *"The Security Distance is* {*Low* | *Medium* | *High*}*"*.

The aggregation function $W_y = g(W_{y_7}, \ldots, W_{y_{13}}, W_{y_{d7}}, \ldots, W_{y_{d13}})$ has been implemented using an expert set of fuzzy (IF–THEN) rules. In this rules, operator AND has been implemented through the minimum, while the operator OR has been implemented through the maximum. The following rules are an example of some of the rules

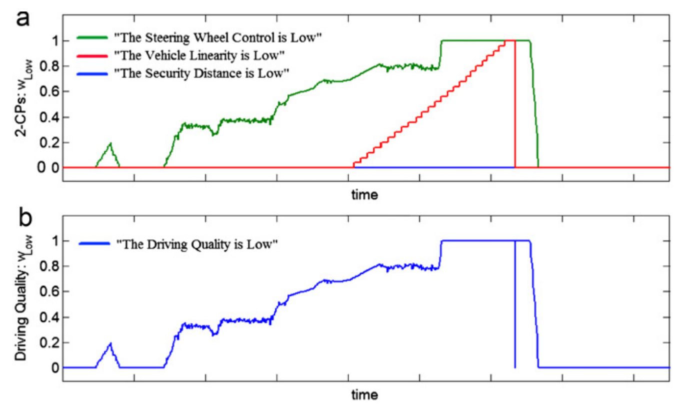


**Fig. 6.** Examples of instantaneous 2*CP* signals over a period of time. (a) The progression of the weights ($w_{Low}$) corresponding to de label "$a_{Low}$" of the parameters $y_1^2, y_2^2, y_3^2$. (b) The progression of $w_{Low}$ corresponding to the parameter $y_4^2$, i.e., Driving Quality.

used to define the *Security Distance* (see Tables 1 and 2 for quickly reference the names of *CPs*).

- IF ($y_7$ is Medium) AND ($y_{13}$ is Low) AND ($y_{10}$ is Medium) AND ($y_9$ is Strong Descendent) THEN $y_3^2$ is Low
- IF ($y_7$ is Medium) AND ($y_{13}$ is Medium) AND ($y_{10}$ is Very Small) THEN $y_3^2$ is Low
- IF ($y_{10}$ is Very Small) AND ($y_{13}$ is Low) AND ($y_{12}$ is not 0) THEN $y_3^2$ is Low
- IF ($y_{10}$ is Very Small) AND ($y_{13}$ is Low) AND ($y_7$ is not Null) THEN $y_3^2$ is Low
- IF ($y_{10}$ is Low) AND ($y_{d3}$ is Rapidly Decreasing) THEN $y_3^2$ is Low
- IF ($y_{10}$ is Low) AND ($y_{d3}$ is Decreasing) THEN $y_3^2$ is Low
- IF ($y_{13}$ is Low) AND ($y_{10}$ is Medium) THEN $y_3^2$ is Medium
- IF ($y_{10}$ is Low) AND ($y_{d10}$ is Slowly Decreasing) THEN $y_3^2$ is Medium
- IF ($y_{10}$ is Big) THEN $y_3^2$ is High

Fig. 6 shows an example of the evolution of the label *Low* at the 2*CPs* ($y_1^2, y_2^2, y_3^2$ and $y_4^2$) over a period of time within a particular simulation where a loss of driving quality happens.

Finally, two more 2*CPs* have been defined in order determine the driving qualities with and without interaction activity.

$y_{i5}^2$ *Driving Quality during interaction*. This perception is an *integrative CP* and it has been computed from inputs ($y_4^2$ and $y_{11}^1$). It provides the accumulated perception of the driving quality over the active interaction period during the simulation exercise. The aggregation function $g$ is the aggregation method *GD* that provides quantified sentences as described in Section 3.4.3.

$y_{i6}^2$ *Driving Quality during non-interaction*. This perception is also an *integrative CP* and it has been computed from inputs ($y_4^2$ and $y_{11}^1$) in order to have the accumulated perception of the driving y: **Driving Quality**:

| QAi \ Ai | Low | Medium | High |
|---|---|---|---|
| Never | **0.53** | 0 | 0 |
| A few times | 0.47 | 0.44 | 0 |
| Sometimes | 0 | **0.55** | 0 |
| Many times | 0 | 0 | 0 |
| Most of the time | 0 | 0 | **0.99** |
| Always | 0 | 0 | 0.01 |

Output sentences:
(0.99, "**Most of the time the Driving Quality has been High.**")
(0.55, "**Sometimes the Driving Quality has been Medium.**")
(0.53, "**Never the Driving Quality has been Low.**")

**Fig. 7.** Example of *CP* $y^2_{i4}$ or $y^2_{i5}$. The weights of the quantified sentences and the resulting output sentences.

quality while there has not been interaction over the simulation exercise.

Although these *CP*s are not used directly to define the top order perception, their output sentences will be required for the final report. Fig. 7 shows an example of the quantified sentences at integrative *CP*s referred to the quality of driving.

The most relevant question to answer by the 2*CP*s mentioned in this section could be about the reasons behind low overall driving quality. Thus, a typical description statement generated with them could be something like:

"*During interaction, sometimes the driving quality has been low, because the vehicle linearity has been low and . . .*"

### 5.2.3. Top order CP: IPD

In this application, the top order *CP* is the *IPD* that the analyzed onboard device has over the driver. This is a more complex type of *integrative CP* with the following elements:

*U* is a temporal series obtained from couples of instantaneous values of $y^2_4$ and $y^1_{11}$ along the duration of a simulation session, i.e., *IPD* is obtained as a combination of distraction rates over intervals of interaction and non-interaction.
*y* is a variable of type 2*CP* describing the *IPD*. Its possible value is expressed by linguistic sentences and their corresponding weights of validity as follows: $(Low, w_{Low})$, $(Medium, w_{Medium})$, $(High, w_{High})$. Here, e.g., *Low* stands for *"The Index of Potential Distraction is Low"*.
*g* is the aggregation function described below.
*T* *"The Index of Potential Distraction is* {*Low* | *Medium* | *High*}*"*.

The *IPD* is calculated as follows:

$$IPD = \frac{D_1}{(D_1 + D_2)} \qquad (2)$$

where $D_1$ and $D_2$ are the distraction rates of intervals with and without interaction respectively.

$D_1$ is calculated as the weighted sum of the cardinalities of labels *Low* and *Medium* on the driving quality $y^2_4$ over the active interaction period.

$$D_1 = k_1 \times \frac{\sum_{t=0}^{l} w_{Low}(t) \times w^{Int}_{Active}(t)}{\sum_{t=0}^{l} w^{Int}_{Active}(t)} + k_2 \times \frac{\sum_{t=0}^{l} w_{Medium}(t) \times w^{Int}_{Active}(t)}{\sum_{t=0}^{l} w^{Int}_{Active}(t)}, \qquad (3)$$

where the values $k_1 = 1$ and $k_2 = (1/2)$ were chosen empirically as a part of the aggregation function design. In this equation, $w_{Low}$ corresponds to the validity degree of linguistic clause "*The Driving Quality is Low*" of $y^2_4$. $w^{Int}_{Active}$ corresponds to the validity degree of sentence "*Interaction is Active*" of $y^1_{14}$ which is 1 when there is interaction and 0 otherwise, as defined in formula (1). In the same way, $D_2$ is calculated as follows:

$$D_2 = k_1 \times \frac{\sum_{t=0}^{l} w_{Low}(t) \times w^{Int}_{Inactive}(t)}{\sum_{t=0}^{l} w^{Int}_{Inactive}(t)} + k_2 \times \frac{\sum_{t=0}^{l} w_{Medium}(t) \times w^{Int}_{Inactive}(t)}{\sum_{t=0}^{l} w^{Int}_{Inactive}(t)}, \qquad (4)$$

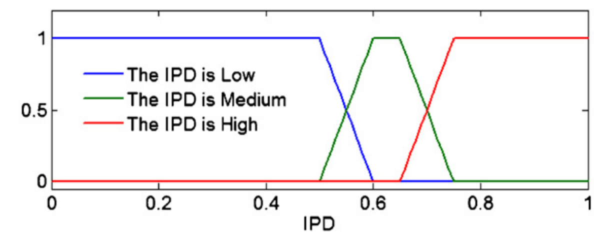


**Fig. 8.** Membership functions used for the verbalization of *IPD*.

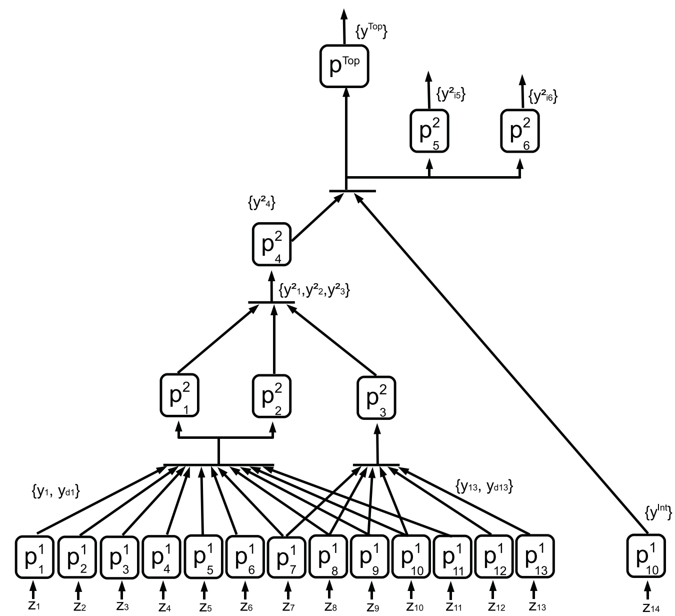

**Fig. 9.** *GLMP* developed to determine the *IPD* of an onboard device.

With this definition for *IPD*, the obtained value will be *Low* while $D_1 < D_2$. *IPD* will be *Medium* while $D_1$ above but still similar to $D_2$, and *IPD* will become *High* as $D_1$ gets considerably bigger than $D_2$ and the driving quality while interaction gets degraded. Fig. 8 shows the membership functions designed to verbalize the obtained *IPD* value.

Fig. 9 shows the *GLMP* developed for this application. As a reference, Tables 1 and 2 show the full list of *CP*s utilized, including the linguistic variables and labels used to describe them.

### 5.3. Identification of distraction events

In order to fulfill the report template as required by the final users, we needed to identify the, so-called, *distraction events*. Here we describe how to identify these events as specific situation types into the information available in the GLMP.

The values of $D_1$ and $D_2$ are an insight to the overall driving quality evaluations during the simulation. However the identification of particular low driving quality events is also important within the application. These individual distraction events allow users analyzing in depth the reasons behind them. The evaluation statements of the *IPD* will be accompanied by the most likely reasons for that judgements.

The identification of incidences is focused on the 2*CPs* $y_5^2$ and $y_6^2$ (*Driving Quality* with and without interaction). For this task a window length (*wl*) is considered within which the normalized cardinalities of labels *Low* and *Medium* were computed, e.g.,

$$CARD(Low) = \frac{1}{l}\sum_{j=k-l}^{k} w_{Low}(j). \tag{5}$$

where $k$ is the current sample, $l$ is the length in samples of the analyzed window, $l = wl \times sr$, and $sr$ is the sample rate. Using expert knowledge, in this application a *wl* of 10 seconds have been selected. The presence of distraction events is defined as CARD(*Low*) > 0.3 or CARD(*Medium*) > 0.6. Note that here, we must make the crisp decision of either including or not including the description of an event in the final report.

### 5.4. Template of the IPD analysis report

As mentioned above, the generation of a report describing the *IPD* within a simulation exercise needs to follow a customized report template in order to highlight the relevant aspects the final user needs.

The report template used for this application represents a typical summarization document defined by human factors experts within the HITO project. This typical summary focuses on the comparison of driving quality with and without interaction in order to determine the *IPD* an onboard device may have. On the other hand, the report focuses on the abnormal events during the driving activity providing the final user more information to determine the type of distractions the onboard device may induce.

Our software application generates a linguistic report including hypothesis about the relation between the distraction events and parameters that triggered each of the events.

The template of the report defined for this application consists of the following sections.

(i) *General observations*: this section holds various subsections to describe in detail the different aspects happened on the simulation exercise.
   (a) *Interaction activity*: this subsection gives information about the interaction activity with the analyzed onboard system during the simulation. It states the number of independent interactions and the accumulated time of interaction. A figure indicating the periods of interaction is also included.
   (b) *Comparison of interaction vs. non-interaction*: the quantified sentences obtained from $y_{i5}^2$ and $y_{i6}^2$ describing the quality of driving are compared trying to highlight the similarities and differences between them. The following sentences show an example:
   – "***Both during interaction and non-interaction***,
   ***Most of the time the Driving Quality has been High***".
   – "***During interaction, Sometimes the Driving Quality has been Low***,
   ***while during non-interaction A few times the Driving Quality has been Low.***"
   Note that the template provides two different linguistic expressions depending on the results. This subsection also provides a relation of distraction events during the simulation indicating the number of events, and for each event, occurrence time and cause of event, e.g.,
   – "***During interaction 1 distraction event happened***:"
   ***Event (1) at 100 seconds: The Vehicle Linearity is Low***"
   (c) *Detailed description of events during interaction*: in this subsection, the route-cause of events during the interaction are explained individually. Depending on the rules defined in each aggregation function, the *GLMP* navigates backwards on its branches solving the inverse problem, in order to deduce which rules have been triggered and by which parameters. Each event has an associated time, and so the generation of statements at the exact time of the event is a straight forward task (we identify the antecedents of the triggered fuzzy rules), e.g.,
   – "***Event (1) approximate time at 100 s.***
   ***The Vehicle Linearity is Low***:
   * ***The Vehicle Speed is Medium.***
   * ***The Percentage of accelerator usage is High.***
   * ***The Vehicle is Not Overtaking.***
   * ***The Lateral Position is Rapidly Decreasing.***"
   Each event description is also accompanied with video frames and figures showing the related input parameters.
(ii) Conclusion: this section provides the estimated *IPD* during the simulation exercise.

Therefore, the final layout of generated reports will be dynamic and depending on the number of distraction events detected during the simulation. Fig. 10 shows an example of an event description within one of the generated reports (see an explanation of this text in the next section).

## 6. Experimentation

The experimentation was based on data of a series of simulation exercises performed with professional drivers in a variety of settings. Within the project, four drivers conducted 8 exercises each in 4 predefined scenarios. These exercises were designed to evaluate 4 different onboard devices. For the validation of the implemented application, we were asked to focus on the "*mobile phone*" as onboard device. Exercises performed by the first driver were used to tune the rule sets and parameters defined within the GLMP.

A standard simulation exercise analyzed within the project contains data of 10–15 min, and considering the sample rate of the measurements (≃50 Hz) and the size of the *GLMP*, the number of sentences generated can amount to hundreds of thousands.

Fig. 10 shows an interesting example of a description of an event occurred during the execution of an simulation exercise in an inter-urban environment. At some point during this simulation, the driver (of the truck) must overtake a group of cyclists that appears on the road. In the particular case of the figure, the numerical information indicates the presence of an object (group of cyclists) in front of the vehicle. The speed of the object is low, while the truck is accelerating hard and not braking. The prolongation of this situation over a period of time is a reason why an incident happens. It is worth noting that the description of this particular event would only appear in a final document in case the event occurred during a positive interaction period as specified in the user requirements.

As shown in Fig. 10, the description of a particular event contains the explanation sentences that describe it and conjectures about the possible reasons that caused it. Then, the first graphic shows the first order parameters around the instant of the event. Secondly the progression of the incidence level signal during the same period is shown. Finally the frame of the simulation video that corresponds to the approximate instant of the incidence is represented (here the group of cyclists can be observed).

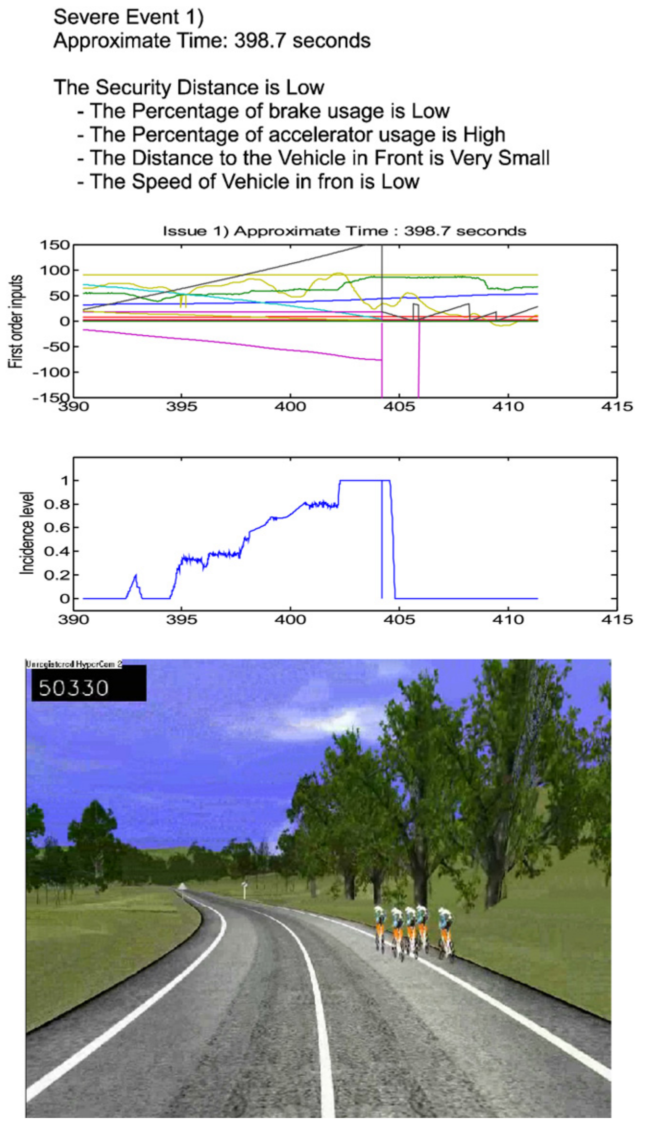


**Fig. 10.** Example of event description within a report that was generated during the experimentation.

At the time of describing the incidence, the rules that triggered the detection of a distraction event are identified. In this sense, the inverse problem concerned with fuzzy relations is investigated here. We propose the diagnosis of occurring incidents through a linguistic approach. In every instant of the simulation, the rules that are activated can be tracked. For example, we can access the combination of parameters in each rule, and the exact variables that trigger and cause the incidence can be reported.

## 7. Validation

One of the critical problems for the designer of this type of computational applications is that of assessing a degree of quality to each automatically generated text. This evaluation is needed to provide him/her with the necessary feedback to improve the final results.

Here in order to obtain a measure of text quality, the generated report was sent to five experts in human factors in the HITO project, who fulfilled a form composed by eight questions. This questionnaire is a result of our own research in this specific field [35]. In this section, we introduce the essentials of our approach and we address the interested reader to the referenced paper. Our source of inspiration are Pragmatics [36] and Systemic Functional Linguistics (SFL) [37]. From Pragmatics, we take the definition of a good communicative act and from SFL we take the structure of language as a system.

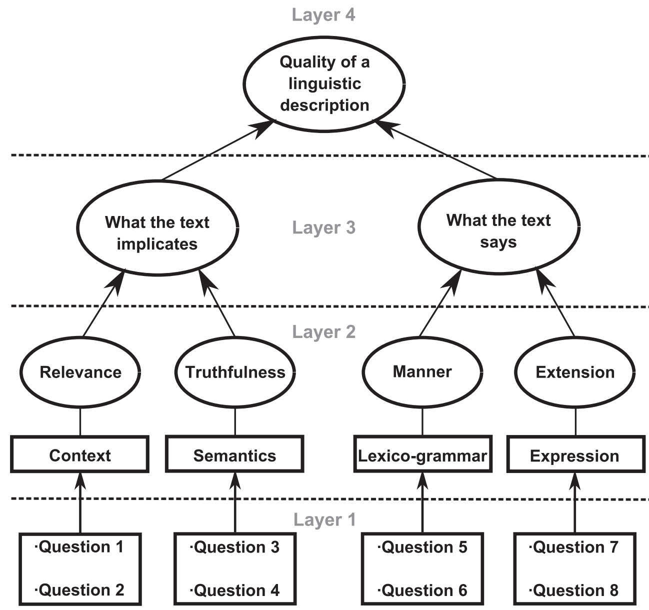


**Fig. 11.** Hierarchical structure of our concept of the quality of a report.

**Table 3**
Questions about quality of the generated reports.

| | |
|---|---|
| 1 | Indicate in which degree the content of this report belongs to the application domain of the HITO project |
| 2 | Indicate in which degree you identify the type of results expressed as the type of results expressed by yourself |
| 3 | After observing the behavior of the driver in the simulator, do you agree with the assessed global quality? |
| 4 | Do you agree with the provided explanations? |
| 5 | Indicate in which degree the vocabulary is used correctly |
| 6 | Indicate in which degree the ideas are correctly ordered to facilitate the comprehension of the report |
| 7 | Indicate in which degree the format of the report, including the use of figures and punctuation, is the most adequate |
| 8 | Indicate in which degree you consider that the extension of the report is the right respect to its content |

Fig. 11 shows a hierarchical structure of concepts about the report quality that are explained using the answers obtained using a questionnaire fulfilled by a human expert. The questions have five possible answers in form of numeric scale of evaluation in [1,5]. Table 3 contains our proposal of questions. Question 1 considers subjective relevance meanwhile question 2 considers the inter-subjective relevance. Questions 3 and 4 deal with evaluating the truthfulness of a report. With the ratings of the relevance and the truthfulness, we obtain the partial rating of the report respect to "what the text implicates". Questions 5 and 6 evaluate if the report uses the adequate vocabulary, if the order of the ideas is the most appropriate or if the used expressions are the right ones. Bad *quantity* reports contains too much or insufficient statements respect to the fact that try to be described; good *quantity* reports contains the right statements to understand the fact. To evaluate this aspect, we propose questions 7 and 8.

As mentioned above, for the validation of the implemented application, we were asked to focus on the "*mobile phone*" as onboard device. These sample gives a good insight of the potential of the developed technology, and the reports generated in these exercises were used to assess the quality of the automatically generated reports. In a typical experimental layout, we selected a group

**Table 4**
Average results of the evaluation.

| | R1 | R2 | R3 | R4 | R5 |
|---|---|---|---|---|---|
| Question 1 | 5 | 5 | 5 | 5 | 5 |
| Question 2 | 4 | 3 | 3 | 5 | 2 |
| Question 3 | 4 | 4 | 4 | 4 | 3 |
| Question 4 | 5 | 4 | 3 | 5 | 3 |
| Question 5 | 3 | 3 | 4 | 4 | 4 |
| Question 6 | 4 | 4 | 2 | 5 | 4 |
| Question 7 | 4 | 3 | 2 | 4 | 3 |
| Question 8 | 4 | 4 | 4 | 5 | 5 |
| Average | 4.1 | 3.8 | 3.4 | 4.6 | 3.7 |

of five people familiar with the HITO project. For each exercise, we gave them: the description of the context of the simulation, the video of the exercise, the automatically generated document and the form to evaluate it. After 45 minutes, they had to fulfill the questionnaire. The results of the different reviewers are shown in Table 4, where R1, R2, R3, R4 and R5 denotes each reviewer that participated in the experiment. The data in Table 4 are an example of the type of practical results obtained during the development of the project. The average denotes the global rating for the report. Two of the five ratings are higher than 4 and the minimum rating is 3.7. The average is calculated as the arithmetic mean of all questions; therefore, all the questions have the same importance for the global rating. It is relevant to note that all the reviewers agree with the maximum rating in question 1, the question that tries to determine if the content of the report belongs to the application domain. Question 7 has the worst ratings, following by question 2 and question 5.

After analyzing these results the designers must interpret them in order to improve the HITO report generator. For instance, the rating of question 7 indicates that the format of the report and the different techniques for expressing the information must be improved, because the reviewers consider that it is not good enough. Question 2, that analyzes the way as the results are expressed, also has a low global rating. Therefore, this is other aspect that the designers should improve to make better reports.

## 8. Conclusion

Linguistic description of phenomena is a very complex challenge. The expected results of this new technology will soon be useful for experts dealing with the monitoring and evaluation of big volumes of data. The exploitation of data generated by simulators is a remarkable example of these applications. Simulators are widely used in a variety of fields and in many situations, domain experts are required in order to observe and evaluate the amount of data acquired in the performed exercises.

In this paper, based on CTP, we contribute to the automatic evaluation of vehicle onboard devices in a simulation environment. Using our solution, experts in the field will save time and resources when analyzing all the data generated in simulation exercises. The system will provide objective reports generated based on unified criteria with the interpretation, arguments and conclusions derived from the data.

We have presented the complete linguistic model and report generation template defined on the HITO project. With this application we fulfill the requirement of automatic evaluation of onboard devices in road transport environment. With respect to our previous works, we have extended the linguistic model and report template in order to meet the final specifications of the project. We have included the detailed description of our approach for automatic text generation and we have included the definition of the *IPD*. We have also included our approach for the quality assessment of the generated text. We seek to evaluate the usability of the generated report from the users' point of view, and we have identified different indicators in a hierarchical form to define the quality of a linguistic report.

Through the use of *assertive CP*, *derivative CP* and *integrative CP* computational perceptions inspired in Control Theory, the GLMP has been clearly adapted to be applied to the representation of temporal evolution of phenomena.

It is worth noting that in this application we have implemented a solution to the, so-called *inverse problem*. The computational system navigates through the GLMP and realizes a backward search in order to determine the rules that were activated and ultimately triggered each event. Thus, conjectures about the possible reasons (variables) that caused the events can be established.

Interestingly, the GLMP paradigm has been useful for the whole project team, including human factors experts, helping to define the new concept of Index of Potential Distraction. Our experience in the HITO project encourages us to apply this paradigm in other application fields.

We have a lot of pending future work in this area. Here we have presented a specific solution to a specific problem of linguistic description of data. In the context of the HITO project, we have created just a first prototype that must be validated and tuned after a period of practical application. Many topics of research remain unexplored in front of us, e.g., develop new types of Perception Mapping, i.e., new types of sentences; develop the mechanisms to select the more relevant sentences in each situation type; generate different linguistic expressions according with the experience of each different user, and develop new methods for assessing the quality of the obtained reports.

We think that the presented automatic linguistic description approach can be applied in a wide variety of domains and in multiple forms. For example, this technology could be used for driving quality assessment in real time. In this case, the driving quality assessment would generate text messages that should be converted to voice. We could also use this technology in simulation training environments. Depending on students' performance in training with respect to established objectives, customized training plans could be proposed automatically.

## Acknowledgments

This work was supported in part by the Spanish Ministry of Science and Innovation (Grants TIN2008-00040, PSS-370100-2007-12, TIN2008-06890-C02-01 and TIN2011-29827-C02-02) and the Spanish Ministry for Education (FPU Fellowship Program).

## References

[1] M. Meyer, J. Booker, Eliciting and Analyzing Expert Judgment: A Practical Guide, Society for Industrial Mathematics, 2001.
[2] M. Riojas, C. Feng, A. Hamilton, J. Rozenblit, Knowledge elicitation for performance assessment in a computerized surgical training system, Applied Soft Computing 11 (2011) 3697–3708.
[3] A. Proper, Intelligent Transportation Systems Benefits: 1999 Update, US Dept. of Transportation, Federal Highway Administration, ITS Joint Program Office, 1999.
[4] L. Tijerina, E. Parmer, M. Goodman, Driver workload assessment of route guidance system destination entry while driving: a test track study, in: Proceedings of the 5th ITS World Congress, 1998, pp. 12–16.
[5] L. Chittaro, L. De Marco, Driver distraction caused by mobile devices: studying and reducing safety risks, in: Proceedings of the International Workshop on Mobile Technologies and Health: Benefits and Risks, Udine, 2004.
[6] R.R. Yager, A new approach to the summarization of data, Information Sciences 28 (1982) 69–86.
[7] R.R. Yager, Fuzzy summaries in database mining, in: Proceedings of the 11th Conference on Artificial Intelligence for Applications, IEEE, 1995, pp. 265–269.
[8] J. Kacprzyk, R. Yager, S. Zadrozny, A fuzzy logic based approach to linguistic summaries of databases, International Journal of Applied Mathematics and Computer Science 10 (2000) 813–834.

[9] J. Kacprzyk, S. Zadrozny, Computing with words is an implementable paradigm: fuzzy queries, linguistic data summaries and natural language generation, IEEE Transactions on Fuzzy Systems 18 (2010) 461–472.
[10] L.A. Zadeh, From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions, IEEE Transactions on Circuits and Systems 45 (1999) 105–119.
[11] L.A. Zadeh, A new direction in AI: toward a computational theory of perceptions, AI Magazine 22 (2001) 73–84.
[12] L. Eciolaza, G. Trivino, B. Delgado, J. Rojas, M. Sevillano, Fuzzy linguistic reporting in driving simulators, in: IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS), IEEE, 2011, pp. 30–37.
[13] L. Eciolaza, G. Trivino, Linguistic reporting of driver behavior: summary and event description, in: Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA), 2011, pp. 148–153.
[14] E. Reiter, R. Dale, Building Natural Language Generation Systems, Cambridge Univesity Press, 2000.
[15] O. Gusikhin, D. Filev, N. Rychtyckyj, Intelligent vehicle systems: applications and new trends, Informatics in Control Automation and Robotics (2008) 3–14.
[16] D. Salvucci, Predicting the effects of in-car interface use on driver performance: an integrated model approach, International Journal of Human–Computer Studies 55 (2001) 85–107.
[17] D. Salvucci, K. Macuga, Predicting the effects of cellular-phone dialing on driver performance, Cognitive Systems Research 3 (2002) 95–102.
[18] Hito cabintec project website, http://www.cabintec.net/proyectos-hito.asp, 2011.
[19] L.A. Zadeh, Toward human level machine intelligence – is it achievable? the need for a paradigm shift, IEEE Computational Intelligence Magazine 1 (2008) 11–22.
[20] J. Lawry, A methodology for computing with words, International Journal of Approximate Reasoning 28 (2001) 51–89.
[21] L.A. Zadeh, The concept of linguistic variable and its application to approximate reasoning, Information Sciences 8 (1975) 199–249.
[22] L.A. Zadeh, A computational approach to fuzzy quantifiers in natural languages, Computing and Mathematics with Applications 9 (1983) 149–184.
[23] A. Bargiela, W. Pedrycz, Granular Computing: An Introduction, Kluwer Academic Publishers, 2003.
[24] L. Lietard, A new definition for linguistic summaries of data, in: Proceedings of International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2008, pp. 506–511.
[25] R. Castillo-Ortega, N. Marí n, D. Sánchez, A fuzzy approach to the linguistic summarization of time series, Journal of Multiple-Valued Logic and Soft Computing 17 (2011) 157–182.
[26] J. Kacprzyk, S. Zadrozny, Computing with words and systemic functional linguistic: linguistic data summaries and natural language generation, in: V.N. Huynh, et al. (Eds.), Integrated Uncertainty Management and Applications, AISC, Springer-Verlag, Berlin, 2010, pp. 23–36.
[27] L. Martinez, Sensory evaluation based on linguistic decision analysis, International Journal of Approximate Reasoning 44 (2007) 148–164.
[28] E. Herrera-Viedma, E. Peis, J.M.M. del Castillo, S. Alonso, K. Anaya, A fuzzy linguistic model to evaluate the quality of web sites that store xml documents, International Journal of Approximate Reasoning 46 (2007) 226–253.
[29] I. Glockner, Fuzzy Quantifiers: A Computational Theory, Springer Verlag, 2006.
[30] F. Dí az-Hermida, A. Bugarí n, S. Barro, Definition and classification of semi-fuzzy quantifiers for the evaluation of fuzzy quantified sentences, International Journal of Approximate Reasoning 34 (2003) 49–88.
[31] S. Peters, D. Westerståhl, Quantifiers in Language and Logic, Oxford University Press, USA, 2006.
[32] K. Ogata, Modern Control Engineering, Prentice Hall, 2009.
[33] M. Delgado, D. Sánchez, M. Vila, Fuzzy cardinality based evaluation of quantified sentences, International Journal of Approximate Reasoning 23 (2000) 23–66.
[34] C. Pappis, M. Sugeno, Fuzzy relational equations and the inverse problem, Fuzzy Sets and Systems 15 (1985) 79–90.
[35] M. Pereira-Farina, L. Eciolaza, G. Trivino, Quality assessment of linguistic description of data, in: ESTYLF2012: XVII Congreso Espanol sobre Tecnologías y Lógica Fuzzy, 2012.
[36] K. Korta, J. Perry, Pragmatics, in: N.Z. Edward (Ed.), The Stanford Encyclopedia of Philosophy, 2011, http://plato.stanford.edu/archives/sum2011/entries/pragmatics/
[37] M.A.K. Halliday, M.I.M. Matthiessen, Construing Experience Through Meaning: A Language-based Approach to Cognition, Cassell London, 1999.

Please cite this article in press as: L. Eciolaza, et al., Automatic linguistic reporting in driving simulation environments, Appl. Soft Comput. J. (2012), http://dx.doi.org/10.1016/j.asoc.2012.09.007