



HAL
open science

Automated vehicles: multivariate analysis of drivers' take-over behaviour

F. Saïd, Christine Chauvin

► **To cite this version:**

F. Saïd, Christine Chauvin. Automated vehicles: multivariate analysis of drivers' take-over behaviour. 13th ICNC-FSKD Conference 2017, Jul 2017, Guilin, China. hal-01617043

HAL Id: hal-01617043

<https://hal.science/hal-01617043v1>

Submitted on 16 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Automated vehicles: multivariate analysis of drivers' take-over behaviour

F. SAID
LMBA, UMR CNRS 6205
Université Bretagne Sud
Lorient, France

C. CHAUVIN
Lab-STICC, UMR CNRS 6285
Université Bretagne Sud
Lorient, France

Abstract—This paper presents a study carried out within the context of Level 3 automated driving. It aims to characterize drivers take-over behaviour by exploring quantitative data related to the driver's performance (vehicle data) and qualitative data (user experience expressed in post-activity interviews). For this purpose, several techniques of multivariate analysis were used such as clustering of variables and units. This study illustrates that user experience is associated with the magnitude of actions related to lateral and longitudinal control; positive experience being associated with smoother actions, whereas negative feelings were associated with rougher ones. Furthermore, two main driver's profiles emerged; they are described by two latent variables, one of which is quality of the actions. Future research will investigate the determinants of these profiles.

Index Terms—Automated driving, human factors, driver behaviour, data clustering, ClustOfVar

I. INTRODUCTION

Automation has grown considerably in the automotive domain during the last decade. More and more driving tasks can be delegated to technology and several manufacturers have announced their intention to market semi-autonomous vehicles by 2020 [1]. These vehicles provide dual-mode operation whereby, on demand, longitudinal and lateral vehicle control can be delegated to the car. They are classified at level 3 of the NHTSA (National Highway Traffic Safety Administration) taxonomy; vehicles at this level of automation enable the driver to cede full control of all safety-critical functions under certain traffic or environmental conditions but changes in those conditions require the transition back to driver control [2]. Several benefits are expected from car automation, such as improvement of traffic flow and safety. However, and paradoxically, highly automated vehicles also raise safety issues and substantive human factors challenges need to be addressed before these forms of automated systems become a practical reality [2]. These challenges concern mainly the transition between automatic and manual modes. When automated driving fails, or is limited, the autonomous mode disengages and the driver is expected to resume manual driving. As has been emphasized [3], it is imperative that drivers react in an appropriate and timely manner for this transition to occur safely, whilst he/she has been engaged in another task. However, previous works [4] have shown the difficulties encountered by the driver to quickly assess the situation and to make a proper decision; these difficulties

are reminiscent of the "out of the loop syndrome", well known in the aviation domain. Beyond these general findings, various studies [5] [6] [7] indicate that drivers' take-over actions may be of a different type and quality depending upon contextual variables (e.g., traffic density, action possibilities, HMI) as well as personal factors (type of secondary tasks, drivers' distraction, drivers' gaze behaviour during automated driving). In line with these works, the present article presents an exploratory study aimed at characterizing drivers' take-over behaviours on the basis of two types of data: quantitative ones (vehicle data) and qualitative ones (user experience expressed by the drivers).

II. METHODS

A. Participants

The study participants were male employees of the Renault Company, ranging from 24 to 61 years of age ($m = 47$ years). All drive a car at least three times a week, use their vehicle cruise control function, and expressed an interest in automated vehicles. Thirty-six participants were recruited; eight of them provided data that could not be retained because of technical issues or discomfort brought on by the simulator. Hence, the data provided by 28 participants were analysed.

B. Experimental design

Each participant faced two automated driving conditions: without augmented reality (AD-woRA) and with augmented reality (AD-RA). Each condition was related to a different final scene to avoid a learning effect. The running order was counterbalanced.

C. Hardware and HMI

The experiment was conducted with a C-Cards driving simulator. In automated driving mode, without augmented reality conditions, participants were provided with a Head-Up Display (HUD) interface. In automated driving mode with augmented reality mode, participants were provided with an interface showing virtual information superimposed onto the real objects that exist in the road environment.

D. Experimental sessions

The experiment was conducted in two phases. In the first discovery phase, lasting one hour, participants were given the opportunity of familiarising themselves with the video simulator, the simulated vehicle, the scenario type and the different kinds of display. The actual test phase included an 11 minutes driving session before the Take-over request (TOR), the take over phase and a 20 minutes interview for each condition. During the driving sessions, participants were invited to play a game on a tablet placed under the windscreen on the right-hand side of the steering wheel. The TOR took place in a special context, when a vehicle in front of the ego-vehicle applied the brakes and the driver needed to change lanes (this goal was indicated on the windscreen).

E. Data collecting and coding

The study is based upon two types of data: vehicle data and post-activity interview data. The vehicle data refer to the quantitative variables which are presented in Table I: time variables (indicative of the drivers' take-over and lane change reaction speed), variables which indicate the smoothness of the action on the vehicle (its quality), and variables indicating the risk level. At the end of the experiment, a post-activity interview was conducted with each participant. This type of interview is designed to elicit a fine-grained description of a past activity carried out by people engaged in professional practice or in executing a task. The verbatim statements were transcribed, and broken down into propositions. The Hoc and Amalberti's coding scheme [8] was used to classify them into classes related to three main cognitive activities: information elaboration (before the interpretation process), diagnostic (mainly interpretation and inference), and decision making. A fourth class was added to take into account the emotional valence (negative or positive) associated with the take-over or the evaluation of the action. This last class relates to the driver's experience; it is labelled UXGlobal in Table I.

F. Processing Data

This paper adopts a two-step multivariate segmentation analysis, where clustering techniques are used for both variables and units. The first step consists in reducing the dimension of the original set of variables. Subsets of correlated variables are detected and the ones which provide the same kind of information are lumped into the same group. The obtained groups reveal the main dimensionalities of the data and a synthetic variable, or latent variable, is calculated for each of them. Classical techniques of dimensionality reduction such as Principal Component Analysis (PCA) or Multivariate Correspondence Analysis (MCA), for quantitative and qualitative variables respectively, are usually used in order to calculate latent variables. However, they exclude the possible presence of links within the set of latent dimensions. In that sense, techniques such as ClustOfVar, which is used in this study, are more powerful because they overcome the orthogonality constraint between the latent variables. After finding homogeneous groups of variables, the second step

TABLE I
VARIABLES : DESCRIPTION OF LABELS

Variables	Description
Vehicle variables	
TORDuration (s)	Time between the TOR and the switch to manual mode.
HandsOn (s)	Time between the TOR and the moment drivers place their hands on the steering wheel
Pbp (s)	Time between the TOR and the moment drivers press the button that deactivates the automatic mode
Accel1st (s)	Time between the TOR and the first acceleration(only variations of at least 1% are considered)
Accel1stSp (%/s)	Assesses the magnitude of the first action upon the accelerator pedal
AccelMaxDep (%/s)	Assesses the greatest magnitude of the action upon the accelerator pedal
Brak1stSp (N/s)	Assesses the first action upon the brake pedal. This variable shows the speed at which drivers press on the pedal, smoothly or abruptly; the more abrupt the drivers' action, the higher the speed.
BrakMaxPedalSpeed (N/s)	Assesses the greatest magnitude of the action upon the braking pedal
IndicatorSet(s)	Time between the TOR and the moment when drivers activate the indicator, thus signalling their intention to change lanes.
LaneChange (s)	Time between the TOR and the moment drivers arrive in the required lane
Vmax (m/s)	The maximum speed of the vehicle is calculated.
MaxAccelResult (m/s ²)	The resulting acceleration is the root sum between the longitudinal (a_x) and lateral (a_y) acceleration $a_{res} = \sqrt{a_x^2 + a_y^2}$
MaxLatAccel (m/s ²)	Vehicle's lateral acceleration(m/s ²).The maximum is calculated;it is observed when drivers perform an abrupt manoeuvre.
Steering rate max (/s)	Measures how smoothly drivers act upon the steering wheel.
DistToEndMergeLane (m)	The distance between the place of lane changing and the place where the manoeuvre would no longer be possible.
THW(s)	THW=Dint/Vego where Dint is the distance separating the two vehicles and Vego is the speed of the ego-vehicle.
TTCAhead	Time To Collision ahead (time remaining before the collision occurs with the vehicle placed in front of the ego-vehicle)
TTCRear (s)	Time To Collision rear
UXGlobal	The user's rating of his driving experience: negative, positive, NR (no response)

of our study consists of segmenting the units via classical clustering algorithms.

1) *ClustofVar*: ClustOfVar [9] is an algorithm for the statistical software R [10]. It detects groups of variables by using a homogeneity criterion that is based on the sum of correlation ratios for qualitative variables, and squared correlations for quantitative variables. Identification of the number of clusters is made from the analysis of aggregation levels and stability of the partitions via the bootstrapped mean-adjusted Rand index.

Then from each cluster a synthetic representative variable is extracted through PCAmix. The latter is a generalization of PCA and MCA that enables the handling of mixed datasets. The extracted new variable is the main principal component of PCAmix applied within each group of variables.

2) *The algorithm:* Consider two data matrices of N_1 standardized quantitative and N_2 qualitative variables on n units, respectively named $\{x_1, \dots, x_{N_1}\}$ and $\{y_1, \dots, y_{N_2}\}$. ClustOfVar follows the usual steps of hierarchical agglomerative methods. Therefore it starts by considering each of the $N = N_1 + N_2$ variables as forming a cluster on its own.

- Start with N partitions.
- Aggregate the two clusters A and B with the smallest similarity given by $d(A, B) = H(A) + H(B) - H(A \cup B)$. The homogeneity of the k -th cluster of variables $C_k \in \{C_1, \dots, C_K\}$ is calculated from $H_{C_k} = \sum_{x_i \in C_k} r^2(x_i, c_k) + \sum_{y_j \in C_k} \eta^2(y_j, u_k) = \lambda_k^1$. In the latter u_k is the first principal component of PCAmix applied to the standardized variables of C_k . The values of r^2 and η^2 are respectively the squared correlation of x_i and u_k , and the correlation ratio of y_j and u_k . The element λ_k^1 is the first eigenvalue of PCAmix applied to C_k . Consider a partition P_K of the initial set in K clusters, and define $H(P_k) = \sum_{i=1}^K H_{C_k} = \lambda_1^1 + \lambda_2^1 + \dots + \lambda_K^1$. The algorithm operates in such a way that each partition in pl clusters maximizes $H(P_K)$ amongst all the possible partitions that can be obtained by aggregating two of the clusters from the previous partition in $p - l + 1$ clusters.
- Stop when one single partition is obtained.

In order to evaluate the number K of suitable clusters, both the inspection of aggregation levels and a bootstrap method to evaluate partitions stability are used. The latter is based on the generation of b bootstrap samples of the N observations. On each bootstrapped sample the steps of ClustOfVar are repeated. Mean-adjusted Rand indices [11] are then computed from each of the b samples.

3) *Clustering of units:* The subsequent clustering of units is based on the first extracted components from each cluster of variables. It is performed via the hierarchical agglomerative criterion of Ward and a k-means type partitioning algorithm. A method based on a bootstrap approach to evaluate the stability of the partitions is used to determine suitable numbers of clusters.

III. RESULTS

This study deals with the data collected in the automated driving condition with augmented reality (AD-RA). Because of missing data, only 24 participants were taken into account.

A. Clusters of variables

To get an idea of the links between quantitative variables, one can plot the correlation circle of the two first PCA dimensions. In line with this, Figure 1 gives an initial overview of groups of correlated and anti-correlated variables among the vehicle variables. However, it does not provide a strict partition of the variables and to that end, a hierarchical partition of

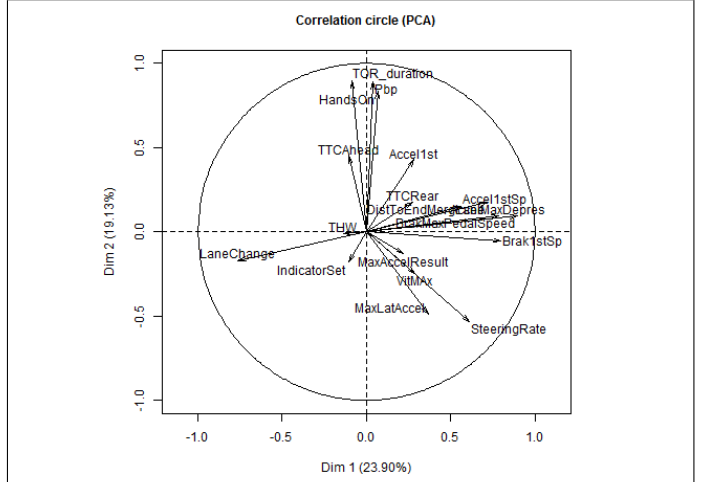


Fig. 1: Vehicle variables

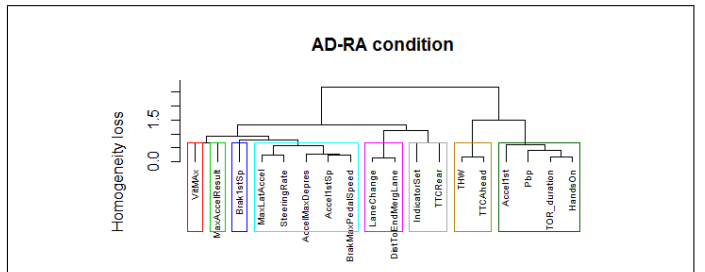


Fig. 2: Dendrogram of the vehicle variables

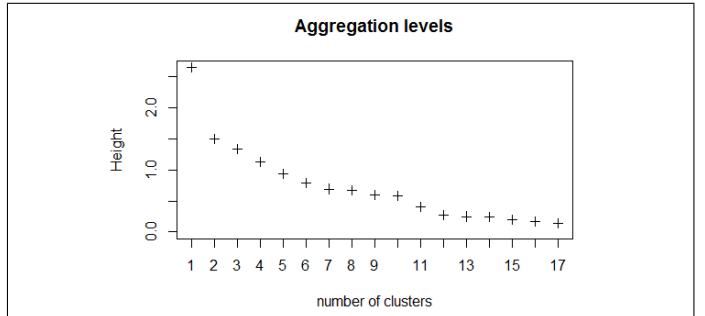


Fig. 3: Aggregation levels of the vehicle variables

the variables was constructed and was consolidated using the kmeans algorithm.

The dendrogram of the hierarchical clustering of the variables is shown in Figure 2. It highlights 8 and 11 as candidates for the number of clusters to retain, as does the examination of aggregation distances in Figure 3. The bootstrapped mean-adjusted Rand criterion in Figure 4 and the boxplots in Figure 5, which show the dispersion of the Rand indices over $B=40$ bootstrap replications, suggest the choice of 11 clusters. However, the gain in cohesion with 8 clusters is still significant (71.9%) so it was decided to favour bigger clusters by partitioning the vehicle variables into 8 groups.

Table II shows the composition of each cluster of variables together with their factor loadings. The latter corresponds to a squared correlation between any variable and the first principal

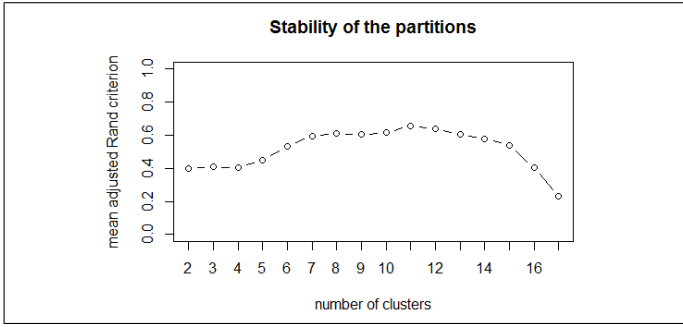


Fig. 4: Stability of the partitions of the vehicle variables

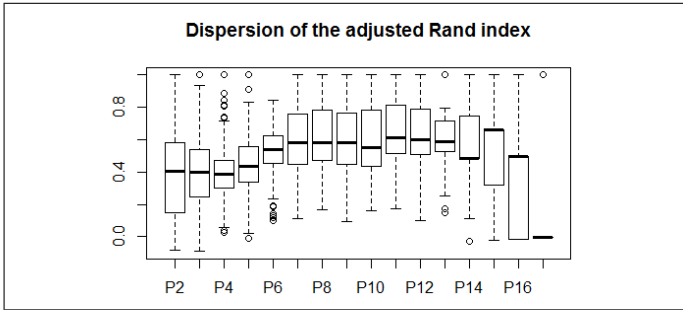


Fig. 5: Dispersion of the adjusted Rand index

component of the cluster to which it belongs.

- Cluster 1 is labelled *Take-over*. It aggregates all the variables which refer to the immediate take-over process. All variables in this group are positively correlated to their latent representative.
- Cluster 2 is labelled *Quality of driving*. It collects the variables which measure the magnitude of the driver's actions on the pedals and the steering wheel. High values of these variables correspond to rough actions.
- Cluster 4 relates to the lateral and rear control of the vehicle; the two variables of this group are anticorrelated.
- Cluster 5 is labelled *Changing lane*. It indicates the timing of the lane change. The latent variable of this group is positively correlated with the variable *Dist-ToEndMergeLane*, whereas it is anticorrelated with the variable *LaneChange*.
- Cluster 8 concerns ahead control of the vehicle.

B. Characterization of the user experience

A normed PCA was performed on the latent variables which represent the clusters described above. The three first PCA dimensions account for 61.9% of the total inertia as can be seen in Table III and the significant correlations between the PCA axes and latent variables help to interpret the PCA axes.

- The first axis opposes participants between those who execute smooth driving and late lane changes (on the left of the axis) and those who execute rough driving and early lane changes (on the right of the axis).
- The second axis opposes participants between those who drive at a moderate pace (low peaks of speed and low maximal resulting acceleration) and those who change lane early and have good control of the rear environment.

TABLE II
CLUSTERS OF VARIABLES

Squared loadings	
Cluster 1: Take-over	
TORDuration	0.83
HandsOn	0.84
Pbp	0.64
Accel1st	0.51
Cluster 2: Quality of driving	
Accel1stSp	0.80
AccelMaxDepres	0.72
BrakMaxPedalSpeed	0.75
MaxLatAccel	0.69
SteeringRate	0.70
Cluster 3: Brak1stSp	
Brak1stSp	1
Cluster 4: Rear&lateral control	
IndicatorSet	0.66
TTCRear	0.66
Cluster 5: Changing lane	
LaneChange	0.93
DistToEndMergLane	0.93
Cluster 6: VitMax	
VitMax	1
Cluster 7: MaxAccelResult	
MaxAccelResult	1
Cluster 8 : Ahead control	
THW	0.9
TTCAhead	0.9

TABLE III
CHARACTERIZATION OF PCA AXES

Dimension 1: 29,54% inertia		
	correlation	p.value
Quality of driving	0.91	0.00
Brak1stSp	0.80	0.00
Changing lane	0.71	0.00
	R^2	pvalue
UXglobal	0.48	0.00
Dimension 2: 17.7% inertia		
	correlation	p.value
Rear&lateral control	0.48	0.03
Changing lane	0.44	0.05
MaxAccelResult	-0.62	0.00
VitMax	-0.67	0.00
	R^2	pvalue
UXglobal	0.37	0.02
Dimension 3: 14.7% inertia		
	correlation	p.value
Ahead control	0.79	0.00
Rear& lateral control	0.46	0.03

- The third axis concerns the general control of the vehicle surroundings (ahead, lateral and rear).

In Figure 6, the participants are represented in the first PCA plane and are tagged with regards to the qualitative variable *UXglobal*. The participants with positive feedback

TABLE IV
CHARACTERIZATION OF THE USER EXPERIENCE

	v.test		
	Dim.1	Dim.2	Dim.3
negative	1.94	-2.30	0.26
NR	1.36	2.26	-0.65
positive	-3.06	0.76	0.21

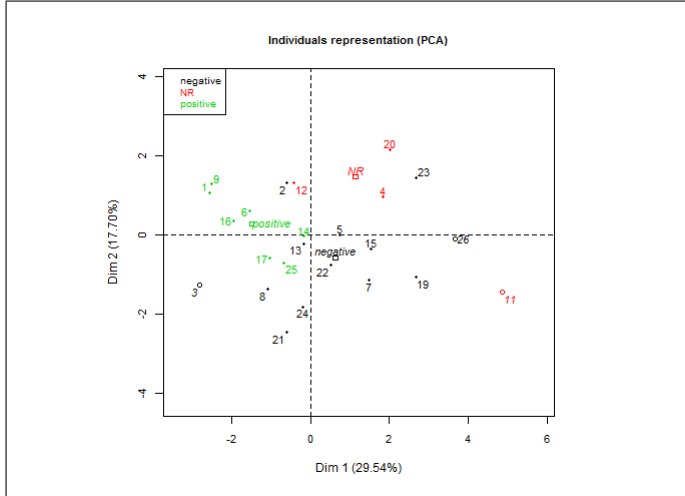


Fig. 6: Representation of the participants in the first PCA plane

form a homogeneous group at the left of the plane, while the participants with negative feedback are scattered along the first axis. The second dimension discriminates between negative and non-response feedback. These observations are confirmed by the tests presented in Table IV. Positive feedback is associated with negative coordinates in the first dimension; negative feedback is associated with negative coordinates in the second dimension, whereas participants who provided no feedback have positive coordinates in the second dimension. To conclude, positive feedback is associated with smoother actions on the pedals and the steering wheel and later lane changes. Negative feedback is associated with higher peaks of speed, higher resulting acceleration and more rapid lane changes.

C. Clustering of the participants

In order to construct the driver profiles, we used the PCA coordinates of the participants to perform hierarchical clustering and consolidated the resulting partition with k-means partitioning.

As suggested by the hierarchical tree in Figure 7, the participants were separated into 2 groups. The scatterplot illustrated in Figure 8 shows an opposition between the two groups relatively to the first dimension, and this observation is confirmed by the tests presented in Table V.

The latent variables which are significantly correlated to the partition are presented in Table VI and they enable the characterization of the two driver profiles:

- driver profile 1: good quality of driving, a smoother first action on the braking pedal, later lane changes;

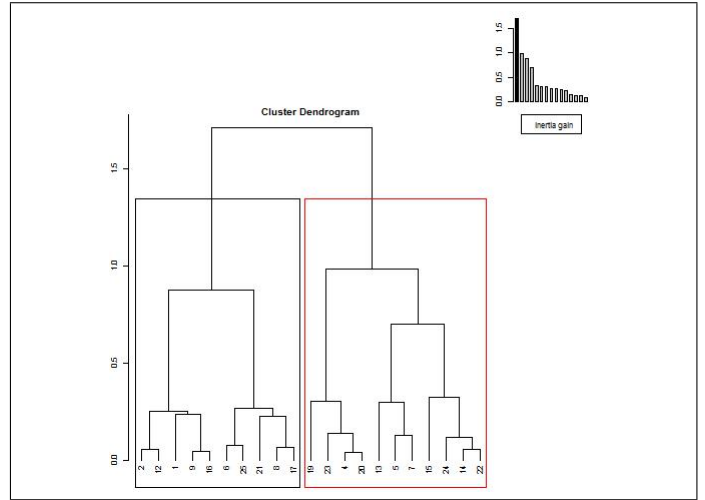


Fig. 7: Dendrogram of the participants

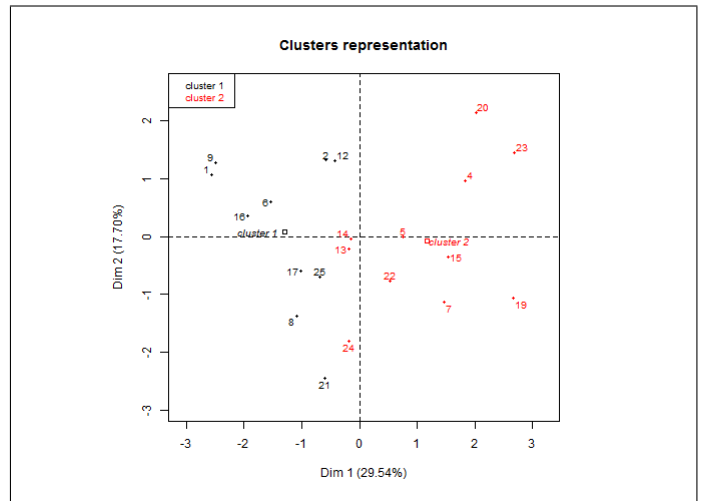


Fig. 8: Representation of the clusters of participants

TABLE V
DESCRIPTION OF THE DRIVER PROFILES WITH PCA AXIS

	v.test	Mean in category	Overall mean	p.value
Driver profile 1				
Dim.1	-3.59	-1.29	0	0.0003
Driver profile 2				
Dim.1	3.59	1.29	0	0.0003

- driver profile 2: poor quality of driving, a rougher first action on the braking pedal, earlier lane changes.

IV. DISCUSSIONS AND CONCLUSION

This study shows that different behaviours may be exhibited after the take-over of an automated vehicle. Two main profiles arise, which are defined by the combination of two main latent variables: the quality of driving (the intensity of the first braking action may be associated to this variable) and the lane change features. A group is characterized by smoother actions and later lane change. In this group, participants mainly expressed a positive user experience. The other group presents

TABLE VI
DESCRIPTION OF THE DRIVER PROFILES WITH LATENT VARIABLES

	v.test	Mean in category	Overall mean	p.value
Driver profile 1				
Quality of driving	-3.69	-1.29	-0.37	2e-04
Brak1stSp	-3.34	-0.76	-0.008	8e-04
Changing lane	-2.63	-0.92	-0.06	0.009
Driver profile 2				
Quality of driving	3.69	0.48	-0.37	2e-04
Brak1stSp	3.34	0.68	-0.008	8e-04
Changing lane	2.63	0.72	-0.06	0.009

opposing features: rougher actions and a faster lane change. One may think that the first ones mastered the situation, whereas the second ones were rushed.

This study illustrates that the perceived quality of the actions performed depends upon the magnitude of the longitudinal (accelerating and braking) and lateral control actions. Several studies have already emphasized the necessity to analyze the take-over of automated vehicles not only in terms of temporal features of the takeover but also in terms of "quality" of actions. The present study confirms the importance of such an approach and suggests a proposal for operationalizing the notion of quality.

From a statistical point of view, the methodology used proved to be useful for revealing amongst the latent traits of the data the ones which are the most significant descriptors of driver behaviours.

This paper focused on vehicle variables ; however the methodology used can handle mixed data and it will be used in the future to process multiple sources data: vehicle, eye-tracking data and qualitative data such as cognitive activities expressed during post-activity interviews. Furthermore, segmentation techniques could enrich this approach by identifying the main determinants of the driver profiles.

ACKNOWLEDGMENT

The authors would like to thank the Renault Company which commissioned and supported this study.

REFERENCES

- [1] C. Diels, "Will autonomous vehicles make us sick?", In Contemporary Ergonomics and Human Factors, S. Sharples and S. Shorrock, Eds. Boca Raton, FL: CRC Press, 2014, pp. 301-307.
- [2] T. E. Trimble, R. Bishop, J.F. Morgan, and M. Blanco, "Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts", Report No. DOT HS 812 043. Washington, DC: National Highway Traffic Safety Administration, 2014.
- [3] V. Dixit, S. Chand and D. Nair, "Autonomous Vehicles: Disengagements, accidents and reaction times", Plos ONE, vol 11, n12, e0168054.
- [4] N. Merat, A. Jamson, F. Lai, and O. Carsten, "Highly automated driving, secondary task performance, and driver state", Human Factors, vol 54, n5, pp. 762-771, 2012.
- [5] L. Lorenz, P. Kerschbaum, and J. Schumann, "Designing take over scenarios for automated driving How does augmented reality support the driver to get back into the loop?", Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 58, No. 1, pp. 1681-1685. SAGE Publications, September 2014.
- [6] J. Radlmayr, C. Gold, L. Lorenz, M. Farid, and K. Bengler, K., "How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving", Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 58, No. 1, pp. 2063-2067. Sage CA: Los Angeles, CA: Sage Publications, September 2014.
- [7] K. Zeeb, A. Buchner, and M. Schrauf, "What determines the take-over time? An integrated model approach of driver take-over after automated driving", Accident Analysis & Prevention, vol. 78, pp. 212-221, 2015.
- [8] J.-M. Hoc, and R. Amalberti, "Analyse des activits cognitives en situation dynamique: d'un cadre thorique une mthode", Le travail humain, vol. 62, pp. 97-129, 1999.
- [9] M. Chavent, V. Kuentz, B. Liquef, J. Saracco, "ClustOfVar: an R package for the clustering of variables", J. Stat. Softw. vol. 50, pp. 1-16, 2012.
- [10] R Core Team, "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing, Vienna, Austria, 2013. [http://cran.r-project.org/doc/manuals/fullrefman.pdf].
- [11] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, *NbClust: An Examination of Indices for Determining the Number of Clusters: NbClust Package*. R package version 1.4 (2013).
- [12] L. Hubert, P. Arabie, "Comparing partitions", J. Classif., pp. 193-208, 1985.
- [13] H.A.L. Kiers, "Simple structure in Component Analysis Techniques for mixtures of qualitative and quantitative variables", Psychometrika, vol. 56, pp. 197-212, 1991.
- [14] V. Kuentz-Simonet, S. Lyser, J. Candau, P. Deuffic, M. Chavent, J. Saracco, "Une approche par classification de variables pour la typologie dobservations: le cas dune enquete agriculture et environnement", Journal de la Socit Franaise de Statistique, vol. 154, pp. 37-63, 2013.
- [15] F. Leisch, "A toolbox for K-centroids cluster analysis, Comput. Stat. Data Anal, vol. 5, pp. 526-544, 2006
- [16] S.H. Liao, P.H. Chu, P.Y. Hsiao, "Data mining techniques and applications - a decade review from 2000 to 2011", Expert Syst. Appl., vol. 39, pp. 11303-11311, 2012.
- [17] K.V. Mardia, J.T. Kent, J.M. Bibby, *Multivariate Analysis*. London: Academic Press Inc., 1979.
- [18] J.H. Ward, "Hierarchical grouping to optimize an objective function", J. Am. Stat. Assoc, vol. 58, pp. 236-244, 1963.
- [19] J.G. Brida et al., "ClustOfVar and the segmentation of cruise passengers from mixed data: Some managerial implications", Knowledge-Based Systems, vol. 70, pp. 128-136, 2014.
- [20] M.-P. Ellies-Oury et al., "An innovative approach combining Animal Performances, nutritional value and sensory quality of meat", Meat Science, vol. 122, pp. 163-172, 2016.
- [21] E. Vigneau, E. M. Qannari, "Clustering of variables around latent components", Digests of Statistics Simulation and Computation, vol. 12, pp. 1131-1150, 2003.
- [22] E. Vigneau et al., "Classification de variables autour de composantes latentes", Revue de Statistique Appliquee, vol. 54, pp. 27-45, 2006.