Annual Reviews in Control 44 (2017) 316–322

Contents lists available at ScienceDirect

Annual Reviews in Control

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

Review article

Towards increased systems resilience: New challenges based on dissonance control for human reliability in Cyber-Physical&Human Systems

F. Vanderhaegen

University of Valenciennes, LAMIH, CNRS UMR 8201, F-59313 Valenciennes, France

ARTICLE INFO

Article history:
Received 15 February 2017
Revised 19 September 2017
Accepted 22 September 2017
Available online 6 October 2017

Keywords:
Dissonance oriented stability analysis
Dissonance control
Human reliability
Resilience
Risk assessment
Cooperation
Learning
Reinforcement
Cyber-Physical&Human Systems (CPHS)

ABSTRACT

This paper discusses concepts and tools for joint human and cyber-physical-systems analysis and control in the view of increasing the whole system resilience. More precisely, it details new challenges for human reliability based on dissonance control of Cyber-Physical&Human Systems (CPHS) to improve the system's resilience. The proposed framework relates to three main topics: the stability analysis in terms of dissonances, the dissonance identification, and the dissonance control. Dissonance oriented stability analysis in this sense consists in determining any conflicting situations resulting from the human behaviors interacting with Cyber-Physical Systems (CPS). Frames of reference support the assessment of stable or unstable gaps among stability shaping factors and the identification of dissonances. Dissonance control consists in reinforcing the frames of reference by applying reinforcement modes. It aims then at accepting or rejecting the identified dissonances by using supports such as expert judgment, feedback of experience, simulation, learning or cooperation. An example in road transportation illustrates the interest of the proposed framework by studying possible dissonances between car drivers and CPS. As automation spreads out in society by generating close interactions with humans, the ideas of the paper will support the design of new analysis and control tools jointly made by researchers from social and control sciences to study the resilience of the whole CPHS in terms of dissonances.

Contents

1. Introduction . . . 316
2. Dissonance oriented stability analysis . . . 317
3. Dissonance identification . . . 319
4. Dissonance control . . . 320
5. Conclusion. . . . 321
Acknowledgements . . . 321
References . . . 321

1. Introduction

Systems have to be designed not for prohibiting individual unsafe acts but for preventing human error occurrence or reducing their potential consequences by specifying adequate barriers or defenses (Reason, 2000). Human error can then be seen as a consequence of a failed defense instead of a cause of an unsafe event. However, more than 70% of accidents remain due to human errors and 100% of them are directly or indirectly linked with human factors (Amalberti, 2013). Moreover, even if a technical system such as a Speed Control System (SCS) is designed to improve the safety or the comfort of the car driver, its use can produce unsafe situations due to the reduction of the inter-distance between cars, the increasing of the reaction time or the decreasing of the human vigilance (Dufour, 2014). This paper proposes some ways to analyze such dilemma between the design and the use of a system. It extends the proposal by presenting some new challenges for assessing and controlling human reliability of Cyber-Physical&Human Systems (CPHS) where a Cyber-Physical System (CPS) or several CPS interact with human operators. It is an extension of the ple-

---

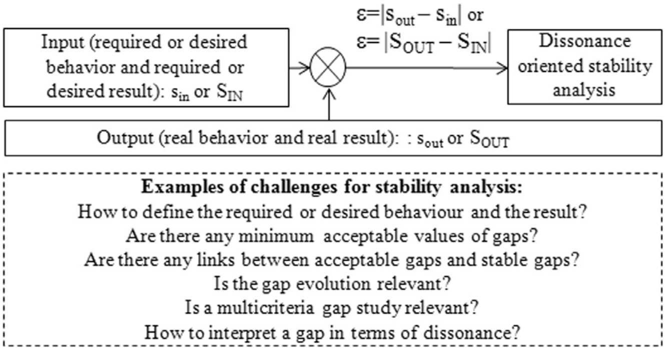
E-mail address: frederic.vanderhaegen@univ-valenciennes.fr

https://doi.org/10.1016/j.arcontrol.2017.09.008

nary session given by the author at the first IFAC conference on CPHS entitled “Human reliability and Cyber-Physical&Human Systems” in Brazil. Based on the author’s experience and on literature reviews, several challenges are discussed and motivate a new framework proposal for human reliability study in CPHS.

Human reliability usually confronts the problem of its definition and its assessment in the course of the design, the analysis or the evaluation of CPHS such as human-machine systems, joint cognitive systems, systems of systems, socio-technical systems, multi-agent systems, manufacturing systems or cybernetic systems. Human reliability of CPHS may be defined by distinguishing two sets of frames of reference or baselines: (1) the frame related to what the human operators are supposed to do, i.e. their prescriptions, (2) the frame related to what they do outside these prescriptions. Human reliability can then be seen as the capacity of human operators to realize successfully the tasks required by their prescriptions and the additional tasks, during an interval of time or at a given time. Human error is usually considered as the negative view of human behaviors: it is the capacity of human operators not to realize correctly their required tasks or the additional tasks. Methods for analyzing human reliability exist and are well explained and discussed on published states-of-the-art (Bell & Hollroyd, 2009; Hickling & Bowie, 2013; Kirwan, 1997a,b; Pan, Lin, & He, 2016; Reer, 2008; Straeter, Dolezal, Arenius, & Athanassiou, 2012; Swain, 1990; Vanderhaegen, 2001, 2010). They consider mainly the first set of tasks, i.e. they study the possible human errors related to what the users are supposed to do. Human reliability assessment methods remain unsuitable or insufficient, and new developments have to be done considering new constraints such as the dynamic evolution of a system upon the time, the variability of a human operator or between human operators, or the creativity of human operators who are capable to modify the use of a system or to invent new uses. Regarding such new requirements for human reliability study, many contributions present the concept of resilience as an important issue for organization management and for controlling criteria such as safety, security, ethics, health or survival (Engle, Castle, & Menon, 1996; Hale & Heijer, 2006; Hollnagel, 2006; Khaitan & McCalley, 2015; Orwin & Wardle, 2004; Pillar, 2016; Ruault, Vanderhaegen, & Kolski, 2013; Seery, 2011; Wreathall, 2006). Resilience is usually linked with the system stability, and it is defined as the ability or the natural mechanisms of a CPHS to adjust its functioning after disturbances or aggressions, in order to maintain its stable state, to come back to a stable state or to recover from an instable state. The more stable a system, the less uncertain the human attitudes related to beliefs and intentions (Petrocelli, Clarkson, Tormala, & Hendrix, 2010). On the other hand, other studies present the organizational stability as an obstacle for being resilient, and the instability as an advantage to survive (Holling, 1973, 1996; Lundberg & Johansson, 2006). Then, a system such as a CPHS with regular important variations that provoke its instability may survive and be resilient for a long period of time, whereas an isolated stable CPHS that does not interact with others may not be resilient when an external aggression occurs and makes it instable.

This paper proposes new challenges for the human reliability study of CPHS based on the above mentioned concept of stability applied to human behaviors. The analysis of human stability is interpreted in terms of dissonances and the successful control of these dissonances makes the CPHS resilient. The concept of dissonance is adapted from Festinger (1957) and Kervern (1994), and a dissonance is a conflict of stability. Three main topics are then discussed in Sections 2–4 respectively: the dissonance oriented stability analysis, the dissonance identification, and the dissonance control. In parallel, a case study based on road transportation illustrates an application of such new ways to treat human reliability of a CPHS by taking into account the integration of different CPS into a car, i.e., a Speed Control System (SCS) and an Adaptive Cruise Control (ACC) that replaces the previous SCS.



Examples of challenges for stability analysis:
How to define the required or desired behaviour and the result?
Are there any minimum acceptable values of gaps?
Are there any links between acceptable gaps and stable gaps?
Is the gap evolution relevant?
Is a multicriteria gap study relevant?
How to interpret a gap in terms of dissonance?

Fig. 1. Dissonance oriented stability analysis challenges.

2. Dissonance oriented stability analysis

Human stability relates to the equilibrium of human behaviors, i.e. human behaviors or their consequences remain relatively constant around a threshold value or an interval of values whatever the disturbances that occur. Out of this equilibrium, human behaviors or their consequences are unstable. The threshold value or the interval of values can be determined qualitatively or quantitatively by taking into account intrinsic and extrinsic factors such as technical factors, human factors, environmental factors or organizational factors. Human factors are physical, cognitive or physiological parameters or their impact factors for instance. Human stability is then analyzed by using these intrinsic or extrinsic factors and by comparing input factors with output ones. Fig. 1 gives a non-exhaustive list of human stability challenges that will be discussed hereafter. The input and output factors relate to human behaviors and to their consequences. Their assessments are noted s_{in} and s_{out} when it is a single value, or noted S_{IN} and S_{OUT} when it is a matrix of successive values related to different measurements. The resulting gap is a single value or a matrix, noted ε. Input and output factors are human stability shaping factors and a given factor can have an impact on the same factor or other factors. Measurement criteria are then required in order to assess and compare them. These measurements aim at detecting stable or unstable gaps of human behaviors or those of their impacts by studying instantaneous values of gaps, the evolution of these values upon the time, the frequency, the duration or the shape of this evolution for instance (Richard, Vanderhaegen, Benard, & Caulier, 2013; Vanderhaegen, 1999a, 2016c). Their analysis can also determine the associated risks of human stability or instability by taking into account instantaneous, variable or regular gaps. The variability and the sustainability of gaps related to the study of irregular or regular evolutions of gaps can require risk analyses on different intervals of time.

A measure of a risk is usually defined as the product between a measure of the occurrence of an undesirable event and a measure of its gravity. As a matter of fact, the risk of human behaviors is sometimes explained as a compromise between several criteria by taking into account good and bad practices (Vanderhaegen & Carsten, 2017). The so-called Benefit-Cost-Deficit (BCD) model takes into account the positive and negative gaps for different criteria (Sedki, Polet, & Vanderhaegen, 2013; Vanderhaegen, 2004; Vanderhaegen, Zieba, & Polet, 2009; Vanderhaegen, Zieba, Polet, & Enjalbert, 2011). The positive gaps are benefits, the negative but acceptable ones are costs and the unacceptable negative ones are deficits or dangers. The BCD parameters can be weighted with a probability of success or of failure of human behaviors