The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences

The Multimodal Tutor

Adaptive Feedback from Multimodal Experiences

Daniele Di Mitri

The Multimodal Tutor

Adaptive Feedback from Multimodal Experiences

The research reported in this thesis was carried out at the Open Universiteit in the Netherlands at the Faculty of Education, formerly known as Welten Institute – Research Centre for Learning, Teaching and Technology,

and under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

SIKS Dissertation Series No. 2020-17

Printed by ProefschriftMaken
Cover design: Daniele Di Mitri
Typeset in LaTeX

ISBN/EAN: 978 94 93211 21 6

Multimodal Tutor

Adaptive Feedback from Multimodal Experiences

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. dr. Th. J. Bastiaens
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 4 september 2020 te Heerlen
om 13:30 uur precies

door

Daniele Di Mitri
geboren op 23 april 1991 te Bari, Italië

Promotores

Prof. dr. H.J. Drachsler
Open Universiteit / DIPF - Leibniz Institute for Research and Information in Education

Prof. dr. M.M. Specht
Open Universiteit / TU Delft

Co-Promotor

Dr. J. Schneider
DIPF - Leibniz Institute for Research and Information in Education

Leden beoordelingscommissie

Prof. dr. M. Kalz
Open Universiteit

Prof. dr. K. Hindriks
Vrije Universiteit Amsterdam

Prof. dr. H.U. Hoppe
Universität Duisburg-Essen

Prof. dr. R. Klamma
Rheinisch-Westfälische Technische Hochschule Aachen

Dr. S. Bromuri
Open Universiteit

Life can only be understood backwards, but it must be lived forwards.

Søren Kierkegaard

Contents

General Introduction

Learning is a fundamental part of human nature. The knowledge acquired from learning new skills helps individuals in changing their cognitive and affective behaviour. Learning is the centre of human growth and development and it is hoped to be the mean for happiness, safety, emancipation, productivity, societal success. *Education*, as the set of all planned learning processes and activities, is the "mean by which men and women deal critically and creatively with reality and discover how to participate in the transformation of their world" (Freire, 1970).

Despite being so important in the development of an individual, learning is not always easy. In 1978, Vygotsky explained the difficulty of learning by introducing the *Zone of Proximal Development* (Vygotsky, 1978) indicating the psychological processes that the learner can reach with the support of knowledgeable guidance. According to Vygotsky, there are certain skills and competencies that the learner can only acquire if given the right support. With the right guidance, each learner can stretch outside of the *zone of comfort* and is able to experience and learn new skills and concepts. Besides external guidance, also internal factors play a determining role in learning success. Those are for example motivation to learn (Pintrich, 1999), the self-determination of an individual (Ryan and Deci, 2000) or meta-cognitive skills like self-regulation (Winne and Hadwin, 1998; Zimmerman, 2002) and the right set of dispositions (Buckingham Shum and Crick, 2012), skills, values and attitudes.

For several decades, educational researchers were busy understanding the "black box" of learning, trying to unveil the underlying dynamics and the factors that lead towards successful learning. More recently the education technology research community has been busy trying to understand the following question: *is there a place for technology to facilitate learning and teaching*?

In modern history, scientists have tried to apply technological tools as a means to investigate and understand complex natural phenomena. For example, in 1609, Galileo Galilei designed and implemented the first scientific telescope by which he admired the cratered surfaces of the moon or the details of the milky way. In 1665, the scientist Robert Hooke inspired the use of microscopes for scientific exploration, which paved the way to the theory of biological evolution. In 1822, Charles Babbage started working on the *Difference Engine*, the ancestor of modern calculators, a machine made to automatically compute the values of polynomial functions.

Scientists made use of technology to solve mathematical problems, understand the complexity of the universe, study the composition of natural elements and living creatures. Leveraging new technologies has also always been a valid research approach chosen by researchers to study and understand human learning.

The first massive implementation of digital technologies in education dates back to the mid-1980s, with the diffusion of the modern personal computer. American universities started sharing course content in the university libraries implementing the so-called *Computer-Based Learning*. Higher education institutions took advantage of the computer by developing distance courses and primitive forms of e-learning systems. In parallel to this, the 1980s saw also a “new spring” for *Artificial Intelligence* (AI) research. The invention of the *back-propagation rule*, which allowed *Artificial Neural Networks* to learn complex, non-linear problems, generated a new wave of enthusiasm. The 1980s were characterised by the surge of *Expert Systems*, computer programs typically written in LISP that modelled specific portions of knowledge. In the domain of education and training, these systems took the name *Intelligent Tutoring Systems (ITS)*, adaptive computer programs which aimed at providing rich interaction with the student (Yazdani, 1986; Anderson et al., 1985). The ITS research introduced the idea of the *Intelligent Tutor*, an intelligent algorithm able to adapt to the individual learner characteristics and that works as an “instructor in the box” (Polson et al., 1988) capable of replacing the human teacher. The AI-ITS vision was both controversial and technically complex to achieve for the 1980s. It did not fully take-off as much as other educational technologies such as e-learning.

In the 1990s, the e-learning systems took further steps of developments. E-learning became more popular as it was less ambitious and more applicable also to more ill-structured subjects, other than mathematics, programming or other natural science. E-learning became a tool that could support *Computer-Supported Collaborative Learning* (Dillenbourg, 1999). The computer in education shifted from being a knowledge diffusion system to a platform which encouraged the sharing and the development of knowledge between groups of learners.

In the 2000s, digital technologies met a fast development also thanks to the fast spreading of the internet and the *World Wide Web*. In education research, the *Technology-Enhanced Learning* (TEL) community emerged. The initial focus of TEL was on e-learning systems, learning objects and multimedia educational resources. While these educational contents were previously only accessible via a personal computer, in the late 2000s they became available for portable computing devices such as smartphones, tablets or laptops. These new technological affordances established the research focus on *ubiquitous and mobile learning* (Sharples et al., 2009), i.e. learning *anywhere at any time* without physical nor geographical location constraints.

In the 2010s, we observed a *data-shift* in education technologies with the rise of the *Learning Analytics* (LA) research community (Ferguson, 2012). The core idea at the basis of LA research was that learners interacting with computer devices leave behind a considerable number of *digital footprints* which can be collected and analysed for describing the learning progress and help to optimise it (Greller and Drachsler, 2012). Ten years after LA research was introduced, the field has moved forward considerably by identifying additional fundamental challenges (Selwyn, 2019). Despite the vast amount of data that can be collected, there is still some confusion on how these data can be harnessed to support learners. One part of LA research aims to foster *self-regulated learning* by stimulating learners to improve their meta-cognitive skills through self-reflection and social comparison with peer learners (Winne, 2017). Nevertheless, the common idea of simply providing learners with LA dashboards for raising their awareness does not seem to change their behaviour and meet their goals (Jivet et al., 2017). Other challenges LA deals with is how to ensure ethics and privacy (Drachsler and Greller, 2016)and how to change and inform learning design with learning analytics and data-driven methods (Schmitz et al., 2017).

Another limitation addressed by the LA community is related to the data source used. So far, the LA data are mostly related to learners interacting with a digital platform (e.g. Learning Management System) utilising mouse and keyboard. LA research – as well as its predecessors – were born nested into the *glass slab era*: the main learning and productivity tools are mediated by a computer screen, a mouse or a keyboard. With such tools, there is little space for interactions with physical objects in the physical world. The lack of physical interactions during learning led to a *reality drift* for learning science. According to the theory of *embodied cognition*, humans have developed their cognitive abilities together with the use of their bodies and that is encoded in the human DNA (Shapiro, 2019). For example, the hands are made for grasping physical objects, or the human senses developed for witnessing sound, smell or light. The limited data sources raise valid questions concerning the *understandability* and *interpretability* of the digital footprints analysed by LA researchers. Trying to derive meaning from limited educational data brings the risk of falling into the *street-light effect* (Freedman, 2010), the common practice in science of searching for answers only in places that are easy to explore.

To include novel data sources and new forms of interaction, a new research focus has emerged within LA research, coined as *Multimodal Learning Analytics* (MMLA) (Blikstein, 2013). The objective of MMLA is to track learning experiences by collecting data from multiple modalities and bridging complex learning behaviours with learning theories and learning strategies (Worsley, 2014). The *multimodal shift* is motivated from a theoretical point of view by the need to achieve more comprehensive evidence and analysis of learning activities taking place in the physical realm such as co-located collaborative learning (e.g. Pijeira-Díaz et al., 2018), psychomotor skills training (e.g. Schneider and Blikstein, 2015; Di Mitri et al., 2019b), dialogic classroom discussions (e.g. D'mello et al., 2015) which were underrepresented in LA research and other data-driven learning research. In parallel, the *multimodal shift* is also stimulated from a technological push given by the latest technology developments (Dillenbourg, 2016). Learning researchers are making use of new technological affordances for gathering evidence about learning behaviour. In recent years, the low costs of sensor devices made them more affordable. Sensors can be found embedded in smartphones, fitness trackers, wrist-based monitors or

Internet of Things devices and provide the possibility to continually measure human behaviour. These devices can collect data streams and measure life aspects such as hours and quality of sleep, working and productivity time, food intake, physiological responses such as heart-rate or electrodermal activity. Multimodal sensors can collect "social signals" – thin slices of interaction which predict and classify physical and non-verbal behaviour also in group dynamics. Multimodality is relatively novel in the field of learning. For this reason, we introduce the metaphor of the *unexplored land* which encloses the promise – or probably the hope – to better understand learning and human behaviour.

In the 2020s, a new kind of educational technology is taking off. With this doctoral thesis, we introduce it under the name of *Multimodal Tutor,* a new approach for generating adaptive feedback from capturing multimodal experiences. The Multimodal Tutor capitalises on the support of multimodal data for understanding learning and human behaviour, pushing it to the next level. It proposes a theoretical and methodological approach to deal with the complexity of multimodal data by combining the support of machine learning, artificial intelligence and human assessment. With this hybrid approach, the Multimodal Tutor carries an advanced promise for learners, making learning more authentic, adaptive and immersive. We argue the Multimodal Tutor may enable to move towards a *learner-centred* and *constructionist* idea of learning, as an active and contextualised process of construction of knowledge (Piaget, 1952). The multimodal approach is learner-centred as it focuses on the full span of human senses and embodied cognitive abilities. It moves away from non-natural interactions introduced by computers or smartphones and it stimulates interactions with the physical world. In the meantime, it tracks information about the learner's physiology, behaviour, and learning context.

The Multimodal Tutor advocates for reuniting two branches of developments in education technology which have been developing in parallel. The first one is Learning Analytics and TEL research that has been focusing primarily on deriving insights from learning data to support human decision making. The second one is AI-ITS research, which for almost three decades has designed, developed and tested artificially intelligent systems that model the knowledge of the learners and guide them through the learning activities domain.

Outline of the Thesis

This doctoral thesis is organised into four parts and seven chapters. Part I describes the "Exploratory mission", characterised by the experiment *Learning Pulse* described in Chapter 1. Learning Pulse unveils the complexity of using multimodal data for learning and paves the way to the Multimodal Tutor. Learning Pulse discovers empirically a series of complex dynamics, of both conceptual and methodological nature, derived by using multimodal data for predicting learning performance.

Part II provides a "Map of Multimodality". Enriched by several lessons learnt with the *Learning Pulse* study, in Chapter 2, we explore the concept of multimodality by analysing existing constructs and by conducting a literature survey. This qualitative research approach leads to the formulation of the *Multimodal Learning Analytics Model* (MLeAM), a conceptual model which serves as the "Map of Multimodality". The MLeAM sheds light on the multimodal feedback loop that the Multimodal Tutor is set to accomplish. However, if the MLeAM indicates the "way to go", it does not say "how to get there". There is, in fact, the need for a better understanding of the problem from a technological standpoint and the formulation of the possible solution. We describe this in chapter 3 with the "Big Five challenges" for the Multimodal Tutor. The size of the enterprise is then more clear, as much as its multifaceted complexity.

In Part III we reflect on the methodological approach needed to address the challenges identified in Part II. This results in the "Preparation of the Navy", a series of tools needed to be developed for our MMLA journey. There is a large expedition to realise, designing and implementing technical infrastructure able to follow the MLeAM, the "map" which led towards the Multimodal Tutor. From there originates the idea of the *Multimodal Pipeline* described in Chapter 5. The Multimodal Pipeline exploits the cyclic nature of the MLeAM and addresses the "Big Five" challenges with a technical infrastructure. The *Multimodal Pipeline* reveals to be the most critical part of the Multimodal Tutor research. The multimodal data streams are complex to align, synchronise and store. We identify a promising solution by combining the Multimodal Pipeline with an already existing MMLA prototype, the *Multimodal Learning Hub* (Schneider et al., 2018).

Chapter 4 focuses on one specific, unsolved aspect of the Multimodal Pipeline: the Data Annotation. From this challenge emerges the idea of creating a *Visual Inspection Tool*, an application for annotating and inspecting multimodal data streams, which allows to "read between the lines". After this achievement, the "navy" is prepared and ready to sail toward the new promising land to apply MMLA research. In this phase, we decide to narrow the focus to the specific domain of Cardiopulmonary Resuscitation Training (CPR). In Chapter 6 we focus on modelling the CPR domain, in particular how to detect multimodal mistakes using machine learning techniques.

Finally, Part IV describes the conclusive "conquest mission" of the CPR Tutor, an instance of the Multimodal Tutor. In Chapter 7, the CPR Tutor is employed in a field study for feedback generation. we report the design, development and experimental testing of the CPR Tutor.

Part I

Exploratory mission

Chapter 1

Learning Pulse

Learning Pulse explores whether using a machine learning approach on multimodal data such as heart rate, step count, weather condition and learning activity can be used to predict learning performance in self-regulated learning settings. An experiment was carried out lasting eight weeks involving PhD students as participants, each of them wearing a Fitbit HR wristband and having their application on their computer recorded during their learning and working activities throughout the day. A software infrastructure for collecting multimodal learning experiences was implemented. As part of this infrastructure, a Data Processing Application was developed to pre-process, analyse and generate predictions to provide feedback to the users about their learning performance. Data from different sources were stored using the xAPI standard into a cloud-based Learning Record Store. The participants of the experiment were asked to rate their learning experience through an Activity Rating Tool indicating their perceived level of productivity, stress, challenge and abilities. These self-reported performance indicators were used as markers to train a Linear Mixed Effect Model to generate learner-specific predictions of the learning performance. We discuss the advantages and limitations of the used approach, highlighting further development points.

This chapter is based on:

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., & Specht, M. (2017). Learning Pulse: a Machine Learning Approach for Predicting Performance in Self-Regulated Learning Using Multimodal Data. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference 2017* (LAK '17) (pp. 188–197). New York, NY, USA. ACM. DOI: 10.1145/3027385.3027447.

1.1 Introduction

The permeation of digital technologies in learning is opening up interesting opportunities for educational research. Flipped classrooms, ubiquitous and mobile learning as other technology-enhanced paradigms of instruction are enabling new data-driven research practices. Mobile devices, social networks, online collaboration tools as well as other digital media are able to generate a *digital ocean* of data (Dicerbo and Behrens, 2014) which can be "explored" to find new patterns and insights. The opportunities that data opens up are unprecedented to educational researchers as they allow to analyse and understand aspects of learning and education which were difficult to grasp before.

The disruption lies primarily in how the evidence is gathered: "data collection is embedded, on-the-fly and ever-present" (Cope and Kalantzis, 2015). Collecting data is not enough to extract useful information: the data must be pre-processed, transformed, integrated with other sources, mined and interpreted. Reporting on historical raw data only does not bring, in most of the cases, added value to the final user. As Li points out (Li, 2015) individuals are already exposed to so many data they risk to "drawn" into data. What is instead more desirable is receiving support *in-the-moment* which can prescribe positive courses of action, especially for twenty-first-century learners which need to orient themselves continuously in an ocean of information with very little guidance (Ferguson and Shum, 2012).

Machine learning and predictive modelling can play a major role in extracting high-level insights which can provide valuable support for learners. Such ability highly depends whether the attributes taken in consideration to describe the learning experiences (the *Input space*) are descriptive for the learning process, they carry enough information to be able to accurately predict a change in the learning performance (the *Output space*). The relation between these two dimensions is further described in section 1.3.1.

The standard data sources in the reviewed predictive applications are most of the time Learning Management Systems (LMS) and the Student Information Systems. Looking only at clickstreams, keystrokes and LMS data alone gives a partial representation of the learning activity, which naturally occurs across several platforms (Suthers and Rosen, 2011). Several authors have pointed out the need to explore data "beyond the LMS" (Kitto et al., 2015) to be able to get more meaningful information on the learning process. We believe that an interesting alternative could be found in the Internet of Things (IoT) and sensor community. Schneider et al. 2015a have listed 82 prototypes of sensors that can be applied for learning. The employment of IoT devices allows collecting real-time and multimodal data about the context of the learning experience.

These considerations have shaped the motivation for the Learning Pulse experiment. The challenges it seeks to answer are the following: (1) define a set of data sources "beyond the LMS"; (2) find an approach to couple multimodal data with individual learning performance; (3) design a system which collects and stores learning experience from different sensors in a cloud-based data store; (4) find a suitable data representation for machine learning; (5) identify a machine learning model for the collected multimodal data.

Learning Pulse's main contribution to the Learning Analytics community consists in outlining the main steps for a new practice to design automated multimodal data collection to provide personalised feedback for learning with the ultimate aim to facilitate prediction and reflection, the two most relevant objectives of learning analytics (Greller and Drachsler, 2012). This proposed practice borrows the modelling approach from the machine learning field and uses it to model, investigate and understand human learning.

1.2 Related Work

Learning Pulse belongs to the cluster of Predictive Learning Analytics applications. The scope of this sub-field in Learning Analytics was framed by the American research institute Educause with a manifesto (ECAR, 2015) reporting some example applications, including Purdue's Signals (Arnold, 2010) or the Student Success System (S3) by Desire To Learn (D2L) (Essa and Ayad, 2012). These applications rely solely on LMS data for predicting academic outcomes or student drop-outs. Learning Pulse goes beyond those Predictive Analytics Applications by using multimodal data from sensors to investigate the learning process.

The field of multimodal data was given more prominence in the last Conference Learning Analytics and Knowledge (LAK16) with the workshop *Cross-LAK: learning analytics across physical and digital spaces* (Martinez-Maldonado et al., 2016). The concept behind Learning Pulse was presented at the Cross-LAK workshop (Di Mitri et al., 2016). In this workshop, several topics were touched: data synchronisation (Echeverría et al., 2016), technology orchestration (Martinez-Maldonado, 2016) or face to face collaboration settings (Wong-Villacres et al., 2016).

With a mission similar to Learning Pulse, a data challenge workshop on Multimodal Learning Analytics (MLA'16) took place at LAK'16 for investigating learning happening on the physical or virtual world through multimodal data including speech, writing, sketching, facial expressions, hand gestures, object manipulation, tool use, artefact building.

Finally, there has been a paper by Pijeira Diaz et. al (Pijeira-Díaz et al., 2016) who used mutimodal data for Computer-Supported Collaborative Learning in a school setting. Although not focused on using machine learning, the link made with *psychophysiology* theory introduces a novel research question, i.e. the possibility to infer psychological states including cognitive, emotional and behavioural phenomena from physiological responses such as sweat regulation, heartbeat or breath (Cacioppo et al., 2007).

1.3 Method

The background exposed in the previous chapter has led to the formulation of an overarching research question:

> How can we store, model and analyse multimodal data to predict performance in human learning? (**RQ-MAIN**)

This main research question leads to three sub-questions:

(**RQ1**) Which architecture allows the collection and storage of multimodal data in a scalable and efficient way?

(**RQ2**) What is the best way to model multimodal data to apply supervised machine learning techniques?

(**RQ3**) Which machine learning model is able to produce learner specific predictions on multimodal data?

To further investigate these research questions, we designed the Learning Pulse experiment that involved nine PhD students as participants and generated a multimodal dataset of approximately ten thousands records.

1.3.1 Approach

While frameworks already exist for standard *within-the-LMS* Predictive Learning Analytics, e.g. the PAR Framework (Wagner and Davis, 2014), there are no structured approaches to treat *beyond-the-LMS* data in the context of multimodal data. For this reason, in this work, a novel approach for predictive applications inspired by machine learning is proposed. The objective is to learn statistical models out of the learning experiences and outcomes. Using a mathematical formalism that corresponds to learning a function f in the equation $y = f(X)$, where X is a vector containing the attributes of one learning experience which work as the input of the function and, y is a particular learning outcome.

By using such an approach, three elements need to be further clarified: (1) the scope of the investigation (the *learning context*); (2) the attributes encompassed by multimodal data (the *Input space*); (3) the learning performance object of the predictions (the *Output space*).

Learning context

The learning context investigated is *self-regulated learning* (SRL) which is defined as "the active process whereby learners set goals for their learning and monitor, regulate, and control their cognition, motivation, and behaviour, guided and constrained by their goals and the contextual features of the environment" (Pintrich Zusho, A., 2007). Self-regulated learners are able to monitor their learning activity by defining strategic goals and that drive them not only to academic success but lead to increased motivation and personal satisfaction (Zimmerman, 2002). There is an overarching difference between self-regulated and non-self-regulated learners: the former are generally more engaged with their learning activities and desire to improve their learning performance (Butler and Winne, 1995). On the contrary, the latter ones are less experienced, they do not perceive the relevance of their learning program and, for this reason, need to be followed closely by a tutor.

Input space

Learning is a complex human process and its success depends on several endogenous (e.g. psychological states) and exogenous factors (e.g. learning contexts). Defining the *Input space* consists of selecting the relevant attributes of the learning process and structuring them into a correct data representation. This modelling task is non-trivial: according to Wong (Wong, 2012) modern "seamless" learning encompasses up to ten different dimensions. In this project, two of them are of main interest: *Space* and *Time*. The Input space can be imagined as the sequence of events happening throughout the learning time across digital and physical environments as shown on the left of figure 1.1.

Learning in a *digital space* means "mediated by a digital medium" i.e. by technological devices like laptops, smartphones or tablets. Digital learning data are easier to collect as most of the digital tools leave traces of their use. On the contrary, learning happening in the *physical space* refers to the learning not mediated by digital technology, like 'reading a book' or 'discussing with a peer'. Although the line between Digital and Physical gets blurred with the pervasiveness of technology, the bulk of the learning activities still happens *offline* and should be "projected into data" through a sensor-based approach to be able to take advantage of those moments.

Time is also a relevant dimension: the *data-driven* approach works best whenever the data collection becomes continuous and unobtrusive for the learner. This requirement inevitably limits the scope of investigation only to tangible events whose values are easy to measure over time. If on the one hand, this constraint makes data collection easier as there is no need to employ time-consuming surveys and questionnaires, on the other hand, this approach does not make it possible to directly capture psychological states which manifest during the learning.

Besides spanning across physical and digital space, the *Input space* of Learning Pulse can be grouped into three layers as shown in figure 1.1: those are (1) *Body* encompassing physiological responses and physical activity, (2) *Learning Activities* (3) and *Learning Context*.

Output space

The *Output space* of the prediction models corresponds to the range of possible learning performances. These outputs are crucial for the machine learning algorithms to distinguish between successful learning moments from the unsuccessful ones. As self-regulated learners decide on their own learning goals and required learning activities, we need performance indicators which go beyond common course grades.

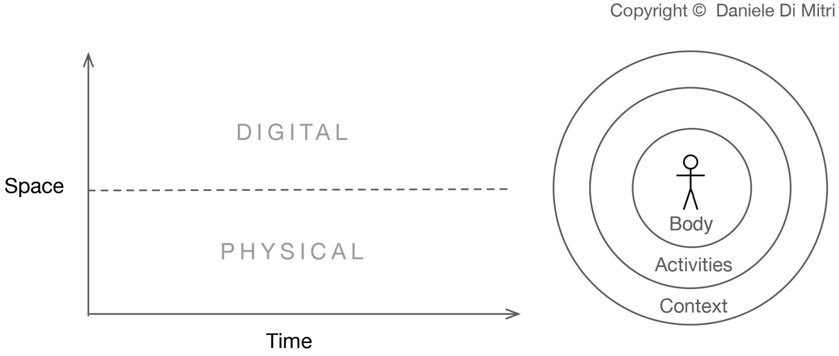


Figure 1.1 Bi-spatial and three-layered Input Space.

An interesting approach to measure learning productivity is the concept of Flow theorised by the Hungarian psychologist Csikszentmihalyi. The Flow is a mental state of operation that individuals experience whenever they are immersed in a state of energised focus, enjoyment and full involvement with their current activity. Being *in the Flow* means feeling in complete absorption with the current activity and being fed by intrinsic motivation rather than extrinsic rewards (Csikszentmihalyi, 1997). In the model theorised by Csikszentmihalyi depicted in figure 1.2, the Flow naturally occurs whenever there is a balance between the level of difficulty of the task (the challenge level is high) and the level of preparation of the individual for the given activity (the abilities are high).

To measure the Flow we applied experience sampling (Larson and Csikszentmihalyi, 1983): the participants reported about their self-perceived learning performance. As self-assessment is strictly subjective it has the advantage to be exclusively based on the learner's personal feelings. If carefully designed, self-assessment can lead to models tailored on personal dispositions. This brings clear advantage in the context of self-regulated learning: what is perceived as good (or productive, stressful etc.) is classified as such, meaning that *what is good* is only *what the learner thinks is good*.

1.3.2 Participants and Tasks

The experiment took place at the Welten Institute of the Open University of the Netherlands involving nine doctoral students as participants, five males and four females, aged between 25 and 35 with a background in different disciplines including computer science, psychology and learning science. PhD students are good self-regulated learners, as they are generally experienced learners and have strong engagement and motivation with their tasks.

All participants were provided with a Fitbit HR wristband and installed the tracking software on their laptops. As sensitive data were collected, every participant signed

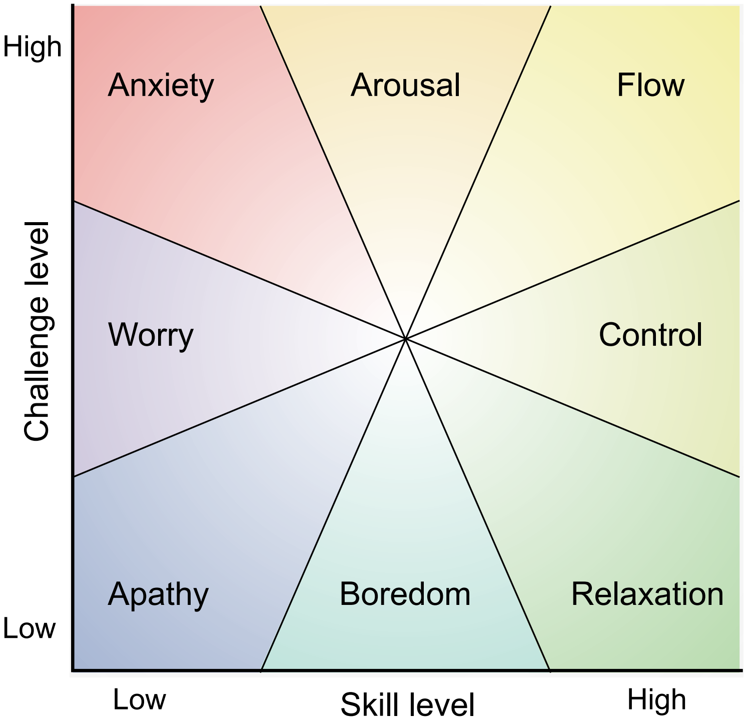


Figure 1.2 Csikszentmihalyi's Flow model.

an informed consent. In addition, to ensure their privacy, their data were anonymised making use of the alias *ARLearn* plus an ID between 1 and 9.

The experimental task requested from the study participants was to continue their typical research activity throughout the day: the only additional action consisted in rating their learning activity every working hour between 7 AM and 7 PM (for the number of hours they worked) through the *Activity Rating Tool* (described in sec. 1.3.4).

The actual experiment lasted for eight weeks and consisted of three phases: 0) Pre-test, (1) Training and (2) Validation.

Phase 0: *Pre-test* System infrastructure was tested in all its functionalities. A presentation was rolled out to introduce the experimental setting and the study's rationale to the participants. Participants were instructed to set-up the data collection software on their laptop as well as the fitness wristband.

Phase 1: *Training* The first phase of the experiment lasted three weeks and consisted of the rating collection: participants have rated their activities hourly. The only visualisation they could see at that point was the ratings during that day. The first phase was named *training* because the collected data and ratings were necessary to train the predictive models.

Phase 2: *Validation* After two weeks of break, the second phase started lasting for another two weeks. In the *Validation* phase, the activity rating collection continued in a Learner Dashboard visualisation. The second phase was called Validation as its purpose was to compare the predicted performance indicators with the actual rated ones and to determine the prediction error.

1.3.3 Data sources

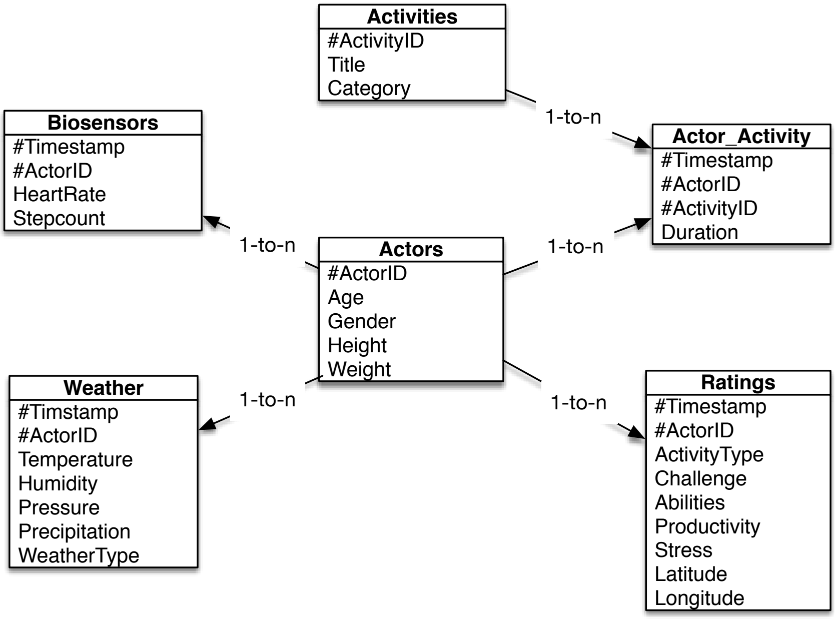


Figure 1.3 The Entity-Relation model of the data.

Biosensors

The physiological responses and physical activity (*Biosensor data* for short) in this study are represented by heart rate and step count respectively. The approach used to track these “bodily changes” consisted of making use of wearable sensors. The decision of the most suitable wearable tracker was dictated by following criteria: (1) heart rate tracking sensor; (2) price per single device; (3) accuracy and reliability of the measurements; (4) comfort and unobtrusiveness; (5) openness of the APIs and data for analysis.

The choice converged to *Fitbit Charge HR*[1]: standing out on the cost-quality trade off, Fitbit HR complied with all the requirements, in particular by offering open access to the collected data through the Fitbit API. Such a way of accessing data was beneficial on the one hand, as the software application developed for the project had to communicate exclusively with the Fitbit cloud datastore - while being agnostic to sensor trackers and their interfaces. The downside, on the other, hand was the dependence to the API specifications: the maximum level of detail available was a heart rate value updates every five seconds and step count update every minute.

It is relevant to point out the difference of the heart rate and step count signals: while the heart rate values are a *continuous* time-series, also called fixed event, the number of steps per minute is a *random event* as it represents a *voluntary human activity* and not an *involuntary process* as the heartbeat. The value of step count at one time point is not dependent on the previous ones (i.e. is random) while the heart rate value at time t surely depends on the value at time $t - 1$.

Learning Activities

To monitor self-directed learning we decided to track PhD students' activities on their laptops, being those the main learning medium in which they perform their PhD activities. Given the variety of learning tasks executed by the participants during the experiment, the actual learning happens across different platforms including software applications, websites, web tools. To capture and represent this heterogeneous complex of digital activities a software tracking tool was installed on the working laptop of the participants. The idea is that the use of particular software or application adds up a valuable piece of information to consider when abstracting the learning process.

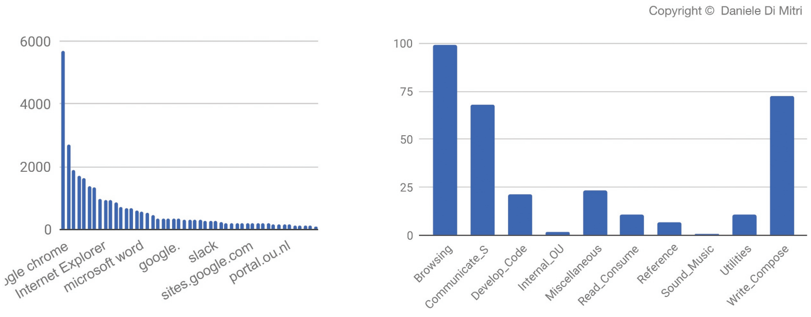
The tool chosen to monitor working efficiency was *RescueTime*, a time management software tool. RescueTime stores an array containing the applications in use by the learner, weighted by their duration in seconds, into a proprietary cloud database every five minutes (maximum level of detail allowed by its API specifications). Each activity in one interval has an activity ID and duration in seconds. The duration ranges between 1 and 300 (max seconds in five minutes), as the zero-valued entries are the applications not used in an interval.

Given the diversity of research topics and learning tasks there is a high intersubject difference on the set of applications used during the learning experience; apart from a few common applications, the majority of applications used are very sparse. To mitigate this problem applications were grouped into categories by hand. The name of the categories chosen were: (1) *Browsing*, (2) *Communicate and Schedule*, (3) *Develop and Code*, (4) *Write and Compose*, (5) *Read and Consume*, (6) *Reference Tools*, (7) *Utilities*, (8) *Miscellaneous*, (9)*Internal Open Universiteit*, (10) *Sound and Music*.

In figure 1.4, the distribution of the applications is compared with their categories. The height of the bars represents the number of executions that application had

[1] https://www.fitbit.com/chargehr

during the experiment, which equals to the presence of that application in one of the five-minute intervals. While in the left-hand chart the *long tail effect* due to the sparsity is quite noticeable, on the right hand side that does not appear.



Copyright © Daniele Di Mitri

Figure 1.4 Plots showing the number of executions per Applications (left), per Application category (right).

Performance indicators

The indicators used in Learning Pulse are four: Stress, Productivity, Challenge and Abilities. The four indicators were collected with the following questions.

1. **Stress**: how stressful was the main activity in this time frame?
2. **Productivity**: how productive was the main activity in this time frame?
3. **Challenge**: how challenging was the main activity in this time frame?
4. **Abilities**: how prepared did you feel in the main activity in this time frame?

Each participant had to rate each of these indicators retroactively with respect to the main activity performed in the time frame being rated. The participants were expected to answer these questions at the end of every working hour from 7AM to 7PM using for each of them a slider in the *Activity Rating Tool* described in section 1.3.4 which translated the rating into an integer ranging from 0 to 100.

The Flow The Flow is operationalised through a single numerical indicator calculated based on the Challenge and Abilities indicators. i identifies a specific learners, while j references a specific time frame. F_{ij} is the Flow score for the learner i^{th} at the time frame j^{th}; A_{ij} and C_{ij} is the level of Abilities and Challenge rated by the learner i^{th} at the time frame j^{th}.

$$F_{ij} = (1 - |A_{ij} - C_{ij}|) * \frac{|A_{ij} + C_{ij}|}{2} \tag{1.1}$$

Figure 1.5 plots the ratings of all the participants throughout the whole experiment in a two-dimensional space, where the x-axis are the level of Abilities and the y-axis is the level of Challenge. Both indicators are expressed as percentages. The dots in the scatter plot are coloured depending to their Flow-value calculated with the Equation 1.

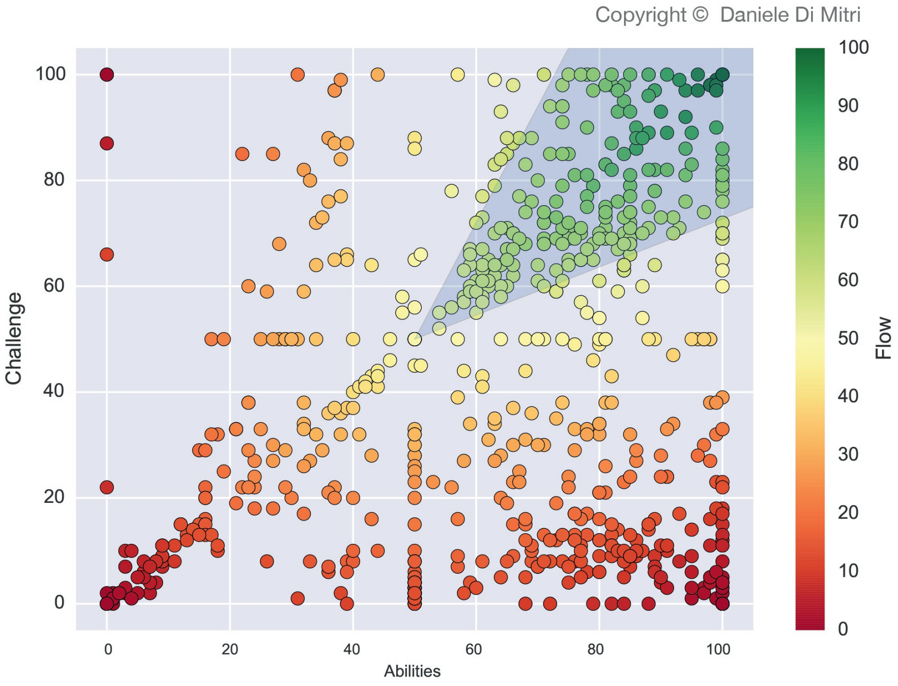


Figure 1.5 Scatter plot of the Flow of all study participants.

The colour scale used for the Flow goes from red over yellow to green recalling the metaphor of a traffic light: high Flow values are green, medium ones are yellow and low Flow values are red. The plot visualises how the equation of the Flow works. The Flow is higher if two conditions apply: (1) the difference between Abilities and Challenge is small, meaning they are close to line $x = y$; (2) the mean between Abilities and Challenge is close to one, meaning the observation falls into the top-right corner of the plot, which corresponds to the Flow zone, as in the original definition of Flow (see figure 1.2).

Besides the four questions also the *Activity Type* was sampled along with the GPS coordinates. The Activity Type was a categorical integer representing the following labels (1) *Reading*, (2) *Writing*, (3) *Meeting*, (4) *Communicating*, (5) *Other*.

The rationale behind this labelling was to have a hint on the nature of the main learning task executed during that time frame. Finally, the GPS coordinates consisted of two floating points which are the latitude and longitude of the location where the

Copyright © Daniele Di Mitri

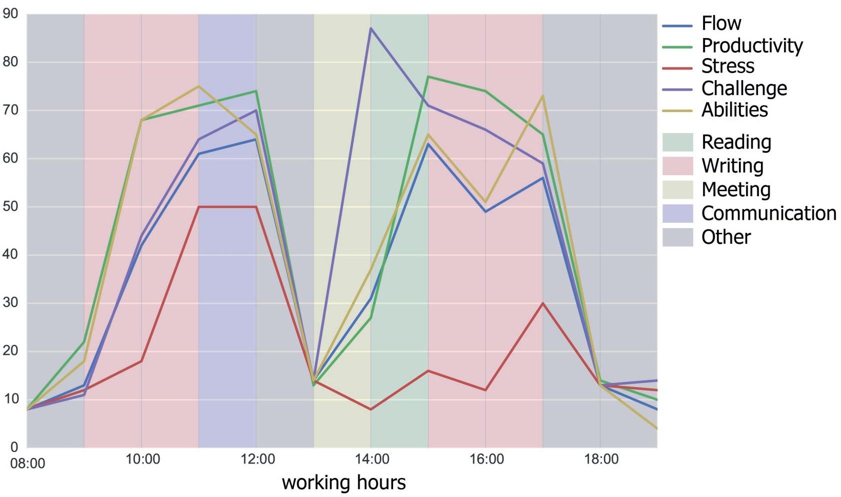


Figure 1.6 Plot showing the ratings given by one participant in one day.

rating was submitted with the *Activity Rating Tool*.

Figure 1.6 shows the ratings of the four indicators of one participant during one day of the experiment, as well as the calculated Flow indicator. The background colours represent the different activity types, as the legend visually indicates.

Environmental context

The third data source is made up of the surrounding context of learning as the environment might also have an impact on the final learning outcomes. The ideal solution would be to track information about the indoor surrounding environment, such as measuring the light intensity, humidity and heat inside the office, thus combining these with the information about the weather.

Given the lack of adequate sensors to employ in the office environment, only the outdoor weather conditions were monitored. For each participant, the GPS coordinates were stored that allowed to call the weather data API through the online service *OpenWeatherMap*[2] and to store weather data specific to the location from where each participant was operating. The weather API was called automatically every ten minutes for each of the nine participants. The attributes extracted from these statements were (1) Temperature, (2) Pressure, (3) Precipitation, (4) Weather Type, with the first three being floating points while the latter is a categorical integer.

[2] https://openweathermap.org/

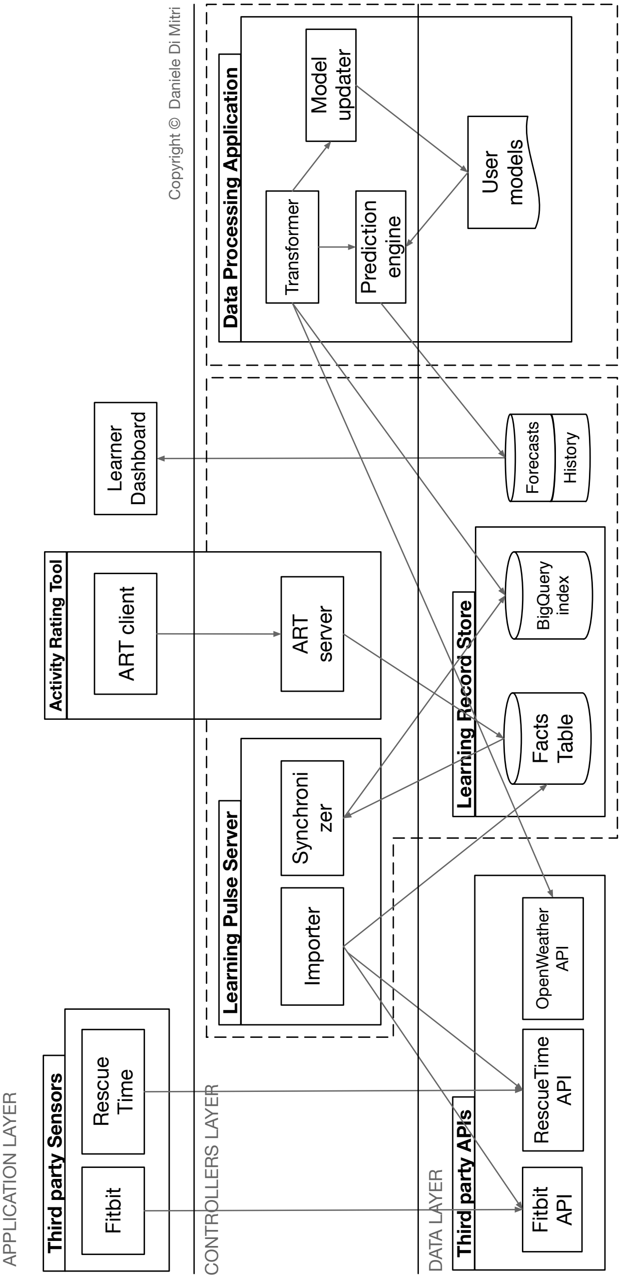


Figure 1.7 System architecture of Learning Pulse.

1.3.4 Architecture

Combining different Data Sources into a central data store and processing them in real-time is not a trivial task. Figure 1.7 presents a transversal view of the system architecture which is divided into three layers.

At the top level, the *Application Layer* groups all the services that the end-user interfaces with including the Fitbit wristband and the RescueTime application here called to *Third Party Sensors*. The Activity Rating Tool (ART) belongs to the same level.

The middle level is the *Controllers Layer* which gathers the back-end components of the Applications. In this layer, as figure 1.7 shows, the software is running on two server infrastructures: the *Cloud* and the *Virtual Machine*. Not reported here are the controllers of the Third Party Sensors and the Learner Dashboard as the System Architecture described here is agnostic towards their implementation. On the Cloud side, there are the *Learning Pulse Server*, a scripting software responsible for importing data from different APIs and storing them into the *Learning Record Store*. In addition, also running on the Cloud, there is the server software of the Activity Rating Tool which connects the client user interface with the database. The scripting software running on the Virtual Machine is the Data Processing Server, which as the name indicates, implements the post-processing operations including data transformation, model fitting and predictions.

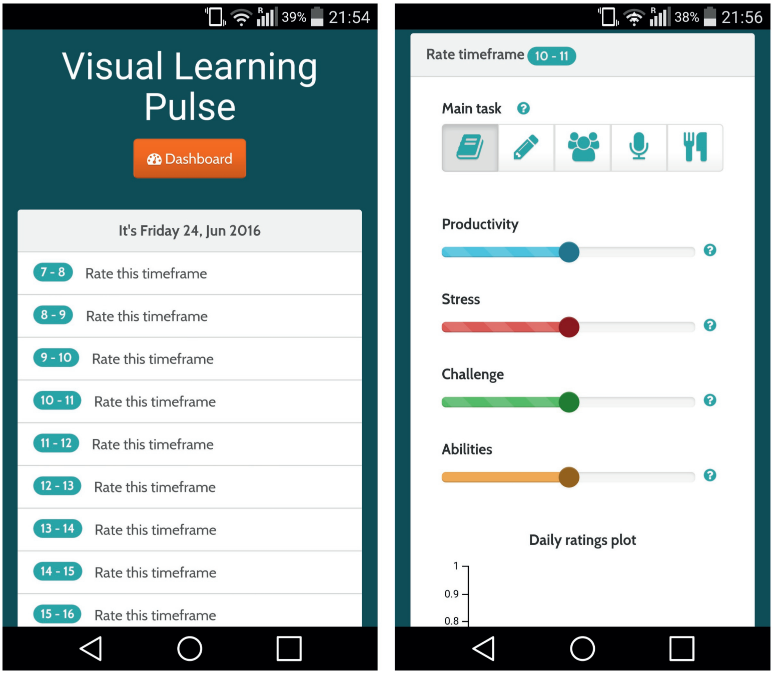
The lowest level is the *Data Layer*. While the Third Party Services use their own APIs which receive regular queries by the importers of the Learning Pulse Server, the main datastore is the Learning Record Store. Consisting of a Fact Table and a Big Query Index, the Learning Record Store is the cloud-based database which collects the data about the learning experience of all participants. It also runs on the Cloud infrastructure and is further described in section 1.3.4.

Even though they are not directly part of the Learning Record Store, also the results of the Data Processing server are pushed into a datastore which is also shown in the Data Layer. This datastore is developed with a non-relational database and collects the predictions (also referred as forecasts) and the transformed representation of the historical data, namely the learning experience data in the Learning Record Store opportunely processed and transformed. Finally, the Data Processing Server makes use of further persistent data, for example the Learners' Models, which are stored locally, reused constantly and regenerated once a day.

Activity Rating Tool

Responsible for collecting the participants' ratings about their learning experience, designed and developed as a scalable web application, the Activity Rating Tool runs App Engine using webapp2 lightweight Python web framework. While the back-end was written in pure Python, the front-end uses Bootstrap[3].

[3] http://getbootstrap.com/



It's Friday 24, Jun 2016

7 - 8 Rate this timeframe
8 - 9 Rate this timeframe
9 - 10 Rate this timeframe
10 - 11 Rate this timeframe
11 - 12 Rate this timeframe
12 - 13 Rate this timeframe
13 - 14 Rate this timeframe
14 - 15 Rate this timeframe
15 - 16 Rate this timeframe

Figure 1.8 Two screenshots of the *Activity Rating Tool*: on left side the list of time frames available for rating, on the right the rating form of a time frame.

The interface of the tool was designed to be as intuitive as possible and to make the rating action quick and easy for the participants considering they needed to use it several times a day. Figure 1.8 shows two screenshots of the application's main page; on the left-hand side, it shows the list of all the past time frames between 7 AM and the hour previous to the current. To rate a time frame the form shown on the right-hand side of figure 1.8 opened. The users are asked to select the Activity Type through five different icons; below, users can input the rating for the four indicators through four sliders, differently coloured for each indicator. Once the desired values are chosen, the sliders translate the position of the slide into an integer between 0 and 100. To prioritise straightforwardness and to avoid information overload, the guiding questions were hidden into a help tool-tip at the right-hand side of the sliders.

Once the participant pressed “Submit" the time frame turned green coloured in the time frame list. The participant could also delete ratings or resubmit in case of errors. Additionally, a Daily Rating Plot is shown just before the “Submit" button which shows the past ratings recorded that day with the purpose of reminding the participant their previous ratings that day in order to support a coherent overall rating.

Learning Pulse Server

The Learning Pulse Server is the script component responsible for pulling the data from the third party APIs and transforming them into learning records and handing out their identifiers. The learning records are first stored into the Fact Table by assigning a UUID (Universally Unique Identifiers). The Learning Pulse Server script and the Fact Table were implemented as application and data store in the Cloud, which allowed to balance the data load on a distributed architecture for scalability purposes. From the Fact Table, the data were synchronised into a Query Index, implemented with a scalable non-relational database, which contrarily to the Fact Table, allowed to query the distributed learning statements with SQL language. The synchronisation between the Fact Table and the Query Index happens using a queue, such that no learning record could get lost.

While the Learning Pulse Server is the application script responsible for pushing and pulling the learning records, the Fact Table and the Query Index together form the LRS. Implementing the LRS with a cloud-based solution allowed to achieve properties such as (1) high availability: the LRS could be reached at any time, with respect to the privileges of the client; (2) high scalability: although the size of the data collected was about 1 Gigabyte the number of learning statements could easily scale up tens or even hundreds of times more; (3) high reliability: the cloud infrastructure chosen provided performance and security.

Experience API

The chosen data format for the learning records was the Experience API (or xAPI) data standard, an open-source API language through which systems send learning information to the LRS. XAPI is a RESTful web service, with a flexible standard which aims at interoperability across systems. The XAPI standard has the format *actor-verb-object* and is generated and exchanged in JSON format, opportunely validated by and stored in the LRS. The main advantage of using xAPI is interoperability: learning data from any system or resource can be captured and eventually queried by the third party authenticated services. For each event captured in Learning Pulse, an xAPI statement template was designed following the Dutch xAPI specification for learning activities (Berg et al., 2016) [4].

1.3.5 Data processing

After being stored in the LRS, learning records were processed, transformed and mined to generate predictions to be shown to the learners. Data collection and Data processing can be seen as two legs which walk side by side, complementing each other's role. The data processing software was named *Data Processing Application*[5] (DPA) and its main responsibilities consisted of (1) fetching the data from the

[4] A list of the statements can be found here http://bit.ly/DutchXAPIreg

[5] The source code of the Data Processing Application is available at https://github.com/WELTEN/learning-pulse-python-app

Learning Record Store; (2) transforming the new data by time resampling and features extraction; (3) learning and exploiting different regression models; and (4) storing the results of the regression.

The DPA needed to run continuously on a server *always-on* without the need for human interaction. Other important requirements for the DPA were the possible integration with other software components (e.g. interfacing with the LRS) and availability of statistical and Machine Learning tools. The final choice converged on using Python as the main programming environment, mainly because of its flexibility and wide support for data analysis.

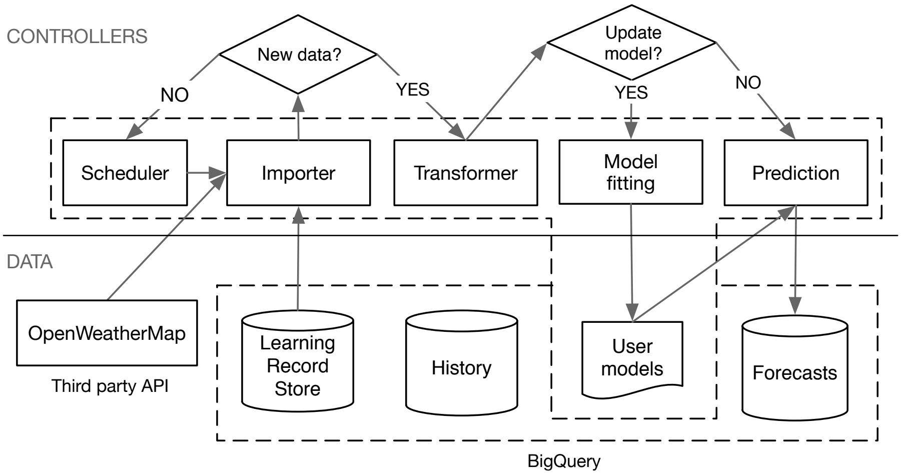


Figure 1.9 The data processing workflow.

For the Data Processing Server, namely the computer infrastructure which hosted the DPA, cloud options were considered including popular cloud IaaS solutions. For financial reasons, the choice directed towards an in-house server solution constituting of a Virtual Machine running an OpenSuse Linux distribution.

The diagram in figure 1.9 shows the data processing workflow, a close-up of the system architecture shown in section 1.3.3. The figure is divided into three layers: the controllers, the data and the visualisations.

Data fetching

A cron-job on the Virtual Machine activated the scheduler every ten minutes, every working day, from 7AM to 7PM. The main task of the scheduler was to query the Learning Record Store and to realise whether new intervals could be formed based on the learning records retrieved. In order to be valid, the learning intervals have to be completed for Biosensor, Activity and Weather data. If any of these data are not available, the execution of the Data Processing Application is interrupted and postponed to the next round. To connect to the Learning Record Store, the DPA uses Pandas' Big Query connector. This interface can authenticate the client (the DPA Python script) to the Big Query service, submit a query and fetch the results that are returned into a *data frame*, the popular data format for structuring tabular data in Pandas.

Multi-instance representation

Each data source had its frequency of data generation: the ratings were submitted every hour, the heart rate was updated every five seconds, the step count every minute, the activities every five minutes and the weather every ten minutes. That resulted in the so-called *relational representation* as for each participant a different number of relations corresponded with all the other entities depending on how frequent their values were updated. Relational representations are not ideal for machine learning as the input space which needs to be examined can become very broad (De Raedt, 2008).

The problem was therefore translated into a *multiple instance representation* where each training sample is a fixed-length time interval. The interval length is determined by how frequently the labels i.e. the ratings, are updated. As the ratings here equal the working hours (say 8 hours), if multiplied by the experiment days (say 15), that would result in the best-case scenario of 120 samples for each participant, which is too small in size for a training set. To overcome this problem the compromise was found selecting 5 minutes long intervals. This decision, however, triggered another problem, what to do with those attributes that are updated more or less frequently. The approach used was different for each entity. Ratings, which are updated hourly, were linearly interpolated; the step count, which is updated every minute, was aggregated with a sum function; the weather, which was updated every 10 minutes, was copied backwards; the activities came already with a five minutes frequency, therefore no action was required. Finally, to represent a five minutes heart rate signal into one or more features, the best solution was to use different aggregate functions, namely: (1) the minimum of the signal, (2) the maximum, (3) the mean, (4) the standard deviation and (5) the average change - i.e. the mean of the absolute value of the difference between two consequent data points. This naive approach consists of plugging in several different features and letting the machine learning algorithm decide which ones are the most influential in predicting the output. It is, however, useful to point out that more sophisticated techniques for feature extraction from the heart rate exist, such as the Heart Rate Variability (Wang and Huang, 2012) or the Sample Entropy.

Data storing

Similarly to the data collection, also the data processing had to be the same. In order not to repeat the processing step of the same data multiple times, it was convenient to store the results of the transformation in a permanent data store, to be able to retrieve it when necessary. To do so a Big Query table was created called *History*:

the name was used to differentiate the transformed historical data with the forecast about the future, whose table is called *Forecasts*.The Big Query was preferred over other solutions since the LRS was developed with the same technology. In addition, Pandas offers an easy Big Query interface, which allows pushing and pulling data easily from the Cloud Database.

1.3.6 Regression approach

As the collected data were longitudinal, the fixed effects showed stochastic behaviour implying that the observations were highly dependent on one another. In formal terms, this means that observing the behaviour of one participant at time t, the output variable y_t is described by the equation $y_t = \alpha + \beta X_t + e_t$. The dependence among the samples means that given a later observation at time $t+1$, the covariance $cov(e_t, e_{t+1}) \neq 0$ with $t \neq t+1$.

As the samples were intercorrelated it was not possible to employ common regression models, as most of these techniques assume that the residuals are independent and identically distributed normal random variables. Treating correlated data as if they were independent can yield wrong *p-values* and incorrect confidence intervals. To overcome this problem the chosen approach was the *Linear Mixed Effect Models* (LMEM).

LMEM relax the dependency constraint of the data and they can both treat data of mixed nature, including fixed and random effects, plus they describe the variations of the response variables with respect to the predictor variables with coefficients that can vary for each group (Lindstrom and Bates, 1988). In formal terms, the LMEM as described by (Laird and Ware, 1982) consist in a n_i-dimensional vector y for the i-th subject:

$$y_i = X_i\beta + Z_i\gamma_i + \epsilon_i, i = 1, ..., M\gamma_i \sim N(0, \Sigma) \tag{1.2}$$

• n_i is the number of samples for subject i
• Y is a n_i dimensional vector of response variables
• X is a $n_i \times k_{fe}$ dimensional matrix of fixed effects coefficients
• β is a k_{fe}-dimensional vector of fixed effects slopes
• Z is a $n_i \times k_{re}$ dimensional matrix of random effects coefficients
• γ is a $k_{re}-$dimensional random vector with mean zero and covariance matrix; each subject gets its independent γ
• ϵ is a n_i-dimensional *within-subject* error with mean 0 and variance Σ^2 with a spherical Gaussian distribution.

1.4 Analysis and Results

At the end of the experimental phase, the transformed dataset presented the following characteristics: a total of 9410 five-minute learning samples, counting for all nine participants. The biggest sample size was *ARLearn5* with 1725 samples, while the one with the smallest number of samples was *ARLearn4* with 514. There were 29 attributes in total.

As a single-output LMEM implementation was chosen, five different models were learnt each of them having as response variable one of the five performance indicators (Abilities, Challenge, Productivity, Stress and Flow). The models were initialised with the following parameters:

• **Fixed Effects:** timeframe, latitude, longitude, weatherConditionId, pressure, temp, humidity, hr_min, hr_avc, hr_mean, hr_std, hr_max

• **Random Effects:** Browsing, Communicate_Schedule, Develop_Code,Internal_OU, Miscellaneous, Read_Consume,Reference, Sound_Music, Utilities, Write_Compose, Steps.

As the rating style of each participant was different, the predicted values were normalised with respect to the learner-specific historical min and max using the following formula.

$$x_{new} = \frac{(x_{max} - x_{min}) * x_i}{100} + x_{min}$$

For the evaluation of the predicted results, we used R-squared, a statistical measurement which scores how close the data are to the regression line and outputs a number from 0 and 1 which measures the *goodness-of-fit* of the model. The results obtained were the following: Stress: 0.32, Challenge: 0.22, Flow score: 0.16, Abilities: 0.08, Productivity: 0.05.

1.5 Discussion

The first question (RQ1) focused on the best architectural setup to process multimodal data. The answer found to the question was satisfactory as architecture design discussed in section 1.3.3 was capable of: (1) importing a great number of learning statements from the sensors and their APIs; (2) feeding the statements into a cloud-based LRS avoiding collisions among them and information loss; (3) combining the statements with the self reports regularly provided by the learners; (4) programmatically transforming the learning statements by extracting relevant attributes and by re-sampling into uniform intervals; (5) fitting the predictive model on historical observations and saving for the reuse with the newer observations and (6) saving the predictions in a separate store to be able to compare with the actual values. On the other hand, the architectural design had some limitations. First of all, it exhibited a real-time syncing issue: the data synchronisation with the wearable trackers was slower than expected; in the best-case scenario, the data about the heart rate and the steps were available in the LRS only 15 to 20 minutes later. Secondly, the Data Processing Server hosting the Data Processing Application was poor in performance: the weak processing power slowed down the data processing and that resulted in long job cycles.

The second research question (RQ2) was concerned with finding the best way to model multimodal data suitable for machine learning. The solution found was to treat the problem using a *Multiple Instance Representation* as detailed in section 1.3.5, i.e. using a tabular representation where each row represents a five-minute learning interval and each column a different attribute. This representation helped to overcome the problems derived from the relational nature of the collected data. Additionally, third party APIs influenced a lot the type of data that is possible to be retrieved from the sensors. An example is the Fitbit Charge HR, whose API only allows to get values of the heart rate every five seconds and no inter-beat distance. This scarcity of available data did not allow to calculate useful measurements on the heart rate, like the Heart Rate Variability which has been proven to be a good predictor for workload stress (Taelman et al., 2009).

The third research question (RQ3) asked which machine learning model for regression is best suited for the heterogeneous type of data. The solution discussed in section 1.3.6 consisted in using the *Linear Mixed Effect Models* as they allow (1) taking into account data specific to each learner; (2) distinguishing between *fixed* and *random* effects; (3) taking categorical data into account. Despite LMEM being the appropriate model for the intended task, the R-squared evaluation test yielded poor prediction accuracies for the five outputs. One possible reason might be the sparsity of random effects, especially those that refer to the least used activity categories (whose distribution is shown in figure 1.4). We observed that while adding up sparse attributes (random effects) as predictors decreases the prediction accuracy, fixed effects improve the general accuracy.

The answers to the three sub research questions provide an answer to the main research question (RQ-MAIN): a way to store, model and analyse multimodal data was successfully found. Nevertheless, the limited significance of the prediction results does not allow us to assert that accurate and learner-specific predictions can be generated. This might have been caused by (1) the combination of multimodal data selected in the experiment; (2) no clear learning task to be executed, high variance of the learning context explored; (3) sparse random effects were still too many as opposed to fixed effects.

1.6 Conclusions

This paper described Learning Pulse, an exploratory study whose aim was to use predictive modelling to generate timely predictions about learners' performance during self-regulated learning by collecting multimodal data about their body, activity and context. Although the prediction accuracy with the data sources and experimental setup chosen in Learning Pulse led to modest results, all the research questions have been answered positively and have led towards new insights on the storing, modelling and processing multimodal data.

We raise some of the unsolved challenges that can be considered a research agenda for future work in the field of Predictive Learning Analytics with “beyond-LMS” multimodal data. The ones identified are: (1) the number of self-reports vs unobtrusiveness; (2) the homogeneity of the learning task specifications; (3) the approach to model random effects; (4) alternative machine learning techniques.

There is a clear trade-off between the frequency of self-reports and the seamlessness of the data collection. The number of self-reports cannot be increased without worsening the quality of the learning process observed. On the other side, having a high number of labels is essential to make supervised machine learning work correctly.

In addition, a more robust way of modelling random effects must be found. The found solution to group them manually into categories is not scalable. Learning is inevitably made up by random effects, i.e. by voluntary and unpredictable actions taken by the learners. The sequence of such events is also important and must be taken into account with appropriate models.

As an alternative to supervised learning techniques, also unsupervised methods can be investigated, as with those methods fine graining the data into small intervals does not generate problems with matching the corresponding labels also the amount of labels is no longer needed.

Regarding the experimental setup, it would be best to have a set of coherent learning tasks that the participants of the experiment need to accomplish, contrarily to as it was done in Learning Pulse, where the participants had completely different tasks, topics and working rhythms. It would be also useful to have a baseline group of participants, which do not have access to the visualisations while another group does have access; that would allow seeing the difference of performance, whether there is an actual increase.

To conclude, Learning Pulse set the first steps towards a new and exciting research direction, the design and the development of predictive learning analytics systems exploiting multimodal data about the learners, their contexts and their activities to predict their current learning state and thus being able to generate timely feedback for learning support.

Part II

Map of Multimodality

Chapter 2

From Signals to Knowledge

Multimodality in learning analytics and learning science is under the spotlight. The landscape of sensors and wearable trackers that can be used for learning support is evolving rapidly, as well as data collection and analysis methods. Multimodal data can now be collected and processed in real-time at an unprecedented scale.With sensors, it is possible to capture observable events of the learning process such as learner's behaviour and the learning context. The learning process, however, consists also of latent attributes, such as the learner's cognition or emotions. These attributes are unobservable to sensors and need to be elicited by human-driven interpretations. We conducted a literature survey of experiments using multimodal data to frame the young research field of multimodal learning analytics. The survey explored the multimodal data used in related studies (the input space) and the learning theories selected (the hypothesis space). The survey led to the formulation of the Multimodal Learning Analytics Model whose main objectives are of (O1) mapping the use of multimodal data to enhance the feedback in a learning context; (O2) showing how to combine machine learning with multimodal data; and (O3) aligning the terminology used in the field of machine learning and learning science.

This chapter is based on:

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4), 338–349. DOI: 10.1111/jcal.12288

2.1 Introduction

With the rise of data-driven techniques to discover insights and generate predictions from the learning process such as learning analytics, the need for 360 degrees data about learners has grown consistently. Combining data coming from multiple sources has become a prominent necessity in learning research and has led to an increased interest in multimodality and consequently into multimodal data analysis. To clarify the concept of multimodality, we use the definition provided by Nigay and Coutaz. The term “multi” refers to “more than one”, whereas the term “modal” stands both for “modality” and for “mode”. The modality is the type of communication channel used by two agents to convey and acquire information that defines the data exchange. The mode is the state that determines the context in which the information is interpreted (Nigay and Coutaz, 1993). The reasons why multimodality in learning is drawing so much attention can be summarized according to four developments.

First of all, multimodality is a consolidated theory. It has been subjected of investigation already for two decades in different fields including functional linguistic, conversational analysis, and social semiotics (Jewitt et al., 2016). Research in multimodal interaction investigated how different modalities interact and complement each other to convey and densify meaning (Norris, 2004). Different experiments using multimodal data in learning scenarios also date back to the early 90s. In 1993, Ambady and Rosenthal found out that college teachers were able to predict students’ end–of-semester results by observing “thin slices” of interactions, that is, looking at their physical and non-verbal behaviour with short video clips (Ambady and Rosenthal, 1992). These early findings paved the way towards a new research hypothesis, the possibility to infer cognitive and social processes by using multiple data sources and social signal processing (Poggi et al., 2012).

Second, multimodal tracking has recently become more feasible. This happens because of recent technological developments such as the Internet of Things, wearable sensors, cloud data storage, and increased computational power for processing and analysing big data sets. To date, sensors can be used to gather high-frequency and fine-grained measurements of micro-level behavioural events as, for example, movement, speech, body language, or physiological responses. The Internet of Things approach, that is, connecting sensors to physical world objects or to human bodies, allows computers to take measurements of the world as well as the physiological phenomena, encoding them into machine-interpretable data.

Third, modelling across physical and digital worlds is a rising need. A general “call formultimodality” has been fostered in computer-supported collaborative learning and learning with interactive surfaces communities (Schneider and Blikstein, 2015). Multimodal data systems are needed to link digital and physical interactions and shed a light on collaborative learning and collective sense-making (Martinez et al., 2011; Pijeira-Díaz et al., 2016). Sensors and wearable trackers can be used in learning settings to collect attributes from face-to-face physical learners’ interactions, such as speech, body movement, and gestures. These bodily micro-actions can be combined with digital interactions recorded with tabletops and stored in log files. A similar need exists in the Learning Analytics & Knowledge, for achieving a more complete picture of the learning process. Such need originates from the fact that traditional data sources, like logs, clickstreams, and content interactions taking place within the learning management system, only represent a small proportion of the learning activities and not the whole learning process (Pardo and Kloos, 2011). Multimodal data, in summary, can mitigate the streetlight effect[1], by adding more street lights and expand the visible area and complete the learner's digital profile in the computer (Heckmann, 2005).

Finally, the multimodal approach is more aligned with the nature of human communication. The use of multiple modalities in human communication is redundant and complementary (Calvo et al., 2015). This reflects also when the human interacts with the computer. Humans communicate their intentions and emotions using multiple modalities such as facial expression, voice intonation, or body movements. When analysing incomplete data sets, especially those having missing data (e.g., due to hardware failures), the information overlap across multiple modalities is convenient because it allows their overall meaning to be preserved (Bosch et al., 2015).

The developments here described paved the way for a new approach in data-driven learning support, that is, the multimodal learning analytics (Blikstein, 2013). MMLA is a research field located at the crossroad between learning science, machine learning. MMLA leverages the advances in multimodal data capture and signal processing to investigate the learning in complex learning environments (Ochoa and Worsley, 2016). MMLA can establish a bridge between complex learning behaviour and learning theories (Worsley, 2014). MMLA can offer new insights into learning spaces and tasks in which learners have open choices to differentiate their learning trajectories by facilitating the provision of feedback (Blikstein, 2013).

Despite the increased interest that the MMLA research field is receiving, it remains a new kind of "data geology," which faces several challenges. Some of these challenges are inherited by the complex and multiform nature of multimodal data. On this extent, the most relevant multimodal data challenges were described by Lahat et al. (2015) and include high-dimensionality, different modality resolutions, noise, missing data, data fusion techniques, and choice of the right model.

MMLA also faces challenges specific to its application domain of education and learning. In this article, we aim to get an overview of the MMLA field and its challenges. First, we proposed a classification framework for MMLA research consisting of input space and hypothesis space divided by the observability line. Thereafter, we conducted a literature survey where we explored MMLA empirical studies (Section 2.2), and we further operationalized the input and the hypothesis spaces. The literature survey helped to identify three main challenges in the field of MMLA: (C1) There is a lack of understanding of how multimodal data relate to learning and how these data

[1] The streetlight effect describes the common practice in science of searching for answers (i.e., the lost key) only into places that are easy to explore,s that is, the streetlights (Freedman, 2010)

can be used to support learners achieving the learning goals; (C2) it is still unclear how to combine human and machine interpretations of multimodal data; and (C3) the fields of machine learning and learning science use different terminologies that are ambiguous and need to be aligned. The surveyed literature allowed us to go a step further and address the exposed challenges by introducing the Multimodal Learning Analytics Model (MLeAM, Section 2.3). MLeAM was designed to fulfil three objectives: (O1) mapping the use of multimodal data to enhance the feedback in a learning context; (O2) showing how to combine machine learning with multimodal data; and (O3) aligning the terminology used in the field of machine learning and learning science.

2.2 Literature Survey

In this section, we first describe the classification framework (Section 2.2.1) used to conduct the literature survey in the field of MMLA. We detail the two main components of the classification framework: the input space and the hypothesis space. In Section 2.2.2, we describe the selection process and criteria adopted to identify the relevant articles. In Section 2.2.3, we present the results of the survey by proposing the taxonomy of multimodal data for learning and the classification table of the hypothesis space. Lastly, in Section 2.2.3, we discuss the results, and we draw the conclusions in terms of future challenges for the MMLA community.

2.2.1 Classification framework

Some aspects of the learning process such as the learner's behaviour can be directly observed and measured by means of sensors. Some other aspects, such as learner's cognition or emotions, are latent attributes that cannot be directly measured by sensors and thus can only be inferred. For our literature survey, we named these aspects as input space and hypothesis space, which is a distinction widely used in machine learning. In the case of human learning, the input space includes, for example, the learner's behaviour and the learning context. These aspects of learning can be captured automatically into multimodal data. It is relevant to point out that sensors have a different viewpoint than humans; sensors are not capable of making interpretations or assigning meaning to the data they collect. The hypothesis space encompasses the range of possible interpretations, that is, attributes not directly observable by sensors but that can also be expressed as data. The hypothesis space includes semantic interpretations of the multimodal data,which can be based on psychological and learning-related constructs such as emotions, beliefs, motivation, cognition, or learning outcomes. These attributes belong to the learner's sense-making process, which in classroom activities remains invisible for educators and researchers (Kim et al., 2011).

The input and hypothesis spaces are therefore conceptually separated by the observability line: a line of separation between the observable evidence and all the possible interpretations. The attributes of both spaces are facets of the same iceberg,

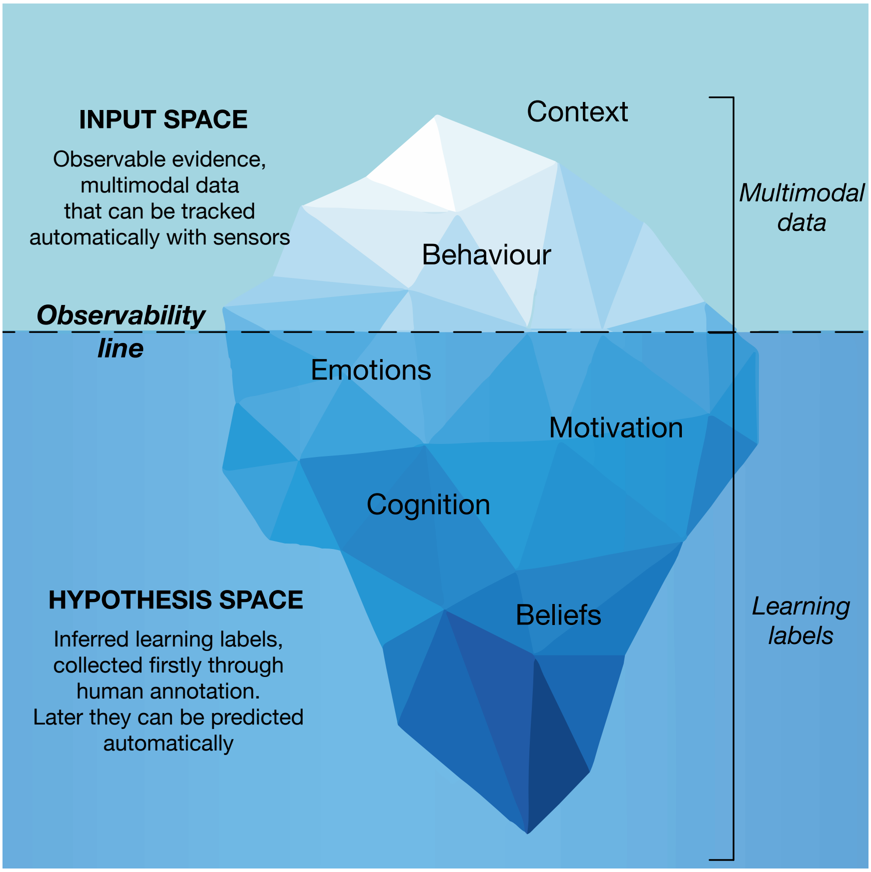

INPUT SPACE
Observable evidence, multimodal data that can be tracked automatically with sensors
Context
Behaviour
Multimodal data
Observability line
Emotions
Motivation
Cognition
Beliefs
HYPOTHESIS SPACE
Inferred learning labels, collected firstly through human annotation. Later they can be predicted automatically
Learning labels

Figure 2.1 The observability line: The multimodal data can capture only the observable attributes.

the ones "above the waterline" are noticeable from the point of view of a generic sensor. While, the attributes "underwater", require multiple levels of interpretation, depending on how deep they stand from the observability line. The distinction between observable/unobservable is conceptual and can vary in practice. Figure 2.1 presents one possible instantiation of this concept. The distinction is useful when employing sensors and using machine-guided interpretations. For computers, the interpretation process, that is moving from the input to the hypothesis space, is increasingly difficult.

Although input and hypothesis spaces are separated for computers and sensors, they are tightly intertwined for humans. Humans can interpret behavioural cues, by reasoning and drawing conclusions, for example, yawning corresponds to boredom or tiredness. Psychological and educational theories tell us how these relationships can be drawn. For example, the *affective-behaviour-cognition* theory connects observed behaviour with emotions and cognition (Ostrom, 1969). Similarly, Damasio proposed the idea of "somatic markers" that are special instances of feelings in the body associated with emotions such as rapid heartbeat is associated with anxiety or nausea is associated with disgust (Damasio et al., 1991). At the biological level, the process of self-regulation as a response to physical and external demands is known as homeostasis, which supports the idea of the human body working as a complex system. An example of this homeostasis for learning is the state of arousal known as the degree of physiological activation and responsiveness caused by a situation or collaborative activities (De Lecea et al., 2012). Low arousal is an indication for a harmful physiological state for learning such as frustration or boredom, whereas high arousal indicates an active or responsive mode that is supportive for learning (Bjork et al., 2013; Pijeira-Díaz et al., 2018).

Input space: Multimodal data

Learning is a complex and multidimensional process (Wong, 2012). Defining the input space, that is, identifying the relevant modalities and extracting informative attributes, is not trivial tasks. To facilitate these tasks, we expand the initial notion of multimodal data for learning by describing their distinctive features.

An important requirement to be fulfilled is that the modalities must be periodically measurable. To explain this, we pick the counterexample of biomarkers testing extensively employed in medicine (Koh and Jeyaratnam, 1998). Analysing samples of blood, body fluids, or tissue, biomarker tests can be used to investigate the genomic structure, the presence of molecules or hormones concentration like dopamine or norepinephrine (noradrenaline). The presence or lack of one of these substances can indicate potential disease or a certain body state. The way these tests are conducted does not allow for continuous measurements and monitoring: For this reason, these dimensions are out of the scope of multimodal data analytics.

The modalities belong to the input space and can be either endogenous or exogenous (behaviour vs. context), depending on if they explain the learner's behaviour or the learning environment affordances that are external but might influence the learning process. The behavioural modalities can be divided between motoric and physiological. Motoric modalities are movements and describe events mainly governed by the somatic nervous system and actuated by the muscles and the skeleton. These modalities are generally deliberated, they should be seen as random events, as there exists no evident correlation between consequent values. Conversely, the physiological modalities that are governed by the autonomic nervous system are generally involuntary, their role is to help to self-regulate, and they should be seen as continuous events. An example is the cardiovascular activity controlled by the heart: The value of the heart rate at one-time point is dependent on the previous values and, for this reason, must fall into a range.

The division between intentional and unintentional events is however not so black and white as it seems. Anderson (2002) illustrates how humans have different levels of cognition like biological, cognitive, rational, or social, and human actions can be classified accordingly to these levels depending on which timescale they take place. The reaction time for an action can span from microseconds for biological reactions to minutes, hours, days, or weeks for social actions. The rational and social actions are pondered; they require enough time to go through different layers of consciousness; for this reason, they are associated with a higher level of intentionality. One example can be standing in a very hot room with closed windows: A common unintentional biological reaction can be starting to sweat, whereas a common rational action can be opening the window. Both actions and reactions can be considered self-regulatory.

At the biological level, the Neurovisceral Integration Model supports the idea of coordination: It shows how the human body works as a complex interconnected system, adapting its functioning according to the stimuli it receives and to goals it wants to reach (Thayer et al., 2009). For example, the mind under effort is associated with physiological arousal and therefore with increased heart rate. The discipline that studies the correlations between physiological activity and psychological states including cognitive, emotional phenomena is called psychophysiology.Cacioppo et al. (2000) found interesting correlations between heart rate accelerations and emotions such as anger, fear, and sadness. For example, an increase in heart rate variability (HRV) is correlated with joy and amusement, whereas the decrease of HRV is correlated with happiness.

Another distinction that can be made is between verbal and nonverbal modalities. Non-verbal expressions are thought to make up to 93% of the meaning during face-to-face communication and social interaction (Mehrabian, 1971). In particular, kinesics, commonly referred as body language and physical appearance, is thought to have an important role, especially during learning. Teachers, for instance, often use kinesics to reinforce the meaning of the words (Leong et al., 2015). Verbal modalities, on the other hand, use natural language as communication and, for this reason, have a much higher interpretation complexity. For an intelligent computer, it is way more complicated to make sense of the meaning of what one person is saying (or writing and drawing) as compared with how she is saying it. The surveyed studies using speech modalities focus on prosodic features rather than discourse analysis.

Hypothesis space: Learning theories

The hypothesis space, a term which is widely used in inductive logic and machine learning, specifies the range of possible states of a phenomenon. In the case of the MMLA field, the hypothesis space lists all the possible interpretations that can be assigned to the observed learning process and are driven by validated learning theories or by psychological constructs. One state in the hypothesis space is a unique value combination of the attributes describing a phenomenon. The learning states are represented by data through the learning labels. The learning labels are typically assigned by human inference to specific time intervals of multimodal data recordings. The act of repeatedly assigning learning labels to multimodal data intervals is called an annotation. The annotation is often the only way to provide the baseline to multimodal data, that is, the truth values that will be used to train the machine learning models and test their accuracy. A careful definition of the hypothesis space weighs a lot in the optimal success of the data-driven solution. Defining the hypothesis space consists in three points: (a) defining actionable components; (b) selecting the most appropriate data representation for the learning labels; and (c) devising an annotation strategy.

Defining actionable components for the hypothesis space The size of the hypothesis space is proportional to its descriptive power, that is, the number of possible interpretations that it describes, but it is inverse proportional to its generalizability. This is the well-known bias-variance trade-off (Friedman, 1997). One good heuristic for deciding the most useful hypothesis space is thinking in terms of actionability. The predicted state in the hypothesis space should support the design of valuable and actionable feedback for the learner. Hence, the hypothesis space specification must be guided by the question: "what is relevant for the learner to know to improve the performance?" The answer to this question is not trivial and can be properly addressed with careful feedback design (Hattie and Timperley, 2007). The machine learning models, alongside predicting the learning labels in the hypothesis space, can contribute by determining the attribute importance (e.g., how much a modality weigh in for the prediction). The attribute importance is the extent by which each attribute contributes to predicting the learning labels in the hypothesis space that can be used for targeted suggestions. Multimodal data can also provide historical values records and can shed a light on the historical changes in both the input and the hypothesis spaces. Predictions, attribute importance, and the historical multimodal records are three integrative elements that can enhance the learner's feedback.

Data representation of the learning labels From a data representation point of view, the learning labels of the hypothesis space can be represented as binary variables (e.g., focused vs. not focused) and can be specified in a numerical scale or as discrete categories (e.g., bored, engaged, and confused). The number of required learning labels depends on the size of the input space, that is, the number of attributes selected by the multiple modalities. In general, the number of labels required to properly run supervised machine learning is still quite high, for example, thousands of labels per individual learner. Many researchers in the machine learning field are currently researching techniques based on transfer learning to minimize the problem of the required labels, for example, using techniques such as pretraining with unlabelled data (Pan and Yang, 2010). The frequency of the annotations can also vary from 10 s to hours.

Annotation strategy Generally, there are two approaches for annotating multimodal data recordings: The first is asking experts to provide the learning labels and the second is asking the learner to fill self-reports on a regular or random basis. Both approaches come with their set of pros and cons, and both are subject to bias. One advantage of using external experts could be not to interfere with the natural task execution flow during learning; the con is that experts are expensive and hard to organize. Self-reports, instead, produce imbalanced class distribution (Hussain et al., 2012), which require some down-sampling approach, which means losing data. Self-reports, however, can be given in-the-moment, leveraging the short memory and, for this reason, producing more trustworthy reports compared with retrospective ratings (Edwards et al., 2017).

2.2.2 Literature survey selection process

Using the concept of observability line, we conducted a literature survey of empirical studies in the field of MMLA. The survey was first aimed to discover both the most frequent modalities and learning theories used in MMLA research and therefore the existing patterns and commonalities in the definition of the input and hypothesis spaces. In this survey, we identify representative MMLA studies, and we used them to specify our conceptual model for MMLA (see Table 2.3). The selected articles were found by going through all the papers of the last 5 years' Learning Analytics & Knowledge conference proceedings (2014–2018), the six editions of the MMLA Data Challenge workshop series (2013–2018), the Learning Analytics Across Physical and Digital Spaces workshop series (2016–2018), and additional publications by influential researchers in the MMLA field. We filtered the retrieved studies by applying two selection criteria: (a) the data set analysed in the studies was generated using more than one modality and (b) the multimodal data were linked to clear learning theories. We obtained a subset of 20 empirical studies fulfilling these criteria. We consider this number to be sufficient for getting an overview of the field; however, we foresee an increase of similar studies in the future.

2.2.3 Results of the literature survey

Following the description of the input space (Table 2.2.1) and the hypothesis space (Table 2.2.1), we further operationalize both spaces with insights gained from the literature survey: in Section 2.2.3, the *Taxonomy of multimodal data for learning* and Section 2.2.3, the *Classification table of the hypothesis space*.

Taxonomy of multimodal data for learning

The Taxonomy of multimodal data for learning is the first approach to organize the complexity of the observable modalities (input space), which can be monitored by sensors and are mentioned in the surveyed studies. This taxonomy is not meant to be an exhaustive classification of the modalities for learning or a technical review of different sensor types. For the latter, we refer to the review of Schneider et al. (2015a) that provides an extensive list of sensors that can be applied in the domain of education.

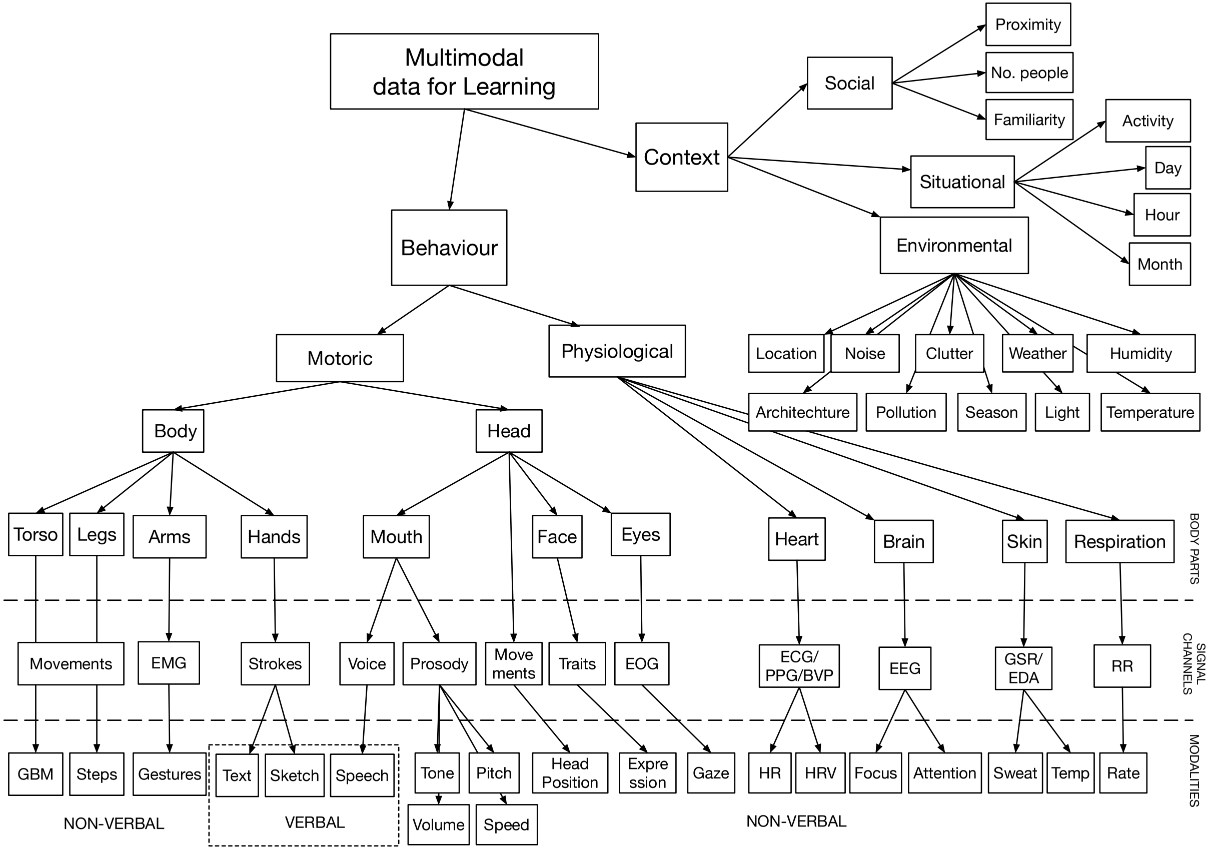


Figure 2.2 Taxonomy of multimodal data for learning. EMG: electormyogram; ECG: electrocardiogram; PPG: photoplethysmography; EEG: electroencephalogram; GSR: galvanic skin response; GBM: gross body movement; HR: heart rate; HRV: heart rate variability; EOG: Electrooculogram; BVP: Blood volume pulse; EDA: Electrodermal activity; RR: Respiration rate.

The taxonomy is presented from the perspective of a generic sensor. The underlying idea is that a sensor can monitor one (or multiple) modalities. We consider here the modality as a measurable property belonging to a specific part of the body or the context. The modalities are communicated through signal channels. Continuous sampling of a signal channel leads towards the longitudinal collection of one (or multiple) modalities. For instance, a microphone (sensor) can sample the voice (channel) to detect speech (modality), or a video camera can track at the same time voice, movements, and facial traits and therefore provide speech, gross body movements (GBMs), and facial expressions. To make an overview of the proposed taxonomy, we analysed the two main branches: (a) behavioural motoric and (b) behavioural physiological modalities by providing meaningful examples found in the surveyed literature of multimodal experiments. For the third main branch, (c) the contextual modalities, we remind to the work of Zimmermann et al. (2005), who propose a framework for context-aware systems in ubiquitous computing that combines personalisation and contextualization.

For simplicity, the motoric modalities can be split between the ones concerning the “body” or the “head.” Part of the subcategory body is the torso, legs, arms, and hands. The movements of the torso can provide GBM, which is typically derived from video cameras. GBM was used by Raca and Dillenbourg (2014) in their study for assessing students’ attention from their body posture, gesturing, and other cues. Similarly, Bosch et al. (2015) used GBM to detect learners’ emotions in combination with facial expression and learning activity. Although movements of the legs can be trackedwith step counters and provide good indicators for physical activity, arms and hands are body parts richer in meaning. Movements of the arms can be detected by video cameras, a popular choice, in this case, is Microsoft Kinect, for gestures and body postures recognition; several studies opted for this solution in the survey, especially those focusing on presentation skills (Barmaki and Hughes, 2015; Echeverría et al., 2014; Schneider et al., 2015b). An alternative to arm movements and gestures can be traced with electromyographic sensors (EMG): Hussain et al. (2012), for instance, used EMG in their study in emotion detection. Finally, hands are probably the parts of the body that can provide the best insights on the learner’s activity: Hands movement can be traced in search for specific hand signs or to track the handling of objects as well as pen strokes or drawings. For instance, Oviatt (2013) gathered a data set known as Math Data Corpus in which they combined analysed pen strokes with modalities captured from video and speech records in group settings with the aim to detect expert from non-expert students.

The motoric modalities of the head include analysis of the facial expressions, eye movements, and speech analysis. These three body parts can provide relevant information to the point that three well established research communities are dedicated to advancing the techniques and methodologies for data acquisition. Facial expressions are highly investigated in learning for emotion recognition in the affective computing research and have been quite extensively used in multimodal human-computer interaction experiments (Alyuz et al., 2016; Bosch et al., 2015; Hussain et al., 2012, e.g.). Eye-tracking is commonly used as an indicator for learners’ attention has also been used with multimodal data sets (Edwards et al., 2017; Prieto et al., 2016; Raca and Dillenbourg, 2014). Finally, an analysis of the speech spans from paralanguage analysis like speaking time, keywords pronounced, or prosodic features like tone and pitch (Prieto et al., 2016) to actual recognition of spoken words in dialogic settings like student-teacher interactions (D'mello et al., 2015). In theory, speech recognition opens up the possibility to transcribe discourse and use natural language processing to look for deeper level semantic interpretations. In practice, due to its high-level technical complexity, discourse analysis is a frontier that we envision in multimodal learning but which has been not yet explored in related works.

The physiological modalities can be also divided into corresponding body parts. For instance, heart, brain, and skin are the main organs of which is possible to derive physiological information. The most popular approaches to detect brain activity is the electroencephalogram (EEG), which measures the difference of potential inside of the brain. EEG was used by Prieto et al. (2016) in combination with eye tracking, from a teacher analytics perspective to predict social plane of interaction and concrete teaching activity. Different techniques can be used to calculate measurements of the heart activity like the heart rate and HRV: the electrocardiogram (ECG) or the photoplethysmography. Galvanic skin response (GSR), also referred as electrodermal activity (EDA), is the measure of electrical conductance of the skin. If the body receives stimuli that are physiologically arousing, the skin conductance increases. Arousal is widely considered to be one of the two main dimensions of an emotional response. Alzoubi et al. (2012) used the combination of EEG, ECG, and galvanic skin response to detecting naturalistic expressions of affect. EDA was used by Pijeira-Díaz et al. (2016) in combination with BVP, heart rate, skin temperature, and pupil size. Heart rate has been used by Di Mitri et al. (2017) to predict Flow in combination with steps and activity data. Edwards et al. (2017) used EDA to detect presence and lack of attention. Hussain et al. (2012) combined ECG, EMG, EDA, and respiration with video features to predict emotions. Also, Grafsgaard et al. (2014) used multimodal analysis to predict emotions combining EDA (skin conductance) with facial expression derived from video, gestures, and posture.

Classification table of the hypothesis space

Table 1 provides a summary of the learning theories found in the selected studies that used multimodal data. The table classifies the studies according to the chosen theoretical construct, hypothesis space specification, data representation type, and annotation method, and it provides a reference to study.

Table 2.1 Classification table of the hypothesis space (Note. N.A.: not applicable).

Construct	Hypothesis space	Representation type	Annotation method	Used by
Emotions in learning	Low, medium, high valence	Numerical	Self-reports with video records	(Hussain et al., 2012)
	Satisfied, bored, confused	Categorical	N.a.	(Alyuz et al., 2016)
	Boredom, confusion, curiosity, delight, flow and surprise.	Categorical	Self-reports	(Alzoubi et al., 2012)
	Confidence, frustration, excitement, and interest	Categorical	Self-reports	(Arroyo et al., 2009)
	Boredom, confusion, delight, engaged concentration and frustration.	Categorical	N.a.	(Bosch et al., 2015)
	Happiness, sadness, surprise, fear, disgust, anger, and neutral.	Categorical	Mimicking	(Bahreini et al., 2015)
	Engagement, frustration, learning	Categorical	Self-reports	(Grafsgaard et al., 2014)
Flow	Flow 0 to 100	Numerical	Self-reports	(Di Mitri et al., 2017)
Attention	N.a.	Categorical	Self-reports	(Raca and Dillenbourg, 2014)
Relevance of the lecture	N.a.	Categorical	Self-reports	(Raca and Dillenbourg, 2014)
Action codes	Build, Plan, Test, Adjust, Undo	Categorical	Clustering	(Worsley and Blikstein, 2013)
Activity types	N.a.	N.a.	Expert	(Prieto et al., 2016)
Social plane of interaction	N.a.	N.a.	Expert	(Prieto et al., 2016)
Expertise	Expert, non-expert; High, medium, low.	Categorical ordinal	N.a.	(Ochoa et al., 2013) (Worsley and Blikstein, 2013)
Activity performance	Good, bad	Categorical	N.a.	(Echeverría et al., 2014)
Cognitive load	Low, high	Categorical ordinal	Expert	(Eveleigh et al., 2010)

The most advanced studies using multimodal data focus on predicting emotions. Emotions as they are considered readouts of physiological changes in the body, changing as the response to certain stimuli. According to the Somatic Marker Hypothesis, physiological changes occur in the body and are passed to the brain when they are interpreted as emotions (Damasio et al., 1991). People adapt their environment and emotional stimuli via the autonomic nervous system responses (Kemper and Lazarus, 1992). It is possible therefore to correlate certain autonomic nervous system activity to emotional states. Emotions are thought to have also an important role in learning (Boekaerts, 2010). Typical emotions during learning are confusion, boredom, engagement, curiosity, interest, surprise, delight, anxiety, and frustration (Hussain et al., 2012). D'Mello (2013) provided a meta-analysis of the incidence of emotions during learning.

A psychological construct used is the *Flow*, a mental state of operation that individuals experience whenever they are immersed in the state of energized focus, enjoyment, and full involvement with their current activity. "Being in the flow" means feeling in complete absorption with the current activity and being fed by intrinsic motivation rather than extrinsic rewards (Csikszentmihalyi, 1997). The Flow naturally occurs whenever there is a balance between the level of difficulty of the task and the level of preparation of the individual for the given activity.

Another construct found in the literature is the one of *Cognitive load* refers to the demands within working memory that occur during learning: Too little load fails to engage learners sufficiently, whereas too much load overruns the capacity of working memory (Van Merriënboer and Sweller, 2005). Eveleigh et al. (2010) measured cognitive load of basketball players using speech during think-aloud protocols and using external experts to annotate through a 9-point Likert scale low or high cognitive load.

Epistemological frames are a way of understanding student reasoning and have to deal with the student motivation toward the learning activity. Examples of these frames are hesitant, calm, active (Andrade and Danish, 2016); talk, flow, action, stress (Worsley and Blikstein, 2015). These frames were also named action codes by Worsley and Blikstein (2013), which aimed to develop a system that based on speech and gesture recognition would be able to detect three levels of expertise in construction building.

2.2.4 Discussion

This literature survey deepens the knowledge about the modalities for learning and learning theories and how these were operationalised in the learning scenarios investigated in related studies. Among the "Taxonomy of multimodal data for learning" and the "Classification table of the hypothesis space," we identified three main challenges of MMLA revealed by the literature survey.

First of all, analysing the literature according to the proposed observability line (Table 2.2.1) evidenced that the MMLA community has not yet clarified how multimodal data can ultimately support learners in their learning process. None of the studies describes how multimodal data can be used to provide actionable feedback or even an intervention to learners. Hence, the first challenge identified is that (C1) there is a lack of understanding of *how multimodal data relates to learning and how these data can be used to support learners achieving the learning goals*.

Second, we noticed that generating analytics with multimodal data and letting humans (learners and teachers) making sense them is increasingly complex. The raw multimodal data are generally very noisy and have a large number of attributes and a low semantic value (Dillenbourg, 2016). When the number of attributes in the data set increases, the data become hard to visualize and to interpret for humans. In contrast, intelligent computer agents are able to deal more efficiently with multimodal data and can be employed to process vast amounts at scale and be trained to perform interpretations. Therefore, the second challenge is that (C2) *it is still unclear how to combine human and machine interpretations of multimodal data.*

Third, the field with MMLA is a field located at the intersection of different disciplines including learning science, machine learning, and social signal processing. We have noticed that learning science and machine learning talk differently about "learning" and that results in very ambiguous meanings and less fruitful discussions. The third challenge identified is that (C3) *the fields of machine learning and learning science use different terminologies which are ambiguous and need to be aligned*.

2.3 The Multimodal Learning Analytics Model

To address the challenges found by the literature survey (Section 2.4), we introduce the MLeAM, a conceptual model for the emerging research field of MMLA.

The design of MLeAM originates by the necessity to make optimal use of multimodal data for supporting learning activities through intelligent tutoring and learning analytics. The intended MLeAM contributions are framed more clearly into the following three main objectives, respectively, addressing the three challenges described in Section 2.2.4.

The first objective of MLeAM (O1) is to map the use of multimodal data to enhance the feedback in a learning context. Although other conceptual models were proposed, such, for example, the Learning Analytics Framework (Greller and Drachsler, 2012), until today, no conceptual model for learning was specifically designed to deal with multimodal data. MLeAM can, therefore, provide more structure to drive further research into the new research field of MMLA and help researchers to design future experiments. With such a structured approach, the community can better identify and describe major challenges that than can be addressed by independent research teams globally.

The second objective (O2) is to show how to combine machine learning with multimodal *data*. Multimodal data have the potential to provide a digital representation of the physical world in a way that both humans and artificial agents can process. The MLeAM shows explicitly for the MMLA community how to best combine human interpretations with machine learning and automatic inference.

The third objective (O3) is to establish a joint terminology across the two main scientific disciplines that the MMLA field combines: learning science, machine learning. With MLeAM, we hope to establish a shared MMLA terminology to make it meaningful for educational researchers but also express well-established terms from the educational world to the machine learning community. To address these objectives, we propose the MLeAM represented in Figure 2.3. Along with the observability line (Section 2.1) separating the input and hypothesis spaces, MLeAM introduces a second orthogonal dimension: the *Mixed reality line*. Mixed reality is defined as the contiguous space where physical and digital worlds meet (Milgram et al., 1994). We believe that the separation between physical and digital world helps to understand the benefit which intelligent computer agents and digital technologies can bring into the learning process. The behaviour of the learners and the feedback transmitted to them happens in the physical world. The multimodal data representation of the modalities and their processing and annotation live in the digital world. The intersection between the observability line and mixed reality line creates four quadrants as represented in Figure 2.3. The transition between these quadrants is guided by a *process* ("P") that generates a *result* ("R"). The model proceeds clockwise iteratively starting from the top centre.

2.3.1 From sensor capturing to multimodal data

The model starts with *(P1) sensor capturing*. This process consists of automatically sampling sensors' recording data from several modalities. These chosen modalities relate to the attributes of the input space (see Section 2.1) such as learner's body position, gaze direction, and facial expression. These data can be extracted from of the learner's behaviour and actions or from the learning environment; in either case, the modalities reside in the physical world. P1 continuously transforms different modalities into their digital representation: multiform data streams that we call *(R1) multimodal data*. A transversal cut into the multimodal data streams corresponds to a digital snapshot of the learner in the learning context at one specific time point. There are three important aspects to be considered when designing a P1 implementation: (a) definition of the used input space: the heuristic selection of the modalities and their data representation; (b) identification of the most suitable sensors to capture the selected modalities for the specific learning scenario; and (c) design and implementation of a sensor architecture, a hardware and software infrastructure for collecting and serializing the data streams from multiple sensors (Di Mitri et al., 2017). The design of the sensor architecture must take care of several technical aspects including sensor network engineering, raw data synchronization, fusion techniques, and data storage logic for sensor data persistence. A similar challenge regarding sensor data collection has also been addressed by Specht (Specht, 2015)

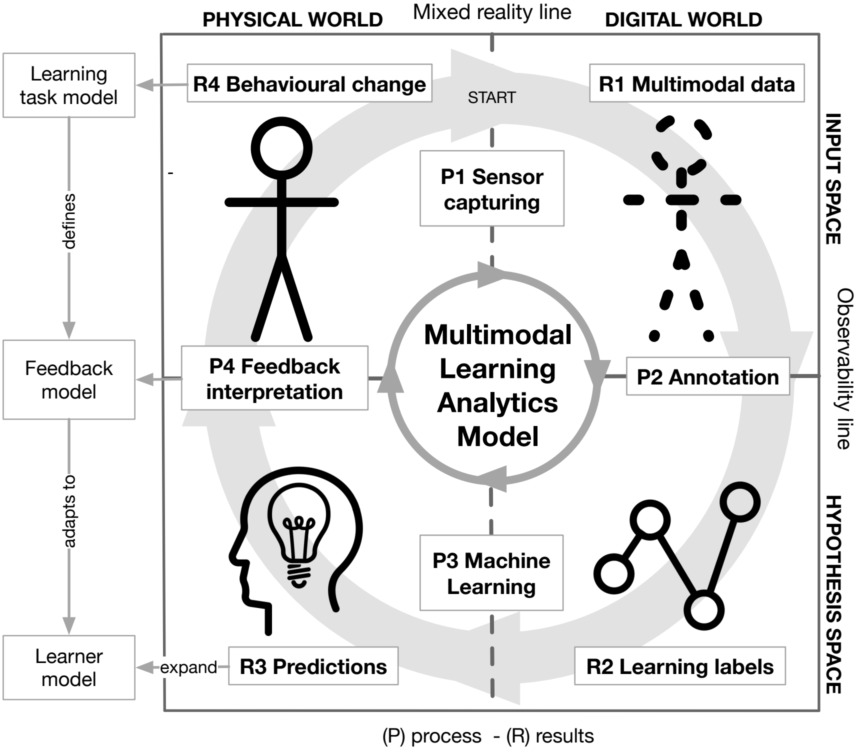


Figure 2.3 Multimodal Learning Analytics Model (MLeAM).

in the AICHE model.

2.3.2 From annotation to learning labels

The second process is the *(P2) annotation*, a repeated procedure driven by human such as an expert or by the learner. P2 aims at enriching the low-semantic multimodal data with human judgments according to some predefined assessment scheme. The scheme is based on the hypothesis space (see Section 2.2.1), that is, the unobservable interpretations that the machine learning algorithms automatically derive from the multimodal data. P2 can be seen as the assessment of a learning task in relation to some learning goals. P2 is achieved through triangulation: A judge is exposed to some human interpretable evidence of the learning task (e.g., videos or direct observation). The judge assigns some *(R2) learning labels* to time segments of the multimodal data. This process P2 annotation allows providing some meaning to some time intervals of the raw data. Similarly, to P1, P2 requires to define all the possible learning labels. This task corresponds to defining the hypothesis space and its data representation. It also requires devising an annotation strategy consisting of a reporting tool and an annotation procedure. The procedure must minimize the interpretation bias to provide the most reliable labels and should take into account the nature of the observed tasks (i.e., the learning context and activities). The minimum number of labels should be decided a priori. That is usually dependent on the estimated number of attributes to be considered in the model.

2.3.3 From machine learning to predictions

The third process is the *(P3) machine learning*. The purposes of supervised machine learning are (a) to learn statistical models (functions) out of observed (R1) multimodal data and manually annotated (R2) learning labels and (b) to generalize on future unobserved data similarly structured to generate *(R3) predictions* (Mohri et al., 2012). The core machine learning task can be expressed with a mathematical formalism, calculating a function: $y = f(X) + \epsilon$ where:

- X is a multimodal observation, input of the function f. X is a vector of n attributes $< x1, ..., xn >$ derived from the multiple learning modalities. All the possible value combinations of X constitute the input space, the domain of f.
- y is the learning label (s), which locate each input observation into the hypothesis space, the range of f of all possible learning labels.
- The function f is a generalization of the relationship between observations X and learning labels y plus some error term ϵ.
- Given a new multimodal observation X_{new}, the prediction task corresponds to calculating the learning label (s) $y_{new} = f(X_{new}) + \epsilon$.

P3 also includes the following iterative steps: (a) *preprocessing*: resampling, handling missing data; (b) *fitting the model* to the data; (c) *post-processing*: selection relevant attributes, tuning the parameters; (d) *validating the generalisability* of the model on new data; and (e) *diagnostics*: deriving relevance to determine the importance that each attribute holds in predicting the learning labels. If the obtained model is trained with reasonable accuracy, the system can be able to predict the learning labels throughout unseen multimodal data. This prediction is a machine-assisted estimation of the learner's standpoint in the learning process. P3 automatizes using machines the annotation procedure that has to be driven by humans. Predictions can be used to enrich the *learner model* and have a more adaptive feedback model for the learners and nudge them towards positive behavioural change. Both the learner model and the feedback models as shown in Figure 2.3 are not part of MLeAM but are connected to and extended by it.

2.3.4 From feedback interpretation to behavioural change

The final process is the *(P4) feedback interpretation* closing the machine-driven feedback loop returned to the learner. The purpose of P4 is to exploit support of the multimodal data and lead to *(R4) behavioural change*. P4 requires a feedback model that has to be designed in advance. Devising an efficient feedback model is not within the scope of MLeAM (see Figure 2.3). The feedback model is highly dependent on the learning activity and is defined by the task model. MLeAM does not deal with any of the feedback dimensions (Mory, 2004) and also does not inform about effective feedback strategies that depend on the learning activity. Nonetheless, MLeAM can be used in combination with different models of feedback with relevant already analysed information about the learners' behaviour and context. Different forms of feedback can be prompted to the learner based on the predictions obtained through MLeAM. The feedback design should be able to facilitate the process of feedback interpretation and lead the learner to some new learning behaviour. Similarly, to the (P2) annotation, the P4 feedback interpretation is fully human-driven.

2.4 Conclusions

In this article, we analysed the emerging field of MMLA. In Section 1, we introduced the origins of this new field according to four main developments. We highlighted the main mission for MMLA: using multimodal data and data-driven techniques for filling the gap between observable learning behaviour and learning theories. In Section 2.2.1, we described two components as input space and the hypothesis space separated by the observability line. We used them as a classification framework to conduct a literature survey (Section 2.2) of MMLA studies. By analysing the related literature, we were able to derive general characteristics of the multimodal data for learning (the input space, Section 2.2.1) and the learning theories and other constructs (the hypothesis space, Section 2.2.1). As a result of the literature survey, we proposed the *Taxonomy of multimodal data for learning* (Section 2.2.3, Figure 2.2) and the *Classification table for the hypothesis space* (Section 2.2.3), Table 2.1). The literature survey also unveiled three main challenges for the MMLA field (Section 2.2.3). We addressed these challenges introducing the MLeAM (Section 2.3), a conceptual model to support the emerging field of MMLA. MLeAM has three main objectives: (O1) mapping the use of multimodal data to enhance the feedback in a learning context; (O2) showing how to combine machine learning with multimodal data; and (O3) aligning the terminology used in the field of machine learning and learning science. We acknowledge that MLeAM is not to be considered in its final stage. In the future, we aim to extend MLeAM with various activities. First of all, it is important to extend the literature survey because some aspects were intentionally not covered, by, for instance, the social dimension of learning, that is, the extent to which both the teacher and the learning peers influence each other, for example, during dialogic learning. We encourage the readers to contribute in expanding the *Taxonomy of multimodal data for learning*[2] (Figure 2.2) and the *Classification table of the hypothesis space*[3] (Table 2.1) with further studies using a combination of different modalities and presenting convincing results in terms of accuracy and their adaptability to different learning settings. Further empirical studies and meta-analysis can also focus on which is the most suitable data representation for each modality;

[2] Available online for comments at http://bit.ly/MLEAMtree

[3] Available online for comments at http://bit.ly/MLEAMtheory

the heuristics for best modality combination; best pairing between modality and available sensors in commerce; and providing guidelines for the data analysis of multimodal data sets (P3 in MLeAM). On this particular point, multimodal data for learning needs best practices to achieve real-time time series analysis and classification, in combination with random events and proper balance between learner specificity and generalization across groups. Baselines for future experiments must be established, preventing to reinvent the wheel every time. This could technically be done on the one hand by extending the current interoperability standards (e.g., *Experience API*-xAPI) to better work with a high-frequency sensor and consequent data analysis. Meaningful baselines can also be software prototypes such as the Multimodal Learning Hub (Schneider et al., 2018), or hardware prototypes that can be used off-the-shelf for data collection: for instance, Process Pad (Salehi et al., 2012) or the Multimodal Selfie (Domínguez et al., 2015), two low-cost devices that can be used in classrooms for capturing multimodal data. Finally, the MLeAM classification evidenced a shortage of studies that focus on feedback and interventions for the learner and their learning process. In particular, more research is needed to invest feedback systems that use timely predictions generated by multimodal data. We encourage further collaboration with feedback experts to discover what kind of feedback is valuable for the learner and is it able to trigger fundamental behavioural changes.

Chapter 3

The Big Five challenges

The analysis of multimodal data in learning is a growing field of research, which has led to the development of different analytics solutions. However, there is no standardised approach to handle multimodal data. In this paper, we describe and outline a solution for five recurrent challenges in the use of multimodal data for supporting learners: the data collection, storing, annotation, processing and exploitation. For each of these challenges, we envision possible solutions.

This chapter is based on:

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018) The Big Five: Addressing Recurrent Multimodal Learning Data Challenges. In R. Martinez-Maldonado et al. (Eds.), *Proceedings of the Second Multimodal Learning Analytics Across (Physical and Digital) Spaces* (CrossMMLA), Vol. 2163. CEUR Proceedings.

3.1 Background

The Learning Analytics & Knowledge (LAK) community has acknowledged the necessity of taking into account physical and co-located learning activities as much as practice-based skills training; it is undeniable that in the classroom and at the workplace these "offline moments" still represent the bulkiest set of learning activities. Bringing these moments into account requires extending the data collection to additional data sources which go beyond the conventional ones, such as online learning systems, Massive Online Open Courses (MOOCs) platforms or student information systems. With the term multimodal data, we refer to the learning data sources collected "beyond user-computer interaction", i.e. those data sources collected during learning moments alternative to the classic desktop-based learning scenario. Although user-computer interaction data could still hold some relevant information, they can be complemented by additional multimodal data; these data can be classified into 1) data describing the learner's behaviour: including motoric and physiological data; 2) data regarding the learning situation, including social context, learning environment and learning activity. Most of these aspects can be monitored through wearable sensors, cameras or Internet of Things (IoT) devices. These tools can capture only what is "visible" by a generic sensor, meaning they generally do not have the ability to reason on the meaning behind the collected data. The observability line – i.e. what is visible by sensors and whatnot, conceptually separates multimodal data by human-driven qualitative interpretations, like expert reports or teacher assessments. The latter, that are more qualitative and human-driven, describe dimensions that the sensors cannot directly observe, such as learning outcomes, cognitive aspects or affective states.

Bridging the gap between learner's complex behavioural patterns with learning theories and other unobservable dimensions is the paramount challenge for multimodal analysis of learning (Worsley, 2014). Multimodal data can be used as historical evidence for the analysis and the description of the learning process: this field of research is called Multimodal Learning Analytics (Blikstein, 2013). The related literature shows the potential to apply a multimodal approach in a variety of learning settings including dialogic learning in teacher-student discourse (D'mello et al., 2015); computer-supported collaborative learning during knowledge-sharing and group discussions (Martinez-Maldonado et al., 2018; Schneider and Blikstein, 2015); in practice-based and open-ended learning tasks, when understanding and executing a practical learning task (Ochoa et al., 2013).

The potential benefits of multimodal data are not only limited to analytics, e.g. human interpretation of dashboards or other visual metaphors. If multimodal data are reliable and correctly addressed and exploited, they can be used as the base to drive machine intelligence and achieve better personalisation and adaptation during learning. Multimodal data is expanding the horizon of the Learning Analytics community and its moving towards Intelligent Tutoring Systems and Artificial Intelligence in Education research communities. For decades the long-term goal of these communities consisted in designing intelligent computer agents empathic to the learners which work as an instructor in the box and that can implement strategies to reduce the difference between experts and student performance (Polson et al., 1988). Multimodal data can facilitate achieving this goal, by equipping intelligent tutors with action-based recognition and reasoning, so that they can deal with open-ended learning tasks in uncontrolled environments.

3.2 Multimodal challenges

The analysis of multimodal data in learning is a fairly new but steadily growing field of research. As the interest tracing learning through the use of multimodal data grows, the opportunities stemming from it become more evident. As some authors have pointed out, the field of MLA faces a set of open challenges that create research gaps that need to be filled (Blikstein and Worsley, 2016). For instance, the LAK community (and its CrossMMLA interests group) still lacks a standardised approach for modelling of the evidence extracted from the learning process and producing valuable feedback with multimodal data. In contrast, multiple tailored ad-hoc solutions have been developed in related researches. A standardised approach to MMLA, in our understanding, should help researchers in setting-up their multimodal experiments by clarifying how the collection, storage, analysis and exploitation of the multimodal data takes place in a pragmatic and scalable manner that can be adopted into real-life educational settings. To contribute to filling this gap, in this paper, we outline five main challenges stemming from the feedback loop empowered by multimodal data and learning analytics. For each of these challenges, we describe possible solutions or approaches.

3.2.1 Data collection

The first step of the journey is the data collection, that being the creation of datasets through multiple sensors and external data sources. The sensors employed are most likely to be produced by different vendors, hence to have different specifications and support. The approach used for data collection must be flexible and extensible to different sensors, it should allow the collection of data at different frequencies and formats. Strongly connected to the collection is the data synchronisation. Proposed solution: to address this challenge, we introduce the LearningHub, a software prototype whose purpose is to synchronise and fuse different streams of multimodal data generated by the multiple sensor-applications. The LearningHub's main role is to deal with the low-level specifications of every sensor offering a customisable interface to start and stop the capturing of a meaningful part of a learning task, i.e. moments definable by atomic actions; we call this an Action Recording. The LearningHub is responsible to collect the updates for every sensor, organising and synchronising them chronologically.

3.2.2 Data storing

The second step is the data storing that encompasses the serialisation, storing and logic for retrieval of the action recordings. This step is crucial to organise the complexity of multimodal data which has multiple formats and big sizes. Proposed solution: The LearningHub channels the data from multiple sensors and provides as output multiple JSON files, which serialise and synchronise the sensor values for each sensor application. The JSON files allow for sensors having multiple attributes with different time frequencies and formats; they work as exchange format documents and provides also the logic to facilitate the action recording for storing and later retrieval.

3.2.3 Data annotation

The data annotation challenge consists in finding a seamless and unobtrusive approach for labelling the learning process, i.e. triangulating the multimodal action recordings with the evidence (e.g. video clips) of the learning activities. The annotation step is rather crucial, as most of the time the meaning of a recording is not trivial to derive just by looking at the sensor values. The format chosen for assigning the semantics to the action recordings is also a relevant issue. Proposed solution: to address this challenge, we propose the Visual Inspection Tool (VIT). The VIT is a web-application prototype for the retrospectively analysis and annotation of multimodal action recordings. The VIT allows to load multimodal datasets, plot them on a common time scale and triangulate them with video recordings of the learning activity. It allows to select a particular timeframe and annotate the multimodal data slice with an Experience API (xAPI) triplet, assigning an actor, a verb and an object. The VIT offers a human-computer interface which helps to deal with the complexity of multimodal data.

3.2.4 Data processing

The data processing steps consist in extracting and aligning the relevant attributes from the “raw” multimodal data and transforming them into a new data representation suitable for exploitation. The data processing steps depend tightly on the data exploitation which is discussed in the next section. Common steps for data processing include data cleaning (e.g. handling missing values, resampling and realigning the time series), feature extraction, dimensionality reduction and normalisation. The challenging side of the data processing for multimodal data is given by the size of the multimodal datasets, the need to process them periodically and the need to process as close to real-time as possible, a relevant aspect especially in the case of immersive feedback generation.

Proposed solution: the idea is to have a Pipeline for multimodal data for learning, a cloud-based application which allows to plan and execute data processing routines (e.g. Spark jobs). These routines should query the Learning Record Store and fetch the all recent/relevant xAPI statements and load into memory all the action recordings connected to each xAPI statement. The raw action recordings will be transformed according to the set of operations specified which will output a transformed action recording which is saved and ready to be fed into a data mining algorithm.

3.2.5 Data exploitation

Through an analysis of the related experiments in the literature using multimodal data in learning settings, we concluded that there are different use cases generally used for enhancing and facilitating the learning process with multimodal data. Proposed solution: we classify the different use cases into five exploitation strategies:

1. *Light-weight feedback*: hard-coded rules and feedback based on heuristics of the form “if sensor value is x then y”;
2. *Replica*: replays of the action recordings, e.g. ghost-tracks of motoric sensors data;
3. *Historical reports*: aggregated visualisations in forms of analytics dashboard, a group of data visualisations that show the historical progress of the sensor recordings in condensed form;
4. *Frequent patterns*: mining of recurrent sensor values occurrences within one or multiple sensor recordings;
5. *Predictions*: estimation of the human-annotated labels during similar action recordings.

The strategies can be used for different purposes and applications. They differ in the level of data processing used and consequently by the methods used for data analysis; these include descriptive statistics, supervised or unsupervised machine learning. For example, light-weight feedback requires simple hard-coded rules; historical reports require visualisations that can be grouped into analytics dashboard; frequent patterns or predictions require training either machine learning models, store them into memory, and use them to estimate the value or the class of a particular target attribute. Historical reports also differ by the effort required by human experts, for example in collecting the labels or for interpreting the visualisations; similarly, the strategies differ by the level of machine reasoning, e.g. between those using machine learning and those which use heuristics.

3.3 Conclusions

In this paper, we have introduced five main challenges connected to the use of multimodal data in learning. These challenges deal with the data collection, storing, annotation, processing and exploitation and constitute important research questions for all the CrossMMLA community. Along with these challenges, we briefly explained some practical solutions. Being these ideas preliminary, we use them as agenda points and research questions for the Cross-Multimodal Learning Analytics (CrossMMLA) research community. We hope that pointing out these challenges can raise interest and awareness in the current research endeavours in the area of multimodal learning analytics.

Part III

Preparation of Navy

Chapter 4

Read Between The Lines

This chapter introduces the Visual Inspection Tool (VIT) which supports researchers in the annotation of multimodal data as well as the processing and exploitation for learning purposes. While most of the existing Multimodal Learning Analytics (MMLA) solutions are tailor-made for specific learning tasks and sensors, the VIT addresses the data annotation for different types of learning tasks that can be captured with a customisable set of sensors flexibly. The VIT supports MMLA researchers in (1) triangulating multimodal data with video recordings; (2) segmenting the multimodal data into time-intervals and adding annotations to the time-intervals; (3) downloading the annotated dataset and using it for multimodal data analysis. The VIT is a crucial component that was so far missing in the available tools for MMLA research.

This chapter is based on:

Di Mitri D., Schneider J., Specht M., & Drachsler H. (2019) Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. In: *Proceedings of the Ninth International Learning Analytics & Knowledge Conference 2019* (LAK '19) (pp. 51–60). New York, NY, USA. ACM. DOI: 10.1145/3303772.3303776.

4.1 Introduction

Multimodal interaction methods are becoming increasingly popular in learning science. User-computer interaction data, traditionally derived from software logs or clickstreams are being enriched by additional data sources. These new data sources are gathered with wearable sensors, Internet of Things (IoT) ubiquitous devices or embedded systems. The driving reason for this shift is to achieve a more comprehensive evidence and analysis of learning activities taking place in the physical realm. In the last decade, research in learning science using data-driven technologies such as learning analytics, educational data mining or intelligent tutoring systems have focused primarily on online, desktop or mobile learning. In these settings, learning activities are mediated by computer devices. In contrast, settings as co-located collaborative learning, practical skills training, or dialogic classroom discussions are less addressed in data-driven learning research. Including novel data sources to investigate these learning activities happening in the physical space is in line with the overarching theme of the Learning Analytics & Knowledge conference 2019, which focuses on 'inclusion'.

In support of these new forms of interaction, within the Learning Analytics community, a new research focus has emerged, coined as multimodal learning analytics (MMLA) (Blikstein, 2013). The objective of MMLA is to track learning experiences by collecting data from multiple modalities and bridging complex learning behaviours with learning theories and learning strategies (Worsley, 2014). MMLA, however, is an emerging research field that needs more evidence: a recent survey analysed eighty-two MMLA papers, forty-six of which are empirical, while the reminder theoretical (Worsley, 2018). Along with more evidence, MMLA needs structured technological practice. Whilst the learning analytics community put considerable efforts into standardisation and interoperability into data collection, analysis and exchange, these efforts did not yet meet multimodal data. Researchers using multimodal interaction approaches face multiple challenges that stem from the complex nature of multimodal data (Lahat et al., 2015). For the case of MMLA, when using multimodal data for learning, the challenges have been grouped into five categories (Di Mitri et al., 2018b).

1. The *data collection*: the approach used for capturing, aggregating and synchronising data from multiple modalities and sensor streams;
2. the *data storing*: the approach used for organising multimodal data which having multiple formats and big sizes, for storing them the and logic for later retrieval;
3. the *data annotation*: the approach for providing meaning to portions of multimodal recordings and collecting human interpretations through expert or self-reports;
4. the *data processing*: various steps for cleaning, aligning, integrating, extracting relevant features from the "raw" multimodal data and transforming them into

a new data representation suitable for exploitation;

5. the *data exploitation*: the approach to ultimately support the learner during the learning process with the predictions and the insights obtained by the multimodal data.

As a recent review concluded, MMLA has yet no structured approach to handling multimodal data for learning gathered from different sources (Shankar et al., 2018). In contrast, multiple tailor-made solutions have been developed in related research adapted for specific use-cases, making the process of MMLA time consuming and expensive. This aspect constitutes, in our view, a great shortcoming for the field which hinders MMLA solutions to be adopted in the practice. Standard approaches could help researchers simplify the setup of their experiments. They would allow practitioners benefiting from more inclusive Learning Analytics that go beyond mouse and keyboard interaction, including learning activities in the physical space. To fill this gap in the current research, we propose a structured approach for MMLA. We first explain in section 4.2 the differences that multimodal bring about compared to traditional user-interaction data (section 4.2.1). We describe the challenges introduced by sensors (section 4.2.2) as well as the potential of the multimodal approach (section 4.2.3). We then review different existing multimodal tools and see how they compare them in terms of their functionalities and purposes and classifying them their features according to five MMLA challenges (section 4.2.4). In section 4.3 we introduce our prototypical solution, consisting of the design and implementation of the Visual Inspection Tool (VIT). The VIT is a generic solution for the third category of MMLA challenges, the data annotation. We list a set of functional requirements for the VIT (section 4.3.1) and we describe how we implement the different components of it (section 4.3.2). We propose a procedure for the data processing (section 4.3.3) and data exploitation (section 4.3.4) and we test the VIT in three use cases (section 4.3.5). In section 4.4, we discuss the findings of our approach positioning the VIT within an integrated workflow, the *Multimodal Learning Analytics Pipeline* which works as a guideline for MMLA researchers. Finally, in section 4.4 we draw conclusions and point to future directions.

4.2 Background

4.2.1 Computer Assisted Learning without mouse and keyboard

To date, learning research communities and practitioners using data-d riven approaches in learning focus on traditional and easy-to-retrieve data sources. These sources include interaction data with the learning platform, such as the Learning Management System (LMS) the Massive Online Open Course (MOOC), data crawled from social media or collected from mobile applications. The accessibility and the richness of these “traditional” data sources has in recent years motivated and inspired learning science research communities, such as the Learning Analytics and Knowledge community (Greller and Drachsler, 2012). These traditional data sources only capture the learner's interactions with the learning system or platform, hardly capture any other moment in which learning occurs. For instance, it was found that the usage of a mobile application used for learning was not a reliable source for measuring the actual time the learner was spending on the learning task (Tabuenca et al., 2015b). The 'traditional' learner-computer interactions are events that share similar properties. For each of those events, it is straightforward to derive *who did what*, as they are already sorted by user, action and activity. Such semantic encoding is used by interoperability standards as Experience API for specifying events such as "user likes post", "user watches video". With these types of events, there is no need to decipher *who* did the action, *when* it was done and *what* kind of action it was. This semantic encoding is easy to be interpreted by humans and for machines. They are easily summarisable with aggregate functions and fed into visualisations or dashboards showing the activity level and the learning effort. But what happens during the learning moments in which the interaction with a computer is not the main activity? For instance, when learning how to use a circular saw in a carpentry school, when elaborating a solution to a challenge in a co-located group discussion, or when participating in classroom discussions with teachers and peers. These are examples of learning activities frequently happening in everyday classroom or workplace settings. Psychomotor skills training, dialogic learning and co-located group interactions are dominated by physical interactions, which remain most of the time *untracked*. The physical interactions are not captured and, therefore, are not included in the datasets used for analysis. In these physical learning scenarios, mobile devices or laptops are not used as main media, but rather as side-tools, for example for taking notes or looking-up web resources. Bringing these moments into account requires a more seamless data collection that goes beyond the direct learner-computer interactions derived from clicks, keystrokes or nested software logs.

4.2.2 Sensors in learning

Beyond mouse and keyboard events there are multimodal sensor data, i.e. data that can be collected with sensors, computer chips, wearable trackers, microphones, cameras and other computer input devices. Consumer-level sensors, wearables and IoT devices introduce new multimodal affordances that can be exploited to collect evidence of learning actions in a wide range of situations. There exist a large variety of sensor devices that can be used in learning contexts. Schneider et al. (Schneider et al., 2015a), have identified more than 23 different sensor devices that have been or can be used in the domain of learning. Sensors can monitor both the physical environment where the learner operates and the learner's behaviour including 360-degrees body movements, physiological responses such as heart rate or body temperature, or interpersonal communication, student-teacher or student-peer discussions. In related literature, sensors have been used to track different modalities such as hand gestures (Ochoa et al., 2013; Worsley and Blikstein, 2018), gross body movements (Bosch et al., 2015), eye-tracking (Prieto et al., 2016; Raca and Dillenbourg, 2014), facial expressions (Arroyo et al., 2009; Bahreini et al., 2015).

Sensors were also used to monitor physiological signals such as heart rate (Hussain et al., 2012; Alzoubi et al., 2012), galvanic skin response (Pijeira-Díaz et al., 2018; Grafsgaard, 2014), brain waves (Alzoubi et al., 2012; Prieto et al., 2016). The recent framework of Limbu et al. proposes how to use sensors to capture the expert performance and re-enact it through Augmented Reality using a variety of approaches, coined as Transfer Mechanisms (Limbu et al., 2018b). Despite the extensive body of literature featuring sensors, still, a very small percentage of studies opted for non-tailored and more generic approaches. In addition, to the best of our knowledge, there is little use of MMLA solutions in actual educational practices. While in many sectors such as automotive, manufacturing, retail, smart living or health care, IoT and sensor-based systems are now very common, similar systems have not been developed yet for education and learning. Compared to other domains, human learning is a more complex and individual process much harder to model. The learning environment for humans is usually open-ended and highly unpredictable. There are many factors which influence learning, including the social relationships with teacher and peers, or the necessity to close the feedback loop to correct and improve learning behaviour.

4.2.3 Multimodal data for personalised learning

Multimodality is a theoretical assumption that can be used to provide more structure in the use of sensors for investigating learning. The idea of using multimodality in learning derives from the theory of embodied communication. According to this theory, humans use their whole body to communicate, they use a multitude of channels to exchange messages in shared contexts (Wachsmuth et al., 2012). Similarly to human-human communication, the multimodal principle can also be employed in human-computer interaction. Sensor-based multimodal interfaces can monitor the variation of different modalities during learning activities: including dialogic learning in the classroom (D'mello et al., 2015), in computer-supported collaborative learning during knowledge-sharing and group discussions (Martinez et al., 2011; Praharaj et al., 2018), practice-based and open-ended learning tasks (Worsley, 2014), when understanding and executing a practical learning task (Ochoa et al., 2013) or training presentation skills (Schneider et al., 2015b). The benefit of multimodal data is, first of all, enriching the learner's digital representation as well as the ones of the learning context, environment or task. Enriched representations can shed more light on cognitive states or metacognitive factors that influence learning. Di Mitri et al. 2018a provided a model, the *Multimodal Learning Analytics Model* that frames the different phases of MMLA and describes how MMLA outcomes can be used for personalised feedback and reflection for learners. Using the multimodal data collected during a learning task, machine learning algorithms can be trained to classify, cluster or predict learning dimensions that are "invisible" to sensors, such as learner's emotion, learner's cognition or outcomes. These predictions can be used to tailor different types of feedback, which can be used to guide the learner towards predefined learning goals, whether this is improving a particular task or stimulating self-reflection. Multimodal data can be used to enhance intelligent tutoring systems

(ITSs) employed in physical learning activities, providing a more accurate student model representation. Figure 4.1 represents this aspect with a typical ITS tripartite structure (1) Knowledge domain, (2) Student model, (3) Pedagogical model (Polson et al., 1988) enhanced by multimodal data collected in the physical realm.

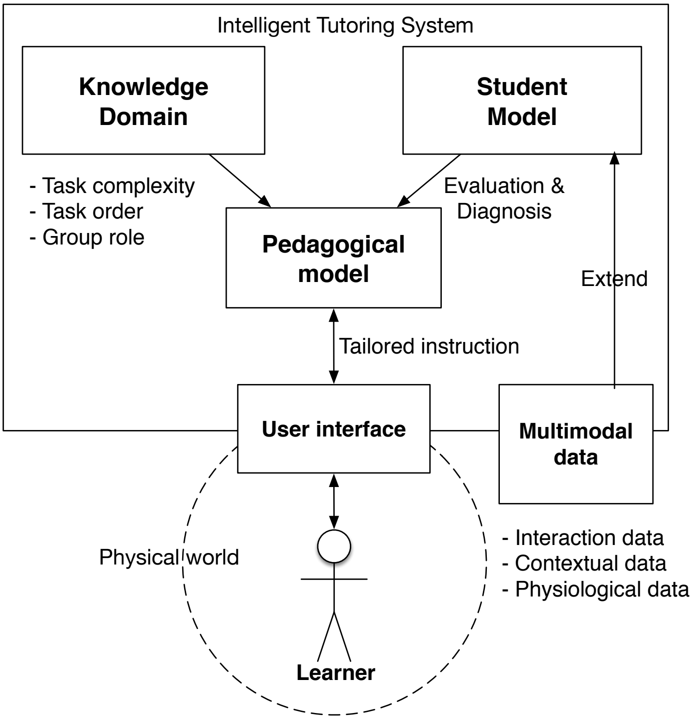


Figure 4.1 Multimodal data can extend the digital representation of the learner in the system.

4.2.4 Tools for Multimodal Data

One relevant aspect of MMLA research is the design of the technical infrastructure and the choice of technical tools to handle the data gathered from multiple modalities. Compared to learner-computer interaction data, sensor data pose a much bigger challenge: single behavioural particles have low semantic meaning if considered singularly (Dillenbourg, 2016). For example, the skeleton data recorded with Microsoft Kinect can record the full-body movements which translate to more than fifty sensor attributes. On the one hand, visually representing raw data values for human inspection is not informative, on the other hand, multimodal data can be processed by computer-based algorithms which need more complex architectures which lack in the learning domain. There exist, however, technical tools in contiguous disciplines to MMLA as social signal processing, which can solve some specific parts of the data-informed cycle. In this section, we review seven of these tools, later we classify them into five categories previously introduced.

Social Signal Interpretation & NovA

The Social Signal Interpretation (SSI) (Wagner et al., 2011) is an open source framework for real-time recognition of social signals during social interactions. SSI supports synchronised data recording from a large range of sensor devices to recognise behavioural cues such as gestures, head movements, and emotional speech. SSI allows the user to collect their own training corpora and obtain personalised models. Through an XML editor, SSI allows to draft and run pipelines by connecting multiple sensors, to data processing techniques (transformations) and visualisations (consumer applications). While SSI has plugins for multiple existing sensors, it does not clarify how to connect new ones. One relevant extension, NovA (NonVerbal behaviour Analyzer) (Baur et al., 2013) allows the user to annotate behavioural cues and measure the quality of the social interactions, in terms of user's levels of engagement and activation in the interaction. SSI is an open source software (GPL license) written in C++.

Lab Streaming Layer

Lab Streaming Layer (LSL) focuses on a unified collection, synchronisation and storing time series data (Kothe et al., 2018). Developed by the Swartz Center for Computational Neuroscience, LSL implements plug-ins for multiple brain-computer interface devices primarily used in neuroscience research such as electroencephalograms (EEG). LSL creates *stream outlets* where streaming data can be published in samples or chunks with regular or irregular sampling ranges. Receiver nodes can also subscribe through *stream inlets*, topics to which computers in the same network can subscribe. LSL uses functions to discover and resolve streams of data in the network. LSL also features a built-in clock that allows assigning timestamps to the collected data samples in order of sub-millisecond accuracy using the Network Time Protocol. LSL uses a custom data format, the eXtensible Data Format (XDF). LSL is open source cross-platform software that offers interfaces in C, C++, Python, Java, C#, Matlab.

Data Curation Framework (DFC)

Data Curation Framework (DFC) focuses on raw sensory data acquisition, curation and monitoring (Amin et al., 2016). It implements an IoT approach for data collection in real time from multiple modalities distributed environment over a cloud-based platform. DFC implements a rule-based anomaly detection system. As the computation is performed over the cloud platform, the interesting side of DFC is the scalability. DFC introduces the concept of user's lifelogs curation through continuous monitoring of the sensory data. The DFC system was created for pervasive health monitoring and, for this reason, it implements algorithms as anomaly detection based on expert hardcoded rules. DFC is written in Java and Javascript and released open source under Apache license 2.0.

ChronoViz

ChronoViz is a tool for the visualisation of time-based data from multiple synchronised streams of data (Fouse et al., 2011). ChronoViz allows researchers to navigate multimodal data, including video, audio, digital nodes and geographic data and add text-based annotations. It is designed to support multiple videos and support log data (e.g. from flight simulator), geographic coordinates, video and audio transcripts or notes were taken on paper with a digital pen. The annotations are overlaid on top of the synchronised graphs and they can be grouped into categories of annotations that have different colours. ChronoViz runs as a local Mac OSX application, its code is open source under Apache license 2.0.

RepoVizz

RepoVizz is a data repository and visualisation tool for storage and user-friendly browsing of music performance multimodal recordings (Mayor et al., 2013). RepoVizz offers means for researchers to access music performance online in a shared multimodal database. RepoVizz supports several formats for audio and video. It also accepts sensor files in CSV format (e.g., motion capture or physiological signals). The data is structured into an XML (repoVizz Struct) which through metadata allows the user to organize multimodal data in a hierarchy. The XML files and all the multimodal data sources are called repoVizz *Datapack* and are uploaded into zip files. It differentiates users types into Producers and Consumers allowing different downloading, uploading and annotating rights. RepoVizz allows annotations for different sound streams to segment different notes to identify different instruments playing in collaborative musical settings such as orchestras. RepoVizz is an open source software written with Javascript-HTML5 front-end, Java backend and MySQL database.

Generalized Intelligent Framework for Tutoring

GIFT is a framework that provides tools, methods and standards for the design and evaluation of computer-based tutoring systems (Sottilare et al., 2012). GIFT is developed by the Learning in Intelligent Tutoring Environments Laboratory, part of the U.S. Army Research Laboratory. GIFT consists of a *sensor module* interfacing with commercial sensors for processing and storing sensor data; a *domain module* providing the content to support the training and assessing the learner's performance against standards; the *pedagogical module* identifying the need for feedback. GIFT uses sensor data to identify the learner's affective, cognitive and psychomotor states. GIFT source code is available via registration.

Multimodal Learning Hub

The Multimodal Learning Hub (LearningHub) is a system that focuses on the data collection and data storing of multimodal learning experiences (Schneider et al., 2018). It uses the concept of *Meaningful Learning Task* (MLT) and introduces a new data format (MLT session file) for data storing and exchange. The LearningHub implements a set of specifications that shape it for certain types of learning activities. It was created to be compatible primarily with commercial devices (e.g. Microsoft Kinect, Leap Motion, Myo Armband) and other sensors with drivers running with the most common operating systems. It focuses on short and meaningful learning activities (10 minutes) and uses a distributed, client-server architecture with a master node controlling and receiving updates from multiple data-provider applications. It also handles video and audio recordings with the main purpose to support the human annotation process. The expected output of the LearningHub is one (or multiple) MLT session files including (1) *one-to-n* multimodal, time-synchronised sensor recordings; (2) a video/audio file providing evidence for retrospective annotations. The LearningHub is open-source and developed in C#.

Table 4.1 Comparison of the existing multimodal tools. (n.a. = not available).

Tool	Collection	Storing	Annotation	Processing	Exploitation	Main purpose
1. Social Signal Interpretation (Wagner et al., 2011)	Multisource, Synchronised streams	No custom format	Using NovA	Custom pipelines, various ML algorithms	n.a.	Human activity recognition
2. Lab Streaming Layer (Kothe et al., 2018)	Multisource, streams	Custom data format(XDF)	n.a.	n.a.	n.a.	Physiological data synch.
3. Data Curation Framework (Amin et al., 2016)	Multisource, synchronised streams	n.a.	n.a.	Anomaly detection	n.a.	Pervasive health-care monitoring
4. ChronoViz (Fouse et al., 2011)	n.a.	n.a.	text-based annotations	n.a.	n.a.	Video coding human interactions
5. RepoVizz (Mayor et al., 2013)	n.a.	Custom data format (repoVizz struct)	text-based annotations	n.a.	n.a.	Visual analysis of multi-user orchestration
6. GIFT (Sottilare et al., 2012)	Multisource, batches	Store in csv format	n.a.	Can be linked with external processing tools	Corrective and personali-sed feedback	Designing ITS
7. Multimodal LearningHub (Schneider et al., 2018)	Multisource, synchronised batches	Custom data format (MLT)	n.a.	n.a.	Corrective feedback	Intelligent Learning Feedback

4.3 Methodology

Among the five categories of challenges of MMLA this paper puts its focus primarily on data annotation and consequently on data processing and exploitation to complement which, as the review of multimodal tools reveals, are the MMLA challenges which are not sufficiently addressed. To address this gap, we formulate the following research questions: (RQ1) *How to design a system to annotate multimodal recordings?* (RQ2)*What processing techniques can be used with annotated multimodal recordings?* (RQ3) *How to exploit annotated multimodal recordings in different learning scenarios?*

4.3.1 Design specifications

Data annotation requires a human evaluator, typically an expert, who can be a researcher or a trainer. The user assigns meaning to portions of the multimodal sensor recording through direct or retrospective observations. This procedure is also known as *triangulation*: the user *views* the interpretable evidence (e.g. video) while *plotting* and visualising synchronised multimodal recordings; thus, the user *adds annotation* transferring meaning to multimodal recordings which are hard to interpret. The annotations should be *customised* according to the chosen annotation scheme and the output - the annotated dataset - can be *downloaded*. Most of the time, multiple experts need *shared access* to the same datasets. Depending on the richness of the sensors, processing the dataset can become computationally expensive, therefore a *scalable* and dynamic allocation of processing power is needed. This scenario can be seen in the form of functional requirements:

(FR1) the user can *plot* and visualise a multimodal recording file, featuring multiple synchronised data streams;

(FR2) the user can *view* video of the session synchronised with the multimodal data;

(FR3) the user can *add annotations* to single time intervals in attribute-value form;

(FR4) the user can add *custom* annotations;

(FR5) the user can *download* the annotations or attach them to the session file;

(FR6) the tool should be compatible with cloud-based solutions for *scalability* and *shared access*.

For the specific challenge of data annotation, the existing tools that best fulfil these requirements are ChronoViz and repoVizz. ChronoViz fulfils FR1 and FR2, however, it is not designed to handle datasets having a large number of sensor attributes. It runs locally and therefore does not fulfil FR6. It allows custom annotations but only consisting of text (not attribute-value) therefore does not fulfil FR3. Similarly, it does not allow to select time intervals but only time points. repoViz instead fulfils FR1, FR2, FR6. When it comes to annotations, however, the tool focuses more on annotating the different contributions of single modalities (e.g. musical instruments) rather than annotating a new corpus to prepare it machine learning or feedback. On this note, SSI and its extension NovA focus on creating custom corpora. One aspect implemented in DCF and repoVizz is the use of web compatible technologies which relate to FR6. repoVizz, contrarily to the other software, implements a web interface while DCF performs operations in the cloud. These decisions allow for better scalability than software running locally on a single computer.

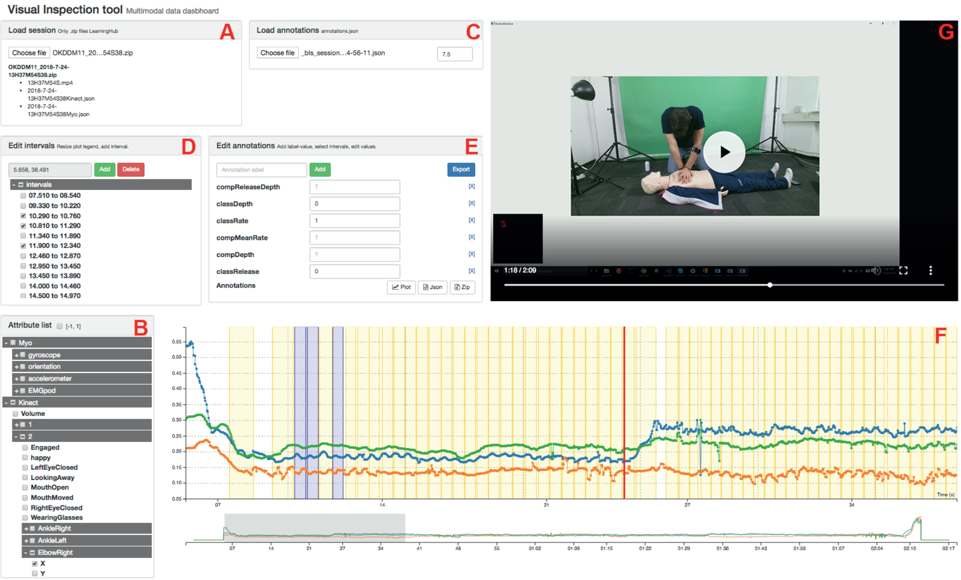
As our solution focuses primarily on multimodal data annotation, we need to apply other solutions for data collection and storing. Among the reviewed tools, the candidates are SSI, LSL and LearningHub. The upside of LSL is the high fidelity data collection and custom format. For SSI are the wider community and support and the variety of plugins. Both do not clearly explain their extensibility with new sensors and their contribution to user feedback. The LearningHub, despite being more recent and experimental, focuses on learning and therefore prioritises the feedback loop. We decide to adopt the LearningHub and its data format (MLT session file) providing synchronised multimodal sessions with video recordings.

4.3.2 Implementation

To fulfil the design requirements, we develop the Visual Inspection Tool. The VIT is developed as a server-less web application in Javascript and HTML5 running on an Internet browser. We chose this technology because it is highly compatible. The VIT allows for plotting datasets and while playing multimedia files such as videos, and high performance is not a requirement for this type of application. VIT can plot the multimodal data of the video recording of the session for triangulation. A comprehensive screenshot is represented in Figure 4.2. VIT contains a parser for the custom MLT data format and is currently able to handle a great number of sensors attributes. In the use cases, we have tested in section 4.3.5, we were able to load up to one hundred different attributes. This number can easily scale higher, since, by defaults, the attributes are not loaded into the chart unless they are clicked by the user.

Session file

The multimodal recordings used as input corresponds to an MLT session file of the LearnigHub. The data format of these sessions is a zip compressed folder. In this version of the session folder, it can be loaded on the panel Session load (Figure 4.2, A, top-left) by picking a session stored on the local drive. If VIT runs on an online server, the session files could also be picked from a cloud repository. When selecting a session an asynchronous parser is activated to scan the content of each file contained in the session zip folder. Following the specs of the LearningHub, the files can be either JSON (Sensor application files) or MP4 file formats. The JSON files will be parsed as described in the next section and the video will be loaded (Figure 4.2, G, top-right). An example of such a file is shown in Listing 1.



Visual Inspection tool Multimodal data dasbhoard

Load session Only .zip files LearningHub

Choose file OKDDM11_20...54S38.zip

OKDDM11_2018-7-24-13H37M54S38.zip
- 13H37M54S.mp4
- 2018-7-24-13H37M54S38Kinect.json
- 2018-7-24-13H37M54S38Myo.json

Load annotations annotations.json

Choose file _bls_session...4-56-11.json 7.5

Edit intervals Resize plot legend, add interval.

5.658, 38.491 Add Delete

- ⊟ intervals
  - □ 07.510 to 08.540
  - □ 09.330 to 10.220
  - ☑ 10.290 to 10.760
  - ☑ 10.810 to 11.290
  - □ 11.340 to 11.890
  - ☑ 11.900 to 12.340
  - □ 12.460 to 12.870
  - □ 12.950 to 13.450
  - □ 13.450 to 13.890
  - □ 14.000 to 14.460
  - □ 14.500 to 14.970

Edit annotations Add label-value, select intervals, edit values.

Annotation label Add Export

compReleaseDepth	?	[X]
classDepth	0	[X]
classRate	1	[X]
compMeanRate	?	[X]
compDepth	?	[X]
classRelease	0	[X]
Annotations	Plot Json Zip	

Attribute list □ [-1, 1]

- - □ Myo
  - + □ gyroscope
  - + □ orientation
  - + □ accelerometer
  - + □ EMGpod
- - ⊟ Kinect
  - □ Volume
  - + □ 1
  - - ⊟ 2
    - □ Engaged
    - □ happy
    - □ LeftEyeClosed
    - □ LookingAway
    - □ MouthOpen
    - □ MouthMoved
    - □ RightEyeClosed
    - □ WearingGlasses
    - + □ AnkleRight
    - + □ AnkleLeft
    - - ⊟ ElbowRight
      - ☑ X
      - □ Y

Figure 4.2 The VIT features the following elements: (A) Session loader MLT session file custom data format; (B) Attribute list; (C) Annotation file loader; (D) Time intervals editor; (E) Annotation editor; (F) Attribute plots; (G) Video player of the recording.

```
{
    "recordingID": "11H57M49S371",
    "applicationName": "Myo",
    "frames": [
    {
        "frameStamp": "00:00:00.0035019",
        "frameAttributes": {
            "orientation_X": "-0.6087646",
            "accelerometer_Y": "-0.3989258",
            "gyroscope_Z": "2.9375",
            "EMGpod_1": "-2",
            "EMGpod_2": "-8",
            "EMGpod_3": "-54",
        }
    }, ... {
```

Listing 4.1 Example of one of the JSON file contained in the session file for the Myo application.

Sensor attributes

Sensor attributes are relevant components of multimodal data since they constitute the facets through which the sensor measures a phenomenon, for example, the electromyogram data of the learner). The sensor attributes are specified in each application file contained in the session file. When the sensor file is loaded in, the sensor attributes are hierarchically sorted in the Attribute list (Figure 4.2, B, bottom-left). Each sensor application (e.g. "Kinect") is one top-node on the list. The parser of the VIT can nest the sensor attributes using underscores. For example, if the parser reads an attribute named "3_Hand_Right_Tip_X", it will nest the attribute in the following hierarchy 3 > Hand > Right > Tip > X. The sensor attributes can be either numerical time-series or categorical series. Multiple sensor attributes can be selected with a checkbox on the Attribute list, this action will update the chart plotting all the selected attributes (Figure 4.2, F, centre-right).

Time intervals

Time intervals allow partitioning the sensor recording into shorter segments that have different purpose and meaning. The user can add a new time interval by selecting a time region under the plot F. Each time interval is represented as light yellow rectangles in the plot, which as a segment of the x-axis (time dimension). The time interval is defined by their *start-time* and *end-time*, which is the relative duration in seconds from the start of the session. All the added time intervals are listed in panel D. The user can click on one or more time intervals on this list and highlight visually the time intervals in the plot which will turn from light yellow to blue. The time intervals are added independently by the sensor attributes. The sensor attributes, however, vary within each time intervals and their value change is important for the annotations. Time intervals can also overlap.

Annotations

The intervals define portions of the sensor recording with a specific meaning which can be assigned through custom properties. We call these properties annotations. In the case of our implementation, the annotations are provided in the time domain rather than the spatial or the frequency domain. The user provides an annotation that is valid for a time interval, i.e. between two time points of the multimodal dataset. In contrast, the user cannot provide an annotation to specific regions of the video or to certain frequencies of the sensor attributes. Each time interval can be annotated with *0-to-n* annotations. The annotations can be both loaded via the Annotation file (Figure 4.2, C, top-centre) or can be defined manually (Figure 4.2, E, centre). All the time intervals share the same set of annotations, this leads to an *Attribute-Value* representation in which the annotated dataset can be seen as a table in which each time interval is a row and each annotation constitutes a column. The value not assigned will be treated as a null value. As time intervals can overlap also annotations can also do the same. This allows for annotations at different levels of granularity, and for example, break down activities into smaller sub-activities or atomic actions.

Annotation procedure

The annotation procedure consists of assigning all the time intervals to the right annotation values. To do so, the user can be helped by the video (if available) in identifying the right affiliation between time intervals and annotations. After the annotation procedure, the user can click "export" which enables three different follow-up actions for the user: download the single annotation file in JSON format. The same file format is accepted by the Load annotation panel (Figure 4.2, C, center-top); download the session file in zip format embedding the annotation file. The file will be renamed "*recordingID_annotated.zip*" add the annotations in the Attribute list (Figure 4.2, B, bottom-left) to be able to plot them along with the sensor attributes. This feature works best with numerical data and can be useful to visualise the variations of annotation values.

4.3.3 Data processing

Data transformation

As output, the VIT produces an annotated dataset "*recordingID_annotated.zip*" which can be downloaded and is ready to be transformed. We have implemented this routine in Python outside of the VIT, since, while the transformation requirements might change, the annotated dataset remains unchanged. In Figure 4.3, we provide a graphical representation of this transformation. The routine takes all the time intervals $(t_1...t_n)$ which are specified in the annotation file and make them as one row in the tabular representation. Each of this row has corresponding annotations $(y_1, ..., y_n)$. Consequently, we use the time intervals as time window on the sensor application files, each of them containing a different set of attributes. This windowed selection gives as output multiple smaller time-series, one for each attribute contained in each application file. In each time-series, we then apply feature selection, to extract numerical features to transform the variable length of the smaller time series. A great number of features can be extracted, spanning from the most classic time-domain features (e.g. min, max, st. deviation, variance, mean) to frequency-domain features (eg. no. peaks, Fast Fourier transform). In addition, there are signal specific features, which vary depending on whether the time series describes a physical movement or a physiological response. The appropriateness of the features must be found in the literature dealing with those particular signals. At the end of this process, the end result is a table with i rows (number of intervals) and $(p*w*n+m)$ columns, where p is the number of attributes for the sensor application files, w is the number of aggregating function ϕ, n is the number of applications and m the target values contained in the annotation file.

Model creation

The techniques for processing multimodal data highly depend on the goal of the system. Here we list a set of techniques to derive models from the transformed data.

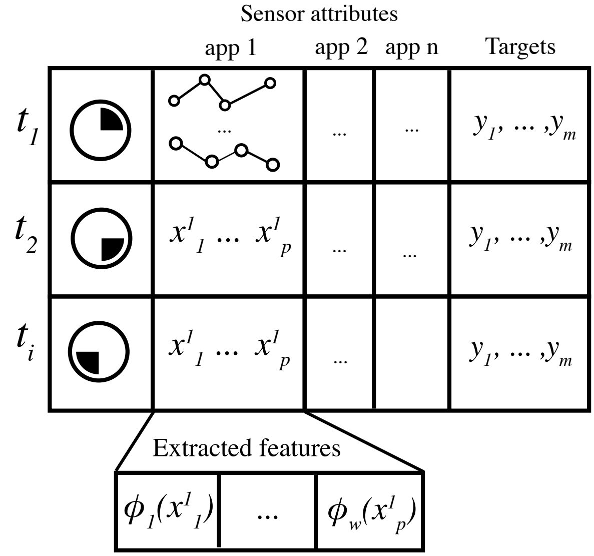


Figure 4.3 Attribute-value transformation of the annotated dataset.

1) Supervised machine learning models are computed *a posteriori* using the labels collected with the annotation as target attributes for prediction or classification. VIT has been designed particularly for using these models. The annotation allows distinguishing optimal from non-optimal learning moments, e.g. if the learning performance meets the learning goals or if the learner commits an error.

2) Unsupervised models do not use labels as the reference system for the optimal learning trajectory, but cluster similar learning moments together. This analytical approach allows discovering patterns of behaviour that are close to each other. Unsupervised techniques can be used in case no annotation scheme can be found. In that case, it could still be useful to use the VIT to segment the learning experience into meaningful parts.

3) Rule-based models, in this case, the sensor thresholds for the ideal performance are presumed *a priori*. This approach is also known as constraint-based models (CBMs). CBMs use constraints in the form of tuples $< C_r, C_s >$, where C_r is the relevant condition and C_s is the satisfaction condition. When C_r applies, then C_s is tested, if not satisfied then feedback is triggered (Kodaganallur and Weitz, 2005). With this approach, VIT does not have a central role and can be bypassed.

4) Probabilistic models such as Bayesian Networks or Structural Equation Modelling. These models also require the VIT and the annotation procedure. Differently from the discriminative models such as supervised or unsupervised learning, these models learn the conditional distribution which can be used for diagnostic or for informing about the relations between different factors that influence the target

Data exploitation strategy	Data processing technique	Feedback medium	Model	Use of the VIT	Actionability	Interpretability
A) Corrective non-adaptive feedback	3) Rule-based models	ITS	Presumed	No	High	High
B) Predictive adaptive feedback	1) Supervised ML models	ITS	Computed	Intervals, annotations	High	Low
C) Patterns identification	2) Unsupervised ML models	Dashboard	Computed	Intervals	Low	Medium
D) Historical reports	Descriptive statistics	Dashboard	No model	No	Low	Low
E) Diagnostic analysis of factors	4) Probabilistic models	Dashboard	Computed	Intervals, annotations	Low	High
F) Comparison	5) Expert-learner comparison	Dashboard	Presumed	Intervals	Low	Medium

Table 4.2 Mapping the processing techniques to exploitation strategies for multimodal data.

attributes.

5) Expert-learner models can compare two or more performances, e.g. learner against the expert's performance to search for the presence of erratic behaviour. This approach requires a pairwise comparison between the expert's performance and trainee's performance. It uses similarity algorithms, such as Dynamic Time Warping considering that sessions might have different execution times. The VIT can be used to segment the multimodal recordings in relevant activities. However, it can become complex when multimodal recordings have hundreds of attributes.

4.3.4 Data exploitation

The data-driven system can have multiple benefits for learning such, for instance: A) providing immediate feedback for correcting erroneous behaviour; B) tailor a more personalised learning experience; C) find patterns in the course of action; D) provide an in-depth overview of the learning process; E) diagnose factors and triggers for learning; F) compare the learner with the expert. All these forms of feedback can be delivered differently, for example, through an ITS or via a dashboard. In Table 4.2, we map the strategies with the data processing techniques discussed in previous section 4.3.3. We classify these strategies based on whether the model is presumed a priori or computed a posteriori; if the output is actionable for the learner (i.e. suggest the learner what exactly to do or improve) and if the model is interpretable (i.e. explains how the model was created).

4.3.5 Technical use cases

To test the VIT in authentic learning scenarios we have used it with three different ITSs designed for three different psychomotor learning scenarios (Figure 4.4) (1) training how to present in public with Presentation Trainer; (2) cardiopulmonary resuscitation training with the CPR Tutor; (3) calligraphy training with the Calligraphy Tutor. For each of these three scenarios, the authors developed an ITS which uses the LearningHub in the backhand for the data collection and synchronisation. This allowed collecting session files with a structure that can be loaded into the VIT.

Presentation trainer

the Presentation Trainer (PT) (Schneider et al., 2015b) is a research prototype designed to support the development of nonverbal communication skills for public speaking (Figure 4.4, top). The multimodal data used are audio and skeleton via Microsoft Kinect. The feedback implemented in the PT is rule-based corrective. The model of ideal presentation has been designed a priori by interviewing presentation experts. The performance of the learner is compared against such a model. PT is only able to detect gestures, not able to differentiate among types of gestures such as iconic, deliberate, unconscious or bigger than usual gestures. The role of the VIT in the data generated by PT is to identify different types of gestures and enhance the feedback of the PT.

Multimodal Tutor for CPR

The Multimodal Tutor for CPR (Di Mitri, 2018) was designed for training people to perform cardiopulmonary resuscitation using patient manikins (Figure 4.4, centre). The tutor uses a multi-sensor setup for tracking the CPR execution and generating personalised feedback The feedback is based on the multimodal data it predicts when the CPR correct posture for chest compression is violated. It uses supervised machine learning models to generate these predictions. The multimodal data considered are trainee's body position (with Microsoft Kinect), electromyogram (with Myo armband) and compression rates data derived from the manikin. The VIT was used in the CPR Tutor for annotating the single chest compressions, if and when errors in the from the CPR correct posture occurred. The following binary target attributes were annotated: *copressionRate* (speed of the compression), *compressionDepth* (depth of the compression), *compressionRelease* (hands correctly released from the manikin), *armsLocked* (whether arms were kept locked), *bodyWeight* (if the whole body weight was used when compressing).

Calligraphy Tutor

the Calligraphy Tutor (Limbu et al., 2018a) application runs on Microsoft Surface tablet and is designed to record the fine motor movements of calligraphy experts (Figure 4.4, bottom). The multimodal data used are pen strokes with Microsoft Surface pen, Myo electromyogram data. The feedback provided guides the calligraphy trainees guides to learn calligraphy for the first time. The VIT was used for annotating letters. Letters consisting of different pen strokes were grouped into different intervals, and their execution was evaluated by a calligraphy expert.

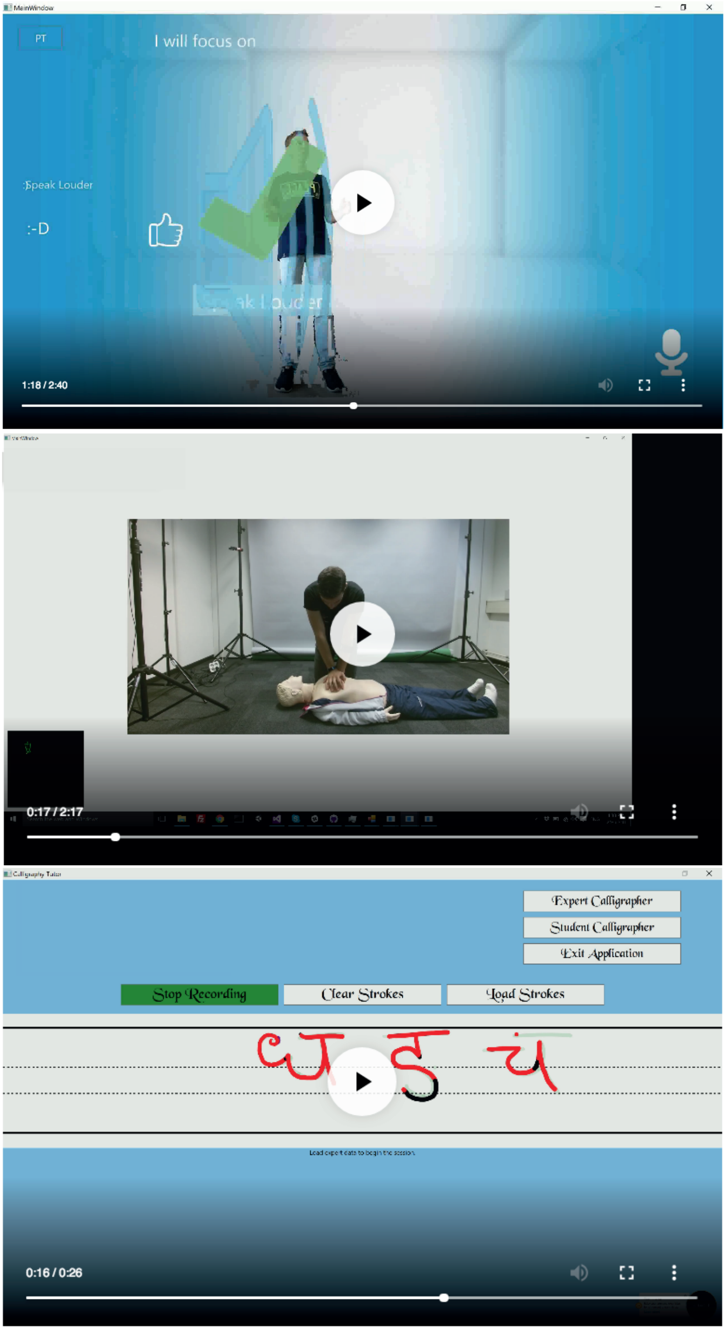


Figure 4.4 Use cases for the VIT: (1) Presentation trainer, (2) CPR Tutor, (3) Calligraphy Tutor.

4.4 Discussion

We answered RQ1 by listing six functional requirements (section 4.3.1) and by describing how the components of the VIT fulfil these requirements (section 4.3.1). The testing of VIT in the three technical use cases (section 4.3.5) shows that VIT in combination with the LearningHub are tools that can be applied in a variety of learning tasks. There are, however, some limitations: the scenarios used as technical use cases are characterised by learning experiences conducted by one learner individually and the learner executes practical tasks which are evident for the sensors. We have not tried VIT for collaborative settings or for learning activities that require cognitive tasks which are difficult to capture with sensors, these areas require future work.

To answer RQ2 we show in section 4.3.2 how to transform the output of the VIT in tabular representation (attribute-value) having the annotated time intervals as individual learning samples. This transformation provides the researcher a dataset ready for being further analysed with a variety of techniques, some of these summarised in section 4.3.3.

Finally, we addressed RQ3 providing a set of use cases where the VIT can be applied and support specific learning tasks (section 4.3.5). The use cases are also enriched with a list of different strategies of exploitation for multimodal data for learning. We consider that the answer to RQ3 is highly relevant for the MMLA community as it contributes to close the feedback loop that begins with multimodal data collection and ends to the learner.

As we evidenced in the review of existing tools in section 4.2.4, the data exploitation, is the part of MMLA that requires more research effort by the community. We also argue that the contribution of the VIT, together with LearningHub and other technical tools for MMLA from our review can be part of an integrated workflow, which can work as a toolkit for MMLA researchers to quickly set up their experiment without having to “reinvent the wheel” and create each time new systems from scratch. We call this the *Multimodal Learning Analytics Pipeline*. The MMLA Pipeline is composed of five steps (as shown in Figure 5.1), corresponding to the five MMLA challenges: (1) collection, (2) storing, (3) processing, (4) annotation and (5) exploitation of multimodal data.

The MMLA Pipeline is a workflow for researchers which allows multimodal tracking of learning activities using wearable sensors, IoT devices, audio and video recordings. There could be multiple routes in the MMLA Pipeline: in Figure 5.1, for example, we show the first four exploitation strategies that we discussed in section 4.3.4: A) Corrective feedback, B) Predictions, C) Patterns, D) Historical reports. The first two routes (A, B) can be implemented within an ITS for, respectively, instantaneous corrections and adaptation. The second two (C, D) provide the possibility to orchestrate the learning activity by detecting the frequency of patterns or to raise awareness on the historical development of the modalities. The use cases we tested in section 4.3.5 can be positioned into the MMLA Pipeline. The Presentation Trainer would take route A), as feedback is directly processed in the LearningHub and it requires no storing, annotation, exploitation. The Tutor for CPR will take the route B), as the VIT is used and therefore prediction models are trained. The Calligraphy Tutor will take route D or C, depending on whether the VIT used for annotating the letters.

4.5 Conclusions

In this paper, we introduced the Visual Inspection Tool, which addresses the data annotation challenge and facilitates data processing and exploitation. In contrast to most of the existing MMLA architectures and tools tailored-made for specific learning tasks and sensors, the VIT allows addressing data annotation generically, for any type of psychomotor learning task that can be captured with a customisable set of sensors. The VIT enables the user (1) to triangulate multimodal data with video recordings; (2) to segment the multimodal data into time intervals and to add annotations to the time intervals; (3) to download the annotated dataset and use the annotations as labels for machine learning predictions. The flexibility introduced by VIT terms of dataset accepted makes it with the LearnigHub a structured research solution solving different challenges introduced MMLA. We called this solution Multimodal Learning Analytics Pipeline. The MMLA Pipeline is a new integrated workflow that works as a toolkit for supporting MMLA researchers to set up new experiments in a variety of learning scenarios. Using components from this toolkit allows reducing developing time to set up experiments and facilitates the transfer of knowledge. In the future, we plan to continue working on the MMLA Pipeline, to further improve this structured approach for MMLA, which can bring the benefits of the multimodal data, machine learning and data analysis also into the learning settings beyond mouse and keyboard happening in the physical space.

Chapter 5

The Multimodal Pipeline

We introduce the Multimodal Pipeline, a prototypical approach for the collection, storing, annotation, processing and exploitation of multimodal data for supporting learning. At the current stage of development, the Multimodal Pipeline consists of two relevant prototypes: 1) Multimodal Learning Hub for the collection and storing of sensor data from multiple applications and 2) the Visual Inspection Tool for visualisation and annotation of the recorded sessions. The Multimodal Pipeline is designed to be a flexible system useful for supporting psychomotor skills in a variety of learning scenarios such as presentation skills, medical simulation with patient manikins or calligraphy learning. The Multimodal Pipeline can be configured to serve different support strategies, including detecting mistakes and prompting live feedback in an intelligent tutoring system or stimulating self-reflection through a learning analytics dashboard.

This chapter is based on:

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2019) Multimodal Pipeline: A generic approach for handling multimodal data for supporting learning. *First workshop on AI-based Multimodal Analytics for Understanding Human Learning in Real-world Educational Contexts* (AIMA4EDU). IJCAI'19 Macau, China.

Note: this contribution received the Best Paper Award at the AIMA4EDU workshop at IJCAI'19.

5.1 Introduction

The diffusion of wearable fitness trackers, sensor-rich smartphones, mixed reality headsets, cameras and Internet of Things devices is introducing new technological affordances that can be leveraged in the field of education and learning. Educational researchers are increasingly embedding multi-sensor and multimodal interfaces and approaches to track learner's behaviour in authentic learning contexts. These new technologies allow moving beyond the typical human-computer interaction, where the user sits in front of a computer, and move towards more immersive and multimodal interactive experiences across spaces through the manipulation of physical and digital objects and environments. This paradigm shift allows a more careful investigation of 'psychomotor' learning activities, i.e. those practical skills that require fine coordination between body and mind. In learning science, learning analytics and human-computer interaction, we are witnessing a drastic increase in the use of multi-sensor interfaces and multimodal data sources (Oviatt et al., 2018). Nevertheless, in these fields of research, the technological solutions chosen to gather multimodal data opted primarily for tailor-made and ad-hoc solutions. Researchers are still required to take many architectural decisions to collect their data set to reach a stage they can collect their datasets to do their investigation.

5.2 Proposed solution

To change this idea, we present our scientific contribution: the *Multimodal Pipeline*, a generic approach for systematically collect, store, annotate, process and exploit multimodal data in a learning scenario. The Multimodal Pipeline enables researchers to design their experiment and quickly obtain synchronised multimodal datasets so that they can focus on the data analysis. The Multimodal Pipeline proposes a technological solution to the different steps. With the Multimodal Pipeline, we aim at addressing the lack of tools and support for the MMLA researchers. The Multimodal Pipeline provides an approach for collecting and exploiting multimodal data to support activities across physical and digital spaces. The Multimodal Pipeline facilitates researchers in setting up their multimodal experiments, reducing setup and configuration time required for collecting meaningful datasets. The multimodal data collected can support researchers to design more accurate student modelling, learning analytics and intelligent machine tutoring. Using the Multimodal Pipeline, researchers can decide to use a set of custom sensors to track different modalities, including behavioural cues or affective states. Hence, researchers can quickly obtain multimodal sessions consisting of synchronised sensor data and video recordings. They can analyse and annotate the sessions recorded and train machine learning algorithms to classify or predict the patterns investigated. A comprehensive overview of the Multimodal Pipeline is given in Figure 1. The Multimodal Pipeline is a cycle consisting of five steps, which propose a solution to the five main MMLA challenges.

(1) The *data collection*: techniques used for capturing, aggregating and synchronizing

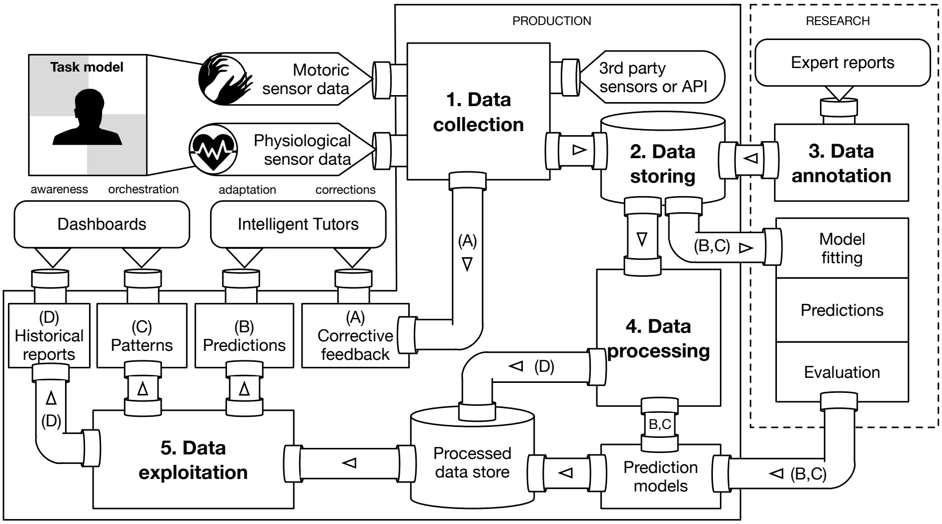


Figure 5.1 Graphical representation of the Multimodal Learning Analytics Pipeline.

data from multiple modalities and sensor streams;

(2) the *data storing*: the approach used for organizing multimodal data which having multiple formats and big sizes, for storing and retrieving them later;

(3) the *data annotation*: how to provide meaning to portions of multimodal recordings and to collect human interpretations through expert or self-reports;

(4) the *data processing*: approach for cleaning, aligning, integrating, extracting relevant features from the 'raw' multimodal data and transforming them into a new data representation suitable for exploitation;

(5) the *data exploitation*: the approach to ultimately support the learner during the learning process with the predictions and the insights obtained by the multimodal data.

The Multimodal Pipeline offers a bird-eye view on the life-cycle of multimodal data that are collected from and used to support the learner. We imagine the Multimodal Pipeline in two phases, the 'research' phase and the 'production' phase. The first one includes several expert-driven operations, such as sensor selections, annotations, model training, parameter tuning. These configurations are used in a later stage of 'production' in which the Multimodal Pipeline is used as the multimodal data backbone infrastructure for collecting the learning data and using them for improving the learning activities. In real-life learning activities, multimodal data can be supportive in different ways. We call these the exploitation strategies. For example, an Intelligent Tutor using the Multimodal Pipeline can prompt instantaneous feedback, nudging the learner towards the desired behaviour. Alternatively, the learner data can be used for retrospective feedback, in the form of an analytics dashboard.

5.3 Technological advantages

Learning activities vary by a significant number of factors. For instance, they can take place inside or outside the classroom, they can be individualised or collaborative, more or less structured. Aiming at creating a system which can support all different combination is an ambitious task. For this reason, we restrict the number of options and better frame the contribution of the Multimodal Pipeline. We use the notion of *Meaningful Learning Task* (MLT), which is an instance of learning activity with a clear 'start' and 'end'. In this time, we define the interval in which the sensor data have to be gathered. We focus on individual psychomotor learning activities, with a maximum of 15 minutes per recording. In the MLT session, the learning activity is recorded through one-to-n sensors having corresponding sensor applications. The learning activity needs to be structured and sequential: it should be possible in one session to identify sequences of smaller steps which can be assessed individually. The assessment or annotation scheme defines the 'goodness' of the learning performance and is highly dependent on the learning activity investigated. It is preferable that the learning task is repetitive, so that it is possible, within one session, to get multiple examples of the same action or movement (e.g. CPR procedure).

5.4 Current prototypes

At the current stage, Multimodal Pipeline consists of two main prototypes: 1) the Multimodal Learning Hub and 2) the Visual Inspection Tool.

5.4.1 The Multimodal Learning Hub

The LearningHub (Schneider et al., 2018) is a research prototype which allows controlling multiple sensor applications. The user can specify one-to-n applications running either in the local machine or in the local computer network. Hence, the user can 'start' and then 'stop' the sensor recording for all the selected applications. Each sensor application will record the data from its connected devices and, once the recording is stopped, it will return a JSON file to the LearningHub with all the sensor updates. Since the LearningHub is activating each application, it can communicate the precise timestamp to all the sensor applications, which allows obtaining the sensor data synchronised with the same clock. In addition to the JSON files, also audio and video can be recorded. All these files will be compressed into a zipped folder: the MLT session. The LearningHub is developed in C# for Windows and released under Open Source[1]. At this moment, there exist a variety of sensor applications library already connected for many existing commercial sensors

[1] Code available at https://github.com/janschneiderou/LearningHub

(Kinect, Myo, Leap, Empatica, Android, etc.). The LearningHub is programmed that is relatively easy integrating a new sensor application.

5.4.2 Visual Inspection Tool

The recorded MLT sessions can be loaded into the Visual Inspection Tool (VIT) (Di Mitri et al., 2019a). VIT allows the manual and semi-automatic annotation of MLT sessions enabling the researcher to 1) triangulate multimodal data with video recordings; 2) to segment the multimodal data into time intervals and to add annotations to the time intervals; 3) to download the annotated dataset and use the annotations as labels for machine learning classification or prediction. The annotations created with the VIT are saved into MLT data-format as the other sensor files. The annotations are treated as an additional sensor application, where each frame is a time interval with relative 'startTime' and 'stopTime' instead of that a single timestamp. Using the standard MLT data-format, the user of the VIT can both define custom annotation schemes or load existing annotation files. Also, the VIT is released with Open Source license[2].

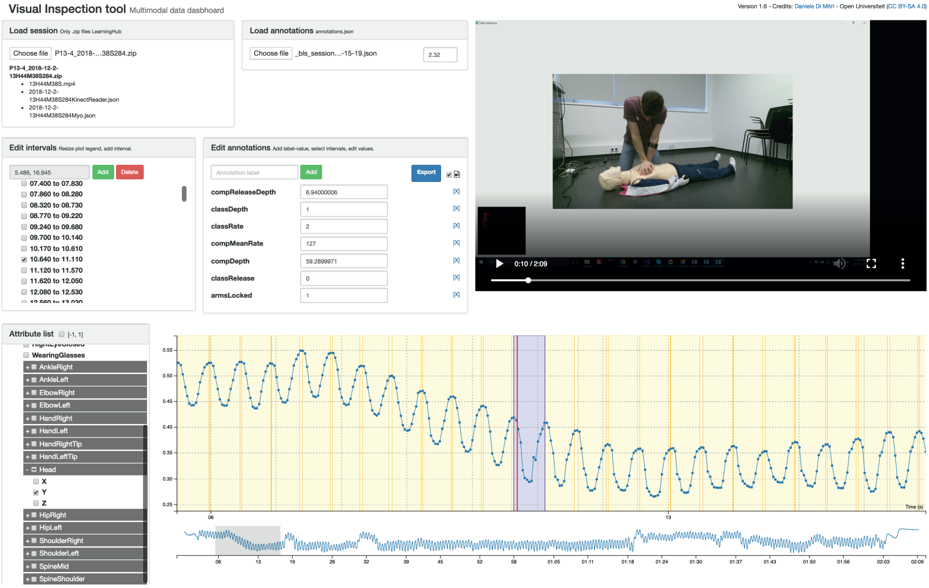
5.5 Practical use cases

The Multimodal can be used for different purposes. In case of the structured task, the Multimodal Pipeline can be used in conjunction with an Intelligent Tutoring System to detect learning mistakes which can be the base for instantaneous and actionable feedback. In the alternative, in the case of less-structured tasks, the data collected with the Multimodal Pipeline can also be summarised into learning analytics dashboards to stimulate reflection from the learner or the teacher. In the following sections, we report three practical use cases in which the multimodal pipeline was used in conjunction with an ITS.

5.5.1 Cardiopulmonary Resuscitation training

We employed the Multimodal Pipeline in the pilot study for the Multimodal Tutor for CPR (Di Mitri, 2018) (figure 5.2). CPR is a highly standardised procedure consisting of repetitive movements. The multi-sensor setup consisted of Microsoft Kinect and Myo armband. In the pilot study, we involved 11 experts and we tracked their body position. We validated the collected data against the performance metrics derived by the ResusciAnne manikin. We also used VIT to annotate additional mistakes currently not tracked by the manikin such as correct locking of the arms and correct use of the body weight. After that, we trained multiple recurrent neural networks each of them achieving a classification accuracy of the performance indicators above 70%. With the trained models, we can implement automatic corrective feedback while the trainee is doing CPR.

[2] Code available at https://github.com/dimstudio/visual-inspection-tool



Visual Inspection tool Multimodal data dasbhoard

Version 1.6 - Credits: Daniele Di Mitri - Open Universiteit (CC BY-SA 4.0)

Load session Only .zip files LearningHub

Choose file P13-4_2018-…38S284.zip

P13-4_2018-12-2-13H44M38S284.zip
- 13H44M38S.mp4
- 2018-12-2-13H44M38S284KinectReader.json
- 2018-12-2-13H44M38S284Myo.json

Load annotations annotations.json

Choose file _bls_session…-15-19.json 2.32

Edit intervals Resize plot legend, add interval.

5.486, 16.945 Add Delete

- □ 07.400 to 07.830
- □ 07.860 to 08.280
- □ 08.320 to 08.730
- □ 08.770 to 09.220
- □ 09.240 to 09.680
- □ 09.700 to 10.140
- □ 10.170 to 10.610
- ☑ 10.640 to 11.110
- □ 11.120 to 11.570
- □ 11.620 to 12.050
- □ 12.080 to 12.530

Edit annotations Add label-value, select intervals, edit values.

Annotation label Add Export ☑

compReleaseDepth	6.94000006	[X]
classDepth	1	[X]
classRate	2	[X]
compMeanRate	127	[X]
compDepth	59.2899971	[X]
classRelease	0	[X]
armsLocked	1	[X]

Attribute list □ [-1, 1]

- □ WearingGlasses
- + AnkleRight
- + AnkleLeft
- + ElbowRight
- + ElbowLeft
- + HandRight
- + HandLeft
- + HandRightTip
- + HandLeftTip
- - Head
  - □ X
  - ☑ Y
  - □ Z
- + HipRight
- + HipLeft
- + ShoulderRight
- + ShoulderLeft
- + SpineMid
- + SpineShoulder

Figure 5.2 Screenshot of the VIT in the CPR use case

5.5.2 Learning a Foreign Alphabet

The second use case considered was learning how to write in a foreign alphabet using the Calligraphy Tutor (Limbu et al., 2018a) (figure 5.3). This tutor allows the expert to write a baseline sentence so that the learner can practice reproduce it with feedback by the tutor. The Calligraphy tutor uses Microsoft Surface and its capacitive pen as well as Myo. The coordinates and pressure of the pen were also combined with myogram and gaze information. The authors used these information to study the features of optimal feedback and correlated it with the user's cognitive load.

5.5.3 Training public skills

The Multimodal Pipeline was also used with the Presentation Trainer (Schneider et al., 2015b) (figure 5.4) a Kinect-based system which gives real-time feedback on different features of the presentation including posture, pauses, volume and hands position. The learners using the presentation trainer were enthusiastic using it as it allowed to receive feedback from practice their presentations. In the Presentation Trainer, the use of multimodal data is two-fold: it is used both for instantaneous and corrective feedback and, at the end of the session, it is presented in form of visual summary for self-reflecting about the performance.

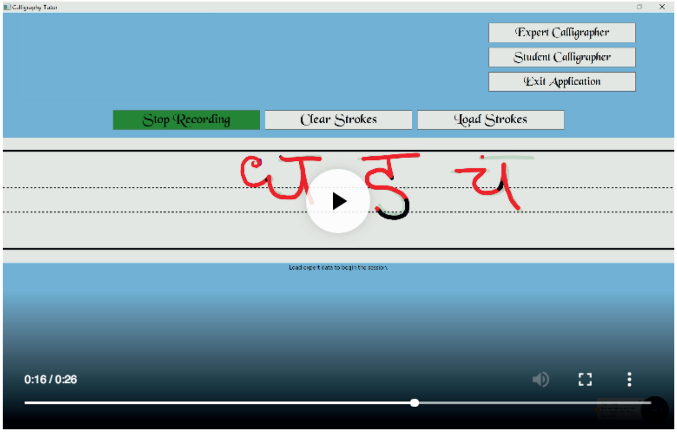


Figure 5.3 Screenshot of the Calligraphy Tutor

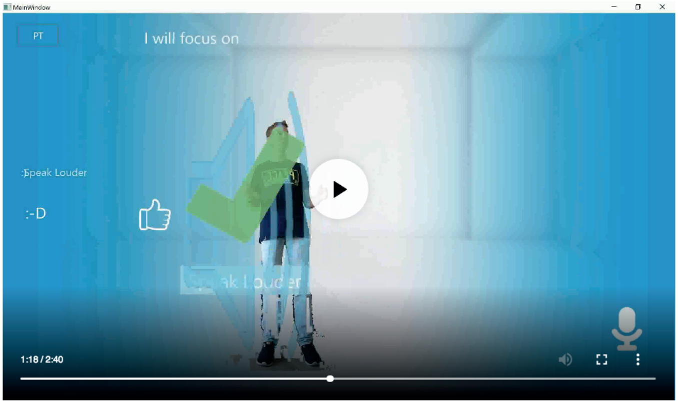


Figure 5.4 Screenshot of the Presentation Trainer

5.6 Future research directions

We plan to progressively improve the Multimodal Pipeline by refining its current components and adding additional ones. We are planning, for instance, to release a data processing called *DataFlow*, which allows to process and run machine learning on the annotated MLT sessions. We are also evaluating the possibility to have a machine learning script for each modality to stack-up modality-dependent classifiers (e.g. one for movements, one for heart rate etc). We are currently working on a runtime feedback engine, the *Multimodal Runtime Framework*, which can channel feedback across sensor applications. In this way, there is a centralised interface where the researcher can set-up feedback rules depending on the task and learning design. Necessary also to point out, the Multimodal Pipeline is a research prototype which has been almost exclusively tested in laboratory settings. We do not exclude in the near future to roll it out in authentic classroom or learning environments.

In addition, with the support of the scientific community in Multimodal Learning Analytics, we are planning to develop additional use cases for the Multimodal Pipeline. For instance, one idea is to collect EEG and electrodermal activity to study visual attention in computer games. Another idea is to develop a smartphone application which can be used by students during classrooms and collects kinematic and interaction data. Moreover, we are optimising the Multimodal Pipeline to also work in collaborative learning situations, using, for example, multiple microphones. This requires a layer of user-identification. With this setup in mind, we are investigating how to extract features from audio and video signals. At this stage, the analysed data are restricted only the sensor data while the videos are used only by the researchers for annotation.

Chapter 6

Detecting Multimodal Mistakes

This study investigates to what extent multimodal data can be used to detect mistakes during Cardiopulmonary Resuscitation (CPR) training. We complemented the Laerdal QCPR ResusciAnne manikin with the Multimodal Tutor for CPR, a multi-sensor system consisting of a Microsoft Kinect for tracking body position and a Myo armband for collecting electromyogram information. We collected multimodal data from 11 medical students, each of them performing two sessions of two minutes chest compressions (CCs). We gathered in total 5254 CCs that were all labelled according to five performance indicators, corresponding to common CPR training mistakes. Three out of five indicators, CC rate, CC depth and CC release, were assessed automatically by the ResusciAnne manikin. The remainder two, related to arms and body position, were annotated manually by the research team. We trained five neural networks for classifying each of the five indicators. The results of the experiment show that multimodal data can provide accurate mistake detection as compared to the ResusciAnne manikin baseline. We also show that the Multimodal Tutor for CPR can detect additional CPR training mistakes such as the correct use of arms and body weight. So far, these mistakes were identified only by human instructors. Finally, to investigate user feedback in the future implementations of the Multimodal Tutor for CPR, we conducted a questionnaire to collect valuable feedback aspects of CPR training.

This chapter is based on:

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2019). Detecting mistakes in CPR training with multimodal data and neural networks. *Sensors*, 19(14), 1–20. DOI: 10.3390/s19143099

6.1 Introduction

Mastering practical skills is a requirement in several professions and working domains. For example, working in construction requires to learn how to use a circular saw; nursing requires learning how to draw blood samples from patients; and repairing clothes requires being able to sew. Practical skill training, also known as psychomotor learning, entails the acquisition of an apprenticeship learning model (Schon, 1983). Typically, in this model, a human expert demonstrates to the learner how to perform a specific task. The learner mimics the expert movements to develop a mental model of the psychomotor skill and, after some practice, this model is automated. For more complex tasks, practical skills are also trained through simulation, allowing the learner to perform the task in an authentic and controlled environment. In simulations, feedback is mostly provided by the human instructor in the debriefing phase after the simulation takes place. Human instructors, however, are not always available to follow each learner step by step, and their time is costly. The lack of instructors leads to the shortage of on-task, real-time and actionable feedback affecting negatively the quality of the training and resulting in longer and less efficient training sessions for the aspiring professionals.

The shortfall of human feedback can be addressed with the employment of Intelligent Tutoring Systems (ITSs), automated computer programs designed to automatically detect learners mistakes and generate personalised adaptive feedback, without requiring the intervention of a human instructor. Although the concept of ITS dates back to decades ago (e.g. Polson et al., 1988), ITS interfaces have been designed as computer or web applications using the standard "mouse and keyboard" interaction paradigm, in which the learner sits in front of a computer. In contrast, practical learning activities take place primarily in physical space or, in some cases, they are complemented with the use of mobile or tablet applications. Reliable tracking of these activities, therefore, requires gathering data from multiple data sources "beyond mouse and keyboard". These various data sources can be seen as modalities of human interaction: speech, biomechanical movements, interaction with peers, manipulation of physical objects, physiological responses or contextual and environmental information. With the diffusion of smartphones, fitness trackers, wearable sensors, Internet of Things, cameras and smart objects (González García et al., 2017), we can nowadays track human behaviour much more easily through multimodal and multi-sensor interfaces (Oviatt et al., 2018). In support to the diffusion of the multimodal approach, comes as well the progress of contiguous scientific research areas such as body sensor networks (Gravina et al., 2017), social signal processing (Vinciarelli et al., 2008), multimodal machine learning (Baltrusaitis et al., 2019) and multimodal data fusion (Lahat et al., 2015).

In the past few years, the multimodal interaction approach has become increasingly popular also in the learning science and learning analytics research communities. This increase of popularity is witnessed by the rising interest in Multimodal Learning Analytics (MMLA) (Blikstein, 2013; Ochoa and Worsley, 2016). MMLA integrates a sensor-based approach with learning analytics, to bridge learning behaviour to complex learning theories (Worsley, 2014). MMLA focuses on how to present learners and teachers insights from multimodal data through visualizations in analytics dashboards (Martinez-Maldonado et al., 2018) or data-storytelling (Echeverria et al., 2018) in support of intuitive decision-making processes in ill-structured learning situations (Cukurova et al., 2019). Recently, the MMLA community focuses more on the challenges of tracking complex learning constructs with multimodal data (Ochoa and Worsley, 2016). Some of the challenges identified include data collection, storing, annotation, processing and exploitation of multimodal data for supporting learning (Di Mitri et al., 2018b).

In this study, we built upon the approach proposed by the MMLA community and further investigated how it can be applied for automatic mistake detection, which, in turn, can be the base for automatic feedback generation for *multimodal tutoring systems* in psychomotor learning scenarios. To do so, we selected cardiopulmonary resuscitation (CPR) training as representative learning task and developed an Intelligent Tutor for CPR using multimodal data (Multimodal Tutor for CPR). This study validated the mistake detection system of the Multimodal Tutor for CPR, a multi-sensor, intelligent tutoring system to train CPR using the ResusciAnne manikins.

The first objective of this study was validating the Multimodal Tutor for CPR according to common performance indicators as implemented in the ResusciAnne manikin. Thus, we used the manikin data as baseline measurements to compare the accuracy of the Multimodal Tutor for CPR to broadly established measures (*Validation*). The second objective was to explore how the Multimodal Tutor for CPR could detect mistakes that are not tracked by the ResusciAnne and are typically only detected by human instructors (*Additional mistake detection*).

CPR is a highly standardised medical procedure with clear performance indicators, therefore there are already various commercial training tools in the market which can assess most of the critical CPR performance indicators automatically. In this study, we leveraged one of these commercial tools, the ResusciAnne QCPR manikin, to derive some key performance metrics of CPR. We complemented the ResusciAnne manikin with a multi-sensor setup consisting of Microsoft Kinect to track the learner's body position, a Myo armband for recording electromyogram data and performance indicators derived from the ResusciAnne manikin. With this setup, we ran a pilot study involving 11 experts. We used different components of the MMLA Pipeline to collect, store, annotate and process the data from the experts. We used the processed data to train five recurrent neural networks, each of them to detect common CPR training mistakes. Three of those mistakes were derived from the ResusciAnne manikin detection system while, the remaining two were annotated manually by the research team.

The paper is structured as follows. In Section 6.2, we introduce the CPR procedure (Section 6.2.1), the concept of ITSs and to what extent they were used in CPR (Section 6.2.2) and we explain the added value of MMLA data for CPR training (Section 6.2.3). Section 6.3 presents related studies. In Section 6.4, we detail the design of this study. In Section 6.5, we report the performance score of the classification models which we discuss in Section 6.6, and answer the two research questions. We also report the results of the participants' questionnaire, aimed at collecting participants' opinions on how to implement effective CPR feedback based on the additional mistakes detected by the models.

6.2 Background

6.2.1 Cardiopulmonary Resuscitation

Cardiopulmonary resuscitation (CPR) is a life-saving technique which is given to someone who is in cardiac arrest. CPR is useful in many emergencies, including a heart attack, near drowning or in the case of stopped heartbeat or breathing. In the present study, we selected CPR as a representative task for the following reasons:

- CPR is a procedure which can be taught singularly to one learner.
- CPR is a highly standardised procedure consisting of a series of predefined steps, that limits the set of possible actions that the learner can take.
- CPR has clear and well-defined criteria to measure the quality of the performance (we use the performance indicators defined by the European CPR Guidelines (Perkins et al., 2015)).
- CPR is a highly relevant skill, which everyone should learn not only medical experts.

The cases of cardiopulmonary arrest are, unfortunately, widespread. The more people are trained to do CPR, the higher the chance of saving lives. For this reason, CPR is currently compulsory in several types of professions, and CPR training is becoming standard practice in several public settings such as schools or public workplaces.

Among the criteria for proper CPR, some indicators, such as correct CC rate, CC depth or CC depth, are more common and tracked automatically by CPR training tools such as the ResusciAnne manikins. Other CPR performance indicators are neglected to down-scale the simulation environment. For this reason, they need to be corrected by human instructors. Examples of these indicators are the use of the body weight or the locking of the arms while doing the CCs. Commercial CPR manikins such as the ResusciAnne manikin do not report on these two indicators, which creates a feedback gap for the learner and higher responsibility for the course instructors.

6.2.2 Intelligent Tutoring Systems

The intelligent tutoring system is a computer program working as “instructor in the box”, meaning that it can provide learners with direct and custom instruction or feedback, without requiring intervention from a human teacher. Traditional ITSs were mostly designed within desktop interfaces for academic education subjects such as geometry and algebra (e.g. Koedinger et al., 1996; Canfield, 2001), and computer science subjects (e.g. Mitrovic and Hausler, 2003). In traditional academic learning, ITS has been proved to be nearly as good as human tutors (VanLehn, 2011), outperforming other instruction methods and learning activities, including traditional classroom instruction, reading printed text or electronic materials, computer-assisted instruction, laboratory or homework assignments (Steenbergen-Hu and Cooper, 2014). Learning academic subjects in desktop-based interfaces is different from learning practical skills in simulations. In the former, the learner's behaviour is represented by the words typed on the keyboard or the clicks on the correct answers. In the latter, the learner's behaviour is determined by the interaction with physical objects using different modalities such as hands movement, gaze or speech. In practical learning scenarios, psychomotor coordination plays a much more prominent role.

In the medical field, some examples of ITS can be found in the literature. The Cardiac Tutor (Eliot and Woolf, 1996) is an ITS for training basic life support tasks such as CPR. It uses clues, verbal advice, and feedback in order to personalise and optimise the learning process. The Collaborative Medical Tutor (COMET) (Suebnukarn and Haddawy, 2007) an intelligent tutoring system for medical problem-based learning that focuses on learners collaboration.

Only a few examples of ITS use multimodal interfaces. D'Mello et al. (2008) enhanced the ITS AutoTutor with a multisensor interface consisting of eye-tracking and posture sensor embedded in the chair where the learner is sitting, so it can detect learners' affective and cognitive states. A comparable setup was also used by Burleson (2007) in the Affective Learning Companion, using a camera for face recognition, learner posture, wrist-based skin conductivity and mouse pressure to detect learners' affective states during game playing. In a more recent study, Schneider et al. (2015b) used a Kinect-based system in the Presentation Trainer, an ITS for training public presentation skills which give both real-time as well as retrospective feedback about the quality of the presentation. Another example is the Calligraphy Tutor by Limbu et al. (2018a), which uses EMG and capacitive pens for training calligraphy and writing in a foreign language.

6.2.3 Multimodal Data for Learning

In the context of education and learning, with the term *multimodal data*, we refer to learner's motoric movements, physiological responses, information of the learning context, environment and activity. Combining data from multiple modalities allows obtaining a more accurate representation of the learning process (Blikstein and Worsley, 2016). Multimodal data, therefore, can be used as historical evidence for the analysis and the description of the learning process (Blikstein, 2013). Multimodal data can be collected using wearable sensors, cameras, Internet of Things (IoT) devices, and computer logs. Research in the field shows there are several existing devices which can be used in the field of learning for collecting data and prompting feedback (Schneider et al., 2015a).

In the field of MMLA, related studies have used multimodal data to investigate learning performances. In the context of classroom activities, Raca and Dillenbourg (2014) analysed student attention from posture and computer vision. D'mello et al. (2015) tracked teacher-student dialogues interaction using audio data. Domínguez et al. (2015) used a tracking device, the *Multimodal Selfie*, to analyse video audio pen strokes of each student.

In group collaboration settings, Ochoa et al. (2013) collected multimodal data including video, audio and pen strokes, to classify expertise. In (Worsley, 2014), researchers recorded video and audio from 13 students building simple structures with sticks and tape. From video data, they derived skeletal position and gesture movements, translating the multimodal transcripts into *action codes* and task-specific patterns.

In the MMLA literature, it is possible to identify two approaches to MMLA: (1) the "analytics" approach which aims at presenting insights to the educational actors, helping them to provide better feedback, particularly useful in ill-structured learning tasks (Cukurova et al., 2019); and (2) the "modelling" approach, used also by the ITS community, which aims at using machine learning to automatically classify or predict learning dimensions. There are, however, examples that combine the two approaches (e.g., (Schneider et al., 2015b; Prieto et al., 2018)), which suggest there is not a "black and white" division but rather a continuum between the ITS and MMLA fields.

In the latter, the intelligent algorithms are fed with the multimodal data associated with task performance measurements and are trained to classify or predict learning goals, training mistakes and prompt on-time automatic and personalised feedback. The modelling approach in the field of MMLA, is described by a recent model, the *Multimodal Learning Analytics Model* (MLeAM) (Di Mitri, 2018).

Although the potentials of MMLA for learning are well documented, practical applications remain a challenge. Multimodal data are messy. To get meaningful and supportive interpretations from multimodal data, intensive data geology steps are required. Moreover, to the best of our knowledge, no standardised procedures exist in this field. In this study, we use an emerging approach, the Multimodal Learning Analytics Pipeline (MMLA Pipeline) (Di Mitri et al., 2019c), in the context of CPR training for creating Multimodal Tutor for CPR. We believe that the selected method can be used as roadmap for the general identification of learners mistakes using multimodal data in authentic training procedures.

6.3 Related Studies

As CPR is a life-saving technique, there is a prosperous research community around the topic, which also publishes in CPR-specific resuscitation journals and conferences. Scouting the related literature, we also found that the idea of using Kinect-based systems for tracking CPR is not new. The study by Semeraro et al. (2012) first piloted a Kinect-based system for providing feedback on CC depth noticing that the depth camera is well suited for the CPR task and that Kinect-based system can improve performance. In the study of Wattanasoontorn et al. (2013), the authors analytically programmed an algorithm to detect the arm posture and CC rate, a weak part of the approach was the calibration process needed to make the detection work. The study in (Wang et al., 2018b) designed a Kinect-based real-time audiovisual feedback device to investigate the relationship among rescuer posture, body weight and CC quality. They tested 100 participants monitoring depth and rate of CC and providing further real-time feedback. The result of this study is that kneeling posture provides better CC than a standing posture and that audio-visual feedback can provide better CC depth, rate, and effective CC ratio. In our study, we proposed the use of a neural network to detect training mistakes in terms of CC rate, CC depth, and CC release, as well as to detect additional training mistakes, not currently tracked by commercial manikins, such as the correct locking of the arms during and the correct use of body posture and body weight during CC. Moreover, in the setup we proposed, we also included a Myo armband to prove the concept of a system learning from multiple modalities. Differently from the previous studies that use tailor-made solutions, the Multimodal Tutor for CPR is a multimodal system that uses generic solutions for data collection described in the MMLA Pipeline.

6.4 Method

This study validated the mistake detection system of the Multimodal Tutor for CPR, a multi-sensor, intelligent tutoring system to train CPR using the ResusciAnne manikins. First, we aimed at validating the Multimodal Tutor for CPR on performance indicators currently implemented in the ResusciAnne using the manikin data as baseline measurements of the CPR performance (RQ1—*Validation*). After that, we explored if the Multimodal Tutor for CPR could detect mistakes not tracked by the ResusciAnne but typically only detected by human instructors (RQ2—*Additional mistake detection*). Our research questions therefore are:

- *Validation*: How accurately can we detect common mistakes in compression rate, compression depth and release depth in CPR training with multimodal sensor data in comparison to the ResusciAnne manikin?
- *Additional mistake detection*: Can we use multimodal data to detect additional CPR training mistakes such as “locking of the arms” and use of “body weight” which are not tracked by ResusciAnne manikin and are only identified by human instructors?

To answer these research questions, we conducted a quantitative observational study in collaboration with AIXTRA, simulation centre of the Uniklink in Aachen, Germany. The experiment involved collecting data from 11 participants. We focused on the quality criteria of the CCs, which is a part of the procedure of the CPR. Each participant performed two sessions of two minutes continuously doing CC, without rescue breaths. For answering RQ1, we used the ResusciAnne manikin data as the baseline measurement to validate the correct mistake detection of the Multimodal Tutor for CPR. To answer RQ2 and mark the presence of additional mistakes in the CPR executions, the research team annotated manually the recorded sessions.

6.4.1 Experimental Setup

The multimodal setup, as represented in Figure 6.1, consisted of the following devices:

- A Microsoft Kinect (v2), depth camera capable of recording three-dimensional skeleton position of the expert, and video record the expert.
- A Myo armband, a Bluetooth device which records electro-myogram and accelerometer data of the person wearing it and provides haptic feedback.
- A Laerdal ResusciAnne QCPR full-body manikin, a popular CPR training manikin optimised for multiple feedback.
- A Laerdal Simpad SkillsReporter, a touchscreen device that couples wirelessly with the ResusciAnne manikin and allows to debrief the overall performance CPR through the assessment of multiple CPR indicators.

The ResusciAnne manikin and its SimPad SkillsReporter are validated CPR instruments, which allow extracting high-quality performance data and they are guideline-compliant according to the official ECR guidelines (Perkins et al., 2015). We used the indicators derived from the SimPad device as our baseline for measuring the quality of the CPR training performance and answer RQ1. On the SimPad device, we used the two-minute-long CC in "evaluation mode".

Among the data that can be retrieved from the SimPad SkillsReporter, we considered the following indicators (the first three in Table 6.1): (1) CC rate; (2) CC depth; and (3) CC release. The remaining two indicators ((4) Arms position; and (5) body position) are not tracked by the SimPad SkillsReporter.

Table 6.1 Considered CPR performance indicators and whether they are detected by the SimPad SkillsReporter and by the human instructor.

Performance Indicator	Ideal Value	SimPad	Instructor
CC rate	100 to 120 compr./min	✓	✓
CC depth	5 to 6 cm	✓	?
CC release	0–1 cm	✓	?
Arms position	elbows locked	✗	✓
Body position	using body weight	✗	✓

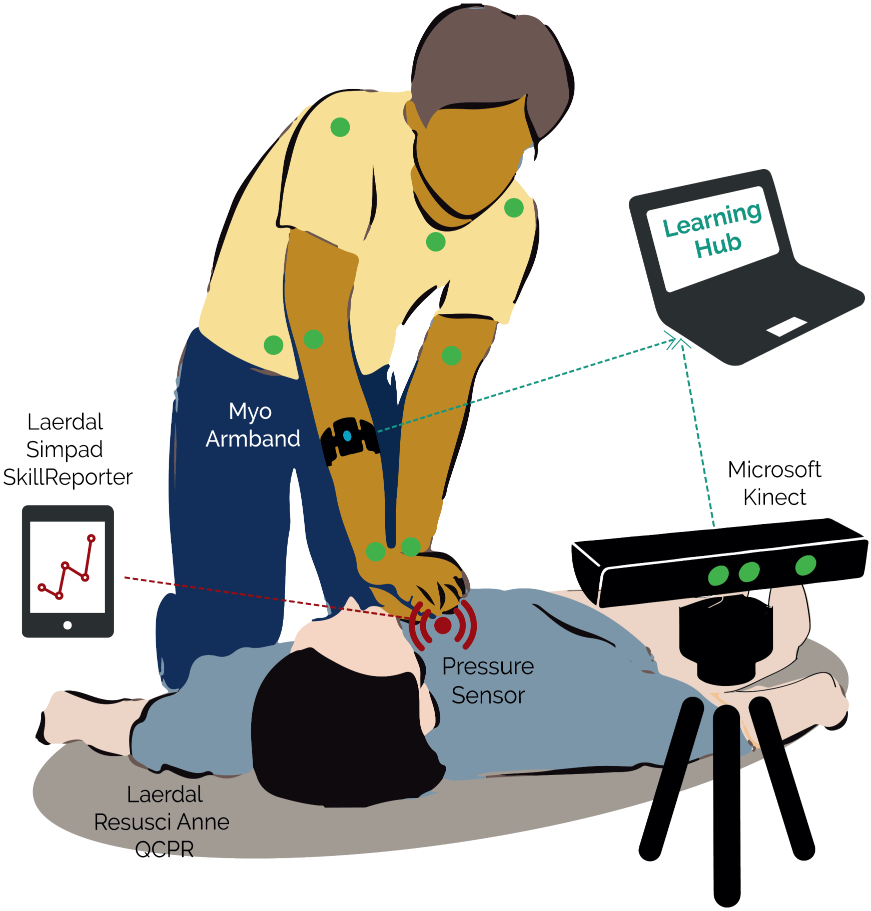


Figure 6.1 The graphic representation of the experimental setting

6.4.2 Participants

CPR is a standard training procedure which requires to be taught by certified trainers. For this reason, we decided that Multimodal Tutor for CPR, at least in the prototyping phase, was not suitable to test complete beginner but rather participants who had previous training knowledge. We selected 14 experts, advanced medical and medical dentist students of the Uniklink Aachen University Hospital. As evidenced in the questionnaire (reported in Section 6.5.3), each participant followed, on average, five CPR training courses. All participants were asked to sign an informed consent letter, describing the experimental details, as well as the data protection guidelines, elaborated following the new European general data protection regulation (GDPR 2016/679). The data collected from the participants were fully anonymised.

6.4.3 Experimental Procedure

The experimental procedure consisted of three phases: (1) prototyping phase; (2) on-site experiment; and (3) analysis. In the first phase, before the on-site experiment, we designed and improved the Multimodal Tutor for CPR iteratively. One or multiple data collection sessions involving one participant, and consequent analysis of the quality of the data collected followed each design iteration improvement of the system. When the prototype reached a satisfactory level, we organised an on-site experiment in cooperation with the University Hospital. Participants were tested individually. Each test consisted of two sessions of two minutes doing CCs separated by a five-minute break, during which the participant had to answer a questionnaire. We recorded each two-minute session separately. Phase 3 was the in-depth data analysis of the data collected. In this phase, we discarded the data from 3 out of 14 participants (6 out of 28 sessions) due to insufficient quality—either caused by faulty Myo readings or incorrect booting of the LearningHub. We narrowed the number of participants considered in the data analysis to 11, for a total of 22 sessions. To identify mistakes "arms properly locked" and correct "body weight", we also recorded five extra sessions with mistakes conducted on purpose by one of the participants.

The technological approach used for the experiment uses the Multimodal Pipeline (Di Mitri et al., 2019c), a workflow for the collection, storage, annotation, analysis and exploitation of multimodal data for supporting learning. We used three existing component implementations of the Multimodal Pipeline. In the next sections, we describe how we used each of these existing components in the experiment of the Multimodal Tutor for CPR.

6.4.4 Data Collection

The data of each session were recorded using the Multimodal Learning Hub (Schneider et al., 2018) (LearningHub), a system for data collection and data storing of multimodal learning experiences using multiple sensors applications. It focuses on short and meaningful learning activities (∼10 min) using a distributed, client–server architecture with a master node controlling and receiving updates from multiple sensor data provider applications. Each sensor application retrieves data updates from a single device, it stores it into a list of frames, and, at the end of the recordings, it sends the list of frames to the LearningHub. In this way, the LearningHub allows collecting data from multiple sensors streams produced at different frequencies. Already various sensor applications have been implemented to work with the LearningHub, both for commercial devices as well as for custom sensor boards. The LearningHub and its sensor data provider applications have all been implemented Open Source. In the Multimodal Tutor for CPR, the LearningHub was used together with three sensor applications, the Kinect data provider, the Myo data provider and a screen recorder. The Kinect data provider collected data from 15 body joints represented as three-dimensional points in space and position features for a total of 60 attributes. The Myo data provider collected accelerometer, orientation, gyroscope and readings from eight EMG sensors for a total of 18 attributes. The screen recorder captured the video of the participant performing the CCs through the point of view of the Kinect. For the data collection, we also used the ResusciAnne manikin data recorded with the SimPad, which provided the baseline assessment of the CPR performance of the participants. This device was not integrated with LearningHub but saved on the local memory of the SimPad, hence transferred via USB to the computer used for the analysis. The SimPad sessions were then manually synchronised with the help of the Visual Inspection Tool (see Section 6.4.6) (Di Mitri et al., 2019a). The most sensitive data, the video recording of the participants, were only included during the annotation phase, exclusively by the research team. The video recordings were eventually taken out from the sessions files making the dataset entirely anonymous. During the experiment, we asked each participant to fill in a short questionnaire soon after the first session. The questionnaire aimed at collecting additional information about the participant's level of previous expertise. We asked questions regarding previous training, kind of feedback received, pros and cons of such feedback and self-perceived performance during the first CPR session. The purpose of this questionnaire was also to gain extra information on how to build useful feedback for the Multimodal Tutor for CPR. The results of this questionnaire are detailed in Section 6.5.3.

6.4.5 Data Storage

The LearningHub uses the concept of *Meaningful Learning Task* introducing a new data format (MLT session file) for data storing and exchange. The MLT session comes as a compressed folder including (1) one or multiple time-synchronised sensor recordings; amd (2) one video/audio of the recorded performance. The sensor recordings were serialised into JSON and have the following properties: an applicationId, an applicationName and a list of frames. The frames have a timestamp and a key-value dictionary of sensor attributes and their corresponding values. In the Multimodal Tutor for CPR, each two-minute CPR session was recorded into a separate MLT session file and stored locally. Each session was 17 Mb and contained initially two JSON files, one for the Myo, one for the Kinect and one MP4 file with the video recording. The example of Myo and Kinect is presented in table view in Table 6.2 and 6.3.

6.4.6 Data Annotation

The CPR annotations were later synchronised with the sessions using the Visual Inspection Tool (VIT) (Di Mitri et al., 2019a). VIT allows the manual and semi-automatic annotation of psychomotor learning tasks which can be captured with a set of sensors. The VIT enables the researcher: (1) to triangulate multimodal data with video recordings; (2) to segment the multimodal data into time intervals and to add annotations to the time intervals; and (3) to download the annotated dataset and use the annotations as labels for machine learning classification or prediction. In addition, the VIT is a software released under Open Source license. The annotations created with the VIT are saved into MLT data-format as the other sensor files. The annotations were treated as an additional sensor application, where each frame is a time interval with relative startTime and stopTime instead of a single timestamp. Using the standard MLT data-format, the user of the VIT can both define custom annotation schemes or load existing annotation files.

A screenshot of the VIT used for the Multimodal Tutor for CPR is given in figure 6.2. We derived the annotation files by the sessions of the SimPad converted into MLT data format. An example of such file is shown in figure 6.3 and in Table 6.4, having an annotation attribute of the three indicators discussed in Table 6.1. As we added the annotation file manually using the VIT, the time-intervals were not synchronised with the other sensor recording. The VIT, however, allowed setting a time offset to align the annotations manually to the sensor recordings. For each of the 22 CPR sessions recorded, we loaded the corresponding annotation file, and synchronised it manually, being guided by the sensor data plots and the video recordings. Hence, we downloaded the annotated session, excluding the video file.

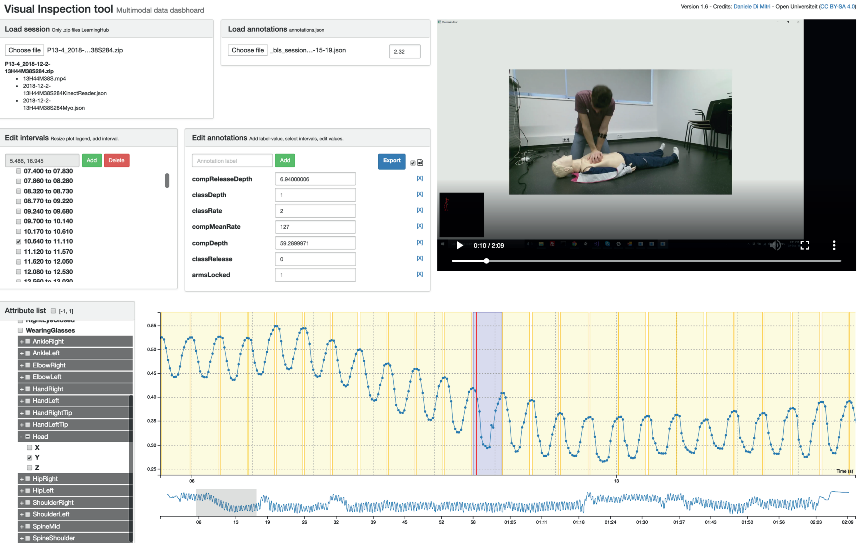
The classification scheme used for the performance indicators summarised in Table 6.1 was based on their ideal interval and on the feedback that the learner would receive. In the case of CC rate, the value of the CC can be either lower than the interval (Class 0) or within the interval (Class 1) or above the interval (Class 2). If labelled with Class 0, the CC rate is "too slow", and then the feedback should be to increase the speed of the CC. If Class 2, the CC rate is "too fast", and then the feedback should be to decrease the speed of the CC. With Class 1, the CC rate is "on point". A similar approach is for CC depth: Class 0 the depth is "too shallow," and the feedback should be to "push harder"; Class 2 is too deep, and the feedback should be to "push less hard"; and Class 1 indicates the CC depth is "on point". CC release, arms position and body position follow instead a binary classification approach, either the task is correctly executed (Class 1) or not correctly executed (Class 0). Once again, we based this decision on the type of feedback the learner should receive, which can be "release the CC completely", "lock your arms" or "use your body weight correctly".

6.4.7 Data Analysis

The 22 annotated datasets were loaded into a Python script using Numpy and Pandas, two data manipulation libraries widely used for statistical analysis. These libraries allow the user to define custom data frames with custom time-indexes ideal for time-series and perform various kinds of vectorised operations. The Python script implemented the following routine. It created a list of all the available sessions, and then iterated and processed each session singularly. The results of the processing are stacked up into a single data frame.

Preparing the Data

The script processes first the annotation file, in order to have all the intervals (i.e., the CCs) into a single data frame DF_{intv}. In the MLT format, the annotation file is different from other sensor files as it is the only JSON file having a list of intervals with start and end, instead of frames with timestamps. The script also computes the



Visual Inspection tool Multimodal data dasbhoard

Version 1.6 - Credits: Daniele Di Mitri - Open Universiteit (CC BY-SA 4.0)

Load session Only .zip files LearningHub

Choose file P13-4_2018-…38S284.zip

P13-4_2018-12-2-13H44M38S284.zip
- 13H44M38S.mp4
- 2018-12-2-13H44M38S284KinectReader.json
- 2018-12-2-13H44M38S284Myo.json

Load annotations annotations.json

Choose file _bls_session…-15-19.json 2.32

Edit intervals Resize plot legend, add interval.

5.486, 16.945

- □ 07.400 to 07.830
- □ 07.860 to 08.280
- □ 08.320 to 08.730
- □ 08.770 to 09.220
- □ 09.240 to 09.680
- □ 09.700 to 10.140
- □ 10.170 to 10.610
- ☑ 10.640 to 11.110
- □ 11.120 to 11.570
- □ 11.620 to 12.050
- □ 12.080 to 12.530
- □ 12.560 to 13.030

Edit annotations Add label-value, select intervals, edit values.

compReleaseDepth	6.94000006	[X]
classDepth	1	[X]
classRate	2	[X]
compMeanRate	127	[X]
compDepth	59.2899971	[X]
classRelease	0	[X]
armsLocked	1	[X]

0:10 / 2:09

Attribute list □ [-1, 1]

- □ RightEyeClosed
- □ WearingGlasses
- + ■ AnkleRight
- + ■ AnkleLeft
- + ■ ElbowRight
- + ■ ElbowLeft
- + ■ HandRight
- + ■ HandLeft
- + ■ HandRightTip
- + ■ HandLeftTip
- - ■ Head
  - □ X
  - ☑ Y
  - □ Z
- + ■ HipRight
- + ■ HipLeft
- + ■ ShoulderRight
- + ■ ShoulderLeft
- + ■ SpineMid
- + ■ SpineShoulder

Figure 6.2 Screenshot of the VIT.

Table 6.2 Table view of the Myo armband data; (2) Kinect camera data; and (3) data from Annotation.json, table view of the JSON in shown in figure 6.3.

frameStamp	EMGPod0	…	EMGPod7	GyrX	AccZ	OriY
00:00:02.0160129	−5.0	…	−4.0	53.0625	−0.137695	−0.199036
00:00:02.0170122	−4.0	…	0.0	45.5625	−0.343750	−0.186706
00:00:02.0297305	−2.0	…	−4.0	45.5625	−0.343750	−0.186706

duration of the interval subtracting *end* with *start*. As the session implement relative timing, where $t_0 = 0$, it is important to add the session date-time to the timer to differentiate between sessions. Consequently, the script processes each of the sensor JSON files. It transforms the list of frames each one with the same set of attributes into a table having the frames as rows and the attributes as columns. It removes the underscores and other special characters from the attribute names and it adds the time-offset. It sets the timestamp as the index and removes the duplicates with the same index. It discards all the attributes whose running total equals zero, which are uninformative attributes. Note that this approach is only applicable to numerical and not categorical data. Finally, it concatenates the results into one single, time-ordered data frame DF_{attr}. From such data frame, some attributes are excluded a priori, as considered of not being informative. In our case, we excluded the data from ankles, hips and head. The hips and ankles as the participants are on their knees when

```
{
    "applicationName": "SimPadAnnotations",
    "recordingID": "bls_session_P0-1",
    "intervals": [
        {
            "start": "00:00:00",
            "end": "00:00:00.660",
            "annotations": {
                "compReleaseDepth": "6.10999966",
                "compMeanRate": "110",
                "compDepth": "59.5200005",
                "classDepth": "1",
                "classRate": "1",
                "classRelease": "0",
                "armsLocked": "1",
                "bodyWeight": "1"
            }
}
```

Figure 6.3 Example of the annotation file derived from the SimPad and loaded in the VIT.

Table 6.3 Table view of the Kinect camera data.

frameStamp	ElbowLeftX	ElbowRightY	HandLeftX	HeadZ	...
00:02.0282288	2.076761	0.275064	−0.520056	1.995384	...
00:02.0607683	2.057799	0.258791	−0.513448	1.996636	...
00:02.0932877	2.019825	0.249333	−0.502858	1.998549	...

doing CPR and heads because sometimes people tend to raise their heads to look up, right or left, which is a movement not influencing the quality of the CCs. At the end of the iteration, we have two data frames:

- DF_{intv}, in which the rows $(t_0..t_n)$ are time intervals (i.e., CCs) and $(y_0, ..., y_m)$ columns are the target indicator values (annotations).
- DF_{attr}, in which the rows $(j_0, ..., j_p)$ are sensor updates and $(a_0, ..., a_q)$ columns are the sensor attributes. As the sensor applications have different update frequencies, DF_{attr} is a sparse matrix having many zeros.

We proceeded to mask DF_{attr} with the time intervals $(t_0..t_n)$, in order to create a new data frame:

Table 6.4 Table view of the JSON in shown in figure 6.3.

start	end	classDepth	classRate	classRelease	compDepth	compMeanRate	compRelease
00:07.070	00:07.730	1	1	0	59.520001	110	6.11
00:07.750	00:08.380	2	0	1	60.910000	98	4.00
00:08.390	00:08.770	1	0	1	60.000000	97	2.00

- DF_{mask}, which has n elements, $(x_0..x_n)$. Each element is an array of $(a_0, ..., a_q)$ time series of about 0.5 s containing the sensor updates for that specific time interval.

The issue with this was that time series in DF_{mask} were of different sizes, not smoothed and with missing values. For this reason, we resampled each of them with equal size ($S = 8$). An example of this resampling is shown in Figure 6.4 (left). The resampling process led us to a tensor of size $(N \times S \times Q)$ where N is the number of intervals, S is the size chosen for the resampling, and Q is the number of attributes. Figure 6.4 (right) shows a graphical representation of the tensor obtained.

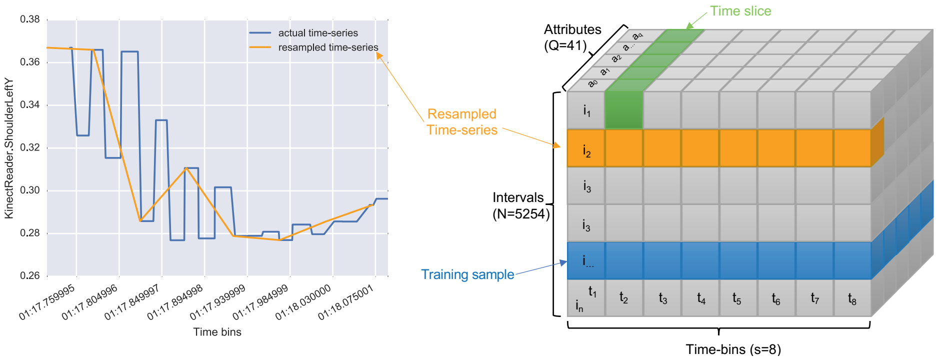


Figure 6.4 Left: example resampling of a time-series interval of the attribute *Kinect.ShoulderLeftY*. Right: a graphic representation of the data transformation into a tensor of shape $(5254 \times 8 \times 41)$.

Training the Neural Networks

For machine learning, we opted for using Keras, an open source neural network library written in Python. We used the Keras implementation of the *Long Short Term Memory networks (LSTM)* (Hochreiter and Schmidhuber, 1997), a type of recurrent neural network (RNN) able to learn over long sequences of data. RNNs are looping networks in which each iteration leaves a footprint, which is used to calculate the following iterations. For this reason, RNNs retain a memory of the previous iterations. When the network updates, however, the memory degrades with a vanishing gradient which can result in the loss of valuable information. To address this issue, LSTM networks use additional long-term memory, where important information are stored to prevent them from degrading over time. The advantage of using LSTM in the CPR domain is able to preserve relevant information throughout the entire CPR session.

We first scaled the tensor's values in a range between 0 and 1. Then, the following sub-routine was applied three times for each target class *classRate*, *classDepth*, and *classRelease*. Then, we split the dataset in 66.6% training and 33.4% test using random shuffling of the training samples. Then, the data were fed into the LSTM neural network with two layers:

- LSTM input layer of size (8×41) feeding into consisting of a hidden layer of 128 units; and
- dense layer sized as unique values of the target class (either 2 or 3).

In Figure 6.5, we present a graphical representation of the configuration of the LSTM neural network. We compiled the LSTM neural network with the training dataset selecting 30 fitting iterations (epochs). As model parameters, the loss function was set using the Sparse Categorical Cross-Entropy was chosen, while we set the accuracy as the performance metrics to evaluate the model. The model was also evaluated using the remaining 33.4% of the dataset. The results obtained are discussed in Section 6.5.

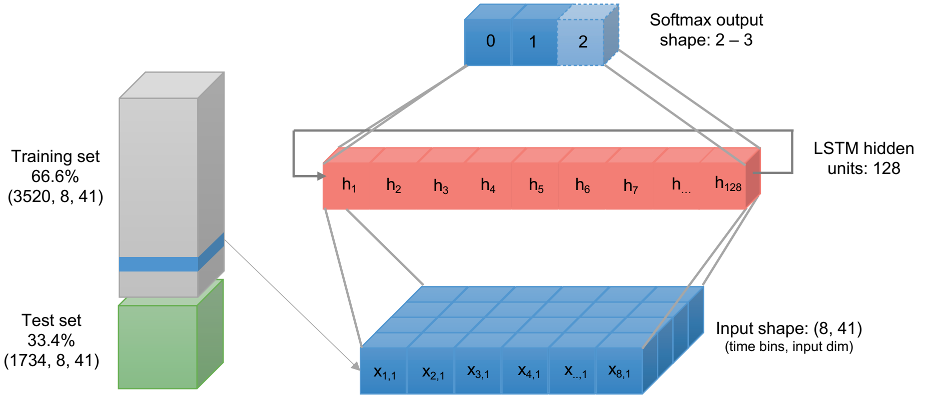

0
1
2
Softmax output shape: 2 – 3
Training set 66.6% (3520, 8, 41)
LSTM hidden units: 128
h_1 h_2 h_3 h_4 h_5 h_6 h_7 $h_{...}$ h_{128}
Test set 33.4% (1734, 8, 41)
Input shape: (8, 41)
(time bins, input dim)
$x_{1,1}$ $x_{2,1}$ $x_{3,1}$ $x_{4,1}$ $x_{...,1}$ $x_{8,1}$

Figure 6.5 Graphical representation of the LSTM neural network configuration.

6.5 Results

The dataset transformed into a tensor of shape $(5254 \times 8 \times 41)$ where 5254 are the learning samples, 41 the attributes and 8 the fixed number of time updates (time-bins) for each attribute. As the ratio chosen between split and test was 66/33, the training set consisted of 3520 samples while the test set in 1734 samples.

The annotation data retrieved from the ResusciAnne manikin are summarised in Table 6.5. The data refer to the total across the 11 participants. While the mean value for compression depth 54.49 mm and for compression release 4.74 mm matches with the CPR guidelines in Table 6.1, the mean compression rate was 121.59 compression/minute, which is slightly above the guideline's range. For this reason, the training mistake which participants seemed to make most often was to compress the chest too fast.

Table 6.5 Annotation data from the SimPad summarised across participants.

Indicator	mean	std	min	max
compDepth (mm)	54.49	6.00	60.22	62.94
compMeanRate (compression/min)	121.59	16.08	133.0	164.0
compReleaseDepth (mm)	4.74	3.64	6.11	30.0
duration (sec)	0.44	0.06	0.48	0.88

We report the individual performance of each participant in the three plots in Figure 6.6. In the case of *classRate* and *classDepth*, the target variables with three possible class-values, it is interesting to acknowledge that each participant shifted almost always in two classes out of three. Participants tended to make only one type of mistake for a particular target, or, in other words, either they were too fast or too slow, but not a combination of both. The performance of a single participant is therefore not representative for the full span of possible mistakes that may occur during training. A more even distribution of mistakes can only be achieved when collecting data from multiple participants.

6.5.1 Neural Network Results

In Figure 6.7, we plot the results of the neural network training of the three classifiers, showing both the loss function values (charts in the top row) and the model's accuracy (charts in the bottom row) through 30 fitting iterations (also called epochs). At each iteration, we compare the results of the training set using 3520 samples (dark line) with the result of the validation set using 1734 samples (light line). We also mark with the red dashed line the inflection point which highlights when the model starts overfitting the training data. In this point, the training loss function (dark blue line) reduces progressively, whereas the validation loss (light blue line) remains more or less stable. In this way, it is possible to identify the iteration where the training of the model has to be stopped before it starts overfitting the data. For *classRate*, this epoch is around the 24th iteration, for *classDepth* around the 19th, and for *classRelease* around the 2nd.

Nonetheless, in Table 6.6, we can see all the three classifier reached an accuracy higher than 70% before overfitting. The most accurate model is the one classifying *classRate*, the binary class indicating if the CCs were executed at the right rate and

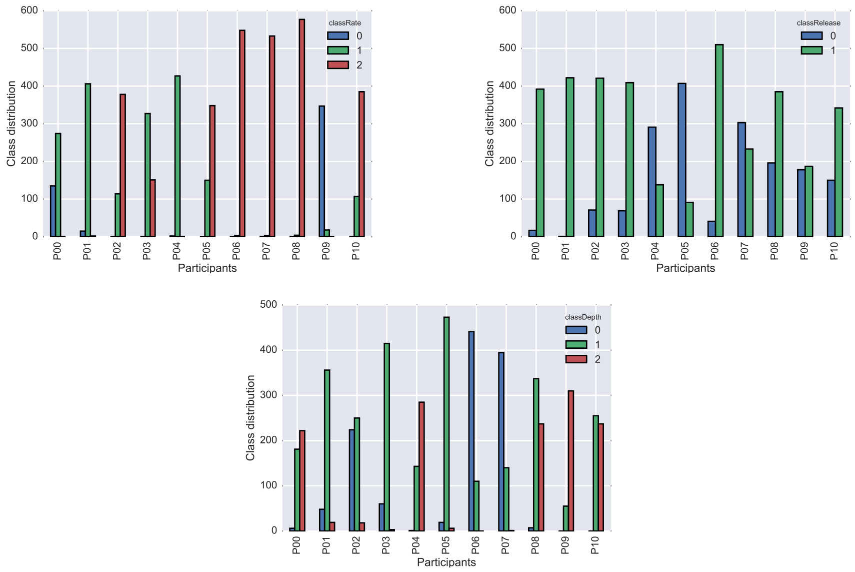


Figure 6.6 Class distribution for each individual participant for: *classRate* (left); *classRelease* (centre); and *classRelease* (right)

speed, followed by *classDepth*, indicating the correct depth of the compression, and *classRelease*, indicating the correct release of the CC. To achieve better accuracy, lower loss function and less premature overfitting, we would need to have more training data.

Table 6.6 Accuracy scores, loss values and ROC-AUC scores for each of the target classes.

	Test accuracy	**Test loss**	**ROC-AUC score**
classRate	0.8650	0.3241	n.a.
classRelease	0.7391	0.5121	0.7305
classDepth	0.7180	0.6144	n.a.

With an accuracy of 86.5%, *classRate* is the best-classified target. A plausible explanation is that the depth-camera sensors, as well as the accelerometer embedded in the Myo, can track temporal-related features such as the acceleration in doing the compression. It is essential to point out that all three models shared the same set of 41 features, all CCs were re-sampled into eight bins and we did not take the actual duration of each CC as an additional feature. The reason *classRelease* and *classDepth* do not perform as well as *classRate* could stem from the fact that the compression and the release are movements of few centimetres and the threshold of correctness is thin to measure, even for a human observer.

The confusion matrices in Figure 6.9 present the result of correct and incorrect classifications of the three models. The tendency we can observe is that the highest number of correct classification (darker colour) happens for class with the highest number of examples (see Figure 6.8 for reference).

6.5.2 Manually Annotated Classes

For the two additional CPR training mistakes of arms not correctly locked (*armsLocked*) and failing to use entire body weight (*bodyWeight*), we used an additional dataset consisting of five sessions with one participant mimicking the two mistakes. The reason we opted for this solution was that all participants in the initial data collection showed very good CPR technique and did not commit these two types of errors.

For the new dataset, we used the same methodology described in Section 6.4. We collected in this case 1107 CCs, 41 attributes and 8 time bins leading to a tensor of size (1107 × 8 × 41). We trained two LSTM neural networks for 100 iterations, the performances are shown in Figure 6.10 both for training and validation loss and accuracy showing also the overfitting point. Before the model starts overfitting, we achieved for *armsLocked* 93.4% accuracy and *bodyWeight* 97.8% accuracy, as shown in Table 6.7. We also show for both target classes the Area Under the Receiver Operating Characteristic Curve (ROC AUC), obtaining 93.7% and 98.3%, respectively. The accuracy of the manual annotation is higher due to the fact that the training errors of arms not correctly locked and incorrect body weight are easier to track with a depth camera. The confusion matrices in Figure 6.11 also show the results of the correct or incorrect classifications.

Table 6.7 Accuracy scores, loss function values and ROC-AUC score for the two manually annotated classes.

	Test Accuracy	**Test Loss**	**ROC-AUC Score**
armsLocked	0.9344	0.1595	0.9371
bodyWeight	0.9781	0.0694	0.9833

6.5.3 Questionnaire Results

The questionnaire entailed seven questions and had 14 respondents. The first question concerned the number of CPR trainings. The answers spanned from 1 to 15 trainings, as shown o in Figure 6.12 (left).

The second question required an open answer and it concerned the type of feedback received during those training. All respondents answered they had received verbal feedback from the CPR trainer. Three respondents mentioned having received feedback also directly from the manikin, and two from the device connected to the manikin reporting the performance. One respondent mentioned the feedback was also written other than verbal, another that was also visual. Three mentioned the feedback was given in real-time feedback while two others retrospectively.

The third question asked about what was the most crucial aspect of the feedback received. Nine respondent agreed that the feedback from the instructor was the most important and that was because the expert also makes an imitation of either the training mistake or the correct position. Another mentioned aspect was to be able to revise their performances looking at the and depth and frequency of the CCs.

The fourth question asked why the feedback was useful. The answers were more diverse, including to get a better understanding of how to execute CPR optimally; helping to keep calm during an emergency; the instructor showing the corrections to adopt; the possibility to correct mistakes on time; being reminded to not lose strength. A respondent answered that it is difficult to realise their mistakes while another asserted that “even if you know the rules, practising, in reality, is different”.

The fifth question asked about any missing aspect of the feedback received in the previous CPR trainings. Three respondents mentioned lack of information of the depth compression, while other respondents remarked the need for real-time feedback because the CPR procedure is tiring and during execution, there is little awareness of the perceived performance.

The sixth question asked the participants desired feedback in CPR. The most mentioned was real-time audio (five times), then real-time visualisations, dashboards at the end and haptic vibrations or augmented reality visuals.

Finally, we asked the participants to rate their CPR performance with a grade from 1 to 10. The results in Figure 6.12 show that the most frequent grade was 7, followed by 8 and 9.

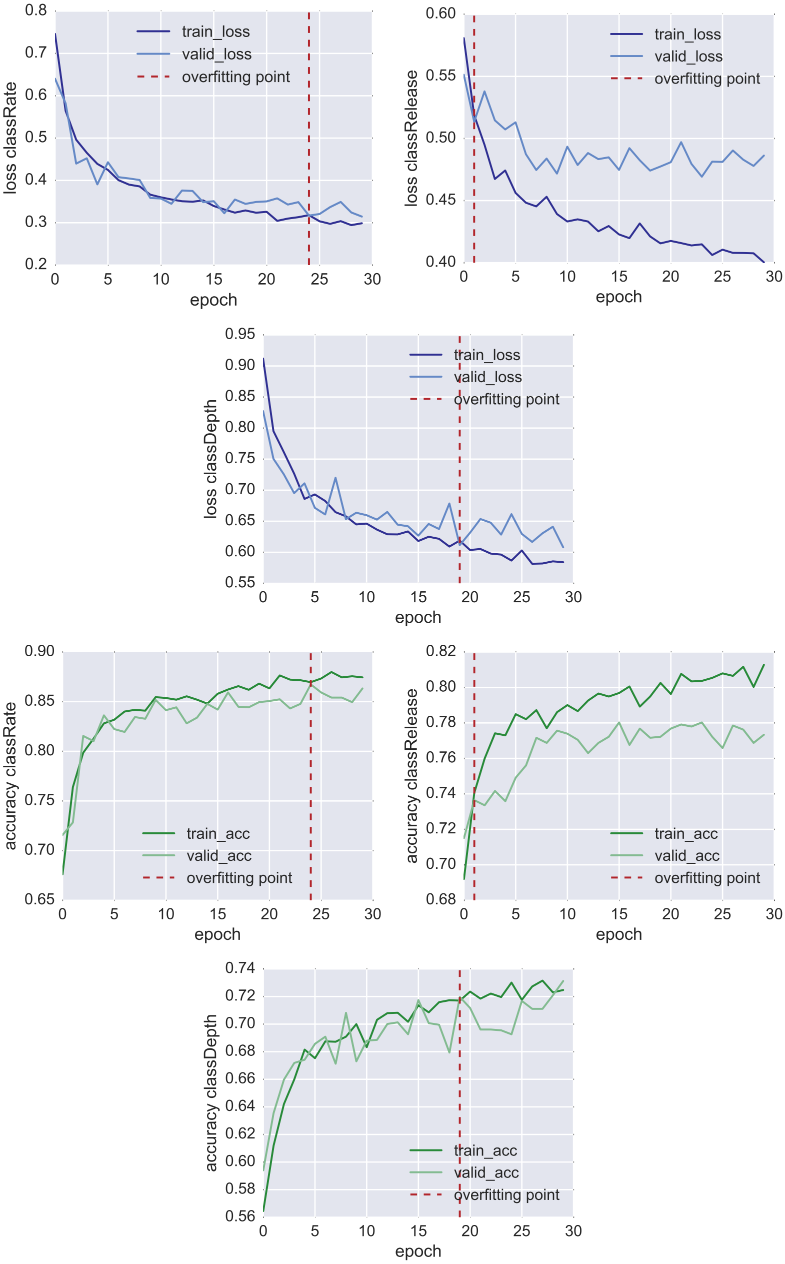


Figure 6.7 Performance values of the three classifiers. Loss function (top) and Accuracy (bottom) of the classifiers: *classRate* (left); *classRelease* (centre); and *classRelease* (right), during training (dark line) and validation (light line). Dashed red lines indicate the "overfitting points", i.e., the epoch in which training loss continues decreasing while the validation loss does not improve anymore.

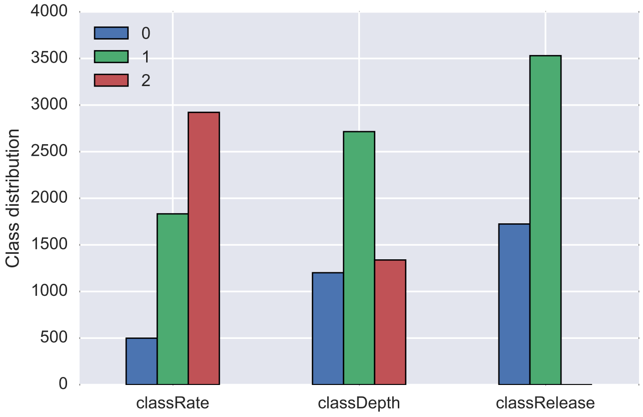


Figure 6.8 Overall class distribution for *classRate*, *classDepth* and *classRelease*.

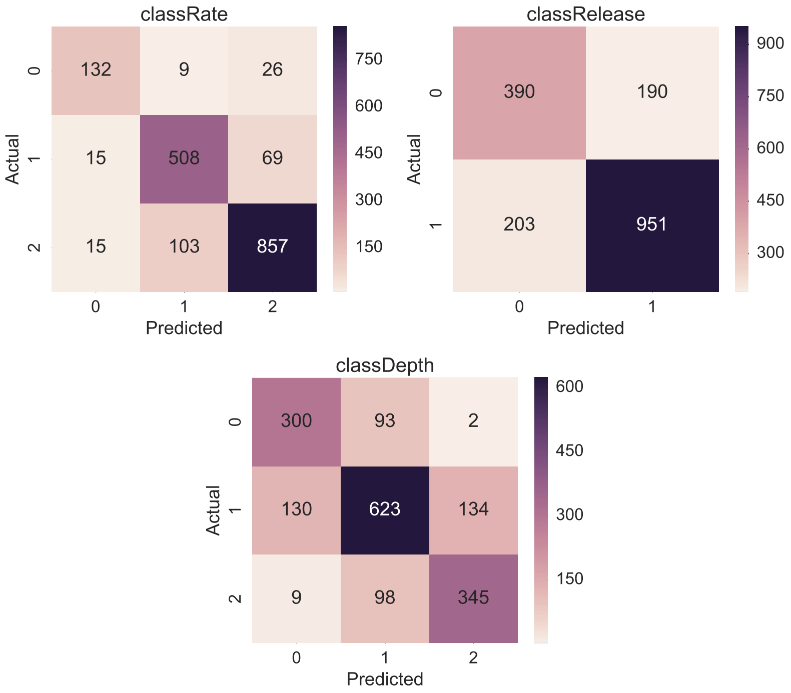


Figure 6.9 Confusion matrices for *classRate*, *classDepth* and *classRelease*. The *y*-axis is the actual class, while the x-axis is the predicted class.

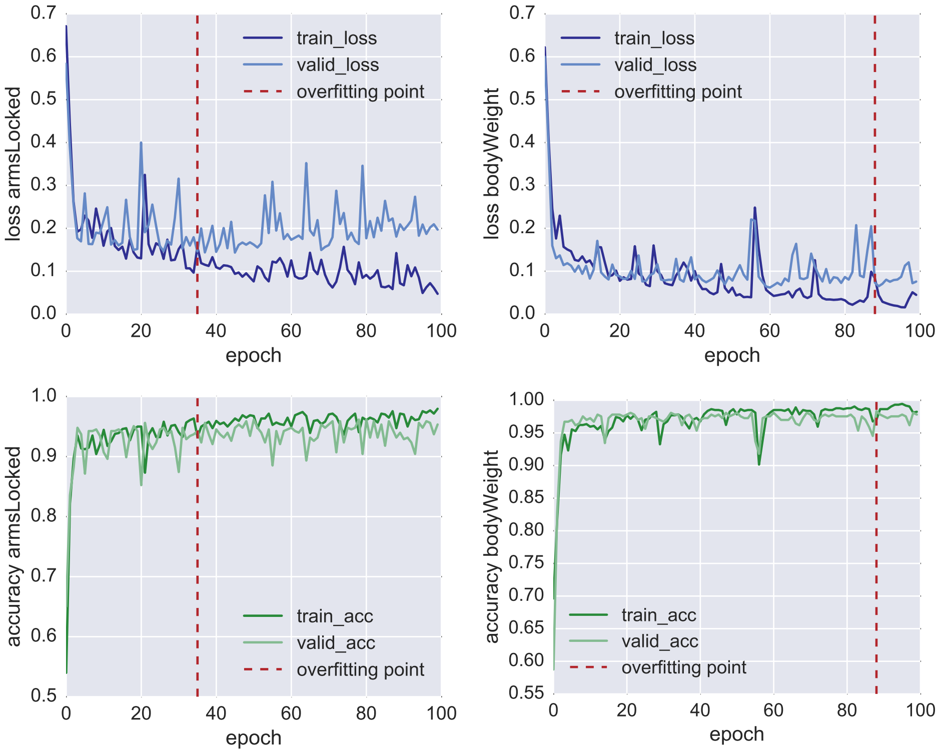


Figure 6.10 Performance values of the classifiers for the two manually annotated classes *armsLocked* and *bodyWeight*. Loss function (top) and Accuracy (bottom) of the classifiers: *armsLocked* (left); and *bodyWeight* (right), during training (dark line) and validation (light line). Dashed red lines indicate the "overfitting points", i.e., the epoch in which training loss continues decreasing while the validation loss does not improve anymore.

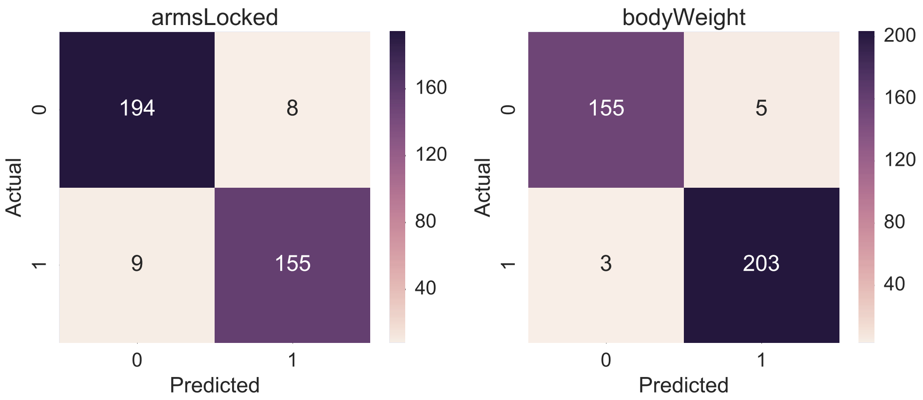


Figure 6.11 Confusion matrices for the manually annotated classes *armsLocked* and *bodyWeight*.

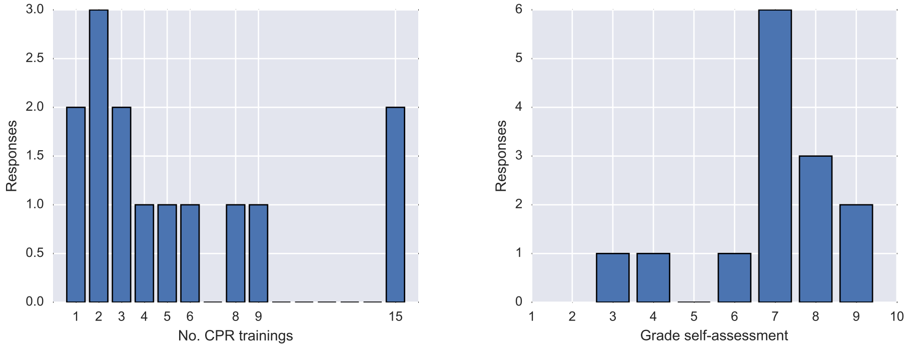


Figure 6.12 (left) Answers on for Question 1, the number of previous CPR trainings for each participant; and (right) the self-assessment ratings given by the participants to their performance.

6.6 Discussion

With the results presented in Section 6.5, we can answer the two research questions of the study as outlined in Section 6.4.

RQ1 focused on the validation of the Multimodal Tutor for CPR and the collected multimodal data against the standardised ResusciAnne manikin measurements (RQ1—*Validation*). With the approach described in Section 6.4, we were able to train three models based on recurrent neural networks, capable of classifying the right compression rate with 86% accuracy, correct compression release with 73% accuracy and correct compression depth with 71% accuracy. Therefore, we can answer RQ1 positively, the Multimodal Tutor for CPR and the collected multimodal data can detect standardised mistakes in CPR training almost as accurately as common CPR training manikins. We also need to point out that the models were not heavily optimised, which makes us confident that they can be improved to achieve even higher accuracy scores.

Concerning the generalisability across learners, all models trained to classify training mistakes tracked with the ResusciAnne manikin generalised well across different users. Training for each participant would also be possible, however, in that case, the data points would have been only around 400 instead of 5200. That would have resulted in a higher accuracy score, but also a higher risk of over-fitting the training data. In general, as CPR is a standard procedure that requires the same and repetitive movements, the individual differences do not seem to hold a strong influence for the classifiers. This condition, however, applies only when the experimental setup is unchanged. In the case the setup changes, then generalising becomes more difficult. For example, if the Kinect were moved to a different location, the values of the sensors would be quite different, or, if a sensor were either added or taken out from the current setup, it would result in a set of different features. A limitation of the approach used therefore is the need to keep the learning activity setup as unchanged as possible.

RQ2 focused on the classification of training mistakes that so far can only be detected by the human instructors and not by the ResusciAnne manikin (RQ2—*Extended mistake detection*). We specifically focused on locked arms and body position mistakes to be detected by the Multimodal Tutor for CPR. To answer this question, we trained our model on additional CPR session where the two training mistakes, *bodyWeight* and *armsLocked*, were mimicked. We achieved an accuracy score of 93% for *armsLocked* and 97% for *bodyWeight*. Based on this high accuracy, we can answer RQ2 positively: the Multimodal Tutor for CPR was able to detect additional CPR training mistakes beyond the mistake detection of the standard ResusciAnne manikin. As with the previous research question, a limitation of our approach is that the set-up must remain unchanged.

6.7 Conclusions

In this paper, we introduce a new approach for detecting CPR training mistakes with multimodal data using neural networks.

The CPR use case was chosen as representative learning task for this study being a standardised and relevant skill. Improving CPR learning feedback could reduce training time and make the CPR procedure more efficient, which would result in greater support for people having a cardiopulmonary arrest and consequently a higher number of lives saved. We designed the Multimodal Tutor for CPR, a multi-sensor setup for CPR, which is a specific implementation of the Multimodal Tutoring System, consisting of a Kinect camera and a Myo armband. We used the Multimodal Tutor for CPR in combination with the ResusciAnne manikin for collecting data from 11 experts doing CPR. We first validated the collected multimodal data upon three performance indicators provided by the ResusciAnne manikin, observing that we can classify accurately the training mistakes on these three standardised indicators. We can further conclude that it is possible to extend the standardised mistake detection to additional training mistakes on performance indicators such as correct locking of the arms and correct body position. Thus far, these mistakes could only be detected by human instructors. After these positive findings regarding the abilities of the Multimodal Tutor for CPR, we envision a follow-up study to investigate different feedback interventions for the learners during CPR training. To facilitate this further research, we asked the participants to fill in a questionnaire to elicit the most relevant aspects of the feedback they received during CPR training the elements which they considered most useful. These principles can be used for the future study of the Multimodal Tutor for CPR, along with the models trained to detect mistakes. In addition, a run-time feedback engine which ensures the multimodal data are captured in real-time and that the feedback is timely.

With the *Multimodal Tutor for CPR*, we demonstrate that multimodal data can be used as the base to drive machine reasoning and adaptation during learning. This can be used both for automatic feedback (modelling approach) as well as for retrospective human feedback (analytics approach).

Among the findings to our research questions, we can also report that the *MMLA Pipeline* for collecting multimodal data from various sensors (see Section 6.4.1) and our MMLA approach for automatically identifying CPR training mistakes, has proven to be highly effective. We are convinced that the suggested approach can be extended to other psychomotor learning tasks. To prove this claim in future works, Multimodal Tutoring Systems for additional psychomotor learning domains need to be developed.

Part IV

Conquest mission

Chapter 7

Keep me in the Loop

We developed the CPR Tutor, a real-time multimodal feedback system for cardiopulmonary resuscitation (CPR) training. The CPR Tutor detects mistakes using recurrent neural networks for real-time time-series classification. From a multimodal data stream consisting of kinematic and electromyographic data, the CPR Tutor system automatically detects the chest compressions, which are then classified and assessed according to five performance indicators. Based on this assessment, the CPR Tutor provides audio feedback to correct the most critical mistakes and improve the CPR performance. To test the validity of the CPR Tutor, we first collected the data corpus from 10 experts used for model training. Hence, to test the impact of the feedback functionality, we ran a user study involving 10 participants. The CPR Tutor pushes forward the current state of the art of real-time multimodal tutors by providing: 1) an architecture design, 2) a methodological approach to design multimodal feedback and 3) a field study on real-time feedback for CPR training.

This chapter is based on:

Di Mitri D., Schneider J., Trebing K., Sopka S., Specht M., Drachsler H. (2020) Real-Time Multimodal Feedback with the CPR Tutor. In: Bittencourt I., Cukurova M., Muldner K., Luckin R., Millàn E. (eds) Artificial Intelligence in Education. AIED 2020. Lecture Notes in Computer Science, vol 12163. Springer, Cham. DOI: 10.1007/978-3-030-52237-7_12

7.1 Introduction

In learning science, there is an increasing interest in collecting and integrating data from multiple modalities and devices to analyse learning behaviour (Blikstein and Worsley, 2016; Ochoa and Worsley, 2016). This phenomenon is witnessed by the rise of multimodal data experiments especially in the contexts of project-based learning (Spikol et al., 2018), lab-based experimentation for skill acquisition (Giannakos et al., 2019), and simulations for mastering psychomotor skills (Santos, 2019). Most of the existing studies using multimodal data for learning stand at the level of "data geology", investigating whether multimodal data can provide evidence of the learning process. In some cases, machine learning models were trained with the collected data for classifying or predicting outcomes such as emotions or learning performance. At the same time, the existing research that uses multimodal and multi-sensor systems for training different types of psychomotor skills feautures neither personalised nor adaptive feedback (Santos, 2016).

In this study, we aimed at overcoming this knowledge gap and by *exploring how multimodal data can be used to support psychomotor skill development by providing real-time feedback*. We followed a design-based research approach: the presented study is based on the insights of (Di Mitri et al., 2019b) (Chapter 6), in which we demonstrated that it is possible to detect common CPR mistakes regarding the quality of the chest compressions (CC) (CC-rate, CC-depth and CC-release). In (Di Mitri et al., 2019b) (Chapter 6), we have also shown that it is possible to extend the common mistake detection of commercial and validated training tools like the Laerdal ResusciAnne manikin with the CPR tutor. We were able to detect the correct locking of the arms while doing CPR and the correct use of the body weight when performing the CCs. The mistake detection models were obtained training multiple recurrent neural networks, using the multimodal data as input and the presence or absence of the CPR mistakes as output. This study extends the previous efforts by embedding the machine learning approaches for mistake detection with a real-time feedback intervention.

7.2 Background

7.2.1 Multimodal data for learning

With the term "multimodal data", we refer to the data sources derived from multimodal and multi-sensor interfaces that go beyond the typical mouse and keyboard interactions (Oviatt et al., 2018). These data sources can be collected using wearable sensors, depth cameras or Internet of Things devices. Example of modalities relevant for modelling a learning task is learner's motoric movements, physiological signals, contextual, environmental or activity-related information (Di Mitri et al., 2018a) (Chapter 2). The exploration of these novel data sources inspired the Multimodal Learning Analytics (MMLA) research (Ochoa and Worsley, 2016), whose common hypothesis is that combining data from multiple modalities allows obtaining a more accurate representation of the learning process and can provide valuable insights to the educational actors, informing them about the learning dynamics and supporting them to design more valuable feedback (Blikstein and Worsley, 2016). The contribution of multimodal data to learning is still a research topic under exploration. Researchers have found out that it can better predict learning performance during desktop-based game playing (Giannakos et al., 2019). The MMLA approach is also thought to be useful for modelling ill-structured learning tasks (Cukurova et al., 2019). Recent MMLA prototypes have been developed for modelling classroom interactions (Ahuja et al., 2019) or for estimating success in group collaboration (Spikol et al., 2018). Multimodal data were also employed for modelling psychomotor tasks and physical learning activities that require complex body coordination (Martinez-Maldonado et al., 2018). Santos et al. reviewed existing studies using sensor-based applications in diverse psychomotor disciplines for training specific movements in different sports and martial arts (Santos, 2019). Limbu et al. reviewed existing studies that modelled the experts to train apprentices using recorded expert performance (Limbu et al., 2018b).

7.2.2 Multimodal Intelligent Tutors

We are interested in the application of multimodal data for providing automatic and real-time feedback. This aim is pursued by the Intelligent Tutoring Systems (ITSs) research. Historically ITSs have been designed for well-structured learning activities in which the task sequence is clearly defined, as well as the assessment criteria and the range of learning mistakes that ITS can detect. Related ITS research looked primarily at meta-cognitive aspects of learning, such as the detection of learners' emotional states (e.g. (D'Mello et al., 2008; Arroyo et al., 2009)). Several ITSs of this kind are reviewed in a recent literature review (Alqahtani and Ramzan, 2019). Most of these studies employed a desktop-based system where the user-interaction takes place with a mouse and keyboard. To find applications of ITSs beyond mouse and keyboard we need to look in the field of medical robotics and surgical simulations into systems like DaVinci. These robots allow aspiring surgeons to train standardised surgical skills in safe environments (Taylor et al., 2016).

7.2.3 Cardiopulmonary Resuscitation (CPR)

In this study, we focus on one of the most frequently applied and well studied medical simulations: Cardiopulmonary Resuscitation. CPR is a lifesaving technique applied in many emergencies, including a heart attack, near drowning or in the case of stopped heartbeat or breathing. CPR is nowadays mandatory not only for healthcare professionals but also for several other professions, especially those more exposed to the general public. CPR training is an individual learning task with a highly standardised procedure consisting of a series of predefined steps and criteria to measure the quality of the performance. We refer to the European CPR Guidelines (Perkins et al., 2015). There exists a variety of commercial tools for supporting CPR training, which can track and assess the CPR execution. A common training tool is the Laerdal ResusciAnne manikins. The ResusciAnne manikins provide only retrospective and non-real-time performance indicators such as CC-rate, CC-depth and CC-release. Other indicators are neglected and that creates a feedback gap for the learner and higher responsibility for the course instructors. Examples of these indicators are the use of the body weight or the locking of the arms while doing the CCs. So far, these mistakes need to be corrected by human instructors.

7.3 System Architecture of the CPR Tutor

The System Architecture of the CPR Tutor implements the five-step approach introduced by the *Multimodal Pipeline* (Di Mitri et al., 2019c) (Chapter 4), a framework for the collection, storing, annotation, processing and exploitation of data from multiple modalities. The System Architecture was optimised to the selected sensors and for the specific task of CPR training. The five steps, proposed by the *Multimodal Pipeline* are numbered in the graphical representation of the System Architecture in Fig. 7.1. The architecture also features three layers: 1) the Presentation Layer interfacing with the user (either the learner or the expert); 2) the Application Layer, implementing the logic of the CPR Tutor; 3) the Data Layer, consisting of the data used by the CPR Tutor. In the CPR Tutor, we can distinguish two main phases which have two corresponding data-flows: 1) the *offline training* of the machine learning models and 2) the *real-time exploitation* in which the real-time feedback system is activated.

7.3.1 Data collection

The first step corresponds to the collection of the data corpus. The main system component responsible for the data collection is the CPR Tutor, a C# application running on a Windows 10 computer. The CPR Tutor collects data from two main devices: 1) the Microsoft Kinect v2 depth camera and 2) the Myo electromyographic (EMG) armband. In the graphic user interface, the user of the CPR Tutor can ‘start’ and ‘stop’ the recording of the session. The CPR Tutor collects the data of the user in front of the camera wearing the Myo. The collected data consist of:

– the 3D kinematic data (x,y,z) of the body joints (excluding ankles and hips)

– the 2D video recording from the Kinect RGB camera,

– 8 EMG sensors values, 3D gyroscope and accelerometer of the Myo.

7.3.2 Data storing

The CPR Tutor adopts the data storing logic of the *Multimodal Learning Hub* (Schneider et al., 2018), a core component of the *Multimodal Pipeline*. As the sensor applications collect data at different frequencies, at the ‘start’ of the session, each sensor application is assigned to a *Recording Object* a data structure an arbitrary number of *Frame Updates*. In the case of the CPR Tutor, there are two main streams coming from the Myo and the Kinect. The *Frame Updates* contain the relative timestamp starting from the moment the user presses the ‘start’ until the ‘stop’ of the session. Each *Frame Update* within the same *Recording Object* shares the same set of sensor attributes, in the case of the CPR Tutor, 8 attributes for Myo and 32 for Kinect, corresponding to the raw features that can be gathered from the public API of the devices. The video stream recording from the Kinect uses a special type of *Recording Object*, specific for video data. At the end of the session, when the user presses ‘stop’, the data gathered in memory in the *Recording Object*s and the *Annotation Object* is automatically serialised into the custom format introduced by the LearningHub: the *MLT Session* (Meaningful Learning Task). For the CPR Tutor, the custom data format consists of a zip folder containing: the Kinect and Myo sensor file, and the 2D video in MP4 format. Serialising the sessions is necessary for creating the data corpus for the offline training of the machine learning models.

7.3.3 Data annotation

The annotation can be carried out by an expert retrospectively using the *Visual Inspection Tool* (VIT) (Di Mitri et al., 2019a). In the VIT, the expert can load the *MLT Session* files one by one to triangulate the video recording with the sensor data. The user can select and plot individual data attributes and inspect visually how they relate to a video recording. The VIT is also a tool for collecting expert annotations. In the case of CPR Tutor, the annotations were given as properties of every single CC. From the SimPad of the ResusciAnne manikin, we extracted the performance metrics of each recorded session. With a Python script, we processed the data from the SimPad in the form of a JSON annotation file, which we added to each recorded session using the VIT. This procedure allowed us to have the performance metrics of the ResusciAnne manikin as “ground truth” for the training the classifiers. As previously mentioned, the Simpad tracks the chest compression performance monitoring three indicators, the correct CC-rate, CC-release and CC-depth. By using the VIT, however, the expert can extend these indicators by adding manually custom annotations, in the form of attribute-value pairs. For this study, we use the target custom classes *armsLocked* and *bodyWeight* corresponding to two performance indicators, currently not tracked by the ResusciAnne manikins.

7.3.4 Data processing

For data processing, we developed a Python script named SharpFlow[1]. This component is used both for the offline training and validation of the mistake detection classifiers as well as for the real-time classification of the single CCs. In the training phase, the entire data corpus (*MLT Sessions* with their annotations) is loaded into memory and transformed into two Pandas data frames, one containing the sensor data the other one containing the annotations. As the sensor data came from devices with different sampling frequencies, the sensor data frame had a great number of missing values. To mitigate this problem, the data frame was resampled into a

[1] Code available on GitHub (https://github.com/dimstudio/SharpFlow)

fixed number corresponding to the median length of each sample. We obtained, therefore, a 3D tensor of shape (#samples × #attributes × #intervals). The dataset was divided in 85% for training and 15% for testing using random shuffling. A part of the training set (15%) was used as validation set. We also applied feature scaling using min-max normalisation with a range of -1 and 1. The scaling was fitted on the training set and applied on the validation and test sets. The model used for classification was a Long-Short Term Memory network (Hochreiter and Schmidhuber, 1997) which is a special type of recurrent-neural network. Implementation was performed using PyTorch. The architecture of the model chosen was a sequence of two stacked LSTM layers followed by two dense layers:

- a first LSTM with input shape 17x52 (#intervals $times$ #attributes) and 128 hidden units;
- a second LSTM with 64 hidden units;
- a fully-connected layer with 32 units with a sigmoid activation function;
- a fully connected layer with 5 hidden units (number of target classes)
- a sigmoid activation.

All of our classes have a binary class, so we use a binary cross entropy loss for optimisation and train for 30 epochs using an Adam optimiser with a learning rate of 0.01.

7.3.5 Real-time exploitation

The real-time data exploitation is the run-time behaviour of the System Architecture. This phase is a continuous loop of communication between the CPR Tutor, the SharpFlow application and the prompting of the feedback. It can be summarised in three phases 1) detection, 2) classification and 3) feedback.

1) Detection. For being able to assess a particular action and possibly detect if some mistake occurs, the CPR Tutor has to be certain that the learner has performed a CC and not something different. The approach chosen for action detection is a rule-based approach. While recording, the CC detector continuously checks the presence of CCs by monitoring the vertical movements of the shoulder joints from the Kinect data. These rules were calibrated manually so that the CC detector finds the beginning and the end of the CCs. At the end of each CC, the CPR Tutor pushes the entire data chunk to SharpFlow via a TCP client.

2) Classification. SharpFlow runs a TCP server implemented in Python which is continuously listening for incoming data chunks by the CPR Tutor. In case of a new chunk, SharpFlow checks if it has a correct data format and if it is not truncated. If so, it resamples the data chunks and feeds them into the min-max scaler loaded from memory, to make sure that also the new instance is normalised correctly. Once ready, the transformed data chunk is fed into the layered LSTMs also saved in memory. The results for each of the five target classes are serialised into a dictionary and sent back to the CPR Tutor where they are saved as annotations of the CC. SharpFlow takes on average 70 milliseconds to classify one CC.

3) Feedback. Every time the CPR Tutor receives a classified CC, it computes a performance and an Error Rate (ER) for each target class. The performance is calculated with a moving average with a window of 10 seconds, meaning it considers only the CCs performed in the previous 10s. The Error Rate is calculated as the inverse of sum of the performance: $ER_j = 1 - \sum_{i=0}^{n} \frac{P_{i,j}}{n}$ where j is one of the five target classes, n is the number of CCs in one time window of 10s. Not all the mistakes in CPR are, however, equally important. For this reason, we handcrafted five feedback thresholds of activation in the form of five rules. If the ER is equal or greater than this threshold the feedback is fired, otherwise, the next rule is checked. The order chosen was the following: $ER_{armsLocked} >= 5$, $ER_{bodyWeight} >= 15$, $ER_{classRate} >= 40$, $ER_{classRelease} >= 50$, $ER_{classDepth} >= 60$. Although every CC is assessed immediately after 0.5s we set the feedback frequency to 10s, to avoid overloading the user with too much feedback. The modality chosen for the feedback was sound, as we considered the auditory sense the least occupied channel while doing CPR. We created the following audio messages for the five target classes:

1. *classRelease*: “release the compression”
2. *classDepth*: “improve compression depth”
3. *armsLocked*: “lock your arms”
4. *bodyWeight*: “use your body weight”
5. *classRate*: *metronome sound at 110 bpm*.

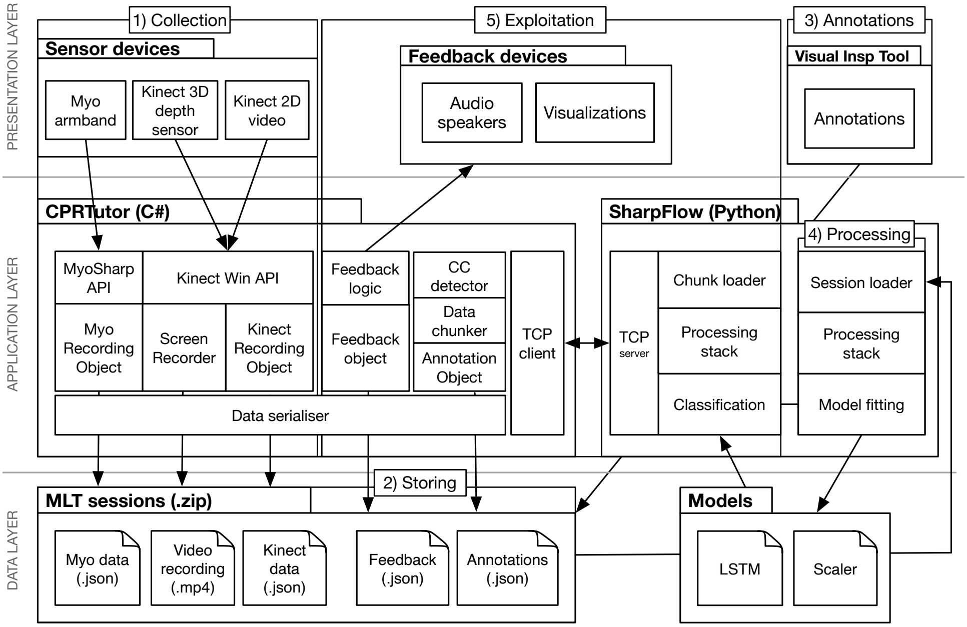


Figure 7.1 The System Architecture of the CPR Tutor.

7.4 Method

In light of the research gap on providing real-time feedback from multimodal systems, we formulated the following research hypothesis which guided our scientific investigation.

H1: The proposed architecture allows the provision of real-time feedback for CPR training.

H2: The real-time feedback of the CPR Tutor has a positive impact on the considered CPR performance indicators.

7.4.1 Study design

To test H1, we developed the CPR tutor with a real-time feedback component based on insights from our design-based research cycle. We planned a quantitative intervention study in collaboration with a major European University Hospital. The study took place in two phases: 1) Expert data collection involving a group of 10 expert participants, in which the data corpus was collected; 2) a Feedback intervention study involving a new group of 10 participants. A snapshot of the study setup for both phases is shown in Fig. 7.2. All participants in the study were asked to sign an informed consent letter detailing all the details of the experiment as well

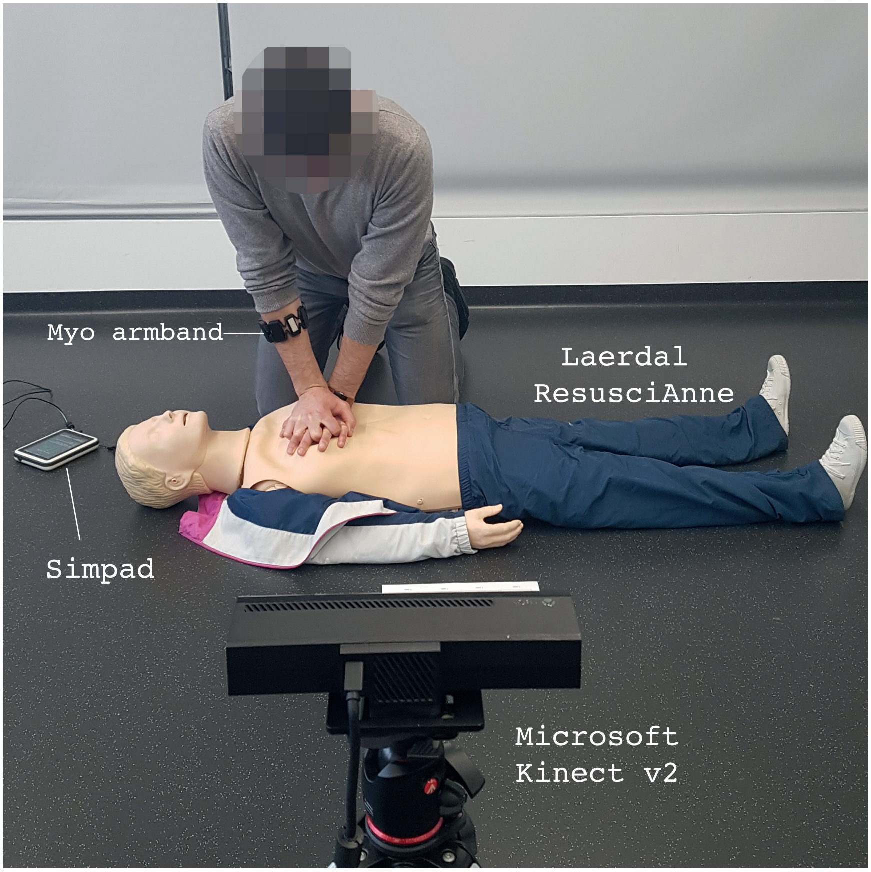


Figure 7.2 Study design of the CPR Tutor.

as the treatment of the collected data in accordance with the new European General Data Protection Regulation (2016/679 EU GDPR).

7.4.2 Phase 1 - Expert data collection

The expert group counted 10 participants (M: 4, F: 6) having an average of 5.3 previous CPR courses per person. We asked the experts to perform 4 sessions of one-minute duration. Two of these sessions, they had to perform correct CPR, while the reminder two sessions they had to perform incorrect executions not locking their arms and not using their body weight. In fact, from the previous study (Di Mitri et al., 2019b) (Chapter 6) we noticed it was difficult to obtain the full span of mistakes the learners can perform. Asking the experts to mimic the mistakes was, thus, the most sensible option for obtaining a dataset with a balanced class distribution. We, therefore, collected around 400 CCs per participant. The one-minute duration was set to prevent that physical fatigue influenced the novice's performance. Once the data collection was completed, we inspected each session individually using the *Visual Inspection Tool*. We annotated the CC detected by the CPR Tutor, by triangulating with the performance metrics from the ResusciAnne manikin. The *bodyWeight* and *armsLocked* were instead annotated manually by one component of the research team.

7.4.3 Phase 2 - Feedback intervention

The feedback intervention phase counted 10 participants (M: 5, F: 5) having an average of 2.3 previous CPR courses per person. Those were not absolute novices but recruited among the group of students that needed to renew their CPR certificate. The last CPR training for these participants was, therefore, older than one year. Each participant in the feedback intervention group performed 2 sessions of 1 minute, one with feedback enabled and one without feedback.

7.5 Results

The collected data corpus from the expert group consisted of 4803 CCs . Each CC was annotated with 5 classes. With the methodology described in sec. 7.3.4, we obtained a tensor of shape $(4803, 17, 52)$. As the distribution of the classes was too unbalanced, the dataset was downsampled to 3434 samples (-28.5%). In Tab. 7.1, we report the new distribution for each target class. In addition, we report the results of the LSTM training reporting for each target class the accuracy, precision, recall and F1-score. In the feedback group, we collected a dataset of 20 sessions from 10 participants with 2223 CCs detected by the CPR Tutor and classified automatically. The feedback function was enabled only in 10 out of 20 sessions. The feedback was fired a total of 16 times. In Tab. 7.2, we report the feedback frequency for each target class and the class distribution for each target class. We generated Error Rate plots for each session. In Fig. 7.3, we provide an example plot of a session having five feedback interventions (vertical dashed lines) matching the same colours of the target classes. Although the Error Rates fluctuate heavily throughout each

Table 7.1 Five target classes distribution and performance of corresponding LSTM models trained on the expert dataset.

class	Class distribution	Accuracy	Precision	Recall	F1-score
classRelease	0: (1475, 42.9%), 1: (1959, 57.1%)	0.905	0.897	0.954	0.925
classDepth	0: (2221, 64.6%), 1: (1213, 35.4%)	0.954	0.955	0.953	0.954
classRate	0: (1457, 42.5%), 1: (1977, 575%)	0.901	0.815	0.819	0.817
armsLocked	0: (1337, 38.9%), 1: (2097, 61.1%)	0.981	0.975	1	0.987
bodyWeight	0: (1206, 35.1%), 1: (2228, 64.9%)	0.97	0.967	0.994	0.98

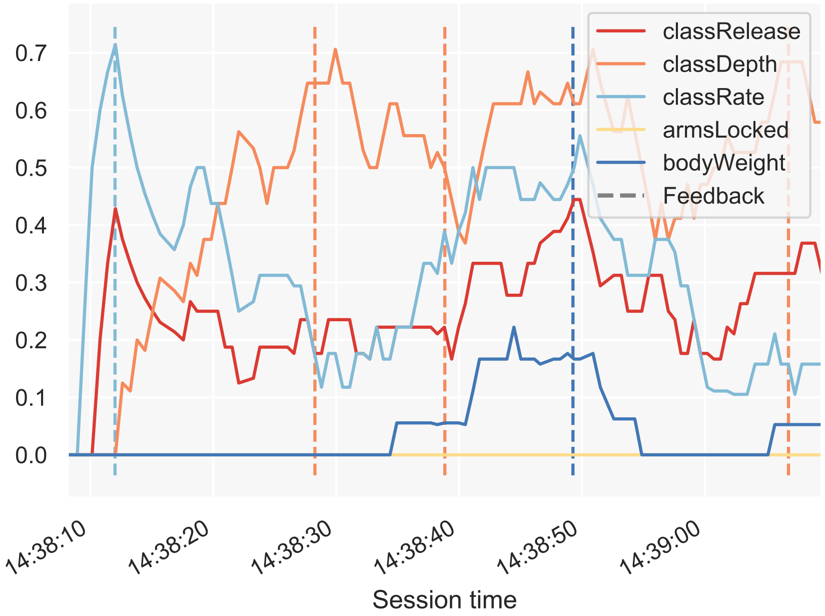


Figure 7.3 Plot of the error rates for one session.

session, we noticed that nearly every time the feedback is fired the Error Rate for the targeted mistake is subject to a drop. We analysed the effect of CPR Tutor feedback by focusing on the short-term changes in Error Rate for the mistakes targeted by the CPR Tutor. In Tab. 7.2, we report the average ERs 10s before and 10s after the audio feedback was fired. We report the average delta of these two values for each target class. For *classRelease*, *classDepth* and *classRate* we notice a decrease of the Error Rate, whereas for *armsLocked* and *bodyWeight* an average increase.

Table 7.2 Average Error Rate for each target class 10s before and 10s after the audio feedback were fired.

Class	Class distribution	Freq. Feedback	ER 10s before feedback	ER 10s after feedback	delta
classRelease	0: (475, 21.4%), 1: (1746, 78.6%)	2	46.60%	33.50%	-13.10%
classDepth	0: (704, 31.7%), 1: (1517, 68.3%)	5	59.80%	55.00%	-4.80%
classRate	0: (475, 21.4%), 1: (1746, 78.6%)	5	44.20%	34.5%	-9.70%
armsLocked	0: (3, 0.1%), 1: (2218, 99.9%)	1	0.6%	5.1%	4.50%
bodyWeight	0: (69, 3.1%), 1: (2152, 96.9%)	3	10.70%	12.90%	2.20%

7.6 Discussion

In H1 we hypothesised that the proposed architecture for real-time feedback is suitable for CPR training. With the System Architecture outlined in sec. 7.3, we implemented a functional system which can be used both for the offline model training of the CPR mistakes as well as for the real-time multimodal data exploitation. The proposed architecture exhibited reactive performances, by classifying one CC in about 70 milliseconds. The System Architecture proposed is the first complete implementation of the *Multimodal Pipeline* (Di Mitri et al., 2019c) and it shows that it is possible to close the feedback loop with real-time multimodal feedback.

In H2 we hypothesised that the CPR Tutor with its real-time feedback function can have a positive impact on the performance indicators considered. With a first intervention feedback study involving 10 participants, we noticed that there is a short-term positive influence of the real-time feedback on the detected performance, witnessed by a decrease of Error Rate in the 10 seconds after the feedback was fired (Tab. 7.2). This effect is confirmed in three out of five target classes. The remaining two classes show opposite behaviours. In these two cases, the increase of Error Rate is smaller as compared to the former target classes. We suppose this behaviour is linked to the extreme class distribution of these two classes. In turn, this distribution can be due to the fact that the participants of the second group were not beginners and therefore not perform common mistakes such as not locking the arms or not using their body weight correctly. These observations cannot be generalised due to the small number of participants tested for the study.

7.7 Conclusions

We presented the design and the development of real-time feedback architecture for CPR Tutor. Building upon existing components, we developed an open-source data processing tool (SharpFlow) which implements a neural network architecture as well as a TCP server for real-time CCs classification. The architecture was employed in a first study aimed at expert data collection and offline training and the second study for real-time feedback intervention allowing us to prove our first hypothesis. Regarding H2, we collected observations that, while cannot be generalised, provide some indication that the feedback of the CPR tutor had a positive influence on the CPR performance on the target classes. To sum up, the architecture used for the CPR Tutor allowed for the provision of real-time multimodal feedback (H1) and the generated feedback seem to have short-term positive influence on the CPR performance on the target classes considered.

General Discussion

This doctoral thesis described the journey of ideation, prototyping and empirical testing of the Multimodal Tutor, a system that supports psychomotor skills acquisition with the support of machine learning and multimodal data. In the introduction chapter, we compare this journey to a *maritime expedition* to the new, promising land of multimodality. The journey consisted on the one hand of the theoretical conceptualisation of multimodality, on the other hand of designing and developing technical prototypes to support the creation of a proof-of-concept of the Multimodal Tutor. This doctoral thesis was divided into four parts.

Part I described the “Exploratory mission”, characterised primarily by the experiment *Learning Pulse* described in Chapter 1.

Part II provided a “Map of Multimodality”. In Chapter 2, we explored the concept of multimodality by analysing existing constructs and by conducting a literature survey. This qualitative research approach led to the formulation of the *Multimodal Learning Analytics Model* (MLeAM). In chapter 3 we described the “Big Five challenges” for the Multimodal Tutor.

In Part III we described the “Preparation of the Navy”, a series of tools needed to be developed for the realisation of the Multimodal Tutor. In Chapter 5 we described the *Multimodal Pipeline*, a generic technical infrastructure which addressed the “Big Five” challenges. The first component of the Multimodal Pipeline was an external tool, the *Multimodal Learning Hub* (Schneider et al., 2018). In Chapter 4, we decided on one specific aspect of the Multimodal Pipeline, the Data Annotation. From this challenge emerged the idea of creating a *Visual Inspection Tool*, an application for annotating and inspecting multimodal data streams, which allowed to “read between the lines”. In Chapter 6 we narrowed the focus to the specific domain of Cardiopulmonary Resuscitation Training (CPR), in particular how to detect multimodal mistakes using machine learning techniques.

Part IV described the conclusive “conquest mission” where the CPR Tutor, an instance of the Multimodal Tutor, was employed in a field study for automatic feedback generation during CPR training. Chapter 7 reported about the design, development and experimental testing of the CPR Tutor.

Main findings

Chapter 1 - Learning Pulse

The first, exploratory study of this doctoral thesis was *Learning Pulse*, described in Chapter 1. Learning Pulse aimed at predicting levels of stress, productivity and level of flow during self-regulated learning. In the study, we gathered multimodal data from nine participants. The data consisted of (1) physiological data (heart-rate and step count) from Fitbit HR wristbands; (2) software applications used on their laptops from RescueTime; and (3) environmental information (temperature, humidity, pressure and geolocation coordinates) using web APIs. In a period of two weeks, the participants had to self-report every working hour via a mobile application, the *Activity Rating Tool*. The data were collected in a *Learning Record Store* using custom *Experience API* (xAPI) triplets. The experimental setup chosen allowed too much diversity of tasks, resulting in an uncontrolled experiment and influencing negatively the quality of the results. Although the nine participants were PhD students of the same department, throughout the two weeks of the data collection, they used different laptops and sets of software applications, which were thus grouped into categories to ease analysis. The collected data were heterogeneous: some attributes such as ‘step-count’ exhibited random behaviour, some other attributes such as ‘heart-rate’ had instead continuous values. To accommodate both types of continuous and random effects we opted for *Linear Mixed Effect Model* (LMEM), a multi-level prediction algorithm typically used for time-series forecasting.

The collection of the labels needed for the data annotation was among the biggest challenges of Learning Pulse. The self-perceived levels of stress, productivity and flow were reported by the participants retrospectively every hour using the Activity Rating Tool. We thus realised that the number of labels was not sufficient for supervised machine learning. For this reason, from each labelled hour, we derived 12 labelled intervals of five minutes. Finally, the data processing approach (Fig. 1.7) was elementary, especially the *Data Processing Application*. The processing pipeline was tailor-made and not flexible, nor reusable for other purposes outside of the experiment. The xAPI format turned out to be a bottleneck when using data exchange and storing of high-frequency sensor data such as heart-rate and step count. Storing each heart-rate update with an xAPI triplet generated a load of redundant information that slowed down the data import and the overall computation. Finally, the poor results in the model accuracy did not allow to explore further the feedback mechanisms.

Findings

- Data collections during long periods need to deal with the task diversity of each user and uncontrolled setups.
- Tracking software applications used by the user leads to diverse sets of attributes for each user, which makes it more difficult to compare them.

- Some modalities are continues variables (e.g. heart-rate), some other are random variables (e.g. step-count), which makes it hard to combine them and analyse them.
- Fixed-time (e.g. hourly) self-reports are not always reliable and are subject to bias.
- There is a trade-off between the number of labels needed for supervised machine learning and the time that humans need to annotate the data.
- Harnessing the potentials of multimodal data require run-time systems such as data processing pipelines instead of data analysis scripts which run only once.
- xAPI is not suitable for storing and exchanging high-frequency sensor data, due to high overhead of the XML format.

Chapter 2 - From Signals To Knowledge

The literature study in Chapter 2 aimed at mapping the state of the art of Multimodal Data for learning, a field which was emerging as *Multimodal Learning Analytics* (MMLA). The exploratory study Learning Pulse in Chapter 1 and the related work done in the field were the main motivations driving this scientific investigation. Surveying the related literature showed that MMLA covered a scattered scientific field and not yet a coherent one. This work contributed to framing the mission of MMLA: using multimodal data and data-driven techniques for filling the gap between observable learning behaviour and learning theories. We coined this mission “from signals to knowledge”. We conducted a literature survey (Section 2.2) of MMLA studies using the proposed *classification framework* in which we separate two main components: the input space and the hypothesis space that are separated by the *observability line*. The literature survey led to the *Taxonomy of multimodal data for learning* and the *Classification table for the hypothesis space*. Surveying the related studies allowed discovering interesting commonalities as, for example, that most of the studies using multimodal data looked primarily at metacognitive dimensions such as the presence of certain emotions in learning.

The literature survey led to propose a new theoretical construct, the *Multimodal Learning Analytics Model* (MLeAM), a conceptual model for supporting the emerging field of MMLA. MLeAM has three main objectives: (1) mapping the use of multimodal data to enhance the feedback in a learning context; (2) showing how to combine machine learning with multimodal data; (3) aligning the terminology used in the field of machine learning and learning science.

Findings

- Sensors can capture observable learning dimensions that include behavioural, activity and contextual data – we refer to this as the *input space*.
- The unobservable learning dimensions such as cognitive, meta-cognitive or emotional aspects stand below the *observability line* – we refer to this as the

hypothesis space.

- Using human-driven data annotation and machine learning makes it possible to infer the unobservable from the observable dimensions. This process is described by the *Multimodal Learning Analytics Model* (MLeAM).
- MLeAM shows how best to exploit machine learning and multimodal data to support human learning.
- The work in MMLA is jeopardised as it cannot yet rely on standardised approaches and techniques.
- Further research efforts must be put in technical prototypes, standardised technical infrastructures, run-time systems and common practices for multimodal data for learning.

Chapter 3 - The 'Big Five' challenges

In Chapter 3, we addressed one structural shortcoming in the MMLA field, as evidenced by the literature survey conducted in Chapter 2: the lack of standardised technical approaches for multimodal data support of learning activities. We claimed that this technical gap is holding back the development of the MMLA field by imposing the MMLA researchers to duplicate efforts in setting up data collection infrastructures and preventing them to focus on data analysis research questions answering. In Chapter 3 the identified technical challenges were grouped into five categories, named the *'Big Five' challenges of Multimodal Learning Analytics* which are the (1) *data collection*, (2) *data storing*, (3) *data annotation*, (4) *data processing* and (5) *data exploitation*. The chapter attempted to provide possible solutions to the challenges which are flexible enough for being employed in different contexts.

Findings

- The technical challenges of MMLA can be grouped into five categories: (1) data collection, (2) data storing, (3) data annotation, (4) data processing and (5) data exploitation.
- The five challenges represent the steps that need to be addressed for implementing a data-driven feedback loop.
- Each of the challenge categories presents a set of sub-challenges which need to be addressed by MMLA researchers.
- Tackling all these challenges together is a complicated research effort.

Additional Research - The Multimodal Learning Hub

As tackling all the five challenges requires a complex effort, we decided to build upon an existing research prototype that a solution for the data collection and synchronisation and the data storing: the **Multimodal Learning Hub** (Schneider et al., 2018). The LearningHub is a platform which can collect data from multiple sensor applications and synchronise them into session files. Although not directly included in this doctoral thesis, the LearningHub is an integral part of the research reported in this doctoral thesis. The biggest research outputs of the LearningHub are (1) a software prototype which can connect to multiple sensor applications running on Windows, and (2) the introduction of a new data storing logic and custom data-format which we coined as *Meaningful Learning Task* (MLT-JSON).

Findings

- Sensor devices have different software systems making the integration of data from multiple sources not trivial.
- Sensors generate data at different frequencies.
- One sensor stream can be composed of several attributes.
- A typical problem of sensor fusion is the time synchronisation of different devices which can be addressed using the LearningHub as a 'master' that decides when the sensor applications should begin collecting the data.
- As continuous data collection is complex and expensive to realise, it is easier adopting a 'batch approach', in which the user can decide when to 'start' and 'stop' the data collection.
- The MLT-JSON format allows creating a document for each sensor device with multiple attributes and stores the data into human-readable format.
- Although MLT-JSON adopts a verbose format (due to repetitive JSON tags), when compressed, its file-size is reduced by 90-95%.

Chapter 4 - Read Between The Lines

In Chapter 4 we focused on one of the five big challenges, the *data annotation*. This challenge deals with how humans can make sense of complex multidimensional data. In this chapter, we proposed a new technical prototype, **the Visual Inspection Tool** (VIT). The VIT allows the researchers to visually inspect and annotate a variety of psychomotor learning tasks that can be captured with a customisable set of sensors. The file format supported by VIT is MLT-JSON, meaning that any recording session recorded with LearningHub can be loaded, visualised and annotated using the VIT. The VIT enables the researcher (1) to triangulate multimodal data with video recordings; (2) to segment the multimodal data into time intervals and to add annotations to the time intervals; (3) to download the annotated dataset and use the annotations as labels for machine learning predictions. Beside generically addressing the data annotation, the VIT also facilitates data processing and exploitation. The VIT is released as Open Source software[2].

[2] Code available on GitHub (https://github.com/dimstudio/visual-inspection-tool)

Findings

- Sensor data are poorly informative when visualised, for this reason, they need to be complemented by evidence interpretable by humans, such as video data without which it is not easy to make sense of what happened in the recorded session.
- The numerical sensor attributes (as opposed to categorical variables) can be visualised as time-series. The visualisation of more than a couple of time-series is tricky for the human eye; manually selecting the attributes to visualise therefore is crucial.
- Audio and video data can be transformed into numerical time-series (e.g. by extracting colours of pixels or audio features) and added in the multimodal dataset.
- The annotation is a human interpretation of the data which apply to a specific time interval with a beginning and end.
- Each time interval (annotation) can consist of multiple attributes, this approach allows the optimal definition of binary and non-binary classes.
- Manually selecting the time-intervals is an expensive task, which should be automated if possible – in the best-case scenario, the human role should be only that of supervising, i.e. correcting and integrating the (semi)-automatic annotations.

Chapter 5 - Multimodal Pipeline

The VIT, as well as the LearnigHub and its custom data format MLT-JSON, constitute a chain of technical reusables which we coined as **the Multimodal Pipeline** and that we described in Chapter 5. The Multimodal Pipeline is an integrated technical workflow that works as a toolkit for supporting MMLA researchers to set up new experiments in a variety of psychomotor learning scenarios. Using components from this toolkit can reduce developing time to set up experiments and it can facilitate and speed up the transfer of research knowledge in the MMLA community. The Multimodal Pipeline connects a set of technical solutions to the “Big Five” challenges described presented in Chapter 5. The Multimodal Pipeline has two main stages, the first one is the ‘offline training’, in which the collected sessions are annotated and the ML models are trained with the collected data. The second stage is the ‘online exploitation’, which corresponds to the ‘run-time’ behaviour of the Multimodal Pipeline.

Findings

- The Multimodal Pipeline describes in technical terms the data-driven feedback cycle proposed by MLeAM in Chapter 2.
- There are two flows of data in the Multimodal Pipeline: the ‘offline-training’ and the ‘online-exploitation’.

- The Data Annotation happens typically before the data processing, as annotations are required for training the models.

- The Data Annotation is not always required. The Multimodal Pipeline can serve different strategies of exploitation for the Multimodal Pipeline, besides predictive feedback using supervised ML (as discussed in Section 4.3.3; these include rule-based corrective feedback, pattern identification, historical reports, diagnostic analysis or expert learner comparison.

- The Multimodal Pipeline can harness multimodal data both for Learning Analytics Dashboards, for example for raising awareness and stimulate orchestration in the learning activities; similarly, it can be embedded in Intelligent Tutors for achieving better adaptation and personalisation of the tutoring experience.

Chapter 6 - Learning Domain: Detecting CPR Mistakes

In Chapter 6 we selected Cardiopulmonary Resuscitation (CPR) as an application case for the Multimodal Tutor. We selected CPR training as a representative learning task for carrying out a study on mistake detection. CPR was chosen primarily because: it is an individual learning task, it is repetitive and highly structured, it has clear performance indicators and because it is a training with high social relevance. Among the different specialisation options that the Multimodal Tutor could take, we decided to focus on the design of a CPR Tutor. We introduced a new approach for detecting CPR training mistakes with multimodal data using neural networks. The proposed system was composed in a multi-sensor setup for CPR, consisting of a Kinect camera and a Myo armband. We used the system in combination with the ResusciAnne manikin for collecting data from 11 experts performing CPR training. We first validated the collected multimodal data upon three performance indicators provided by the ResusciAnne manikin, observing that we can classify accurately the training mistakes on these three standardised indicators. We further concluded that it is possible to extend the standardised mistake detection to additional training mistakes on performance indicators such as correct locking of the arms and correct body position. So far, those mistakes could only be detected by human instructors.

Findings

- The quality of the data training corpus is crucial for ensuring solid model training. Collecting data and training classifiers for a small number of participant leads to very specific models that do not generalise well. Diversity and amount of training data is the key.

- There is no gold-number in the number of annotated samples (chest-compressions - CC) which needs to be collected; there is, however, a dependency with the number of attributes that will be considered.

- Given that the samples (CCs) have different duration, it is important to resample to a fixed number of bins applying some trimming.

- Applying normalisation and min-max scaling of all attributes is important for achieving the best result, this has to follow the activation function used in the neural networks.
- Increasing the number of input attributes (e.g. adding new modalities) increases the classification accuracy of the model; these attributes work as regularisation factor, adding more 'background noise' to the model and making it more robust.
- Neural Networks seem robust in accepting heterogeneous input while converging to good results.
- It is difficult to capture the span of all possible mistake with a restricted number of participants; each participant tends to make only a small subset of mistakes; the solution found was asking participants to mimic some types of mistakes.
- The task structure of 2 sessions of performing 2 minutes of CC is a tiring task for the participants.
- Body size is different among participants and it has an effect on sensor wearing; for instance, people with thinner forearms had some trouble wearing the Myo which was too loose.

Chapter 7 - Keep Me In The Loop

In Chapter 7, we presented the design and the development of real-time feedback architecture for the CPR Tutor. To complete the chain of flexible technical solutions proposed by the Multimodal Pipeline, we developed *SharpFlow*[3], an open-source data processing tool. SharpFlow supports the MLT-JSON format used as well by the VIT and the LearningHub. The data serialised in this format are transformed by SharpFlow into a tensor representation and fed into a Recurrent Neural Network architecture which is trained to classify the different target classes contained in the annotation files. SharpFlow also implements the two data-flows of *offline training* and *online exploitation*. SharpFlow achieves the latter using a TCP server for classifying in real-time every new chest compression. In Chapter 7, the architecture was first employed in an Expert Study involving 10 participants, aimed at training the mistake classification models, and second in a User Study involving 10 additional participants in which the CPR Tutor was prompting real-time feedback interventions.

Findings

- Learning from experts is complicated as experts do not make enough mistakes; instances of mistakes are needed to train the machine learning algorithm; in Chapter 7 we asked the experts to mimic some common mistakes.
- The amount of training data collected from 10 experts was limited; while the findings could not be generalised, they provided some indication that the

[3] Code available on GitHub (https://github.com/dimstudio/SharpFlow)

feedback of the CPR tutor had a positive influence on the CPR performance on the target classes.

- The proposed architecture used for the CPR Tutor allowed for successful provision of real-time multimodal feedback.
- The generated feedback seemed to have a short-term positive influence on the CPR performance on the target classes considered.
- There is a hierarchy among the performance indicators: some mistakes are less frequent but more critical than others and they need to be corrected first; some other mistakes are more frequent, but not so critical.
- Imbalanced class distribution is a real problem; there seems to be an amplifying effect: the majority class in the training set tends to prevail even more in the test set and in the classification of new instances.
- Down-sampling is not trivial; as we had five target classes, down-sampling one class would also affect the other ones; finding a fair balance among the classes was hard.
- Oversampling seemed not trivial either with time-series, generating fake data could undermine the prior class distribution.
- Highest feedback frequency was set to 10s intervals; more frequent feedback would distract or confuse the participant.
- The feedback messages must be explained to the participants beforehand so that they know what to expect and what each message means preventing confusion.
- The SharpFlow online exploitation was very fast (70ms for classifying each instance); in this way, the overall system was not heavily disrupted every time it had to assess every CC.
- For the longer-term influence of the feedback on the target performance indicators we would need to (1) collect data from more participants; (2) increase the number of sessions per participant; (3) select participants with less experience so their performance is not optimal and feedback is fired more frequently.

Contributions of thesis

This doctoral thesis describes the genesis from the ideation phase to final testing of the Multimodal Tutor, a system for adaptive learning experiences from multimodal data capturing. To illustrate the contribution of the thesis, we make use of the metaphor of the *reversed anchor*, as shown in figure 7.4. The top-side of the figure is the *crown* of the anchor which is its widest part and corresponds to the conceptual framing of this thesis, the *Multimodal Learning Analytics Model* (MLeAM). The crown has two ‘arms’. The first arm is the *theoretical ground* corresponding to the learning theories such as *psychomotor learning* and *embodied cognition*. The other arm is the *technology drive*, the new technological affordances such as multimodal and multi-sensor interfaces. The horizontal bar also called ‘stock’ of the anchor, is the *technical approach* for the Multimodal Tutor, i.e. the Multimodal Pipeline. Finally, the narrowest part of the anchor is its ‘ring’ which is the specific ‘learning domain’ we decided to select to demonstrate the functionality of the Multimodal Tutor, i.e. Cardiopulmonary Resuscitation training. With this approach, the wider top of the anchor specialises into a narrower bottom, working as a ‘hook’ for the work described in this doctoral thesis.

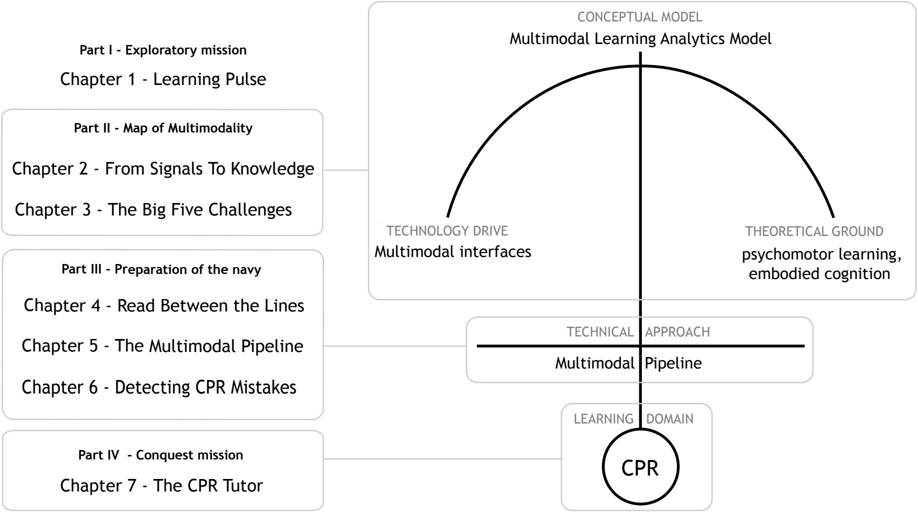

Part I - Exploratory mission
Chapter 1 - Learning Pulse
Part II - Map of Multimodality
Chapter 2 - From Signals To Knowledge
Chapter 3 - The Big Five Challenges
Part III - Preparation of the navy
Chapter 4 - Read Between the Lines
Chapter 5 - The Multimodal Pipeline
Chapter 6 - Detecting CPR Mistakes
Part IV - Conquest mission
Chapter 7 - The CPR Tutor
CONCEPTUAL MODEL
Multimodal Learning Analytics Model
TECHNOLOGY DRIVE
Multimodal interfaces
THEORETICAL GROUND
psychomotor learning, embodied cognition
TECHNICAL APPROACH
Multimodal Pipeline
LEARNING DOMAIN
CPR

Figure 7.4 The structure of this doctoral thesis can be represented as a reversed anchor.

The Multimodal Tutor presents a set of advantages for the MMLA community. It builds on top of a new proposed technological framework, the *Multimodal Pipeline*, which, in turn, is composed of a chain of technological prototypes such as the (1) *Multimodal Learning Hub*, (2) the *Visual Inspection Tool* and (3) *SharpFlow*. All these tools adopt the same data-exchange format (MLT-JSON) and are released under the Creative Commons - ShareAlike 4.0 International license[4].

The main advantage for the MMLA researcher of using such tools is that there is no longer a need to re-invent solutions for data collection, synchronisation, storing, annotation, processing. The MMLA researcher can focus on more specific aspects of their experiments, such as deciding which sensor configuration to use, depending on which modalities need to be monitored or, similarly, deciding what hypothesis to formulate, what unobservable dimensions of learning have to be assessed and how

[4] https://creativecommons.org/licenses/by-sa/4.0/

these dimensions can be translated into an annotation scheme. MMLA researchers can ultimately focus on modelling the learning task, on what the sets of atomic actions are, or on what pedagogical and feedback intervention is suitable for correcting or optimising the performance in each of these actions. Using the Multimodal Tutor, and its underpinning technological frameworks (Multimodal Pipeline) and conceptual model (Multimodal Learning Analytics Model), provides *flexibility* and *multipurposeness*, pushes forward the entire MMLA field. By explaining how to support learning with the use of multimodal data, the Multimodal Tutor generates scientific added value for different data-driven learning research communities, like the ones for Learning Analytics & Knowledge and the Intelligent Tutoring System / Artificial Intelligence in Education. Ultimately, the Multimodal Tutor sets the way for more emerging fields of research such as *Hybrid Intelligence* (Kamar, 2016) or *Social Artificial Intelligence* or Social Robotics (Kanda and Ishiguro, 2017) that focus on how to best interface human communication with artificial (robotic) intelligence.

Limitations of the thesis

Among many advancements in MMLA research, the Multimodal Tutor still carries some limitations. First and foremost, the Multimodal Tutor still consists of a set of research prototypes not ready to be launched in the market as fully working products. To achieve production-ready software there has to be extensive testing, quality-checking or control of the existing functionalities. Within the research applications of the Multimodal Tutor, there exist also additional limitations which can be divided into different levels: (1) learning domain level, (2) hardware level, (3) software level, (4) data level and (5) model level.

At the **learning domain level**, we have been focusing primarily on CPR training, which is a common type of medical simulation. Related research using the components of the Pipeline have been created for Presentation Trainer (Schneider et al., 2015b), Calligraphy Tutor (Limbu et al., 2018a), Tennis Table Tutor (forthcoming). We group all these learning tasks as *individual psychomotor learning tasks in the physical space*, i.e. practical training tasks where the learner has to individually master skills that require a high level of psychomotor coordination that take place in the physical realm. For this reason, in this subset, we intentionally left out learning scenarios such as *cognitive learning*, i.e. tasks that require more reasoning and cognitive abilities, or *social learning*, i.e. tasks that require interaction by multiple actors and or by groups, or *distance and online learning*, including activities mediated by mouse and keyboards. We decided to narrow the focus to make the research contribution of the Multimodal Tutor more evident to the community. At the same time, we believe the boundaries of these scenarios are blurry, therefore the proposed categorisation may run into inconsistency. As specified in the next section, we firmly believe that in the future the Multimodal Tutor can evolve to support also different types of learning scenarios outside of its current focus. Modelling of the learning task is a fundamental part to assess how the Multimodal Tutor can be most supportive. Psychomotor learning tasks can differ primarily by two factors: (1) by their *repetitiveness* and (2) by their *structuredness*. Learning how to perform chest compressions during CPR is a highly repetitive learning task, as the learner needs to perform repetitive movements; at the same time, CPR is highly structured, as there are very clear performance indicators that define the characteristics of a good CPR performance. These two characteristics make CPR an ideal application scenario for the Multimodal Tutor. On the contrary, the learning domain of calligraphy or foreign alphabet learning used in the Calligraphy Tutor consists of training repetitive tasks without such clear-cut performance indicators. The domain of public speaking of the Presentation Trainer consists of diverse and not repetitive movements which lack clear performance indicators for assessment.

At the **hardware-level**, sensors can influence importantly the quality of the collected data, the quality of the model training and thus of the feedback. In the CPR Tutor, as well as in related reference application scenarios, we opted for commercial sensor devices in place of custom made boards. When compared to custom-made boards, sensor devices such as Microsoft Kinect, Myo Armband or Fitbit HR have the advantage of being widely tested, of providing high-level drivers and having an API to easily connect and offer wide community support. Still, however, the commercial devices have known limitations in terms of precision. In this doctoral thesis, we realised that the choice of the sensor setup should be based on compromises between precision, easiness of use and relevance for the learning task investigated.

The third level concerns the limitations at the **software level**. The CPR Tutor and the LearningHub have been programmed using C# programming language that runs on Microsoft Windows 10 machines. The reason for such choice was to make the best use of Microsoft devices like Kinect. The VIT has been developed in Javascript and HTML 5, but tested primarily with Google Chrome browser. SharpFlow has been developed using Python 3.7. These choices could compromise the portability of the software components on different operating systems, browsers or platforms.

The fourth level of limitation is at the **data level**. As mentioned earlier, the precision and quality of the sensor devices can influence the quality of the data gathered. However, the data limitations lay also in the choice of the participant size and the diversity of these participants. Participants can have different body sizes, different ways to approach the task and different physiological responses. We call this the *inter-subject variability* among the participants. This variability can be mitigated by training a model with a diverse population, which can generalise their behavioural characteristics. There is, however, always the risk that the general model flushes out individual peculiarities. As an alternative, it is possible to train one classifier for each participant. The drawback of this approach is that the models will be suitable only for one person and not generalisable to new participants.

Finally, some limitations can stand at the **model level**. There are several limitations to using the supervised machine learning approach. Such an approach is optimal when having a high number of annotated training samples available. In the case of CPR, the more collected CCs, the more robust and general neural networks can be trained for mistake classification. Similarly, to set a clear division line between correct and incorrect learning performance, the learning task must have clear performance indicators. For example, in CPR the compression rate needs to have between 100 and 120 beats per minute for being optimal. The drawback of the supervised learning is well known by the machine learning research community and there are alternative ways that can be explored to reduce the amount of annotated samples needed, those are unsupervised learning, one-shot learning or transfer-learning techniques. Concerning the use of Recurrent Neural Networks, aside from the amount of training data, the other common limitation is the tendency of *overfitting* the training set. Besides dividing the collected data set between training, test, validation and performing cross-validation at the level of the training samples, it is important doing it at the subject level too. For example, it would be useful to *hold-one-participant-out*, to make sure that the data of one or more participants are completely new and unseen by the model.

Sailing into new horizons

In this doctoral thesis, the limitations can be seen as a research agenda for the future implementations of the Multimodal Tutor. Future research endeavours should go both in the theoretical and in the technical directions. From the theoretical standpoint, as evidenced in the literature survey in Chapter 2, future works of the Multimodal Tutor should also look into empirical studies and meta-analysis to focus on the most suitable data representation for each modality and propose guidelines for efficient modality combination. It could be useful knowing what is the best between modality and available sensors in commerce; providing guidelines for the data analysis of multimodal data sets.

Social Learning

Moreover, the Multimodal Tutor "of the future", the Multimodal Pipeline will improve and evolve as a concept to accommodate more reference application scenarios. For instance, one aspect deliberately left out both from the theoretical and from the application side, is the *social dimension of learning*: the extent by which the teacher and the learning peers influence each other in a social context. For example, during collaborative learning or physical classroom activities, social learning is of paramount importance. We think of the implementation of the Multimodal Tutor in the *Classroom of the Future*. Along the line of experimentation proposed by the *EduSense* prototype (Ahuja et al., 2019), the Classroom of The Future will embed a run-time framework which controls different sensors for example installed in laptops, chairs or desk and connects to various actuators such as the projector, the smart board, some lights. The purpose is to automatically orchestrate learning activities in the classroom. For this purpose, a renewed conceptualisation of the Multimodal Pipeline as a framework that runs continually on run-time is needed (Schneider et al., 2019). From such a system, not only learners could profit, but also teachers, for example, the system could identifying students at-risk. Along this line, the system

Lumilo provides an inspiring example of real-time teaching support using augmented reality, by identifying and signalling students at-risk to teachers with the help of "virtual hands" (Holstein et al., 2018).

In the Cloud

From the technical point of view, future implementation of the Multimodal Tutor can move away from collecting short and high-frequency data sessions towards longer data collection periods which can last days or weeks. In our vision, the Multimodal Tutor can become a *learning companion* that supports the learner throughout the entire duration of a course until the target skill is properly mastered. For this reason, we imagine future personalised learning technologies like the Multimodal Tutor can be *on-demand*, *wherever* and *whenever* the learner needs them. The functionalities of the Multimodal Tutor should be embedded in personal devices such as smartphones or smartwatches which can be at the learner's fingertips. To become fully ubiquitous, the Multimodal Tutor needs to better leverage cloud-based technologies. In that case, the learner would need only a device and an internet connection for using the functionalities of the Multimodal Tutor for learning support. Given the great amount and the data gathered from the sensors, sending the complete streams to the cloud might be an overhead for the network infrastructure. An option alternative to cloud computing that should be explored is *fog computing* (Bonomi et al., 2012), in which only relevant data or decisions are sent to the online server.

User Experience

Finally, future research of the Multimodal Tutor should look at how to improve the user experience from the learner perspective. As argued in this doctoral thesis, self-reports, questionnaires and user-ratings are important for collecting the learning labels necessary for annotating the multimodal experiences and for allowing the system to learn from historical data. Repeatedly asking the learner to answer a questionnaire or to submit a report, can become, nevertheless, a quite tiring task. For being able to mature the Multimodal Tutor from a research to a productivity tool, stratagems have to be thought to maximise its usability and user retention.

Ethics and Privacy

Connected to the user-experience, another paramount issue is to ensure user privacy when collecting high-frequency and highly personal multimodal data. Future Multimodal Tutor applications need to be designed with better *privacy* features. For instance, they need to implement multiple privacy layers, consisting of features such as end-to-end encryption, authentication or distributed data saving. The Multimodal Tutor should connect and use the concept of *Trusted Learning Analytics* (Drachsler and Greller, 2016). The learner has to become the ultimate authority over the data and the algorithms. The technology embedded in the Multimodal Tutor should ultimately support and improve learning rather than judge and punish the learner.

References

Ahuja, K., Agarwal, Y., Kim, D., Xhakaj, F., Varga, V., Xie, A., Zhang, S., Townsend, J. E., Harrison, C., and Ogan, A. (2019). EduSense: Practical Classroom Sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–26.

Alqahtani, F. and Ramzan, N. (2019). Comparison and efficacy of synergistic intelligent tutoring systems with human physiological response. *Sensors (Switzerland)*, 19(3).

Alyuz, N., Okur, E., Oktay, E., Genc, U., Aslan, S., Mete, S. E., Arnrich, B., and Esme, A. A. (2016). Semi-supervised model personalization for improved detection of learner's emotional engagement. *Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016*, pages 100–107.

Alzoubi, O., D'Mello, S. K., and Calvo, R. A. (2012). Detecting naturalistic expressions of nonbasic affect using physiological signals. *IEEE Transactions on Affective Computing*, 3(3):298–310.

Ambady, N. and Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2):256–274.

Amin, M. B., Banos, O., Khan, W. A., Bilal, H. S. M., Gong, J., Bui, D. M., Cho, S. H., Hussain, S., Ali, T., Akhtar, U., Chung, T. C., and Lee, S. (2016). On curating multimodal sensory data for health and wellness platforms. *Sensors (Switzerland)*, 16(7):1–27.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 16(07):1–2.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, 26(1):85–112.

Anderson, J. R., Franklin Boyle, C., and Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228(4698):456–462.

Andrade, A. and Danish, J. A. (2016). Using Multimodal Learning Analytics to Model Student Behaviour: A Systematic Analysis of Behavioural Framing. *Journal of Learning Analytics*, 3(2):282–306.

Arnold, K. E. (2010). Signals: Applying Academic Analytics. *EDUCAUSE Quarterly*, 33(1):87–92.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., and Christopherson, R. (2009). Emotion sensors go to school. *Frontiers in Artificial Intelligence and Applications*, 200(1):17–24.

Bahreini, K., Nadolski, R., and Westera, W. (2015). Improved multimodal emotion recognition for better game-based learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9221:107–120.

Bakharia, A., Heathcote, E., and Dawson, S. (2009). Social networks adapting pedagogical practice: SNAPP. *Proceedings ASCILITE 2009, Auckland*, pages 49–51.

Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2):423–443.

Barmaki, R. and Hughes, C. E. (2015). Providing real-time feedback for student teachers in a virtual rehearsal environment. *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, pages 531–537.

Baur, T., Damian, I., Lingenfelser, F., Wagner, J., and André, E. (2013). NovA: Automated analysis of nonverbal signals in social interactions. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8212 LNCS, pages 160–171. Springer, Cham, Cham, Switzerland.

Berg, A., Scheffel, M., Drachsler, H., Ternier, S., and Specht, M. (2016). Dutch cooking with xAPI recipes the good, the bad, and the consistent. In *Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016*, pages 234–236.

Bjork, R. A., Dunlosky, J., and Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, 64(1):417–444.

Black, P. and Dylan, W. (2009). Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, 21(1):5–31.

Blikstein, P. (2013). Multimodal learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, pages 102–106, New York, USA. ACM.

Blikstein, P. and Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: Using Computational Technologies to Measure Complex Learning Tasks. *Journal of Learning Analytics*, 3(2):220–238.

Boekaerts, M. (2010). The crucial role of motivation and emotion in classroom learning. In Dumont, H., Istance, D., and Benavides, F., editors, *The Nature of Learning: Using Research to Inspire Practice*, pages 91–112. OECD Publishing.

Bonomi, F., Milito, R., Zhu, J., and Addepalli, S. (2012). Fog computing and its role in the internet of things. In *MCC'12 - Proceedings of the 1st ACM Mobile Cloud Computing Workshop*, pages 13–15, New York, New York, USA. ACM Press.

Booth, M. (2012). Learning Analytics: The New Black. *Educause Review*, 47:52–53.

Börner, D. (2013). *Ambient Learning Displays*. PhD thesis, Open Universiteit.

Börner, D., Tabuenca, B., Storm, J., Happe, S., and Specht, M. (2015). Tangible Interactive Ambient Display Prototypes to Support Learning Scenarios. *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '14*, pages 721–726.

Bosch, N., Chen, H., Baker, R., Shute, V., and D'Mello, S. (2015). Multimodal Affect Detection in the Wild. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*, pages 645–649, New York, USA. ACM.

Boucsein, W. and Backs, R. W. (2000). Engineering psychophysiology: issues and applications. page 400.

Brown, M., Dehoney, J., and Millichap, N. (2015). The Next Generation Digital Learning Environment: A Report on Research. Technical report, Educase.

Brown, T. M. and Fee, E. (2002). Walter Bradford Cannon. *American Journal of Public Health*, 92(10):1594–1595.

Buckingham Shum, S. (2011). Learning Analytics: Notes on the Future The lost key. *Slideshare*.

Buckingham Shum, S. (2015). The Connected Intelligence Centre: Human-Centered Analytics for UTS.

Buckingham Shum, S. and Crick, R. D. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In *ACM International Conference Proceeding Series*.

Burleson, W. (2007). Affective learning companions. *Educational Technology*, 47(1):28.

Butler, D. L. and Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3):245–281.

Cacioppo, J., Tassinary, L. G., and Berntson, G. G. (2007). *The Handbook of Psychophysiology*, volume 44. Cambridge University Press.

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (2000). Handbook fo Psychophysiology. *book*, page 21.

Calvo, R., D'Mello, S., Gratch, J., and Kappas, A. (2015). The Oxford handbook of affective computing.

Campbell, J. P., DeBlois, P. B., and Oblinger, D. G. (2007). Academic Analytics: A New Tool for a New Era. *Educause Review*, 42(August 2007):40–57.

Canfield, W. (2001). ALEKS: a Web-based intelligent tutoring system. *Mathematics and Computer Education*, 35(2):152–158.

Clow, D. (2012). The learning analytics cycle. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, page 134.

Cope, B. and Kalantzis, M. (2015). Interpreting Evidence-of-Learning: Educational research in the era of big data. *Open Review of Educational Research*, 2(1):218–239.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. Basic Books.

Cukurova, M., Kent, C., and Luckin, R. (2019). Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, page bjet.12829.

Damasio, A. R., Tranel, D., and Damasio, H. C. (1991). Somatic markers and the guidance of behavior: Theory and preliminary testing. In *Frontal Lobe Function and Dysfunction*, pages 217–229. Oxford University Press.

De Lecea, L., Carter, M. E., and Adamantidis, A. (2012). Shining light on wakefulness and arousal.

De Raedt, L. (2008). *Logical and relational learning*, volume 5249 LNAI. Heidelberg, Springer-Verlag Berlin.

Dekker, G., Pechenizkiy, M., and Vleeshouwers, J. (2009). Predicting students drop out: A case study. *EDM'09 - Educational Data Mining 2009: 2nd International Conference on Educational Data Mining*, pages 41–50.

Di Mitri, D. (2018). Multimodal tutor for CPR. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10948 LNAI, pages 513–516, Cham, Switzerland. Springer International Publishing.

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., and Specht, M. (2016). Learning Pulse: using Wearable Biosensors and Learning Analytics to Investigate and Predict Learning Success in Self-regulated Learning. In *CEUR proceedings*, pages 34–39.

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., and Specht, M. (2017). Learning Pulse: a machine learning approach for predicting performance in self-regulated learning using multimodal data. In *LAK '17 Proceedings of the 7th International Conference on Learning Analytics and Knowledge*, pages 188–197, New York, NY, USA. ACM.

Di Mitri, D., Schneider, J., Klemke, R., Specht, M., and Drachsler, H. (2019a). Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19*, pages 51–60, New York, NY, USA. ACM.

Di Mitri, D., Schneider, J., Specht, M., and Drachsler, H. (2018a). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4):338–349.

Di Mitri, D., Schneider, J., Specht, M., and Drachsler, H. (2018b). The Big Five: Addressing Recurrent Multimodal Learning Data Challenges. In Martinez-Maldonado Roberto, editor, *Proceedings of the Second Multimodal Learning Analytics Across (Physical and Digital) Spaces (CrossMMLA)*, page 6, Aachen. CEUR Workshop Proceedings.

Di Mitri, D., Schneider, J., Specht, M., and Drachsler, H. (2019b). Detecting mistakes in CPR training with multimodal data and neural networks. *Sensors (Switzerland)*, 19(14):1–20.

Di Mitri, D., Schneider, J., Specht, M., and Drachsler, H. (2019c). Multimodal Pipeline: A generic approach for handling multimodal data for supporting learning. In *AIMA4EDU Workshop in IJCAI 2019 AI-based Multimodal Analytics for Understanding Human Learning in Real-world Educational Contexts*, pages 2–4.

Dicerbo, K. E. and Behrens, J. T. (2014). Impacts of the Digital Ocean on Education Impacts of the Digital Ocean on Education About the Authors. Technical Report February.

Dietz-Uhler, B. and Hurn, J. (2013). Using learning analytics to predict (and improve) student success: a faculty perspective. *Journal of Interactive Online Learning*, 12(1):17–26.

Dillenbourg, P. (1999). What do you mean by collaborative leraning? In *Collaborative learning: Cognitive and computational approaches*. Oxford: Elsevier.

Dillenbourg, P. (2016). The Evolution of Research on Digital Education. *International Journal of Artificial Intelligence in Education*, 26(2):544–560.

D'Mello, S. (2013). A Selective Meta-Analysis on the Relative Incidence of Discrete Affective States During Learning With Technology. *Journal of Educational Psychology*, 105(4):1082–1099.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El Kaliouby, R., Picard, R. W., and Graesser, A. (2008). AutoTutor detects and responds to learners affective and cognitive states. *IEEE Transactions on Education*, 48(4):612–618.

D'mello, S., Olney, A., Blanchard, N., Sun, X., Ward, B., Samei, B., and Kelly, S. (2015). Multimodal Capture of Teacher-Student Interactions for Automated Dialogic Analysis in Live Classrooms. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 557–566, New York, USA. ACM.

Domínguez, F., Echeverría, V., Chiluiza, K., and Ochoa, X. (2015). Multimodal Selfies : Designing a Multimodal Recording Device for Students in Traditional Classrooms.

Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pages 567–574.

Drachsler, H. and Greller, W. (2016). Privacy and Learning Analytics - it ’ s a DELICATE issue. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK’16)*, pages 89–98. ACM Press.

ECAR (2015). The Predictive Learning Analytics Revolution: Leveraging Learning Data for Student Success. *ECAR working group paper.*, pages 1–23.

Echeverría, V., Avendaño, A., Chiluiza, K., Vásquez, A., and Ochoa, X. (2014). Presentation Skills Estimation Based on Video and Kinect Data Analysis. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics*, pages 53–60.

Echeverría, V., Domínguez, F., and Chiluiza, K. (2016). Towards a distributed framework to analyze multimodal data. *CEUR Workshop Proceedings*, 1601:52–57.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., and Shum, S. B. (2018). Driving data storytelling from learning design. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK ’18*, 2018(March):131–140.

Edwards, A. A., Massicci, A., Sridharan, S., Geigel, J., Wang, L., Bailey, R., and Alm, C. O. (2017). Sensor-based Methodological Observations for Studying Online Learning. In *Proceedings of the 2017 ACM Workshop on Intelligent Interfaces for Ubiquitous and Smart Learning - SmartLearn ’17*, pages 25–30, New York, USA. ACM.

Eliot, C. and Woolf, B. P. (1996). An intelligent learning environment for advanced cardiac life support. *Proceedings of the AMIA Annual Fall Symposium*, pages 7–11.

Ericsson (2016). Ericsson Mobility Report.

Essa, A. and Ayad, H. (2012). Improving student success using predictive models and data visualisations. *Research in Learning Technology*, 5:58–70.

Europa, O. E. (2014). Learning Analytics and Assessment. In Duval, E. and Koskinen, T., editors, *eLearning Papers*, number 36, pages 1–48. P.A.U. Education, S.L.

Eveleigh, G. S., Ruiz, N., Liu, G., Yin, B., Farrow, D., and Chen, F. (2010). Teaching Athletes Cognitive Skills : Detecting Cognitive Load in Speech Input. *Training*, pages 2–5.

Ferguson, R. (2012). The state of learning analytics in 2012: a review and future challenges. *Technical Report KMI-12-01*, 4(March):18.

Ferguson, R. and Shum, S. B. (2012). Social learning analytics. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK ’12*, pages 23–33.

Fouse, A., Weibel, N., Hutchins, E., and Hollan, J. D. (2011). ChronoViz : A system for supporting navigation of time-coded data. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems - CHI EA '11*, page 299, New York, NY, USA. ACM.

Freedman, D. H. (2010). Why Scientific Studies Are So Often Wrong : The Streetlight Effect.

Freire, P. (1970). *Pedagogy of the Oppressed*. The Continuum International Publishing Group Inc.

Friedman, J. H. (1997). On Bias, Variance, 0/1‚ÄîLoss, and the Curse-of-Dimensionality. *Data Mining and Knowledge Discovery*, 1(1):55–77.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., and Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48(February):108–119.

Goetz, T. (2011). Harnessing the power of feedback loops.

González García, C., Meana Llorián, D., Pelayo G-Bustelo, C., and Cueva-Lovelle, J. M. (2017). A review about Smart Objects, Sensors, and Actuators. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(3):7.

Grafsgaard, J. F. (2014). Multimodal Analysis and Modeling of Nonverbal Behaviors during Tutoring. In *International Conference on Multimodal Interaction*, pages 404–408, New York, USA. ACM.

Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., and Lester, J. C. (2014). Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue. *Proceedings of the Seventh International Conference on Educational Data Mining*, 123(24):122–129.

Gravina, R., Alinia, P., Ghasemzadeh, H., and Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion*, 35:1339–1351.

Greller, W. and Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, 15(3):42–57.

Hattie, J. and Timperley, H. (2007). The power of feedback. [References]. *Review of Educational Research*, .77(1):16–7.

Heckmann, D. (2005). *Ubiquitous User Modeling*. PhD thesis, Technical University Eindhoven.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Holstein, K., McLaren, B. M., and Aleven, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10947 LNAI:154–168.

Hussain, M. S., Monkaresi, H., and Calvo, R. A. (2012). Categorical vs. dimensional representations in multimodal affect detection during learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7315 LNCS:78–83.

J. Baron, J. Whitmer, S. S. (2015). Predictive Learning Analytics: Fueling Actionable Intelligence.

Jayaprakash, S., Moody, E. W., Lauria, E. J. M., Regan, J. R., and Baron, J. D. (2014). Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative. *Journal of Learning Analytics*, 1(1):6–47.

Jewitt, C., Bezemer, J., and O'Halloran, K. (2016). *Introducing multimodality*. Routledge.

Jivet, I., Scheffel, M., Drachsler, H., and Specht, M. (2017). Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. In *EC-TEL: European Conference on Technology Enhanced Learning*. Springer.

Jones, C. (2004). Networks and learning: communities, practices and the metaphor of networks. *Alt-J*, 12(1):81–93.

Kamar, E. (2016). Directions in hybrid intelligence: Complementing AI systems with human intelligence. *IJCAI International Joint Conference on Artificial Intelligence*, 2016-Janua:4070–4073.

Kanda, T. and Ishiguro, H. (2017). *Human-Robot Interaction in Social Robotics*. CRC Press.

Kemper, T. D. and Lazarus, R. S. (1992). Emotion and Adaptation. *Contemporary Sociology*, 21(4):522.

Kim, J., Meltzer, C., Salehi, S., and Blikstein, P. (2011). Process pad: A multimedia multi-touch learning platform. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS'11*, pages 272–273.

Kitto, K., Cross, S., Waters, Z., and Lupton, M. (2015). Learning Analytics beyond the LMS: the Connected Learning Analytics Toolkit. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK '15). ACM, New York, NY, USA*, pages 11–15.

Kodaganallur, V. and Weitz, R. R. (2005). A comparison of model-tracing and constraint-based intelligent tutoring paradigms. *International Journal of Artificial Intelligence in Education*, 15(2):117–144.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. A. (1996). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education (IJAIED)*.

Koh, D. and Jeyaratnam, J. (1998). Biomarkers, screening and ethics. *Occupational Medicine*, 48(1):27–30.

Kothe, C., Grivich, M., Brunner, C., and Medine, D. (2018). Lab Streaming Layer.

Lahat, D., Adali, T., and Jutten, C. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE*, 103(9):1449–1477.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–74.

Larson, R. and Csikszentmihalyi, M. (1983). The Experience Sampling Method.

Leong, C. W., Chen, L., Feng, G., Lee, C. M., and Mulholland, M. (2015). Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits. In *ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction*, number 3, pages 547–556. ACM.

Li, I. (2015). Beyond Reflecting on Personal Data: Predictive Personal Informatics. In *Beyond Personal Informatics: Designing for Experiences with Data CHI 2015*, pages 1–5.

Limbu, B., Schneider, J., Klemke, R., and Specht, M. (2018a). Augmentation of practice with expert performance data: Presenting a calligraphy use case. In *3rd International Conference on Smart Learning Ecosystem and Regional Development - The interplay of data, technology, place and people*, pages 1–13.

Limbu, B. H., Jarodzka, H., Klemke, R., and Specht, M. (2018b). Using sensors and augmented reality to train apprentices using recorded expert performance: A systematic literature review. *Educational Research Review*, 25(June 2017):1–22.

Lindstrom, M. and Bates, D. (1988). Newton-Raphson and EM Algorithms for Linear Models for Repeated-Measures Data. *Journal of the American Statistical Association*, 83(404):1014–1022.

Lins, C., Eckhoff, D., Klausen, A., Hellmers, S., Hein, A., and Fudickar, S. (2019). Cardiopulmonary resuscitation quality parameters from motion capture data using Differential Evolution fitting of sinusoids. *Applied Soft Computing Journal*, 79:300–309.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg.

M. Bienkowski, F. Mingyu, B. M. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. *U.S. Department of Education Office of Educational Technology. Center for Technology in Learning SRI International*.

Martinez, R., Collins, A., Kay, J., and Yacef, K. (2011). Who did what? Who said that? In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces - ITS '11*, page 172, New York, USA. ACM Press.

Martinez-Maldonado, R. (2016). Seeing learning analytics tools as orchestration technologies: Towards supporting learning activities across physical and digital spaces. *CEUR Workshop Proceedings*, 1601:70–73.

Martinez-Maldonado, R., Echeverria, V., Santos, O. C., Santos, A. D. P. D., and Yacef, K. (2018). Physical learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, number May, pages 375–379, New York, NY, USA. ACM.

Martinez-Maldonado, R., Suthers, D., Aljohani, N. R., Hernandez-Leo, D., Kitto, K., Pardo, A., Charleer, S., and Ogata, H. (2016). Cross-LAK: Learning analytics across physical and digital spaces. In *ACM International Conference Proceeding Series*, pages 486–487.

Mayor, O., Llimona, Q., Marchini, M., Papiotis, P., and Maestre, E. (2013). repoVizz: A Framework for Remote Storage, Browsing, Annotation, and Exchange of Multimodal D ata. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 415–416, New York, USA. ACM Press.

Mehrabian, A. (1971). *Silent messages*. Wadsworth Publishing Company.

Milgram, P., Takemura, H., Utsumi, a., and Kishino, F. (1994). Mixed Reality (MR) Reality-Virtuality (RV) Continuum. *Systems Research*, 2351(Telemanipulator and Telepresence Technologies):282–292.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59(4):439–483.

Mitrovic, A. and Hausler, K. (2003). An Intelligent SQL Tutor on the Web. *International Journal of Artificial Intelligence in Education (IOS Press)*, 13:173–197.

Mohamed Chatti.,, A., Anna Dyckhoff.,, L., Ulrik, S., and Hendrik, T. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6):1–22.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2012). *Foundations of machine learning*. MIT Press.

Mory, E. H. (2004). Feedback research revisited. *Handbook of research on educational communications and technology*, 45:745–784.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2):133–142.

Nicol, D. J. and Macfarlane-Dick, D. (2006). Formative assessment and self- regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2):199–218.

Nigay, L. and Coutaz, J. (1993). A design space for multimodal systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*, number January 1993, pages 172–178, New York, New York, USA. ACM Press.

Norris, S. (2004). *Analyzing Multimodal Interaction*. Number 1. Routledge.

North Dakota University (2014). Predictive Analytics Reporting: Using technology to help students succeed.

Novak, T. P., Hoffman, D. L., and Yung, Y. F. (1998). Measuring the flow construct in online environments: a structural modeling approach. *Unpublished manuscript*, pages 1–48.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., and Castells, J. (2013). Expertise estimation based on simple multimodal features. In *Proceedings of the 15th ACM International Conference on Multimodal Iteraction (ICMI '13)*, pages 583–590, New York, USA. ACM Press.

Ochoa, X. and Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, 3(2):213–219.

Oller, R. (2012). The Future of Mobile Learning (Research Bulletin). *EDUCAUSE Center for Applied Research,*.

Open University, U. (2015). Policy on Ethical use of Student Data for Learning Analytics. Technical Report September 2014, Open University UK.

Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, 5(1):12–30.

Oviatt, S. (2013). Problem solving, domain expertise and learning: Ground-truth performance results for math data corpus. *2013 15th ACM International Conference on Multimodal Interaction, ICMI 2013*, pages 569–574.

Oviatt, S., Cohen, A., Weibel, N., Hang, K., and Thompson, K. (2013). Multimodal Learning Analytics Data Resources : Description of Math Data Corpus and Coded Documents. Technical report, University of Sydney.

Oviatt, S., Schuller, B., Cohen, P. R., Sonntag, D., Potamianos, G., and Krüger, A. (2018). *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 2*. [s.n.].

Paas, F., Tuovinen, J. E., Tabbers, H., and Gerven, P. W. M. V. (2010). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1):63–71.

Pan, S. J. and Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.

Pardo, A. and Kloos, C. D. (2011). Stepping out of the box: towards analytics outside the learning management system. *1st International Conference on Learning Analytics and Knowledge (LAK11)*, pages 163–167.

Perkins, G. D., Handley, A. J., Koster, R. W., Castrén, M., Smyth, M. A., Olasveengen, T., Monsieurs, K. G., Raffay, V., Gräsner, J.-T. T., Wenzel, V., Ristagno, G., Soar, J., Bossaert, L. L., Caballero, A., Cassan, P., Granja, C., Sandroni, C., Zideman, D. A., Nolan, J. P., Maconochie, I., and Greif, R. (2015). European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. *Resuscitation*, 95:81–99.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press,, second edi edition.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., and Kirschner, P. A. (2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, pages 64–73.

Pijeira-Díaz, H. J., Drachsler, H., Kirschner, P. A., and Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning*, 34(4):397–408.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31(6):459–470.

Pintrich Zusho, A., P. R. (2007). Student motivation and self-regulated learning in the college classroom. In *The scholarship of teaching and learning in higher education: An evidence-based perspective*, pages 731–810. Springer.

Poggi, I., D'Errico, F., and Vinciarelli, A. (2012). Social signals: From theory to applications.

Polakow, V. (1944). The Politics of Education: Culture Power and Liberation. *Phenomenology + Pedagogy*, pages 86–89.

Polson, M. C., Richardson, J. J., and Soloway, E. (1988). *Foundations of intelligent tutoring systems*. Erlbaum Associates Inc., Hillsdale, NJ, USA.

Praharaj, S., Scheffel, M., Drachsler, H., and Specht, M. (2018). Multimodal Analytics for Real-time Feedback in Co-located Collaboration. *Ec-Tel*, 1:187–201.

Prieto, L., Sharma, K., Kidzinski, Ł., Rodríguez-Triana, M., and Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of Computer Assisted Learning*.

Prieto, L. P., Sharma, K., Dillenbourg, P., and Jesús, M. (2016). Teaching analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, pages 148–157, New York, USA. ACM.

Raca, M. and Dillenbourg, P. (2014). Holistic Analysis of the Classroom. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge - MLA '14*, 740(Cv):13–20.

Roll, I. and Winne, P. H. (2015). Understanding , evaluating , and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2:7–12.

Rosmalen, P. V. (2014). Instructional Designs for Real- time Feedback. Technical Report October, Open Universiteit, Heerlen, The Netherlands.

Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*.

Salehi, S., Kim, J., Meltzer, C., and Blikstein, P. (2012). Process pad: A low-cost multi-touch platform to facilitate multimodal documentation of complex learning. *Proceedings of the 6th International Conference on Tangible, Embedded and Embodied Interaction, TEI 2012*, 1(212):257–262.

Santos, O. C. (2016). Training the Body: The Potential of AIED to Support Personalized Motor Skills Learning. *International Journal of Artificial Intelligence in Education*, 26(2):730–755.

Santos, O. C. (2019). Artificial Intelligence in Psychomotor Learning: Modeling Human Motion from Inertial Sensor Data. *International Journal on Artificial Intelligence Tools*, 28(04):1940006.

Scheffel, M., Drachsler, H., Stoyanov, S., and Specht, M. (2014). Quality Indicators for Learning Analytics. *Journal of Educational Technology & Society*, 17(4):117–132.

Schmitz, M., van Limbeek, E., Greller, W., Sloep, P., and Drachsler, H. (2017). Opportunities and challenges in using learning analytics in learning design. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.

Schneider, B. and Blikstein, P. (2015). Unraveling Students' Interaction Around a Tangible Interface using Multimodal Learning Analytics. *JEDM - Journal of Educational Data Mining*, 7(3):89–116.

Schneider, J., Börner, D., van Rosmalen, P., and Specht, M. (2015a). Augmenting the Senses: A Review on Sensor-Based Learning Support. *Sensors*, 15(2):4097–4133.

Schneider, J., Börner, D., van Rosmalen, P., and Specht, M. (2015b). Presentation Trainer, your Public Speaking Multimodal Coach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*, pages 539–546, New York, USA. ACM.

Schneider, J., Di Mitri, D., Limbu, B., and Drachsler, H. (2018). Multimodal Learning Hub: A Tool for Capturing Customizable Multimodal Learning Experiences. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11082 LNCS, pages 45–58, Cham, Switzerland. Springer.

Schneider, J., Mitri, D. D., Drachsler, H., and Specht, M. (2019). Multimodal Learning Analytics Runtime Framework. In *Proceedings of the Third Multimodal Learning Analytics Across (Physical and Digital) Spaces (CrossMMLA).*, pages 1–6.

Schon, D. (1983). The Reflective Practitioner. *New York: Basic*.

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, 6(3):11–19.

Semeraro, F., Frisoli, A., Loconsole, C., Bannò, F., Tammaro, G., Imbriaco, G., Marchetti, L., and Cerchiari, E. L. (2012). A new Kinect-based system for the analysis of performances in cardiopulmonary resuscitation (CPR) training. *Resuscitation*, 83:e20.

Shankar, S. K., Prieto, L. P., Rodriguez-Triana, M. J., and Ruiz-Calleja, A. (2018). A review of multimodal learning analytics architectures.

Shapiro, L. (2019). *Embodied Cognition*. Routledge, Routledge, 2 edition.

Sharples, M., Arnedillo-Sánchez, I., Milrad, M., and Vavoula, G. (2009). Mobile learning: Small devices, big issues. In *Technology-Enhanced Learning: Principles and Products*, pages 233–249. Springer Netherlands.

Shaun, R., Baker, J. D., and Inventado, P. S. (2014). Chapter 4: Educational Data Mining and Learning Analytics. *Springer*, Chapter 4:61–75.

Siemens, G. (2012). Learning Analytics: Envisioning a Research Discipline and a Domain of Practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, number May, page 4, New York, New York, USA. ACM Press.

Siemens, G. and Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46:30–32.

Singley, M. and Lam, R. (2005). The classroom sentinel: supporting data-driven decision-making in the classroom. *Proceedings of the 14th international conference on World Wide Web*, pages 315–321.

Slade, S. and Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist*, 57(10):1510–1529.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., and Holden, H. K. (2012). Adaptive Tutoring & the Generalized Intelligent Framework for Tutoring.

Specht, M. (2015). Connecting Learning Contexts with Ambient Information Channels. In *Seamless Learning in the Age of Mobile Connectivity*, pages 121–140. Springer Singapore, Singapore.

Spikol, D., Ruffaldi, E., Dabisias, G., and Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, 34(4):366–377.

Steenbergen-Hu, S. and Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2):331–347.

Steve Leibson (2008). IPV6: How Many IP Addresses Can Dance on the Head of a Pin?

Suebnukarn, S. and Haddawy, P. (2007). COMET : A Collaborative for Medical Problem-Based Learning. *IEEE Intelligent Systems*, 22(4):70–77.

Suthers, D. and Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, pages 64–74. ACM.

Swan, M. (2012). Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks*, 1(3):217–253.

Tabuenca, B., Börner, D., Kalz, M., and Specht, M. (2015a). User-Modelled Ambient Feedback for Self-regulated Learning. In Conole, G., Klobučar, T., Rensing, C., Konert, J., and Lavoué, E., editors, *10th European Conference on Technology Enhanced Learning, EC-TEL 2015*, pages 535–539, Toledo. Springer.

Tabuenca, B., Kalz, M., Drachsler, H., and Specht, M. (2015b). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, 89:53–74.

Tabuenca, B., Kalz, M., and Specht, M. (2014). NFC LearnTracker : Seamless support for learning with mobile and sensor technology. *In Proceedings of the 4th European Immersive Education Summit. Journal of Immersive Education (JiED)*.

Tabuenca, B., Kalz, M., and Specht, M. (2015c). Tap it again, Sam: Harmonizing Personal Environments towards Lifelong Learning. *International Journal of Advanced Corporate Learning*, 8(1):16–23.

Taelman, J., Vandeput, S., Spaepen, a., and Huffel, S. V. (2009). Influence of Mental Stress on Heart Rate and Heart Rate Variability. *Ecifmbe 2008*, 29(1):1366–1369.

Taylor, R. H., Menciassi, A., Fichtinger, G., Fiorini, P., and Dario, P. (2016). Medical Robotics and Computer-Integrated Surgery. In *Springer Handbook of Robotics*, pages 1657–1684. Springer International Publishing, Cham.

Tempelaar, D. T., Rienties, B., and Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47:157–167.

Thayer, J. F., Hansen, A. L., Saus-Rose, E., and Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, 37(2):141–153.

Van Merrienboer, J. J., Clark, R. E., and De Croock, M. B. (2002). Blueprints for complex learning: The 4C/ID-model.

Van Merriënboer, J. J. G. and Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17(2):147–177.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4):197–221.

Vatrapu, R., Teplovs, C., Fujita, N., and Bull, S. (2011). Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, pages 93–98.

Vinciarelli, A., Pantic, M., Bourlard, H., and Pentland, A. (2008). Social signal processing: state-of-the-art and future perspectives of an emerging domain. *Proceedings of the 16th ACM international conference on Multimedia*, 27(November 2008):1061–1070.

Vygotsky, L. (1978). Interaction between learning and development.

Wachsmuth, I., Lenzen, M., and Knoblich, G. (2012). *Embodied Communication in Humans and Machines*. Oxford University Press, Oxford, UK.

Wagner, E. and Davis, B. (2014). The Predictive Analytics Reporting (PAR) Framework , WCET. *Educase Review Online*, pages 1–8.

Wagner, J., Lingenfelser, F., and André, E. (2011). The social signal interpretation framework (SSI) for real time signal processing and recognition. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 3245–3248, Makuhari, Chiba, JP. International Speech Communication Association.

Wang, H. M. and Huang, S. C. (2012). SDNN/RMSSD as a surrogate for LF/HF: A revised investigation. *Modelling and Simulation in Engineering*.

Wang, J.-C., Liao, W.-I., and Tsai, S.-H. (2018a). Potential pros and cons of the kinect-based real-time audiovisual feedback device during in-hospital cardiopulmonary resuscitation. *American Journal of Emergency Medicine*, 36(2):319–320.

Wang, J.-C., Tsai, S.-H., Chen, Y. L., Chu, S.-J., and Liao, W.-I. (2018b). Kinect-based real-time audiovisual feedback device improves CPR quality of lower-body-weight rescuers. *American Journal of Emergency Medicine*, 36(4):577–582.

Wattanasoontorn, V., Magdics, M., Boada, I., and Sbert, M. (2013). A kinect-based system for cardiopulmonary resuscitation simulation: A pilot study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8101 LNCS:51–63.

West, D. M. (2012). Big Data for Education: Data Mining, Data Analytics, and Web Dashboards.

Winne, P. H. (2017). Learning Analytics for Self-Regulated Learning. *Handbook of Learning Analytics*, pages 241–249.

Winne, P. H. and Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In Hacker, D. J., Dunlosky, J., and Graesser, A. C., editors, *Metacognition in Educational Theory and Practice*, pages 277–304. Lawrence Erlbaum Associates Publishers.

Wong, L.-H. (2012). A learner-centric view of mobile seamless learning. *British Journal of Educational Technology*, 43(1):E19–E23.

Wong-Villacres, M., Granda, R., Ortiz, M., and Chiluiza, K. (2016). Exploring the impact of a tabletop-generated group work feedback on students' collaborative skills. *CEUR Workshop Proceedings*, 1601:58–64.

Worsley, M. (2014). Multimodal Learning Analytics as a Tool for Bridging Learning Theory and Complex Learning Behaviors. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge - MLA '14*, pages 1–4, New York, USA. ACM.

Worsley, M. (2018). Multimodal learning analytics' past, present, and, potential futures. In *CEUR Workshop Proceedings*, volume 2163, pages 1–16, Aachen, Germany. CEUR Workshop Proceedings.

Worsley, M. and Blikstein, P. (2013). Towards the Development of Multimodal Action Based Assessment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, pages 94–101.

Worsley, M. and Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, pages 360–367.

Worsley, M. and Blikstein, P. (2018). A Multimodal Analysis of Making. *International Journal of Artificial Intelligence in Education*, 28(3):385–419.

Yazdani, M. (1986). Intelligent tutoring systems: An overview. *Expert Systems*, 3(3):154–163.

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, 41(August):64–70.

Zimmermann, A., Specht, M., and Lorenz, A. (2005). Personalization and context management. *User Modelling and User-Adapted Interaction*, 15(3-4):275–302.

List of Tables

List of Figures

Summary

This doctoral thesis describes the journey of ideation, prototyping and empirical testing of the *Multimodal Tutor*, a system designed for providing digital feedback that supports psychomotor skills acquisition using learning and multimodal data capturing. The feedback is given in real-time with the machine-driven assessment of the learner's task execution. The predictions are tailored by supervised machine learning models trained with human-annotated samples. The doctoral thesis is organised into four parts and seven chapters.

Chapter 1 describes the exploratory experiment *Learning Pulse*, a study which has aimed at predicting levels of stress, productivity and level of flow during self-regulated learning. We have gathered multimodal data from nine participants, PhD students of the Open Universiteit's TELI department. The data consisted of (1) physiological data (heart-rate and step count) from Fitbit HR wristbands; (2) software applications used on their laptops from RescueTime; and (3) environmental information (temperature, humidity, pressure and geolocation coordinates) using web APIs. During the two weeks, the participants had to self-report every working hour via a mobile application, the *Activity Rating Tool*. The data have been collected in a *Learning Record Store* using custom *Experience API* (xAPI) triplets. Throughout the two weeks of data collection, the nine participants have used different laptops and sets of software applications, which have been thus grouped into categories to ease analysis. As the collected data were heterogeneous, we opted for the *Linear Mixed Effect Model* (LMEM), a multi-level prediction algorithm. The poor results in the model accuracy did not allow to explore further the feedback mechanisms. This outcome led us to reconsider several methodological decisions, setting the basis of the following experiments.

The literature study in Chapter 2 has aimed at mapping the state of the art of Multimodal Data for learning, a field which was emerging as *Multimodal Learning Analytics* (MMLA). Surveying the related literature has shown that MMLA covered a scattered scientific field and not yet a coherent one. This work has contributed to framing the mission of MMLA: *using multimodal data and data-driven techniques for filling the gap between observable learning behaviour and learning theories*. We have conducted a literature survey (Section 2.2) of MMLA studies using the proposed *classification framework* in which we have separated two main components: the *input space* and the *hypothesis space* that are separated by the *observability line*. The results

of the literature survey have led to the *Taxonomy of multimodal data for learning*, to the *Classification table for the hypothesis space* and to a conceptual model for supporting the emerging field of MMLA the ***Multimodal Learning Analytics Model*** (MLeAM). MLeAM has three main objectives: (1) mapping the use of multimodal data to enhance the feedback in a learning context; (2) showing how to combine machine learning with multimodal data; (3) aligning the terminology used in the field of machine learning and learning science.

In Chapter 3, we have addressed another the structural shortcomings in the MMLA field: the lack of standardised technical approaches for multimodal data support of learning activities. This aspect is holding back the development of the MMLA field by imposing the MMLA researchers to duplicate efforts in setting up data collection infrastructures and by preventing them to focus on data analysis. In Chapter 3, the identified technical challenges have been grouped into five categories, named the *'Big Five' challenges of Multimodal Learning Analytics* which are the (1) *data collection*, (2) *data storing*, (3) *data annotation*, (4) *data processing* and (5) *data exploitation*.

In Chapter 4 we have focused on one of the five big challenges, the *data annotation*. This challenge deals with *how humans can make sense of complex multidimensional data*. In this chapter, we have proposed a new technical prototype, the *Visual Inspection Tool* (VIT). The VIT allows the researchers to visually inspect and annotate a variety of psychomotor learning tasks that can be captured with a customisable set of sensors. The VIT enables the researcher (1) to triangulate multimodal data with video recordings; (2) to segment the multimodal data into time intervals and to add annotations to the time intervals; (3) to download the annotated dataset and use the annotations as labels for machine learning predictions. Beside generically addressing the data annotation, the VIT also facilitates data processing and exploitation. The VIT is released as Open Source software[5].

In Chapter 5, we have coined the term *Multimodal Pipeline*, a chain of technical reusable components which includes the VIT, the Multimodal Learning Hub and its custom data format MLT-JSON. The Multimodal Pipeline is an integrated technical workflow that works as a toolkit for supporting MMLA researchers to set up new experiments in a variety of psychomotor learning scenarios. We argue that using components from this toolkit can reduce developing time to set up experiments and it can facilitate and speed up the transfer of research knowledge in the MMLA community. The Multimodal Pipeline connects a set of technical solutions to the “Big Five” challenges described presented in Chapter 5. The Multimodal Pipeline has two main stages: the ‘offline training’, in which the collected sessions are annotated and the ML models are trained with the collected data; the ‘online exploitation’, which corresponds to the ‘run-time’ behaviour of the Multimodal Pipeline.

In Chapter 6 we have selected Cardiopulmonary Resuscitation (CPR) as an application case for the Multimodal Tutor, a representative learning task for carrying out a study on mistake detection. CPR was chosen mainly because: it is an individual learn-

[5] Code available on GitHub (https://github.com/dimstudio/visual-inspection-tool)

ing task, it is repetitive and highly structured, it has clear performance indicators and because it is a training with high social relevance. We introduced a new approach for detecting CPR training mistakes with multimodal data using neural networks. The proposed system is composed in a multi-sensor setup for CPR, consisting of a Kinect camera and a Myo armband. We have used the system in combination with the ResusciAnne manikin for collecting data from 11 experts performing CPR training. We have first validated then collected multimodal data upon three performance indicators provided by the ResusciAnne manikin, observing that we can classify accurately the training mistakes on these three standardised indicators. We further have concluded that it is possible to extend the standardised mistake detection to additional training mistakes on performance indicators such as correct locking of the arms and correct body position. So far, those mistakes could only be detected by human instructors.

In Chapter 7, we have presented the design and the development of real-time feedback architecture for the CPR Tutor. To complete the chain of flexible technical solutions proposed by the Multimodal Pipeline, we have developed *SharpFlow*[6], an open-source data processing tool. SharpFlow supports the MLT-JSON format used as well by the VIT and the LearningHub. The data serialised in this format are transformed by SharpFlow into a tensor representation and fed into a Recurrent Neural Network architecture which is trained to classify the different target classes specified by the human annotations. SharpFlow implements the two data-flows of *offline training* and *online exploitation*. The latter is achieved using a TCP server for classifying in real-time every new chest compression. In Chapter 7, the architecture has been first employed in an Expert Study involving 10 participants, aimed at training the mistake classification models, and second in a User Study involving 10 additional participants in which the CPR Tutor was prompting real-time feedback interventions. The analysis of the detected mistakes in the User Study has proved that there is a short-term positive effect on the CPR mistakes targeted by the automatic feedback.

The discussion of continues by presenting the findings of each study of this doctoral thesis, along with the contributions and the limitations of this research. With its underpinning technological frameworks (Multimodal Pipeline) and conceptual model (Multimodal Learning Analytics Model), the Multimodal Tutor pushes forward the entire MMLA field, by explaining how to support learning with the use of multimodal data. The Multimodal Tutor generates scientific added value for different data-driven learning research communities including the Learning Analytics & Knowledge and the Intelligent Tutoring System/Artificial Intelligence in Education. Ultimately, the Multimodal Tutor sets the way for more emerging fields of research such as *Hybrid Intelligence*, *Social Artificial Intelligence* or *Human-AI Teaming*, that focus on how to best interface human communication with artificial (robotic) intelligence.

[6] Code available on GitHub (https://github.com/dimstudio/SharpFlow)

Samenvatting

Deze dissertatie beschrijft de reis van ideevorming, prototyping en de empirische testen van de *Multimodal Tutor*, eeen systeem dat is ontworpen om het verwerven van psychomotorische vaardigheden te ondersteunen door middel van het vastleggen van, en feedback geven over, multimodale gegevens tijdens het leren. De feedback wordt in real-time gegeven met een machine gestuurde beoordeling van de uitvoering van de taak van de leerling. De voorspellingen worden op maat gemaakt met behulp van bewaakte modellen voor machine learning die worden getraind met door mensen geannoteerde voorbeelden. De doctoraalscriptie is opgezet in vier delen en zeven hoofdstukken.

Hoofdstuk 1 een studie die gericht is op het voorspellen van het stress niveau, de productiviteit en het niveau van doorstroming tijdens het zelfregulerend leren. We hebben multimodale gegevens verzameld van negen deelnemers, promovendi van de afdeling TELI van de Open Universiteit. De data bestond uit (1) fysiologische gegevens (hartslag en aantal stappen) via Fitbit HR- polsbandjes; (2) softwaretoepassingen van RescueTime op hun laptops; en (3) omgevingsinformatie (temperatuur, vochtigheid, luchtdruk en geolocatiecoördinaten) via web-API‚Äôs. Gedurende twee weken moesten de deelnemers elk werkuur zelf rapporteren via een mobiele applicatie, de *Activity Rating Tool*. De gegevens zijn verzameld in een *Learning Record Store* met behulp van aangepaste *Experience API* (xAPI) statements. Gedurende de twee weken van dataverzameling hebben de negen deelnemers gebruik gemaakt van verschillende laptops en sets van softwareapplicaties, die gegroepeerd zijn in categorieën om de analyse te vergemakkelijken. Omdat de verzamelde gegevens heterogeen waren, hebben we gekozen voor het *Linear Mixed Effect Model* (LMEM), een voorspellingsalgoritme met meerdere niveaus. De slechte resultaten in het model lieten niet toe om de terugkoppelingsmechanismen verder te onderzoeken. Dit resultaat leidde tot het heroverwegen van een aantal methodologische beslissingen, die de basis vormden voor de volgende experimenten.

De literatuurstudie in hoofdstuk 2 heeft zich gericht op het in kaart brengen van de stand van zaken van de techniek van Multimodale Data voor het leren, een veld nu beter bekend als *Multimodal Learning Analytics* (MMLA). Uit onderzoek van gerelat eerde literatuur is gebleken dat MMLA een versnipperd en nog niet samenhangend wetenschappelijk gebied bestrijkt. Dit werk heeft bijgedragen aan het formuleren van de missie van MMLA: het *gebruik van multimodale data en data* *gestuurde technieken om de kloof tussen waarneembaar leergedrag en leertheorieën te dichten*. We hebben een literatuuronderzoek (Section 2.2) van MMLA-studies uitgevoerd aan de hand van het voorgestelde *classificatieraamwerk* waarin we twee hoofdcomponenten hebben gescheiden: de *input space* en de *hypothesis space* die door de *observability line* worden gescheiden. De resultaten van de literatuurstudie hebben geleid tot de *Taxonomy of multimodal data for learning*, tot de *Classification table for the hypothesis space* en tot een conceptueel model ter ondersteuning van het opkomende veld van MMLA het ***Multimodal Learning Analytics Model*** (MLeAM). MLeAM heeft drie hoofddoelstellingen: 1) het in kaart brengen van het gebruik van multimodale data om de feedback in een leercontext te verbeteren; (2) laten zien hoe machine learning kan worden gecombineerd met multimodale gegevens; en (3) de terminologie die wordt gebruikt op het gebied van machine learning en de leerwetenschap op elkaar afstemmen.

In hoofdstuk 3, hebben we nog een andere structurele tekortkoming op het gebied van MMLA aangepakt: het gebrek aan gestandaardiseerde technische benaderingen voor multimodale gegevensondersteuning van leeractiviteiten. Dit aspect belemmert de ontwikkeling van het MMLA-veld. Het zet de MMLA-onderzoekers ertoe aan om dubbel werk te doen bij het opzetten van infrastructuren voor gegevensverzameling en het voorkomt dat ze zich concentreren op gegevensanalyse. In hoofdstuk 3, tzijn de geïdentificeerde technische uitdagingen gegroepeerd in vijf categorieën, genaamd de *'Big Five' challenges of Multimodal Learning Analytics* die zijn: (1) *dataverzameling*, (2) *(2) gegevensopslag*, (3) *gegevensannotatie*, (4) *gegevensverwerking* en (5) *gegevensexploitatie*.

In hoofdstuk 4 whebben we ons gericht op een van de vijf grote uitdagingen, de *data annotatie*. Deze uitdaging gaat over *hoe mensen complexe multidimensionale gegevens kunnen begrijpen*. In dit hoofdstuk hebben we een nieuw technisch prototype voorgesteld, de *Visual Inspection Tool* (VIT). De VIT stelt de onderzoekers in staat om verschillende psychomotorische leertaken die kunnen worden vastgelegd met een aanpasbare set van sensoren, visueel te inspecteren en te annoteren.. De VIT stelt de onderzoeker in staat om (1) multimodale data te trianguleren met videoopnames; (2) de multimodale data te segmenteren in tijdsintervallen en annotaties toe te voegen aan de tijdsintervallen; en (3) de geannoteerde dataset te downloaden en de annotaties te gebruiken als labels voor machine learning voorspelling. Naast de algemene aanpak van de data annotatie vergemakkelijkt de VIT ook de verwerking en exploitatie van de gegevens. De VIT wordt uitgebracht als Open Source software[7].

In hoofdstuk 5, hebben we de term *Multimodal Pipeline*, een keten van technische herbruikbare componenten die de VIT, de Multimodal Learnig Hub en het aangepaste gegevensformaat MLT-JSON omvat, uitgewerkt. De Multimodal Pipeline is een geïntegreerde technische workflow die werkt als een toolkit om MMLA- onderzoekers te ondersteunen bij het opzetten van nieuwe experimenten in diverse psychomotorische leerscenario's. Wij stellen dat het gebruik van componenten uit deze toolkit de ontwikkeltijd voor het opzetten van experimenten kan verkorten en

[7] Code beschikbaar op GitHub (https://github.com/dimstudio/visual-inspection-tool)

de overdracht van onderzoekskennis in de MMLA-community kan vergemakkelijken en versnellen. De Multimodal Pipeline verbindt een reeks technische oplossingen met de "Big Five" uitdagingen die in hoofdstuk 5 worden beschreven. De Multimodal Pipeline kent twee hoofdfasen: de 'offline training', waarbij de verzamelde sessies worden geannoteerd en de ML-modellen worden getraind met de verzamelde gegevens; en (2) de 'online exploitatie', die overeenkomt met het 'run-time' gedrag van de Multimodal Pipeline.

In hoofdstuk 6 hebben we Cardiopulmonale Reanimatie (CPR) geselecteerd als een toepassing voor de Multimodale Tutor, omdat dit een representatieve leeropdracht voor het uitvoeren van een onderzoek naar foutdetectie is. Reanimatie is vooral gekozen omdat: het een individuele leeropdracht is, het is repetitief en zeer gestructureerd is, het heeft duidelijke prestatie-indicatoren heeft en het een training is met een hoge sociale relevantie. We hebben een nieuwe aanpak geïntroduceerd voor het opsporen van reanimatiefouten met behulp van multimodale gegevens met behulp van neurale netwerken. Het voorgestelde systeem is samengesteld in een multi-sensor opstelling voor reanimatie, bestaande uit een Kinectcamera en een Myo armband. We hebben het systeem gebruikt in combinatie met de ResusciAnne-testpop voor het verzamelen van gegevens van 11 experts die reanimatietraining geven. We hebben eerst de multimodale gegevens gevalideerd, en vervolgens de gegevens verzameld op basis van drie prestatie-indicatoren die door de ResusciAnne-testpop zijn verstrekt, waarbij we hebben opgemerkt dat we de trainingsfouten op deze drie gestandaardiseerde indicatoren nauwkeurig kunnen classificeren. We hebben verder geconcludeerd dat het mogelijk is om de gestandaardiseerde foutdetectie uit te breiden naar extra trainingsfouten op prestatie-indicatoren zoals correcte vergrendeling van de armen en correcte lichaamshouding. Tot nu toe konden deze fouten alleen door menselijke instructeurs worden opgespoord.

In hoofdstuk 7, hebben we het ontwerp en de ontwikkeling van een real-time feedbackarchitectuur voor de reanimatietutor gepresenteerd. Om de keten van flexibele technische oplossingen die de Multimodal Pipeline voorstelt te vervolledigen, hebben we *SharpFlow*[8], aontwikkeld, een open source dataverwerkingstool. SharpFlow ondersteunt het MLT-JSON-formaat dat ook door de VIT en de LearningHub wordt gebruikt. De in dit formaat opgeslagen gegevens worden door SharpFlow getransformeerd in een tensorrepresentatie en worden gevoed met een Recurrent Neural Network architectuur die is getraind in het classificeren van de verschillende doelklassen die door de menselijke annotaties worden gespecificeerd. SharpFlow implementeert de twee datastromen van *offline training* en *online exploitatie*. De laatste wordt bereikt met behulp van een TCP-server voor het classificeren in real-time van elke nieuwe borstcompressie. In hoofdstuk 7, is de architectuur eerst toegepast in een Expert Study met 10 deelnemers, gericht op het trainen van de foutclassificatiemodellen, en vervolgens in een User Study met 10 extra deelnemers, waarbij de reanimatietrainer realtime feedbackinterventies uitvoerde. De analyse van de gedetecteerde fouten in de Gebruikersstudie heeft aangetoond dat er op

[8] Code beschikbaar op GitHub (https://github.com/dimstudio/SharpFlow)

korte termijn een positief effect is op de reanimatiefouten waarop de automatische feedback is gericht.

De discussie hierover wordt voortgezet met een presentatie van de bevindingen van elke studie van dit proefschrift, samen met de bijdragen en de beperkingen van dit onderzoek. De Multimodale Tutor duwt met zijn ondersteunende technologische kaders (Multimodal Pipeline) en conceptueel model (Multimodal Learning Analytics Model) het hele MMLA-veld naar voren, door uit te leggen hoe het leren met behulp van multimodale data kan worden ondersteund. De Multimodale Tutor genereert wetenschappelijke meerwaarde voor verschillende data gestuurde leer-onderzoeks community, waaronder de Learning Analytics & Knowledge en het Intelligent Tutoring System/Artificial Intelligence in Education community. Uiteindelijk zet de Multimodale Tutor de weg vrij voor meer opkomende onderzoeksgebieden zoals *Hybride Intelligentie*, *Social Artificial Intelligence* of *Human-AI Teaming*, die zich richten op hoe de menselijke communicatie het beste kan worden gekoppeld aan kunstmatige (robotachtige) intelligentie.

Riassunto

Questa tesi di dottorato descrive il percorso di ideazione, prototipazione e sperimentazione empirica del *Multimodal Tutor*, un sistema progettato per fornire feedback digitale e supportare l'acquisizione di competenze psicomotorie tramite l'analisi di dati multimodali. Il feedback viene fornito dal sistema in tempo reale mentre l'allievo esegue un compito psicomotorio. Il sistema effettua una valutazione automatica dell'allievo tramite la rilevazione e la classificazione di errori procedurali per via da modelli di machine learning supervisionati. Tali modelli sono addestrati con campioni annotati manulamente. La tesi di dottorato è organizzata in quattro parti e sette capitoli.

Il capitolo 1, riporta uno studio che ha lo scopo di predire i livelli di stress, la produttività e il livello di Flow durante l'apprendimento auto-regolato. Abbiamo raccolto dati multimodali da nove partecipanti, dottorandi del dipartimento TELI della Open University of The Netherlands. I dati multimodali raccolti consistevano in (1) dati fisiologici (frequenza cardiaca e conteggio dei passi) reccolti attraverso i braccialetti Fitbit HR; (2) dati software, registrati con l'applicazione RescueTime utilizzata ed installata sui computer di ogni partecipante; (3) dati relativi ad informazioni ambientali (temperatura, umidità, pressione e coordinate di geolocalizzazione) utilizzando varie web-API. Al contempo, i partecipanti dovevano riportatare i valori percepiti di produttività stress e Flow al termine di ogni singola ora di lavoro utilizzando un'applicazione mobile *Activity Rating Tool*. I dati sono stati raccolti in un *Learning Record Store* utilizzando statements *Experience API* (xAPI) personalizzate. Durante le due settimane di raccolta dati, i nove partecipanti hanno utilizzato diversi computer portatili e set di applicazioni software, che sono stati così raggruppati in categorie per facilitarne l'analisi. Poichè i dati raccolti erano eterogenei, nell'analizzarli abbiamo optato per il *Linear Mixed Effect Model* (LMEM), un algoritmo di previsione multilivello. Gli scarsi risultati nell'accuratezza del modello non hanno permesso di esplorare ulteriormente i meccanismi di feedback. Questo risultato ci ha portato a riconsiderare diverse decisioni metodologiche, ponendo le basi per i seccessivi esperimenti, dettagliati nei capitoli successivi di questa tesi.

Il capitolo 2 riguarda lo studio della letteratura mirato a identificare e mappare lo stato dell'arte dei dati multimodali per l'apprendimento, un campo che sta emergendo con il nome di *Multimodal Learning Analytics* (MMLA). Questo studio ha dimostrato che MMLA copre un campo scientifico disperso e non ancora coerente.

Sottolineare questa mancanza ha contribuito a inquadrare la missione di MMLA: *utilizzare dati multimodali e tecniche guidate dai dati per colmare il divario tra il comportamento di apprendimento osservabile e le teorie dell'apprendimento*. L'indagine della letteratura (Section 2.2) é stato condotta utilizzando un *framework di classificazione* composto da due componenti principali: lo *spazio degli input* e lo *spazio delle iposesi*. Queste sono separate dalla *linea di osservabilità*. I risultati dell'indagine di letteratura hanno portato alla *Tassonomia dei dati multimodali per l'apprendimento*, alla *tabella di classificazione per lo spazio delle ipotesi* e ad un modello concettuale a supporto del campo emergente di MMLA, il *Multimodal Learning Analytics Model* (MLeAM). MLeAM ha tre obiettivi principali: (1) mappare l'uso dei dati multimodali per migliorare il feedback, dato agli allievi, in un contesto di apprendimento; (2) mostrare come combinare l'apprendimento automatico con dati multimodali; (3) allineare la terminologia utilizzata nel campo dell'apprendimento automatico e della scienza dell'apprendimento.

Nel capitolo 3, abbiamo affrontato un'altra delle carenze strutturali nel campo della MMLA: la mancanza di approcci tecnici standardizzati per il supporto di dati multimodali per le attività di apprendimento. Questo aspetto sta frenando lo sviluppo del settore MMLA imponendo ai ricercatori MMLA di duplicare gli sforzi nella creazione di infrastrutture di raccolta dati e impedendo loro di concentrarsi sull'analisi dei dati. Nel capitolo 3, le sfide tecniche identificate sono state raggruppate in cinque categorie, denominate le *cinque grandi sfide delle Multimodal Learning Analytics* ovvero: la (1) *raccolta di dati*, (2) la *memorizzazione dei dati*, (3) l'*annotazione dei dati*, (4) l'*elaborazione dei dati* e (5) lo *sfruttamento dei dati*.

Nel capitolo 4, ci siamo concentrati su una delle cinque grandi sfide, l'*annotazione dei dati*. Questa sfida riguarda il *modo in cui gli esseri umani possono dare un senso a dati multidimensionali complessi*. In questo capitolo abbiamo proposto un nuovo prototipo tecnico, il *Visual Inspection Tool* (VIT). Il VIT permette ai ricercatori di ispezionare visivamente e annotare una varietà di compiti di apprendimento psicomotorio che possono essere catturati con un set di sensori personalizzabili. Il VIT consente al ricercatore (1) di triangolare i dati multimodali con registrazioni video; (2) di segmentare i dati multimodali in intervalli di tempo e di aggiungere annotazioni; (3) di scaricare il set di dati annotati e di utilizzare le annotazioni come 'labels' per i modelli di machine learning. Oltre ad affrontare genericamente l'annotazione dei dati, il VIT facilita anche l'elaborazione e lo sfruttamento dei dati. Il VIT è rilasciato come software Open Source[9].

Nel capitolo 5, abbiamo coniato il termine *Multimodal Pipeline*, una catena di componenti tecnici riutilizzabili che comprende il VIT, il Multimodal Learnig Hub e il suo formato dati personalizzato MLT-JSON. La Multimodal Pipeline è un framework tecnico integrato che funziona come un toolkit per supportare i ricercatori MMLA nell'impostare nuovi esperimenti in una varietà di scenari di apprendimento psicomotorio. In questo capitolo sosteniamo che l'utilizzo dei componenti di questo toolkit può rendere piú efficace il lavoro del ricercatore, riducendo i tempi di sviluppo per

[9] Codice disponibile su GitHub (https://github.com/dimstudio/visual-inspection-tool)

impostare gli esperimenti e può facilitare e accelerare il trasferimento delle conoscenze della ricerca nella comunità MMLA. La Multimodal Pipeline collega una serie di soluzioni tecniche alle "cinque grandi sfide" descritte nel capitolo 5. La Multimodal Pipeline ha due fasi principali: (1) il "training offline", in cui le sessioni raccolte vengono annotate e i modelli di appredndimento vengono addestrati con i dati raccolti; (2) lo "sfruttamento online", che corrisponde al comportamento "run-time" della Multimodal Pipeline.

Nel capitolo 6, abbiamo selezionato la Rianimazione Cardiopolmonare (RCP) come task rappresentativo per la realizzazione di uno studio sul rilevamento degli errori di esecuzione/performance dell'allievo. La RCP è stata scelta principalmente perché: è un compito di apprendimento individuale, è ripetitivo e altamente strutturato, ha chiari indicatori di performance e perché è una formazione ad alta rilevanza sociale. In questo studio abbiamo introdotto un nuovo approccio per il rilevamento degli errori di esecuzione nell'apprendimentio di RCP con dati multimodali che utilizzano le reti neurali. Il sistema proposto è composto da un setup multisensore per la RCP, costituito da una camera Kinect e una fascia da braccio Myo. Abbiamo utilizzato il sistema in combinazione con il manichino ResusciAnne per raccogliere i dati di 11 esperti che hanno eseguito la RCP. I dati dati multimodali sono stati prima convalidati e poi raccolti su tre indicatori di performance forniti dal manichino ResusciAnne. Abbiamo osservato che é possibile classificare accuratamente gli errori di apprendimento su questi tre indicatori di performance. Questo ci ha permesso di concludere che è possibile estendere la rilevazione ad ulteriori errori di apprendimento RCP, quiali: il corretto bloccaggio delle braccia e la corretta posizione del corpo. Errori, che fino ad ora, potevano essere rilevati solo da istruttori in carne ed ossa.

Nel capitolo 7, abbiamo presentato la progettazione e lo sviluppo dell'architettura di feedback in tempo reale per il tutor della RCP. Per completare la catena di soluzioni tecniche flessibili proposte dalla Multimodal Pipeline, abbiamo sviluppato *SharpFlow*[10], uno strumento di elaborazione dati open source. SharpFlow supporta il formato MLT-JSON utilizzato anche da VIT e dal LearningHub. I dati serializzati in questo formato vengono trasformati da SharpFlow in un 'tensor' e inseriti in un'architettura di Rete Neurale Ricorrente che è addestrata a classificare le diverse classi di destinazione specificate dalle annotazioni. SharpFlow implementa i due flussi di dati della *training offline* e dello *sfruttamento online*. Quest'ultimo è ottenuto utilizzando un server TCP per classificare in tempo reale ogni nuova compressione toracica. Nel capitolo 7, l'architettura è stata impiegata per la prima volta in uno studio con esperti che ha coinvolto 10 partecipanti, finalizzato al training dei modelli di classificazione degli errori, e per la seconda volta in uno studio con utenti che ha coinvolto altri 10 partecipanti. Il Tutor RCP ha fornito interventi di feedback in tempo reale. L'analisi degli errori rilevati nello Studio Utente ha dimostrato che vi è un effetto positivo a breve termine sugli errori di RCP oggetto del feedback automatico.

[10] Codice disponibile su GitHub (https://github.com/dimstudio/SharpFlow)

La tesi di dottorato si conclude con la discussione che presenta i risultati di ogni studio insieme con i contributi e i limiti di questa ricerca. Con il suo framework tecnologico di base (Multimodal Pipeline) e il modello concettuale (Multimodal Learning Analytics Model), il Multimodal Tutor permette l'avanzamento dell'intero campo MMLA, spiegando come supportare l'apprendimento con l'uso di dati multimodali. Il Multimodal Tutor genera valore scientifico aggiunto per diverse comunità di ricerca sull'apprendimento, tra cui la comunità di Learning Analytics & Knowledge e la comunità di l'Intelligent Tutoring System/Artificial Intelligence in Education. In definitiva, il Tutor Multimodale spiana la strada a campi di ricerca più emergenti come l'*Hybrid Intelligence*, l' *Intelligenza Artificiale Sociale* o *Human-AI Teaming*, che si concentrano su come interfacciare al meglio la comunicazione umana con l'intelligenza artificiale.

Acknowledgements

This PhD thesis would have not been possible without the outstanding support of the many people that accompanied and supported me throughout this long journey.

The first ‘thank you’ goes to Hendrik, who was my first contact at the OU and the reason why I joined the TELI group as a master intern in fall 2015. Hendrik believed in me from day one. He provided me with all the freedom to explore, make mistakes and succeed. He pushed me to follow my research intuitions and he motivated me even when I was losing confidence in my ideas.

The second ‘thank you’ goes to Marcus my other promoter. Marcus was for me an inspiring leader and visionary researcher. Marcus trusted and supported my ideas providing me with opportunities to expand. Marcus was the reason why I stayed with the TELI group to conduct my PhD research. He set the fruitful grounds on which I could conduct my research, with a flexible, open and collaborative culture which remains my gold standard for academia.

The other person who mostly supported me in this journey is Jan. More than just a supervisor, Jan is for me a mentor and a friend. Armed with TEL research experience, kindness and lots of patience. Jan helped me to improve throughout my entire PhD journey. He was my reference point, always reachable and helpful when I needed help. Jan has a critical mind and doesn’t like conventionality. With his mindset helped me question things, always stay relevant and original.

The other ‘thank you’ goes to Maren and Roland, who were involved in several aspects of my PhD trajectory offering me support on several levels. Maren was my ‘guide’ in the ‘academic jungle’, having always ready for me a how-to’s for all types of procedures such as conference preparation, paper submissions, ethical approvals and helping me in countless moments in which I was stuck. Roland was the ‘mediator’ and ‘networker’ whom as topic leader in the Multimodal Learning Experience (MLX) group, provided me with project opportunities, feedback, suggestions and lots of laughs and enjoyable moments.

Other special mention goes to Alessandra and Johan, who were my ‘guardian angels’. In several occasions, they offered me practical advice, psychological support inside and outside the office. They were always there watching my back just like a family.

I would like to thank the other TELI PhDs Bibeg, Ioana, Angel, Martine, Sambit, Liqin; the colleagues Stefaan, Olga, Ellen, Howard, Slavi, Jeroen, etc.; former colleagues who left as Marco, Dirk or Esther. Thank you to the yOUng network, especially to the other board members Laurie, Manon, Mari and Katya. ‘Thank you’ also to

Marina, Mieke and the rest of the OU staff/secretariat and colleagues which with their support made my journey possible and pleasant.

A special thanks also to the more recent OU colleagues from CAROU where I recently moved to continue my postdoctoral research: Martine H., Stefano, Wiebke, Alex, Deniz, Elianne, Gerard, Petru etc. This group has a team spirit and welcomed me and my ideas on the team. I am very glad I can follow up my research with them.

I need to thank the EATEL research community and its JTEL Summer School, in which constituted the biggest part of my scientific network the last four years. Becoming the local chair of the Summer School in 2019 and hosting the Summer School in my home town was lots of fun. Special thank you the people involved Mikhail, Christian, Maria, Alejandro, and all the JTELers. The other research community which 'nurtured' me was the Learning Analytics & Knowledge (LAK). Recurrent appointment each year, attending the LAK conference helped me grow. Special thanks to the workshop communities in particular, the Alan, Gabor and the Learning Analytics Hackathon organisers as well as Roberto, Daniel and the CrossMMLA community.

Outside of the research field, I also need to thank the people that most supported me during this process. In the first place my girlfriend Sophie, who backed me up in all the ups and downs I faced in the long process. Sophie was my emotional support and my work-life balance, remembering me to take breaks, visit friends and enjoy life aside from just working. She prevented I ended up in isolation, she has cheered with me my accomplishments and turn my worries into productive mindset.

A big thank you goes to my mother and my two sisters Paola and Silvia, my beautiful family who made me always believed that I could reach this big achievement.

Lastly, this thesis is dedicated to my dear father, who first introduced me to computer science. In his spare time, my father liked fishing with his small boat, taking me along from time to time. He once explained to me the benefit of using a 'reversed anchor'. He told me that while an anchor is necessary to keep the boat stable against the water streams, the risk is that the anchor's arms get stuck in the seabed's rocks. He then taught me to first tie the rope to the bottom of the anchor, then to connect the rope to the top extremity using a thinner nylon string. In this way, when getting stuck, a strong pull would break the nylon string making possible to raise the anchor by dragging it from the bottom and allowing to successfully continue the navigation.

Funding sources

The research projects presented in this dissertation were mainly funded by (1) SafePAT, Interreg Euregion Meuse-Rhine project SafePAT (http://www.safepat.eu/); (2) WEKIT Project funded by Horizon 2020 research and innovation programme under the grant agreement number: 687669 (https://www.wekit.eu).

SIKS Dissertation Series

The complete list of dissertations carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems from 1998 on can be found at `http://www.siks.nl/dissertations.php`.

2011

2011-01 Botond Cseke (RUN)
Variational Algorithms for Bayesian Inference in Latent Gaussian Models.

2011-02 Nick Tinnemeier(UU)
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language.

2011-03 Jan Martijn van der Werf (TUE)
Compositional Design and Verification of Component-Based Information Systems.

2011-04 Hado Philip van Hasselt (UU)
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms.

2011-05 Bas van de Raadt (VU)
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

2011-06 Yiwen Wang(TUE)
Semantically-Enhanced Recommendations in Cultural Heritage.

2011-07 Yujia Cao (UT
Multimodal Information Presentation for High Load Human Computer Interaction.

2011-08 Nieske Vergunst (UU)
BDI-based Generation of Robust Task-Oriented Dialogues.

2011-09 Tim de Jong (OU)
Contextualised Mobile Media for Learning.

2011-10 Bart Bogaert (UVT)
Cloud Content Contention.

2011-11 Dhaval Vyas (UT)
Designing for Awareness: An Experience-focused HCI Perspective.

2011-12 Carmen Bratosin (TUE)
Grid Architecture for Distributed Process Mining.

2011-13 Xiaoyu Mao (UVT)
Airport under Control; Multiagent Scheduling for Airport Ground Handling.

2011-14 Milan Lovric(EUR)
Behavioral Finance and Agent-Based Artificial Markets.

2011-15 Marijn Koolen (UVA)
The Meaning of Structure: the Value of Link Evidence for Information Retrieval.

2011-16 Maarten Schadd (UM)
Selective Search in Games of Different Complexity.

2011-17 Jiyin He (UVA)
Exploring Topic Structure: Coherence, Diversity and Relatedness.

2011-18 Mark Ponsen (UM)
Strategic Decision-Making in complex games.

2011-19 Ellen Rusman (OU)
The Mind ' s Eye on Personal Profiles.

2011-20 Qing Gu (VU)
Guiding service-oriented software engineering - A view-based approach.

2011-21 Linda Terlouw (TUD)
Modularization and Specification of Service-Oriented Systems.

2011-22 Junte Zhang (UVA)
System Evaluation of Archival Description and Access.

2011-23 Wouter Weerkamp (UVA)
Finding People and their Utterances in Social Media.

2011-24 Herwin van Welbergen (UT)
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior.

2011-25 Syed Waqar ul Qounain Jaffry (VU)
Analysis and Validation of Models for Trust Dynamics.

2011-26 Matthijs Aart Pontier (VU)
Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots.

2011-27 Aniel Bhulai (VU)
Dynamic website optimization through autonomous management of design patterns.

2011-28 Rianne Kaptein (UVA)
Effective Focused Retrieval by Exploiting Query Context and Document Structure.

2011-29 Faisal Kamiran (TUE)
Discrimination-aware Classification.

2011-30 Egon van den Broek (UT)
Affective Signal Processing (ASP): Unraveling the mystery of emotions.

2011-31 Ludo Waltman (EUR)
Computational and Game-Theoretic Approaches for Modeling Bounded Rationality.

2011-32 Nees-Jan van Eck (EUR)
Methodological Advances in Bibliometric Mapping of Science.

2011-33 Tom van der Weide (UU)
Arguing to Motivate Decisions.

2011-34 Paolo Turrini (UU)
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations.

2011-35 Maaike Harbers (UU)
Explaining Agent Behavior in Virtual Training.

2011-36 Erik van der Spek (UU)
Experiments in serious game design: a cognitive approach.

2011-37 Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference.

2011-38 Nyree Lemmens (UM)
Bee-inspired Distributed Optimization.

2011-39 Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games.

2011-40 Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development.

2011-41 Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control.

2011-42 Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution.

2011-43 Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge.

2011-44 Boris Reuderink (UT)
Robust Brain-Computer Interfaces.

2011-45 Herman Stehouwer (UVT)
Statistical Language Models for Alternative Sequence Selection.

2011-46 Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work.

2011-47 Azizi Bin Ab Aziz(VU)
Exploring Computational Models for Intelligent Support of Persons with Depression.

2011-48 Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent.

2011-49 Andreea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality.

2012

2012-01 Terry Kakeeto (UVT)
Relationship Marketing for SMEs in Uganda.

2012-02 Muhammad Umair(VU)
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models.

2012-03 Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories.

2012-04 Jurriaan Souer (UU)
Development of Content Management System-based Web Applications.

2012-05 Marijn Plomp (UU)
Maturing Interorganisational Information Systems.

2012-06 Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks.

2012-07 Rianne van Lambalgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions.

2012-08 Gerben de Vries (UVA)
Kernel Methods for Vessel Trajectories.

2012-09 Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms.

2012-10 David Smits (TUE)
Towards a Generic Distributed Adaptive Hypermedia Environment.

2012-11 J.C.B. Rantham Prabhakara (TUE)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics.

2012-12 Kees van der Sluijs (TUE)
Model Driven Design and Data Integration in Semantic Web Information Systems.

2012-13 Suleman Shahid (UVT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions.

2012-14 Evgeny Knutov(TUE)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems.

2012-15 Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes..

2012-16 Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment.

2012-17 Amal Elgammal (UVT)
Towards a Comprehensive Framework for Business Process Compliance.

2012-18 Eltjo Poort (VU)
Improving Solution Architecting Practices.

2012-19 Helen Schonenberg (TUE)
What's Next? Operational Support for Business Process Execution.

2012-20 Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing.

2012-21 Roberto Cornacchia (TUD)
Querying Sparse Matrices for Information Retrieval.

2012-22 Thijs Vis (UVT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?.

2012-23 Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction.

2012-24 Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval.

2012-25 Silja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application.

2012-26 Emile de Maat (UVA)
Making Sense of Legal Text.

2012-27 Hayrettin Gurkok (UT
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games.

2012-28 Nancy Pascall (UVT)
Engendering Technology Empowering Women.

2012-29 Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval.

2012-30 Alina Pommeranz (TUD)
Designing Human-Centered Systems for Reflective Decision Making.

2012-31 Emily Bagarukayo (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure.

2012-32 Wietske Visser (TUD)
Qualitative multi-criteria preference representation and reasoning.

2012-33 Rory Sie (OUN)
Coalitions in Cooperation Networks (COCOON).

2012-34 Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications.

2012-35 Evert Haasdijk (VU)
Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics.

2012-36 Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes.

2012-37 Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation.

2012-38 Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms.

2012-39 Hassan Fatemi (UT)
Risk-aware design of value and coordination networks.

2012-40 Agus Gunawan (UVT)
Information Access for SMEs in Indonesia.

2012-41 Sebastian Kelle (OU)
Game Design Patterns for Learning.

2012-42 Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning.

2012-43 Withdrawn
.

2012-44 Anna Tordai (VU)
On Combining Alignment Techniques.

2012-45 Benedikt Kratz (UVT)
A Model and Language for Business-aware Transactions.

2012-46 Simon Carter (UVA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation.

2012-47 Manos Tsagkias (UVA)
Mining Social Media: Tracking Content and Predicting Behavior.

2012-48 Jorn Bakker (TUE)
Handling Abrupt Changes in Evolving Time-series Data.

2012-49 Michael Kaisers (UM)
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions.

2012-50 Steven van Kervel (TUD)
Ontology driven Enterprise Information Systems Engineering.

2012-51 Jeroen de Jong (TUD)
Heuristics in Dynamic Scheduling; a practical framework with a case study in elevator dispatching.

2013

2013-01 Viorel Milea (EUR)
News Analytics for Financial Decision Support.

2013-02 Erietta Liarou (CWI)
MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing.

2013-03 Szymon Klarman (VU)
Reasoning with Contexts in Description Logics.

2013-04 Chetan Yadati(TUD)
Coordinating autonomous planning and scheduling.

2013-05 Dulce Pumareja (UT)
Groupware Requirements Evolutions Patterns.

2013-06 Romulo Goncalves(CWI)
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience.

2013-07 Giel van Lankveld (UVT)
Quantifying Individual Player Differences.

2013-08 Robbert-Jan Merk(VU)
Making enemies: cognitive modeling for opponent agents in fighter pilot simulators.

2013-09 Fabio Gori (RUN)
Metagenomic Data Analysis: Computational Methods and Applications.

2013-10 Jeewanie Jayasinghe Arachchige(UVT)
A Unified Modeling Framework for Service Design..

2013-11 Evangelos Pournaras(TUD)
Multi-level Reconfigurable Self-organization in Overlay Services.

2013-12 Maryam Razavian(VU)
Knowledge-driven Migration to Services.

2013-13 Mohammad Safiri(UT)
Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly.

2013-14 Jafar Tanha (UVA)
Ensemble Approaches to Semi-Supervised Learning Learning.

2013-15 Daniel Hennes (UM)
Multiagent Learning - Dynamic Games and Applications.

2013-16 Eric Kok (UU)
Exploring the practical benefits of argumentation in multi-agent deliberation.

2013-17 Koen Kok (VU)
The PowerMatcher: Smart Coordination for the Smart Electricity Grid.

2013-18 Jeroen Janssens (UVT)
"*Outlier Selection and One-Class Classification*".

2013-19 Renze Steenhuizen (TUD)
Coordinated Multi-Agent Planning and Scheduling.

2013-20 Katja Hofmann (UVA)
Fast and Reliable Online Learning to Rank for Information Retrieval.

2013-21 Sander Wubben (UVT)
Text-to-text generation by monolingual machine translation.

2013-22 Tom Claassen (RUN)
Causal Discovery and Logic.

2013-23 Patricio de Alencar Silva(UVT)
Value Activity Monitoring.

2013-24 Haitham Bou Ammar (UM)
Automated Transfer in Reinforcement Learning.

2013-25 Agnieszka Anna Latoszek-Berendsen (UM)
Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System.

2013-26 Alireza Zarghami (UT)
Architectural Support for Dynamic Homecare Service Provisioning.

2013-27 Mohammad Huq (UT)
Inference-based Framework Managing Data Provenance.

2013-28 Frans van der Sluis (UT)
When Complexity becomes Interesting: An Inquiry into the Information eXperience.

2013-29 Iwan de Kok (UT)
Listening Heads.

2013-30 Joyce Nakatumba (TUE)
Resource-Aware Business Process Management: Analysis and Support.

2013-31 Dinh Khoa Nguyen (UVT)
Blueprint Model and Language for Engineering Cloud Applications.

2013-32 Kamakshi Rajagopal (OUN)
Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development.

2013-33 Qi Gao (TUD)
User Modeling and Personalization in the Microblogging Sphere.

2013-34 Kien Tjin-Kam-Jet (UT)
Distributed Deep Web Search.

2013-35 Abdallah El Ali (UVA)
Minimal Mobile Human Computer Interaction.

2013-36 Than Lam Hoang (TUe)
Pattern Mining in Data Streams.

2013-37 Dirk Borner (OUN)
Ambient Learning Displays.

2013-38 Eelco den Heijer (VU)
Autonomous Evolutionary Art.

2013-39 Joop de Jong (TUD)
A Method for Enterprise Ontology based Design of Enterprise Information Systems.

2013-40 Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games.

2013-41 Jochem Liem (UVA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning.

2013-42 Leon Planken (TUD)
Algorithms for Simple Temporal Reasoning.

2013-43 Marc Bron (UVA)
Exploration and Contextualization through Interaction and Concepts.

2014

2014-01 Nicola Barile (UU)
Studies in Learning Monotone Models from Data.

2014-02 Fiona Tuliyano (RUN)
Combining System Dynamics with a Domain Modeling Method.

2014-03 Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions.

2014-04 Hanna Jochmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation.

2014-05 Jurriaan van Reijsen (UU)
Knowledge Perspectives on Advancing Dynamic Capability.

2014-06 Damian Tamburri (VU)
Supporting Networked Software Development.

2014-07 Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior.

2014-08 Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints.

2014-09 Philip Jackson (UVT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language.

2014-10 Ivan Salvador Razo Zapata (VU)
Service Value Networks.

2014-11 Janneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support.

2014-12 Willem van Willigen (VU)
Look Ma, No Hands: Aspects of Autonomous Vehicle Control.

2014-13 Arlette van Wissen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains.

2014-14 Yangyang Shi (TUD)
Language Models With Meta-information.

2014-15 Natalya Mogles (VU)
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare.

2014-16 Krystyna Milian (VU)
Supporting trial recruitment and design by automatically interpreting eligibility criteria.

2014-17 Kathrin Dentler (VU)
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability.

2014-18 Mattijs Ghijsen (UVA)
Methods and Models for the Design and Study of Dynamic Agent Organizations.

2014-19 Vinicius Ramos (TUE)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support.

2014-20 Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link.

2014-21 Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments.

2014-22 Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training.

2014-23 Eleftherios Sidirourgos (UVA/CWI)
Space Efficient Indexes for the Big Data Era.

2014-24 Davide Ceolin (VU)
Trusting Semi-structured Web Data.

2014-25 Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction.

2014-26 Tim Baarslag (TUD)
What to Bid and When to Stop.

2014-27 Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty.

2014-28 Anna Chmielowiec (VU)
Decentralized k-Clique Matching.

2014-29 Jaap Kabbedijk (UU)
Variability in Multi-Tenant Enterprise Software.

2014-30 Peter de Cock (UVT)
Anticipating Criminal Behaviour.

2014-31 Leo van Moergestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support.

2014-32 Naser Ayat (UVA)
On Entity Resolution in Probabilistic Data.

2014-33 Tesfa Tegegne (RUN)
Service Discovery in eHealth.

2014-34 Christina Manteli(VU)
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems..

2014-35 Joost van Oijen (UU)
Cognitive Agents in Virtual Worlds: A Middleware Design Approach.

2014-36 Joos Buijs (TUE)
Flexible Evolutionary Algorithms for Mining Structured Process Models..

2014-37 Maral Dadvar (UT)
Experts and Machines United Against Cyberbullying.

2014-38 Danny Plass-Oude Bos (UT)
Making brain-computer interfaces better: improving usability through post-processing..

2014-39 Jasmina Maric (UVT)
Web Communities, Immigration, and Social Capital.

2014-40 Walter Omona (RUN)
A Framework for Knowledge Management Using ICT in Higher Education..

2014-41 Frederic Hogenboom (EUR)
Automated Detection of Financial Events in News Text.

2014-42 Carsten Eijckhof (CWI/TUD)
Contextual Multidimensional Relevance Models.

2014-43 Kevin Vlaanderen (UU)
Supporting Process Improvement using Method Increments.

2014-44 Paulien Meesters (UVT)
Intelligent Blauw: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden..

2014-45 Birgit Schmitz (OUN)
Mobile Games for Learning: A Pattern-Based Approach.

2014-46 Ke Tao (TUD)
Social Web Data Analytics: Relevance, Redundancy, Diversity.

2014-47 Shangsong Liang (UVA)
Fusion and Diversification in Information Retrieval.

2015

2015-01 Niels Netten (UVA)
Machine Learning for Relevance of Information in Crisis Response.

2015-02 Faiza Bukhsh (UVT)
Smart auditing: Innovative Compliance Checking in Customs Controls.

2015-03 Twan van Laarhoven (RUN)
Machine learning for network data.

2015-04 Howard Spoelstra (OUN)
Collaborations in Open Learning Environments.

2015-05 Christoph Bosch (UT)
Cryptographically Enforced Search Pattern Hiding.

2015-06 Farideh Heidari (TUD)

Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes.

2015-07 Maria-Hendrike Peetz (UVA)
Time-Aware Online Reputation Analysis.

2015-08 Jie Jiang (TUD)
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions.

2015-09 Randy Klaassen (UT)
HCI Perspectives on Behavior Change Support Systems.

2015-10 Henry Hermans (OUN)
OpenU: design of an integrated system to support lifelong learning.

2015-11 Yongming Luo (TUE)
Designing algorithms for big graph datasets: A study of computing bisimulation and joins.

2015-12 Julie M. Birkholz (VU)
Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks Promotor: Prof. dr. P. Groenewegen (VU), Prof. dr. J.H. Akkermans (VU).

2015-13 Giuseppe Procaccianti (VU)
Energy-Efficient Software.

2015-14 Bart van Straalen (UT)
A cognitive approach to modeling bad news conversations.

2015-15 Klaas Andries de Graaf (VU)
Ontology-based Software Architecture Documentation.

2015-16 Changyun Wei (UT)
Cognitive Coordination for Cooperative Multi-Robot Teamwork.

2015-17 Andre van Cleeff (UT)
Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs.

2015-18 Holger Pirk (CWI)
Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories.

2015-19 Bernardo Tabuenca (OUN)
Ubiquitous Technology for Lifelong Learners.

2015-20 Lois Vanhee (UU)
Using Culture and Values to Support Flexible Coordination Using Culture and Values to Support Flexible Coordination.

2015-21 Sibren Fetter (OUN)
Using Culture and Values to Support Flexible CoordinationUsing Peer-Support to Expand and Stabilize Online Learning.

2015-22 Zhemin Zhu (UT)
Co-occurrence Rate Networks - Towards separate training for undirected graphical models.

2015-23 Luit Gazendam (VU)
Using Culture and Values to Support Flexible CoordinationCataloguer Support in Cultural Heritage.

2015-24 Richard Berendsen (UVA)
Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation.

2015-25 Steven Woudenberg (UU)
Bayesian Tools for Early Disease Detection.

2015-26 Alexander Hogenboom (EUR)
Sentiment Analysis of Text Guided by Semantics and Structure.

2015-27 Sandor Heman (CWI)
Updating compressed colomn stores.

2015-28 Janet Bagorogoza (TiU)
Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO.

2015-29 Hendrik Baier (UM)
Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains.

2015-30 Kiavash Bahreini (OU)
Real-time Multimodal Emotion Recognition in E-Learning.

2015-31 Yakup Koc (TUD)
On the robustness of Power Grids.

2015-32 Jerome Gard (UL)
Corporate Venture Management in SMEs.

2015-33 Frederik Schadd (UM)
Ontology Mapping with Auxiliary Resources.

2015-34 Victor de Graaff (UT)
Geosocial Recommender Systems.

2015-35 Jungxao Xu (TUD)
Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction.

2016

2016-01 Syed Saiden Abbas (RUN)
Recognition of Shapes by Humans and Machines.

2016-015 Steffen Michels (RUN)
Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments.

2016-017 Berend Weel (VU)
Towards Embodied Evolution of Robot Organisms.

2016-019 Julia Efremova (TUE)
Mining Social Structures from Genealogical Data.

2016-02 Michiel Meulendijk (UU)
Optimizing medication reviews through decision support: prescribing a better pill to swallow.

2016-03 Maya Sappelli (RUN)
Knowledge Work in Context: User Centered Knowledge Worker Support.

2016-04 Laurens Rietveld (VU)
Publishing and Consuming Linked Data.

2016-05 Evgeny Sherkhonov (UVA)
Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers.

2016-06 Michel Wilson (TUD)
Robust scheduling in an uncertain environment.

2016-07 Jeroen de Man (VU)
Measuring and modeling negative emotions for virtual training.

2016-08 Matje van de Camp (TiU)
A Link to the Past: Constructing Historical Social Networks from Unstructured Data.

2016-09 Archana Nottamkandath (VU)
Trusting Crowdsourced Information on Cultural Artefacts.

2016-10 George Karafotias (VU)
Parameter Control for Evolutionary Algorithms.

2016-11 Anne Schuth (UVA)
Search Engines that Learn from Their Users.

2016-12 Max Knobbout (UU)
Logics for Modelling and Verifying Normative Multi-Agent Systems.

2016-13 Nana Baah Gyan (VU)
The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach.

2016-14 Ravi Khadka(UU)
Revisiting Legacy Software System Modernization.

2016-16 Guangliang Li (UVA)
Socially Intelligent Autonomous Agents that Learn from Human Reward.

2016-18 Albert Merono Penuela (VU)
Refining Statistical Data on the Web.

2016-20 Daan Odijk (VU)
Context & Semantics in News & Web Search.

2016-21 Alejandro Moreno Celleri (UT)
From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground.

2016-22 Grace Lewis (VU)
Software Architecture Strategies for Cyber-Foraging Systems.

2016-23 Fei Cai (UVA)
Query Auto Completion in Information Retrieval.

2016-24 Brend Wanders (UT)
Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach.

2016-25 Y. Kiseleva (TUE)
Using Contextual Information to Understand Searching and Browsing Behavior.

2016-26 Dilhan J. Thilakarathne (VU)
In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains.

2016-27 Wen Li (TUD)
Understanding Geo-spatial Information on Social Media.

2016-28 Mingxin Zhang (TUD)
Large-scale agent-based social simulation: A study on epidemic prediction and control.

2016-29 Nicolas Honing (TUD)
Understanding Geo-spatial Information on Social Media.

2016-30 Ruud Mattheij (UVT)
The Eyes Have IT.

2016-31 Mohammad Khelghati (UT)
Deep web content monitoring.

2016-32 Eelco Vriezekolk (UVT)
Assessing Telecommunication Service Availability Risks for Crisis Organisations.

2016-33 Peter Bloem (UVA)
Single Sample Statistics, exercises in learning from just one example.

2016-34 Dennis Schunselaar (TUE)
Configurable Process Trees: Elicitation, Analysis, and Enactment.

2016-35 Zhaochun Ren (UVA)
Monitoring Social Media: Summarization, Classification and Recommendation.

2016-36 Daphne Karreman (UT)
Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies.

2016-37 Giovanni Sileno (UVA)
Aligning Law and Action - a conceptual and computational inquiry.

2016-38 Andrea Minuto (UT)
Materials that matter - Smart Materials meet Art & Interaction Design.

2016-39 Merijn Bruijnes
Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect.

2016-40 Christian Detweiler (TUD)
Accounting for Values in Design.

2016-41 Thomas King (TUD)
Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance.

2016-42 Spyros Martzoukos (UVA)
Combinatorial and compositional aspects of bilingual aligned corpora.

2016-43 Saskia Koldijk (RUN)
Context-Aware Support for Stress Self-Management: From Theory to Practice.

2016-44 Thibault Sellam (UVA)
Automatic assistants for database exploration.

2016-45 Bram van Laar (UT)
Experiencing Brain-Computer Interface Control.

2016-46 Jorge Gallego Perez (UT)
Robots to Make you Happy.

2016-47 Christina Weber (UL)
Real-time foresight - Preparedness for dynamic innovation networks.

2016-48 Tanja Buttler (TUD)
Collecting Lessons Learned.

2016-49 Gleb Polevoy (TUD)
Participation and Interaction in Projects: A Game-Theoretic Analysis.

2016-50 Yan Wang (UVT)
The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains.

2017

2017-01 Jan-Jaap Oerlemans (UL)
Investigating Cybercrime.

2017-02 Sjoerd Timmer (UU)
Designing and Understanding Forensic Bayesian Networks using Argumentation.

2017-03 Daniel Harold Telgen (UU)
Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines.

2017-04 Mrunal Gawade (CWI)
Multi-core parallelism in a column-store.

2017-05 Mahdieh Shadi (UVA)
Collaboration Behavior; Enhancement in Co-development.

2017-06 Damir Vandic (EUR)
Intelligent Information Systems for Web Product Search.

2017-07 Roel Bertens (UU)
Insight in Information: from Abstract to Anomaly.

2017-08 Rob Konijn (VU)
Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery.

2017-09 Dong Nguyen (UT)
Text as Social and Cultural Data: A Computational Perspective on Variation in Text.

2017-10 Robby van Delden (UT)
(Steering) Interactive Play Behavior.

2017-11 Florian Kunneman (RUN)
Modelling patterns of time and emotion in Twitter #anticipointment.

2017-12 Sander Leemans (TUE)
Robust Process Mining with Guarantees.

2017-13 Gijs Huisman (UT)
Social Touch Technology - Extending the reach of social touch through haptic technology.

2017-14 Shoshannah Tekofsky (UVT)
You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior.

2017-15 Peter Berck (RUN)
Memory-Based Text Correction.

2017-16 Aleksandr Chuklin (UVA)
Understanding and Modeling Users of Modern Search Engines.

2017-17 Daniel Dimov (UL)
Crowdsourced Online Dispute Resolution.

2017-18 Ridho Reinanda (UVA)
Entity Associations for Search.

2017-19 Jeroen Vuurens (TUD)
Proximity of Terms, Texts and Semantic Vectors in Information Retrieval.

2017-20 Mohammadbashir Sedighi (TUD)
Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility.

2017-21 Jeroen Linssen (UT)
Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds).

2017-22 Sara Magliacane (VU)
Logics for causal inference under uncertainty.

2017-23 David Graus (UVA)
Entities of Interest — Discovery in Digital Traces.

2017-24 Chang Wang (TUD)
Use of Affordances for Efficient Robot Learning.

2017-25 Veruska Zamborlini (VUA)
Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search.

2017-26 Merel Jung (UT)
Socially intelligent robots that understand and respond to human touch.

2017-27 Michiel Joosse (UT)
Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors.

2017-28 John Klein (VU)
Architecture Practices for Complex Contexts.

2017-29 Adel Alhuraibi (UVT)
From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT.

2017-30 Wilma Latuny (UVT)
The Power of Facial Expressions.

2017-31 Ben Ruijl (UL)
Advances in computational methods for QFT calculations.

2017-32 Thaer Samar (RUN)
Access to and Retrievability of Content in Web Archives.

2017-33 Brigit van Loggem (OU)
Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity.

2017-34 Maren Scheffel (OUN)
The Evaluation Framework for Learning Analytics.

2017-35 Martine de Vos (VU)
Interpreting natural science spreadsheets.

2017-36 Yuanhao Guo (UL)
Shape Analysis for Phenotype Characterisation from High-throughput Imaging.

2017-37 Alejandro Montes Garcia (TUE)
WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy.

2017-38 Abdullah Kayal (TUD)
Normative Social Applications.

2017-39 Sara Ahmadi (RUN)
Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR.

2017-40 Altaf Hussain Abro (VUA)
Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems".

2017-41 Adnan Manzoor (VUA)
Minding a Healthy Lifestyle:An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle.

2017-42 Elena Sokolova (RUN)
Causal discovery from mixed and missing data with applications on ADHD datasets.

2017-43 Maaike de Boer (RUN)
Semantic Mapping in Video Retrieval.

2017-44 Garm Lucassen (UU)
Understanding User Stories - Computational Linguistics in Agile Requirements Engineering.

2017-45 Bas Testerink (UU)
Decentralized Runtime Norm Enforcement.

2017-46 Jan Schneider (OU)
Sensor-based Learning Support.

2017-47 Yie Yang (TUD)
Crowd Knowledge Creation Acceleration.

2017-48 Angel Suarez (OU)
Collaborative inquiry-based learning.

2018

2018-01 Han van der Aa (VU)
Comparing and Aligning Process Representations.

2018-02 Felix Mannhardt (TUE)
Multi-perspective Process Mining.

2018-03 Steven Bosems (UT)
Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction.

2018-04 Jordan Janeiro (TUD)
Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks.

2018-05 Hugo Huurdeman (UVA)
Supporting the Complex Dynamics of the Information Seeking Process.

2018-06 Dan Ionita (UT)
Model-Driven Information Security Risk Assessment of Socio-Technical Systems.

2018-07 Jieting Luo (UU)
A formal account of opportunism in multi-agent systems.

2018-08 Rick Smetsers (RUN)
Advances in Model Learning for Software Systems.

2018-09 Xu Xie (TUD)
Data Assimilation in Discrete Event Simulations.

2018-10 Julienka Mollee (VUA)
Moving forward: supporting physical activity behavior change through intelligent technology.

2018-11 Mahdi Sargolzaei (UVA)
Enabling Framework for Service-oriented Collaborative Networks.

2018-12 Xixi Lu (TUE)
Using behavioral context in process mining.

2018-13 Seyed Amin Tabatabaei (VUA)
Computing a Sustainable Future: Exploring the added value of computational models for increasing the use of renewable energy in the residential sector.

2018-14 Bart Joosten (UVT)
Detecting Social Signals with Spatiotemporal Gabor Filters.

2018-15 Naser Davarzani (UM)
Biomarker discovery in heart failure.

2018-16 Jaebok Kim (UT)
Automatic recognition of engagement and emotion in a group of children.

2018-17 Jianpeng Zhang (TUE)
On Graph Sample Clustering.

2018-18 Henriette Nakad (UL)
De Notaris en Private Rechtspraak.

2018-19 Minh Duc Pham (VUA)
Emergent relational schemas for RDF.

2018-20 Manxia Liu (RUN)
Time and Bayesian Networks.

2018-21 Aad Slootmaker (OU)
EMERGO: a generic platform for authoring and play-

ing scenario-based serious games.

2018-22 Eric Fernandes de Mello Araujo (VUA)
Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks.

2018-23 Kim Schouten (EUR)
Semantics-driven Aspect-Based Sentiment Analysis.

2018-24 Jered Vroon (UT)
Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots.

2018-25 Riste Gligorov (VUA)
Serious Games in Audio-Visual Collections.

2018-26 Roelof de Vries (UT)
Theory-Based And Tailor-Made: Motivational Messages for Behavior Change Technology.

2018-27 Maikel Leemans (TUE)
Hierarchical Process Mining for Scalable Software Analysis.

2018-28 Christian Willemse (UT)
Social Touch Technologies: How they feel and how they make you feel.

2018-29 Yu Gu (UVT)
Emotion Recognition from Mandarin Speech.

2018-30 Wouter Beek (VU)
The K in semantic web stands for knowledge: scaling semantics to the web.

2019

2019-01 Rob van Eijk (UL)
Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification.

2019-02 Emmanuelle Beauxis- Aussalet (CWI, UU)
Statistics and Visualizations for Assessing Class Size Uncertainty.

2019-03 Eduardo Gonzalez Lopez de Murillas (TUE)
Process Mining on Databases: Extracting Event Data from Real Life Data Sources.

2019-04 Ridho Rahmadi (RUN)
Finding stable causal structures from clinical data.

2019-05 Sebastiaan van Zelst (TUE)
Process Mining with Streaming Data.

2019-06 Chris Dijkshoorn (VU)
Nichesourcing for Improving Access to Linked Cultural Heritage Datasets.

2019-07 Soude Fazeli (TUD)
Recommender Systems in Social Learning Platforms.

2019-08 Frits de Nijs (TUD)
Resource-constrained Multi-agent Markov Decision Processes.

2019-09 Fahimeh Alizadeh Moghaddam (UVA)
Self-adaptation for energy efficiency in software systems.

2019-10 Qing Chuan Ye (EUR)
Multi-objective Optimization Methods for Allocation and Prediction.

2019-11 Yue Zhao (TUD)
Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs.

2019-12 Jacqueline Heinerman (VU)
Better Together.

2019-13 Guanliang Chen (TUD)
MOOC Analytics: Learner Modeling and Content Generation.

2019-14 Daniel Davis (TUD)
Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses.

2019-15 Erwin Walraven (TUD)
Planning under Uncertainty in Constrained and Partially Observable Environments.

2019-16 Guangming Li (TUE)
Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models.

2019-17 Ali Hurriyetoglu (RUN)
Extracting actionable information from microtexts.

2019-18 Gerard Wagenaar (UU)
Artefacts in Agile Team Communication.

2019-19 Vincent Koeman (TUD)
Tools for Developing Cognitive Agents.

2019-20 Chide Groenouwe (UU)
Fostering technically augmented human collective intelligence.

2019-21 Cong Liu (TUE)
Software Data Analytics: Architectural Model Discovery and Design Pattern Detection.

2019-22 Martin van den Berg (VU)
Improving IT Decisions with Enterprise Architecture.

2019-23 Qin Lin (TUD)
Intelligent Control Systems: Learning, Interpreting, Verification.

2019-24 Anca Dumitrache (VU)
Truth in Disagreement- Crowdsourcing Labeled Data for Natural Language Processing.

2019-25 Emiel van Miltenburg (VU)
Pragmatic factors in (automatic) image description.

2019-26 Prince Singh (UT)
An Integration Platform for Synchromodal Transport.

2019-27 Alessandra Antonaci (OUN)
The Gamification Design Process applied to (Massive) Open Online Courses.

2019-28 Esther Kuindersma (UL)
Cleared for take-off:Game-based learning to prepare airline pilots for critical situations.

2019-29 Daniel Formolo (VU)
Using virtual agents for simulation and training of social skills in safety-critical circumstances.

2019-30 Vahid Yazdanpanah (UT)
Multiagent Industrial Symbiosis Systems.

2019-31 Milan Jelisavcic (VU)
Alive and Kicking: Baby Steps in Robotics.

2019-32 Chiara Sironi (UM)
Monte-Carlo Tree Search for Artificial General Intelligence in Games.

2019-33 Anil Yaman (TUE)
Evolution of Biologically Inspired Learning in Artificial Neural Networks.

2019-34 Negar Ahmadi (TUE)
EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES.

2019-35 Lisa Facey-Shaw (OUN)
Gamification with digital badges in learning programming.

2019-36 Kevin Ackermans (OUN)
Designing Video-Enhanced Rubrics to Master Complex Skills.

2019-37 Jian Fang (TUD)
Database Acceleration on FPGAs.

2019-38 Akos Kadar (OUN)
Learning visually grounded and multilingual representations.

2020

2020-01 Armon Toubman (UL)
Calculated Moves: Generating Air Combat Behaviour.

2020-02 Marcos de Paula Bueno (UL)
Unraveling Temporal Processes using Probabilistic Graphical Models.

2020-03 Mostafa Deghani (UvA)
Learning with Imperfect Supervision for Language Understanding.

2020-04 Maarten van Gompel (RUN)
Context as Linguistic Bridges.

2020-05 Yulong Pei (TUE)
On local and global structure mining.

2020-06 Preethu Rose Anish (UT)
Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support.

2020-07 Wim van der Vegt (OUN)
Towards a software architecture for reusable game components.

2020-08 Ali Mirsoleimani (UL)
Structured Parallel Programming for Monte Carlo Tree Search.

2020-09 Myriam Traub (UU)
Measuring Tool Bias & Improving Data Quality for Digital Humanities Research.

2020-10 Alifah Syamsiyah (TUE)
In-database Preprocessing for Process Mining.

2020-11 Sepideh Mesbah (TUD)
Semantic-Enhanced Training Data Augmentation-Methods for Long-Tail Entity Recognition Models.

2020-12 Ward van Breda (VU)
Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment.

2020-13 Marco Virgolin (CWI/ TUD)
Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming.

2020-14 Mark Raasveldt (CWI/UL)
Integrating Analytics with Relational Databases.

2020-15 Georgiadis Konstantinos (OU)
Smart CAT: Machine Learning for Configurable Assessments in Serious Games.

2020-16 Ilona Wilmont (RUN)
Cognitive Aspects of Conceptual Modelling.

2020-17 Daniele Di Mitri (OU)
The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences.

Open Universiteit