COGNTIVE 2024

The Sixteenth International Conference on Advanced Cognitive Technologies and Applications

ISBN: 978-1-68558-157-2

April 14 - 18, 2024

Venice, Italy

COGNITIVE 2024 Editors

Jérôme Dinet, Université de Lorraine, France

COGNITIVE 2024

Forward

The Sixteenth International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2024), held on April 14 – 18, 2024, targeted advanced concepts, solutions and applications of artificial intelligence, knowledge processing, agents, as key-players, and autonomy as manifestation of self-organized entities and systems. The advances in applying ontology and semantics concepts, web-oriented agents, ambient intelligence, and coordination between autonomous entities led to different solutions on knowledge discovery, learning, and social solutions.

The conference had the following tracks:

- Brain information processing and informatics
- Artificial intelligence and cognition
- Agent-based adaptive systems
- Applications
- Autonomous systems and autonomy-oriented computing
- Hot topics on cognitive science

Similar to the previous edition, this event attracted excellent contributions and active participation from all over the world. We were very pleased to receive top quality contributions.

We take here the opportunity to warmly thank all the members of the COGNITIVE 2024 technical program committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and effort to contribute to COGNITIVE 2024. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations and sponsors. We also gratefully thank the members of the COGNITIVE 2024 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope COGNITIVE 2024 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of cognitive technologies and applications. We also hope that Venice provided a pleasant environment during the conference and everyone saved some time to enjoy enjoy this beautiful city.

COGNITIVE 2024 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COGNITIVE 2024 Steering Committee

Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA
Thomas Ågotnes, University of Bergen, Norway
Muneo Kitajima, Nagaoka University of Technology (Emeritus), Japan

COGNITIVE 2024 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

COGNITIVE 2024

Committee

COGNITIVE 2024 General Chair

Jaime Lloret Mauri, Universitat Politecnica de Valencia, Spain

COGNITIVE 2024 Steering Committee

Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA
Thomas Ågotnes, University of Bergen, Norway
Muneo Kitajima, Nagaoka University of Technology (Emeritus), Japan

COGNITIVE 2024 Publicity Chair

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

COGNITIVE 2024 Technical Program Committee

Witold Abramowicz, University of Economics and Business, Poland
Thomas Agotnes, University of Bergen, Norway
Vered Aharonson, University of the Witwatersrand, Johannesburg, South Africa
Thangarajah Akilan, Lakehead University, Canada
Piotr Artiemjew, University of Warmia and Masuria in Olsztyn, Poland
Divya B, SSNCE, India
Thierry Bellet, University Gustave Eiffel, France
Petr Berka, University of Economics, Prague, Czech Republic
Ateet Bhalla, Independent Consultant, India
Mahdi Bidar, University of Regina, Canada
Guy Andre Boy, CentraleSupélec, LGI, Paris Saclay University / ESTIA Institute of Technology, France
Dilyana Budakova, Technical University of Sofia - Branch Plovdiv, Bulgaria
Valerie Camps, Paul Sabatier University - IRIT, Toulouse, France
Yaser Chaaban, Leibniz University of Hanover, Germany
Olga Chernavskaya, P. N. Lebeved Physical Institute, Moscow, Russia
Helder Coelho, Universidade de Lisboa, Portugal
Igor Val Danilov, Academic Center for Coherent Intelligence, Latvia
Angel P. del Pobil, Jaume I University, Spain
Soumyabrata Dev, University College Dublin, Ireland
Jerome Dinet, University of Lorraine, France
Serena Doria, University G.d'Annunzio Chieti, Italy
Piero Dominici, University of Perugia, Italy
Jayfus Tucker Doswell, The Juxtopia Group, Inc., USA

Table of Contents

Work in Progress: Promoting Self-Regulation for Children with Autistic Spectrum Disorder with Robots

Rychalski Marie
Association Jean-Baptiste Thiéry
Maxéville, France
Lab of Psychology and Neurosciences
Lorraine University
Nancy, France
email: marie.rychalski@univ-lorraine.fr

Manukyan Armand
Association Jean-Baptiste Thiéry
Maxéville, France
Lab of Psychology and Neurosciences
Lorraine University
Nancy, France
email: armand.manukyan@jbthiery.asso.fr

Dinet Jérôme
Lab of Psychology and Neurosciences
Lorraine University
Nancy, France
email: jerome.dinet@univ-lorraine.fr

***Abstract*—The main challenge in educating children with specific needs, particularly those with Autistic Spectrum Disorder (ASD), revolves around their adjustment to the learning environment and their cognitive functions, notably through information processing, executive function control and emotional regulation, which refer to Self-Regulation (SR). SR deficits address significant hurdles, not only for children with ASD but also for teachers who encounter difficulties trying to implement self-regulation systems. The simplicity of robots, such as facial expressions or monotone voice would create less stressful interactions for children with ASD and would improve SR in educational context. Integration models like the 4A model aim to gradually introduce these tools into the classroom. This experimental study, conducted in an ecological settings (which were not studied in literature), has a dual purpose: first, to evaluate the acceptability of robotic agents among teachers and educators; second, to explore how these artificial agents can enhance self-regulation development for children with ASD. This study will have both, qualitatives and quantitatives measures. We expect that Self-Regulation of a child with ASD using a robot depends on the functionality of the robot. Moreover, the most effective and acceptable robot for teachers will be easy to program, have an intuitive interface, and possess hybrid particularities. Finally, we expect that these robots will support children's self-regulation by a learning effect over the long-term. The aim of this study is to deliver a guide for robots' utilization for SR, and to generalize it in school context.**

***Keywords*—Self-Regulation ; Autistic Spectrum Disorder ; Robotic ; Children ; Teachers ; Learnings.**

I. Introduction

Autistic Spectrum Disorder (ASD) is a public health issue with 1% of the population. It is characterized by "Persistent deficits in social communication and social interaction" and "Restricted, repetitive patterns of behavior, interests, or activities". Presence or absence must be specified of: "intellectual impairment, language impairment, medical or genetic condition or environmental factor, another neurodevelopmental, mental or behavior disorder, or catatonia" [1]. However, one of the main difficulties for children with ASD is their adaptation to the environment and their cognitive functioning, notably through emotional regulation, executive function control and information processing, which refer to Self-Regulation (SR). This concept of SR can be explained as a conscious process by which the person orients their own thoughts, actions, and emotions for a specific goal [2]. SR is essential in learning, more particularly in school learning. In fact, when an individual can regulate their own learning, they can develop autonomy and thinking skills in success of the task to fulfil. Therefore, SR behaviors are aiming children to develop autonomy in learning path [3]. In addition, Whitman [4] explains that people with Intellectual Disabilities (ID) show a bigger deficit in SR, this makes it more difficult to generalize knowledge in unfamiliar situations. For that reason, this leads to adaptation difficulties. Finally, Nader-Grosbois and Thomée [5] complete this statement by explaining that other difficulties are directly related to SR difficulties, such as planning actions, maintain attention, choosing cognitive strategies or staying involved in the task. All these difficulties impact teachers as well, by complex conditions at work with lack of time or feeling of stress. This complicates the implementation of a SR system in a class context for all students and of a new tool's integration [6]. SR deficit damages children's learning and then, has negative impacts on the feeling of efficiency and the support of teachers and professionals.

A. Relationships between Use, Usability and Acceptance of Robots

To prevent professional and children's difficulties, the use of a robotic device, as an addition of the adult's support and adapted for children, could be a practical and a relevant solution. To help teachers and professionals in their support to children, robot would offer help in children SR development, which simplify their learning and teacher's practice. Many advantages of robotic use have been demonstrated in literature to work on executive's functions among the audiences mentioned, which are affected by SR. In fact, Dinet and al. [7] used a Cozmo robot in an ecologic context in Autistic Elementary Education Unit (AEEU) among 9 years old children. Robot allows to improve interaction between robot and children and improve caregiver's perceptions on their own capacity and children's. Cozmo's presence also helped to increase interactions peer-to-peer in role-playing task. In addition, Oesch and al. [8] show the benefits of using the NAO robot in lexical acquisition of children with ASD, highlighting the increase of gaze, lexicon and response time. Finally, Azizi and al. [9] highlighted a greater involvement of children with Learning

 ISBN: 978-1-68558-157-2

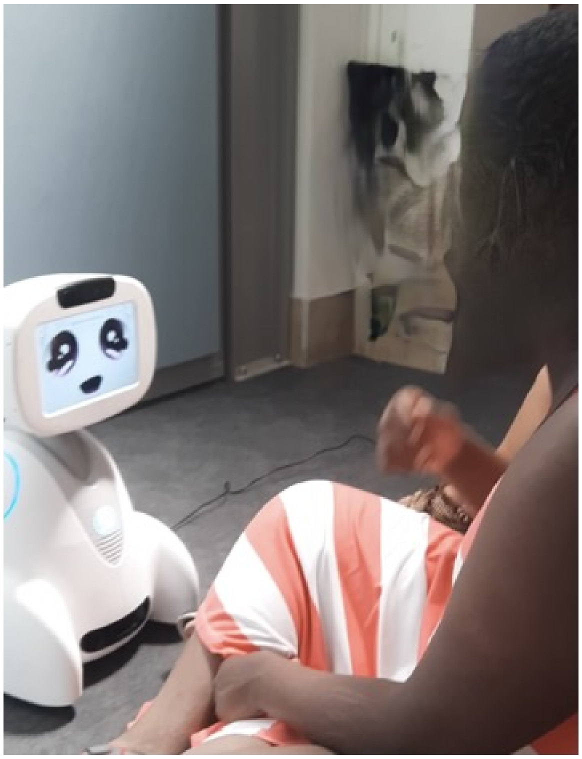

Figure 1: A young child interacting with the Buddy robot

Disabilities (LD) when the task is performed in the presence of the robot (QTRobot). Authors conclude that a social robot can facilitate children's commitment and attention in a educational framework. Both of these capacities are affected by SR, which supports the use of robot in our context. It also provides a way of helping professionals in the field who support children. For example, Figure 1 shows Buddy robot in interaction with a young child.

Nowadays, there are many types of robots on the market (toys, humanoids, animals, machine, etc.). In every type of robots, there are advantages and disadvantages in their use, so it is important to take an interest in a numeric tool integration context for special needs children. In fact, simple robots and animals' ones show their advantages among ASD children, in social skills by facilitating emotional comprehension had to less complex facial expressions. Also, a reduction of sensorial overload and of level of anxiety usually feel with humans have been shown. Concerning android type of robots, they showed no increased capacity on acquired skills over the long term, although some communication improvements have been shown [10]. Moreover, most of experiments were conducted in a laboratory context, which does not allow integration and utilization of tool in an ecological context, suitable and efficient for students and professionals. In a laboratory context, difficulties of teachers' handling, use and acceptability are not considered. Robots can work on each aspect of SR (emotional, behavioural and cognitive) but there are not experimentation which study the effect of robot on global SR yet, this is the goal of this experimentation. Moreover, clinical observation highlights the difficult acceptance and integration of robots by professionals in the field (teachers and educators) in their practice. This emphasize the need for a thorough study of the factors and determinants of the acceptance of robots by teachers in their environment and their practice. To better understand the role of the different parameters explaining the acceptability, and finally the use of a robotic system, the 4A model [11] is the most relevant theoretical background.

B. The 4A Model: Acceptability, Acceptation, Approval, Appropriation

As Figure 2 shows, the 4A model [11] [12] provides an explanation of the temporal process of appropriation of a digital device, such as a robot (for a complete presentation of the model, see [12] [11]. Several studies related to the TAM theory [13] [14] [15] or the UTAUT theory [16] [17] [18] describe the role of attitudes and opinions on future acceptability and acceptance of digital devices [13] [14] [15]. But even if all these studies related to TAM or UTAUT theories provide very interesting results, they have four important limitations that prevent to generalize results: (i) data are often collected by using questionnaires and surveys, i.e., only attitudes, opinions and verbalization are collected; (ii) data are often collected during only one-shot setting, and thus do not investigate the longitudinal and temporal process of appropriation across the time; (iii) they assume that the effective use of a digital device means that this device is accepted; (iv) context and environments (professional, physical, social) are never considered.

The 4A model has several advantages:

- This 4A model allows to better understand the relationships between attitudes, opinions and effective behaviours;
- The temporal and longitudinal dimensions related to the appropriation are included by distinguishing before and after the implementation of the device in the context;
- If attitudes determine effective behaviours, the 4A model proposes that behaviours can also have an impact on attitudes by a reciprocal action (i.e., retro-feedback);
- According to the 4A model, an effective use of a device does not necessarily mean that this device is accepted: in some cases, the use is forced and thus, does not indicate that the device is really accepted.

This 4A model is the only one model that considers representations, cognitive biases, as well as the tool's ease of use and adaptability, offering insights into the integration process. This model is also interesting from an ecological point of view by its consideration of professional's perceptions of robots and their interaction with them. The 4A model highlights that the acceptance of the tool impacts its adoption and incorporation. Hence, professional's view of the robot, its ease of use and the associated usage-related challenges serve as perspective factors for its practical utilization. A progressive handling of the tool allows to facilitate teachers' comprehension and to focus on the use to offer an efficient support, with less workload for professionals.

 ISBN: 978-1-68558-157-2

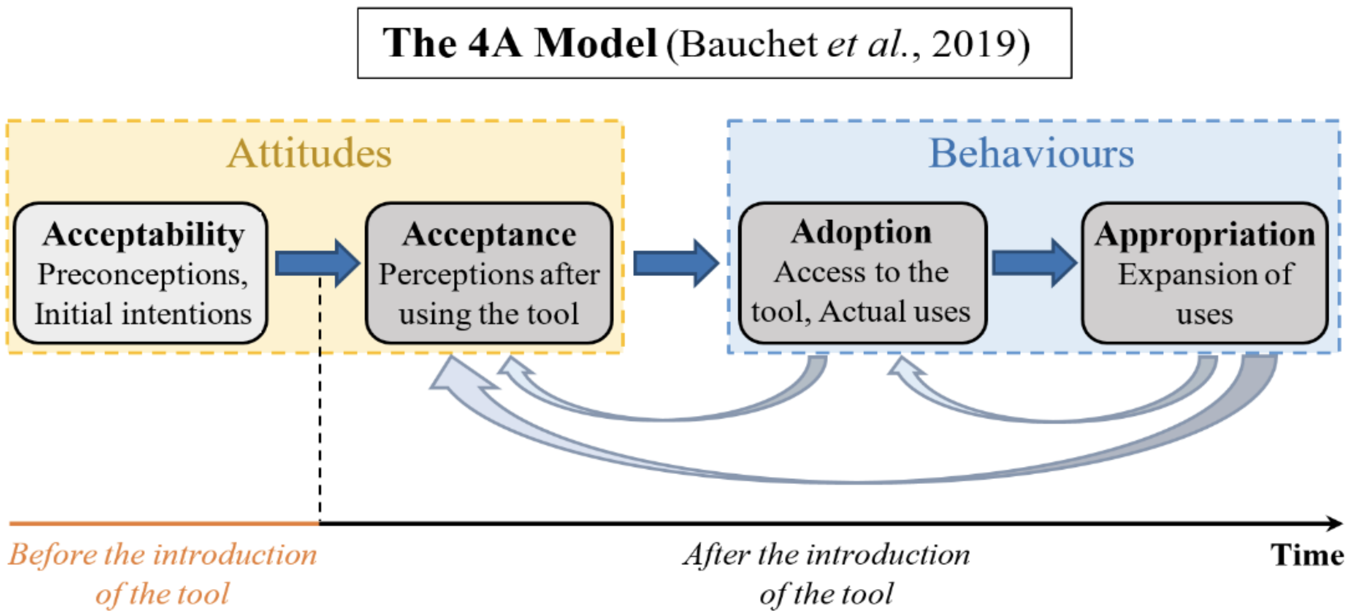

Figure 2: The 4A Model [12] [11]

This research project is based on the premise that robots can be beneficial in developing self-regulation in children with autism. To this end, it emphasizes the importance of comparing the benefits of different robots in an ecological context, and the need to study the factors of acceptability to teachers, as Figure 3 shows.

Following this introduction which presented concepts of ASD and SR, the relationships between Use, Usability and Acceptance and completed by a presentation of the 4A model, the article is structured as follows: in section 2, methodology will be presented, specifying firstly the problematic, secondly hypothesis, thirdly the population and finally the material used. In section 3, discussion and conclusion will be proposed.

II. Method

This study aims in one hand to evaluate a robotic agent's acceptability criteria among teachers and educators. In the other hand, it purposes to the agent's integration in educational system to a self-regulation development of Children with Autistic Spectrum Disorder (ASD). This study will begin by focus groups and questionnaires with specialized professionals, to study needs and expectations. The aim is to create a collaboration between both and collect the point of view of different context on inclusion, difficulties, and intervention with children.

The second step of this work will be to study the robots' efficiency both on self-regulation for children and on integration for professionals (depending on the results of the precedent step by carrying out a data analysis between step 1 and 2). For them, the goal is to determine several features, such as: ease of use, predictability, flexibility, difficulties, and robots' representation. For children, the goal is to increase self-regulation capacities, more precisely on executives' functions (inhibition, visual attention, and tasks memorization). This stage will consist of the development of activities on executives' functions with robot to work on self-regulation in a school context.

The third step will consist of passing new focus groups and questionnaires, to study the efficiency of robots' integration in school context and the impact on inclusion, from professionals' point of view. In those new focus groups, questions will be raised about: representation, integration, permanence over time, generalizing learning and perpetuation.

The fourth step of this study will be data analysis of step 2 and 3. Focus group analysis will provide qualitative measures for children and professionals. Self-regulation analysis will provide quantitative measures for children on cognitive and behavioral self-regulation. There will be different statistical models used, such as Pearson correlation and repeated measures ANOVA.

A. *Research Question*

All our previous thoughts lead us to study the integration of robots into a teaching system. Therefore, we investigate the question : to what extent can a robot be integrated into a teaching system adapted to develop the self-regulation of a child with an Autistic Spectrum Disorder?

B. *Hypothesis*

H1: SR of a child with ASD using a robot depends on the robot's functionalities.

H2: Robot facilitates the SR of children through a long-term learning effect.

H3: The integration of robots depends on the user experience (the user's acceptance of the tool).

 ISBN: 978-1-68558-157-2

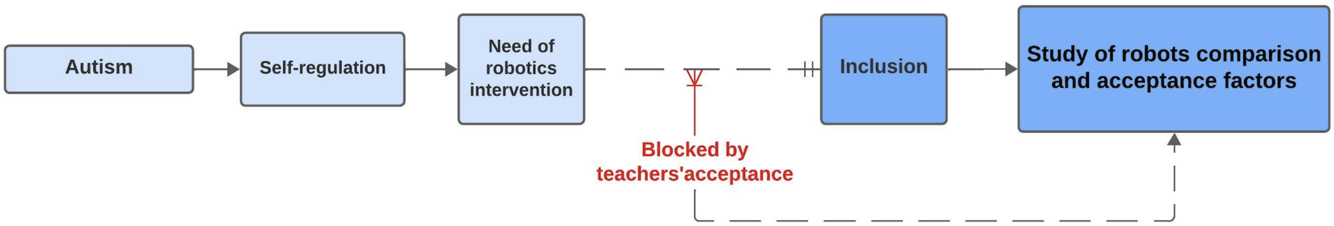

Figure 3: Proposition of a visual presentation of our subject

C. Population

To answer this question, our population will be composed of 50 to 70 children between 3 and 11 years old, with ASD. Data collection will be in educational units of National Education in the Nancy-Metz Academy perimeter and in educational units of J.B. Thiéry Association, in collaboration with pedagogic and educational team.

D. Material

This experiment consists of evaluating 3 dimensions: Children's SR, professionals' perception, and robot's acceptability. Mixed evaluation method will be used to evaluate these 3 dimensions, by using qualitative and quantitative measures (interviews, questionnaires, observation grid). Measures and evaluation scales are under development. The days of presence with the robot will be agreed with the structure's professionals. The Association J.B Thiéry owns several robots: Leka, NAO and Buddy. These robots will be used as part of this study and will be brought in on the agreed days. The project will be studied by an Ethics Committee and will be the subject of a free and informed consent addressed to the parents. Data will be anonymized.

III. Conclusion and Future Work

A fundamental issue for children with ASD is their difficulties in adapting to the learning environment and managing cognitive functions like information processing, executive control, and emotional regulation, which are part of self-regulation (SR). Self-regulation, as the ability to establish strategies and implement means to achieve a goal, is increasingly emerging as a major deficit in autism, and a target for strategic support. The effectiveness of certain robots for developing self-regulation functions in controlled environments means that they need to be evaluated in depth in schools. However, their acceptance by professionals has been hampered by a number of obstacles, such as representations or high cognitive costs. The aim of the study presented in this article is to observe the nature of these obstacles, and to determine the best ways to overcome them. At the moment, a guide is being produced to facilitate robot's utilization and to determine which capacities of children are involved. This will be used to generalized the use of robots in educational context.

Acknowledgment

We would like to thank the Association Jean-Baptiste Thiery in Maxéville (France) for its support and the provision of an experimentation area and robots.

References

[1] A. P. Association *et al.*, "Diagnostic and statistical manual of mental disorders: Dsm-5," 2013.
[2] W. Hofmann and K. Vohs, "Desire and self-regulation," *Handbook of self-regulation: Research, theory, and applications*, pp. 76–94, 2016.
[3] J. Veijalainen, J. Reunamo, and A. Alijoki, "Children's self-regulation skills in the finnish day care environment," *Journal of Early Childhood Education Research*, vol. 6, no. 1, pp. 89–107, 2017.
[4] T. L. Whitman, "Self-regulation and mental retardation.," *American Journal of Mental Retardation: AJMR*, vol. 94, no. 4, pp. 347–62, 1990.
[5] N. Grosbois and C. Thomée, "How can we study the self-regulation of children with mental retardation and parental self-regulation in media and non-media learning situations?," 2006.
[6] D. Ruaud, "Teacher stress," *Administration et Éducation*, no. 4, pp. 79–85, 2022.
[7] Dinet *et al.*, "A hybrid robotics system for children with autism: A longitudinal study of social interaction inside the classroom," in *Proceedings of the 17th "Ergonomie et Informatique Avancée" Conference*, pp. 1–8, 2021.
[8] F. Oesch, A. Piquard-Kipffer, and A. Manukyan, "The benefits of using the nao robot in the lexical acquisition of children with asd : a study of four students in an autism elementary teaching unit in france," in *13th Autism-Europe International Congress 2022" Conference*, pp. 1–2, 2023.
[9] Azizi *et al.*, "User evaluation of social robots as a tool in one-to-one instructional settings for students with learning disabilities," in *International Conference on Social Robotics*, pp. 146–159, Springer, 2022.
[10] Kumazaki *et al.*, "Optimal robot for intervention for individuals with autism spectrum disorders," *Psychiatry and Clinical Neurosciences*, vol. 74, no. 11, pp. 581–586, 2020.
[11] C. Bauchet, B. Hubert, and J. Dinet, "Between acceptance and appropriation of integrated digital tools in the education system: The 4a model," in *13ème colloque international RIPSYDEVE La psychologie du développement et de l'éducation pour le 21ème siècle: nouveaux objets, espaces et temporalités*, pp. 158–161, 2020.
[12] C. Bauchet, B. Hubert, and J. Dinet, "From acceptability of digital change to appropriation of technology: The 4a model," in *17th EARA Conference" Adolescence in a rapidly changing world"*, pp. 1–2, 2020.
[13] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.
[14] F. D. Davis, "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts," *International journal of man-machine studies*, vol. 38, no. 3, pp. 475–487, 1993.
[15] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models," *Management science*, vol. 35, no. 8, pp. 982–1003, 1989.
[16] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management science*, vol. 46, no. 2, pp. 186–204, 2000.

 ISBN: 978-1-68558-157-2

[17] V. Venkatesh, “Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model,” *Information systems research*, vol. 11, no. 4, pp. 342–365, 2000.

[18] V. Venkatesh, F. Davis, and M. G. Morris, “Dead or alive? the development, trajectory and future of technology adoption research,” *The Development, Trajectory and Future of Technology Adoption Research (April 27, 2007). Venkatesh, V., Davis, FD, and Morris, MG “Dead or Alive*, pp. 267–286, 2007.

Restricting In-variance and Co-variance of Representations for Adversarial Defense in Vision Transformers

Jeevithan Alagurajah
School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, LA 70504, U.S.A.
Email: jeevithan.alagurajah1@louisiana.edu

Chee-Hung Henry Chu[a,b]
[a]Informatics Research Institute
[b]School of Computing and Informatics
University of Louisiana at Lafayette
Lafayette, LA 70504, U.S.A.
Email: chu@louisiana.edu

***Abstract*—The development of trustworthy and secure AI applications is a fundamental step towards building AI systems that can reliably operate in the real world, where they may face malicious attempts to manipulate them. Deep neural networks, despite their impressive image classification accuracy, are vulnerable to even small, imperceptible changes called adversarial attacks, causing their performance to plummet. Existing defenses often struggle when attackers have full knowledge of the model (white-box attacks) and craft even stronger perturbations. To address this, the Adversarial Invariant and Co-Variance Restriction (AICR) loss function was recently proposed. The AICR loss function forces clean and noisy images from the same class to have similar activation patterns in convolutional neural networks, essentially making them harder for attackers to differentiate. Given the superior performance of Vision Transformers (ViTs) in image classification, we adapted the AICR loss to train ViTs and investigated its effectiveness against gradient-based attacks. Our experiments show that ViTs trained with AICR loss achieve a significant improvement in accuracy compared to those trained with the standard cross-entropy loss, demonstrating the effectiveness of AICR in enhancing ViT's resilience against adversarial attacks.**

***Index Terms*—Vision transformers, adversarial training, adversarial defense, image classification**

I. INTRODUCTION

While Deep Neural Networks (DNNs) have achieved impressive performance in various domains such as computer vision [1]–[4], natural language processing [5]–[7], speech recognition [8]–[10], and reinforcement learning [11]–[13], their vulnerability to adversarial attacks remains a critical concern. Adversarial attacks involve creating imperceptible modifications to input data that can fool DNNs into making incorrect predictions. This vulnerability raises serious questions about the reliability and accountability of DNNs, particularly in such applications as autonomous vehicles, security and surveillance, and environmental monitoring. Designing AI systems that are robust and generalizable against adversarial attacks is therefore an urgent and crucial challenge, thus requiring further research to ensure the safe and responsible deployment of DNNs across various critical domains.

In recent years, numerous defense methods have been proposed to combat adversarial attacks. These methods can be broadly categorized into two types: reactive and proactive defenses. Reactive defense methods focus on modifying or transforming inputs to counter specific attack strategies. Examples of such methods include input smoothing, input clipping, and adversarial training with specific noise distributions [14]–[17]. However, these methods are often limited in their effectiveness against unknown attack strategies and may not generalize well to diverse adversarial perturbations. Proactive defense methods, on the other hand, aim to inherently improve the robustness of the model itself. This is achieved by modifying model parameters, network architectures, training objectives, or utilizing adversarial training with diverse noise distributions. Proactive methods are generally more versatile and provide wider protection against various attacks, making them more widely adopted in practical applications.

A novel adversarial training framework using the so-called Adversarial Invariant and Co-Variance Restriction (AICR) loss function was proposed [18] that incorporates two objectives to enhance robustness and generalization: maximizing class separation and minimizing intra-class variance. The first objective utilizes an attractive and repulsive mechanism at different representation levels. This mechanism encourages samples from the same class to cluster together (attractive) and pushes samples from different classes apart (repulsive). This promotes a model that naturally separates classes, making it more resilient to noise and achieving better generalization. The second objective aims to minimize the variation between adversarial and clean images within the same class. This is achieved by maximizing the correlation and minimizing the redundancy in their intermediate representations. This ensures that visually similar samples are projected to the same region in the multidimensional space, making them difficult to fool with adversarial attacks within a given budget. This approach was shown (i) to achieve state-of-the-art robustness against a wide range of strong adversarial attacks under the strongest first-order attack and (ii) to maintain the highest robust performance under black-box settings using the CNN6 net [19] and the ResNet-110 [4] Convolutional Neural Networks (CNNs).

Vision Transformers (ViTs) [20] is a different DNN ap-

 ISBN: 978-1-68558-157-2

proach towards image classification compared to CNNs. Unlike CNNs, which process images pixel-by-pixel, ViTs first divide an image into patches and treat these patches as tokens. The network then learns by examining the relationships between these tokens, similar to how it learns relationships between words in a sentence. This allows ViTs to capture long-range dependencies and global context within an image. CNNs have a strong inductive bias towards local spatial relationships due to their convolutional structure. This helps them learn effectively even with smaller datasets. ViTs rely on attention mechanisms, which have a weaker inductive bias towards spatial structure. This means they are more flexible but may require more data or specific techniques to guide them towards learning the important features. When training on smaller datasets, ViTs are more likely to overfit, which increases its reliance on model regularization or data augmentation.

While the AICR loss function has demonstrated superior adversarial defense performance on CNNs like Resnet-110, its effectiveness on ViTs remains unexplored. ViTs have recently gained significant traction for their impressive image classification capabilities. However, their vulnerability to adversarial attacks is only beginning to be addressed. Given the recent surge in interest and potential benefits of ViTs, we propose incorporating them into the AICR objective function to assess their adversarial defense performance.

The rest of this paper is organized as follows. Relevant related work is reviewed in Section II. In Section III, we describe our proposed method and our testing plan. We present our experimental results in Section IV. Finally, we draw our conclusion and suggest future work in Section V.

II. Related Work

Deep learning algorithms are vulnerable to adversarial perturbations, which are carefully crafted inputs that can cause the model to make incorrect predictions. Several defense algorithms have been proposed to counter such attacks, and these can be broadly categorized into two main approaches: input transformation and model modification.

Input transformation methods attempt to modify the input data in a way that makes it more difficult for the adversary to craft effective perturbations [17], [21]. For example, one common technique is to add noise to the input data. This can make it more difficult for the adversary to find perturbations that have a significant impact on the model's output.

Model modification methods attempt to make the model itself more robust to adversarial perturbations [22]–[24]. One common technique is adversarial training, which involves training the model on a dataset that includes both clean and adversarial examples. This can help the model to learn how to better distinguish between clean and adversarial inputs.

Given the recent competitive performance of ViTs for image classification tasks, the robustness of these methods against adversarial attacks has received more attention. ViT architectures can be broadly classified as vanilla and hybrid. Vanilla ViTs are pure attention-based and are computationally less expensive. Hybrid ViTs combine CNNs and attention by incorporating both convolutional layers and self-attention modules. This leverages the strengths of both approaches, viz. CNNs for extracting local features and attention for capturing global relationships. The robustness of vanilla and of hybrid ViTs against adversarial attacks were found to be different [25]. While both vanilla and hybrid ViTs are tougher against adversarial attacks compared to regular CNNs, it was shown that vanilla ViTs but not the hybrids resisted defenses aimed at high-frequency features, suggesting potential differences in how they process information.

The adversarial robustness of ViTs has been explored by focusing on their unique building blocks [26]. It was shown that significant improvement in their ability to resist deception is feasible by randomly hiding information within these blocks during training.

A different approach to achieve adversarial robustness in ViT is to modify the training recipe [27]. Traditionally, training ViTs relies heavily on data augmentation. While effective for normal training, this approach was shown to hurt performance during adversarial training. Instead, omitting data augmentation and incorporating specific techniques like ε-warmup and bigger weight decay significantly improves the robustness of ViTs.

Adversarial training on the ViT architecture is computationally expensive. An attention-guided adversarial training method was introduced to trade off computational efficiency and adversarial robustness [28]. This method identifies and removes unimportant parts of an image during training, focusing the model's attention on crucial areas. This significantly speeds up training while maintaining or even improving robustness.

III. Method

We propose to incorporate the AICR loss function [18] into the ViT architecture and test the performance against white-box attacks. We summarize the characteristics of the ViT architecture next.

A. The Vision Transformer (ViT)

Transformers, originally developed for natural language processing, excel at image classification with their ability to capture long-range dependencies and contextual relationships within images [20]. The image is first divided into smaller patches, each of which is converted into a fixed-length embedding vector, capturing essential information about its color, texture, and other features. Additional information about the relative position of each patch within the image is incorporated into the embedding vectors.

In the transformer encoder's self-attention mechanism, each patch embedding attends to all other patch embeddings, allowing it to learn how relevant each patch is to itself and other parts of the image. It does so by transforming the patch embedding into three separate vectors: a query, a key, and a value. Each query vector is compared to all other key vectors using a dot product operation to generate a matrix of attention scores, where each score represents the similarity between a pair of patches. Each attention score is normalized

 ISBN: 978-1-68558-157-2

using a softmax function, turning it into a weight. This weight indicates the relative importance of each patch to the query patch. The values of all patches are multiplied by their respective weights and then summed up. This results in a new vector that represents the query patch based on the information from all other patches, weighted by their relevance.

The multi-head attention is formed by repeating the self-attention process multiple times in parallel, using different, randomly initialized matrices for generating queries, keys, and values. The attended representations from all heads are concatenated together, which is is then transformed by a final linear layer to produce the output of the multi-head attention layer.

A feed-forward network further processes the information extracted by the self-attention layers, adding non-linearity to increase the model's expressiveness. Residual connections and layer normalization help stabilize the training process to improve the overall performance of the model.

After the final transformer encoder layer, the output vector representing the entire image is passed through a classification head, which is typically a Multi-Layer Perceptron (MLP) that predicts the probability of the image belonging to each class.

B. Loss Function for ViT Adversarial Training

In [18], it was shown that the AICR loss provides an effective and robust defense against state-of-the-art white-box attacks and black-box settings. We adapt the AICR loss from its constituent loss functions for the ViT architecture as follows.

Let K be the number of classes of a given data set distribution $\mathcal{D}$ and N be the number of samples in the data set. For an image classification task, we formulate a deep neural network as $\mathcal{F}_\theta(x)$, where θ is the trainable parameters and x is the input image. The DNN outputs a feature representation $h_x \in \mathbb{R}^d$ for input x which is then used for classification in a multiclass classifier $Z = [z_k] \in \mathbb{R}^{d\times K}$, where $k = 1, \cdots, K$. To train the model we minimize an objective function to minimize θ and Z.

By maximizing the similarity between intermediate representations and minimizing redundancy, a so-called variance loss function was introduced to enforce local compactness between images and their adversarial counterparts. This effectively removes unnecessary information from the input data by decorrelating the clean image and its adversarial counterpart and ensuring all variables have similar variances. Consequently, both the clean image and its adversarial counterpart retain minimal but sufficient representations for accurate classification.

To further enhance the local compactness of features, a convolutional generator network G_Ψ is employed to map the intermediate layer of the discriminator network into a new feature space with reduced redundancy. The G_Ψ mapping is learned in an end-to-end manner by minimizing the cross-correlation between the clean and adversarial features while maintaining their individual variances. A square matrix Q is

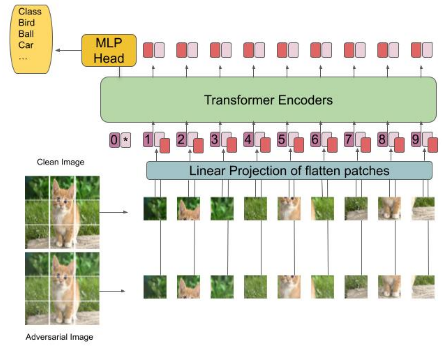

Fig. 1. The ViT architecture modified to learn jointly from $\mathcal{L}_{CE}$ and $\mathcal{L}_{var}$.

computed between the clean and its corresponding adversarial image in terms of G_Ψ:

$$Q = \frac{G_\Psi(h_x) \times G_\Psi(h'_x)}{\sqrt{G_\Psi(h_x)}\,\sqrt{G_\Psi(h'_x)}}. \tag{1}$$

Here, $Q \in \mathbb{R}^{d\times d}$, where d is the dimension of the output of G_Ψ. Intuitively, the diagonal elements of matrix Q represent the dot-product between clean and corresponding adversarial images while off-diagonal terms represent the covariance between them.

Minimizing this objective correlates clean and adversary counterparts and encourages to have non-redundant information while being closer in intermediate layers. This allows the centers to have contrasting representations and promotes maximum separation between classes. The variance loss function is then defined as

$$\mathcal{L}_{var} = \sum_i (1 - Q_{ii})^2 + \lambda \sum_i \sum_{j\neq i} Q_{ij}^2 \tag{2}$$

where λ is the trade-off parameter of the invariant (diagonal) and the redundancy (off-diagonal) terms of the matrix.

The overall prediction accuracy of an input-label pair (x, y) is ensured by the commonly used softmax Cross-Entropy (CE) loss $\mathcal{L}_{CE}(x, y)$. A third loss function is the so-called attract-repulsive loss function, which is not applicable to the ViT architecture.

The loss function that encourages the same classes to be mapped closer and different classes to be mapped farther from each other by a large margin is developed as follows. The modified AICR loss function is used in the training of a ViT as illustrated in Figure 1. The AICR loss function is modified for the ViT architecture by combining the cross entropy loss function and variance loss function for an input-label pair (x, y) and the corresponding adversarial image x' as:

$$\mathcal{L}(x, x', y) = \sum_i^n (\mathcal{L}_{CE}(x_i, x'_i, y_i) + \alpha \times \mathcal{L}_{var}(h_i^l, h_i'^l, y_i)) \tag{3}$$

where

$$h^l = \mathcal{G}^l_\phi(\mathcal{F}^l_\theta(x)) \quad \text{and} \quad h'^l = \mathcal{G}^l_\phi(\mathcal{F}^l_\theta(x')),$$

and α is the regularizing term for the contrastive centroid loss, $\mathcal{G}_\phi$ is the auxiliary function that maps intermediate layers to a lower dimension output, and n denotes the number of layers.

IV. Experimental Results

We evaluated the variance loss [18] with the ViT [20] architecture. We tested two different variants of ViT network that incorporate variance loss in the final representations. One of them incorporates $\mathcal{L}_{var}$ in the final classification head, identified as *ViT-C*; and the other variant uses $\mathcal{L}_{var}$ on representations of the final patches except the classification head, and it is identified as *ViT-All*.

We illustrate attention shift due to adversarial attacks when a network is trained solely by the CE loss and when a network is trained by the AICR loss as follows. The Grad-CAM [29] helps us understand why a model predicts a certain class by using an attention map to highlight the areas in the image that most influence its decision. In Figure 2, we show the attention maps of some sample images that have been trained using $\mathcal{L}_{CE}$, the cross entropy loss. In Figure 3, we show the corresponding attention maps of the adversarial images. The shifts due to the adversarial attack are evident by comparing the two images.

The reason attention shifts are markers of suscpetibility to adversarial attacks is as follows. Many adversarial attacks work by subtly manipulating an image in a way that causes the deep learning model to shift its attention to irrelevant or misleading parts of the image. This attention shift can lead to misclassifications. Deep learning networks that are robust to adversarial attacks tend to exhibit smaller or less significant shifts in their attention when presented with adversarial examples. This suggests that a model maintaining its focus on the correct features is less likely to be fooled. If a deep learning network keeps its attention on the right parts of the image even when attacked, it has a higher chance of correctly identifying the object despite the adversary's attempts to mislead it.

In Figure 4, we show the attention maps of some sample images that have been trained using $\mathcal{L}_{AICR}$, the loss function that combines cross entropy and variance loss. In Figure 5, we show the corresponding attention maps of the adversarial images. The lack of shift due to the adversarial attack are evident by comparing the two images.

In an adversarial setting, there are two main threat models. In *white-box* attacks, the adverser has complete knowledge of the target model including model architecture and objective function used for training and parameters. *Black-box* attacks, on the other hand, feed adversarial noise to the input images during inference time, and it is crafted without any knowledge of target model. Following the attack settings in [30], we crafted adversarial examples in a non-targeted way with respect to allowed perturbation ϵ for gradient based attacks, i.e., FGSM, BIM, PGD, MIM. The number of iterations for BIM, MIM, PGD were set to 10 with a step size of $\epsilon/10$. We

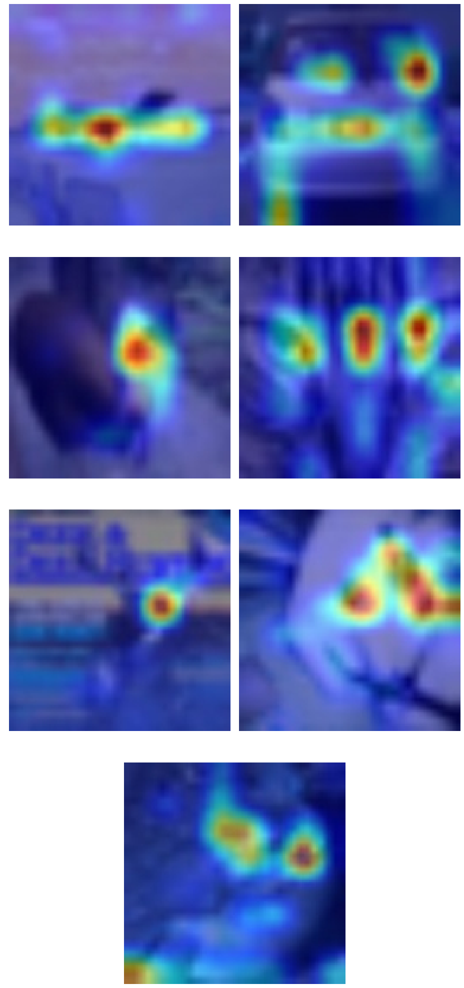

Fig. 2. Attention maps by Grad-CAM of clean CIFAR-10 images; model trained by CE loss.

compare the accuracy of ViT network with $\mathcal{L}_{var}$ at different settings of ViT on the CIFAR 10 data set for white box attacks. Results in Table I show that using variance loss increased the robustness of the model compared to the model that does not use variance loss.

TABLE I
Accuracy of vision transformers under adversarial attacks

Attacks	ϵ	ViT	*ViT-C*	*ViT-All*
No-attack	-	**80.1**	78.9	79.6
FGSM	0.1	15.2	15.8	**16.2**
	0.2	2.7	1.8	**3.6**
PGD	0.1	8.5	**9.9**	9.2
	0.2	0.15	**0.33**	0.16
BIM	0.1	8.4	**9.9**	9.1
	0.2	0.15	**0.33**	0.16
MIM	0.1	8.8	**10.3**	9.6
	0.2	0.17	**0.37**	0.22

Here *ViT-C* refers to the model trained with $\mathcal{L}_{var}$ only on the classification head and *ViT-All* refers to the model with patch representations trained optimized with $\mathcal{L}_{var}$. All models are trained jointly with adversarial samples and clean samples.

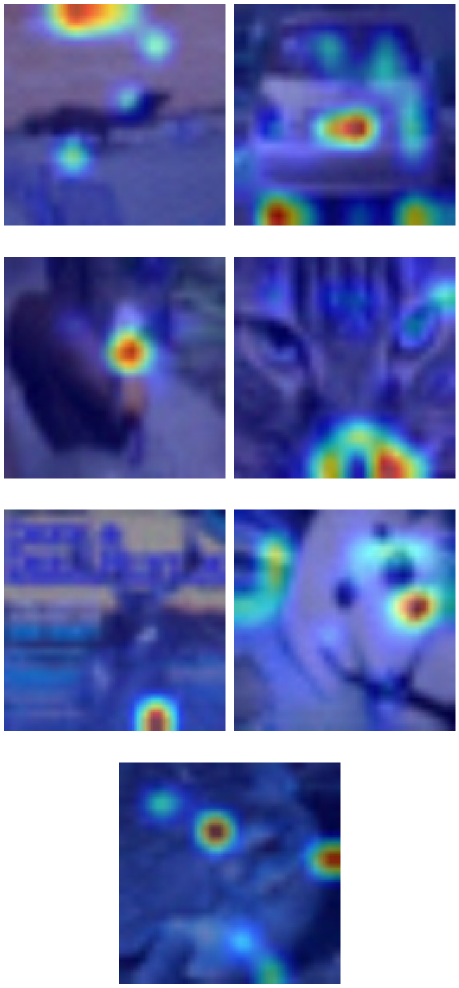

Fig. 3. Illustration of attention shift using Grad-CAM on model trained by CE loss. Adversarial images are obtained crafted with PGD (ϵ = 0.03); see clean images in Figure 2 for comparison.

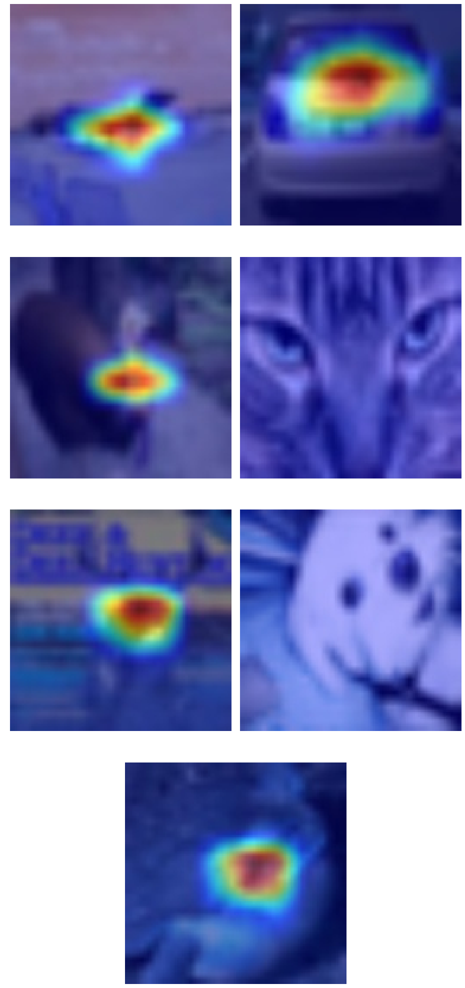

Fig. 4. Attention maps by Grad-CAM of clean CIFAR-10 images; model trained by AICR loss.

We can see that there were performance drops by *ViT-C* and *ViT-All* when there was no attack. At all levels of attacks, the ViT trained with the modified AICR loss performed better than the ViT trained with only cross entropy loss. The *ViT-All* network performed better under the FGSM attack while the *ViT-C* network performed better the PGD, BIM, and MIM attacks. Under the FGSM attacks, *ViT-All* improved by 6.57% to 33.33% for $\epsilon = 0.1$ and 0.2, respectively. Under the PGD, BIM, and MIM attacks, *ViT-C* improved by about 17% to 120% for $\epsilon = 0.1$ and 0.2, respectively.

V. Conclusion and Future Work

The AICR loss function has previously shown a significant improvement in the robustness of CNNs against adversarial attacks, particularly in such tasks as image classification. Motivated by ViT's superior performance in such tasks, we adapted the AICR loss and investigated its effectiveness in training ViTs against gradient-based attacks such as PGD and BIM. Our experiments revealed negligible changes in the attention distribution of ViTs trained with modified AICR loss compared to cross-entropy, indicating stable attention patterns. Furthermore, ViTs trained with AICR loss achieved a 33% to 120% improvement in accuracy compared to cross-entropy, demonstrating its effectiveness in enhancing ViT's resilience against adversarial attacks.

A promising direction of future work might focus on exploiting attention shift as a marker of adversarial vulnerability. Encouraging models to maintain consistent attention patterns between clean and adversarial examples during training is a promising defense strategy, aimed at improving robustness against attention-based attacks. Networks can be explicitly trained with adversarial examples that are designed to shift attention. This helps the model learn to recognize these tricks and maintain stability. A longer term goal is to develop more reliable connections between attention and adversarial robustness.

Acknowledgments

The authors thank the anonymous reviewers whose suggestions helped to improve the clarity of our paper. This work is supported in part by the U.S. National Science Foun-

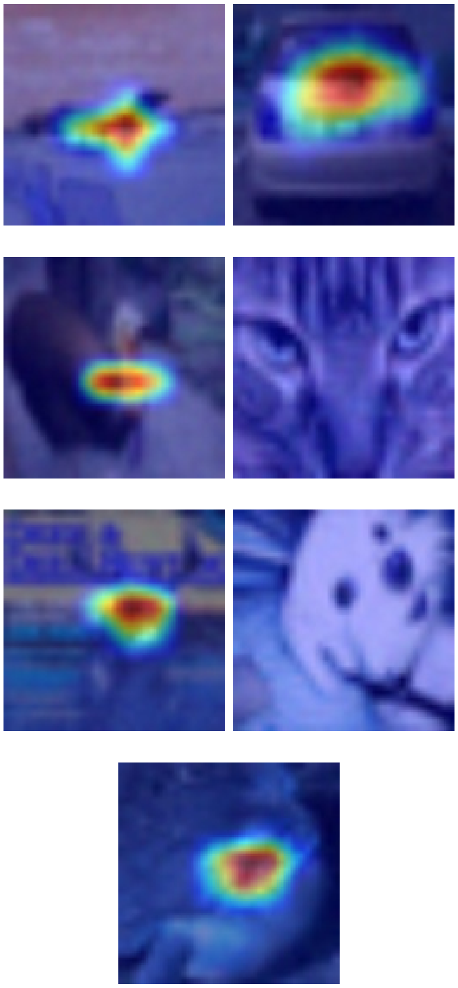

Fig. 5. Illustration of attention shift using Grad-CAM on model trained by AICR loss. Adversarial images are obtained crafted with PGD (ϵ = 0.03); compare with clean images in Figure 4.

dation under grant number OIA-1946231 and the Louisiana Board of Regents for the Louisiana Materials Design Alliance (LAMDA).

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026–1034, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[6] W. Wang and J. Gang, "Application of convolutional neural network in natural language processing," in *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, pp. 64–70, 2018.

[7] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized Bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[8] L. Deng and X. Li, "Machine learning paradigms for speech recognition: An overview," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 1060–1089, 2013.

[9] O. Abdel-Hamid, A.-r. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533–1545, 2014.

[10] A. Gulati, J. Qin, C.-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, *et al.*, "Conformer: Convolution-augmented transformer for speech recognition," *arXiv preprint arXiv:2005.08100*, 2020.

[11] L. Buşoniu, R. Babuška, and B. De Schutter, "Multi-agent reinforcement learning: An overview," *Innovations in multi-agent systems and applications-1*, pp. 183–221, 2010.

[12] H. Van Hasselt, "Reinforcement learning in continuous state and action spaces," in *Reinforcement learning*, pp. 207–251, Springer, 2012.

[13] "A cnn-based policy for optimizing continuous action control by learning state sequences," *Neurocomputing*, vol. 468, pp. 286–295, 2022.

[14] N. Das, M. Shanbhogue, S.-T. Chen, F. Hohman, L. Chen, M. E. Kounavis, and D. H. Chau, "Keeping the bad guys out: Protecting and vaccinating deep learning with jpeg compression," *arXiv preprint arXiv:1705.02900*, 2017.

[15] Y. Luo, X. Boix, G. Roig, T. Poggio, and Q. Zhao, "Foveation-based mechanisms alleviate adversarial examples," *arXiv preprint arXiv:1511.06292*, 2015.

[16] A. Mustafa, S. H. Khan, M. Hayat, J. Shen, and L. Shao, "Image super-resolution as a defense against adversarial attacks," *IEEE Transactions on Image Processing*, vol. 29, pp. 1711–1724, 2019.

[17] C. Xie, J. Wang, Z. Zhang, Z. Ren, and A. Yuille, "Mitigating adversarial effects through randomization," *arXiv preprint arXiv:1711.01991*, 2017.

[18] J. Alagurajah and C.-H. H. Chu, "Adversarial defense by restricting invariance and co-variance of representations," in *The 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22)*, 2022.

[19] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *European conference on computer vision*, pp. 499–515, Springer, 2016.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[21] C. Guo, M. Rana, M. Cisse, and L. Van Der Maaten, "Countering adversarial images using input transformations," *arXiv preprint arXiv:1711.00117*, 2017.

[22] T. Miyato, S.-i. Maeda, M. Koyama, K. Nakae, and S. Ishii, "Distributional smoothing with virtual adversarial training," *arXiv preprint arXiv:1507.00677*, 2015.

[23] F. Yu, C. Liu, Y. Wang, L. Zhao, and X. Chen, "Interpreting adversarial robustness: A view from decision surface in input space," *arXiv preprint arXiv:1810.00144*, 2018.

[24] N. Akhtar, A. Mian, N. Kardan, and M. Shah, "Advances in adversarial attacks and defenses in computer vision: A survey," *IEEE Access*, 2021.

[25] A. Aldahdooh, W. Hamidouche, and O. Deforges, "Reveal of vision transformers robustness against adversarial attacks," *arXiv preprint arXiv:2106.03734*, 2021.

[26] Y. Mo, D. Wu, Y. Wang, Y. Guo, and Y. Wang, "When adversarial training meets vision transformers: Recipes from training to architecture," *arXiv preprint arXiv:2210.07540*, 2022.

[27] E. Debenedetti, V. Sehwag, and P. Mittal, "A light recipe to train robust vision transformers," *arXiv preprint arXiv:2209.07399*, 2023.

[28] B. Wu, J. Gu, Z. Li, D. Cai, X. He, and W. Liu, "Towards efficient adversarial training on vision transformers," in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*, (Berlin, Heidelberg), p. 307–325, Springer-Verlag, 2022.

[29] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[30] H.-Y. Chen, J.-H. Liang, S.-C. Chang, J.-Y. Pan, Y.-T. Chen, W. Wei, and D.-C. Juan, "Improving adversarial robustness via guided complement entropy," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4881–4889, 2019.

Development of Children's Crossing Skills in Urban Area: Impact of Age and Traffic Density on Visual Exploration

Jordan Solt
University of Lorraine, 2LPN
Luxembourg-ville
Email: solt.jordan@gmail.com

Aurelie Mailloux
Reims hospital, URCA
Reims, France
Email: aurelie.mailloux@univ-reims.fr

Jerome Dinet
2LPN and Chair BEHAVIOUR
Nancy, France
Email: jerome.dinet@univ-lorraine.fr

Samuel Ferreira Da Silva
Lorraine University and Chair BEHAVIOUR
Metz, France
Email: samuel.ferreira-da-silva@univ-lorraine.fr

Muneo Kitajima
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

Gaelle Nicolas
Lorraine University and Chair BEHAVIOUR
Metz, France
Email: gaelle.nicolas@univ-lorraine.fr

***Abstract*—Pedestrian trauma represents a significant proportion of all road traumas, young pedestrian being over-represented in all these road traumas. From a cognitive point of view, road crossing ability is a high and complex mental activity because the individual has to process dynamic and complex information extracted from his/her surrounding environment, to make a decision (i.e., where and how to cross), and safe pedestrians must possess and utilize advanced cognitive skills. More precisely, there are two major problems for young pedestrians to make the decision about when and where it is safe to cross the street : gap selection and assessment of inter-vehicular gap. An experimental study conducted with forty children aged 3-10 years and twenty-two adults has been conducted to investigate the impact of one individual factor (Age) and one environmental factor (Traffic density) on decision making (i.e., "to cross" or "not to cross a street"), time spent to make decision (in milliseconds) and on visual exploration using eye-tracking techniques of urban scenes displayed on a computerized screen. Main results showed that (i) Traffic density has a significant impact on performance and visual exploration, (ii) Age has a significant impact on time spent to make decision and visual exploration and (iii) there is an interaction between Age and Traffic density.**

***Keywords*—Child; Pedestrian; Visual exploration; Risk; Hazard; Eye-Tracking**

I. INTRODUCTION

Pedestrian trauma represents a significant proportion of all road traumas, young pedestrian being over-represented in all these road traumas. In particular, the safety of child pedestrians is of concern, given that a sizable proportion of pedestrians killed and seriously injured involve children and the special value society places on its youth [1][2].

In section 1, context related to accidents with young pedestrians and factors influencing children's crossing skills are presented. In section 2, method of the experiment conducted with our participants are described. Finally, in section 3, theoretical and methodological implications related to the changes in visual strategy occurring around the age of 7-8 years are discussed.

A. Context

Around the world, the number of pedestrians killed increase. Young pedestrians are particularly concerned by these accidents: According to the official data issued from the Traffic Safety Facts, on average, three children were killed and an estimated 502 children were injured every day in the U.S. in traffic crashes. In 2019 and 2020, there were respectively 181 and 177 children killed in pedestrian accidents. Most were toddlers (between the ages of 1-3) and young children (4-7). In fact, an estimated 1 in 5 children killed in car accidents were pedestrians, i.e., just walking on the sidewalk or crossing a street whatever the country [3][4].

At ages 6-10 years, children are at highest risk of pedestrian collision, most likely due to the beginning of independent unsupervised travel at a time when their road strategies, skills and understanding are not yet fully developed. Whatever the country, research suggests that children between the ages of 6 to 10 are at highest risk of death and injury, with an estimated minimum four times the risk of collision compared to adult pedestrians [5]. Until the age of 6-7 years, children are under active adult supervision, i.e., parents hold their child's hand when crossing roads together. Even if every year many pedestrians are injured or killed in traffic accidents in rural parts of the country [6], pedestrian safety is being considered as a serious traffic safety problem in urban and suburban settings [7][8]. Thus, children more than adults, are at risk as pedestrians, often due to their own actions and behaviors. So the question is: "Why do young pedestrians not adopt safety behaviors specially during street crossing?"

 ISBN: 978-1-68558-157-2

B. Factors Influencing Children's Crossing Skills and Gap Selection

From a cognitive point of view, road crossing ability is a high and complex mental activity because the individual has to process dynamic and complex information from his/her surrounding environment, to make a decision (i.e., where and how to cross). Safe pedestrians must possess and utilize advanced cognitive skills [9][10]. Crossing decisions include whether or not to enter the roadway, the place to cross, the path to take, how fast to travel, and how the driver might react. A sound decision on whether to enter the roadway should be based upon recall (experience) and monitoring of the traffic detected, including the distance, speed, and anticipated direction of vehicles and the opportunities provided by various gaps in traffic [11]. The time that has elapsed while making the decision also needs to be incorporated. Successful crossing performance also requires reliable estimation of the pedestrian's walking speed, peak capabilities, and distance to the other side of the road or a traffic island. Integrating all these aspects is difficult for the child, especially one inexperienced in traffic, and result in a longer decision making time: In fact, a 5 year old child requires about twice as long to reach a pedestrian decision as an adult.'and This leaves even less time to execute an imperfectly planned crossing [9][10][12].

A vast amount of research suggests that children's development of cognitive skills is significantly related to increased pedestrian safety and that relevant skills improve as children get older [13][14][15]. Of course, it is not a single cognitive skill that influences safety. Instead, it is the combined development of a number of different cognitive processes that are linked to safe pedestrian behavior. Those processes also overlap with other developing skills, such as perceptual (visual and auditory essentially) and motor abilities.

As children develop, specific pedestrian injury risks change [12][15][16][17][18][19]. More precisely, toddlers (ages 1–2) are most likely to be injured in driveways, where drivers moving backward are unable to see them [20], while adolescents are at risk due to walking at night with poor visibility, walking while intoxicated or walking while distracted by phones [21]. Our paper focuses on children between those two phases, in ages 6 through 12. During this stage of development, most pedestrian injuries occur in mid-block areas, where children enter into the middle of the street and are struck by moving vehicles, or at intersections [22]. As Schwebel and his colleagues said, if some incidents are "dart-out" situations where children enter the street quickly, without thought (i.e., to chase a person, toy, or pet, or to meet someone on the other side of the street), the majority of the incidents/collisions are the result of poor judgment by the child, i.e., s/he believes it to be safe, and enters the street when in fact the situation is not safe [19].

Several studies showed that gap selection and assessment of inter-vehicular gap by young pedestrians are two major problems for young pedestrians to make the decision about when and where it is safe to cross the road [23][24][25]. Inter-vehicular gap is both temporal and spatial because these two parameters are crucial to make the decision in relation to available gaps in the traffic [26]. More precisely, judgement of whether a gap in the traffic is sufficient to safely cross requires the determination of the time gap of the nearest vehicle with the planned crossing line and the assessment of whether this time gap exceeds the time required to cross the road. So, children aged below 10 years have relatively poor skills at reliably setting safe distance gap thresholds, and thus do not consistently make safe crossing decisions [27][28][29][30][31][32].

But, very few authors concentrated on visual exploration of young pedestrians during crossing activity. For instance, Whitebread and her colleagues examined the relationships between pedestrian skills and visual search strategies for young pedestrians [33]. According to their findings, major changes in strategy occurred around the age of 7-8 years. This change expressed in the frequency and pattern of looking at different directions, having a sophisticated 'last-minute' checking approach, exhaustive visual search strategy, and the speed of making the crossing decision. In the same way, Tapiro and her colleagues examined children's visual search strategies in hazardous road-crossing situations [29]. A sample of 33 young participants (ages 7-13) and 21 adults observed 18 different road-crossing scenarios in a 180 degrees dome shaped mixed reality simulator. Gaze data was collected while participants made the crossing decisions. Their results showed that age group, limited field of view, and the presence of moving vehicles affect significantly the way pedestrians allocate their attention in the scene. Therefore, the authors deduce that adults tend to spend relatively more time in further peripheral areas of interest than younger pedestrians do. It was also found that the oldest child age group (11-13 years old) demonstrated more resemblance to the adults in their visual scanning strategy, which can indicate a learning process that originates from gaining experience and maturation. Nevertheless, all participants in these previous studies are 7 years old and above. In this experiment, we collect data with eye-tracking from younger pedestrians (3 to 10 years old) to better understand the visual exploration of urban scenes.

II. METHOD

This experimental study conducted with forty children aged 3-10 years and twenty-two adults is aiming to investigate the impact of one individual factor (Age) and one environmental factor (Traffic density) on decision making (i.e., "to cross" or "not to cross a street"), time spent to make decision (in milliseconds) and on visual exploration of urban scenes displayed on a computerized screen. Eye-tracking technique is used to collect precise data about gaze exploration of each participant.

A. Participants

Sixty-two French participants were recruited to participate in this study. Children are issued from four different age groups: Seven pupils are from Grade 1 (boys, 100 percent; mean age = 3.86 years; SD = 0.37 years), nineteen pupils are recruited from Grade 3 (boys, 56.8 percent; mean age = 6.89 years; SD = 0.31 years), fifteen pupils are recruited from Grade 5 (boys, 60 percent; mean age = 9.87 years; SD = 0.51 years), and twenty-one participants are adults (men, 47.6 percent; mean age = 26.71 years; SD = 8.22 years). All children are issued in the same elementary school located in the mid-town.

All participants are French native speakers and the majority (82.1 percents) lives in urban area. Moreover, even if the majority of adult participants (81 percents) have their driving license, they admit to go to work essentially by using public transportation (61.9 percents) or by walk (38.1 percents). All the children are recruited in the same primary school located in the mid-town. All parents agreed to their children participate. No participant has severe visual impairment and no cognitive impairment. There is no difference between groups according to the visual memory and attention capacities (Table 1).

B. Independent and Dependent Variables

In our study, we investigated the impact of one individual factor (Age) and one environmental factor (Traffic density) on three behavioural indicators:

- The decision (i.e., "to cross" versus "not to cross the street");
- The time spent in milliseconds to make this decision;
- The visual exploration of specific Areas of Interest (AoI) of urban scenes displayed on pictures (Figure 1);

Thus, two independent factors were manipulated, the first one being intra-subject ("Age", with four modalities: Grade 1, Grade 3, Grade 5, and adults) and the second one being inter-subject ("Traffic density", with three modalities: Low, Moderate, and High). In other words, our experimental plan was: Participant $<$ Age 4 $>$ * Traffic density 3

C. Material

Assessment of Cognitive Abilities. Each participant was asked to complete several sub-scales extracted from the Wechsler scales to assess their cognitive abilities. For the youngest participants (Grade 1), "Coding scale" and "Digit span scale" extracted from the WPPSI-V have been used. For the two other groups of children (Grade 3 and Grade 5), they are the same sub-tests used but extracted from the WISC-V. For adults, four sub-scales extracted from the WAIS-V have been used: "Digit span scale", "Arithmetic scale", "Coding scale", and "Symbol scale". All these sub-scales were chosen because they are very sensitive to the visual memory and attention capacities.

Urban Scenes. Each participant was individually asked to examine different urban scenes displayed on a computerized screen before to make a decision for each urban scene, i.e., "to cross" or "not to cross the street". Three traffic densities have been used to investigate the impact of this factor on decision-making and visual exploration: Low, Moderate, and High. Figure 1 shows an example for each of these modalities. For each of the traffic density (Low, Moderate, and High), four different urban scenes. These urban scenes were chosen by four judges after they evaluated and categorized a lot of pictures in these three traffic conditions: Low traffic density ("Low"; e.g., one other pedestrian and two vehicles far), moderate traffic density ("Moderate"; e.g., several other pedestrians and different kinds of vehicles), and high traffic density ("High"; e.g., a lot of vehicles near and far).

Each participant was asked to examine 12 different static pictures of urban scenes, the order of presentation being counterbalanced to avoid order effect on responses (i.e., "to cross" or "not to cross the street"). On-line eye-tracking data for each participant were collected during participants examined urban scenes, by using the eye-tracking techniques. The Tobii T120, with a 17 inch monitor integrated, was used to collect visual exploration of urban scenes by our participants.

D. Procedure

The procedure has four distinct and successive steps:

- Training session. First, each participant was invited to seat in front of a computer (Tobii T120, with a 17 inch monitor integrated) and the same instructions are given: (a) different images will appear on the screen, one by one; (b) s/he must to analyse the urban scenes carefully because s/he was asked to decide if s/he crosses or not the street; (c) when s/he made the decision, s/he was asked to say "stop" and s/he can give his/her decision orally. Different pictures (not used in the following experiment) are used during a training session;
- Experimental session for visual exploration and decision-making. If the participant has no problem with the procedure and has no question, the experiment can begin with the urban scenes related to the three conditions (Low, Moderate, High);
- Assessment of cognitive abilities. Just after the end of the experimental session, each participant was asked to complete sub-scales extracted from the Wechsler scales to assess their visual memory and attention capacities;
- Length of time the subject is expected to participate
- Researchers ensured that those participating in research will not be caused distress;
- End of the experiment. Finally, each participant was asked to complete a survey to provide some several demographic information, is thanked and each child receives a packet of sweets.

Note that for children, the experiment was always conducted in the same quiet room located in the school, dedicated to the experiment. The experimenter was always the same.

E. Design and Data Analysis

First, we examined the impact of our two independent variables ("Group age" and "Traffic condition") on the one hand, decision (i.e., "I am crossing" or "I am not crossing"), and on the other hand, time spent to make this decision (in milliseconds). So the design of this first part of analyses is the following factorial design: Group age (4) (Grade 1, Grade 3, Grade 5, Adult) X Traffic density (3) (Low, Moderate, High), with "Age group" as between-subjects factor and "Traffic density" as within-subjects factor.

 ISBN: 978-1-68558-157-2

TABLE I
MEAN (AND STANDARD DEVIATION) OF FIXATION DURATION FOR EACH AGE GROUP, TRAFFIC CONDITION FOR EACH AREAS OF INTERESTS (AOI)

	Mean of Memory Span (SD)	Mean of Processing Speed (SD)
Grade 1 ($n = 7$)	8.5 (2.7)	9.5 (3.4)
Grade 3 ($n = 19$)	10.5 (4.5)	12.8 (5.1)
Grade 5 ($n = 15$)	10.9 (2.9)	10.2 (3.9)
Adult ($n = 21$)	9.1 (1.9)	10.6 (2.8)

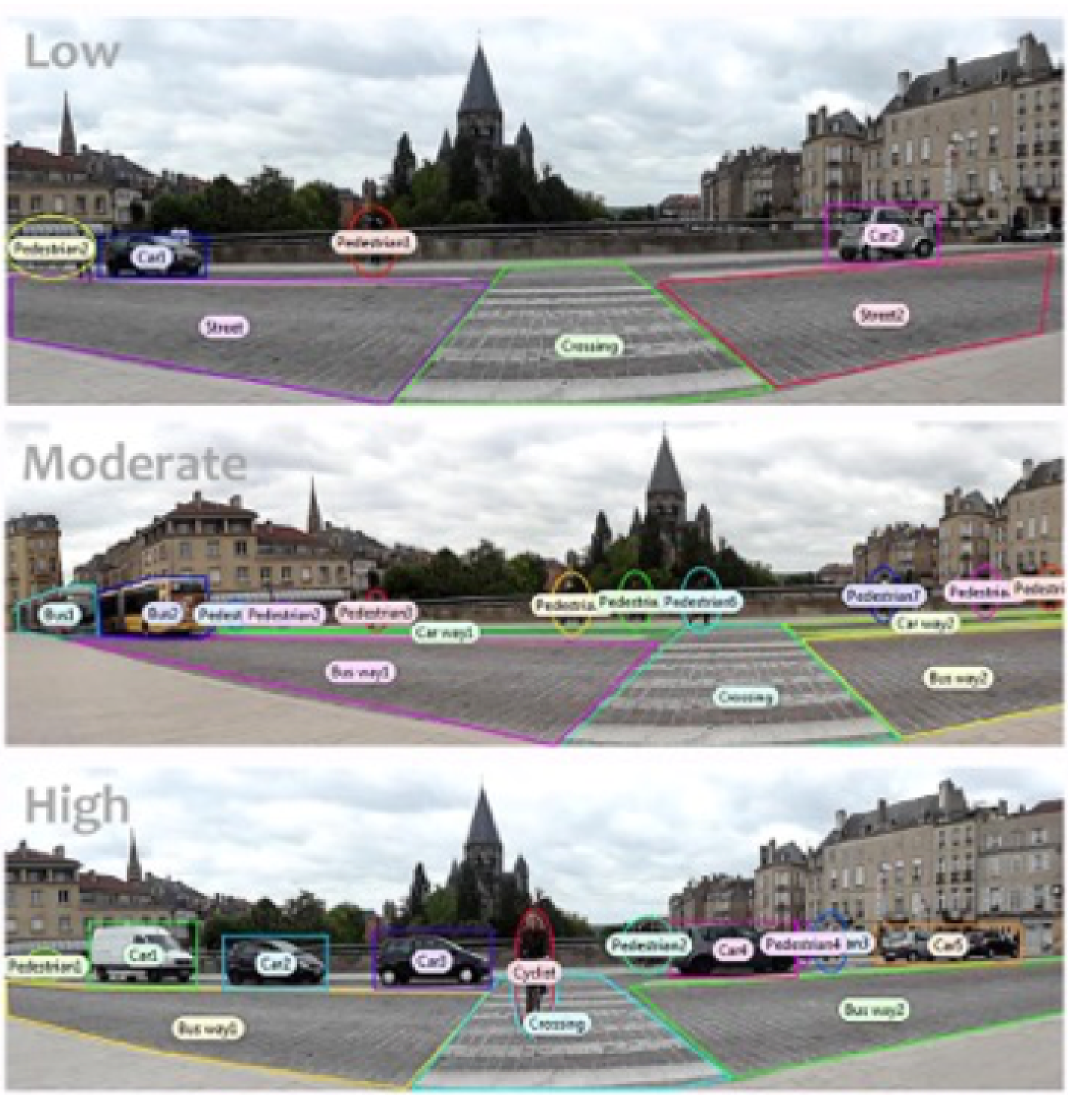

Figure 1. The different Areas of Interest (AoI) in the three Traffic density conditions

Second, we examined the visual exploration on specific Areas of Interest (AoI) predefined for urban scenes (i.e., "Pedestrians", "Sidewalk", "Car", "Car way", Bus", "Bus way", and "Crossing"). Figure 1 shows these six different AoI. The design of this second part of analyses is the following factorial design: Age group (4) (Grade 1, Grade 3, Grade 5, Adult) X Traffic density (3) (Low, Moderate, High) X AoI (6) (Pedestrians, Sidewalk, Car, Car way, Bus", Bus way, Crossing) with "Age group" as between-subjects factor and "Traffic density" and "AoI" as within-subjects factors.

F. Ethics

All adults' participants provided written informed consent for their participation in this study, and all legal parents of children provided the same informed consent. Moreover, the responsible of the school provided also her consent. Before providing the written consent, all adults' participants, legal parents of children and the director of the school where the research has been conducted received the same information relating to the following points:

- A statement that participation is voluntary and that refusal to participate will not result in any consequences or any loss of benefits that the person is otherwise entitled to receive;
- The precise purpose of the research;
- The procedure and material involved in the research;
- Benefits of the research to society and possibly to the individual human subject;
- Length of time the subject is expected to participate
- Researchers ensured that those participating in research will not be caused distress;
- Subjects' right to confidentiality and the right to withdraw from the study at any time without any consequences;
- After the research is over, each participant (adults or children) are able to discuss the procedure and the findings with the psychologist.

III. RESULTS

The experiment based on eye-tracking techniques aimed to investigate the impact of one individual factor (Age) and one environmental factor (Traffic density) on three behavioural indicators related to competencies of very young pedestrians (aged 3-10 years). Several interesting results have been obtained.

A. Impact of Age group and Traffic Density on Decision-Making

The decision made by each participant ("I cross" versus "I do not cross") in front of each urban scene has been collected (Table 2). For each if the three Traffic density conditions (Low, Moderate, High), statistical analyses revealed only one significant impact of Age group in high traffic condition (F(3-58) = 2,858, p = .045).

B. Impact of Age group and Traffic Density on Decision-Making Time

As Table 3 shows, the time spent to make decision decreased with age. Statistical analyses confirmed that Age group had a significant impact on this time spent to make decision (F(3-58) = 8,75, p < .001). Time spent to make decision for the youngest participants (Grade 1, Mean = 8829,68) was superior than time spent by all the other participants (Grade 3, M = 5240,98, F(3-58) = 2,934, p = .005; Grade 5, Mean = 4694,68, F(3-58) = 3,265, p < .005; Adults, Mean = 2797,82, F(3-58) = 4,996, p < .001). In the same way, time spent to make decision for participants aged to 6-7 years (Grade 3, Mean = 5240,98) was superior than time spent for adults (Mean = 2797,82), the difference being significant (F(3-58) = 2,789, p = .007). Finally, time spent to make decision for participants in grade 5 (Mean = 4694,68) was superior than time spent adults (Mean = 2797,82), the difference being significant (F(3-58) = 2,028, p = .047) Traffic condition had also a significant impact on time spent to make decision (F(2-116) = 7,67, p = .001). As Table 2 shows, time spent to make decision in low traffic condition (Mean = 4311,31) was inferior than time spent in high traffic condition (Mean = 5278,16), the difference being significant (F(2-116) = 7,67, p = .002). Finally, there was a interaction between Age group and Traffic condition (F(6-116) = 2,73, p = .016) on the time spent to make a decision.

C. Impact of Age Group and Traffic Density on Global Visual Exploration

There was a global impact of Age group on total fixation duration (Table 4; F(3-58) = 8,475, p < .001). Specifically, Age group had a significant impact for Low traffic density (F(3-56) = 2,980, p = .039) and Moderate traffic density (F(3-56) = 9,422, p = .001) but had no impact on High traffic density (F(3-50) = 2,695, p < .056):

- For Low traffic density condition, mean fixation duration for children recruited in Grade 1 was higher compared to Adults (respectively, Mean = 0.3614 and Mean = 0.2665; t(56) = 2,183, p = .033). In the same way, children recruited in Grade 3 have more longer fixation duration compared to Adults (respectively, Mean = 0.3514 and Mean = 0.2665; t(56) = 2,639, p = .011). Adults had the fastest fixings but that was significant only that in comparison with Grade 1 and Grade 3;
- For Moderate traffic density condition, adults (M = 0.29) had shorter fixation duration compared to Grade 1 (respectively, Mean = 0.29 and Mean = 0.3692; t(56) = 2,293, p = .026) and compared to Grade 3 (Mean = 0.3579; t(56) = 2,656, p = .01). Children issued from Grade 5 spent significantly less time to make decision than Grade 1 (Mean = 0.3692; t(56) = 3,950, p = .000), compared to Grade 3 (Mean = 0.3579) (t(56) = 4,76, p = .000) and compared to adults (Mean = 0.29) (t(56) = -2,345, p = .023);
- For High traffic density condition, only one Age group was concerned by significant differences: Children issued

 ISBN: 978-1-68558-157-2

TABLE II
NUMBER (AND PERCENTAGE) OF PEDESTRIANS CROSSING THE STREET FOR EACH AGE GROUP AND THE THREE TRAFFIC DENSITY CONDITIONS (LOW, MODERATE, HIGH)

	Low	Moderate	High
Grade 1 ($n = 7$)	3 (42.8)	2 (28.5)	1 (14.9)
Grade 2 ($n = 19$)	12 (63.6)	2 (10.5)	0 (-)
Grade 2 ($n = 15$)	10 (66.6)	3 (20)	3 (20)
Adult ($n = 21$)	9 (42.8)	3 (14.2)	5 (23.8)

TABLE III
MEAN (AND STANDARD DEVIATION) OF TIME SPENT TO MAKE DECISION (I.E., "TO CROSS" *versus* "NOT TO CROSS") FOR EACH AGE GROUP AND EACH TRAFFIC DENSITY CONDITION (LOW, MODERATE, HIGH)

	Low	Moderate	High	Mean (SD)
Grade 1 ($n = 7$)	7854 (6399)	7183 (5081)	11450 (10673)	8829 (6812)
Grade 3 ($n = 19$)	4921 (2292)	5337 (1803)	5464 (3140)	5240 (2030)
Grade 5 ($n = 15$)	3804 (1616)	4454 (2175)	5825 (3902)	4694 (1981)
Adult ($n = 21$)	2940 (1708)	2791 (1245)	2661 (1592)	2797 (1345)
Total mean (SD) (N = 62)	4311 (3066)	4469 (2672)	5278 (5025)	-

TABLE IV
MEAN (AND STANDARD DEVIATION) OF TOTAL FIXATION DURATION FOR EACH AGE GROUP AND THE THREE TRAFFIC DENSITY CONDITIONS (LOW, MODERATE, HIGH)

	Low	Moderate	High	Mean (SD)
Grade 1 ($n = 7$)	0.316 (0.08)	0.362 (0.06)	0.408 (0.124)	0.379 (0.09)
Grade 2 ($n = 19$)	0.351 (0.145)	0.357 (0.10)	0.321 (0.14)	0.343 (0.13)
Grade 5 ($n = 15$)	0.303 (0.07)	0.266 (0.06)	0.267 (0.12)	0.265 (0.08)
Adult ($n = 21$)	0.266 (0.05)	0.290 (0.04)	0.278 (0.04)	0.297 (0.06)
Total mean (SD) (N = 62)	0.320 (0.09)	0.311 (0.07)	0.318 (0.11)	-

from Grade 1 spent significantly more time to make decision compared to Grade 5, compared to Grade 1 (respectively, Mean = 0.4081 and Mean = 0.2673; t(50) = 2,521, p = .015) and compared to Adults (Mean = 0.05893; t (50) = 2,54, p = .014). In other words, in the high traffic density condition, children issued from Grade 1 were the slowest.

D. Impact of Age and Traffic Density on Examination for Each Area of Interest (AoI)

As Figure 2 shows, visual fixation duration time was significantly superior for two of the different Areas of Interest (AoI) predefined: the car way (F(3-43) = 4,191, p = .011) and the crossing (F(3-55) = 3,891, p = .014).

Moreover, Age group had a significant impact on distribution of fixation time only for these two of the different Ares of Interest (AoI) predefined. Fixation duration time on the car way was superior for Grade 1 compared to Grade 5 (respectively, Mean = 0,3625 and Mean = 0,2444; t(43) = 2,426, p = .02) and compared to Adults (Mean = 0,2311; t(43) = 2,626, p = .012). And fixation duration time on the car way was superior for Grade 3 compared to Grade 5 (respectively, Mean = 0,3291 and Mean = 0,2444; t(43) = 2,329, p = .025) and compared to Adults (Mean = 0,2311; t(43) = 2,569, p = .014). The pattern of results was identical for the crossing. Fixation duration time on the car way was superior for Grade 1 compared to Grade 5 (respectively, Mean = 0,3729 and Mean = 0,2713; t(55) = 2,3, p = .025) and compared to Adults (Mean = 0,2478; t(55) = 2,932, p = .005). And fixation duration time on the car way was superior for Grade 3 compared to Adults 5 (respectively, Mean = 0,3248 and Mean = 0,2444; t(55) = 2,425, p = .019).

Even if there were the only significant differences, some interesting tendencies can be remarked in the Figure 2 for other AoI such as "Pedestrians", "Cars" and "Bus way". For these three other AoI, fixation duration time for Adults group is always inferior.

There exist some significant interactions between Age group and Traffic density condition on these fixation duration means for the two main AoI ("Car way" and "Crossing"):

- For Low traffic density condition, fixation duration for younger participants (recruited in Grade 1) was significantly superior than fixation duration for children recruited in Grade 3 specially for "Car way" (respectively,

 ISBN: 978-1-68558-157-2

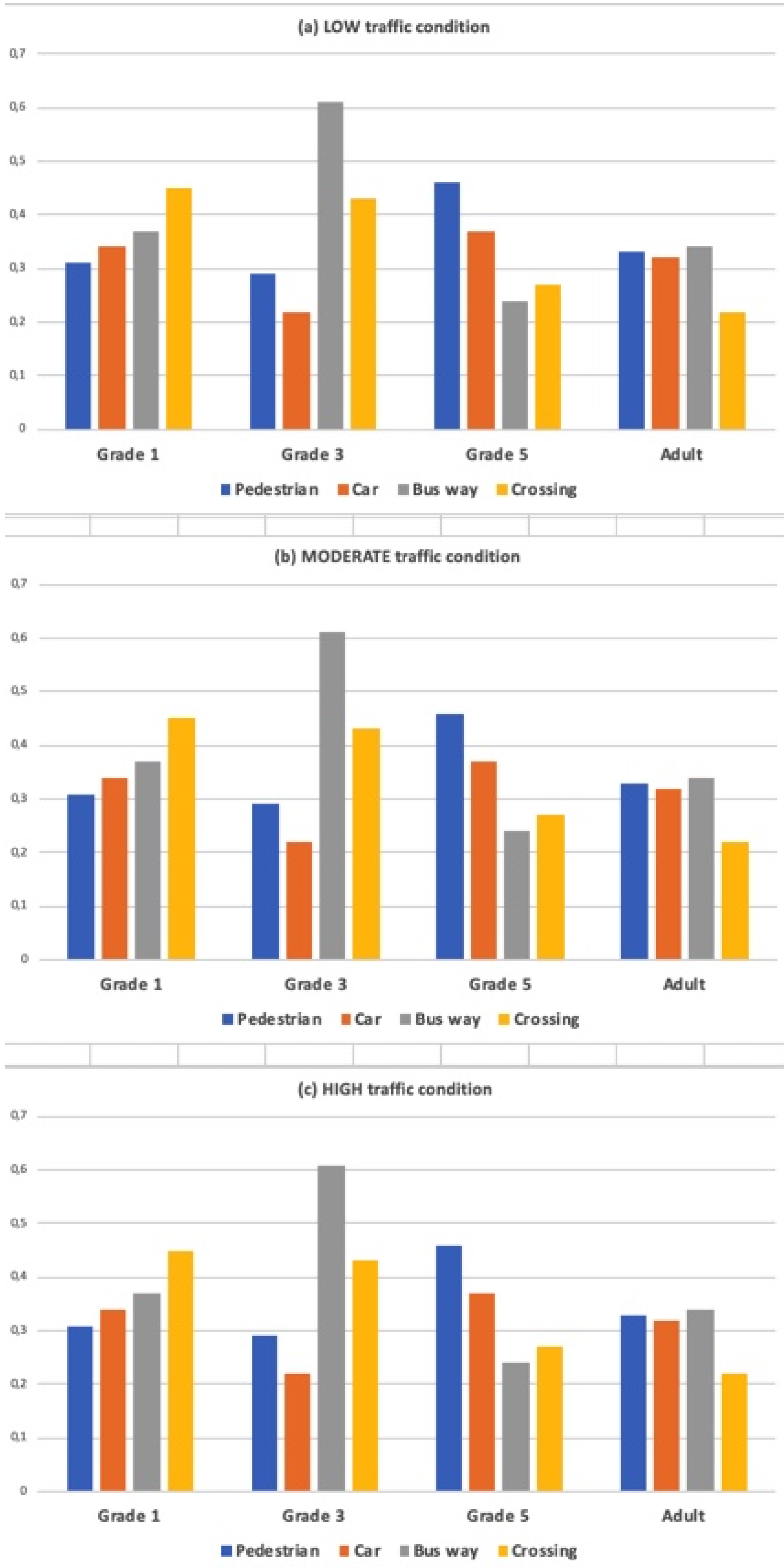

Figure 2. Mean of visual fixation duration for each Age Group (Grade 1, Grade 3, Grade 5, Adult), Traffic Condition (LOW / MODERATE / HIGH) for each Areas of Interests (AOI : Pedestrian / Car / Bus way / Crossing)

 ISBN: 978-1-68558-157-2

Mean = 0,45 and Mean = 0,263; t(54) = 2,023, p = .048);

- In the same way, for Moderate traffic density condition, fixation duration for younger participants (recruited in Grade 1) was also significantly superior than fixation duration for children recruited in Grade 3 (respectively, Mean = 0,768 and Mean = 0,438; t(55) = 3,218, p = .002);
- Finally, for High density traffic condition, fixation duration for children issued from Grade 1 was also superior than fixation duration specially for "cars" (respectively, Mean = 0,430 and Mean = 0,290; t(56) = 2,019, p = .048). The crossing site was extensively explored by the youngest participants (Grade 1, Mean = 0,442) compared to Adults (Mean = 0,229; (t(50) = 2,413, p = .020) and compared to children recruited in Grade 5 (Mean = 0,374; (t(50) = 2,857, p = .006).

IV. Discussion

Several interesting results have been obtained in this experiment. First, the Traffic density has a significant impact on decision made by all the participants. When there is much information in the urban scene (High traffic condition), less participants decide to cross the street, whatever the Age. Second, the Age has a significant impact on time spent to make decision. The decision-making time decreases when the age increases. This result confirms the fact that the age has a strong impact on decision making in pedestrians' skills a process which develops and becomes increasingly effective with the age [34][35][36]. Third, there is an interaction between Age and Traffic density: The decision-making time decreases when the age increases specially when there is much information in the urban scene (i.e., High traffic density condition).

From a theoretical point of view, our results show how the pedestrian's skills would be dependent on the development of at least two simultaneous capabilities: visual exploration strategy and cognitive processing abilities. First, the visual sampling strategies tend to be systematic in younger, not focusing on specific areas or strategic areas and, with age, the visual exploration strategy is specified and is interested in the peripheral areas of the visual field [29]. This development led to a more accurate and relevant information extraction from visual environment in urban areas [23][24][25]. Second, cognitive development allows greater information processing capacity [13][14][15][15], thus taking a more rapid and effective decision. From a theoretical point of view the use of poor visual strategy combined with a cognitive inability to process so many information that explains more time decision-making among young pedestrians in a dense traffic environment.

Several methodological limits prevent us to generalize the results obtained. First, the experiment was conducted inside the school, which resulted in to cause a feeling of observation. The pupils often sought to provide "the good answer" whereas we are interested in their own answer. If the experiment were led in the school, it was a question above all of preserving a medium familiar and reassuring for the pupils. Second, stimuli used in our experiment were only visual and the information in peripheral vision necessarily decreased by the size of the screen. But, for ethical and technical reasons, it was not possible to carry out the experiment in real outdoor environment. Third, stimuli used in our experiment were static (i.e., pictures): So, in our actual new studies, dynamic stimuli (i.e., videos) will be used to introduce dynamic factors, such as motion of vehicles and motion of other pedestrians in the scene. Moreover, even if visual information are crucial, we will add sounds in the experimental material to place participants in a more naturalistic setting. Our study tends to demonstrate on the one hand, that the development of pedestrian skills is essentially based on visual exploration of surrounding environment and on the other hand, these skills increase with the development of more general cognitive abilities

V. Conclusion and Future Works

By using an experimental approach and eye-tracking techniques, our study aimed to investigate the impact of one individual factor (Age) and one environmental factor (Traffic density) on three behavioural indicators related to competencies of very young pedestrians (aged 3-10 years): (i) the decision (i.e., "to cross" versus "not to cross the street"), (ii) the time spent in milliseconds to make this decision and (iii) the visual exploration of urban scenes displayed on pictures.

This study is the first one to our knowledge which investigates visual exploration of urban scenes for very young children (under 4 years old). Using eye-tracking technique is interesting for several reasons. Visual exploration is a irrepressible behavior. Specifically, for young children with limited language capacity, the use of eye-tracking allows comparison with older children and adults. As we reported previously, To our knowledge, no study has looked at the visual exploration of such young pedestrians (under 4 years). Young audiences are more difficult to approach ethically. Younger cannot be put in a situation in real conditions, accompanied due to their motor skills development.

Acknowledgment

Thanks are due to all children, their parents and educational staff.

References

[1] D. N. Lee, D. S. Young, and C. M. McLaughlin, "A roadside simulation of road crossing for children," Ergonomics, vol. 27, no. 12, 1984, pp. 1271–1281.

[2] G. Lee, Y. Park, J. Kim, and G.-H. Cho, "Association between intersection characteristics and perceived crash risk among school-aged children," Accident Analysis & Prevention, vol. 97, 2016, pp. 111–121.

[3] E. K. Adanu, R. Dzinyela, and W. Agyemang, "A comprehensive study of child pedestrian crash outcomes in ghana," Accident Analysis & Prevention, vol. 189, 2023, p. 107146.

[4] K. Koekemoer, M. Van Gesselleen, A. Van Niekerk, R. Govender, and A. B. Van As, "Child pedestrian safety knowledge, behaviour and road injury in cape town, south africa," Accident Analysis & Prevention, vol. 99, 2017, pp. 202–209.

[5] M. Struik et al., "Pedestrian accident project report no. 4: Literature review of factors contributing to pedestrian accidents," Tech. Rep., 1988.

[6] J. N. Ivan, P. E. Garder, and S. S. Zajac, Finding strategies to improve pedestrian safety in rural areas. United States. Dept. of Transportation, 2001.
[7] C. DiMaggio and M. Durkin, "Child pedestrian injury in an urban setting descriptive epidemiology," Academic emergency medicine, vol. 9, no. 1, 2002, pp. 54–62.
[8] W. W. Hunter, J. C. Stutts, W. E. Pein, C. L. Cox et al., "Pedestrian and bicycle crash types of the early 1990's," Turner-Fairbank Highway Research Center, Tech. Rep., 1996.
[9] J. D. Demetre, "Applying developmental psychology to children's road safety: Problems and prospects," Journal of applied developmental psychology, vol. 18, no. 2, 1997, pp. 263–270.
[10] J. D. Demetre et al., "Errors in young children's decisions about traffic gaps: Experiments with roadside simulations," British Journal of psychology, vol. 83, no. 2, 1992, pp. 189–202.
[11] R. A. Schieber and N. Thompson, "Developmental risk factors for childhood pedestrian injuries." Injury Prevention, vol. 2, no. 3, 1996, p. 228.
[12] J. L. Schofer et al., "Child pedestrian injury taxonomy based on visibility and action," Accident Analysis & Prevention, vol. 27, no. 3, 1995, pp. 317–333.
[13] T. Pitcairn and T. Edlmann, "Individual differences in road crossing ability in young children and adults," British Journal of Psychology, vol. 91, no. 3, 2000, pp. 391–410.
[14] A. Meir and T. Oron-Gilad, "Understanding complex traffic road scenes: The case of child-pedestrians' hazard perception," Journal of safety research, vol. 72, 2020, pp. 111–126.
[15] A. Meir, T. Oron-Gilad, and Y. Parmet, "Are child-pedestrians able to identify hazardous traffic situations? measuring their abilities in a virtual reality environment," Safety science, vol. 80, 2015, pp. 33–40.
[16] J. M. Plumert, "Relations between children's overestimation of their physical abilities and accident proneness." Developmental Psychology, vol. 31, no. 5, 1995, p. 866.
[17] J. M. Plumert and J. K. Kearney, "How do children perceive and act on dynamic affordances in crossing traffic-filled roads?" Child development perspectives, vol. 8, no. 4, 2014, pp. 207–212.
[18] J. M. Plumert, J. K. Kearney, and J. F. Cremer, "Children's perception of gap affordances: bicycling across traffic-filled intersections in an immersive virtual environment," Child development, vol. 75, no. 4, 2004, pp. 1243–1253.
[19] D. C. Schwebel, A. L. Davis, and E. E. O'Neal, "Child pedestrian injury: A review of behavioral risks and preventive strategies," American journal of lifestyle medicine, vol. 6, no. 4, 2012, pp. 292–302.
[20] R. J. Brison, K. Wicklund, and B. Mueller, "Fatal pedestrian injuries to young children: A different pattern of injury," Journal of safety research, vol. 20, no. 1, 1989, pp. 41–41.
[21] D. A. Sleet, M. F. Ballesteros, and N. N. Borse, "A review of unintentional injuries in adolescents," Annual review of public health, vol. 31, 2010, pp. 195–212.
[22] J. Warsh, L. Rothman, M. Slater, C. Steverango, and A. Howard, "Are school zones effective? an examination of motor vehicle versus child pedestrian crashes near schools," Injury prevention, vol. 15, no. 4, 2009, p. 226.
[23] M. Congiu et al., "Child pedestrians: Factors associated with ability to cross roads safely and development of a training package," Victoria: Monash University Accident Research Centre (MUARC), 2008.
[24] J. A. Oxley, E. Ihsen, B. N. Fildes, J. L. Charlton, and R. H. Day, "Crossing roads safely: an experimental study of age differences in gap selection by pedestrians," Accident Analysis & Prevention, vol. 37, no. 5, 2005, pp. 962–971.
[25] J. Oxley, B. Fildes, E. Ihsen, J. Charlton, and R. Day, "Simulation of the road crossing task for older and younger adult pedestrians: a validation study," in ROAD SAFETY RESEARCH AND ENFORCEMENT CONFERENCE, 1997, HOBART, TASMANIA, AUSTRALIA, 1997.
[26] G. Simpson, L. Johnston, and M. Richardson, "An investigation of road crossing in a virtual environment," Accident Analysis & Prevention, vol. 35, no. 5, 2003, pp. 787–796.
[27] M. L. Connelly, H. M. Conaglen, B. S. Parsonson, and R. B. Isler, "Child pedestrians' crossing gap thresholds," Accident Analysis & Prevention, vol. 30, no. 4, 1998, pp. 443–453.
[28] M. L. Connelly, R. Isler, and B. S. Parsonson, "Child pedestrians' judgments of safe crossing gaps at three different vehicle approach speeds: A preliminary study," Education and Treatment of Children, 1996, pp. 19–29.
[29] H. Tapiro, A. Meir, Y. Parmet, and T. Oron-Gilad, "Visual search strategies of child-pedestrians in road crossing tasks," Proceedings of the Human Factors and Ergonomics Society Europe, 2014, pp. 119–130.
[30] S. Fotios, J. Uttley, and B. Yang, "Using eye-tracking to identify pedestrians' critical visual tasks. part 2. fixation on pedestrians," Lighting Research & Technology, vol. 47, no. 2, 2015, pp. 149–160.
[31] T. Foulsham, E. Walker, and A. Kingstone, "The where, what and when of gaze allocation in the lab and the natural environment," Vision research, vol. 51, no. 17, 2011, pp. 1920–1931.
[32] A. C. Gallup et al., "Visual attention and the acquisition of information in human crowds," Proceedings of the National Academy of Sciences, vol. 109, no. 19, 2012, pp. 7245–7250.
[33] D. Whitebread and K. Neilson, "The contribution of visual search strategies to the development of pedestrian skills by 4-11 year-old children," British journal of educational psychology, vol. 70, no. 4, 2000, pp. 539–557.
[34] R. Case, "The structure and process of intellectual development," International Journal of Psychology, vol. 22, no. 5-6, 1987, pp. 571–607.
[35] R. Case, "Theories of learning and theories of development," Educational Psychologist, vol. 28, no. 3, 1993, pp. 219–233.
[36] Z. Tabibi, K. Pfeffer, and J. T. Sharif, "The influence of demographic factors, processing speed and short-term memory on iranian children's pedestrian skills," Accident Analysis & Prevention, vol. 47, 2012, pp. 87–93.

Practice Stages for a Proficient Piano Player to Complete a Piece: Understanding the Process based on Two Minds

Katsuko T. Nakahira
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: katsuko@vos.nagaokaut.ac.jp

Muneo Kitajima
Nagaoka University of Technology
Nagaoka, Niigata, Japan
Email: mkitajima@kjs.nagaokaut.ac.jp

Makoto Toyota
T-Method
Chiba, Japan
Email: pubmtoyota@mac.com

***Abstract*— Research into instrumental music performance has garnered significant attention, particularly regarding the intricate interplay of perceptual-cognitive-motor interactions, knowledge application, and the cognitive representation of musical structure. Understanding these dynamics holds promise for enhancing instruction and aiding learners in their journey towards mastering instrumental performance and practice. However, grasping the learning process necessitates more than just comprehending the individual cognitive mechanisms at play; it requires a holistic approach that considers the cognitive architecture enabling the integration of these processes. In this paper, based on the MHP/RT framework proposed by Kitajima and CCE research method which based on the MHP/RT principles, we attempt to understand the process of proficiency in music performance by proficient piano players as a brain model based on the coordination of perception, cognition, and movement, and the concept of Two Mind. Initially, we modeled the cognitive process of piano performance proficiency, and ethnographically described the process of proficiency in music performance for selected elite monitors. The descriptions are analyzed and compared with the model of cognitive processes and actual behaviors in performance proficiency. The description of which perspectives can/cannot be interpreted by the model based on the MHP/RT were considered. Finally, a series of piano playing exercises and lessons are analyzed from the perspectives of the Two Minds process, and the knowledge system (implicit/explicit) utilized. Through the analysis, the relationship between acquired knowledge and cognitive ability and Two Minds is considered. The findings suggest that the proficiency process of instrumental music performance exhibits a kind of phase transition. It involves not only a gradual shift from prolonged, System 2-driven mechanical training towards an intuitive, System 1-driven unconscious expression but also deviations from this pattern. Therefore, it is imperative for players to thoroughly comprehend their perception of the entire piece (System 2) while also fostering a sense of ease and naturalness in performance akin to unconscious expression (System 1) for the listener.**

Keywords: Proficient Piano Player; Cognitive Process; Two Minds; MHP/RT; Ethnological Study;

I. Introduction

Instrumental performance has attracted attention as a result of the interaction of perceptual/cognitive and motor abilities. Numerous studies focus on the process of instrumental performance proficiency. The goal of this study is to understand the proficiency process of instrumental performance, which has the possibility of providing better instruction to a performance learner.

Palmer [1] summarizes empirical research on instrumental performance in terms of conceptual interpretation formation, control over motor actions, interpretive transfer as perception, and structural disambiguation. Lehmann and Ericsson [2] focus on the development of instrumental performance skills at the level reached by high school students and amateurs. In their study, they posit that the method of practice is particularly important in improving the level of instrumental performance. A study that focused on the subjectivity factor of instrumental performance practice itself, shares a different perspective; Araújo [3] conducted an online questionnaire survey of self-regulated practice behaviors pertaining to advanced musicians, from which, he indicates that practice organization, personal resources, and external resources are important factors. For understanding proficiency in instrumental performance, Chaffin et al. [4][5] applied the protocol analysis method, investigating the characteristics of a concert pianist's performance of a piece of music, in addition to the characteristics of the music. They categorized elements of the instrumental performance in three basic dimensions (fingering, high difficulty, and familiarity with the note form), four interpretive dimensions (phrasing, dynamics, tempo, pedal), and three expressive dimensions (basic, interpretative, expressive). Through the categorization process, a possibility of the existence of image for desired representation of the music from the beginning, so-called a "*big picture*" was found.

Focusing on how to practice instrumental music performance, as Palmer [1] mentioned, an individual's cognitive representation of musical structure is important in terms of specific errors and knowledge utilization in instrumental music performance. To understand this, it is not only sufficient to understand the cognitive mechanisms for individual perceptual, cognitive, and motor processes, but also research from the perspective of cognitive architecture, which enables these processes to be handled in an integrated manner.

There are several cognitive architectures concerning the interaction between perceptual/cognitive and motor abilities, however, we apply the Model Human Processor with Realtime Constraints (MHP/RT) proposed by Kitajima et al. [6][7][8] for this study. MHP/RT is a cognitive architecture, which is constructed by extending the concept of Two Minds (Kahneman [9][10]) to reproduce the perceptual, cognitive, and motor processes as well as memory processes at work in

 ISBN: 978-1-68558-157-2

everyday action selection. The MHP/RT has been applied to the comprehension of language utilization and the process of creating ceramic artworks [11][12]. For the latter study, the MHP/RT is applied with a companion field study methodology called Cognitive Chrono-Ethnography (CCE) [8][13]. CCE is a research methodology utilized to clarify the process of development concerning how a specific individual has acquired the behavior selection characteristics at the present time, and the development process of the behavior selection characteristics at the site where the behavior is executed based on the behavior selection mechanism on a time axis, which is specified by the MHP/RT. The implementation of CCE requires appropriate subjects – elite monitors – who are ideal subjects for the purpose of the particular research.

In this article, we attempt to understand the process of proficiency in music performance by applying CCE, underpinned by the MHP/RT's underlying concept of the Two Minds, such as the interplay between the unconscious process of System 1 and conscious process of System 2. In Section II, the cognitive process in piano performance proficiency based on the MHP/RT is modeled, which provides the basis of the CCE. In Section III, the process of proficiency in music performance for selected elite monitors is described. In Section IV, the cognitive process model and actual behavior in performance proficiency is compared, and the points that can be interpreted by the model, the points that cannot be interpreted by the model, and the implications from the MHP/RT perspective are thoroughly discussed.

II. Cognitive Processes Leading to Proficiency in Piano Performance

Playing piano involves processes such as reading the score and creating its mental representations and retrieving knowledge from long-term memory related to the representation, which comprise a variety of information necessary to establish links between the representation of visual information on the score and the concrete hand/finger movements to be conducted on the instrument. Long-term memory consists of chunks for establishing these links, which develop with practice from an initial configuration with inefficient linkage to an advanced one with effective linkage, corresponding to the state of proficiency. This section provides a theoretical description for the development process of the chunk structure.

A. *Initial State: Initial Chunks in Long-Term Memory*

The chunk structure, within long-term memory at the beginning of reading a score, is a set of chunks that have been acquired as knowledge and stored in long-term memory. Let C_{mus} be the chunk set that must be stored, the chunk set C_{LM} that exists in long-term memory at a certain time t is a subset of C_{mus}. C_{mus} is composed of the following, based on the smallest element c_i $(1 \leq i \leq n_c(t))$:

- Chunks composed of the minimum element $n_c(t)$ only,
- Larger chunks composed of $n_e(t)$ $(1 < n_e \leq n_c)$ minimum elements, without duplication, and

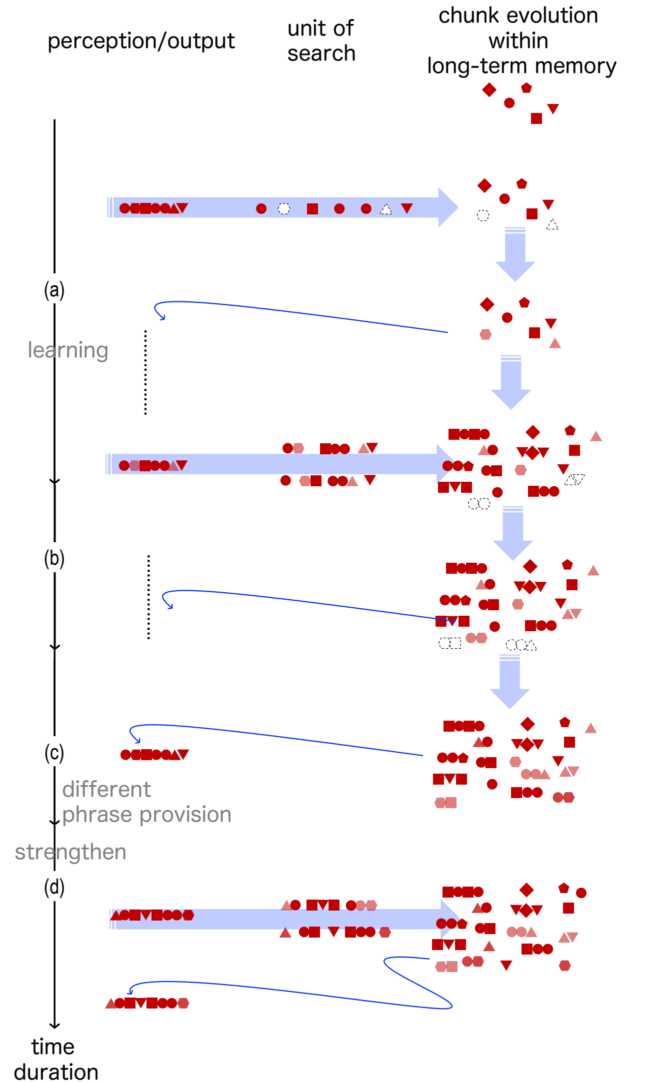

Figure 1. The relation between perception/ourput and chunk evolution within long-term memory

- Still larger chunks consisting of $n_{e'}(t)$ $(1 < n'_e \leq n_c)$ minimum elements, with duplication allowed.

In addition, C_{LM} consists of the relation:

$$C_{LM} \quad = \quad \{\mathbf{c} \mid c_i \in C_{mus},\ 1 \leq i \leq n_{LM}(t)\}$$

The internal structure of C_{LM} evolves as a learning and strengthening process as the number of chunks it contains increases with practice.

B. *State (a): Recognition of Individual Notes or Short Phrases*

When reading a new score of music, the perceived sequence of notes is divided into known notes or short phrases. When the learner encounters an unknown phrase, it is stored as a new chunk. The layer (a) in Figure 1 exhibits this state. A sequence of notes $S(t)$ perceived at t consists of n_p elements. When $S(t)$ is initially read, $S(t)$ is separated by n_p individual chunks c_j, and the score reading process commences. When an unknown element $c_{j'}$ appears, $c_{j'}$ is newly stored in long-term

ISBN: 978-1-68558-157-2

memory (black dashed line in the figure). As the score reading proceeds in this manner, the reading of each n_p element proceeds smoothly, and the newly stored $c_{j'}$ is additionally stored and fixed in memory. In this state, the learner plays these phrases with a pause – each c_j plays with intermittent, so that it can only be played with an awareness of partial cohesion.

C. State (b) and (c): Recognizing Multiple Chunks Simultaneously

When a sequence of notes can be recognized as individual notes or short phrases, the same $S(t)$ is perceived, but several c_j are lumped together and recognized as a novel chunk (phrase) in order to play the music significantly smoother. The layer (b) in Figure 1 exhibits this state. When the learner perceives this unknown combination of c_j's as a set, it is stored as a new chunk (black dashed line in the figure). At this time, the size of the chunk is larger than that of the state (a), enabling the learner to perform with an awareness of longer chunks. In order to be aware of the large phrases, training is also conducted to recognize $S(t)$ more reliably by separating the elements of $S(t)$, and c_j's, in various ways. When the learner perceives an unknown c_j combination, the combination is newly stored in the long-term memory (black dashed figure in Figure 1). Through repeated training, the number of chunks (phrases) formed by the combination of c_j that existed prior to the training increases in long-term memory, and the learner's chunk set structure incrementally approaches C_{mus}. Finally, the learner's chunk set structure in long-term memory is reached at the state (c), and the presented sequence of notes can be recognized as a single chunk. If the learner's condition reaches the state (c), the learner's skill is regarded as "acquiring the ability to perform $S(t)$ with proficiency."

D. State (d): Efforts toward more Reliable Chunking

When the structure of C_{LM} is saturated, even if a sequence of notes is novel to the user, it can be perceived as a known sequence of notes by devising alternative segmentations for c_j. Assuming that a new sequence of notes $S(t')$ consisting only of chunk groups in C_{LM} is perceived, in this regard, the recognition of $S(t')$ is divided by utilizing the chunk elements in long-term memory. Since all the chunks are known, reading will commence without much effort being required. The layer (d) in Figure 1 exhibits this state. In this case, the chunks in long-term memory are simply strengthened.

E. Summary

As the above state is repeated, more C_{LM} is accumulated in long-term memory, and even when it is presented with a complex piece of music, the user can be confident that "this musical piece can be performed". Therefore, as C_{LM} increases in the fasion described above, the more musical pieces the learner practices, the more proficient the learner becomes, and the more musical pieces the learner is able to perform. However, in actual performance, there are two types of practice: one is to perform without making mistakes even if it takes a longer time, i.e., a phase of musical score reading, and the other is to perform without stopping to have the audience experience a smooth performance. A performance aimed at continuously avoiding mistakes involves different cognitive processes in terms of System 1 and System 2. System 1 is controlled by the feed-forward process and is compatible with the latter; while System 2 is controlled by the feedback process, i.e., the conscious process that monitors the outcome of System 1's performance to correct any errors, and is compatible with the former. The process of utilizing chunks should be different in these cases. The next section describes an example of how the cognitive processes, leading to performance proficiency described above, appears in actual performance proficiency.

III. An Example of Proficiency Process of Music Performance by a Proficient Piano Player

In this section, we describe a CCE study focusing on a single elite monitor, following the study conducted by Kitajima et al. [12] to understand the skill of a traditional craft artist and how the skill is passed down from generation to generation, as well as how the process by which a proficient piano player reaches the expected performance level through practice of a given piece of music. We call the elite monitor, i.e., the proficient amateur piano performer, P^3, and consider the situation where P^3 tries to achieve a high level of performance perfection through practice. The characteristics of the score that P^3 is aiming for, i.e., the target score abbreviated as TS, with reference to P^3's performance skill level is elucidated. Subsequently, the study enumerates the elements included in the practice to be conducted to achieve TS, and elucidate the development of the practice over time and the content of the practice elements associated with it.

Here, the role of P^3 is taken by the first author. The core of the CCE analysis – describing P^3's experience – has operated as stated below. In order to avoid a biased analysis, when P^3 made an ethnographic analysis, P^3 asked the instructor the meaning of musical suggestion or cognitive meaning with regard to playing piano training method given by instructor. For representation of the CCE analysis, P^3 wrote down the experience series and the initial proposed model. Subsequently, the other two authors, who are professionals with the CCE, meticulously investigated the proposed model which P^3 proposed. Finally, the authors adopt the representation which all authors judged to be acceptable.

A. Main Objectives of a Skilled Piano Learner

In general, there are two main objectives when an adult learner attempts to acquire proficiency in musical performance.

1) Internal factor, such as genuinely wishing to become proficient for strong motives, e.g., favorite piece of music, wanting to perform it, and select a piece for a competition, etc.
2) External factor, i.e., a piece assigned for a competition or given for practice

It depends on which objective the learner set, but here we target the “to be made best performance at the competition” in 1). In this instance, P^3 can select a piece of his/her own will, but the target performance achievement is to pass at least the regional qualifying round of the piano competition (with a required score is 70/80 or higher), and preferably the regional finals (with a required score is 80/86 or higher).

B. Flow of Music Proficiency to Reach Competition Stage

Figure 2 represents the general proficiency process of a musical performance. Given that it takes a long time, anywhere from six months to one year, to become proficient in a music performance, the most important process is the selection of the music to be performed. Basically, there are two important perspectives of selection with regards to music and performing: whether or not the piece is appropriate for the player’s performance skill level, and whether or not the player prefers the piece. However, in the case of P^3 who can participate in competitions, there is a lot of freedom in music selection, which means the performance skill level is not a constraint. Hence, P^3 asked her instructor for several candidate pieces that would be suitable for her own timbre and expressive characteristics. On top of that, P^3 herself selected the music to be performed through the following process :

- Give the score a once-over,
- Try out playing the initial few pages (where most of the music motifs are available), and confirming whether or not they can play the piece to the end, and
- Listen to a professional performance and determine if you can grasp the image of the piece.

After the piece for competition is selected, the learner practices playing it to the end so that the framework of the piece can be imagined. Then, the learner makes *Analise* with the outcome of practice. Post-completing the *Analise*, she fixes the image that expresses fluent performance, and additional interpretation as well as the necessary skills for performance expression. Subsequently, she will go to the competition performance. Details of each process are described in the subsections to follow.

C. Details of the Processes and Mapping on Two Minds

1) Score Selection toward Practice: There are various ways to select a music piece for competition. In a competition which is not given a set piece of music and in which the goal is to perform well in the competition qualifying round and the finals, there are a number of points to consider in the selection of the music piece. In addition to selecting pieces and considering the level of difficulty, there are some other selection points. In the case of P^3, the following procedure was utilized to select pieces at an appropriate level.

1) Ask her instructor to list some candidate pieces:
   There are two reasons for this. One is to avoid selecting pieces of an inappropriate level for the competition. The other is to have an outsider recommend a piece that is suitable for the color of P^3 from a third party’s perspective.
2) Read the scores giving a once-over to the end to get an image of the music, and narrow down the candidate pieces to 2∼3:
   In the case of P^3, the key points in narrowing down the candidate pieces are basically two points : whether the feeling of the music fits, and whether the image of the music can be grasped by reading the scores once-over.
3) Read and perform the initial few pages of the piece (up to the point where the initial and subsequent motifs appear):
   The mechanics utilized in the actual performance are quite different from the image, and even if the instructor thinks “She can perform this,” it is rare for P^3 to find with the mechanics that “her cognitive or motor reaction rejects”.This process is designed to prevent such mismatches.
4) Select a piece of music that she is convinced she could perform well.
   In the case of P^3, the selection is made focusing on the music that immediately comes to mind concerning “what she wants to express” when the motifs are performed.

Thus, the selection of music, which is the initial step in music practice, is often determined to a large extent from a System 1 perspective. However, in order for System 1 to function, the skills that the body has acquired are as a result of long-term System 2 training. This is because it is necessary to determine which of the chunk configurations (a)∼(d) in Figure 1 are utilized in the piece, and to start from the point where she checks the degree of practice required to become proficient at the piece. If the majority of the motifs are in the state exhibited in Figure 1(a), it will take a considerable period of time to become proficient in the piece, and depending on the situation, the player may have to give up. It is also difficult to receive assurance that “I can perform the piece”. Figure 1(b) and (c) are more likely to be able to perform with the music with proficiency, as it is easy to obtain the confidence that “I can perform the piece”. Therefore, it is expected to be the target of music selection.

2) Transition of Instructional Contents: The process required to complete a musical performance can be divided into two main categories: musical score reading and compositional expression. The musical score reading is a practice stage in which mechanics – motor system – play a major role. Compositional expression practice is the stage, where musical interpretation, i.e., the player’s expression tailored to his/her sensitivity, and technique for the expression, plays a major role. There are significant disparities between the two practices.

In the case of mechanically trained music reading, the focus is on accurate keystroke execution. Therefore, the main task of practice is to reproduce the exact note value, pitch, and interval for each note head. In simple words, the primary focus of practice is to count the lengths accurately, to check the details of pitches, and pitches described in the score, and to check accidentals, articulation marks, ornaments, pedal marks, etc. The utilization of knowledge in this process is basically

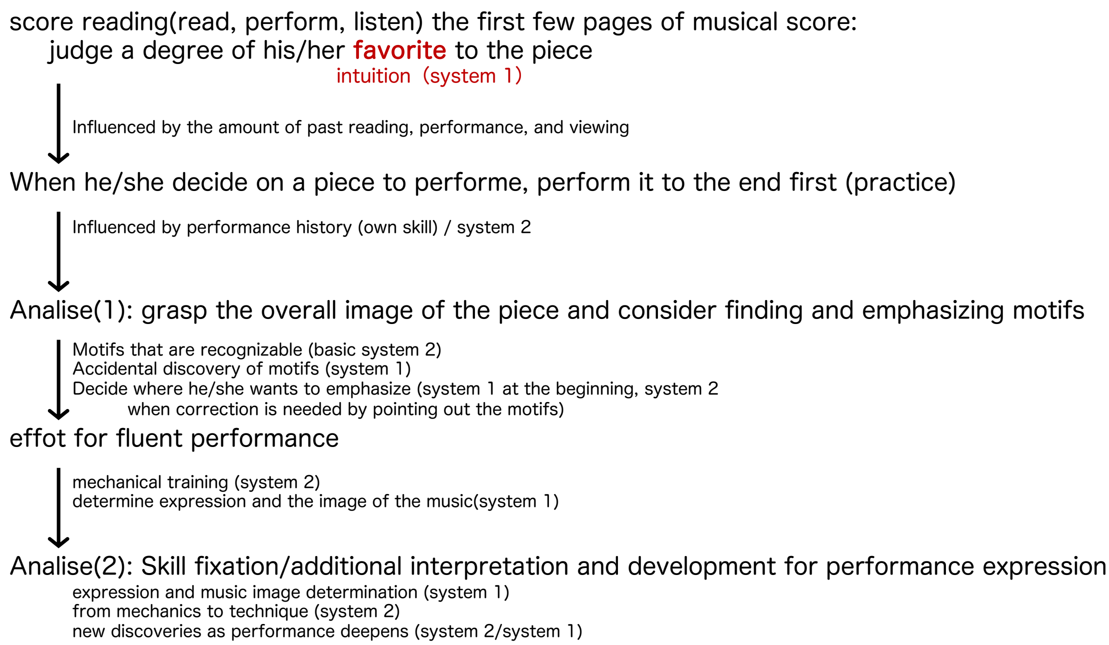

Figure 2. Flow to proficiency in music performance

centered on (a) and (b) in Figure 1, and is mainly a System 2 process in terms of practicing to play the sequence of notes exactly as described in the score.

Conversely, in the case of the musical score reading, which trains techniques for compositional expression, a variety of control with regards to the fingers and cognition is required, such as how far to play a note sequence as a whole, how to add dynamics, and which notes to insist on. Simply put, it is a prerequisite that the player has already finished *Analise* the piece and that the player's image of the entire piece has been established. The two elements are not independent, which means the existence of accurate mechanics enables the player to confidently express music utilizing this technique.

It is also necessary to learn the mechanics required to make the technique more precise, for example, the dynamic technique and the techniques required to change timbre. In this sense, it is a cooperative activity between cognitive and motor processes. When teaching these cooperative activities, the instructor decides on the contents of instruction in the following manner with regards to listening to the player's performance.

- Understand what the player wants to emphasize and what kind of expression he/she wants to express from the performance.
- Imagine what the player wants to do but does not seem to be able to do.
- Point out obvious deviations from the interpretation of the performance as described in the score, and give a more natural interpretation.

Of course, if a player has sufficient ability, he/she can improve these items by self-regulation post-recording his/her own performance. However, different from students who are beginners when it comes to performing, there is a limit to self-regulation improvement in the field where advanced performance is required. For this reason, suggestions from the instructor play an important role for a player's proficiency.

The instructor suggests more exercises that would contribute to the formation of chunks as opposed to the movement. For instance, changing the playing speed between stressed and unstressed parts (contributes to the formation of chunks), practicing rhythm (contributes to the formation of fingering chunks), and giving more accent than necessary to notes that should be emphasized (contributes to the formation of chunks in the imagery of the music). The primary utilization of knowledge in such exercises is exhibited in Figure 1 (c), and primarily consists of combining the smallest elements c_j that may appear in a piece of music in as long a phrase as possible, in order to be aware of the motifs of the music piece.

Given that this is an expression of how the player feels about the music, it is not necessarily a System 2 process, but is gradually shifted to a System 1 process. Repeat the performance expression in the System 1 process as trial and error until the player's intention is well conveyed. The player repeats the pattern that successfully shares the expression he/she wants to share in the System 2 process to fix the expression. In addition, although System 2 and System 1 repeatedly appear during practice, there will be situations where "System 2 $<$ System 1." This is a time when unconscious performances increase and dramatic improvements in performance expressions occur.

As player's technique improves, he/she gradually discovers new discoveries and desires for additional expression in the piece. As player's techniques improve, he/she can make new discoveries for motifs/notes significance, and grow his/her

 ISBN: 978-1-68558-157-2

appetite regarding compositional expression. Some of these improvements can be made solely by P^3, while others can only be made with the advice of the instructor. In any case, the final regulation for the competition will be made by repeating such improvements. At this time, the utilization of knowledge increases in the System 1 process in order to challenge a variety of expressions. In addition, even without the System 2 process, the approach to the state known as "the body remembers" and enables various expressions to be challenged.

IV. Discussion Based on Two Minds

A. Overview of Annual Lessons

The following is a summary of the practice sessions described in Section III, contrasted with the duration of the lessons. In order to take lessons, the learner makes practices about one hour per practice. The number of practice sessions is generally two to three times per week, depending on the situation at the time. One to two weeks prior to the competition, practice sessions occurred almost every day.

- 11 months prior to the qualifiers of the competition (C_P): Selection of pieces
  Play a few pages of several music pieces and select the pieces that suit the player's favorite
- Six months prior to C_P post-selection of music pieces: score reading (T_{C1}).
  Basically, the students practice developing techniques in some parts while focusing on the mechanics. It takes about three months to reach the level of playing through the whole piece, and the playing speed is two to four times slower than the specified speed.
- Six to three months prior to C_P: Transition to the expression of musical ideas (T_{C2}).
  ** By this time, the mechanics are 80% complete, so the main focus is on practicing to develop the techniques necessary for compositional expression.
- Three months prior to C_P, completion of the compositional expression:
  Completion of the musical compositional expression · constructing the music image (T_{C3}).
- 1 month prior to $C_P \sim C_P$: final adjustment for the regional qualifying round. (T_{C4}).
- Post C_P to the primary line of the competition: if you pass the qualifying round, practice for the regional finals (T_{C5}).

A total of 25 lessons were given. Each lesson lasted approximately 1.5 hours.

B. The Relation Between the Flow of Playing Perfection and Two Minds

Once a series of experiences had been performed, the second trial for attending the competition may be able to utilize the prior experience to finish the piece at a faster pace. The items from stage 2 (practice) analise(1) to the effort for fluent performance in Figure 2, or $T_{C1} \sim T_{C2}$ in terms of the lesson schedule, are basically affected by the experience. It is possible to reach the stage of mechanical performance as reproducing with midi, through an expreience such as earlier through participating in competitions repeatedly, taking lessons for many years, and so on. These changes are continuous, i.e., the degree of improvement increases monotonically as a function of the number of performances.

However, additional interpretation and deepening of the performance beyond that point may not be successfully achieved by simply repeating the process. In P^3's participation in the competition, the performance around two to one month prior to the competition qualifier (T_{C3}) undergoes a large change every year, which cannot be explained by the passage of time alone. By this time, the mechanical performance is almost complete in a form that is approximately 1.5 times less than the speed at which it is played on the day of the competition, but it is far from sufficient completion, and the so-called "composition expression and understanding." Around the transition from T_{C2} to T_{C3}, there is a significant change in the recognition of musical motifs and a shift to the recognition of larger motifs and the expression of *Dynamik* including expression marks. Other changes in timbre, for instance, from soft to hard sounds, are also observed.

This situation is further analyzed from the perspective of the disparities between the characteristic times of System 2 and System 1. The period of T_{C1} is a practice process in which the System 2 process is dominant. The time scale for practice per phrase is primarily the cognitive band in Newell's Time Scale of Human Action [14], since the phrase itself is not very long. The time span of the cognitive band is about ∼10[s]. Given that information is exchanged between the working memory and long-term memory in about 10 seconds of very short chunks, all knowledge is likely to be recognized only as fragments. Therefore, even if one were to predict the next chunk that will appear during the performance of a piece of music, only a few chunks exist which is able to collation, and even if many chunks can make connected collation, only a few percent of the entire piece can be predicted, making it difficult to see the entire piece.

By repeatedly practicing a very short chunk, the body remembers new chunks in the order of ease with regards to memorizing. If a similar chunk had been utilized in the past, it is recognized as a "meme" and the chunk becomes an active meme [15]. At this stage, the chunk is considered an action-level meme. Conversely, even if a chunk exists in long-term memory, if it is never invoked again, the chunk is no longer imitated and becomes an extinct meme, therefore making it inactive. From the above, for a learner like P^3 who cannot engage in constant piano practice, score reading at the competition level will require an enormous amount of time.

However, by the time the T_{C1} period had elapsed, the information per chunk is considerably larger. Therefore, during T_{C2}, chunks of the larger size are available for the cognitive processes in the cognitive band. The number of chunks available for cognitive process, invoked chunks, is getting longer and longer, and their coverage is getting longer. As a result, the number of operations utilizing the working memory and

TABLE I. PHASE CLASSIFICATION OF KNOWLEDGE/COGNITIVE PROCESSES AND DEGREE OF INFLUENCE

		process		knowledge		environment
Phase	Subphase	System 1	System 2	tacit	explicit	outsider intervention
decide piece	offer candidate	*	*		*	**
	once-over	*	**	*	**	**
	playing trial	**	*	**	**	
	listning	***	*	**	*	*
	select piece	***	*	***	*	*
score reading	fingering		***		***	
	score reading		***		***	
analise(1)	recognize motif		***		***	*
	set enphasis		***		***	*
	find motif of serendipity	***	*	*	***	
expression	mechanic	*	***	**	***	***
	construct image	***	**	**	**	
	transfer expression	**	**	*	***	**
analise(2)	confirm expression	**	***	**	***	***
	confirm image	**	***	***	**	***
	Technic		***		***	***
	performance deepening/serendipity	***	***	*	***	
final stage	fragmentation and reintegration	**	**	**	***	

long-term memory for a unit time will be gradually increased, and the addition of information to the chunks in long-term memory will be accelerated. In simple words, it is thought that the easily accessible active meme will change to behavior-level meme [15]. In this process, the time when a knowledge group is composed of only an appropriate chunk size may be approximately the time toward T_{C3}.

By the time T_{C3} is entered, the number of movements to call chunks from long-term memory is considered to be considerably reduced. As a result, cognitive-motor coordination is conducted more unconsciously. If all the chunk invocation patterns are optimized, almost all the performances will be performed unconsciously by System 1, and an abrupt phase transition from the T_{C2} state will occur. As a result, one should feel at least a dramatic improvement in their ability for good finger movement.

In the case of P^3, the pieces learned in the last three years, including the time of writing this article, were as follows:

- 2 years ago :
  Partita BWV 826, composed by J. S. Bach (score A)
- 1 years ago :
  Allegro Appassionate op.70, composed by Charles Camille Saint-Saëns (score B), Allemande in French Suites BWV 812, composed by J. S. Bach
- now :
  piano sonata op. 14 first movement, composed by Sergei Sergeyevich Prokofiev (score C), Allegro in Italian concert BWV 971, composed by J. S. Bach

Each of them spent about a year memorizing the scores prior to the competition. Despite the difference in the compositional age, compositional structure, and knowledge required, score A received 76 points and score B received 79 points in the final piano competition. This indicates that the learners' performance skills themselves were well-developed, even though they performed different types of music. In simple words, the examples of the experience in Section III can be considered to have a certain universality.

C. The Relation between Knowledge/Cognition Process and Two Minds

Finally, we discuss the relationship between the Two Minds and the knowledge as well as cognitive abilities acquired through a series of piano practice and lessons. Table I exhibits the results of subjective evaluation for each flow subphase in Figure 2. The items are: the process of the Two Minds, the knowledge system utilized (implicit/explicit), and the subjective evaluation of the degree of intervention by others. The higher the number of $*$, the stronger the effect on the item.

At initial glance, one might think that instrumental music performance is a continuous shift from long time-consuming mechanical training by System 2 (inference) to unconsciousness of musical expression including System 1 (intuition). However, in fact, this is not true.

For instance, in the case of the music selection phase, many factors are involved in the decision-making process, including player: 1) preference (System 1), 2) matching with performance ability (System 1/2), and 3) matching with the ability to read music (System 2), etc. It depends on the situation at that time which of these factors should be prioritized. In simple terms, if motivation is a given priority, preference is given priority, and if ability is given priority, a little more weight is given to the performance ability or reading ability. This indicates that the process of proficiency in instrumental performance is not determined solely by preference or ability. Conversely, song selection, although often neglected at the initial glance, is the most important phase as it is deeply related to the motivation of the student when he or she begins to practice. In the case of the piano beginner, the instructor often selects pieces at an appropriate level, but in the case of a proficient amateur learner, the selection requirements for the score selection are reduced to some extent. Therefore, the degree of freedom of parameters is high, and the decision-making process involves a mixture of perceptual processes to trigger preference by listening to the sound source, perceptual-cognitive processes to compare with the reading ability by score reading, cognitive-motor processes to consider the performance ability, and processes to coordinate all of these. Therefore, the ability to select appropriate music can be regarded as an important ability.

This also applies to the score selection process. It is easy to assume that a System 2 process takes precedence in *Analise* as well, since it requires a precise analysis of the music. However, various cognitive processes are intricately related as follows: Recognizing the motive and searching for methods to emphasize it (System 2), determination of the expression method that is perceived as effective (System 2/1), new expressions discovered by chance (System 1), and so on. Therefore, not only an orderly musical interpretation but also a balance with the impression is important. In particular, when representing a piece of music, it is necessary to "see the big picture", i.e., the following items must be fulfilled at the same time.

- The player must have a complete understanding of how to perceive the entire piece (System 2).
- The player's natural behavior as if he/she were performing it unconsciously, which should be comfortable for the listener (System 1).

Therefore, it is necessary to understand the process of coordination between System 2 and System 1.

V. Conclusion and Future Works

In this study, based on the MHP/RT cognitive architecture and its companion field study methodology, CCE, we attempted to understand the process of proficiency in music performance by proficient piano players as a brain model based on the coordination of perception, cognition, and movement, as well as the Two Minds.

In Section II, we theoretically explained the development process of the chunk structure that exists in the long-term memory, which is the most important part of the piano playing process – score reading and piano playing mechanics/technics. There is a structure, which consists of many small units of chunks in the long-term memory, and links are attached between chunks through practice. As a result, larger chunks are formed. The study argues that the proficient state refers to this state.

In Section III, we ethnographically described the piano practice and proficiency process with P^3 as an example, aiming at participation in the competition. We exhibited that there are four major components: selecting score (System 1), practice (System 2), *Analise*(System 1/ 2), and the effort for fluent performance (System 1/2).

In Section IV, a series of piano playing exercises and lessons were analyzed from the perspectives of the Two Minds, the knowledge system utilized (implicit/explicit), and the intervention of others. Post the analysis, the relationship between the acquired knowledge and cognitive abilities as well as the Two Minds was examined by incorporating the idea of the active meme. The results suggest that instrumental music performance requires both a complete understanding of how the player perceives the entire piece (System 2) and natural behavior that is comfortable for the listener (System 1), as if the player were playing unconsciously.

As an application, we can consider various educational support measures for performance proficiency by understanding the actual growth process of chunks and the player's proficiency process in more detail based on cognitive architecture. In recent years, there have been increasing opportunities for adults who are not professions of instrumental music performance to enjoy music as a hobby as amateurs. While he/she is not a professional with regards to instrumental performance, one of the elements necessary for proficiency, "motivation to practice" and "support for its maintenance", is left solely to the desire of the learner to play this piece, not to the instructor. In this situation, if learners cannot overcome the difficulties they encounter when practicing instrumental music, they may give up the hobby of instrumental music itself. However, if the instructor can appropriately understand the difficulties that the learner cannot overcome, and can demonstrate to the learner how to increase the possibility of overcoming the difficulties, the withdrawal rate of the learner may be reduced. We believe that this study will contribute to the research from this perspective.

The majority of prior research on the process of proficiency in musical performance has focused on the understanding of cognitive mechanisms for individual perceptual, cognitive, and motor processes. Research on the cognitive mechanisms of individual processes is primarily suitable for understanding proficiency or the process of developing literacy, in terms of how beginners can play music. This study's findings can apply to constructing efficient training methods for the novice learner.

However, learner's playing skill shifts slowly with time, so that it is necessary to improve teaching content and methods based on the learner's proficiency. In case the learner's goal level with regards to attending the competition, is not only the improvement of literacy but also the process of proficiency in the "big picture" of a piece of music. In order to establish such a sophisticated instructional method for individual cases, we need a method for analyzing successful/failed cases based on the empirical rules of instruction, and the resulting cognitive model of the learner. In this case, it is necessary to go into the resonance with past performance and appreciation activities, and there are many areas that cannot be elucidated only by the prior cognitive architecture. As one of the solutions to this problem, understanding performance proficiency utilizing a brain model based on the Two Minds is considered to be effective. As a future issue, we believe that further research based on this study will enable, for instance, remote performance instruction of musical pieces at a higher level.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 19K12232, 20H04290, 22K02840, 22K02885, 23K11334. The author would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

[1] C. Palmer, “Music performance,” Annual review of psychology, vol. 48, 02 1997, pp. 115–38.

[2] A. Lehmann and K. Ericsson, “Research on expert performance and deliberate practice: Implications for the education of amateur musicians and music students,” Psychomusicology: A Journal of Research in Music Cognition, vol. 16, 04 1997.

[3] M. V. Araújo, “Measuring self-regulated practice behaviours in highly skilled musicians,” Psychology of Music, vol. 44, no. 2, 2016, pp. 278–292. [Online]. Available: https://doi.org/10.1177/0305735614567554

[4] R. Chaffin, G. Imreh, A. F. Lemieux, and C. Chen, “”seeing the big picture”: Piano practice as expert problem solving,” Music Perception, vol. 20, no. 4, 2003, pp. 465–490.

[5] R. Chaffin and L. Topher, “Practicing perfection: How concert soloists prepare for performance,” Advances in Cognitive Psychology, vol. 2, 01 2006, pp. 113–130.

[6] M. Kitajima and M. Toyota, “Decision-making and action selection in Two Minds: An analysis based on Model Human Processor with Realtime Constraints (MHP/RT),” Biologically Inspired Cognitive Architectures, vol. 5, 2013, pp. 82–93.

[7] M. Kitajima and M. Toyota, “Simulating navigation behaviour based on the architecture model Model Human Processor with Real-Time Constraints (MHP/RT),” Behaviour & Information Technology, vol. 31, no. 1, 2012, pp. 41–58.

[8] M. Kitajima, Memory and Action Selection in Human-Machine Interaction. Wiley-ISTE, 2016.

[9] D. Kahneman, “A perspective on judgment and choice,” American Psychologist, vol. 58, no. 9, 2003, pp. 697–720.

[10] D. Kahneman, Thinking, Fast and Slow. New York, NY: Farrar, Straus and Giroux, 2011.

[11] M. Kitajima et al., “Language and Image in Behavioral Ecology,” in COGNITIVE 2022 : The Fourteenth International Conference on Advanced Cognitive Technologies and Applications, 2022, pp. 1–10.

[12] M. Kitajima, M. Toyota, and J. Dinet, “Art and Brain with Kazuo Takiguchi - Revealing the Meme Structure from the Process of Creating Traditional Crafts -,” in COGNITIVE 2023 : The Fifteenth International Conference on Advanced Cognitive Technologies and Applications, 2023, pp. 1–10.

[13] M. Kitajima, “Cognitive Chrono-Ethnography (CCE): A Behavioral Study Methodology Underpinned by the Cognitive Architecture, MHP/RT,” in Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society, 2019, pp. 55–56.

[14] V. Monaco, “Classification and authentication of one-dimensional behavioral biometrics,” 09 2014.

[15] M. Kitajima, M. Toyota, and J. Dinet, “How Resonance Works for Development and Propagation of Memes,” International Journal on Advances in Systems and Measurements, vol. 14, 2021, pp. 148–161.

A Metadata Model for Harmonising Engineering Research Data Across Process and Laboratory Boundaries

Martin Zinner*, Felix Conrad*, Kim Feldhoff*, Hajo Wiemer*, Jens Weller†, and Steffen Ihlenfeldt*‡
*Chair of Machine Tools Development and Adaptive Controls (LWM)
Institute of Mechatronic Engineering (IMD)
Technische Universität Dresden
Dresden, Germany
Email: {martin.zinner1, felix.conrad, kim.feldhoff, hajo.wiemer, steffen.ihlenfeldt}@tu-dresden.de
†Symate GmbH
Dresden, Germany
Email: jens.weller@symate.de
‡Fraunhofer Institute for Machine Tools and Forming Technology (IWU)
Dresden, Germany

***Abstract*—The availability of precise and comprehensive experimental data in science and technology is crucial for the usability of Artificial Intelligence (AI) models. A digitally analysable, system-independent representation of datasets is essential for enabling the deployment of data-driven applications across different platforms. We propose a metadata model based on domain-specific languages and terminologies, that allows researchers to focus on data provision by reducing routine activities rather than attempting to align with other research groups. Furthermore, it enables fast and efficient integration of new partners from different laboratories and disciplines. To conclude, our approach supports a paradigm shift away from more or less subjectively designed individualistic conceptions in handling research data towards objectively established harmonised solutions. The approach is illustrated for an Interdisciplinary Research Training Group, in which researchers from more than 10 different departments are involved with the main research profiles, such as textile and polymer technology and material sciences.**

***Index Terms*—Metadata Model; FAIR Principles; Research Data Management; Ontology; Machine Learning; Domain-Specific Technical Languages.**

I. Introduction

Metadata is data about data [1], i.e., metadata provides information about one or more aspects of the data. This layered structure enhances the ability to capture subtle data relationships, thereby improving data management and analysis. To work effectively with metadata, organisations can use several key tools and technologies, including taxonomies, ontologies and semantics.

Ontologies and taxonomies are key tools used by researchers to understand and retrieve large quantities of scientific and engineering data. However, the management and application of ontologies themselves can prove to be a daunting task. Although similar in function, ontologies and taxonomies differ in complexity. Taxonomies have a hierarchical structure and use only parent-child relationships, while ontologies are considerably more complicated [2] [3]. In simple terms, an ontology represents the structured and formal knowledge related to a specific domain. The semantic system uses clear and understandable representations of concepts, relationships, and rules to develop knowledge. It is not possible to rely entirely on database programmers or data engineers to build a system, that considers target applications, such as materials or production technologies. They lack domain-specific knowledge, which is fundamental for characterising the associations between concepts. Therefore, to acquire domain knowledge, it is necessary to seek guidance from various domain experts [4].

Over the past decade, Machine Learning (ML) has gained significance in the fields of materials engineering. ML is a subset of the broader category of Artificial Intelligence (AI) that involves the development of algorithms and models that enable systems to learn and improve from data without explicit programming. AI encompasses a wider range of technologies integrated into a system that aims to facilitate reasoning, learning, and problem solving to address complex problems.

ML algorithms analyse vast amounts of data, extract insights, and use them to inform decision-making [5] to detect and extrapolate patterns. ML is becoming increasingly popular worldwide owing to the growing demand for data analysis solutions [6]. However, they also require large amounts of data, which may not be meaningful in many areas, partly due to the need for elaborate large-scale laboratory tests. There has been an increase in the utilisation of ML methodologies in materials science research [7]. Research suggests that the limited practicality of AI in certain domain-specific contexts, partially because of the need for elaborate laboratory tests on a large scale, is a significant obstacle to its application.

Focusing on industrial requirements, we developed a novel approach for the applicability of AI techniques, termed Usable Artificial Intelligence (UAI) [8]. At present, data-driven, machine learning, and artificial intelligence methods are not fully utilised to solve the associated technical challenges, especially in industrial applications, despite versatile development progress. This is mainly due to the limited practicability of AI solutions. Technical practitioners frequently depend on interdisciplinary collaboration with data science specialists to fully exploit the capabilities of AI methods [9]. In our work, a flexible, tractable, scalable, and adaptable technique

 ISBN: 978-1-68558-157-2

for constructing anticipatory models has been introduced [8] and it is demonstrated on a use-cases [9].

Multi-Task Learning (MTL) methodology, which is novel in materials informatics, can be utilised for example to learn and forecast various polymer characteristics simultaneously, efficiently and effectively [10]. MTL is a machine learning approach, in which multiple tasks are trained simultaneously, optimising multiple loss functions simultaneously. Rather than training independent models for each task, we allow a single model to learn to complete all tasks at once. In this process, the model uses all available data across different tasks to learn generalised representations of data that are useful in multiple contexts [11]. For example, multitask models can be utilised to overcome the data scarcity in polymer datasets. This approach is expected to become the preferred technique for training materials data [10].

Additionally, in other fields, existing predictive models struggle to capture the complex relationships between mechanical characteristics and behaviour. These studies used ML to predict the mechanical properties of carbon nanotube-reinforced cement composites [12]. Successful training, validation, and testing of ML and Deep Learning (DL) models require significant amounts of relevant data [13].

According to a survey, data scientists spend most of their time cleaning and organising data (60 %), collecting datasets (19 %), and mining data for patterns (9 %). Messy data are by far the most time-consuming aspect of typical data scientist's workflow [14].

There is an urgent need to enhance the infrastructure that facilitates the reuse of educational data [15]. In addition, it is necessary to consider that data governance is fundamental for other activities besides data within any Information Technology (IT) establishment.

Through analysis, it can be concluded that the difficulty of identifying, collecting, retaining, and granting access to all relevant data for organisations at an acceptable cost is significant. Data integration is a long-standing issue in data management, and the above observations attest to its continuing importance [16]. It is important to tap into the full potential of data to create added value. This provides new insights and justifies the costly initially data collection.

A. *Motivation*

In recent years, data-driven methods have significantly improved various engineering tasks by providing valuable insights, pattern recognition, and identification of the underlying relationships in complex datasets. This has led to remarkable progress and numerous potential data-driven applications, including production engineering [17] and materials science [18]. However, the availability and usability of underlying data are fundamental to the application of these methods.

In engineering, proper documentation of research data is highly significant as experiments are often complex, intricate and elaborate. Inadequate data documentation can lead to the misinterpretation of experiments by other researchers and/or unnecessary repetition of already completed experiments, with data that are publicly accessible in repositories. High-quality data documentation is crucial for researchers seeking to understand the relationships among the processes, structures, and properties of manufactured components. This is sought and increasingly demanded by public project sponsors, such as the German Research Foundation (DFG).

Multi-stage manufacturing in the process chains is common for many products. Cross-process data analysis can be used to identify relationships in process chains. A prerequisite for this is that an evaluable, comprehensive and well-documented global dataset is available [19]–[23]. Nevertheless, acquiring such a dataset across process boundaries presents a formidable obstacle, owing to the distinct handling of individual process steps by different partners.

To facilitate cross-platform implementation of AI models, a digitally analysable, system-independent dataset representation is necessary. These datasets can be combined to form a unique dataset that represents different system properties, ultimately enabling holistic data-driven modelling, for example, through MTL or transfer learning. This will enable the harmonisation of workflows across diverse domains, thereby facilitating communication between areas of expertise or specialists themselves. An overarching strategy is key to aligning different approaches ensuring tha the experimental data are reusable without modification.

We propose a strategy that allows specialists to focus on data provision by reducing routine activities rather than aligning with similar groups. This strategy enables researchers to focus on their experiments and research questions. The objective was to document research data across process boundaries, thereby enabling researchers to maintain their perspectives during data preparation and documentation. Metadata schemas with synonyms grounded on ontologies or taxonomies guarantee that research data that is understandable, usable for further analyses, interoperable across laboratory boundaries, replicable at a qualitative level, complete, and of superior quality.

In conclusion, the main motive of our research is to support data-driven analysis and modelling, including comparisons across laboratory boundaries. Datasets from different laboratories are to be merged, for example, for round robin-tests, etc. For multi-stage process chains this reveals overarching correlations in the overall dataset. Taking into account the FAIR principles [15], i.e., Findability (F), Accessibility (A), Interoperability (I), and Reusability (R), third party researchers will be able to understand and analyse datasets from disciplines that are unfamiliar to them for their own research questions.

B. *Challenges*

Effectively documenting data across processes and laboratory boundaries presents a key challenge. At the heart of these challenges is the need for data to comply with the guidelines of good scientific practice and the FAIR principles. Each domain possesses its own technical language and unique working culture that needs to be integrated, while allowing researchers to retain their languages. A clear and concise example of the objective can be illustrated by the symbols and units of measurement used for the tensile strength in tensile tests. The standards differ in the symbols for the tensile strength, which are used for different materials. The ISO 1920-4 standard for the tensile strength of concrete specifies f_{ct} as a symbol for tensile strength, ISO 6892-1 for metals specifies R_m, ISO 527-1 for plastics specifies σ_m, and the RILEM TC 232-TDT [24] technical guideline for textile-reinforced plastics specifies σ_{cu} as a symbol for tensile strength. In addition, the frequently used units of measurement differ, which are MPa (Megapascal) and GPa (Gigapascal).

A common technical language that allows researchers from different domains to communicate effectively is relevant, without the necessity for a uniform overarching technical language across all laboratories. Local technical terminology should be compatible without the need for a uniform overarching technical language across all laboratories. This can improve recognition and reduce expenses, furthermore, it is necessary for interoperability. We are not aware of any other method for integrating data records. The completeness of reporting is also critical. Researchers from various disciplines consider – due to their specific research questions – different quantities to be significant, leading to incomplete and inconsistent data documentation across process and laboratory boundaries. Therefore, complete data documentation is relevant for the subsequent use of data by third parties and adherence to the principles of good scientific practice is indispensable to ensure accuracy. Sometimes, if experiments were carried out a long time ago, it is not always possible to remember the details, especially because staff turnover in the research sector is very high.

C. Aim

The main objective of this paper is to provide a workable approach towards synchronised documentation of research data within the engineering sector across various phases while meeting all the requirements regarding the FAIR principles. It is specifically geared towards enabling researchers to maintain their domain-specific perspective during the data preparation and documentation phases. The goal is to develop a methodology that is achievable, extensible, and effective for promoting cross-platform functionality. The deployment of AI models is facilitated through a digitally analysable, system-independent presentation of training datasets that enable cross-process data analysis.

D. Contribution

We present a solution concept in which research data can be documented based on a subject-specific ontology. The feasibility of the concept is illustrated using an example of the documentation of compression tests as part of the joint GRK2250 project [25]. The concept is largely independent of the above project and can therefore be easily transferred to other collaborative projects. Subsequent data-driven modelling is outside the scope of this study.

To conclude, our strategy supports a paradigm shift from more or less subjectively designed individualistic conceptions to the handling of research data towards objectively established harmonised solutions. The motivation for this work is the importance of harmonised data preparation and subsequent documentation in the engineering field. The impetus for this work comes from recognising the fundamental significance of standardising data preparation and subsequent documentation in the engineering domain.

E. Paper organisation

The remainder of the paper is structured as follows: Section II provides an overview of existing work related to the described problem. A description of our strategy is presented in Section III. In Section IV, the strategy is illustrated for an Interdisciplinary Research Training Group, in which researchers from more than 10 different departments are investigating mineral-bonded composites for improved structural impact safety.

The presentation of the main results and discussions based on these results constitute the content of Section V. Section VI summarises our contributions and draws perspectives for future work.

II. Related work

This section offers an overview of the existing approaches to metadata schemes for research software. Whereas some publications focus specifically on metadata, others introduce software ontologies that can serve as a vocabulary for research software. A recent summary [26] of existing approaches to metadata schemes for research software includes DataCite [27], CodeMeta [28], and EngMeta [29], etc. The international consortium DataCite was founded in late 2009, to address the ever-increasing amount of digital research data. The objectives of the consortium include promoting the acceptance of research data to facilitate data archiving and enabling future studies to verify and repurpose the results.

CodeMeta is a community driven metadata standard for research software based on the schema.org. Various crosswalks to other metadata schemes already exist. CodeMeta contains multiple elements, some focusing on technical details, such as file size or supported operating systems and others including administrative information, such as licenses. The metadata standard does not have mandatory elements. It supports the use of uniform research identifiers for authors and contributors as well as licenses. Content-specific metadata are limited to the application categories and keywords.

EngMeta is an XML-based formal definition of the information required to find, understand, reproduce, and reuse data from engineering disciplines [29]. It uses a metadata schema for the description of engineering research data and the documentation of the entire research process, including the people, software, instruments and computing environment involved, as well as the methods used and their parameters [30] [31].

In general, the more precise the data documentation that can model a specialist area, the more suitable it is. This means that general ontologies, on which knowledge databases, such as WikiData [32] or DBpedia [33] are based, are only suitable to a limited extent for use in highly specialised fields of application, such as additive manufacturing. The European Materials Modelling Ontology (EMMO) [34] is an approach for standardising technical terms in applied sciences, particularly in materials science. It can be used to model experiments and simulations.

OntoSoft [35] captured scientific software metadata, and expanded them using machine-readable descriptions of the expected content of the inputs and outputs of the software. The EDAM ontology contributes to open science by enabling the semantic annotation of processed data, thus making the data more understandable, findable, and comparable [36]. Software Ontology (SWO) [37] has been developed to extend the EDAM ontology to describe software in this research area [38]. SWO includes licences, programming languages, and data formats as taxonomies. In contrast to OntoSoft, the use of taxonomies improves the usability of semantic web applications and links [39].

Several universal metadata standards are available in the literature, and metadata schemes have been used in online retail for over a decade. More than 100 metadata standards were visualised in [40]. Additionally, metadata standards related to engineering domains are available, such as EngMeta [31]. However, metadata templates for specific experiments are lacking. Even in experiments standardised according to the German industry standard (DIN), there is no guidance on what metadata should be stored. The standards focus on the execution of experiments rather than on managing the data collected during the experiments.

There are already a number of different Research Data Infrastructures (RDIs) for collaborative projects in the fields of engineering sciences, such as the Karlsruhe Digital Infrastructure for Materials Science (Kadi4Mat) [41]. The software includes many features for data management and collaborative work in joint projects (including web-based access, fine-grained role management, creation of reproducible workflows, and publication of research data). This basic functionality can be easily extended using plug-ins. Metadata schemas with key-value pairs are commonly used to document data in a machine-readable form [42], usually stored in XML or JSON format.

III. STRATEGY

We begin by examining some basic concepts. There is a continuous need to enhance infrastructure that supports the reuse of research data. To this end, a concise and measurable set of principles has been developed to govern the reusability of research data, known as FAIR Data Principles [15]. These foundational principles, namely Findability (F), Accessibility (A), Interoperability (I), and Reusability (R) are guidelines for those wishing to improve the quality of their data. However, they also have wider applicability, as researchers who wish to share and reuse their data can benefit from them. They can also be used by professional data publishers, who offer services and expertise in this area.

It is important to note that these values did not end. Data sharing and collaboration are important elements of scientific research. Researchers must share and collaborate in order to broaden their knowledge and perspectives. They must rely on each other's data and interpretations without bias. However, researchers must always maintain objectivity and balance when using technical terminology and adhering to conventional academic structures. It should be applied not only to data in the traditional sense, but also to the algorithms, tools and workflows that produce it. The emphasis on fairness, which applies to both human and machine activities is the focus of the FAIR Guiding Principles. Good data management is not an end in itself, but rather the key to knowledge discovery and innovation, and to the subsequent integration and reuse of data and knowledge by the community after data publication [15].

An ontology describes the structure of data, including classes, properties, and relationships within a particular field of knowledge, ensuring consistency and understanding of the data model. Description Logics (DLs) provides fundamental concepts and information about this family of logic, which has become increasingly important in recent years as the formal basis for most contemporary applications. The Web Ontology Language (OWL) family includes expressive ontology languages [43] [44] An ontology is expressed using OWL 2 QL (query logic). A query is expressed using SPARQL, a mapping is expressed using R2RML, whilst SPARQL is the standard query language for RDF data. Ontologies offer several advantages over the relational and object models. They allow a strict definition of conceptual schemas and enable systems to understand the semantics of the data [45].

The proposed approach is based on an information structure that includes keys (term classes) and values (concrete expressions of terms). The keys are derived from ontology, whereas the values reflect the potential forms of metadata, such as the value ranges for numerical properties, etc. Examples of process-specific metadata include characteristics such as the ultimate tensile strength in tensile tests.

- Schema Structure: Metadata schemas are created for all investigations, which serve as a template for recording the metadata. These schemas are divided into chapters, whereby some chapters are the same for all investigations and other chapters are adapted to the respective investigation.
- Ontologies: Key names are generated based on domain-specific ontology. Key names are used in process-specific metadata schemas to document the research data. Keys and values are filled with specific terms from the taxonomy and ontology to obtain concrete values. Specific ontologies and taxonomies exist for each domain or process.
- Thesaurus: The creation of a common language that encompasses all processes of the involved domains is necessary. A thesaurus is a domain-specific dictionary of synonyms that lists technical terms that have the same meaning or are similar to the technical terms. This helps ensure that an individual researcher can maintain his familiar terminology, even in an interdisciplinary

 ISBN: 978-1-68558-157-2

environment. Constant readjustment of new partners is unnecessary. Nevertheless, it is ensured that the data remain comprehensible and, therefore, usable for other researchers, both internally and externally. The researcher's own language is linked to the master language through synonyms.

The metadata schema has a clearly defined structure. The static metadata chapters have the same keys for all investigations and are mandatory. Dynamic metadata chapters must be redefined for each investigation along with domain experts to ensure the reproducibility of the investigation.

Key names are created using a domain-specific ontology and then used in metadata schemas that are specific to the research process. To ensure specific, concrete values, specific terms obtained from the ontology are used to populate the keys and values of metadata schemas. A universal language that covers all relevant domains is essential. Each domain has its own specific ontologies that must be merged into a global ontology. Domain-specific dictionaries of synonyms (thesauri) enable researchers to use familiar terminology even in interdisciplinary settings, as they list technical terms that have the same or similar meanings. This eliminates the need to constantly adapt to new partners.

The resulting data can easily be shared with other researchers. Thesauri were included to allow the laboratories to retain their preferred terminology. In addition, researchers complete metadata schemas in their native languages. Nevertheless, the information is presented objectively to facilitate understanding by researchers inside and outside the laboratory. The researcher's own language is linked to the master language using synonyms.

IV. Use cases: An excerpt

In order to exemplify our concept as outlined in Section III, we present a research case, i.e., the Interdisciplinary Research Training Group "Graduiertenkolleg 2250" (GRK 2250) [9], [25], which is dedicated to the investigation of mineral-bonded composites for improved structural impact safety [8]. This project provides an overview of common procedures, such as experiments, tests, numerical simulations, and manufacturing. Another special feature of the project is that three cohorts are planned to be established, with each cohort being worked on by a different team. This makes data transfer very important.

The proposed solution is based on an objective information structure known as metadata schema. This schema comprises keys representing term classes and values that express the meaning of the term more concretely.

GRK 2250 was established in 2017 and is currently funded by the German Research Foundation (DFG). Researchers from nine different departments and four faculties at TU Dresden and the Leibniz Institute of Polymer Research, IPF Dresden, were involved in the program organised in three consecutive three-year periods. Each cohort comprised 12 researchers representing the main research profiles, such as textile technology, polymer and material sciences, construction materials, structural engineering, continuum mechanics, numerical modelling, 3D optical monitoring techniques, sustainability, resilience, and machine learning. The scope of research ranges from the microscale to the structural scale and includes experimental, numerical and data-based investigations. Examples of investigations at the microscale include fibre pull-out tests and corresponding simulations. At the structural level, for example, drop tower tests are carried out in a 10-metre drop tower with plates measuring 1.5 metres by 1.5 metres by 30 centimetres and accompanied by corresponding simulations.

The current status of research data management within GRK 2250 shows considerable variation in the amount of data, ranging from a few megabytes to several hundred megabytes per experiment. The cumulative data volume of 3 terabytes was stored from 15 test systems in six different laboratories. Each test system was conducted for 20 to 300 experiments. To support this diverse dataset, the research data infrastructure consisted of a shared drive that could be accessed by all project partners and Excel spreadsheets dedicated to the documentation of the research data. The data documentation workflow involves manually storing the research data in appropriate folders within the group drive. Researchers manually entered the metadata into an Excel spreadsheet, which was automatically named according to a specific scheme. The completed Excel file was then saved to the corresponding research data folder on a shared drive.

The availability of appropriate data in materials science has a major impact on the performance of the applied AI models [6] [7]. Therefore, data management is particularly important to the usability of AI models. To consider and analyse cross-process relationships, a global view of the dataset in an analysable form is required. This requires well-documented data, which can be combined into a global dataset.

Interdisciplinary research networks bring researchers together with different specialised ontologies. To work together effectively in this case, i.e., in order to be able to understand the data of the research partners, a common ontology is required. For example, a measure for the amount of textile per unit for textile reinforcement of concrete is typical weight per unit area (unit: g/m^2) in textile engineering, whereas, in civil engineering, it is the cross-sectional area per linear metre (unit: m^2/m).

The current solution envisages a top-down approach, i.e., there is a group/person responsible for metadata management within the research network. This defined the standard ontology used in the research network. Enforcing the use of a specified standard ontology may lead to poor overall acceptance by participants. This is because the ontology may overwrite terms that have been established in their domain for years. It is important to consider the difficulty of pushing changes through.

The participants/working groups may be involved in other projects or networks in addition to the specific research network. These projects/networks may have agreed upon a different standard ontology. Therefore, conflicts may arise, since researchers must constantly switch between those ontologies.

As already mentioned, the acceptance of specialists of new ontologies is poor, as acceptance decreases the risk of errors increases. Our response to this wrong-headed development are summarised in Figure 1. This example illustrates the basic

 ISBN: 978-1-68558-157-2

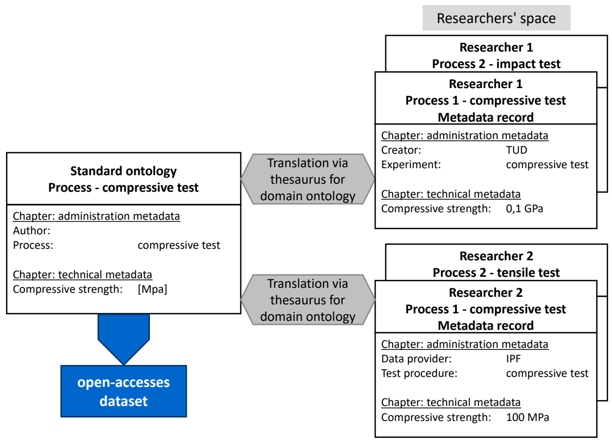

Figure 1. Symbolic representation of the metadata management workflow accessing a mapping thesaurus.

difficulty to overcome the different representation of metadata. As depicted, researchers from different institutes use different terminology for “Author” and different units (MPa/GPa) to record the compression strength, hence the values are also different. The basic idea is to introduce an intermediate layer (thesaurus) as the translation layer. This allows each researcher to use his own “laboratory ontology” or “researcher ontology”. This is then used for the overall project, and translated into the “standard ontology” of the research network.

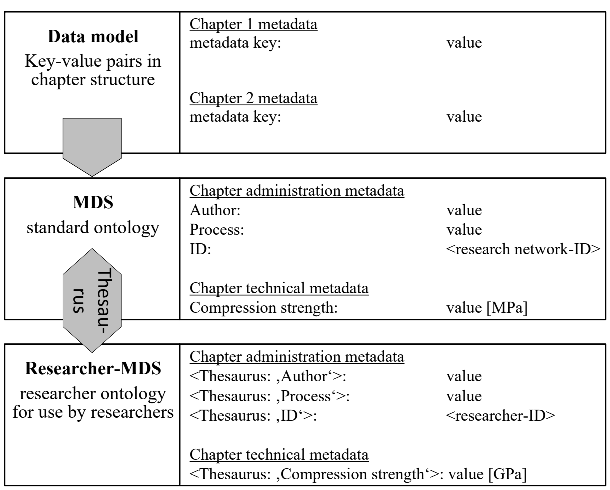

Figure 2. Extended symbolic representation of the metadata management workflow

The data model illustrated in Figure 2 as the general structure of a Metadata Schema (MDS) is divided into several chapters. These chapters contain metadata-key/metadata-value pairs that represent the metadata itself, a method that can generally be used To investigate specific chapters, the keys must be redefined for each investigation. As exemplified, MDS is divided into administrative and technical metadata.

The technical metadata schema (investigation specific) has to be adapted to the specific use case of data generation, whereas the administrative metadata (investigation-independent) has general validity.

There should be a standard ontology agreed upon in the research network. Accordingly, this ontology is also used when data are made publicly available. Standards already in use should be used to facilitate communication with external parties, for example, EngMeta [31] or the European Materials Modelling Ontology (EMMO) [46].

The Metadata Schema is a general structure that is then applied to the specific experiments/data. It can be partitioned into “static” and “dynamic” metadata, the static metadata is similar for all data (i.e., general data, e.g., author, date, etc.), whereas, the dynamic metadata is experiment-specific (i.e., specifically adapted to the experiment, e.g., temperature, test speed, etc.). Adjustment of dynamic metadata should be performed together with domain experts. The MDS is linked to the standard ontology via a translation layer. This allows the researcher to view the schema in their usual domain-specific language. To share the data, the metadata are translated back into a standard ontology and can therefore be understood by everyone.

An initial solution proposal and translation of the ontology has already been published [9]. Furthermore, it explains in more detail what data and metadata management must be able to do in order to support data-driven applications. Figure 3 presents an overview of the FAIR Data Principles proposed by Wilkinson. These principles were expanded to include the principle of “usability” to ensure their practical implementation.

For the extension of the FAIR principles of Wilkinson et al. [15] to “usable FAIR” as depicted in Figure 3, the GRK is working with the company Symate [47] to extend their software Detact. Detact is a cloud-based software for collecting data from various sources along the process chains for subsequent automated data analysis. As stated on the home page from Detact: “Now we are able to merge the material data across disciplines and from different data sources. This leads us to completely new industrial planning and control systems in the sense of the Fourth Industrial Revolution (‘Industry 4.0’)” (Leibniz IPF) [48]. A major aspect of this study was the development of a process modeller.

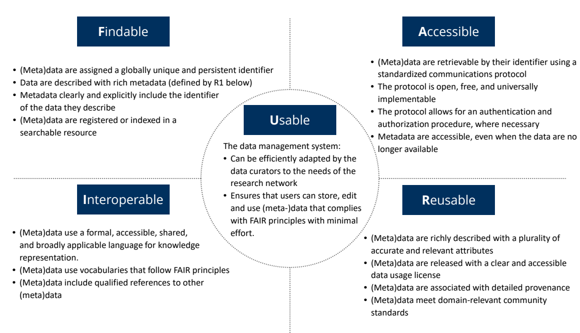

Figure 3. Extension of the FAIR principles of Wilkinson et al. [15] to “usable FAIR”.

Figure 4 shows the planned process modeller used to generate the metadata. In addition to conducting the experiments, the researcher created a process data model for the entire experimental process chain. This model consists of four blocks

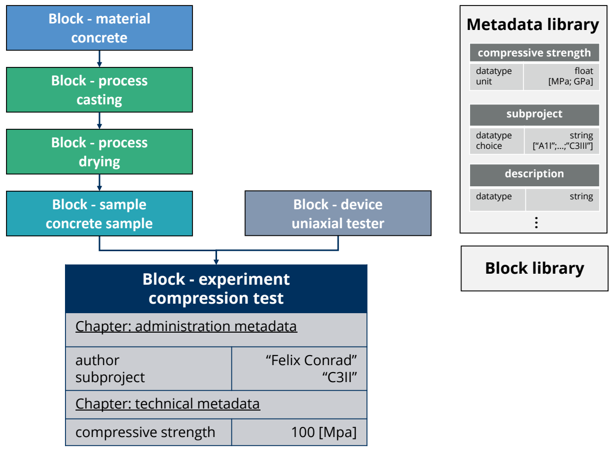

Figure 4. Schematic representation of the process modeller diagram for the use case "uniaxial tensile test of a textile-reinforced concrete test specimen".

(Material, Processes, Devices, and Experiment). The metadata associated with each block can be recorded within that block. All the blocks had the same metadata structure, as depicted in our example, and both administrative and technical metadata were included.

In this particular instance (depicted in Figure 4), the compression test of a concrete test specimen was modelled. The material used is precisely defined in the first block "material concrete", in which the exact composition is recorded. This is followed by the casting of the material defined in the block "process casting" block and the subsequent drying of the material defined in the block "process drying". These steps lead to the final test specimen. The test is then carried out in a testing machine represented by block "device uniaxial tester". The test procedure can then be recorded in the block "experiment compression test".

Figure 4 shows only the metadata of the Block 'Experiment - compressive test' for the purpose of simplification. However, all other blocks also have metadata in the same structure. The corresponding metadata keys can be added to the blocks from a predefined library, known as the metadata library. The experimental process chain for the concrete compression tests, as shown in Figure 4, is relatively simple and short. In this case, the metadata could be included in a single scheme for the concrete compression test experiment, without the process data model.

In real-life applications, the process chains are often much longer and more complex, as shown, for example, by the work in the research network Research Training Group GRK 2250 "Mineral-bonded composites for enhanced structural impact safety" [25], where the process chain begins with mixing the concrete and fibre production and extends to the reinforcement of an existing component with textile-reinforced concrete. Another example is the Collaborative Research Centre SFB 639 "Textile-reinforced composite components for function-integrating multi-material design in complex lightweight applications" [49], where the process chain ranges from the fibre and plastic to the complex function-integrated component. In addition, as shown in the example, the complexity increases very quickly if we use a composite material, such as textile-reinforced concrete (TRC). The block "material concrete" would be accompanied by the block "material - textile", which together would then form the block "material - TRC". The block "material textile" itself can also have upstream blocks, like "process textile manufacturing", "process fibre formation", etc.

This information is required to manage trial data according to FAIR principles. Attempting to fit all of these metadata into a single schema without a process data model will most likely result in important parameters being omitted, hence the need for the process data model. According to the example shown in Figure 4, the drying time of concrete has a significant influence on its strength. If the drying time was not recorded, the data would be essentially unusable for further analysis as a major influence is not recorded.

All individual blocks and overall process data models can be collected and shared in a library, to unify the workflows of the cooperating researchers. Using the process data model, the workflows of other researchers can be easily understood and adopted. Here, the process data model makes communication between the researchers easier. The process data model provides the instruction on how the experiment/entire workflow is to be carried out. Standardisation of workflows is a fundamental step towards sustainable data management and research. Standardised workflows are significant for the comparability of experiments and, thus, for the reusability of the series of experiments. Currently, researchers create their own workflows during their research. Many experiments have not yet been standardised because they require a new type of experimental setup or a new type of material, the production of which is not yet standardised. Even small variations in the manufacturing process can significantly impact the target properties. This makes it difficult to compare/reuse the data. The standardisation of workflows also reduces unnecessary duplicate developments in common process steps.

With the implementation of the process data model in Detact, researchers do not have to scroll through "metadata lists", but can use the graphical version of the process model. Access to Detact can occur via a web browser and metadata can be recorded directly in the laboratory. The comment function can be used to mark deviations that would otherwise not be recorded (e.g., "Concrete stickier than usual today"). Using the process model, metadata can be captured quickly and easily and subsequently recorded. A new test can be initialised with default values and can be easily adapted. Experience has shown that metadata are currently insufficiently (not completely) captured. The perceived costs for a complete recording are too high (see point "Usability") and can be significantly reduced with the "process model".

Based on this approach, it is possible to merge different experiments with the corresponding data sets into a global data set. The merging of the three different experiments into one dataset is sketched in Figure 5. The example above was adapted from the research project GRK 2250 [25]. The purpose

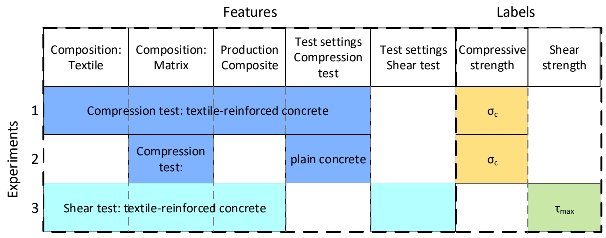

Figure 5. Schematic illustration of a section of the overall dataset of the research network.

of creating a combined dataset is to gain a comprehensive understanding of the material from a data perspective, such that in order to map the material behaviour, data-driven models can be trained on the basis of this broad dataset. To achieve this, the data from various material tests and, thus, properties are to be combined. If the same material is analysed in different tests, a corresponding dataset can be created, as outlined in Figure 5.

The benefit of using such a dataset is the wider range of information available for modelling, and the ability to determine the interactions between multiple influencing parameters. This allows the creation of models that can simultaneously map several material properties.

The importance of the metadata management methods outlined in this section is obvious, as experiments in research networks are frequently conducted by various researchers in different laboratories. Therefore, merging experimental data into a single dataset is feasible using a collaborative data management approach. The following three experiments were conducted:

1) Determination of compressive strength of textile reinforced concrete.
2) Determination of compressive strength of unreinforced concrete.
3) Determination of the tensile strength of textile reinforced concrete.

Each test had its own data space, which sometimes overlapped. The data available for each experiment are indicated by the coloured boxes. The absence of colour markings indicates the absence of the specific data. For example, unreinforced concrete does not have data on textile reinforcement, as this is not present and is outside the scope of the investigation. In this context, features refer to descriptive elements of the experiment, whereas labels denote the outcome of the experiment. Terms features and labels were chosen because they are typical terms in the field of machine learning. In detail, the representation for the experiments is as follows:

- Experiments 1 and 3 share the same features to describe textile-reinforced concrete, but have different features to describe the experiment itself, as there are different tests to determine the compressive and shear strength. In addition, different properties were determined, which are labelled here.
- Experiments 1 and 2 have the same features to describe the experiment and also have the same label (the same material property that is determined). However, in experiment 2 the features that are related to textile reinforcement are missing, since they do not apply to the experiment.
- Experiments 2 and 3 shared the same features for describing the concrete matrix. They have different features to describe different tests and, accordingly, different labels. In experiment 2, the features related to the textile reinforcement were missing.

V. OUTLINE OF THE RESULTS

The proposed strategy helps bridge different domain-specific languages and working cultures by providing a common language that all researchers and engineers from different domains can understand. This is achieved by using metadata. In particular, the metadata model helps unify physical units and terms. Data are stored and documented such that data from different processes along a process chain can be merged, resulting in a single overall dataset. Therefore, cross-process data analysis methods can be applied. The solution approach allows merging of research data in the following ways:

- Merging data from similar processes provided by different institutions or fields.
- Merging data from different processes along a process chain.

A multi-level metadata model connects different domain-specific languages by defining a common set of concepts and relationships that can be used in different domains. The model provides a method to manage metadata by defining a set of rules on how metadata should be structured and stored.

The drawbacks of the metadata model are as follows:

- The creation of a metadata template relies on the assistance of metadata experts, whereas the goal should be for researchers or domain experts to use/create it independently.
- The outcome depends on the domain experts. In the end, the solution approach allows for the creation of global datasets in an analysable manner. In this way, the interoperability and collaboration among different research groups in the engineering domain will be improved.
- The common language was created by surveying data providers to establish a shared technical vocabulary.

The proposed model can be used as a framework for managing digital objects in other research domains, such as the social sciences and natural sciences. Additionally, further research can be conducted to explore how this metamodel can be integrated into existing research data management systems or how it can be improved to better meet the needs of different users. Overall, the model provides a promising approach for addressing the challenges of research data management and improving collaboration among researchers and engineers from different domains. This solution ensures that the data are captured in a comprehensible manner through clear documentation, thus, it can be understood by other researchers too. The interoperability of data across laboratory boundaries is ensured by the proper identity management of components, processes, and machines across these boundaries.

This makes the data available for subsequent data-driven analysis across the laboratory and process boundaries. The analysis results based on the documented research data can be reproduced at a qualitatively high level owing to the detailed data documentation.

VI. Conclusion and Future Research Perspective

The proposed strategy helps bypass different working cultures by providing a harmonised approach that all researchers and engineers from different domains can understand. This is achieved by using metadata enhanced by the development of adequate ontologies. In particular, the metadata model helps to store and document data, such that data from different processes along a process chain can be merged, enabling cross-process data analysis methods.

As a use case approach, this article also summarises the existing requirements of the GRK 2250 joint project for practicable research data management and presents a solution concept for the investigation of mineral-bonded composites within the GRK 2250 project. The concept is largely independent of GRK 2250 and can therefore be easily transferred to other collaborative projects.

Experience has shown that researchers need support in setting up a structured process data model. It is difficult for them to identify all the metadata that needs to be recorded so that the experiment is documented in a repeatable manner. As a result, it can happen that important influencing factors in the experiments were not recorded and the generated data is hardly reusable. The structured process data model (as exemplified in Figure 4) is intended to help identify all necessary steps and influences.

To consider and analyse cross-process relationships, a global view of the dataset in an analysable form is required. This requires well documented data that can be combined into global datasets [6] [7], as subsequent data-driven modelling is not part of this study.

The aspect of "usability" is not fully covered in this paper. It is still an open question as to whether and how it can be satisfactorily solved, since setting up a metadata management system that covers the FAIR principles is one thing, but in the end it only works if all participants are on the board. According to the experience gained through the use cases presented, it often fails in the end because there is an overhead due to the metadata management system for the individual researcher.

The aim of the Industrial Ontologies Foundry (IOF) initiative [50], is similar to that proposed for the OBO Foundry (for biomedicine) [51]. In both cases, commitment to a standard upper-level ontology plays a key role in supporting harmonisation. This upper-level ontology is termed Basic Formal Ontology (BFO) [52]. Consideration of how the current effort relates to this wider effort to curate and facilitate access to industrial ontologies might be a useful focus area for future research.

Acknowledgement

This research was partially funded by the German Research Foundation within the Research Training Group GRK2250/2 - Project C3 (grant number 287321140), by the Sächsische Aufbaubank (SAB), through the European Regional Development Fund (ERDF) and co-financed with tax revenue based on the budget approved by the parliament of the Free State of Saxony, Germany within the research project "AMTwin" (grant number 100373343), by the German Federal Ministry for Economic Affairs and Climate Protection (BMWK) based on the decisions made by the German Bundestag within the joint research projects "SWaT" (grant number 20M2112F) and "LaSt" (grant number 20M2118F), and by the BMWK in the funding guideline "Digitization of the vehicle manufacturers and supplier industry" in the funding framework "Future investments in vehicle manufacturers and the supplier industry", financed by the European Union, and supervised by the project sponsor VDI Technologiezentrum GmbH within the joint research project "Werk 4.0" (grant number 13IK022K). We thank the anonymous reviewers for their valuable feedback, which has significantly enhanced the quality of this paper.

References

[1] J. Riley, "Understanding metadata," *Washington DC, United States: National Information Standards Organization (http://www. niso. org/publications/press/UnderstandingMetadata. pdf)*, vol. 23, pp. 7–10, 2017, retrieved: April 2024.

[2] M. Tainter. (2020) What is the difference between a taxonomy and an ontology? retrieved: April 2024. [Online]. Available: https://www.copyright.com/blog/taxonomy-vs-ontology/

[3] JANZZ.technology. (2019) Ontologies and the Semantic Web. retrieved: April 2024. [Online]. Available: https://janzz.technology/ontology-and-taxonomy-stop-comparing-things-that-are-incomparable/

[4] I. Horrocks. (2008) Ontologies and the Semantic Web. retrieved: April 2024. [Online]. Available: http://www.cs.ox.ac.uk/ian.horrocks/Publications/download/2008/Horr08a.pdf

[5] Google Cloud. (2023) Artificial intelligence (AI) vs. machine learning (ML). retrieved: April 2024. [Online]. Available: https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning

[6] F. Conrad, M. Mälzer, M. Schwarzenberger, H. Wiemer, and S. Ihlenfeldt, "Benchmarking AutoML for regression tasks on small tabular data in materials design," *Scientific Reports*, vol. 12, no. 1, p. 19350, 2022, number: 1 Publisher: Nature Publishing Group, retrieved: April 2024. [Online]. Available: https://www.nature.com/articles/s41598-022-23327-1

[7] Y. Zhang and C. Ling, "A strategy to apply machine learning to small datasets in materials science," *npj Computational Materials*, vol. 4, no. 1, pp. 1–8, 2018, number: 1 Publisher: Nature Publishing Group, retrieved: April 2024. [Online]. Available: https://www.nature.com/articles/s41524-018-0081-z

[8] H. Wiemer *et al.*, "Need for uai–anatomy of the paradigm of usable artificial intelligence for domain-specific ai applicability," *Multimodal Technologies and Interaction*, vol. 7, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2414-4088/7/3/27

[9] ——, "Illustration of the usable AI paradigm in production-engineering implementation settings," in *Artificial Intelligence in HCI*, ser. Lecture Notes in Computer Science, H. Degen and S. Ntoa, Eds. Springer Nature Switzerland, 2023, pp. 640–661, retrieved: April 2024.

[10] C. Kuenneth, A. C. Rajan, H. Tran, L. Chen, C. Kim, and R. Ramprasad, "Polymer informatics with multi-task learning," *Patterns*, vol. 2, no. 4, p. 100238, 2021, retrieved: April 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389921000581

[11] D. Soni. (2021) Multi-task learning in Machine Learning. retrieved: April 2024. [Online]. Available: https://towardsdatascience.com/multi-task-learning-in-machine-learning-20a37c796c9c

[12] J. Huang, J. Liew, and K. Liew, "Data-driven machine learning approach for exploring and assessing mechanical properties of carbon nanotube-reinforced cement composites," *Composite Structures*, vol. 267, p. 113917, 2021, retrieved: April 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0263822321003779

 ISBN: 978-1-68558-157-2

[13] F. Conrad, E. Boos, M. Mälzer, H. Wiemer, and S. Ihlenfeldt, "Impact of data sampling on performance and robustness of machine learning models in production engineering," in *Production at the Leading Edge of Technology*, ser. Lecture Notes in Production Engineering, M. Liewald, A. Verl, T. Bauernhansl, and H.-C. Möhring, Eds. Springer International Publishing, 2023, pp. 463–472, retrieved: April 2024.

[14] CrowdFlower. (2016) Data Science Report. retrieved: April 2024. [Online]. Available: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

[15] M. D. Wilkinson *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160018, 2016, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/sdata201618

[16] G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati, "Using ontologies for semantic data integration," *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, pp. 187–202, 2018.

[17] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, "Machine learning in manufacturing: advantages, challenges, and applications," *Production & Manufacturing Research*, vol. 4, no. 1, pp. 23–45, 2016.

[18] D. Morgan and R. Jacobs, "Opportunities and challenges for machine learning in materials science," *Annual Review of Materials Research*, vol. 50, pp. 71–103, 2020.

[19] N. Angrist, H. A. Patrinos, and M. Schlotter, "An expansion of a global data set on educational quality: a focus on achievement in developing countries," *World Bank Policy Research Working Paper*, no. 6536, 2013, retrieved: April 2024. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2295861

[20] S. Lindersson, L. Brandimarte, J. Mård, and G. Di Baldassarre, "A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies," *Wiley Interdisciplinary Reviews: Water*, vol. 7, no. 3, p. e1424, 2020, retrieved: April 2024. [Online]. Available: https://doi.org/10.1002/wat2.1424

[21] J. T. Abatzoglou, S. Z. Dobrowski, S. A. Parks, and K. C. Hegewisch, "Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015," *Scientific data*, vol. 5, no. 1, pp. 1–12, 2018, retrieved: April 2024. [Online]. Available: https://www.nature.com/articles/sdata2017191

[22] K. Großmann, H. Wiemer, and K. K. Großmann, "Methods for modelling and analysing process chains for supporting the development of new technologies," *Procedia Materials Science*, vol. 2, pp. 34–42, 2013, retrieved: April 2024. [Online]. Available: https://doi.org/10.1016/j.mspro.2013.02.005

[23] B. Awiszus *et al.*, "A holistic methodology to evaluate process chains." *Springer Berlin Heidelberg, part of Springer Nature 2022, L. Kroll (Ed.), Multifunctional Lightweight Structures - Resource Efficiency by MERGE of Key Enabling Technologies*, pp. 236–260, 2022, retrieved: April 2024. [Online]. Available: https://doi.org/10.1007/978-3-662-62217-9

[24] RILEM Technical Committee 232-TDT (Wolfgang Brameshuber), "Recommendation of RILEM TC 232-TDT: Test methods and design of textile reinforced concrete," *Materials and Structures*, vol. 49, no. 12, pp. 4923–4927, 2016. [Online]. Available: https://doi.org/10.1617/s11527-016-0839-z

[25] I. Curosu, V. Mechtcherine, M. Hering, and M. Curbach, "Mineral-bonded composites for enhanced structural impact safety–overview of the format, goals and achievements of the research training group grk 2250," *Bayonne, Frankreich*, 2019, retrieved: April 2024.

[26] S. Ferenz and A. Nieße, "Towards improved findability of energy research software by introducing a metadata-based registry," *ing.grid*, 2023.

[27] DataCite Metadata Working Group. (2016) DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. retrieved: April 2024. [Online]. Available: http://doi.org/10.5438/0012

[28] T. Habermann, "Mapping iso 19115-1 geographic metadata standards to codemeta," *PeerJ Computer Science*, vol. 5, p. e174, 2019, retrieved: April 2024. [Online]. Available: https://doi.org/10.7717/peerj-cs.174

[29] Metadata Standards Catalog. (2019) EngMeta. retrieved: April 2024. [Online]. Available: https://rdamsc.bath.ac.uk/msc/m100

[30] Informations- und Kommunikationszentrum der Universität Stuttgart (IZUS). (2019) EngMeta - Beschreibung von Forschungsdaten; Eng: EngMeta - Description of research data. retrieved: April 2024. [Online]. Available: https://www.izus.uni-stuttgart.de/fokus/engmeta/

[31] B. Schembera and D. Iglezakis, "EngMeta: Metadata for Computational Engineering," *International Journal of Metadata, Semantics and Ontologies*, vol. 14, no. 1, pp. 26–38, 2020, retrieved: April 2024. [Online]. Available: https://arxiv.org/pdf/2005.01637

[32] D. Vrandečić and M. Krötzsch, "Wikidata: a free collaborative knowledgebase," *Communications of the ACM*, vol. 57, no. 10, pp. 78–85, 2014, retrieved: April 2024. [Online]. Available: https://doi.org/10.1145/2629489

[33] J. Lehmann *et al.*, "Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia," *Semantic web*, vol. 6, no. 2, pp. 167–195, 2015.

[34] C. to emmo repo. (2024) Elementary multiperspective material ontology (emmo): A top-level ontology for applied sciences. retrieved: April 2024. [Online]. Available: https://github.com/emmo-repo/

[35] Y. Gil, V. Ratnakar, and D. Garijo, "Ontosoft: Capturing scientific software metadata," in *Proceedings of the 8th International Conference on Knowledge Capture*, 2015, pp. 1–4, retrieved: April 2024. [Online]. Available: https://doi.org/10.1145/2815833.2816955

[36] M. Black *et al.* (2021) Edam: The bioscientific data analysis ontology (update 2021)[version 1; not peer reviewed]. retrieved: April 2024. [Online]. Available: https://bora.uib.no/bora-xmlui/handle/11250/2988255

[37] M. Copeland, A. Brown, H. E. Parkinson, R. Stevens, and J. Malone, "The swo project: A case study for applying agile ontology engineering methods for community driven ontologies." *ICBO*, vol. 7, p. 2012, 2012, retrieved: April 2024. [Online]. Available: http://ceur-ws.org/Vol-897/session4-paper20.pdf

[38] J. Malone *et al.*, "The Software Ontology (SWO): a resource for reproducibility in biomedical data analysis, curation and digital preservation," *Journal of biomedical semantics*, vol. 5, no. 1, pp. 1–13, 2014, retrieved: April 2024. [Online]. Available: https://doi.org/10.1186/2041-1480-5-25

[39] A.-L. Lamprecht *et al.*, "Towards fair principles for research software," *Data Science*, vol. 3, no. 1, pp. 37–59, 2020, retrieved: April 2024. [Online]. Available: https://doi.org/10.3233/DS-190026

[40] J. Riley, "Seeing Standards: A Visualization of the Metadata Universe," 2018, retrieved: April 2024. [Online]. Available: https://doi.org/10.5683/SP2/UOHPVH

[41] N. Brandt *et al.*, "Kadi4mat: A research data infrastructure for materials science," *Data Science Journal*, vol. 20, pp. 8–8, 2021, retrieved: April 2024. [Online]. Available: https://doi.org/10.5334/dsj-2021-008

[42] S. Büttner, H.-C. Hobohm, and L. Müller, "Research data management," in *Handbuch Forschungsdatenmanagement.-Hrsg. von Stephan Büttner, Hans-Christoph Hobohm, Lars Müller.-Bad Honnef: Bock u. Herchen, 2011.-ISBN 978-3-88347-283-6*, 2011, retrieved: April 2024. [Online]. Available: https://opus4.kobv.de/opus4-fhpotsdam/files/192/1.1_Research_Data_Management.pdf

[43] S. Rudolph, "Foundations of description logics," in *Reasoning Web International Summer School*. Springer, 2011, pp. 76–136.

[44] F. Baader, *The description logic handbook: Theory, implementation and applications*. Cambridge university press, 2003.

[45] H. Kamal and B. Fouzia, "From relational databases to ontology-based databases," in *International Conference on Enterprise Information Systems*, vol. 2. SCITEPRESS, 2013, pp. 289–297.

[46] G. Goldbeck, E. Ghedini, A. Hashibon, G. Schmitz, and J. Friis, "A reference language and ontology for materials modelling and interoperability," 2019, retrieved: April 2024. [Online]. Available: https://publica.fraunhofer.de/handle/publica/406693

[47] (2024) Home Page Symate GmbH. retrieved: April 2024. [Online]. Available: https://www.symate.de/

[48] A. of detact.com. (2024) Home Page Detact. retrieved: April 2024. [Online]. Available: https://www.detact.com/en/

[49] N. Modler *et al.*, "Novel hybrid yarn textile thermoplastic composites for function-integrating multi-material lightweight design," *Advanced Engineering Materials*, vol. 18, no. 3, pp. 361–368, 2016, retrieved: April 2024. [Online]. Available: https://doi.org/10.1002/adem.201600028

[50] Industrial Ontologies Foundry. retrieved: April 2024. [Online]. Available: https://oagi.org/pages/industrial-ontologies

[51] Open Biological and Biomedical Ontology Foundry Community development of interoperable ontologies for the biological sciences. retrieved: April 2024. [Online]. Available: https://obofoundry.org/

[52] M. Drobnjakovic, B. Kulvatunyou, F. Ameri, C. Will, B. Smith, and A. Jones, "The industrial ontologies foundry (IOF) core ontology," *FOMI 2022: 12th International Workshop on Formal Ontologies meet Industry, September 12-15, 2022, Tarbes, France*, 2022, retrieved: April 2024. [Online]. Available: https://ceur-ws.org/Vol-3240/paper3.pdf