Thèse n° 8062

Analysis and Remediation of Handwriting difficulties

Présentée le 18 septembre 2020

à la Faculté informatique et communications
Laboratoire d'ergonomie éducative
Programme doctoral en robotique, contrôle et systèmes intelligents

pour l'obtention du grade de Docteur ès Sciences

par

Thibault Lucien Christian ASSELBORN

Acceptée sur proposition du jury

Dr D. Gillet, président du jury
Prof. P. Dillenbourg, directeur de thèse
Prof. U. Maurer, rapporteur
Prof. B. Scassellati, rapporteur
Prof. T. Kaser, rapporteuse

École
polytechnique
fédérale
de Lausanne

2020

Acknowledgements

My deepest gratitude goes to my thesis director Professor **Pierre Dillenbourg** who gave me the opportunity to conduct research in the field of education. Pierre is the greatest professor one can ever hope to have, giving so much trust, support and freedom to his students that they are ready to give the best of themselves on a daily basis. Throughout these years, Pierre has always been there to support me, both in my work but also during the darkest hours of my life (I am obviously thinking of the resignation of Thibault Pinot in the tour de France 2019). I am more than happy to continue sharing great moments with him, whether it is in the CHILI lab or on a ski lift.

I would also like to thank my Jury members: Prof. Denis Gillet, Tanja Käser, Urs Maurer and Brian Scassellati for their availability and for taking the time to evaluate my work.

The CHILI team was an incredible source of motivation throughout the years allowing me to meet so many special people, many of them became my friends. I would like to thank all these people without whom this thesis would have not been possible: **Mina Shirvani Boroujeni**, **Brabrabar Bruno** for being an incredible postdoc despite being a Ferrari fan and for correcting all my thesis, **Florence Colomb** the true boss of the lab, **Jules Courtois**, **Louis Faucon** the board game master, **Thomas Gargot** the cracking machine, **Thanasis Hadzilacos** and **Stian Haklev**. Special thank to my friend **Alexis Jacq** for all these zombies killed and all these insane parties, **Wafa Johal** for welcoming me in the lab in 2016 and for being a great supervisor, **Hala Khodr** - my neighbor and friend with a voice being quite a view, **Soheil Kianzad** - my friend that I will visit in Toronto, **Łukasz Kidziński**, **Kevin Gonyop Kim** (seeyoseean), **Aditi Kothiyal**, **Kshitij**, **Lorenzo Lucignano**, **Khalil Mrini**, the super elegant **Jaja Nasir** that always has a war to fight ;), the nicest person in the world **Utku Norman**, **Jennifer Olsen**, **Catharine Oertel**, my friend **Ayberk Özgur**, my friend and happiness manager of the lab **Arzu Güneysu Özgur**, **Dorsa Safaei**, **Joseph Vavalà**, **Elmira Yadollahi**, the great AOE2 player **Ramtin Yazdanian**, **Teresa Yeo** and **Konrad Żołna**.

I wish to truly thanks to my friends, colleagues, past and current members of the Dynamico team: the very talented **Joëlle Aeschlimann**, **Sara Alemanno**, the very kind and brilliant **Laurent Boatto**, my friend and too well payed "stagiaire" **Lucas Burget**, the Swift machine **Tim Duran**, the very kind **Corinne Lebourgeois** for whom I was willing to document everything. Special thanks to one of my oldest and definitely one of my best friends **Anthony Peguet** that unfortunately supports a crappy football team, **Lucas Scuderi**, **Nuria Unterlohner**, **Valentin Viennot**, and obviously my great friend **Sven Viquerat**, son of **Philippe Viquerat**.

I would also like to thanks all the semester and master students that worked with me during this thesis: **Lea Bommottet**, **Cyril Cadoux**, **Mateo Chapatte**, the brilliant (but a bit too stressed) **Laïla El-Hamamsy**, **Nicolas El Maalouly**, **Pawel Golinski**, **Natasha Klingenbrunn** and **Saumya Shah**.

Special thanks to my Badminton team, the LUC Badminton that allows me to release a lot of the pressure accumulated during my PhD and where I am spending some amazing moments.

Thanks to all my childhood friends for all the good moments spent together. Especially thanks to my priceless friends from France, that I have unfortunately not enough time to visit, I am obviously thinking of **Matthieu Guffroy**, **Jeremy Cruz**, **Marin Dauriat**, **Damien Ducret**, **Fabien Michoud**, **Michel Noir**, **Amandine Millet Noir**, **Franck Jacquier**, **Adeline Morel**, **Noemie Greffier** and all the others.

Throughout these years, I receive an incredible support from my friends. I deeply thank my brother in life **Benoit Desbiolles** for being such an amazing person. Benoit is one of the most talented individual I was lucky to meet in my life. He first have this rare and incredible strength to be, whatever the circumstances, always positive and happy which makes him such a wonderful person to hang up with. He is also insanely good in whatever he is doing, first at work collecting the awards, papers and highest grades with this unbearable humility, but also in music ("eternal roses") and of course sport. We had so much insane moments, sharing so much things that I sincerely don't know what I will do when you will be in Boston. Anyway, I sincerely thank my brother Benoit for being such a wonderful person and for having positively enhanced my life in so many different aspects, and I am looking forward to continue this life with you by my side. I would also like to thank my very dear friend **Walid Amanhoud**. Walid is one of the most hardworking person I ever met in my life, and is truly a source of experience, showing me the real value of work. He was a huge source of support during the Math exercise sessions starting from the first year of my Bachelor to the last year of my PhD when I was teaching myself these Math courses. We had so many strong moments together, at EPFL but also in Evian or London. I really hope it will continue like this. Special thank to another of my best friends **Cedric Ponce Cinnamon**. Since the moment we met in this flat with "JO sa sera pas c'tannée", Cedric and I basically shared everything, from countless movie nights, to so many week-ends or holidays together. He is one of the only person in my life with whom I can truly share everything, and for sure, this thesis wouldn't have been possible without the support of this incredible friend. I would also like to thank my friend **Morgane Huard**, for all of our discussion about literature, my friend **Pilar Championne Benavides** for being such a supportive friend, talented pianist and an inexhaustible source of gossips. Thanks to my very dear friend and great orchestrator **Sina Shahmoradi**, always kind and supportive and for having such a different and original view on the world. Thanks to my priceless friend **Le Compte Thomas Debarre** for all our ping-pong, baby-foot (I will have my revenge) and badminton sessions together, but above all for always being such a kind, positive and cheerful friend. Thanks to my friend **Timon tirailleur Ganguillet**, one of the most intelligent and cultured person that I know (and being the only one not understanding at which point he is). Thanks to **Mathilde Fratou Willemin** for her incredible strength to remove all the clouds from my life, for having one of the biggest heart I know and for being such a great personality. Finally thank to **Elena Batzianouli**, for supporting me and for sharing my life during two of the most intense years of my life. Definitely, nothing would have been possible without her.

To finish I would like to thank my familly, starting from my aunts, uncles and cousins. In particular thanks to my great cousin **Chloé Goulon**, that I am not seeing often enough. Thanks to my mom **Sabine Goulon** for the countless sacrifices she made without hesitation for the sole purpose of giving me the opportunity to do what I want in my daily life. Thank ou for being so kind, and for making our house in Taverole the house of happiness. Thanks to my dad **François Asselborn** that pushed me to become better every day, and without whom I would never have had the life I have today. You really are a great source of inspiration, both for your creativity, kindness and inexhaustible energy. Thanks to my brother **Theo Asselborn** for being my role model since my first day. I admire your unique passion, serenity and self confidence. The time spent next to you (way not enough) makes me a better and calmer person. Thanks to **Sophie Asselborn** for being such a lovely sister. I don't have enough words to tell you how much you are a great person and how much you are important for me. Finally I would like to thank my grand father **Christian Goulon**, for his immense kindness, for sharing his great passion for engineering, but above all for making Dombasle such a source of joy and love.

Lausanne, 28 July 2020 T. A.

Abstract

Even with correct training, up to 25% of children never master handwriting like their peers. While research shows a correlation between handwriting difficulties and school failure, these difficulties can also impact children in their self esteem and behavioral development. Since the mastery of handwriting requires a lot of different skills, it is never easy to understand where a given child is facing difficulties, nor how exactly to help him/her overcome them. For this reason, it is of prime importance to detect and understand children's handwriting difficulties the earliest possible in order to propose the most effective remediation possible.

In this thesis, we first introduce a modernize version of the currently adopted handwriting tests, that show clear limitations in the era of digitalization. Indeed, the nature itself of these tests, conducted on paper, restricts them to the analysis of the final static aspect of handwriting. Its dynamics, found to be very important, is therefore hidden and cannot be taken into consideration. For this reason, we designed in collaboration with therapists several features that describe different aspects of handwriting, which are not limited to static but also capture kinematic, pressure and tilt. The designed features have the main advantage to describe very low level aspects of handwriting, which makes them quite independent of the writing content. We verified this hypothesis by giving the proof of concept that our model for automatic detection of handwriting difficulties can be translated from the latin to the the cyrillic alphabet. In the same way, we demonstrated that our model can also, given retraining, be used on paper or directly on digital tablets, like iPads. Finally, we introduced our iPad-based test allowing to extract the multidimensional handwriting profile of the child. This test aims to answer the first of the afore-discussed problems, by allowing to extract the specific strengths and weaknesses of a child, in less than a minute, on different aspects and at different granularities.

The second part of this thesis tackles the problem of designing remediation activities for handwriting difficulties. We designed activities specifically targeting the handwriting aspects identified by the model and obtained a preliminary proof of concept that serious games targeting specific skills of handwriting (e.g. pressure, kinematic, tilt, ...) can have a positive impact on the overall quality of handwriting. Finally the two last chapters of this thesis tackle two corollary, but still crucial questions related to handwriting

remediation. After the integration of these remediation activities in a Child-Robot Interaction scenario, in which the child is the teacher of the robot, we gave the proof of concept of the importance of the design of robot behaviors towards social acceptance with children, something especially important knowing the importance of the child's perception of the robot and the interaction with it in such a scenario. Finally, in the last Chapter, we investigated whether it is possible to "remediate some handwriting difficulties by preventing them", i.e., by supporting pre-school children in the acquisition of the fundamental visuo-motor coordination skills required by handwriting.

Keywords: Handwriting, Machine learning, Robots for Learning, Handwriting analysis, Handwriting remediation.

Résumé

Même avec une formation correcte, jusqu'à 25% des enfants ne maîtrisent pas l'écriture comme ils le devraient. Alors que la recherche montre une corrélation entre les difficultés d'écriture et le niveau scolaire général, ces difficultés peuvent également avoir un impact sur l'estime de soi et le comportement des enfants hors de l'école. La maîtrise de l'écriture requérant des compétences nombreuses et variées, il est donc complexe de comprendre la nature même des difficultés d'écriture qu'un enfant peut présenter et de savoir comment l'aider à les surmonter. Pour cette raison, il est primordial de détecter et comprendre les difficultés d'écriture le plus tôt possible afin de proposer la remédiation la plus efficace possible.

Nous présentons en premier lieu une version modernisée des tests d'écriture actuellement utilisés, présentant des limites évidentes à l'ère de la numérisation. En effet, la nature même de ces tests, réalisés sur papier, les restreint à une analyse statique de l'écriture. Une part importante de l'information, notamment la dynamique de l'écriture, reste donc inaccessible. Pour cette raison, nous avons conçu, en collaboration avec des thérapeutes, plus d'une centaine de caractéristiques décrivant différents aspects de l'écriture, ne se limitant pas au côté statique, mais décrivant également la cinématique, la pression ou l'inclinaison de l'outil scripteur par exemple. Ces caractéristiques ont l'avantage de décrire des aspects bas niveau de l'écriture, ce qui les rendent indépendantes de son contenu. Nous vérifions cette hypothèse en montrant que notre modèle permet la détection des difficultés d'écriture aussi bien avec l'alphabet Latin qu'avec le Cyrillique. De la même manière, nous montrons que notre modèle peut également être utilisé sur papier ou directement sur tablettes numériques (e.g. iPads). Finalement, toutes ces étapes nous ont permis de développer une nouvelle génération de tests d'écriture, basés sur iPads, permettant d'extraire le profil multidimensionnel d'écriture de l'enfant. Ce test permet d'extraire les forces et faiblesses particulières d'un enfant, en moins d'une minute, sur différents aspects et à différentes granularités.

Dans la deuxième partie de cette thèse, nous nous intéressons à la remédiation des problèmes d'écriture. En premier lieu, nous avons conçu plusieurs jeux permettant aux enfants d'entrainer les aspects spécifiques de l'écriture identifiés comme problèmatiques par le modèle. Nous montrons que ces activités ciblant et mobilisant des compétences spécifiques de l'écriture (que ce soit en lien avec la pression, le tremblement, la vitesse,

...) peuvent avoir un impact positif sur la qualité globale de l'écriture. Enfin, les deux derniers chapitres de cette thèse abordent deux questions corollaires, liées à la remédiation de l'écriture en exploitant les capacités que les robots éducatifs peuvent nous offrir. Premièrement, après avoir intégré nos activités de remédiation dans un scénario d'intéraction entre l'enfant et un robot (appellé CoWriter), nous montrons que modifier le comportement de ce dernier permet d'augmenter son acceptation sociale, permettant ainsi un engagement à plus long terme de l'enfant dans la tâche d'écriture. Enfin, nous cherchons à savoir s'il est possible de remédier à certaines difficultés d'écriture en les anticipant, en aidant les enfants d'âge préscolaire à acquérir les compétences fondamentales de coordination visuo-motrice requises lors de l'écriture.

Mots clefs: Ecriture, Apprentissage automatique, Robots pour l'apprentissage, Analyse de l'écriture, Remediation de l'écriture.

Contents

List of Figures

List of Tables

0 Thesis Organization

This thesis is organized into four main parts, each of which consists of multiple chapters.

In **Part One** is written the Context and Introduction of this thesis.

In **Part Two**, we introduce the work conducted towards the analysis of handwriting difficulties on digital tablets.

In **Part Three**, we present the tools designed to re-mediate handwriting difficulties. In particular, we present how the capabilities of robots can be exploited in order to enhance handwriting remediations under specific contexts.

In **Part Four**, we draw the synthesis of this thesis, giving insights on the main contributions, the limitations as well as insights on the future direction one wanting to continue this work should take.

Introduction Part I

1 Context

Up to now, educational robots were used in the learning of a an extensive variety of disciplines. However, the most common of these areas are undoubtedly programming, robotic engineering and other closely related subjects. An abundance of works including robots for learning in this discipline have been conducted over the years in formal education but also informal scenarios such as workshops, competitions and at-home learning beginning from pre-kindergarden [7, 8], kindergarden [7, 9, 10, 11], elementary school [9, 12, 13, 14, 15], high school to the university [16, 17, 18, 19].

Social robots can be used in education and take the role of tutors or peer learners. Within the application they were used, they have been mainly used to enhance the interaction between the child and the robot in order to achieve outcomes similar to those of human tutoring on restricted tasks. For instance, we can cite examples where Nao or Pepper were used as teaching agents to transmit high level concepts such as the ability to read or speak another language [20, 21].

Within the Cowriter project, the vision is a bit different with a will to take advantage of the physical capabilities of a humanoid robot to teach a physical skill. This project focuses on the skill of handwriting, one of the most difficult physical skills one should acquire during his/her life. Concretely, the rationale behind this project is to use a humanoid robot (NAO) in the role of a learner: the robot will write on a tablet a word with a bad handwriting and will seek the help of the child to progress (see Figure 1.1). The approach used is based on the learning by teaching paradigm: children become the teachers of an humanoid robot requiring help to better write. This approach results into several benefits: it boosts the children' self-esteem (which is especially important for children with handwriting difficulties as we will see in Chapter 2). It also gets them to practice handwriting without even noticing, and engage them into a particular interaction with the robot called the Protégé effect. Because they unconsciously feel that they are somehow responsible if the robot does not succeed in improving its writing skills, they commit to the interaction, and make particular efforts to figure out what is difficult for

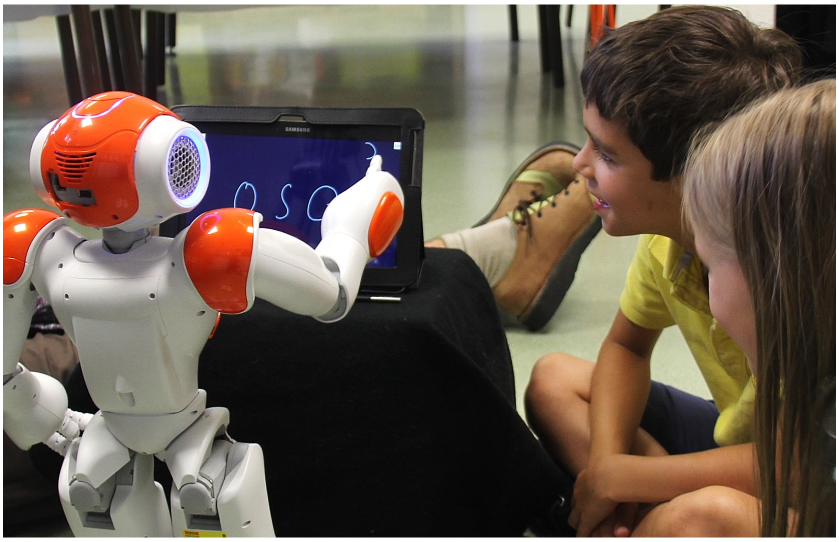

Figure 1.1: The robot is used within the CoWriter project where the robot is learning from a child the ability to write.

the robot, thus developing their meta-cognitive skills and reflecting on their own errors. We can see in Figure 1.2 a summary of the different cognitive and affective processes involved within the Cowriter setup.

However and as we will see in the next Chapter of this thesis, handwriting is a very complex skill to be learned with very different sub-skills and cognitive processes involved. The CoWriter project, if exploiting very interesting capabilities of the humanoid robot to enhance the interaction between the child and the robot, clearly present potential for improvement, especially in the understanding of the child's handwriting problems as well as in the diversity of the remediation activities proposed. For example, in the current activity, the word that the robot will write and thus that the child will train is chosen by a human. Under this context how can we be sure that this word is the best suited for the child's problems? How can we be sure that the best for a child, failing at school and avoiding all handwriting activities, is to write? That no other activities wouldn't be more suited for his/her particular problems?

It is under this context that this thesis is born. One of the goals, being described in the second Part of this thesis, is to have a wider understanding of the child's handwriting, to be able to extract his/her strengths and weaknesses in different aspects of handwriting. The second objective, described in the third Part, is to build an adaptive system for the remediation of handwriting difficulties, comprising different activities (including the CoWriter one), each of them focusing on a specific aspect of handwriting and chosen in function of the analysis done beforehand.

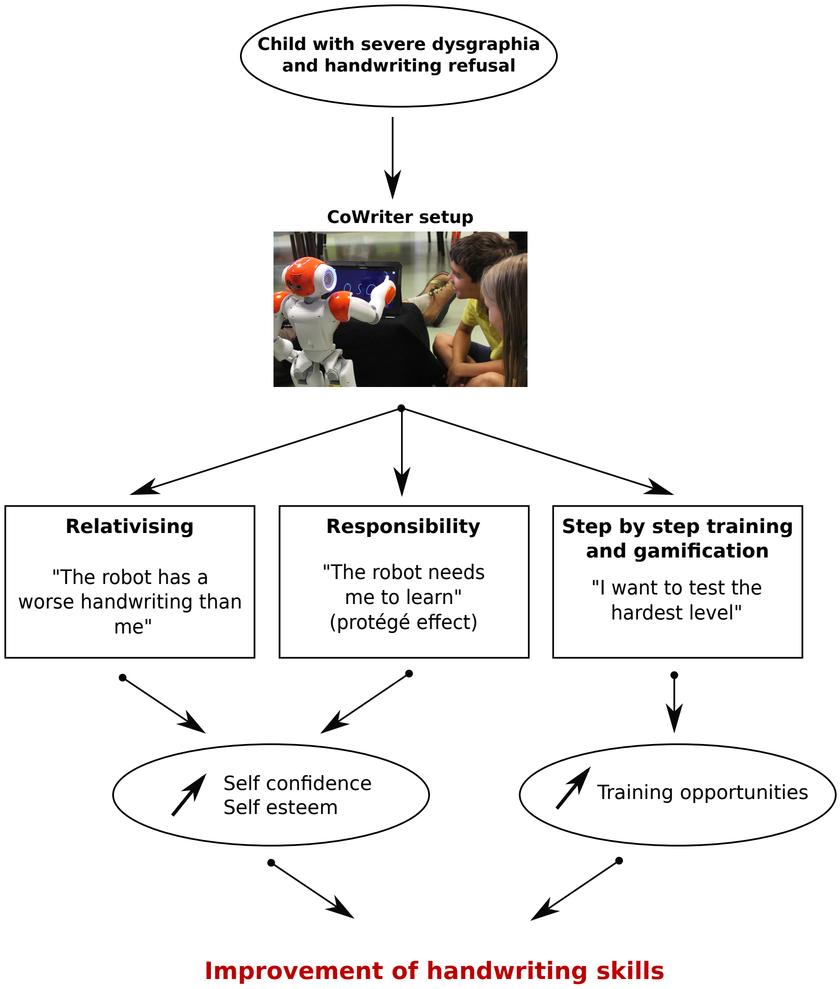

Figure 1.2: Cognitive and affective processes involved in the Cowriter setup.

2 Handwriting

Writing: the making of History

"Writing is the painting of the voice"

-Voltaire

Writing is such a central aspect in our History that its invention marks its beginning. It started more than 6000 years ago, when, for the first time, simple pictographs were drawn in a rock: it was the end of pre-history and the very beginning of our **H**istory.

Little is known about the civilizations that lived before the invention of handwriting for the very reason that all the information from these ancestral times couldn't be recorded. More than a way of communication or expression, writing is a means of memorization and transmission: an externalization of information from the humans' brains to an external support. The invention of writing brought consequences so deep and diverse to not be fully understood yet, from the invention of money to the creation of commercial routes, from the first written laws to new ways of transmission of information leading to a profound reshape of the links and exchanges between individuals.

A profound mutation happened during a major event of Renaissance with the invention of printing with *Gutenberg* and its famous first printed bible. The invention of printing submerges modern countries with millions of books that will change deeply the society and, in particular, education with the way information are transmitted. For the first time in history, the totality of information was available in the nearest library, paving the way to major changes and discoveries. This revolution also leads to an externalization of human memory, from our head to the printed pages of books: the time printing first spreads marks the first appearance of an *artificial memory.* Under this context, one of our cognitive capacity, the faculty to memorize information, got considerably reduced, since the necessity of this faculty was drastically reduced. It is precisely this intuition about the invention of printing that led the French philosopher Montaigne to say *"Better have the head well done than a head well filled"*. Paradoxically, it is the printing revolution that strengthened the importance of the skill of handwriting, since transcribing the information received on an external support became a necessity to avoid loosing it due to our vanished memory.

Finally, a new revolution is happening some century later and is still occurring now with the entry in the digital era. Our society is getting affected at least as much as in the two first revolutions (invention of writing and printing). Education, with the apparition of new powerful tools bringing a new externalization of our memory, this time not in books and libraries, but inside our own pockets with smart-phones and tablets, is and will be deeply affected: it is not only how things are taught but more importantly what needs to be taught that will have to be re-thought. Under this context, how will handwriting be affected ? On the one hand, appearing more useful than ever in order to record the increasingly faster-vanishing information escaping our deficient memory, on the other hand, risking to be replaced by other means of externalization (taking notes on a computer or a tablet): what will be its fare in the era of digitalization? More and more people are swapping their pen and paper to computer or tablets on their everyday life. For example, a recent British survey related that one in three respondents had not written anything by hand in the previous six months [22]. Some countries even started to banish cursive handwriting from school, as in the case of Finland, one of the first countries to replace cursive handwriting classes in favor of keyboard skills [23]. In the United States the same trend can be noticed since cursive handwriting has also been dropped from the Common Core Curriculum Standards. As a consequence, forty three states no longer require the teaching of cursive handwriting in public schools.

If the importance of handwriting for adults in their everyday life is undeniably decreasing, it would be too premature to make the same observation for school children. Indeed, different researches show the importance of handwriting for the child's development in other skills including reading, memorization or understanding for instance. In [24], James et al. compare the effect of handwriting and typing on a keyboard on the brain of pre-school children. They noticed that brain activation during letter perception is influenced in a different and important way by previous handwriting of letters or previous typing or tracing of those same letters, suggesting that handwriting is important for the early development of brain regions known to underlie successful reading. In [25], Mueller et al. showed that students who took notes on laptop performed worse on conceptual questions than students who took notes longhand (handwriting). The authors suggest that when the notes are taken on a laptop, users tend to transcribe lectures verbatim rather than processing information and re-framing it in their own words, for example when the notes are taken longhand, which contributes to memorizing and understanding.

In any case, handwriting currently remains an essential skill to be acquired during childhood education since it is the basis of core educational activities such as taking notes, story composition or self expression for example [26, 27, 28]. Several studies show the link between good handwriting skills and success in school education [29, 30]. Indeed, with the increasing cognitive demand for school work throughout their curriculum, children facing handwriting difficulties rapidly become unable to face simultaneous efforts such as handwriting, grammar, orthography, and composition. These students may then rapidly face more general learning difficulties and even experience situations of failure [29, 30].

Handwriting Acquisition

Learning the ability to write can be seen as one of the most important skills a young child acquires in the early school years. Handwriting is a complex task involving cognitive, perceptual, attentional, linguistic, and fine motor skills [31, 29, 32]. That is why, even in the case of normally developing children, learning handwriting spans a long period, between the ages of 5 (preschool) and 15 [33, 34, 35, 36, 37, 38].

Formal acquisition of handwriting initially begins at pre-school where children starts to draw basic shapes on substrates. According to Kellogg et al. [39], common elements can be found in the early scribblings of young children. These common elements which she called *Basic Scribbles* consists of various lines with different orientations, circular movements, zigzags and other similar movements and are found to be central elements in handwriting. Therefore, it is likely that the ability learned at young age to form these basic shapes can be transfered to the more precise and systematic way to form letters at the time handwriting is initially taught [39]. Other researches have supported the notion that children learn graphic elements through drawing [40, 41, 42] which appears particularly intuitive given the similarities between these two systems in terms of fine motor hand movements. Regardless of the differences that might exist between handwriting and drawing [43], many studies support the fact that children will often mix handwriting and drawing during their young age. As shown by Adi-Japha et al. [44], the specificity of writing over drawing only emerges after six years of age, probably because of the increase of writing practice. Children generally adopt the same motor rules as the ones they use to produce the standard geometrical forms [45] that are taught in the very early stage of school education while drawing. This is one of the reasons why several research studies define handwriting readiness on the basis of the child's ability to copy several standard geometric forms [46, 47].

Another important aspect of handwriting that starts at preschool is the acquisition of the visual perceptions of letters, called the grapheme. Children should acquire the visual perception of letters that will allow them to discriminate between the grapheme of different letters [48, 49]. In order to have a complete representation of the letter, children should also acquire visual-motor coordination skills to produce the dynamics of the letter, also called the ductus [50] which represents the "ability to integrate the visual images of letters or shapes with the appropriate motor response" [51, 52, 53]. To teach these concepts, it is shown that using different sensory informations ranging from audio, visual to kinesthetic feedback may be important [50, 54, 55, 56]. It is for this reason that teachers commonly use various techniques where children can experience these different sensory informations such as drawing letters in sand, feeling the grapheme of letters craved in a piece of wood, describing verbally the letters or building the letter with play-dough [57, 27] as illustrated in Figure 2.1.

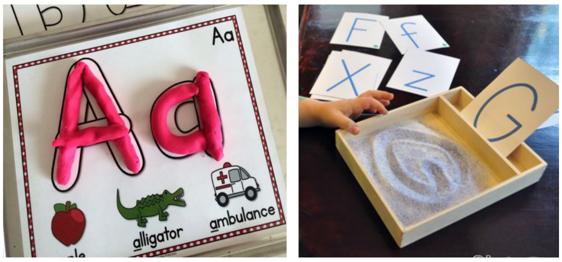

Figure 2.1: Example of techniques used to teach visual perception and visual-motor coordination of letters. Left: the letters are formed using play-dough. Right: the child draws the letter in a sandbox.

Formal handwriting with pen and paper substrates starts around the age of 5.5 years old. It then takes a number of years for children to acquire an automated handwriting as it involves multiple skills such as the ones described above, but also fine motor skills (important for a rapid and effective manipulation of the pen) as well as several high level cognitive elements associated with language production [58]. During these years, handwriting initially evolves on a qualitative level, especially during the first ages of development (between first grade and fifth grade), and reaches its peak in grade 6 [59, 60, 61, 62]. Handwriting then mainly evolves on a quantitative level (handwriting speed mainly evolve starting from grade 4) [36, 63]. It is interesting to notice a gender effect in handwriting acquisition with girls presenting slightly higher handwriting quality and speed compared to their male peers, although no effect of handedness has been reported so far.

Handwriting Difficulties

Consequences of Poor Handwriting

As school children progress through their curriculum, the cognitive demand associated with the work requested increases. In that context, children facing handwriting difficulties quickly become unable to face simultaneous efforts such as handwriting, orthography, grammar or composition. These students may then rapidly face more general learning difficulties and sometimes face situations of failure [29, 30]. As an example of handwriting difficulties leading to more general difficulties, we can cite the the process of composition: while composing, the writer will concentrate his efforts on the process of handwriting itself (such as how to form a particular letter) which will lead to less space in working memory being available for the core task leading him to forget the ideas to express in the composition [64, 65].

In addition, the proper development of handwriting is not only an essential criterion for school success but may also has critical consequences on the behavioral development of children. Poor handwriting skills may affect the way children are perceived by their teachers and peers. A few studies show that handwriting can often be used to judge children in schools: despite similar content, students with poor handwriting will be assigned lower grades compared to those with legible handwriting [66, 67]. As illustrated by Feder et al. [68], "handwriting performance has a widespread effect on the child's own image, academic achievement, attitude and behavior". Under these conditions, handwriting difficulties usually generate avoidance of written tasks, and may eventually result in increased anxiety and low self-esteem, which in turns leads children to avoid training opportunities and ultimately bring school refusal. Children may thus often be trapped in a vicious circle such as the one presented in Figure 2.2.

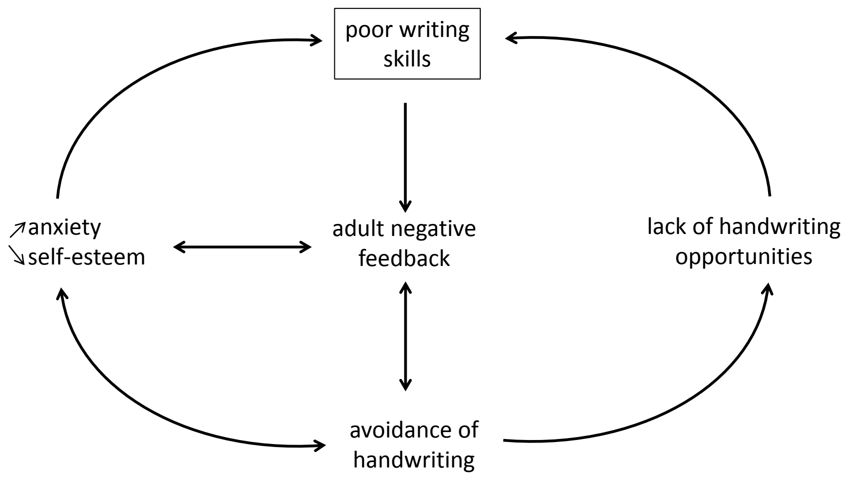

Figure 2.2: Psychopathological model of handwriting difficulties. A vicious circle can appear due to anxiety and lack of practice that can worsen handwriting. Adapted from [1].

Even with proper training, up to one-third of children never truly master the skill of handwriting [61, 69]. When these difficulties are severe, we speak of dysgraphia. According to the definition proposed by Hamstra-Bletz and Blöte [4], dysgraphia is a disturbance in the production of written language, related to the mechanics of writing. The causes of dysgraphia are quite heterogeneous, depending on both biological factors (e.g. motor development, type of motor deficits related to developmental coordination disorder, potential comorbidities with dyslexia, attention-deficit hyperactivity disorder) and social (e.g. writing habits, relationship with parents) factors. Dysgraphia is far from being singular since nearly 8.6% of the population in France is considered dysgraphic [61]. Early detection of handwriting difficulties and especially dysgraphia has been shown to facilitate effective treatment [29]. For this reason, many tests allowing to detect handwriting difficulties have been developed in different languages and alphabets [70, 3].

Current Tests to Detect Handwriting Difficulties

Assessing handwriting difficulties is not a new challenge. The first methods invented to assess the legibility or readability of handwriting were born at the beginning of the twentieth century. The very first scaling method was developed by Thorndike [71] in 1910. His contribution was seen of great important ”not only to the experimental pedagogy but to the entire movement for the scientific study of education” [2]. Thorndike himself compared his invention to the thermometer: ”Just as it was impossible to measure temperature beyond the very hot, hot, warm, cold, etc., of subjective opinion, so it had been impossible to estimate the quality of handwriting except by such vague standards as one’s personal opinion that given samples were very bad, bad, very good, etc.” [2].

Until now, we can find two approaches to evaluate handwriting. The first is a global holistic method that evaluates the handwriting quality as a whole, while the second measures it according to several predefined criteria [72].

The global holistic approach analyses the handwriting by comparing the sample with other examples sorted depending on the quality. The scale developed by Ayres [2] is a typical example of such an approach: his method gives the possibility to teachers to grade handwriting by comparing the sample with eight handwriting references of increasing quality as can be seen in Figure 2.3. If several updated scales tried to decrease the level of subjectivity associated with this approach [73, 74, 75], the assessment of the handwriting quality skill relies on subjective judgment by the teachers.

The second approach analyses handwriting according to several predefined criteria (e.g., size of letters, spacing between words, linestraightness). The evaluation is then made by individually grading these criteria and summing the sub-scores. A large amount of tests developed in the past 40 years are using this approach. Among the most used ones, we can cite the Evaluation Tool of Children Handwriting (ETCH-C) [76] for the Latin alphabet or the Hebrew Handwriting Evaluation (HHE) for the Hebrew alphabet [77]. The most commonly used test in French speaking countries is the concise Evaluation Scale for Children Handwriting (BHK) test [61]. This test was initially developed in

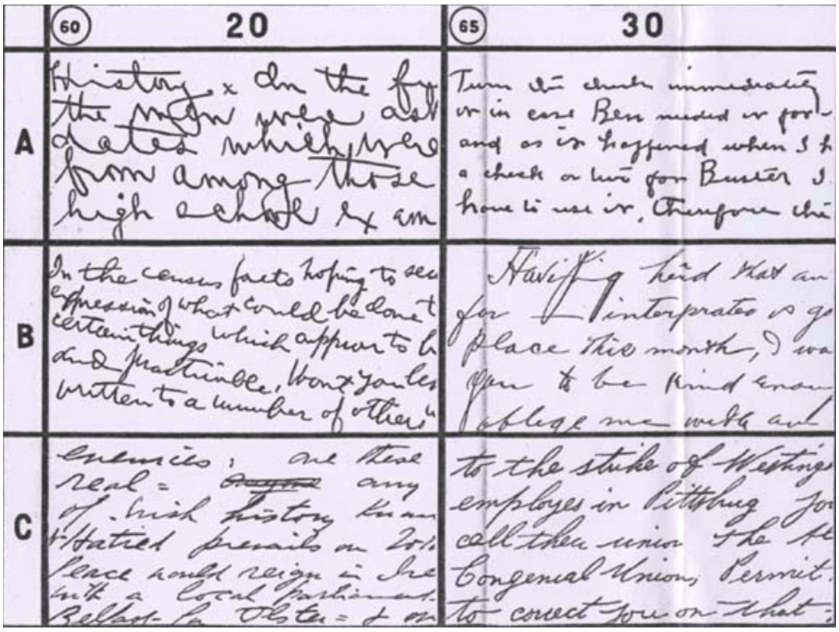

Figure 2.3: Handwriting samples of different quality in the scale developed by Ayres in [2]

the Netherlands [60], and then adapted to other languages including French [61]. This test consists of copying a text beginning with simple monosyllabic words and evolving towards more complex words for five minutes onto a blank paper. Different features reflecting handwriting quality (e.g., letter form, size, alignment, spacing, ...) are scored to generate a quality score as can be seen in Figure 2.4. The final diagnosis of the child handwriting depends on this quality score but also on the child's gender and age. A summary of different tests used to detect handwriting difficulties can be found in Table 2.1.

As shown in Table 2.1, these tests are heterogeneous as they were specifically designed to analyse the handwriting quality for a specific alphabet or age range. In addition, these tests are based on handwriting from different writing tasks (as illustrated in the core task column in Table 2.1), which might imply high variability in the results. Finally, most of the tests bases the analysis on the final handwriting product and do not take into account the handwritten production. Only the final, static product of handwriting is used for analysis, disregarding the handwriting dynamic, tilt, and, in most cases, pressure.

The rapid emergence of digital tablets in the last decade allowed to partially tackle some of these problems. The handwriting analysis not only based on the final product of handwriting (the static image), but also on its dynamics is now possible. Only a few techniques, have then been proposed to classify the handwriting legibility, including the dynamics of the process, opening new perspectives in the analysis of dysgraphia. For instance, Pagliarini et al. [82] collected handwriting samples on digital tablets and used

CRITÈRES						TOTAL
1. Ecriture grande						
2. Inclinaison de la marge vers la droite						
Phrases	1	2	3	4	5	
3. Lignes non planes						
4. Mots serrés						
5. Ecriture chaotique						
6. Liens interrompus entre les lettres						
7. Télescopages						
8. Variation dans la grandeur des lettres troncs						
9. Hauteur relative incorrecte						
10. Distorsion des lettres						
11. Formes de lettres ambiguës						
12. Lettres retouchées						
13. Hésitations et tremblements						
					Score Total	

Figure 2.4: Psychopathological model of handwriting difficulties. Vicious circle can appear due to anxiety and lack of practice that can worsen handwriting. Adapted from [1]

quantitative methods to find patterns showing potential future writing impairments at a very early age. In [70], Mekyska et al. collected data of 54 third-grade Israeli children and used the 10-item questionnaire for Hebrew handwriting proficiency (HPSQ) as a ground truth. They then trained a Random Forest model to classify dysgraphic from non dysgraphic children. Despite the low number of subjects involved in their study, they managed to reach a remarkable accuracy showing the potential of such an approach. In [83], Rosenblum et al. performed a similar work by developing a method for automatic identification and characterization of dysgraphia. They trained machine learning models which are capable of discriminating dysgraphic products from proficient products with approximately 90% accuracy. It has to be noticed that none of the described solutions can be found as a commercialized product.

Handwriting Pedagogy

One of the main objectives for primary school teachers is to help children develop a "good" handwriting characterized by legibility and style. In the same time, training in handwriting must also aim at producing good writing with useful speed.

One of the first steps to ensure the development of good writing skills for children is to

	Age range [y.o.]	Test duration [min]	Scoring duration [min]	Alphabet	Language	Number of items	Dynamic of handwriting	Pressure	Tilt	Speed	Posture	Writing task
Ajuriaguerra [4]	6-12	2	5	Latin	French	37	✗	✓*	✗	✓	✓	WT1
BHK [78]	6-12	5	10	Latin	Multi-language	13	✗	✗	✗	✓	✗	WT2
BHK-teenager [79]	12-18	5	10	Latin	Multi-language	9	✗	✗	✗	✓	✗	WT2
DASH [80]	9-16	20	10	Latin	English	5	✗	✗	✗	✓	✗	WT3
HHE [81]	6-18	5	0	Hebrew	Hebrew	10	✗	✗	✗	✗	✗	WT4

Table 2.1: Summary of different tests used to diagnose dysgraphia. WT1: Copy a sentence several times, request of quality and speed, WT2: Copy a long text for 5 min, WT3: Copy a sentence several times, alphabet, geometric figures and composition, WT4: Copy a text containing all letters. *some pressure aspects of handwriting are assessed thanks to carbon paper.
Ajuriaguerra scale (E scale): is a well spread test evaluating the quality of the writing depending on speed and precision. It has a special focus on the posture and style of pen grasping of the child.
Concise Evaluation Scale for Children's Handwriting (BHK): is the reference test to diagnose dysgraphia for the Latin alphabet language [4, 5, 6].
BHK for teenagers: has also been created using the same principles.
Detailed Assessment of Speed of Handwriting (DASH test): evaluates the quality and the speed of the writing in different conditions (quality, speed, writing with a free topic of the child choice).
Hebrew Handwriting Evaluation (HHE): that examines Hebrew handwriting products and assess the legibility through both global and analytic measures.

provide them formal instructions [84]. A common belief is that systematic instruction is essential for proper handwriting development especially for children who do not write instinctively [85]. Indeed, the authors showed that a lack of formal handwriting instructions may be problematic, especially for children already having underdeveloped foundation skills.

However and despite the well known importance of handwriting in children later achievements, little is known concerning the approaches used by teachers and therapists to teach handwriting. A survey by Graham et al. in 2007 [86] reviewed the teaching practice about handwriting in the United States of America. The results showed that 90% of school teachers are teaching handwriting during their lessons. In average, school teachers spend 70 minutes teaching handwriting during a typical week. However, the considerable variability of handwriting practice (Std. = 55 minutes) across the different schools shows that no conformity seems to exist in the way handwriting is taught. This variability was found again in the frequency of the teaching sessions within a week, in the way teachers are providing these lessons (in an individual, group or class level) as well as in the teaching methods used.

When children demonstrate severe handwriting difficulties and when school remediation do not result in enough advancement, teachers often refer children to occupational therapists [87, 88]. The occupational therapist's unique role is to analyze handwriting difficulties in terms of underlying deficits in postural, motor, sensory integrative, sensorimotor, perceptual or behavioural factors which may be interfering with legible handwriting performance [89, 90]. A survey by Feder et al. [68] across Canada showed that once again no consensus seems to exist since therapists tend to use their own methods for the treatment. A variety of approaches ranging from motor learning, perceptual-motor, sensory integrative and cognitive-behavioral techniques can be used for the remediation, with no consensus on what is the most effective treatment strategy [91].

Under this context, it appears that teaching the skill of handwriting can be extremely complicated due to the many sub-skills involved in the learning process and the number of different remediation activities possible. The emergence of digital tablets appears to be a credible solution in the analysis of handwriting difficulties, but also in the way handwriting problems can be intermediated.

Handwriting Analysis Part II

As expressed in its title, this thesis aims at proposing a new method to detect and remediate handwriting difficulties. Before being able to decipher on the adequate remediation (as developed in *Part Three*) for any given child, it is important to develop a solid understanding of what aspects children present handwriting problems on. The Second Part of this thesis proposes a new approach to detect and analyze handwriting problems based on features describing handwriting on different multidimensional aspects. In particular, our analysis is conducted on a tablet based test allowing to extract the dynamics of handwriting, making hidden but very relevant information emerge as shown in Chapter 3 of this thesis. The method allowing handwriting analysis is then described in Chapter 4. A key characteristic of the proposed method is that it analyses handwriting based on low level features capturing almost physiological aspects of handwriting, independent of the writing content. In Chapter 5, we show how our method can be transfered to other alphabets (from Latin to Cyrillic), paving the way for a cross-lingual model for the detection of handwriting difficulties.

The results shown in Chapter 4 rely on data collected using a Wacom tablet on top of which a sheet of paper was attached. This setup is quite limiting since the device is expensive and needs to be connected to a laptop. A test directly running on iPads (considerably cheaper and more widespread than Wacom tablets) would have a potentially much greater impact on the society. To this end, in Chapter 6, we demonstrate that our method retains its accuracy, given proper retraining of the machine learning models, also when applied on handwriting directly performed on a tablet.

Finally, the last chapter of this Part presents the iPad based test developed to extract the multidimensional handwriting profile of the child. In particular, the developed test allows to extract the specific weaknesses of children in a couple of seconds, on different aspects and at different granularities, and serves as a basis for the recommendation system aiming to propose a handwriting remediation tailored to the specific handwriting profile of the child that is discussed in the *Part three* of this thesis.

3 Handwriting is Dynamic

Introduction

As extensively shown in the Introduction of this thesis, the tests currently adopted to detect handwriting problems are, for the majority of them, pen and paper based tests. In that sense, the handwriting analysis is limited to the final static image of the child's handwriting. Some very informative handwriting aspects, such as the handwriting dynamics is therefore not used in the analysis of handwriting problems.

In this Chapter, we investigate the importance of the handwriting dynamics towards the detection of handwriting problems. To that end, we compared the efficiency of two models. The first one: a Recurrent Neural Network (RNN), works with sequential data and is therefore able to exploit the handwriting dynamic, while the second one, a Convolutional Neural Network (CNN), takes static images as input, and is thus restricted to a static analysis of handwriting.

This work corresponds to the following publication/patent:
K. Żołna, T. Asselborn, C. Jolly, L. Casteran, M. N'Guyen-Morel, W. Johal and P Dillenboug, "The Dynamics of Handwriting Improves the Automated Diagnosis of Dysgraphia", *arXiv preprint*, **2019** [92].

K. Żołna, T. Asselborn and W. Johal, "Method of Handwriting Character Recognition Confirmation", International Patent App. WO2019/101338.

Method

Participants

The present study was conducted in accordance with the Declaration of Helsinki and was approved by the University ethics committee (agreement nb 2016-01-05-79). It was conducted with the understanding and written consent of each child's parents, the oral consent of each child and in accordance with the ethics convention between the academic organisation (LPNC-CNRS) and educational organisations. A total of 971 TD (Typically Developing) children were recruited in 14 schools from various Grenoble suburbs to ensure different socio-economic environments (*TD dataset*). 43 classes were included from pre-school to fifth-grade. None of the TD children included in the study presented known learning problems or neuromotor disorders. Twenty-four dysgraphic children recruited at the Learning Disorders Center of Grenoble hospital (Centre Référent des Troubles du Langage et des Apprentissages, CHU Grenoble) were also included in the study (*D dataset*) and were all diagnosed as dysgraphic based on their BHK scores.

Data Collection

Children were asked to write cursively, without a time limit, the 26 letters of the alphabet, as well as 10 digits, randomly dictated. Two procedures following different dictations were performed, the first in the middle of the school year (January-February) and the second at the end of the school year (May-June). We checked that the dictation order did not affect children performances. Dictations were performed on a sheet of paper placed on a Wacom Intuos 4 A5 USB graphic tablet (sampling frequency = 200 Hz; spatial resolution = 0.25 mm). The sheet of paper was used to place the children in their usual handwriting conditions as asking them to write directly on the tactile surface might be different due to the different friction coefficient [93]. All tracks were monitored using the Scribble software developed in the LPNC laboratory [94, 95].

Handwriting Analysis and Identification of Dysgraphia

Scarcity of annotated data

Children are taught handwriting in a standardised way from the beginning of their school education. As a result, the vast majority of the recorded glyphs corresponds to the standard of what was taught in school (resulting in an over-representation of "good" examples). In other words, the class of glyphs that we are particularly interested in ("bad" examples) are underrepresented.

Since there are only a few ways to write a letter correctly and many ways to do it incorrectly, we have a significant variety of instances that belong to the underrepresented class (illegible glyphs), as compared to the relatively small variance in the overrepresented class.

For these reasons, the discrimination between children with and without dysgraphia appears to be challenging. We argue that using a black box to solve this problem would require a considerable amount of data, which, in practice, we find unfeasible to collect. We are thus left to work in a regime of underrepresented positive cases (bad handwriting), meaning that a more sophisticated solution needs to be developed. The solution we propose is inspired by the principle of transfer learning.

Transfer learning

Transfer learning is an approach where a new task is solved through the transfer of knowledge from a related proxy problem that is simpler to solve [96].

In the case of handwriting, a small number of shapes is defined to be letters or digits, and people are trained to write and read these shapes. The glyphs written by dysgraphic children tend to be harder to decode as they are not written conventionally. This is why we believe that the problem of identifying dysgraphic children may be solved by using an approach that mimics humans. Concretely, we decided to tackle the problem by first training a recognising model to classify glyphs and then use its prediction to discriminate dysgraphic children. The underlying rationale is that if the model fails to predict which glyph is the given one, then the writer has a higher probability to be dysgraphic. In other words, the recognizing model addresses the simpler problem of assessing the legibility of any given glyph. It is important to notice that this approach is plausible only if the recognising model performs well when presented with legible glyphs. In other words, the diagnosing model based on a poorly performing recognising model would wrongly label all children as dysgraphic. Hence, a crucial requirement of our approach is to make sure that the recognising model performs well enough when glyphs written by non-dysgraphic children are provided as the input.

In synthesis, since we are in the regime of scarcity of annotated data for dysgraphic children, we decided to use a straightforward diagnosing model (taking an average of given predictions) built on top of an advanced recognizing model (that may leverage the massive data we have for non-dysgraphic children).

Assessing glyphs' legibility

We divided our data into three disjoint sets:

- *Training set*: 80% of *TD dataset*.
- *Validation set*: 20% of *TD dataset*.

- *Dysgraphic set*: all the glyphs written by dysgraphic children (*D dataset*).

Hence, the recognizing model is trained on examples which are believed to be more legible, since they have been produced by children not diagnosed as dysgraphic.

The model's input consists of a single glyph. There is a class for every glyph (36 classes). Depending on the model used, the input may be the trajectory followed by the child's pen to produce the glyph (point-by-point for the RNN case) or the static image corresponding to the glyph (final, visual results for the CNN). This means that the recognizing model treats the problem as a multi-class classification balanced by the design of the data collecting procedure (each child has to write down all 36 glyphs).

In real-life applications, a child would be asked to write down a given glyph (i.e., the ground truth for the glyph is always known) and, even if the ground truth is known, we let the recognizing model predicts the glyph's label. Afterwards, the discrepancy between the ground truth and the model's prediction is measured.

Assuming that the model discriminates the glyphs properly with high accuracy, a high discrepancy means that glyphs are not legible and the child is likely to be dysgraphic. The glyph-level scores are averaged over all 36 glyphs written by the child, and this value is understood as a statistic for evaluating writing proficiency, which we call the *D* statistic (*dysgraphic statistic*). Concretely, the more similar the glyphs are to their ground truth classes, the higher the *D* statistic of the child is, and therefore the higher are the chances that he/she is not dysgraphic. Taking an unweighted average is a simple, arbitrary choice and more sophisticated methods may be invented.

Final dysgraphic prediction

The lower the value of *D statistic*, the more likely the child is dysgraphic. However, to solve our main problem, we still have to answer the following question:

Given a value of D statistic*, is the writer dysgraphic or not?*

In other words, we have to find the threshold (the critical value) of *D statistic* that divides the children into two groups (dysgraphic and non-dysgraphic children). In the original BHK test, children are considered dysgraphic if they obtain a score beyond two standard deviations from the normative group. In practice and according to the statistics in French speaking countries, it was found that 8.6% of the population is dysgraphic [97]. We therefore, decided to use the same approach to compute the threshold such that the 8.6% of all children, i.e. those who obtained the lowest values of *D statistic*, would be considered as dysgraphic by our diagnosing model. Of course, this threshold value depends on the recognising model. The main property of this approach for fixing the threshold is that, on average, 8.6% of children are labelled dysgraphic by our test, precisely like the original BHK test.

Discrimination procedure

To summarise, the discrimination procedure for a given child consists of the following steps, as can be seen in Figure 3.1:

1. The child is asked to write down one by one the 36 glyphs on the given digital tablet.

2. The recognising model, which is trained to recognize glyphs, is executed on all glyphs and a single value (the probability that the drawn glyph is the one requested) for each input glyph is obtained. A higher value represents a greater legibility of the glyph.

3. All 36 scores are averaged giving the *D statistic*.

4. The value of the *D statistic* describes the child's handwriting proficiency. It is then compared with a threshold calculated in such a way that, on average, 8.6% of children score are below. A child with a score below this threshold is labelled dysgraphic.

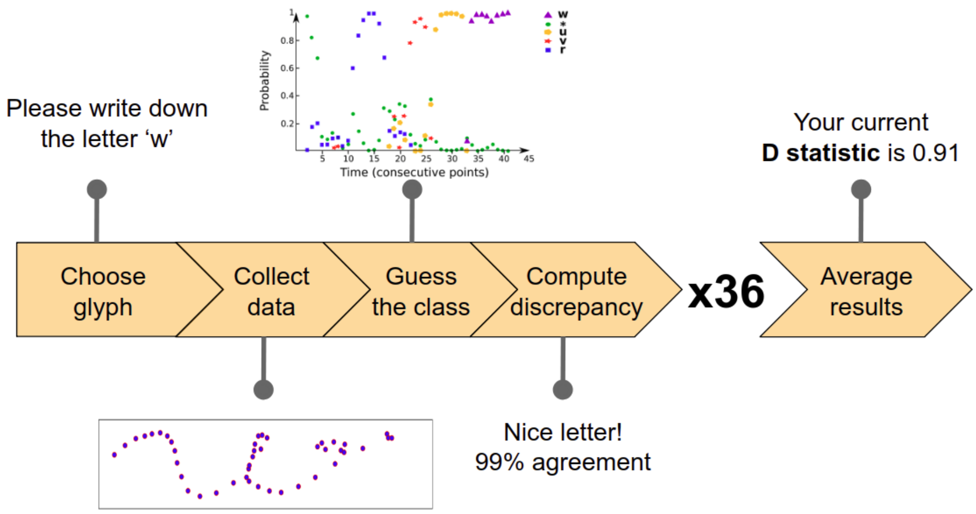

Figure 3.1: Pipeline of our method to diagnose dysgraphia using models assessing legibility of a given glyph. The Figure on top of the module *Guess the class* corresponds to Figure 3.5. Additional information are provided in its caption.

Results

In this section we address the following question:

Do 8.6% of all children suspected by our prototype represent the group of children truly diagnosed as dysgraphic by the original BHK test?

We show in this section that our model provides very promising results, especially when it uses the dynamics of handwriting as input.

Convolutional Neural Network

In this section, we present the results obtained using a Convolutional Neural Network, a category of neural network shown to be very effective in tasks, such as image recognition and classification. Such a network works only with static 2D data, so, contrary to the Recurrent Neural Network, the dynamics of handwriting (i.e., the timeframe) cannot be taken into account. Only the final trace is used by the model to predict the label of the data.

A k-fold cross-validation [98] (with $k = 5$) over the TD dataset produced the graph presented in Figure 3.2.

In this Figure, the x-axis denotes quantiles and the y-axis includes k-averaged *D statistic* values. By design, 8.6% of validation results are below the threshold, which separates the children into dysgraphic and non-dysgraphic groups. This means that in the case of perfect performance of the recognizing model, all dysgraphic children should score below the threshold. It can be seen that even if the score of dysgraphic children is significantly lower than the one of TD children (which does not guarantee all these children are non-dysgraphic), we failed to obtain a clear separation of dysgraphic and non dysgraphic children: only 25% of the dysgraphic are below the threshold line suggesting that the model can successfully identify approximately a quarter of the dysgraphic children.

Recurrent Neural Network

We present the results obtained using a Recurrent Neural Networks which utilises as input the sequence of points produced by the child to draw the glyph to take into account the dynamics of writing. We argue that this temporal information provides valuable insights so that this model is advantageous over other methods only using static images (including the original BHK test).

The same analysis described above for the CNN is done over the RNN, and the results are shown in Figure 3.3.

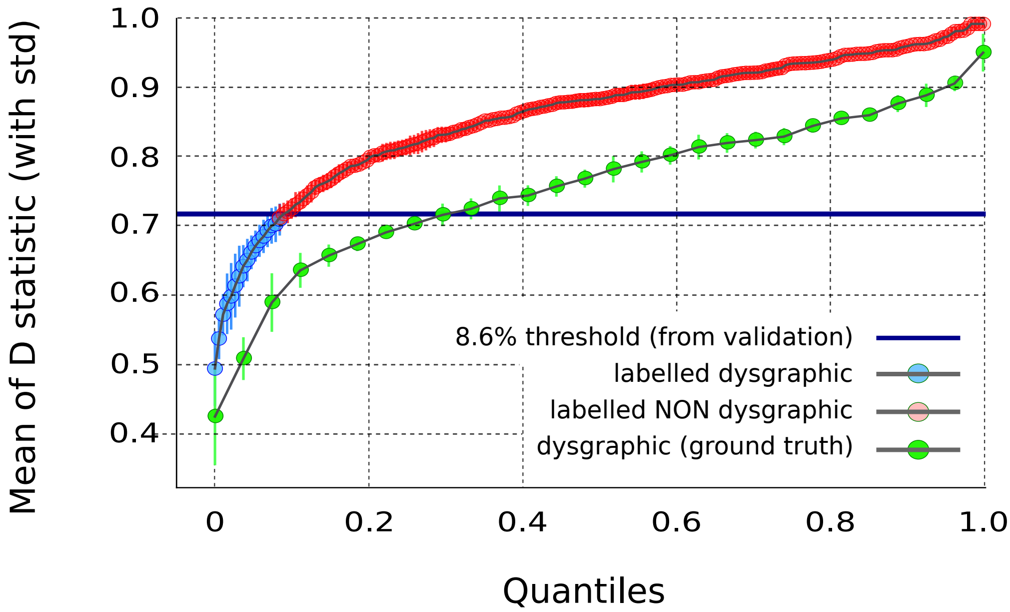

Figure 3.2: Quantile function of *D statistic* on validation and dysgraphic sets for the CNN recognizing model. The computed threshold is drawn as a blue line, and each point represents the *D statistic* for one child (averaged over all folds, the error bar represents the standard deviation). By design, 8.6% of the validation results are below the threshold, which divides the children into dysgraphic (blue points) and non-dysgraphic (red points) groups such that, in the perfect case, all dysgraphic children should score below the threshold. As the Figure shows, the CNN model is not very efficient as less than 30% of dysgraphic children are correctly diagnosed.

We can see that only two dysgraphic children obtain a D statistic score which is not below the threshold, and hence more than 90% of dysgraphic children are correctly diagnosed. This means that even in this simplistic form (that analyses only single glyphs), the proposed diagnostic method achieves performances very close to the original BHK test. This result is impressive as even without any further research, we can prepare a model capable of identifying more than 90% of all dysgraphic children, without the need for the laborious review of an human expert. Additionally, this digital approach for labeling children is not burdened by possible subjectivity and biases introduced by human examiners. Hence, this work is a milestone in achieving our long-term mission of creating a digital test assessing handwriting difficulties.

This finding suggests the superiority of the RNN over the CNN for this task (25% of children correctly labelled for the CNN, against more than 90% in the case of the RNN). We believe that this difference in accuracy is due to the incorporation of the movement's dynamics in the RNN input. This hypothesis will be explored in the next section.

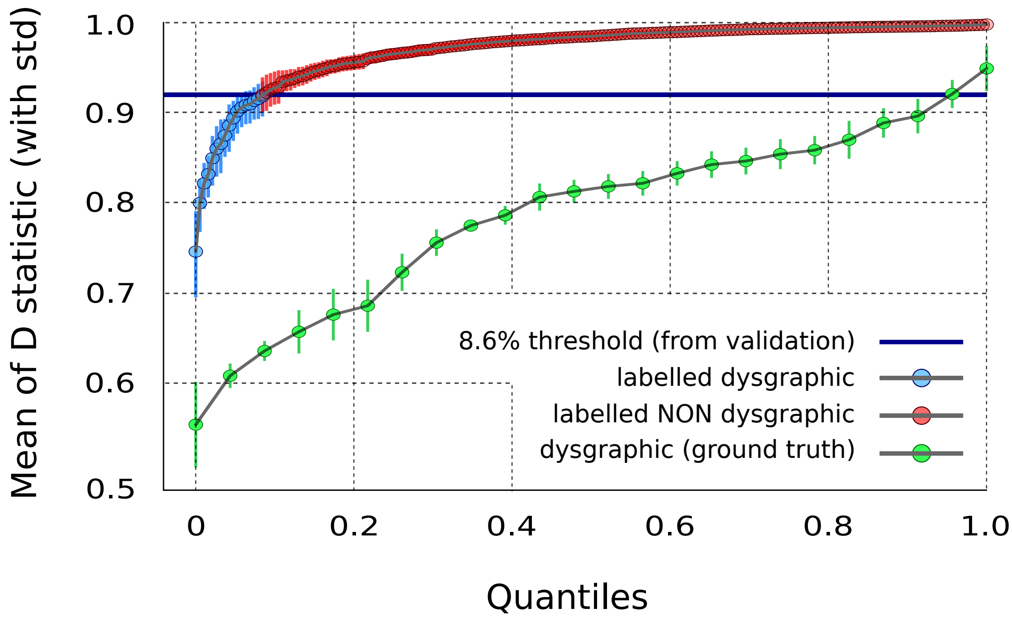

Figure 3.3: Quantile function of averaged *D statistic* on validation and dysgraphic sets for the RNN recognizing model. The blue horizontal line represents the computed threshold. Each point represents the *D statistic* for one child (averaged over all folds, the error bar represents the standard deviation). By design, 8.6% of the validation results are below the threshold which separates the children into dysgraphic (blue points) and non-dysgraphic (red points) groups. In the case of perfect accuracy, all dysgraphic children should score below the threshold. As the Figure shows, the RNN model seems way more efficient than the CNN since more than 90% of dysgraphic children are correctly diagnosed.

Handwriting Dynamics Reduce Confusion Between Visually Similar Glyphs

Some glyphs might look very similar when we only have access to the final trace. This is, for example, the case between the letters 'e' and 'l' or between the letter 'g' and the digit '9'. Thus, we hypothesize that the final trace is sometimes not enough to discriminate between these classes, and, as a consequence, that the dynamic of handwriting gain even bigger importance in such cases. For those glyphs whose final traces look similar, we except the CNN to make inaccurate classifications, while the RNN would yield better results due to its access to the movement dynamics. We believe that this is one of the reasons explaining the superiority of the RNN model over CNN. Figure 3.4 plots the confusion rates between items in the six most similar pairs of glyphs (representing the pairs of glyphs with the greatest confusion) for the CNN and RNN models. If we can see only six pairs of glyphs in Figure 3.4, it is because four are the same for both models. The confusion for a pair of glyph A-B is the number of misclassification (B instead of A or A instead of B) over the entire set of As and Bs. Please notice that the number of misclassification of A in B over all As and the number of misclassification of B in A over all Bs are different. However, as those values were similar in our case, we decided to

use their average to obtain the final misclassification values plotted in Figure 3.4. For example, the confusion of 16% for the CNN model seen between the letters 'e' and 'l' means that a misclassification of 'e' as 'l', or of 'l' as 'e' occurs in approximately 16% of the cases.
This graph appears to support our hypothesis as we can see that the RNN generates fewer confusions between the glyphs that looks *visually* similar, probably due to its access to the handwriting dynamics. Concerning all other pairs of glyphs, that do not look *visually* similar, the confusion rates of the two models are essentially the same and close to 0%.

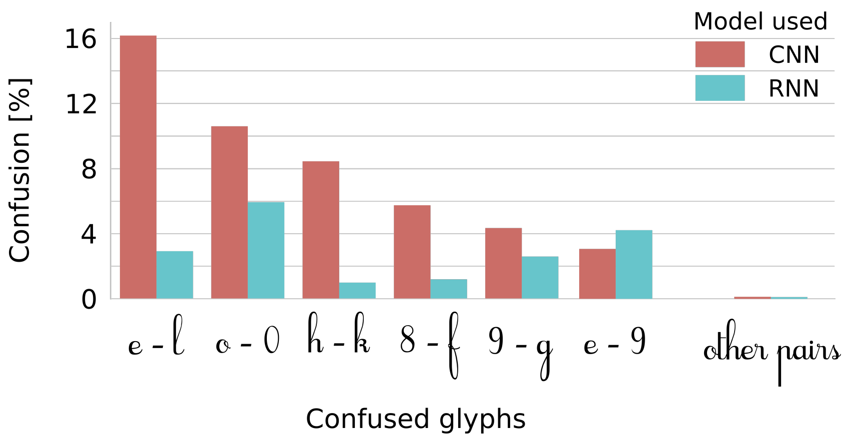

Figure 3.4: Confusion rates for the six most similar pairs of glyphs for the RNN and CNN models.

Additional class

We noticed that it is beneficial to add an extra class to the training set, which we call the star or, simply, the '*' class. This class consists of random hybrids of real glyphs obtained by combining two or three drawings to create a non-existing glyph, such that the beginning part of the first glyph trajectory is associated to the middle part of the second one, for example.

This data augmentation makes our model suspicious about glyphs that look strange and labels them as '*'. In other words, our model cannot assume that the analysed object is a glyph. We believe this strategy can help in situations when a dysgraphic child would perform poorly on a glyph but still makes a good guess since the mistake is easy to figure out. Examples include triple 'w' or a letter 'i' with additional dots.

As can be seen in Figure 3.5, employing the dynamics in the model not only improves the predictive power but also makes the model more interpretable.

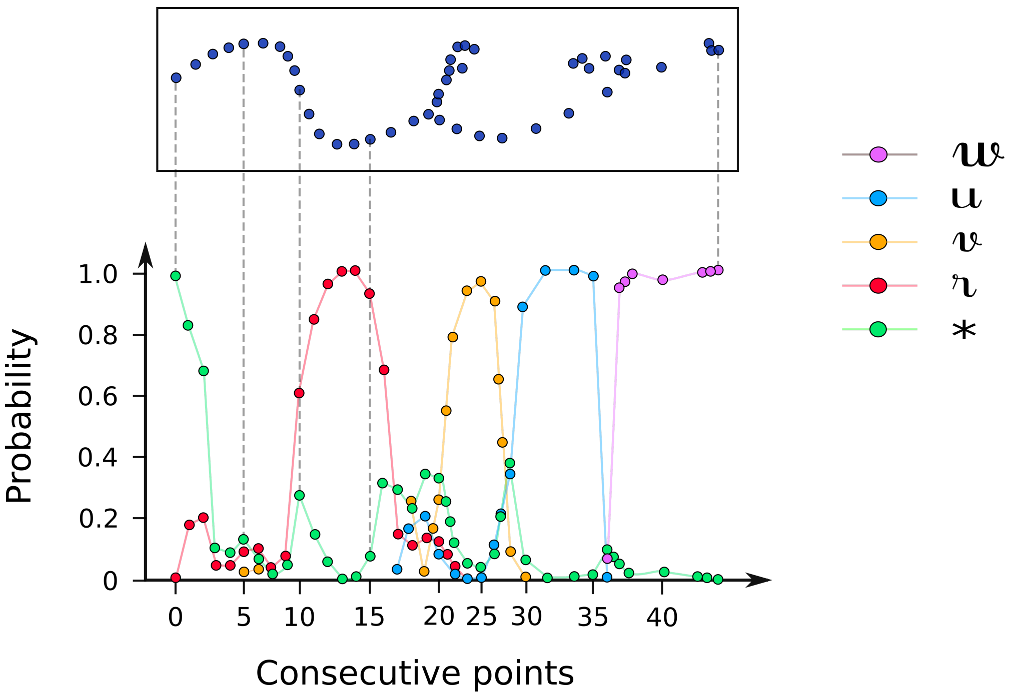

Figure 3.5: An example of glyph 'w' (top) and associated sequence of RNN label predictions (bottom). In the beginning, the most probable label is the star class, since existing glyphs are always longer than a couple of points. Afterward, the model switches to 'r' because the first curve of the glyph 'w' looks similar to the beginning of the glyph 'r'. The model further evolves to 'v', 'u', and finally converge to the correct answer 'w'.

Conclusion and Discussion

the proposed tablet-based test is a first step towards a system to detect handwriting difficulties. We saw that the use of digital tablets has several advantages. With the model presented, the detection of handwriting problems can be done online (compared to 15 minutes of correction by an expert in the case of the BHK test). The model also removes human subjectivity, as it cannot be biased by any of the external parameters that can influence a human. More important, digital tablets allow to explore handwriting on a totally new aspect: its dynamics. In this study, we showed that adding the dynamics of handwriting considerably enhanced the performance of our recognition model, and reduced the confusion between visually similar glyphs. This is the main reason suggesting that developing a digital approach is valuable for this field as the dynamics of the movement is difficult to interpret with the final trace alone, or by a human, even if an expert is observing the child while writing.

At the same time, it is important to notice that this model presents several limitations. First, it is based on the analysis of glyphs while handwriting does not only rely on how a glyph is written but also how they are connected to form a word, a sentence or a text. For instance, within the 13 criteria of the BHK test analyzing handwriting, only 5 are related to single letters. A model allowing to extract the handwriting profile of the child should thus take care of the handwriting as a whole and not only focus on the analysis of letters. Second, even if the proposed model presents interesting insights on explaining handwriting, it is still suffering from a lack of interpretability since it works as a black box.

The next chapter presents our efforts to address the aforementioned limitations.

4 Multidimensional Analysis of Handwriting Difficulties

Introduction

In this Chapter, we present a method allowing the multi-dimensional detection of handwriting difficulties requiring only a commodity tablet. To this end, we modeled data of 298 children including 56 with dysgraphia. Children performed the Concise Evaluation Scale for Children (BHK) test on a digital tablet covered with a sheet of paper. In an effort towards the interpretability of the model describing handwriting from a clinical point of view, we extracted 63 handwriting features describing various aspects of handwriting and used the Random Forest classifier to diagnose dysgraphia

Our method achieved 94% sensibility and 98.9% specificity. Given the statistics on intra-rater and inter-rater levels of agreement in the classical BHK test, our technique has comparable accuracy to experts with numerous advantages including deeper analysis and considerably reduced analysis time.

This work correspond to the following publication:
T. Asselborn, T. Gargot, L. Kidziński, W. Johal, D. Cohen, C. Jolly and P. Dillenboug, "Automated human-level diagnosis of dysgraphia using a consumer tablet", *Nature Digital Medicine*, **2018** [3].

Method

Participants

A total of 242 Typically Developing (TD) children were recruited in 14 primary schools from various suburbs of Grenoble (FRA) to ensure differing socio-economic environments (TD dataset). Children from the first to fifth grade were recruited from 43 classes. None of the TD children included in the study presented known learning disabilities or neuro-motor disorders.

	Dataset	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Age±SD in years	**TD**	6.77±0.29	7.80±0.30	8.75±0.27	9.85±0.35	10.90±0.38
	D	6.83±0.10	7.92±0.67	8.97±0.53	10.08±0.67	10.98±0.61
Males/Females	**TD**	23/26	22/29	21/20	22/26	28/25
	D	2/0	14/6	9/3	14/1	7/0
Right/Left handed	**TD**	41/8	48/3	38/3	41/7	46/7
	D	2/0	16/4	11/1	13/2	7/0

Table 4.1: Summary statistics of the participants involved in this study.

The study also includes 56 dysgraphic children (D dataset) recruited at the Learning Disorders Clinic of Grenoble Hospital (Centre Referent des Troubles du Langage et des Apprentissages, Centre-Hospitalier—Universitaire Grenoble). They were all diagnosed as dysgraphic based on their BHK scores. The scores were assigned by a single rater.

In order to analyse the combined dataset (dysgraphic and non dysgraphic children altogether), we first needed to make sure the age distribution in the two groups was similar. The Kolmogorov-Smirnov test showed no statistical difference (p=0.32) in terms of ages between the two distributions (D and TD datasets). Based on this result and the qualitative assessment of the Q-Q plot (see Figure B.1 in Appendix), we concluded that there was no evidence of a difference between these two distributions and they could be treated jointly in the analysis.

Table 4.1 regroups statistics about the participants of the study.

Data Collection

The 298 children (*TD dataset* + *D dataset*) involved in this study performed the BHK test by writing on a sheet of paper fixed on a Wacom graphic tablet (sampling frequency = 200 Hz; spatial resolution = 0.25 mm). A Wacom Intuos 4 tablet was used for the TD set, and a Wacom Intuos 3 for the D set. Pressure data were carefully calibrated between the two tablets to ensure homogeneity. Details about the calibration can be found in the appendix.

Collecting handwriting parameters include: the x and y coordinates of the pen, the pressure and the pen tilt for every time frame at a maximum sampling frequency of 200Hz, in addition to the age, gender and laterality of the writer.

The BHK tests of the dysgraphic children (*D dataset*) were rated by one expert (always the same) from the hospital of Grenoble. None of the children in the *TD dataset* were evaluated for dysgraphia which means that some of them might be dysgraphic as well.

Features Engineering

To gain a deeper understanding of the relevance of dynamics of handwriting towards dysgraphia diagnosis, and increase interpretability of our model's results, we tried to extract the spectrum of features that could describe handwriting on different aspects such as static, dynamic, tilt or pressure.

The 53 hand-crafted features can be distinguished in 4 categories.

- *Static features* – purely geometric characteristics of a written text.
- *Kinematic features* – dynamics of the path of handwriting.
- *Pressure features* – characteristics of the pressure recorded between the pen tip and the tablet surface.
- *Tilt features* – characteristics of the pen tilt.

The complete description of these features can be found in Appendix C.

Results

Performance of the Model

As described previously, our database is not balanced in terms of positive and negative examples (242 *TD* children versus 56 *D* children), which can skew the model towards a larger subpopulation.

In order to avoid this problem, our data were divided into two disjoint sets.

- *Training set* – 70% of *TD dataset* and 70% of *D dataset.*
- *Testing set* – 30% of *TD dataset* and 30% of *D dataset.*

Due to the unbalance between positive (dysgraphic children) and negative examples (non-dysgraphic children) in the database, the overall accuracy might be misleading (i.e. a model that always predicts non-dysgraphia will be $\sim 75\%$ accurate). Following machine learning literature, we report the **F1-Score**. The F1-Score is the harmonic mean of Precision and Recall. Therefore, the score takes both false positives and false negatives into account, making it more comparable across studies with different proportions of classes. The F1-score is defined as:

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

where:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative} \qquad Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

In our case, the *True Positive* ratio corresponds to the proportion of dysgraphic children correctly labeled dysgraphic, while the *False Negative* ratio refers to the proportion of dysgraphic children incorrectly labeled non-dysgraphic. Finally the *False Positive* ratio defines the proportion of non-dysgraphic children incorrectly labeled dysgraphic.

A k-fold cross validation [98] (with k = 25) on a Random Forest classifier [99] was performed in order to test our model.

For our model, after the 25-fold cross-validation, we obtained a F1-score of 97.50% (Std. of 3.32%). We found this result very satisfactory given the small number of dysgraphic recordings used for training the model. We also conjecture that a larger sample will improve generalizability and robustness of the model.

Robustness of the Model

To validate the robustness of the test, we measured how much data per user we need to use in order to accurately predict dysgraphia. To that end, we trained the model, using only the first seconds of each example, while keeping the same pipeline for training a k-fold validation (k=25). In Figure 4.1, we present the F1-Score as a function of the number of seconds used to train the model. For example, 15 seconds means that only the data recorded during the first 15 seconds of the recordings were used to train the model. We can see that with an analysis window of 15 seconds, the results are already satisfactory (F1-Score of 77.21%) but the high standard deviation (10.34%) may indicate that the model encounter problems of generalisation. With a window of 50 seconds, the F1-Score reaches 93.93% while the standard deviation drops to 4.6%. Afterwards, the results improves only slightly.

We believe that this result indicates the robustness of the proposed features. Indeed, even with the noise generated by the restricted duration (for example 15 seconds) of the test used (a lesser number of examples means less statistical significance), our model still manages to obtain relevant information from the features extracted. This is an indirect benefit compared to the BHK test, in which features need the whole text to be computed.

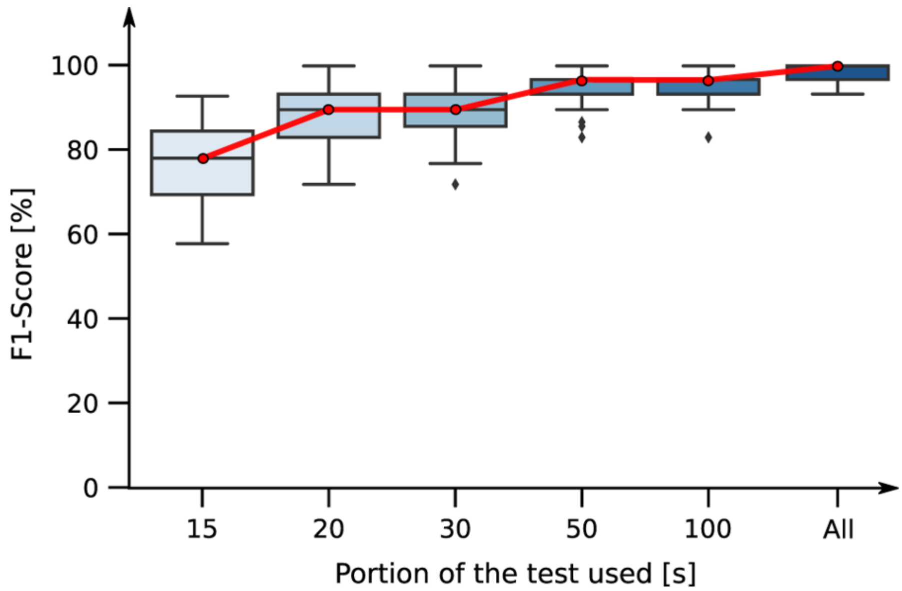

Figure 4.1: Box plot representing the F1-score as a function of the temporal duration of the portion of the examples used for training and testing. We used Random Forest for classification and k-fold cross validation with $k = 25$ for training and testing.

Features Importance

In order to identify which are the most important features found towards discrimination of dysgraphic and non-dysgraphic children, we use the Gini importance [99] since it is known to be one of the most common choices in the literature. Table 4.2 presents the 8 most important features for the random forest model (averaged over the 25 folds). Features related to frequencies (the ones computing medians and bandwidths) seem to be very relevant as 6 out of 8 of the most important features are related to frequencies. Features from all 4 of our categories are represented among the 8 most discriminative features: 3 are kinematic features, 3 are tilt features, 1 is related to pen pressure and 1 is a static feature. In the next section we will further analyze these features.

We notice that only one of these features, the space between strokes, could be potentially extracted if we only have access to the final output of handwriting (like with the BHK test). This confirms the hypothesis raised in Chapter 3 were we showed that dynamic brings very relevant information to the analysis of handwriting. This also reassures us about the value of digital tablets to assess handwriting, as such devices give us access to important information previously inaccessible to clinicians analyzing pen-and-paper tests like the BHK.

Rank	**Category**	**Name**	**Importance** (SD.)[%]
1	Kinematic	(#19) Median of Power Spectral of Speed Freq.	15.71 (9.06)
2	Kinematic	(#18) Bandwidth of Power Spectral of Speed Freq.	12.08 (8.00)
3	Pressure	(#27) Mean Speed of Pressure Change	9.81 (6.52)
4	Static	(#3) Space Between Strokes	7.45 (6.73)
5	Tilt	(#58) Dist. to Mean of Speed of Tilt-X Change Freq.	6.07 (4.30)
6	Kinematic	(#22) Dist. to Mean of Speed Change Freq.	5.18 (4.73)
7	Tilt	(#54) Bandwidth of Speed of Tilt-X Change Freq.	4.10 (4.64)
8	Tilt	(#60) Median of Power Spectral of Tilt-Y Change Freq.	2.97 (3.33)

Table 4.2: We report the ranks, feature categories, and importance averaged for the 25 folds (with standard deviation over all folds) of the 8 most discriminative features.

Conclusion and Discussion

Clinical Features Analysis

Static features

The most discriminative static feature we found was the (#8) *Bandwidth of tremors frequencies* (see left graph of Figure 4.2). This feature represents the range of tremors frequencies found in the handwriting of the writer under investigation. A high value for this feature means that many tremors were extracted from the handwriting of children with dyshraphia. In Figure 4.3-a, we present an example of handwriting from a non-dysgraphic child (on the top left) and from a dysgraphic child (on the top right). The handwriting of the non-dysgraphic child appears to be visually smooth. Conversely, the handwriting of the dysgraphic writer is not smooth, we can easily see some shaking at a high frequency (like in the apostrophe or at the end of the *"a"*). We hypothesize that this characteristic results in an important value of the (#8) *Bandwidth of Tremors Frequencies* feature. The dysgraphic child hesitates more in the drawing of the letters and it is harder for them to smoothly control the pen. Shakiness is thus indicative of poor motor, control resulting in more noise or indicative of prior knowledge of the letter. Interestingly, this feature is related to the BHK item *hesitation and shaking.* According to the therapists, a score of 0/5 (highest score) was obtain by the non-dysgraphic child (left sample presented in Figure 4.3-a) for this feature while a score of 3/5 (low score) was obtained by the dysgraphic child (right sample presented in Figure 4.3-a).

Concerning the (#3) *Space Between Strokes* feature, we can see in the right graph of Figure 4.2 that non-dysgraphic children tend to put more space between the letters they write. This is related to the BHK item called *narrow words.* This BHK item indicates pathology if it is not possible to put the letter *"o"* between each pair of words, meaning that not enough space is left. Moreover, in case of dysgraphia, the writing is hardly automatized, leading to irregular spaces between strokes.

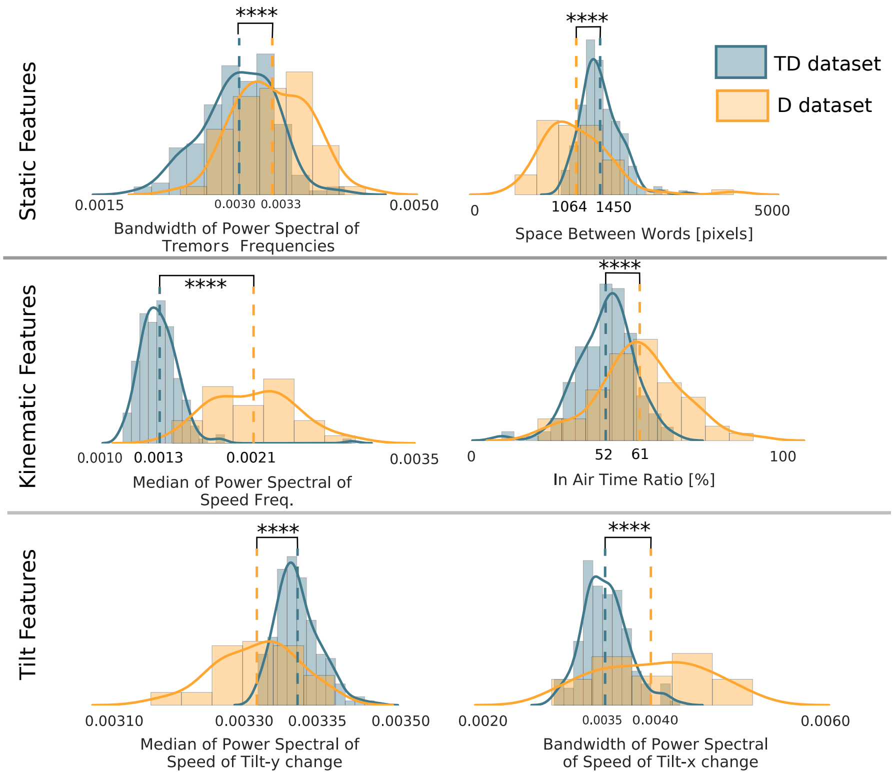

Figure 4.2: Distribution of the dysgraphic children (D dataset) and the non-dysgraphic children (TD dataset). For static features: (#8) Bandwidth of Tremors Frequencies and the Space Between Strokes features. For kinematic features: (#19) Median of Power Spectral of Speed Frequencies and the (#23) In Air Time Ratio features. For tilt features: (#60) Median of Power Spectral of Speed of Tilt-y change and the (#54) Bandwidth of Power Spectral of Speed of Tilt-x change features. Significant differences between the two distributions are found on all listed features (Wilcoxon test).

Kinematic features

The most discriminative kinematic feature we found was the (#19) *Median of the Power Spectral of Speed Frequencies*. This feature indicates that the speed frequencies of the dysgraphic children are shifted toward high frequencies (see center left graph of Figure 4.2). In Figure 4.3-b, we present an example of handwriting coming from a non-dysgraphic child on the left and from a dysgraphic child on the right. The color corresponds to the handwriting speed when the points were recorded. In the dysgraphic child's handwriting we observe very fast changes of speed (high acceleration and deceleration) contrary to what can be found in the handwriting of the non-dysgraphic writer. It is interesting to notice that the features linked with the acceleration were not found discriminative as these sudden changes of speed are local (compensated by long period of constant speed). These sudden changes of speed are translated in high frequency during the Fourier transform. This feature relates to the fact that we can find more saccades during the handwriting of the dysgraphic child due to lack of automation and control of his/her hand movement. In the case of the BHK test, the only feature related to the kinematics of handwriting is the number of characters written after 5 minutes. The results of this very basic feature show that the dysgraphic children are slower than the non-dysgraphic ones.

Another interesting feature that was found discriminative was the (#23) *In-Air-Time Ratio* (proportion of time spent with the pen not touching the surface of the tablet) as can be seen in the right center graph of Figure 4.2. This result appears to be in line with previous findings [100].

Tilt features

As can be seen in the bottom row of Figure 4.2, the frequencies extracted from the speed of tilt change are very discriminative of dysgraphia. Concerning the Tilt-y, in contrast to what was observed for other categories of features, we can see that the non-dysgraphic children seem to exhibit higher frequencies during their handwriting. This finding is highlighted in Figure 4.3-c. For every point recorded, the color of the trace is representing the speed of the Tilt-y change. We can see that the dysgraphic child stays very constant maintaining the Tilt-y of its pen (showing almost no variations in the Tilt-y change speed, with small absolute value). The non-dysgraphic child, in contrast, presents very fast variations in his/her tilt-y change speed (big acceleration and deceleration). These very sudden variations are translated in high frequencies in the Fourier domain shifting the median of the power spectral to high frequencies. Thus, we can infer that the TD child is able to change very frequently and quickly the tilt of his/her pen in the y axis, whereas the dysgraphic child displays less tilt-y variation abilities, probably due to a more constraint and rigid pen grip.

Concerning the Tilt-x, although the distribution of dysgraphic children is very spread (see the right graph of Figure 4.2), the dysgraphic children seem to present a larger range of frequencies in their handwriting concerning the speed of Tilt-x change than typically developing children. This means that they are not constant in the way they move their pen in the ZX plane (see Figure C.4 in appendix). This finding is highlighted in Figure 4.3 bottom. For every point recorded, the color of the trace is representing the speed of the tilt-x change. We can see more variations in the speed of Tilt-x change for the dysgraphic child compared to the one of the non-dysgraphic child who seems to present more control in his/her movement. Contrary to the pen tilt in the direction perpendicular to the handwriting global direction (perpendicular to the lines of the paper sheet), proficient writers exhibit less variations (more control) in the speed of tilt change (and also the tilt itself) in the direction of the lines of the paper sheet.

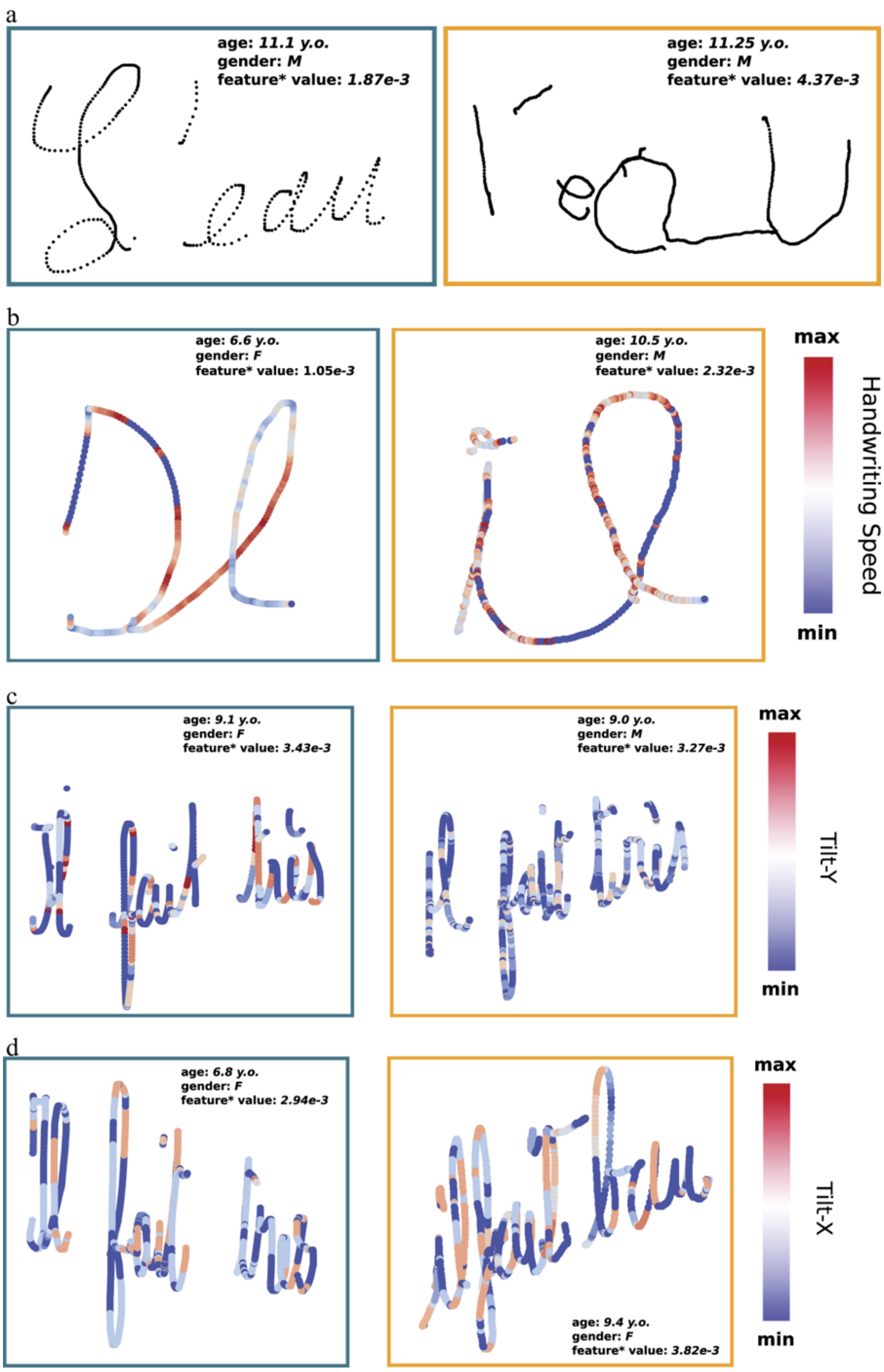

Figure 4.3: A comparison of different metrics for a non-dysgraphic child (left) and a child with dysgraphia (right).

Main Findings

Following the conclusions obtained in Chapter 3, this model assess handwriting not only from a static perspective, but also takes advantage of the dynamic-, pressure- and tilt-related information to deliver a precise diagnosis of dysgraphia: on average, 95.6% (standard deviation of 4.29%) of the writers with severe handwriting problems were correctly diagnosed by our model, with 0.6% (standard deviation of 3.6%) of false positive rate. The final model reached F1-Score of 97.72% (standard deviation of 2.85%). In addition, our diagnosis system has the advantage to be very cheap (beside the cost of the tablet) and fast requiring only a few milliseconds to deliver the diagnosis compared to 15 minutes for the BHK test, allowing an online usage. Our method also reduces subjectivity as the model is permeable to all the external parameters that can bias a human. It should be noticed that the inter-rater correlation in the correction of the BHK test is of 0.89. Since our algorithm outperforms this value, we conclude that the algorithm learned to mimic the rater. These findings suggests that adding data from other raters should not only reduce the bias but also allow us to surpass the accuracy of each individual rater.

If the model presented in this chapter is efficient in detecting severe handwriting problems, it still does not allow to extract the child handwriting profile which is needed to understand the specific weaknesses of the child and propose proper remediation. Although the proposed features are a step forward towards interpretability, they still cannot be interpreted in their raw forms since they present complex interaction with the child's age or gender for example as will be discussed in later chapters. A more complex system, involving several additional steps built in order to give more interpretability to the features is discussed in Chapter 7.

It is interesting to see that among the 53 features used by our model, most of them are very technical and "low level", i.e. measuring micro-skills in terms of writing mechanics. In that way, our test is more robust to differences in handwriting style, language, or understanding of the text by the subject. Indeed, these features (for example the 3 most discriminative features: the (#18) *Bandwidth of Power Spectral of Speed Frequencies*, the (#19) *Median of Power Spectral of Speed Frequencies* and the (#27) *Mean Speed of Pressure Change*) can be interpreted the same way independently of the language or the handwriting direction. For example, languages written from right to left, like Hebrew, or from left to write, like French, still share the same low-level characteristics. This idea will be further developed in the next Chapter of this thesis, where we are interested in transferring the model from the Latin alphabet to the Cyrillic one.

5 Extending Diagnosis to Other Alphabets

Introduction

In this Chapter, we aim to investigate how a model assessing handwriting difficulties based on the features described in the previous Chapter can be transfered from one alphabet to another. The hypothesis raised in the previous Chapter is that the features on which the model is based are describing very low level, almost physiological, aspect of handwriting and thus independent of the writing content. Under this assumption, they should be able to capture writing problems not only regardless of the written sentence, but also regardless of the language and alphabet in which it is written.

Kazakhstan recently adopted a state program for the development and functioning of languages for 2011-2020. This new tri-lingual education policy aims for the Kazakhs to develop fluency in three languages: Kazakh, Russian and English. Additionally, a recent motion to transfer the Kazakh language from Cyrillic to the Latin alphabet was approved by the Kazakh authorities in October 2017 [101]. While there are clear reasons for these reforms, there are numerous risks surrounding the transfer, including increasing inequalities in educational services (e.g., preferences for Russian-speaking schools), causing illiteracy in adults in their native language and bringing about disinterest and lack of motivation in reading and writing in Kazakh among Kazakh and non-Kazakh children and adults. Additionally, the change in alphabet may have a deep impact on the handwriting of Kazakh children since Cyrillic and Latin letters are very different (e.g., only four Cyrillic letters also appear in the Latin alphabet, i.e., a, e, i and o). Since handwriting is a central aspect of school education in the majority of countries, including Kazakhstan, it is very important to understand what impacts the change in alphabet may have on children's school education in order to prevent or anticipate possible problems. In line with this societal concern, we aim to investigate to what extent a person can transfer his/her handwriting from one alphabet to another, i.e. whether the specificity of handwriting is kept (whether the children's particularities are preserved regardless whether they are writing with the Cyrillic or Latin alphabet). This is a required condition to create a cross-lingual model for the detection of handwriting difficulties.

In this Chapter, we conducted a study with 190 children aged 6-11 years old who are native in the Kazakh language. The children were asked to copy a short text in both the Cyrillic and Latin alphabets onto a tablet. In the next sections, we present the analysis used to compare the children's handwriting in the two scripts in terms of the various aspects of handwriting (the static, kinematic, pressure and tilt features introduced in the previous Chapter) as a function of age. One of the objectives of the analysis is to understand the different types of learning transfer appearing between the two alphabets in function of the grades along the different aspects of handwriting. We also explored how the handwriting specificity found in one alphabet can be translated in the other alphabet. Our results show that our features are able to coherently describe the handwriting quality in the two alphabets, thus confirming our hypothesis of features independence from the writing content, and opening the possibility to create a cross-lingual model for the detection of handwriting difficulties.

This work correspond to following publication (submitted):

T. Asselborn, W. Johal, B. Tleubayev, Z. Zhexenova, P. Dillenboug and A. Sandygulova, "The transferability of handwriting skills: from the Cyrillic to the Latin alphabet", *Nature Science of Learning*, **2019**

Method

Participants

A sample of 200 children was recruited from local primary schools in Astana, the capital of Kazakhstan. The children came from diverse socio-economic backgrounds and were all native Kazakh speakers and writers. The children were 6-11 years old (M = 8.48, SD = 1.2). Out of the 200 children who volunteered, data for 10 children were not complete and thus could not be included in the analysis, yielding a total of 190 (90 boys and 100 girls) valid and complete observations. The remaining children were enrolled in four grades: 29 children were first graders, 71 students were in grade 2, 41 children were in third grade and 49 students were in grade 4. Table 5.1 presents an overview of the sample of participating children.

Ethics Statement

This research was approved by Nazarbayev University's ethics committee. Informed consent was obtained in writing from all children and their parents. Supporting information included an assent form for children and an informed consent form for parents or guardians. Children received a brief explanation of the purpose of the study and the procedures involved in data collection. Assent and consent forms were distributed to children in their classrooms in the presence of their teachers. Children were asked to show the forms to their parents at home and submit them to their teachers, who then collected the forms for us during the days that followed.

Data collection

Each child was individually taken outside of classroom for approximately fifteen minutes and led to an empty classroom. It was explained that they had to write on a Wacom Cintiq Pro 13 tablet using its stylus and that the task was to copy the text shown on a piece of paper that was positioned in front of them. There were two versions of the same five sentences, i.e., one in each script. The text was designed to include diverse letters, in particular many specific Kazakh characters that are not present in English.

A data collection tool was used to save the children's demographic information (age, gender, grade, laterality) and the two handwritten texts, separately. As children finished writing the first five sentences in one script, they were told to press the save icon and then write the next five sentences in another alphabet.

The design of our experiment replicated the conditions which will be present in 2020: at the moment of data collection, children had spent different amounts of time practicing handwriting according to their grade. Children in grade 1 had spent approximately 6 months learning to write in Cyrillic script. For each subsequent grade level, we add an additional year of experience (12 months). Thus, children in grade 2 had spent approximately 18 months writing in Cyrillic, third graders had spent around 30 months writing in Cyrillic and fourth-graders had spent around 42 months (3 years and 6 months) writing in Cyrillic. The children practiced handwriting for 6 hours per week, starting from simple shapes and moving to Cyrillic letters after approximately 6 weeks. Children of all grades had 2 hours of English per week where they also started writing English letters after approximately 6 weeks, and the learning time changed with grade level in a manner similar to the learning time for the Cyrillic script (i.e., 6 months of handwriting in English for the first graders). However, none of the children had been introduced to a Latin-based Kazakh alphabet (extended Latin alphabet with six additional letters) and its associated handwriting. Thus, in contrast to the Cyrillic script, the learning time for the new script was the same across all grade levels, i.e. it was equal to 0 (see Table 5.1).

Features extraction

In the work presented in Chapter 4, 63 handwriting features were defined and used to train a random forest classifier to diagnose severe handwriting problems (dysgraphia). In

	Grade 1	Grade 2	Grade 3	Grade 4
Male/Female	13/16	33/38	21/20	23/26
Age (std.) [years]	6.94 (0.4)	7.94 (0.4)	8.88 (0.4)	10.16 (0.6)
Left handed/Right handed	27/2	68/3	37/4	46/3
Cyrillic average learning time [months]	6	18	30	42
Latin average learning time [months]	0	0	0	0

Table 5.1: Summary statistics for the participants involved in this study. Learning pace was 6 hours per week for Kazakh handwriting classes in Cyrillic alphabet and 2 hours per week for English classes (with Latin alphabet). No time was devoted to Kazakh handwriting classes in Latin alphabet.

this work, we only used the features that were found to be the most important in the aforementioned model, according to the Gini importance metric (averaged with a k-fold cross validation, $k = 25$). To maintain a good balance and compare the different groups of features, we selected the most important features from each of the following four groups that we distinguished: static, pressure, kinematic, and tilt. The complete description of these features can be found in Appendix C. Please notice that every possible mismatch between the features importance presented in Chapter 4 and this one are due to the retraining of the model with new data that happened in the meantime.

Results

Absolute Features difference

The descriptions of the features used in this study as well as their clinical interpretations with regards to handwriting can be found in Appendix C as well as in Chapter 4.

For every feature and for all grades (from 1st to 4th), a Wilcoxon test was conducted to detect if the distribution of a given feature in the Cyrillic alphabet is similar to the distribution of this same feature in the Latin alphabet. This statistical test was used since not all the distributions were found to follow a normal distribution. The results can be found in Table 5.2. Additionally, the means and standard deviations for all the distributions (by grade and alphabet) can be found in Appendix D.

Before discussing the results, it is important to notice that the children involved in this experiment were bilingual from birth (Kazakh and Russian languages), attended a Kazakh-speaking school and studied English as a foreign language for a few hours per week starting from grade 1. The study was conducted in Spring 2019 after the change in script had already been announced by the Kazakh authorities (February 2018) but before it was applied. Thus, none of the participating children had received any formal training in the new Latin-based Kazakh script since pilot studies were planned for the 2019-2020 academic year. Within this context, we were able to replicate the conditions that will be present at the moment of transition (see Table 5.1).

Different types of features are noticeably different in their co-evolution across the grades within the two alphabets.

Transferable Features Affected by Alphabet Differences

Significant differences possibly stemming from intrinsic differences between the two alphabets were observed. Since the learning time difference between the two alphabets grows by grade (see Table 5.1), features that are different for all the grades (across four developmental ages) might be labelled as related to intrinsic alphabet differences (a typical example of such a feature can be found in Figure 5.2-C). Three of the four pressure and kinematic features, as well as one static feature, belong to this category.

The *Mean Speed of Pressure Change* (#27) and *Max Speed of Pressure Change* (#28) were found to be significantly higher when children were writing in the Latin alphabet compared to when they were writing in the Cyrillic alphabet across all grades (see Figure 5.2-B and Table D.1). These results are surprising since these features were found to be strongly inversely correlated with handwriting difficulties as shown in Chapter 4 (high values for these features can be translated into higher degrees of handwriting automation). The results found here appear to be abnormal since the mean scores for these features were higher when the children were writing in Latin, leading us to believe that the differences in these features were the result of intrinsic differences between the two alphabets. The same effect was found for the two most discriminative features used to diagnose handwriting difficulties in the Latin alphabet in Chapter 4: the distributions of the *Median of the Power Spectral of Speed Frequencies* (#19) and the *Bandwidth of the Power Spectral of Speed Frequencies* (#18) were found to be significantly different in Latin and Cyrillic for most of the grades.

It is also very interesting to note the evolution of these features with grade. As illustrated in Figure 5.2-B, for both alphabets, a shift appears in the direction of handwriting proficiency with grade (the max speed of pressure change increases as children move from lower to higher grades). Even when there is an absolute significant difference between the two alphabets for this feature, its evolution follows the same path for both alphabets. In other words, this feature is likely to still be an indicator of handwriting automation for both alphabets, provided that the alphabet-intrinsic shift is taken into account. This phenomena is observed in all the features described in this Section.

The *Nb of Peaks of Pressure Change Per Sec* (#31) was found to be consistently higher when children were writing with the Cyrillic alphabet compared to when they were writing with the Latin alphabet, regardless of grade (see Table in Appendix D). The *Median of the Power Spectral of Pressure Frequencies* (#33), a proxy for handwriting automation, was also found to be significantly different between the two alphabets whatever the grade. The difference, however, does not appear abnormal: if we interpret these features with

Feature		Grade 1 W stat	Grade 1 pvalue	Grade 2 W stat	Grade 2 pvalue	Grade 3 W stat	Grade 3 pvalue	Grade 4 W stat	Grade 4 pvalue
Bandwidth Tremolo - **#8**	Static	120	0.065	750	< 5e−2	220	0.065	510	0.95
Median Tremolo - **#9**		100.0	< 5e−2	630.0	< 1e−3	290.0	0.069	530.0	0.41
Space Between Strokes - **#3**		110.0	< 5e−2	390.0	< 1e−3	130.0	< 1e−3	430.0	0.076
Handwriting Moment - **#1**		70.0	< 1e−2	730.0	< 1e−2	280.0	< 5e−2	420.0	< 5e−2
Handwriting Density - **#4**		200.0	0.79	1300.0	0.97	220.0	< 1e−2	430.0	0.074
Mean Velocity - **#13**	Kinematic	200.0	0.63	980.0	0.088	210.0	< 1e−2	320.0	< 1e−2
Max Velocity - **#14**		110.0	< 5e−2	260.0	< 1e−3	80.0	< 1e−3	200.0	< 1e−3
In-Air-Time Ratio - **#23**		38.0	< 1e−2	420.0	0.92	140.0	0.84	150.0	0.24
Bandwidth Speed - **#18**		160.0	0.6	710.0	< 5e−2	330.0	0.76	260.0	< 5e−2
Median Speed - **#19**		160.0	0.24	810.0	< 1e−2	280.0	< 5e−2	270.0	< 1e−3
Mean Pressure - **#24**	Pressure	110.0	< 5e−2	1200.0	0.5	390.0	0.58	440.0	0.094
Mean Speed of Pressure Change - **#27**		120.0	< 5e−2	380.0	< 1e−3	180.0	< 1e−2	190.0	< 1e−3
Max Speed of Pressure Change - **#28**		190.0	0.55	660.0	< 1e−3	160.0	< 1e−3	280.0	< 1e−3
Nb of Peaks of Pressure Change per secs - **#31**		2.0	< 1e−3	47.0	< 1e−3	55.0	< 1e−3	64.0	< 1e−3
Median Pressure - **#33**		50.0	< 1e−3	430.0	< 1e−3	180.0	< 1e−3	240.0	< 1e−3
Std. of of Tilt-X - **#40**	Tilt	140.0	0.082	280.0	< 1e−3	150.0	< 1e−3	200.0	< 1e−3
Std. of Speed of Tilt-X Change - **#46**		130.0	0.056	720.0	< 1e−2	200.0	< 1e−2	210.0	< 1e−3
Median Tilt-y - **#60**		120.0	< 5e−2	630.0	< 1e−3	310.0	0.12	470.0	0.17

Table 5.2: Feature differences between the two alphabets (Cyrillic and Latin) across grades. For each feature, a Wilcoxon test was conducted to see if there was a statistical difference in the means of the two distributions. Shapes of red denote the level of statistical significance of each comparison.

the Latin model developed in Chapter 4, the handwriting using the Cyrillic alphabet is found to be of higher quality. The *Maximum Velocity* (#14), always higher when children were writing in Cyrillic, can also be interpreted in the same manner (see Table in Appendix D).

Two tilt features can also be sorted into this category. The *Std. of Tilt-X* (#40) and the *Std. of Speed of Tilt-X change* (#46) were always different between the two alphabets. Once again, children writing with the Cyrillic alphabet exhibited a higher standard deviation, which appears to be abnormal. In this sense, we believe that the difference comes from the alphabets requiring different pen manipulation styles (and thus pen tilts) during writing.

Finally, the *Space Between Strokes* (#3) feature was found to be consistently higher when children wrote in Cyrillic compared to Latin (see Table in Appendix D). This feature computes the mean distance between strokes, which, in general, is the distance between words. In the case in which a word is composed of letters requiring writers to raise their pen in the middle, the distance traveled between the moment the pen rises and the moment the pen touches the surface again is taken into account in the computation of this feature. We believe that the differences observed here signify more than an intrinsic difference between the two alphabets and that they are related to the difference in the learning time spent on each alphabet. Indeed, since our study provided their first opportunity to write in Latin, some children tended to write the letters one after the other with less continuity than they used when writing in Cyrillic (as can be seen in Figure 5.1). Hence, there tended to be small distances between letters in words when writing with the Latin alphabet, which, when averaged, we believe, were responsible for the noticed difference.

Features with Slow Transfer Between Alphabets

Other features, such as the *Mean Velocity* (#13), presented interesting trends. Figure 5.2-A displays the distributions for the *Mean Velocity* (#13) for the two alphabets for grades 1 to 4. Note that the mean of the distribution for the Cyrillic alphabet is always higher than the mean for the Latin alphabet. This difference appears to be a result of the difference in learning time between the two alphabets. We also see that the two distributions are shifting to the right (in the direction of handwriting proficiency) with grade, meaning that, in both alphabets, the children wrote faster and faster with grade, which is both an indicator and a direct consequence of the level of automation in their handwriting. We believe that this trend is related to the learning transfer between the two alphabets. Note also that the mean velocity is increasing more rapidly for the Cyrillic alphabet. We presume that the increasing difference between the two distributions is due to a "leak" in the learning transfer from one alphabet to the other.

The *handwriting density* (#4) follows the same trend: in both alphabets, children's writing became smaller and smaller with grade (an indicator of handwriting quality), but the difference between alphabets grew (children wrote significantly smaller in Cyrillic than in Latin in grade 3 and marginally smaller in Cyrillic in grade 4).

Features Which are Transferable When the Level of Automation is Higher

Unlike those described in the previous section, some of the features exhibited decreasing differences between the two alphabets as grade progresses. For instance, this trend can be seen in the two features describing shakiness: the *Bandwidth of the Power Spectral of Tremor Frequencies* (#8) and the *Median of the Power Spectral of Tremor Frequencies* (#9). As illustrated in Figure 5.2-D, the general trend show a decrease in the *Median of the Power Spectral of Tremor Frequencies* (#9) with grade, meaning that, for both alphabets, the average level of shakiness decreased with age, which is a direct consequence of the increasing level of automation children acquire. Interestingly, the difference between the two alphabets seems to decrease: children writing in Latin presented a significantly higher level of shakiness compared to when writing in Cyrillic in the lower grades (1st and second grade), while the difference is nearly non-existent by grade 4. Hence, the transfer between the two alphabets is in line with the level of automation. In other words, as automation increases, better control of the pen in one alphabet is beneficial also for the other alphabet.

Other features seem to follow the same trend. Such is the case for the so-called *In-Air-Time ratio* (#23), a feature that was found to be highly correlated with handwriting problems in various studies conducted using the Latin alphabet [3] as well as the Hebrew alphabet [100, 83]. We can see that the *In-Air-Time Ratio* (#23) is always smaller for the Cyrillic alphabet (see Table in Appendix D), which is a sign of a better knowledge of the Cyrillic alphabet by heart and of better motor program (memory) of the letters. Although the difference between the two alphabets was statistically significant for the lowest grade, no differences were found for subsequent grades.

The *Median of the Power Spectral of Tilt-y Frequencies* (#60) also follows the same trend. The median was always higher when the children were writing in Cyrillic (which is in line with the results reported in Chapter 4). Moreover, we can see that the difference between the two alphabets decreased with grade.

Finally, the *Handwriting Moment* (#1) also falls into this category. Children were writing in "straight lines" to a greater degree when writing in the Cyrillic alphabet, which appears to be a normal result (writing in a straight line is an indicator of handwriting quality). The difference in "straightness" between the two scripts was found to decrease with grade.

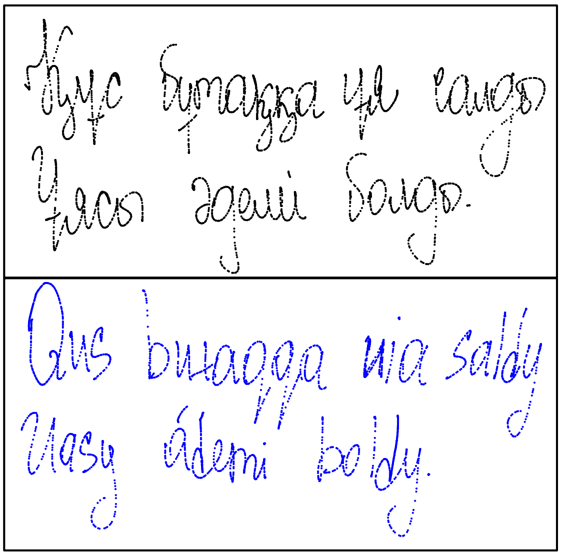

Figure 5.1: Two handwriting samples produced by the same child. The text written in the Cyrillic alphabet is in black, while the text written in the Latin alphabet appears in blue.

Relative Feature Difference

In this Section, we want to assess whether the change of alphabet brought about handwriting differences relative to other children. Concretely, we will no longer focus on raw data differences but rather on the relative differences between individuals. In other words, we are interested in investigating if children who presented a low value for one feature in one of the alphabets **relative to the other children** (of the study) still presented a low value for this same feature in the other alphabet **relative to the other children** (of the study).

In Table 5.3, the Spearman rank correlation is reported for all features and grades. The Spearman Rank correlation coefficient is based on the ranked values for each variable rather than the raw data. The correlation between two variables (e.g., feature X in Cyrillic and feature X in Latin) will be high when the observations have similar ranks (i.e., the relative position labels of the observations within the feature X: 1st, 2nd, 3rd and so forth) in the two alphabets. In other words, if a child has a low (high) value for one feature compared to other children in Latin and still has a low (high) value of this same feature in Cyrillic, the Spearman rank correlation will be high.

In general, most features seem to be highly correlated between the two alphabets. Indeed, 90% of the correlations we computed were significant. This is a very interesting result since it shows that children are able to preserve their handwriting specificity from one alphabet to another. The implication of such a result will be developed in the Discussion. Note that the correlation levels increase with grade for the majority of the features. We

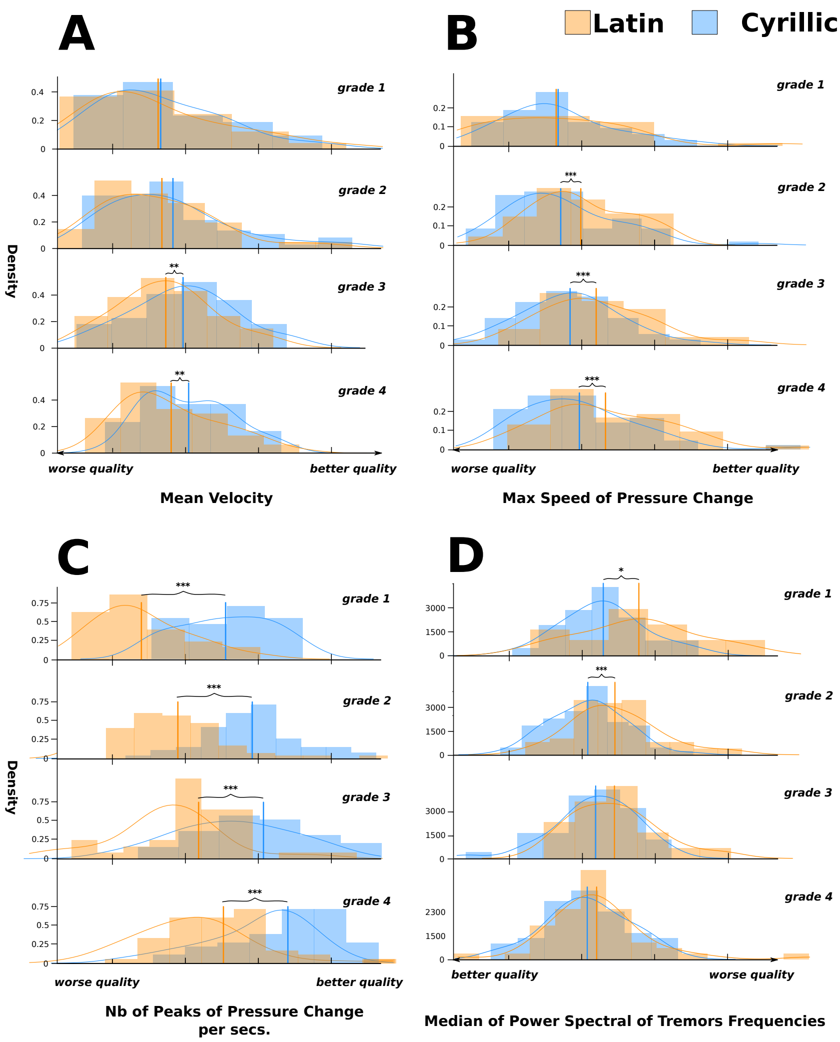

Figure 5.2: Distributions of specific features according to grade for the two alphabets (Cyrillic alphabet in blue, Latin alphabet in orange). The vertical lines represent the means of the distributions. **A**: Mean Velocity feature. **B**: Mean Speed of the Pressure Change feature. **C**: Nb of the Peaks of Pressure Change Per Sec. **D**: Median of the Power Spectral of Tremor Frequencies. The handwriting quality aspect along the x-axis was computed according to the Latin standard (see Chapter 4).

believe that this result is a consequence of the increased level of automation shown by children as they age. Indeed, as they reach higher grades, their movements are more controlled, resulting in less variability in the way they write. In other words, a child in first grade has a tendency to write the same letter differently each time (i.e. with different pressure pattern, different speed, different acceleration and so on), while a child in fourth grade will always write it in the same manner. In this sense, children are able to transfer their motor skills from one alphabet to the other, e.g., a specific pattern of pressure used to write one shape in an alphabet will be used to write a similar shape in another alphabet.

Conclusion and Discussion

Two types of features were used in this study to describe handwriting: static features, which refer to the graphical traces of handwriting, and dynamic features, which describe the kinematics, pressure and tilt aspects of handwriting. The 18 features used in this study were all found to explain the quality of handwriting for the Latin alphabet in the study conducted by Asselborn et al. [3] and presented in Chapter 4. We can see a significant absolute difference between the Latin and Cyrillic alphabets for the majority of the features across four grades. However, it is more interesting to follow the evolution of these features over the years, knowing that children in grades 1 to 4, respectively have 0.5 to 3.5 years of Cyrillic practice, a similarly evolving but much smaller experience with using the Latin alphabet to writing in English, and no experience at all with writing Kazakh with the Latin alphabet (replicating conditions to be found at the moment of introduction of the new alphabet). Figure 5.2 shows three different modalities of co-evolution of the handwriting in the two alphabets over grades, indicating different forms of learning transfer.

In quarter A of Figure 5.2 the differences between features increase between grades 1 and 4: both improve, but Cyrillic improves faster. This evolution reveals that some features, such as average handwriting velocity, improve in Latin, but not as fast as they do in Cyrillic, which we interpret as a slow transfer of these skills rather than a nonexistent transfer.

In quarters B and C of Figure 5.2 The differences in features are approximately constant across grades, i.e., the features improve in both the Cyrillic and Latin alphabets and in a similar fashion. This type of evolution is evidenced by most of the pressure and tilt features. The stability of the differences across grades probably originates from the intrinsic differences between the two alphabets. We tentatively infer three conclusions from these data: (1) Once again, we observe that the dynamic features of handwriting are of prime importance for analyzing handwriting quality since they reveal aspects of

Feature		Grade 1 corr	Grade 1 pvalue	Grade 2 corr	Grade 2 pvalue	Grade 3 corr	Grade 3 stat	Grade 4 corr	Grade 4 pvalue
Bandwidth Tremolo - **#8**	**Static**	0.011	0.96	0.029	0.81	0.3	0.059	0.014	0.93
Median Tremolo - **#9**		0.21	0.27	0.13	0.27	0.29	0.065	0.31	$< 5e{-}2$
Space Between Words - **#3**		0.77	$< 1e{-}3$	0.58	$< 1e{-}3$	0.52	$< 1e{-}3$	0.46	$< 1e{-}3$
Handwriting Moment - **#1**		0.35	0.066	0.42	$< 1e{-}3$	0.68	$< 1e{-}3$	0.56	$< 1e{-}3$
Handwriting Density - **#4**		0.73	$< 1e{-}3$	0.75	$< 1e{-}3$	0.73	$< 1e{-}3$	0.74	$< 1e{-}3$
Mean Velocity - **#13**	**Kinematic**	0.72	$< 1e{-}3$	0.72	$< 1e{-}3$	0.69	$< 1e{-}3$	0.69	$< 1e{-}3$
Max Velocity - **#14**		0.4	$< 5e{-}2$	0.55	$< 1e{-}3$	0.37	$< 5e{-}2$	0.34	$< 5e{-}2$
In-Air-Time Ratio - **#23**		0.45	$< 5e{-}2$	0.53	$< 1e{-}3$	0.4	$< 1e{-}2$	0.53	$< 1e{-}3$
Bandwidth Speed - **#18**		0.52	$< 1e{-}2$	0.53	$< 1e{-}3$	0.22	0.17	0.53	$< 1e{-}3$
Median Speed - **#19**		0.22	0.25	0.56	$< 1e{-}3$	0.54	$< 1e{-}3$	0.4	$< 1e{-}2$
Mean Pressure - **#24**	**Pressure**	0.95	$< 1e{-}3$	0.68	$< 1e{-}3$	0.76	$< 1e{-}3$	0.81	$< 1e{-}3$
Mean Speed of Pressure Change - **#27**		0.67	$< 1e{-}3$	0.57	$< 1e{-}3$	0.37	$< 5e{-}2$	0.54	$< 1e{-}3$
Max Speed of Pressure Change - **#28**		0.84	$< 1e{-}3$	0.55	$< 1e{-}3$	0.69	$< 1e{-}3$	0.54	$< 1e{-}3$
Nb of Peaks of Pressure Change per secs - **#31**		0.68	$< 1e{-}3$	0.5	$< 1e{-}3$	0.47	$< 1e{-}2$	0.37	$< 1e{-}2$
Median Pressure - **#33**		-0.045	0.82	0.29	$< 5e{-}2$	0.49	$< 1e{-}2$	0.41	$< 1e{-}2$
Std. Tilt-X - **#40**	**Tilt**	0.53	$< 1e{-}2$	0.58	$< 1e{-}3$	0.6	$< 1e{-}3$	0.7	$< 1e{-}3$
Std. of Speed of Tilt-X change - **#46**		0.71	$< 1e{-}3$	0.7	$< 1e{-}3$	0.58	$< 1e{-}3$	0.7	$< 1e{-}3$
Median Tilt-Y - **#60**		-0.25	0.19	-0.012	0.92	0.22	0.16	0.37	$< 1e{-}2$

Table 5.3: Feature similarity between the two alphabets (Cyrillic and Latin) for different grades. For each feature, a Spearman's rank correlation was calculated in order to examine the similarities between the two distributions. The green color denotes the level of statistical significance.

the writing process that the final graph fails to disclose. (2) As these features evolve in a similar way across grades in both alphabets, we conclude that the metrics we developed for Latin handwriting remain valid for Cyrillic handwriting. (3) Since these features indicate an improvement in writing in Latin with age, despite the fact that all children had the same experience in Latin in all grades, we may also conclude that a positive transfer of fine motor skills occurs between the two alphabets: any handwriting amelioration (captured by a positive evolution of a feature) in the Cyrillic alphabet seems to bring a positive amelioration in the Latin alphabet too.

In quarter D of Figure 5.2 The differences in features decrease between grades 1 and 4: both improve, but Latin seems to catch up with Cyrillic in terms of quality. This type of evolution is seen, for example, in two features that capture the shakiness of handwriting. Such results may indicate that the transfer is to some extent dependent on the level of handwriting automation: better control of the pen in one alphabet is beneficial in the other alphabet as well. These results can also be interpreted as indicators of the transfer of fine motor control skills from Cyrillic to Latin.

The results also show that the majority of the features are correlated in term of rank across the two alphabets, meaning that a child presenting a low value for a feature (compared to other children) in one alphabet will also present a low value for this same feature in the other alphabet. In this sense, we can also see that children are able to transfer their motor skills from one alphabet to the other: a specific pattern of pressure used to write one shape in an alphabet will be used to write a similar shape in another alphabet.

To translate these scientific findings into societal conclusions, our data do not seem to reveal a major risk of increasing handwriting difficulties due to the change in alphabet. Instead of a potential negative transfer of skills in which the automation acquired in one alphabet would interfere with the skills required for a second, it seems that the skills acquired in one are beneficial in the other. In other words, children seem to be able to transfer their handwriting abilities from one script to another. In that sense, even though adaptation will still be needed, learning won't be restarted from scratch when the change of alphabet will occur. Thus, the impact of simultaneous digraphia on the quality of handwriting seems to be less of an issue in comparison to simultaneous bigraphism as in the case of Lebanon for French and Arabic languages. However, it might be due to having Latin as the direct transliteration of the Cyrillic script in contrast to multidimensional differences between Arabic and Latin scripts [102]. It is important to note that these results were extracted from a copying task. It would be dangerous to generalize the results found in this study to other writing tasks such as spontaneous writing or dictation where other skills are involved. In addition, the transfer is slower for some features than for others, indicating the need for a progressive transition. Moreover, our study focused only on writing skills without taking into consideration the possible other risks raised by such a radical change in the culture of a country.

As our results suggest, the features used to analyze handwriting seem to be transferable from one alphabet to another, i.e., they can be used to diagnose handwriting difficulties in the Cyrillic alphabet in the same way they are used to diagnose handwriting problems in the Latin alphabet. Therefore, a qualitatively different model does not appear to be necessary to analyze handwriting in the two different alphabets since the features seem to be independent of the written content, paving the way for the creation of a cross-lingual model for the detection of handwriting difficulties.

Future work will need to be conducted in this direction: it will be necessary to collect a new database in which traces in the two alphabets are produced by children with equal handwriting exposure to Cyrillic and Latin. Under such conditions, we would hope to find that the majority of the features are similar between the two alphabets. For the remaining features, for which significant differences still appear between the two alphabets, we could learn the function mapping the differences between the two distributions. With this function, the model we use to diagnose handwriting difficulties could be transferred from Latin to Cyrillic, i.e., it could then be used to detect handwriting problems in the Cyrillic alphabet, with the same rationale outlined in Chapter 4.

6 Transferring Handwriting Analysis From Paper to Tablets

Introduction

In the three previous chapters, the handwriting samples were acquired using a Wacom tablet on top of which a sheet of paper was attached. As a consequence, a test designed to assess handwriting difficulties trained with handwriting samples of this type will be restricted to a small category of high-end, expensive and computer dependent digital tablets that only few institutions and private users can afford. In this Chapter, we aim to assess the feasibility of conducting a handwriting test directly on the surface of a digital tablet (and no longer on paper), since a test available on a regular digital tablet (e.g. on an iPad) would have a considerably broader use and higher impact on the society.

Only a few studies studies have been so far conducted to access the impact that the change of handwriting support can have on children handwriting. The study conducted by Alamargot et al. [93] is among the few studies interested in this subject. The authors asked 14 second grade and 14 ninth grade students to write the alphabet as well as their own name on tablet and on paper. Results show that students (no matter their grades) tend to produce a less legible handwriting while writing on a tablet support. In addition, the written characters were found significantly larger and the handwriting speed was found significantly lower when the students were using the tablet support. Concerning the pressure applied on the tablet, the results were found different depending on the student's grade. The ninth graders exhibited a higher pressure while using the tablet compared to the paper support while the opposite was found true for the second graders. Another interesting study comes from Gerth et al [103] who compared the handwriting performance of 25 adults on a tablet and on paper support. Their results reveal differences between the two medias that were partially task dependent. In addition, the authors also showed that adults were able to adapt their handwriting execution quickly to the smoother surface of the tablet, by modulating their pen pressure and increasing their writing size. In [104], the authors recruited 25 preschoolers, 27 second graders and 25 adults and compared their handwriting on the two surfaces with multiple tasks. The tasks were designed to assess different aspects of handwriting, namely grapho-motor abilities, visuo-motor abilities and handwriting performance. Once again, their results show significant differences between writing on a tablet compared to the paper. As shown by Gerth et al. [104], these differences were also found partially task dependent.

In this chapter, we propose to analyze the impact of the change of support from a multidimensional point of view (in term of pressure, speed, static and tilt aspects) thanks to our low level features describing handwriting on several aspects (static, kinematic, pressure and tilt), as described in Chapter 4. In particular, beside the absolute difference coming with the change of support, we are interested in seeing how the specificity of handwriting is affected by the change of support, and whether the model for the classification of handwriting difficulties remains valid with the change of handwriting support.

This work correspond to the following submitted publication:
T. Asselborn, and P. Dillenboug, "Do tablets provide a reliable measure of writing skills ?", *Computer Standards and Interfaces*, **2020**.

Method

Data Collection

In order to compare the quality of handwriting on tablet and on paper, we conducted a within subject study with two groups of children. Twenty-seven 5th grade children (10 males, 17 females, age: M = 11.29, SD = 0.53, 1 left-handed) and ten 1st grade children (6 males, 4 females, age: M = 7.48, SD = 0.68, 2 left-handed) from a primary school in France participated to this experiment. Each of them was asked to copy the 36 glyphs (the 26 letters of the alphabet and the 10 digits) displayed (in a random order) on the screen of a tablet computer (as can be seen in Figure 6.1). Each child was asked to write cursively (as taught in France) all these glyphs **both** on the tactile surface of a tablet computer and on paper (fixed on the digital tablet as explained in Section 6.2). The fact that they started to write in the tablet condition or in the paper condition was assigned randomly.

A software application has been developed specifically for this experiment[1] and was used on a Wacom Cintiq Pro graphic tablet. The application allows to save the child's profile (age, gender, laterality and grade) and to record various handwriting information. At a frequency of 200Hz, the time stamp, pen's coordinate (x and y), pressure and pen tilt were recorded. The Wacom system logged the data measuring the pen tilt with two different angles, which are referred in this paper as the Tilt-x and Tilt-y angles (see Figure C.4). Both angles are measured in the range between -40 to 40 degrees. The

[1] https://github.com/asselbor/alphabet.git

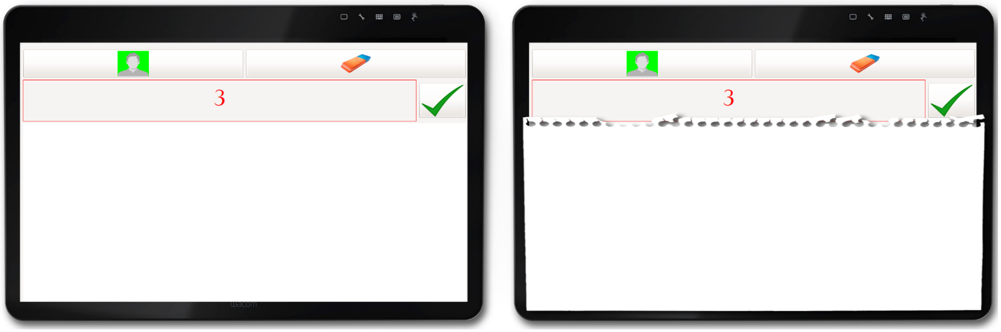

Figure 6.1: Left: Application running on the Wacom Cintiq Pro tablet computer used to acquired data (tablet condition). Right: same device with a pre-cut A4 paper placed on top of the tablet (paper condition). In this example, the child needs to write cursively the digit 3.

Tilt-X reflects the inclination of the pen in the direction of the written line, and the Tilt-Y reflects the inclination of the pen below the writing line. Concerning the pressure, the Wacom system logged the data measuring the pressure between the tablet and the pen nib. The pressure is recorded on 8192 levels of pressure converted in this paper in the range between 0 (no pressure) and 1 (maximum pressure that can be recorded by the system).

As illustrated in Figure 6.1, when a child was asked to write on the paper support, a pre-cut piece of paper was fixed in the touch pad. The top part of the screen was left uncovered by the paper in order to display the glyph to be written and the save button. By doing so, data could be recorded in the same way of the tablet condition. To allow the use of the same Wacom pen on paper, we designed a new nib allowing ink deposit on the paper, while also acting as a digital pen in the same way as before.

In the work presented in Chapter 4, 63 handwriting features were defined and used to train a random forest classifier to detect handwriting difficulties. In this work, as done in the work presented in the previous Chapter, we only used the features that were found to be the most important in the aforementioned random forest model according to the Gini importance metric (averaged with a k-fold cross validation, $k = 25$). To maintain a good balance and compare the different groups of features, we selected the most important features from each of the following four groups that we distinguished: static, pressure, kinematic, and tilt. In the following paragraphs, we briefly provide their respective definitions. A detailed description of these features can be found in Appendix C. Please notice that every possible mismatch between the features importance presented in Chapter 4 and this one are due to the retraining of the model with new data that happened in the meantime.

Results

Absolute Feature Difference

For each feature under consideration, a Wilcoxon test was conducted to detect if its distribution in the paper condition is similar to its distribution in the tablet condition. This statistical test was used since not all the distributions where found to follow a normal distribution. The results can be found in Table 6.1, column *Absolute*. The means and standard deviations for all the distributions can also be found in this Table.

Static Features

The change of handwriting support does not seem to bring any difference concerning the level of shakiness. Indeed, no statistically significant difference was found for the two features describing shakiness in handwriting (namely the *Bandwidth and Median of Power Spectral of Tremor Frequencies* (#8, #9)) in function of the support used. However, the *handwriting density* (#4) was found to be higher when children where writing on paper compared to when writing on tablet. This result appears to be in line with the work conducted by Alamargot et al. [93] who also noticed an amplification in the handwriting's movements, translating in an increase of the letter size (and thus lower density) when children were writing on tablet. We can see an illustration of such finding in Figure 6.2 where the difference in size of the letter "h" written on paper and on tablet appears evident.

Kinematic Features

All the features describing handwriting on a kinematic level were found to be statistically different when children were writing on paper or tablet. This is most probably a direct consequence of the difference in terms of friction between the two handwriting supports. The tablet being more slippery than the paper, it appears normal that the *Mean Velocity* (#13) and *Max Velocity* (#14) were found higher when children were writing directly on the tablet surface (as can be seen in Figure 6.2). Other studies found similar results [104, 93]. The *Nb of Peaks of Speed Change per secs* (#17) and the *Median of Power Spectral of Speed Change Frequencies* (#19) are both measuring the level of automation and were found to be very strongly positively correlated with handwriting proficiency in our previous study reported in Chapter 4. Results concerning these two features show that children were more comfortable when writing on paper. For example, we can see in Figure 6.2 that children writing on paper are able to easily and quickly change their handwriting speed (from low to high speed) while the low friction coefficient of the tablet makes this change more difficult in the tablet condition.

Tilt Features

The features describing the tilt aspects of handwriting were also found to be affected by the change of handwriting support. The *Std. of Tilt-X* (#40), *Std. of Speed of Tilt-X Change* (#46) and the *Median of Power Spectral of Speed of Tilt-Y Frequencies* (#60) are three features found to be describing the level of handwriting automation (see Chapter 4 for additional information). Results once again show that children appear more comfortable when writing on paper compared to when writing directly on the more slippery surface of the tablet.

Pressure Features

The last category of features describes handwriting on pressure aspect. Two features, the *Mean Speed of Pressure Change* (#27) and the *Max Speed of Pressure Change* (#28), were found to be statistically different when children were changing their handwriting support (with higher pressure when writing on tablet as illustrated in Figure 6.2). These two features, correlated with the handwriting automation show again a more difficult control of the pen for children writing on the tablet. The last two features, the *Median of Power Spectral of Speed of Pressure Frequencies* (#33) and the *Nb of Peaks of Speed of Pressure Change per secs* (#31) were not found to be statistically different depending on the type of handwriting support used.

Relative Feature Difference

In this Section, we want to assess whether the change of handwriting support brought handwriting difference relative to other children. Following the same rationale adopted in the last Chapter to compare different alphabets, we no longer focus on raw data differences but rather on the relative differences between individuals. In other words, we are interested in investigating if children who present a low value for one feature when writing on paper **relative to the other children** (of the study) still present a low value for this same feature when writing directly on the tablet **relative to the other children** (of the study). If this condition was found to be valid, it would mean that children handwriting specificity is kept with the change of handwriting support and therefore also the validity of the model proposed to diagnose handwriting difficulties.

In Table 6.1, column *Relative*, the Spearman rank correlation is reported for all features. The Spearman Rank correlation coefficient is based on the ranked values for each variable rather than the raw data. The correlation between two variables (e.g., feature X in the paper condition and feature X in the tablet condition) will be high when the observations have similar ranks (i.e., the similar ordering of the labels of the observations within the feature X: 1st, 2nd, 3rd and so forth) for the two handwriting support. In other words, if a child presents a low (high) value for one feature compared to other children when

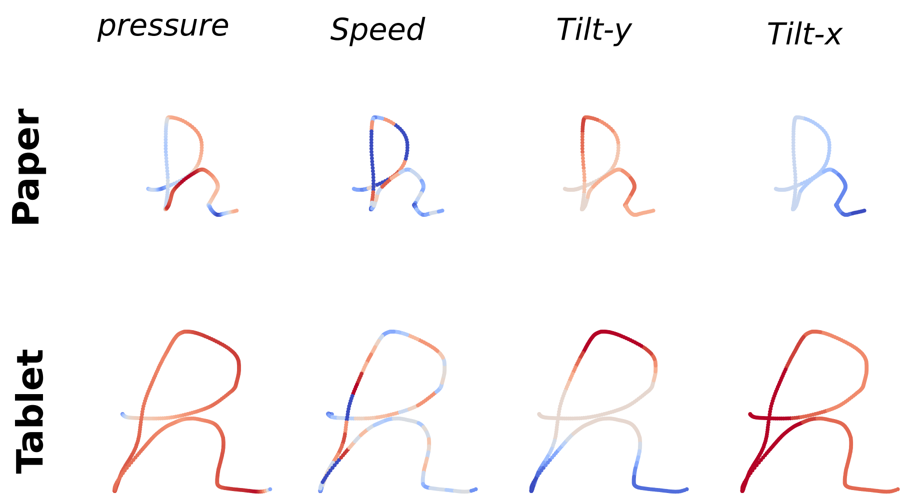

Figure 6.2: Examples of a glyph written on paper and on tablet.
For the pressure column, the color denotes to the amount of pressure applied on the support when the point was recorded: blue means low pressure while red means high pressure.
For the speed column, the color denotes to the handwriting speed when the point was recorded: blue means low speed while red means high speed.
For the Tilt-y column, the color denotes to the value of the Tilt-y angle (see Figure C.4): blue means a small Tilt-y angle while red means big angle.
For the Tilt-x column, the color denotes to the value of the Tilt-x angle (see Figure C.4): blue means a small Tilt-x angle while red means big angle.

writing on paper and still presents a low (high) value of this same feature when writing on tablet, the Spearman rank correlation will be high.

In general, features seem to be highly correlated between the two handwriting supports. Indeed, only one feature (namely the *Median of Power Spectral of Speed of Pressure Frequencies* (#33)) wasn't found to be strongly correlated. This is a very interesting result since it shows that children are able to transfer their handwriting specificity from one handwriting support to another. The implication of such a result will be further discussed in the following Section.

Conclusion and Discussion

In the present chapter, we investigated whether the writing surface (paper condition versus tablet condition) affects the handwriting characteristics. A comparison between the two types of surface for the specific features of handwriting found to be relevant in Chapter 4 was done to investigate the effect of the surface on different handwriting aspects.

Our results demonstrate important absolute differences between the two conditions. Significant differences were in fact found for several of the features considered in this study and describing handwriting on the static, kinematic, pressure and tilt aspects. In line with other studies [93, 104], we showed that the different friction coefficient of the two supports has a clear impact on the handwriting production. For instance, the lower friction coefficient of the tablet results in less force being applied in the opposite direction of the handwriting movement altering the handwriting kinematics (*kinematic aspect altered*). We also hypothesized that the difference in friction coefficient leads children to compensate for their proprio-kinesthetic feedback by putting more pressure (as can be seen in the column pressure of Figure 6.2) on their tablet in order to stabilize the movement (*pressure aspect altered*). This decrease of proprio-kinesthetc feedback can also be compensated by an amplification of handwriting movement (as can be seen in Figure 6.2) (*static aspect altered*) or by an adjustment of the pen grasp (leading to change the pen's angle in order to reduce the effect of the lesser tablet friction coefficient of the tablet, as illustrated in the columns *Tilt-x*, *Tilt-y* of Figure 6.2) (*tilt aspects altered*).

Beside the absolute differences between features, results also show that almost all the features are strongly correlated in ranking across the two conditions (tablet and paper). A child with a low value for a feature (compared to the other children) when writing on paper will also have a low value compared to the other children for this same feature when writing on a tablet. In this sense, we can see that children are able to preserve their handwriting specificity when changing the handwriting support: a child having a shaky handwriting on paper compared to other children, will also have a shaky handwriting on tablet. This result confirms the fact that children are indeed affected by the handwriting

	Paper	Tablet	Absolute		Relative	
	mean (std)	mean (std)	W	p value	corr	p value
Bandwidth Tremolo **#8 *1e-1**	6.44 (0.08)	6.43 (0.10)	54	0.26	0.57	<0.001
Median Tremolo **#9 *1e2**	1.18 (0.06)	1.18 (0.07)	301	0.44	0.28	0.09
Handwriting Density **#4 *1e-2**	8.10 (3.99)	5.53 (4.28)	65	<0.001	0.65	<0.001
Mean Velocity **#13 *1e1**	0.88 (0.37)	1.26 (0.61)	18	<0.001	0.73	<0.001
Max Velocity **#14 *1e1**	3.67 (1.57)	4.91 (2.32)	36	<0.001	0.75	<0.001
Nb Peaks Speed Change per secs **#17 *1e1**	3.29 (0.25)	3.14 (0.29)	100	<0.001	0.74	<0.001
Median Speed **#19 *1e2**	1.16 (0.12)	1.10 (0.14)	121	<0.001	0.71	<0.001
Mean Speed of Pressure Change **#27 *1e4**	1.52 (0.59)	2.64 (1.08)	26	<0.001	0.48	<0.01
Max Speed of Pressure Change **#28 *1e3**	2.71 (0.54)	3.20 (0.83)	126	<0.001	0.26	0.12
Nb Peaks Speed Pressure Change per secs **#31 *1e1**	2.38 (1.72)	3.00 (4.35)	339	0.85	0.41	<0.05
Median Pressure **#33 *1e2**	1.27 (0.05)	1.28 (0.07)	266	0.20	0.17	0.31
Std. of Tilt-X **#40 *1e-1**	3.47 (1.07)	3.09 (1.11)	233	0.07	0.34	<0.05
Std. of Speed of Tilt-X Change **#46 *1e3**	8.01 (2.75)	8.50 (2.87)	248	0.12	0.57	<0.001
Median Tilt-Y **#60 *1e2**	1.24 (0.08)	1.12 (0.07)	174	<0.01	0.40	<0.05

Table 6.1: Feature difference between the two handwriting condition (Paper and Tablet). Column **Absolute**: For each feature, a Wilcoxon test was conducted to see if there was a statistical difference in the means of the two distributions. The red color denotes the level of statistical significance of each test. Column **Relative**: For each feature, a Spearman's rank correlation was calculated in order to examine the similarities between the two distributions (columns **Relative**). The green color denotes the level of statistical significance between the two handwriting support. The features means and standard deviations for the two conditions are also reported. For visibility purposes, some features values have been multiplied by the factor next to the feature name.

support, but that all children are affected in the same way (i.e. there is a consistent shift in the distribution of features). In that sense, although a retraining of the model proposed in Chapter 4 is needed (since an absolute difference between the two conditions exists), it appears possible to detect handwriting problems on a tablet support in the same way for the paper support. The aspects of handwriting allowing us to assess its quality in fact stay the same regardless whether children are writing on paper or tablet.

A model assessing handwriting difficulties directly running on a tablet would have a considerably greater impact on the society, since it could be easily deployed in consumer tablets like iPads available everywhere in the world. Based on the proof of concept obtained in this chapter, such a model is presented in the next Chapter of this thesis.

7 Extracting the Handwriting Profile of the Child

Introduction

Based on the work introduced in the previous chapters of this Part, we present, in this Chapter, a modernized method allowing to break away from the BHK legacy.

Our method leans on the features assessing handwriting on multidimensional data introduced in Chapter 4. Far from the current rule-based tests, our method uses data-driven strategies to model the complexity of our features with statistical findings coming from factual data. In addition, our method allows to assess handwriting on different granularity levels from an individual feature level, a categorical level (static, pressure, kinematic and tilt), to a global level.
Given its integration in consumer tablets, our method could be of great use for therapists but also school teachers and even parents since it would allow a cheap, frequent, deep, multidimensional, fast and free of human bias analysis of handwriting difficulties bringing new possibilities to adapt the remediation based on the specific handwriting problems detected and the monitoring of the child's progresses.

This chapter corresponds to the following publication:
T. Asselborn, M. Chapatte, and P. Dillenboug, "Extending the Spectrum of Dysgraphia: A Data Driven Strategy to Estimate Handwriting Quality", *Scientific Reports*, **2020**.

Method

Participants

The present study was conducted in accordance with the Helsinki Declaration. It was approved by the EPFL Ethics Committee (Agreement 043-2019) and conducted with the understanding and written consent of each child's parents and the oral consent of each child. A consent to publish was obtained from each child's parent/guardian (legally authorised representative of each participant).
A total of 390 typically developing (TD) children were recruited from five primary schools in Haute-Savoie, France. Children ranging in age from five to twelve years of age were recruited from 30 classes. None of these children presented known learning disabilities or neuro-motor disorders.
A total of 58 children presenting severe handwriting difficulties (dysgraphia) were also recruited from therapy centers. Patients of seven grapho-therapists (from the Graphydis Association) located in France were also involved in this study. All of these children were diagnosed dysgraphic (severe handwriting difficulties) according to their BHK scores, previously conducted on paper by therapists.

It is important to note that, unlike to the children recruited from the therapy centers, none of the children recruited in the schools were assessed for dysgraphia with the BHK test, which means that some of these children might present severe handwriting difficulties, as well.

Data collection

The 448 children (school children and children with dysgraphia) involved in this study were asked to write the first five sentences used in the BHK test on an iPad-based application specifically designed for this purpose (Dynamico application v1.0). This application used the Apple pencil.

The data were collected using the Dynamico software (CHILI laboratory EPFL), which recorded the x and y coordinates as well as the pressure, azimuth and altitude angle of the pen at a frequency of 240 Hz. In addition, the gender, age and laterality of the writer were saved.

Based on previous work [3], we extracted, for each child involved in this study, a set of features describing handwriting based on different aspects and sorted into four main categories (static, kinematic, pressure and tilt). These features were found to successfully predict severe handwriting difficulties in the study published by Asselborn et al. [3]. All these features are described in Appendix C.

Results

Every child involved in this study was asked to write the five first sentences used in the BHK test on a digital tablet (iPad). With the data extracted (x and y coordinates and well as the pressure, tilt and azimuth angle of the pen), we then computed, for each child, a set of 63 features describing their handwriting.

From the Feature Value to the Feature Score

The features describing handwriting and the interactions of these features with grade and age are very complex.

When computing the features from a handwriting sample, we obtain a range of values that cannot be interpreted directly since they measure handwriting characteristics that vary as functions of a child's age and gender. The handwriting of a five year-old child will be obviously very different than the handwriting of a 12 year-old one. Handwriting also differs by sex [105, 106]. Although current handwriting tests do assess handwriting by grade, we propose to extend this concept by treating age as a continuous variable since it is evident that handwriting evolves at a more rapid pace than annually. To do so, we used a third-degree polynomial function to interpolate each feature for both genders **separately**. The means (f_{mean}) and standard deviations (f_{std}) of the features as a function of age are provided in Figure 7.1 (for a given feature). Hence, a score between 0 and 1 can be computed for each feature according to the deviation from the average (Z-score) via the following equations. This score takes into account the age and gender of the child.

$$Z - score = \frac{|feature - f_{mean}(age)|}{f_{std}(age)}$$

$$score = e^{-\alpha * Z - score^{\beta}}$$

Where α and β are two positive parameters.

In other words, the score indicates how a feature value differs from the mean for children of the same age and gender, making it easily interpretable for therapists. It should be noted that all the children who participated in this study were recruited from French schools, meaning that they all started to learn handwriting in the same grade. In some European countries, children may start to learn handwriting at a different age (e.g., at the age of 7 instead of 6). Hence, at the same age, it is possible for children to have been exposed to different amounts of training (i.e., a 6 year-old child can have a better handwriting than a 7 year-old child). For this reason, we believe that, in addition to gender, different regressions should be used for different countries or regions.

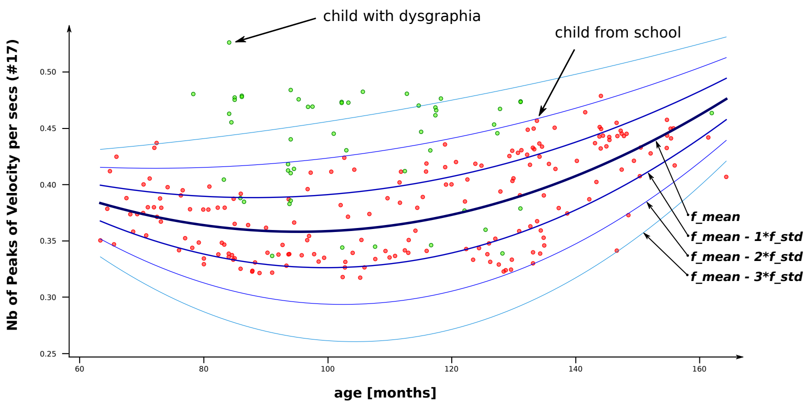

Figure 7.1: The value of the feature: *Number of Peaks in Velocity per second* (#17) as a function of age is plotted for all individuals in our database. The red points represent children recruited in schools, while the green points represent children with dysgraphia. The interpolated function representing the mean (f_{mean}) is plotted as well as the interpolated function representing the standard deviation (f_{std}). Notice that the points for the children with dysgraphia seem to be located further from the mean compared to the points for their peers.

Projecting Features' scores with the PCA

As shown in Chapter 4, the features used in this study have advanced discriminatory power when predicting severe handwriting difficulties (dysgraphia). For example, the kinematic and pressure features were found to be far more important than static or tilt features when predicting severe handwriting difficulties. Hence, a numerical scale for estimating handwriting quality based on these features should take into account this difference in terms of predictive importance. However, the feature importance calculations used by Asselborn et al. [3] are linked implicitly with the design of the BHK test since a feature's importance is based on its power to detect the severe handwriting difficulties (dysgraphia) diagnosed with the BHK test.

In order to design a scale totally independent of the BHK test (and more generally independent of any existing tests), we used the PCA algorithm to determine the importance of the features to predict handwriting quality in a data-driven way. Indeed, as the PCA algorithm projects features into a lower-dimensional space by maximizing the variance within the database, the transformation associated with it will result in the discovery of the combination of features explaining most of the differences in handwriting in our database from those **independently** of any current handwriting tests.

Figure 7.2 shows the amount of variance explained by the three first axes of the PCA algorithm as well as the composition in terms of absolute feature importance sorted into

the four categories for the three first axes. For reasons of clarity, we kept only the six most important features for each category (24 in total) to plot the figure.
We can see that kinematic features seems to be the most important in terms of explaining the variability in our database. For instance, the first axis, explaining 39.4% of the variability in our database, is composed of 52.5% kinematic features. The second axis, representing 25.9% of the variability in our database, is composed of a majority (60.2%) of features belonging to the pressure category. Finally, the third axis, representing 22.6% of the variability in our database seems to be mainly composed of kinematic and pressure features. Therefore, the tilt and static features do not appear to be meaningful in explaining handwriting differences between children, at least as opposed to the kinematic and pressure features. The implications of this finding will be explored further in the *Discussion* section.

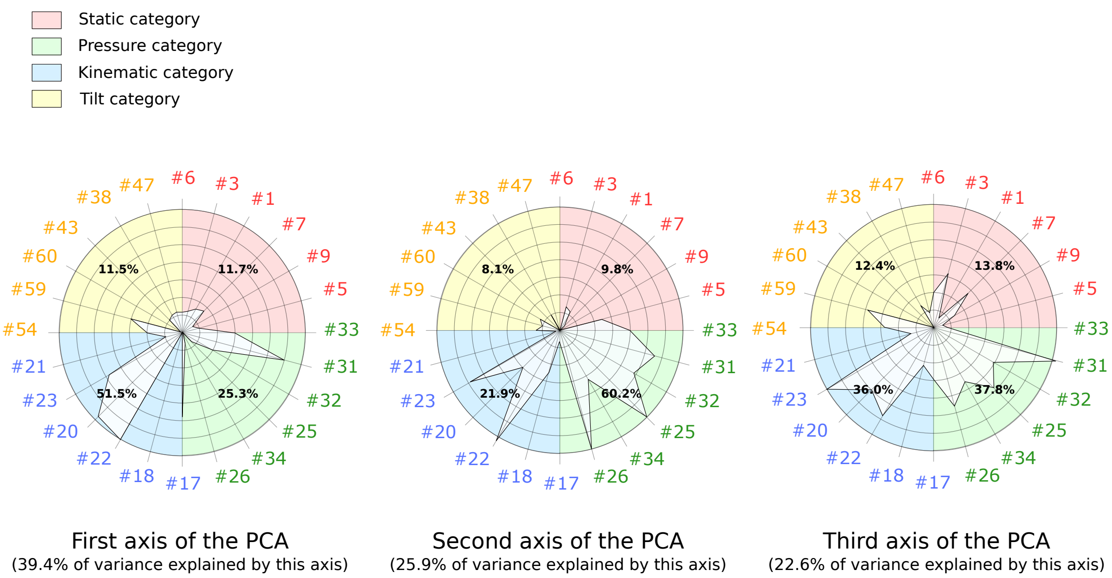

Figure 7.2: The amount of variability explained by the three first axes of the PCA. The absolute feature importance for each of the three first axes is also represented (in the radar chart) and was sorted according to the four categories. For reasons of clarity, the graphs were plotted with only the six most important features for each category (static, kinematic, tilt and pressure). The features are represented by numbers. The correspondence between the numbers and the feature names and descriptions can be found in the *Method* section.

Unsupervised Clustering

We used a K-means algorithm to cluster the individuals from our database (represented by their features projected with the PCA algorithm) into different groups. For reason of interpretability, we limited the number of cluster to two while hypothesizing that the algorithm would find clusters based on handwriting quality. In Figure 7.3, we can see the two clusters (represented by different colors) of individuals projected into the three first axes of the PCA under different projection angles. In order to assess the quality of our clustering, we then checked to see if the dysgraphic children from our database were put into the same cluster. The results showed that around 92% of the children with dysgraphia were clustered together (in the red cluster), while only five children with dysgraphia wound up in the wrong group (blue group). With regards to children without known handwriting difficulties, 86% of them were clustered together (in the blue cluster), while handwriting difficulties were detected in 14% of them (the ones in the red cluster). It is interesting to note that this percentage appears to be comparable to the statistics for French-speaking countries, where approximately 10% of the children in the general population have severe handwriting difficulties (dysgraphia) [60].

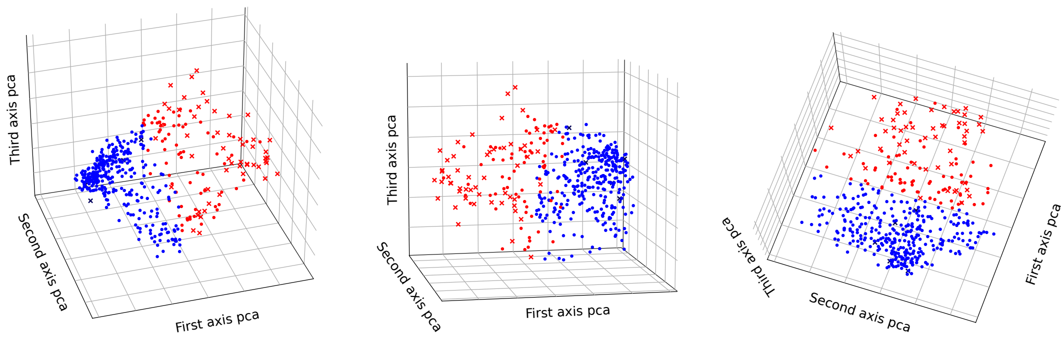

Figure 7.3: K-Means Algorithm applied to our three-dimensional dataset (defined by the features projected on the three first axes of the PCA), with two clusters under three different projections. The crosses represent children with severe handwriting difficulties (as determined by their BHK scores), whereas the points represent children recruited in schools. The colors represent the two different clusters.

The same process can also be repeated (PCA + K-Means) while taking into account only features from one single category. In such a manner, we obtained a model clustering children according to specific handwriting skills related to a category (static, tilt, pressure and kinematic skills only). We then related these clustering models to the two groups of children present in our database (children with dysgraphia and school children) to see how the models managed to link difficulties related to a specific skill (e.g., kinematic skill) with overall handwriting quality. The results are found in Table 7.1. When analyzing the results in this table, it is important to realize that the handwriting skills assessed by our categories can be independent. For example, it is possible for children with kinematic deficiencies to have no difficulties with pressure. Hence, it is possible to observe a lower sensitivity value for the "categories" than the "total" skills. However, it is still interesting to analyze the sensitivity of the categories since doing so provides insight into the percentage of children with dysgraphia presenting issues in a specific skill. For instance, 78% of children with dysgraphia present kinematic difficulties, while 58% of them display issues with pressure skills. The implications of this finding will be explored in the *Discussion* section.

Category	Sensitivity	Specificity
Total	0.91	0.90
Kinematic	0.78	0.82
Pressure	0.58	0.72
Tilt	0.45	0.68
Static	0.36	0.76

Table 7.1: Results of different K-Means clustering models when sub-features belonging to a given category are used (*Total, Kinematic, Pressure, Tilt and Static*) compared to the labeling in our database (children with dysgraphia vs school children).

From a Binary to a Numerical Scale

One of the way to escape a binary determination (with or without severe handwriting difficulties) with the K-Means clustering algorithm is to exploit the parameters found by the algorithm in order to cluster the data points. The K-means algorithm identifies k cluster centroids (k = 2 in our case) and then allocates each data point to its nearest centroid. In such a manner, the centroid of the cluster grouping typically developing children should represent the handwriting of a *typical* child without handwriting difficulties. The distance between one individual and this centroid location would then be a good indicator of their handwriting quality. In Figure 7.4, we can see the scores for both school children and children with dysgraphia. We can see that the scores of school children (0.69 ± 0.13) are statistically higher than the scores of children with dysgraphia (0.41 ± 0.13). A Wilcoxon-Mann-Whitney statistical test ($U = 1917.0, p = 9.67e - 25$) was used to test statistical significance since the data were not found to be normally distributed. On the basis of the distribution of scores for the school children, we were then able to create threshold values dividing handwriting difficulties into five categories in such a way that: 2% of school children were below the very severe threshold (to find the *vs* threshold), 8.6% of school children were below the severe threshold (to find the *s* threshold) (which is the current dysgraphia threshold), 15% of children were below the moderate threshold (to find the *m* threshold) and 25% of children were below the light difficulty threshold (to find the *l* threshold). It should be noted that these thresholds were set as an example, meaning that the number of categories as well as the values of the thresholds defining them can be defined freely.

Conclusion and Discussion

In this chapter, we propose various scales to evaluate handwriting difficulties in a modernized way. These scales exploit the multi-modal data obtained with digital tablets and appear to be disruptive for several reasons.

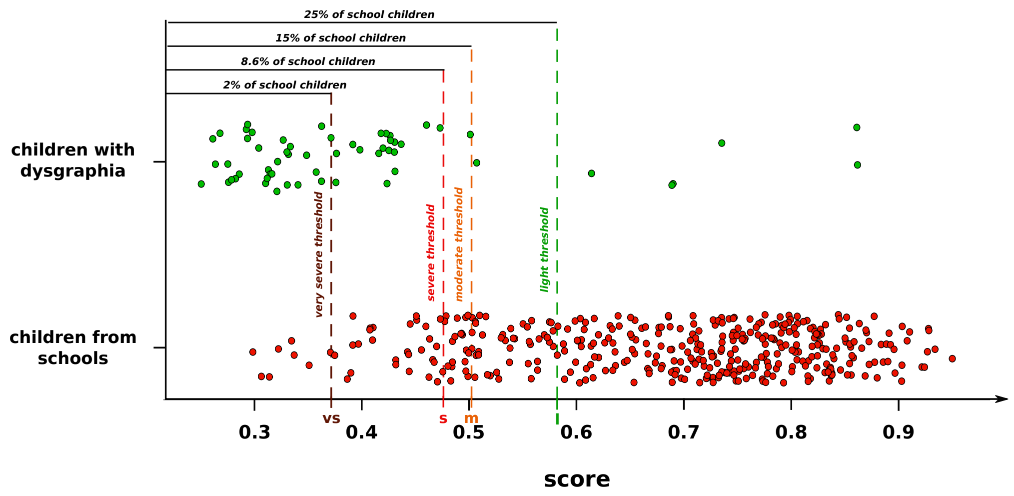

Figure 7.4: Scores computed for all the children in our database. The red points represent children recruited from schools, while the green points represent children with dysgraphia. Four threshold values (*very severe*, *severe*, *moderate*, *light*) were used to divide handwriting difficulties into five categories with 2% of school children below the *very severe* threshold (this allowed us to compute the *vs* value), 8.6% of school children below the *severe* threshold (this allowed us to compute the *s* value, which is the current dysgraphia threshold), 15% of children below the *moderate* threshold (this allowed us to compute the *m* value) and 25% of children below the *light* threshold (this allowed us to compute the *l* value). It is important to note that these thresholds have been set as examples, meaning that any other values can be used. The score is proportional to the handwriting quality

The current methods used in assessing handwriting difficulties, such as the gold standard test in French-speaking countries (BHK) [60], are all rule-based systems built on rules designed by a group of experts in handwriting difficulties. Unlike data-driven systems, the consequences of rule-based systems are determined by humans and their subjective knowledge of such complex problems. In addition, a rule-based system will not change or update on its own, which is a limitation, as the educational system and handwriting, in particular, are evolving constantly. In addition, data-driven models are built from the data, meaning that they do not reduce a given problem to a set of "limited" rules extracted from human subjectivity but rather try to model the problem's complexity with statistical findings coming from factual data.

That being said, it would certainly be a great mistake to neglect years of research and the experience of hundreds of researchers and clinicians interested in this field and build an entirely new handwriting scale. Combining both the advantages of rule-based and data-driven models, our method leans on the crucial expertise of therapists and researchers and exploits machine learning techniques to glean complex and objective knowledge about handwriting difficulties. Indeed, the features used by our model were designed with the help of therapists and therefore aim to capture the current knowledge

and rules adept to explain handwriting difficulties. We then combine these rules with a data-driven approach (unsupervised machine learning) to find the handwriting aspects (and their interactions) best explaining the differences in handwriting between children. In such a manner, our data-driven strategy based on multi-modal features provides an objective way to evaluate handwriting difficulties but still captures the current knowledge of experts in this field.
The results provided by our method were then compared with the results of the BHK test conducted on paper. A clear correlation was found since more than 91% of children with dysgraphia (according to the BHK test [60]) were identified with our method, while 87% of the children recruited from schools were assessed as having no handwriting problems. It is interesting to note that 13% of the children recruited in schools were assessed as presenting handwriting difficulties according to our method, corresponding to the 10% of the population in French-speaking countries considered with dysgraphia [60].
Secondly, our method allows handwriting quality to be measured on a numeric scale, allowing a new categorization of handwriting difficulties. Thanks to our numerical scale, we were able to define new categories according to the level of handwriting difficulties, namely *very severe*, *severe* (corresponding to the current dysgraphia level in French-speaking countries), *moderate* and *light*.

Thirdly, by repeating the process achieved to obtain global handwriting scores but with just the features from a single category (static, pressure, kinematic and tilt), we were able to design new scales evaluating handwriting for these categories. As can be seen in Table 7.1, if the kinematic (78% of sensitivity) and, to some extent, the pressure (58% of sensitivity) scales appears to be correlated with the overall handwriting difficulties, then neither the static or tilt scale will provide a clear distinction between children with and without handwriting difficulties. This finding shows that the current handwriting tests, which only take the static aspect of handwriting in consideration, are clearly limited since the other aspects of handwriting (kinematic and pressure), which appear to be more important (as can be seen in Figure 7.2 and Table 7.1), are not used. This result also shows that when the handwriting skills (e.g., the pressure or kinematic skills) are considered separately, the results are not as good as when they are considered altogether. For this reason, we believe that a scale assessing handwriting as a combination of skills, such as the one presented in this study, presents added value to therapists and school teachers.
Furthermore, the combination of feature, category and global scores that can be extracted thanks to our method could also be of great use for therapists since handwriting can be assessed according to different aspects and at different granularities (as can be seen in Figure 7.5). This flexibility is particularly interesting since the handwriting skills measured by our categories (e.g., kinematic skills, pressure skills, and so on) appear to be quite independent. Hence, measuring any of these skills separately fails to provide a solid measure of a child's handwriting proficiency (as shown in Table 7.1). In other words, a child with severe handwriting difficulties does not necessarily have difficulties with

all the skills assessed by our categories (as illustrated in Figure 7.5). Likewise, a child without handwriting difficulties is not necessarily proficient in all the skills assessed by our categories. Thus, our method allows, thanks to the extraction of a child's handwriting profile, a deeper analysis of handwriting, making it useful for therapists and school teachers in terms of, for example, making informed decisions concerning the type of remediation that should be applied in the case of their patient/student with regards to their specific handwriting problems.

Another great advantage of the scales described in this paper is that there are based on features describing handwriting on a very low level, measuring almost physiological aspect of handwriting. These features are thus independent of the writing content meaning that, contrary to the majority of currently existing tests, the text (and to some extent the alphabet as shown in Chapter 5) used does not matter. In that sense, we can avoid situation of over-training biasing tests results (children learn the text and not handwriting). As we can reach a good measure of our features in only 30 seconds of data (as can be seen in Figure 4.1), compared to 15 minutes for the BHK test for example [60], it is now possible to assess handwriting online and at a very high frequency (several times a week for instance) and to therefore be able to monitor the progress of the child in a totally new way. In some extend, it would be possible to design an application running on a consumer tablet (e.g. Ipad) allowing a cheap, frequent, deep, multidimensional and fast analysis of handwriting, suggesting remediation activities according to the specific handwriting problems detected and allowing to monitor the child's progress on a weekly basis.

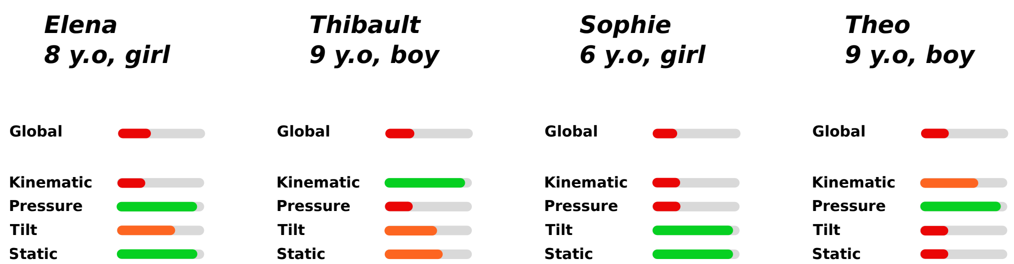

Figure 7.5: Different handwriting profile of children with severe handwriting difficulties. We can see that a same condition (dysgraphia) can be expressed in very different ways with children presenting difficulties in different and independent skills (kinematic, Pressure, Tilt and Static).

Finally, it is important to understand that the scale presented in this paper assess handwriting only on the motor skills. However, many over factors should be considered when it comes to understand a human being. In particular, the emotional aspect of the child is not assessed in our method. Our method therefore does not ambition to provide a global comprehensive measure of the patient/student handwriting but rather aims to provide a complementary analysis tool for therapist or school teacher to help them better understand their patient/student on factors previously unexplored.

This chapter closes Part II dedicated to the analysis of handwriting difficulties. Based on the model presented in this chapter, we now move to the third Part of this thesis focusing on the remediation to handwriting difficulties.

Handwriting Remediation Part III

Part Three of this thesis focuses on the remediation of handwriting difficulties. In particular, we show how robotics can be a great addition to enhance handwriting remediation in some specific context. We present in this part some examples of activities designed to exploit the capabilities that robots can offer.

Some children present handwriting difficulties so severe to generate in them anxiety and avoidance at the idea of writing, which negatively reinforces their difficulties and hinders any chance of improvement. This avoidance also affects the writing opportunities that are compulsory in a training process. For these children, the usual school exercises can therefore become to big of a challenge. For these specific cases, the use of robotics can be an interesting solution. Chapters 8 show the work conducted in the context of the Cowriter project, to design a 30-weeks long interaction between a robot and a child with severe handwriting difficulties, in a clinical environment. Novel, ad-hoc activities targeting specific important handwriting problems identified in the previous Part were used. In Chapter 9, we investigate the importance of the design of robot behaviors towards social acceptance with children, something especially critical knowing the importance of the Child-Robot Interaction within the CoWriter project.

Finally, Chapter 10 shows an example of how the multi-sensory information (e.g. audio, visual and kinesthetic) provided by a robot can help children to learn important handwriting aspects (namely, visual perception and visual-motor coordination) usually learned during pre-school, to hopefully prevent handwriting difficulties to arise later in their school curriculum.

8 Clinical study

Introduction

In this Chapter, we present the study in which we tested for the first time our remediation system and activities in the context of a long term interaction (30 weeks) with one child called R. We first introduce our different remediation activities, each of them designed to remediate a specific handwriting skill, as identified in Part 2 of this thesis (e.g. targeting the pressure, speed or tilt aspect of handwriting).

In addition to testing the overall acceptability of the system in a clinical setting and within a long term interaction, one of the underlying questions that we try to answer in this chapter is whether working on handwriting skills separately can have a positive impact on handwriting as a whole. In other words, does working on the pressure, kinematic, tilt aspect separately have a positive impact on the quality of handwriting?

In order answer this question, data were collected thanks to a digital tablet that allowed us to see the child's progression though the activities. BHK tests were conducted at regular intervals during the 30 weeks that allowed us to observe a remarkable clinical progression both in terms of quality and speed. Specifically, while at the beginning of the therapy the BHKs conducted by R. couldn't be recorded due to a too low handwriting quality, R. obtained scores at the end of the 30 weeks that were not pathological anymore (above the dysgraphia threshold).

This work corresponds to the following submitted publication:

T. Gargot, T. Asselborn, I. Zammouri, J. Brunelle, P. Dillenbourg, J. Nasir, W. Johal, D. Archambault, M Chetouani, D. Cohen and S. Anzalone, "It's not the robot who learns, it's me: Treating severe dysgraphia using Child-Robot Interaction and gaming", *Science Robotics*, **2020**.

Method

Participant

R. was a 8-year-old boy when he was assessed for very severe handwriting difficulties (severe dysgraphia). At that time, R. was facing a situation of failure since he was refusing any handwriting activity at school (he was trying to break his pen during handwriting due to frustration and anger). He had to repeat his first grade due to extreme difficulties concerning handwriting acquisition. Family history shows that R's father and mother also presented dyslexia.

R's early development was marked by psychomotor agitation. He received physiotherapy at the age of 1 year. Oral language was normal. At 5 years old, he started attending a classroom with special education during mornings and received the support of an adult in a classic classroom environment during the afternoons. At 6 years old, a diagnosis of ADHD was confirmed in a specialized clinic and R. started to receive a treatment.

When R. was admitted in the department of pedo-psychiatry of *la pitié salpetrière* hospital, we conducted an in depth assessment. In particular, he got diagnosed with an Attention Deficit with Hyperactivity disorder (ADHD) and a Developmental Coordination Disorder (DCD) with severe dysgraphia and dyslexia that was impairing for schooling. At 8 years old, he was refusing any kind of handwriting activities.

Experimental Design and Metrics

One of our goal during this study was to assess longitudinally how R. would behave during successive therapeutic sessions using the system. To do so, we monitored each sessions and registered several metrics, either clinical or digital. In particular, we aimed to assess: (i) the acceptability and feasibility of the system, software and set up in a clinical setting. (ii) how handwriting can be improved (if any improvement) when training handwriting sub-skills independently during remediation activities and through multiple sessions according to the proficiency in the remediation activities and the clinical BHK quality and speed score. To assess handwriting quality, we conducted a BHK test at the beginning of every 5 sessions. Each clinical BHK was randomly and blindly scored by an expert rater.

Design of Activities

Different activities were developed for handwriting skills remediation purposes in close collaboration with therapists. Each activity was designed as a game, targeting a specific aspect of handwriting, and is meant to be adaptive, meaning that the design of the game can be changed in order to adapt its difficulty. All the activities are described below.

The Co-Writer Activity

The Co-writer activity has been described in the first chapter of this thesis (Chapter 1). During this activity, a robot writes a word (as can be seen in Figure 8.1) with a bad handwriting. The goal of the child is to help the robot improve its handwriting by showing a "good example". The robot then learns from the child's handwriting and adapt its handwriting accordingly.

The difficulty of the activity can be adapted by changing the following parameters:

- The chosen word (longer or shorter, with more or less difficult words).
- The speed at which the robot learns.

The following metrics tracking the proficiency of the child can be recorded:

- The length of the word, being a proxy of the word difficulty.
- The letters composing the words, being a proxy of the word difficulty.

The Kinematic Activity

This activity targets the mastery of the kinematic aspect of handwriting. In it, the child needs to reach the end of a path in which hidden letters are placed, as can be seen in Figure 8.1 in the case of the letter "a". By doing so, the child will learn, without noticing, to trace the letters that are hidden in the path. To motivate the child to the task, a robot head starts moving along the path as soon as the child starts: whoever arrives first at the end of the path wins. A visual feedback indicating if the child is inside or outside the path, represented as a colored halo around the position of the pen (as can be seen in Figure 8.1) was also added.

The difficulty of the activity can be adapted by changing the following parameters:

- The letter hidden in the path.
- The width of the path.
- The speed of the chasing robot.

The following metrics tracking the proficiency of the child can be recorded:

- If the child arrives before or after the robot, being a proxy of speed.

- The time required by the child to reach the end of the path, being a proxy of speed.
- The ratio between the number of points recorded outside the path and inside the path, being a proxy of precision.

The Pressure Activity

This activity targets the mastery of the pressure aspect of handwriting. In it, the child is controlling a robot's head along a maze (as can be seen in Figure 8.1) by moving the pen from left to right (between the *start* sign and the finish line) to control the x position of the robot while the y position is controlled by the amount of pressure the child applies with the pencil onto the tablet (increasing the pressure makes the robot head move downwards, decreasing it makes it move upwards). In order to avoid the obstacles within the game, the child therefore needs to learn to control the amount of pressure he is applying on the tablet.

The difficulty of the activity can be adapted by changing the following parameters:

- The width of the passages (the gap between two peaks).
- The number of peaks.

The following metrics tracking the proficiency of the child can be recorded:

- The number of collision that occurred before success.
- The time required by the child to reach the end of the maze.

The Tilt Activity

This activity targets the mastery of the tilt aspect of handwriting. The child is using the pen like a joystick to control the robot head (as can be seen in Figure 8.1) along the x and y axis. The goal of the activity is to capture the battery in order to recharge the robot while avoiding the bombs.

The difficulty of the activity can be adapted by changing the following parameters:

- The number of bombs.
- The position of the bombs

The following parameters tracking the proficiency of the child can be recorded:

- The number of collision with the bombs.
- The time spent before success.

The Rainbow Activity

The rainbow activity is designed to be used with therapists. In a first phase, the therapist writes a word, or a small text in the upper part of the screen (as can be seen in Figure 8.1). Each time the pen is lifted from the tablet, the color of the link is changed. The child then is asked to write the same word with the goal of reproducing the same rainbow of colors. If the colors match between the two words (the one written by the child and the one written by the therapist), it means that the child could write properly using pauses and liaisons.

The difficulty of the activity can be adapted by changing the following parameters:

- The chosen word of sentence.

The following metrics tracking the proficiency of the child can be required:

- The difference between the number of strokes reccorded by the therapist and the child.

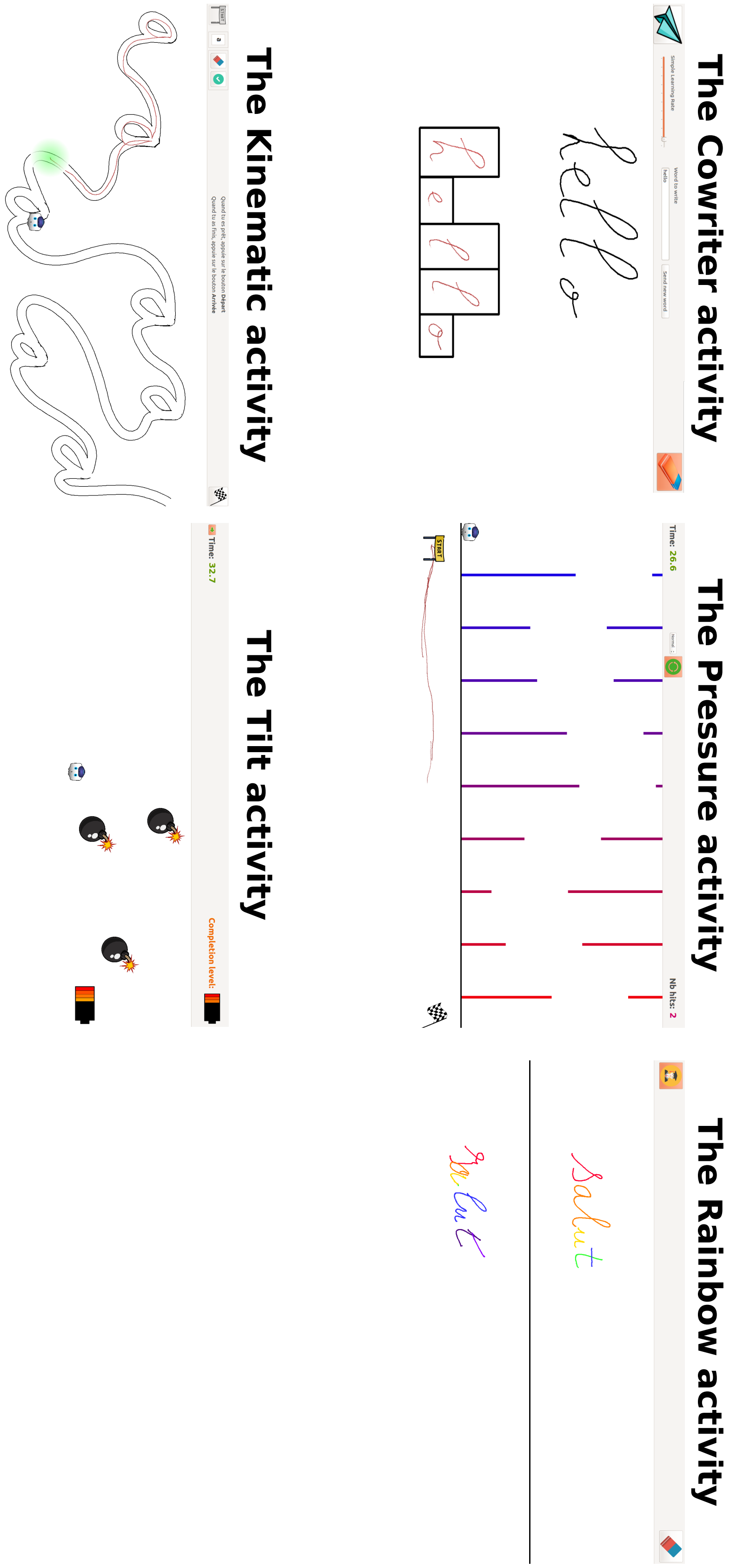

Figure 8.1: The different remediation activities used in the long term study with R.

Results

Activities assessment

One way to follow the progress of the child is through the remediation activities, by following the evolutions of the corresponding, afore-listed metrics along the 30 weeks R. was involved in the therapy.

In Figure 8.2, we can see the evolution of the length of the words that the child writes during the *Co-Writer activity*. The word length is a good indicator of handwriting difficulties since R. could initially only handle simple monosyllabic words. We can see that at the beginning of the therapy, R. was writing simple short words composed of simple letters like "maman", and progressively moved on to writing longer and harder words like "jamais" at week 10 or "football" or "serpent" at week 30.

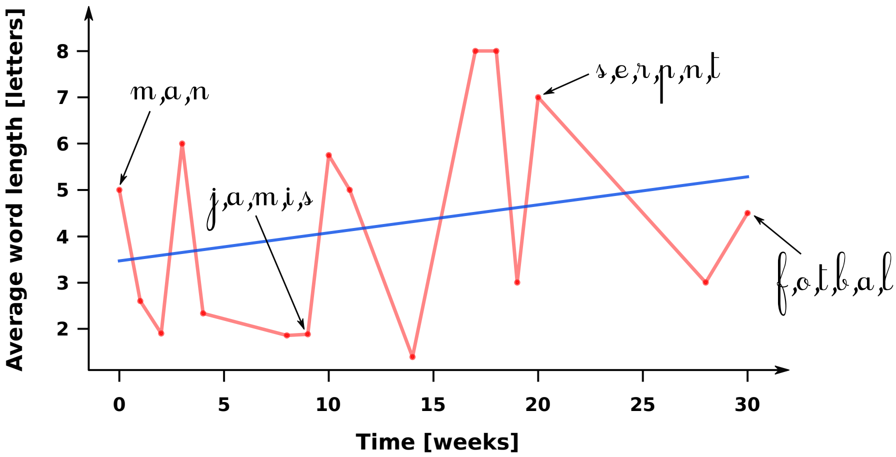

Figure 8.2: Average number of letters in the words written by the child on each session (one session per week). The blue line represents the evolution of the average word length computed with a linear regression. To also illustrate the evolution in letter difficulty, the letter used in some of the sessions are also reported.

In Figure 8.3, we can see the evolution of the progress metrics related to the *Kinematic activity* during the 30 weeks of R.'s therapy. Despite some fluctuations (some due to an increase of the robot's speed during week 10 and 20), we can see an overall increase both in the success ratio (ratio between the number of points recorded outside and inside the path) which we value to be a proxy of precision, as well as in the child's handwriting speed.

In Figure 8.4, we can see the evolution of the progress metrics related to the *Pressure activity* along the 30 weeks R. was involved in the experiment. We can see that the time to reach the end of the maze (a proxy of R.'s proficiency in the exercise) stays relatively constant in average (around 15 seconds) despite a clear increase in the exercise difficulty. This shows an improvement in the performance of R. along the 30 weeks of therapy.

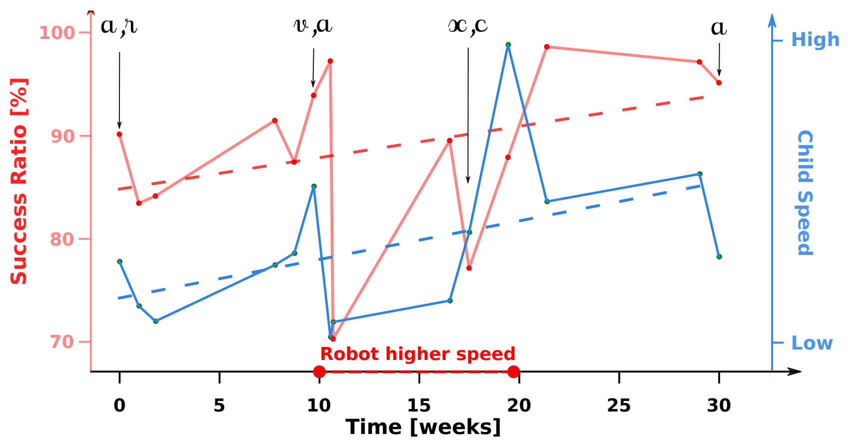

Figure 8.3: Evolution of the metrics related to the *Kinematic activity*. In red, the Success Ratio (ratio between the number of points recorded outside and inside the path) is presented along the 30 weeks while in blue we can see the child's speed computed as a number of pixel per seconds. The dashed lines represent the linear interpolations of both the success ratio and child's speed. During weeks 10 and 20, the robot's speed was increased by the therapists.

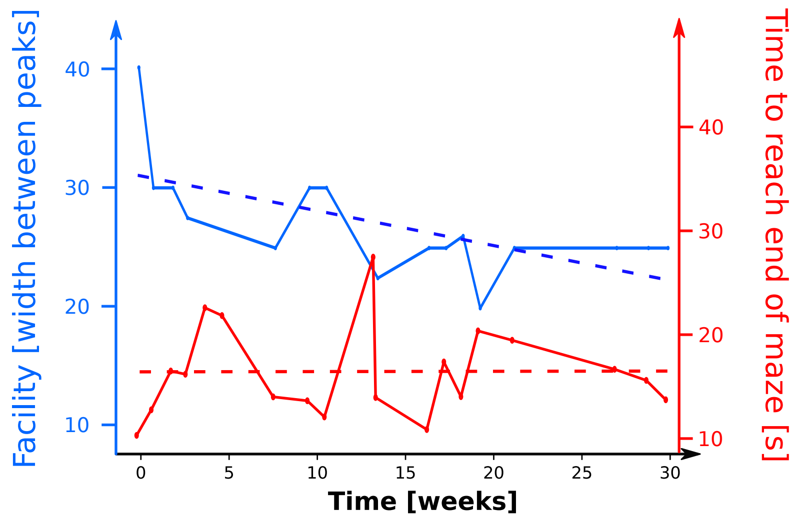

Figure 8.4: Evolution of the metrics related to the *Pressure activity*. In red, the time to reach the end of the maze is presented along the 30 weeks, while in blue we can see the width between the peaks which is a proxy of the maze difficulty. The dashed lines represent the linear interpolation of both the maze's difficulty and the time to reach the end of the maze.

In Figure 8.5, we can see the evolution of the time R. was taking to collect the five batteries composing one gameplay in the *Tilt activity* (see Section 8.2 for additional information). Contrary to the other activities, no significant progress seems to be made by R. along the 30 weeks.

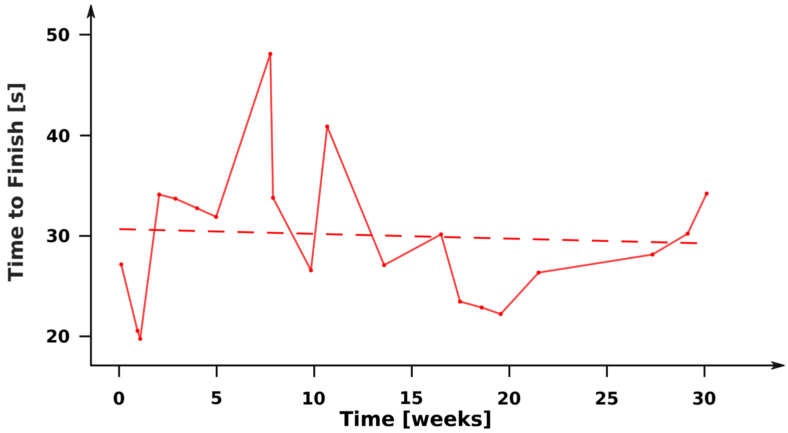

Figure 8.5: Evolution of the time required by R. to finish the *Tilt activity* along the 30 weeks he was involved in the therapy. The dashed lines represent the linear interpolation of the time to finish the activity.

Clinical Assessment

Figure 8.6 shows the evolution of the BHK scores (both quality and speed) during the 30 weeks R. has been using the system. While it was impossible to score the BHK test conducted during the first week of the therapy due to a too low quality (see Figure 8.7), the following weeks register a remarkable progression for both the quality and speed score. As expected, we can see that whenever R. tried to go faster, quality decreased for a brief period of time. Moreover, we can see that, even if the z-score indicates that R. still presents negative Z-scores (meaning lower quality and speed than an average kid of the same age and gender), both the quality and speed score, by the end of the therapy sessions, have exceeded the dysgraphia threshold (defined as a Z-score of -2) meaning that in view of this result, R. would not be considered dysgraphic anymore according to the BHK test. At the end of the 30 weeks of the therapy, R. was ready to go back to a regular school where he received special education. We can see in Figure 8.7 different examples of BHKs conducted by R. during his therapy sessions. As can be seen, the quality of the first BHK conducted 9 months before the first session is of too low quality and thus non scorable. Starting from the beginning of the therapy, we can see both an improvement in the handwriting quality and in the speed since the number of words copied by R. constantly increases.

Conclusion and Discussion

In this chapter, we introduce a system aiming at remediating handwriting difficulties, that was tested in a 30-weeks long study with one child name R.. One of our focus was concerning the acceptability and usability of the system in a long term interaction. The progressive and significant re-engagement of R. with handwriting, leading him to move

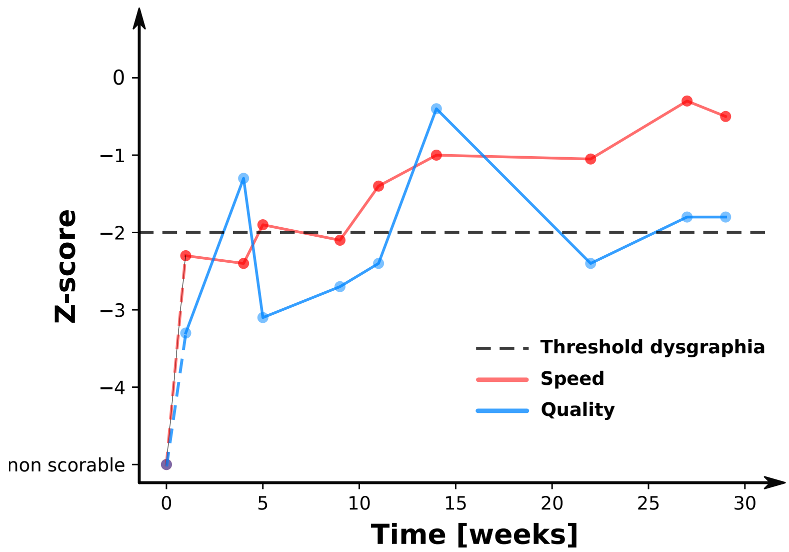

Figure 8.6: Evolution of the Clinical BHK scores (speed and quality scores) during occupation therapy sessions. The z-score shows how many standard deviation the handwriting quality is compared to other children of the same gender and age. By definition of the BHK test, a child is diagnosed dysgraphic if his score is below -2.

from a situation of total refusal to write to being able to rejoin a regular school with special education, is a preliminary proof that our system can be successfully used in a long term interaction to remediate handwriting difficulties. A video has been made, following R. along the sessions and can be found at this link[1].

One way to control the progression of the child is though the metrics that were extracted during the different remediation activities. Following this idea, we tracked the progresses in three of the four remediation activities namely the *Co-Writer*, *Pressure* and *Kinematic* activities. Our hypothesis is that an improvement along one activity should be associated with a similar improvement along the corresponding aspect of handwriting. In other words, if R. is presenting improvements in the control of the pressure in the activity *Pressure*, we hypothesize that this translates in an improvement in the control of pressure during handwriting and thus leads to an improvement of the global quality of handwriting.

At a global level, this hypothesis seems to be confirmed: the steady improvements observed in three of the activities (shown in Figures 8.2, 8.3 and 8.4) and aligned with the remarkable improvement in the BHK quality and speed scores (shown in Figure 8.6 and 8.7). The improvements were such that, by the end of the therapy, R. was not considered dysgraphic anymore according to the BHK test and could reintegrate regular school (where he was still receiving special education). In addition, the therapists reported a clear motivational improvement since R. was very strict with the scheduled (asking to leave as soon as the session was over) at the beginning of the therapy, and

[1] https://vimeo.com/359343838/0c72433f2b

didn't complain anymore after a few sessions. He was even quite proud to be part of the experiment and showed the system to another child in handwriting reeducation. That being said, it is important to remark that the progress made by R. during the sessions is hardly exclusively caused by the proposed system A likely concurrent, and possibly even greater cause, could have been the quality of the human interaction between R. and the therapists

As anticipated in the introduction, the model developed in Chapter 7 couldn't be used in this study since it was trained on data acquired on iPads tablets, while this study relied on a Wacom system such as the one described in Chapter 4. The adaptation of the activities' difficulty through the different metrics as well as the suggestion of remediation activities was done by the therapists rather than automatically. The IReCHeCk project, funded by the Swiss NSF and the French ANR, started on February 2020, exactly aims at extending the preliminary study reported here, by (i) integrating the remediation activities here presented with the assessment model proposed in Chapter 7, (ii) expanding the interaction analysis, by monitoring behavioral cues on the child and accordingly adapting the robot's own behavior, and (iii) assessing the usability of the whole system, by testing it both in clinical and school environment. In the same way, the Dynamico project, introduced in the next Part of this thesis, is also currently conducted with the goal of creating a lighter version of the remediation system here proposed, fully integrated with the model for automated assessment of handwriting difficulties, running on iPads, and not requiring a robot, in order to have a broader use and higher impact on the society.

This Chapter suggests that robots can be a useful and beneficial element in handwriting activities, specifically considering the use of a social robot in the context of remediation of handwriting difficulties. Considering the importance of the social aspect of the interaction between the child and the robot, especially the importance of the child's perception of the robot, the next Chapter of this thesis explores the importance of the design of robot behaviors for a better acceptance by children.

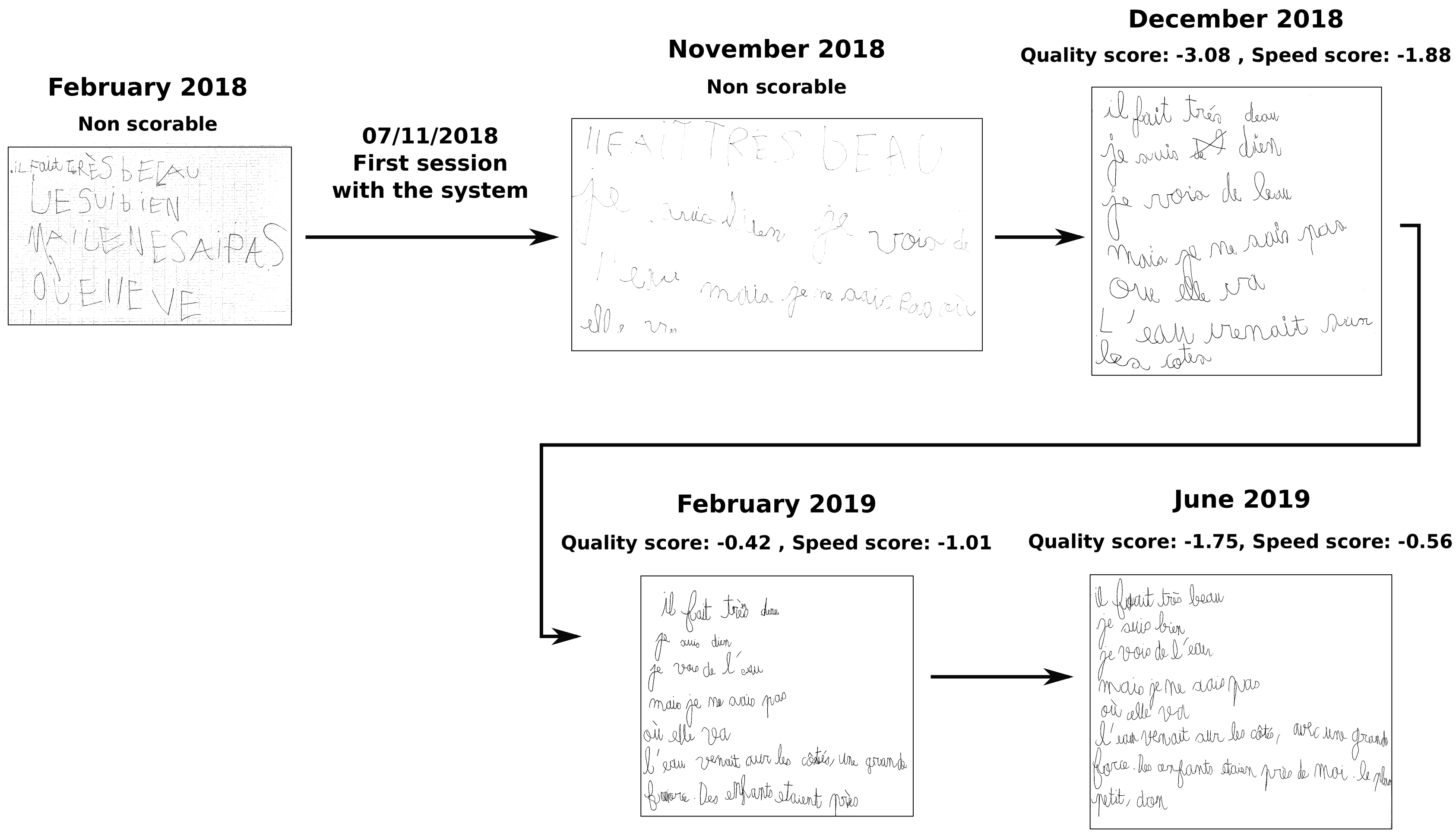

Figure 8.7: Examples of different BHKs conducted by R. before and during the therapy sessions. The Quality and Speed score is given for the different BHKs.

9 Enhancing learner's engagement

Introduction

As already seen in the first Chapter of this thesis, the Co-Writer research project aims to help children with handwriting difficulties with an original approach: the child plays the role of the teacher and a robot is a student requiring help to improve its handwriting. This approach is called learning by teaching and presents several advantages. First, it brings a positive reinforcement of the child's self esteem (which is often particularly low for these children) as he/she becomes the one who "knows and teaches" and is no longer the worst student in the classroom [107]. Secondly, we can observe a huge gain of motivation as the child, feeling responsible of the robot, is committed to the task with a higher intensity compared to when practicing in a normal situation. This particular interaction where children feel responsible of the robot is called the Protégé effect [108]. For this psychological aspect to work, the child however needs to believe he/she can teach the robot, that its effort are useful. In other words, for the protégé effect to be fully functional, the robot needs to be credible in its role of peer learner, exhibiting human-like characteristics and behaviors and should not be cold and static as a robot. One of the limitations concerning the Child-Robot Interaction within the Co-Writer project comes from the unnatural-looking idle behaviour of the robot (i.e., when the robot does not have anything to do, which actually accounts for the majority of the experiment time). In an effort towards improving the credibility of the robot as a peer learner, and thus increase the efficacy of the Co-Writer scenario as a handwriting remediation activity, in this chapter we address the following questions: What can be done during these idle moments, to make the robot act in a natural way.

This chapter present the work conducted in order to make the robot "look more human" with the introduction of subtle movements during idle moments and corresponds to the following publication:

T. Asselborn, W. Johal and P. Dillenboug, "Keep on moving! Exploring anthropomorphic effects of motion during idle moments", *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, **2017** [109].

Context

Over the years, research has used humanoid robots in various contexts pertaining to child-robot interactions, such as autism therapy [110, 111] or educational scenarios [112, 21]. Research findings suggest that the exhibition of social intelligence is a key factor for the acceptance of assistive robots in daily usage [113, 114, 115, 116]. In order to endow human-robot communication, research aims to improve robots with human-like abilities in their social expression and intelligence. Noticeable progress has been made for functional tasks, such as holding conversations [117], catching objects [118] or expressing emotions [119, 120]. While a large corpus of literature aims at improving the robot's performance during a task, a much smaller attention has been given to the robot's behavior off a task, i.e. during what we call *idle moments*.

Non-verbal movements can be categorized into five classes [121]: (1) *emblems*, (2) *illustrators*, (3) *affect displays*, (4) *regulators* and (5) *adaptors*. The first four are communicative movements, used with a semantic goal such as the "OK" sign (1), pointing towards an object (2), nodding to show comprehension (3), or smiling (4) for example. Movements belonging to the last category are postural or other non-verbal movements that are often performed during idle moments (we are still constantly moving even when we are not engaged in a particular task). Adaptors often occur at a low level awareness. They include self touching behaviors (such as twisting of the hair, scratching or swinging of the legs) and object-adaptor (such as tapping of a pen or pushing one's glasses up one's nose) [122]. Adaptor movements can be performed at different frequency, depending on personality and context, and are considered to be social signals of communication. In robotics, it is commonly admitted that these micro-movements can help in making the interaction more natural and the robot more credible [123, 124]. Indeed, the absence of movements during idle moments may make them look like statues - frozen while waiting. However, their impact on anthropomorphism and on task performances hasn't been studied yet.

In this chapter, we explore children's perception of robots (with or without adaptor movements) in terms of anthropomorphism and performance. We hypothesize that a robot displaying adaptor movements will be perceived to be more human and friendly than another one displaying a static posture during idle periods. Conversely, we also make the hypothesis that a high intensity of idle motion can be disruptive and thus reduce children's performances in an attention demanding task such as a memory game.

We designed a within subject experiment in which children were playing a memory game against two Nao humanoid robots. The two robots performed functional motions in the exact same way, but only one of them displayed adaptor movements. We compared attentional behaviour of the children by recording their gaze, and evaluated their subjective perception with a questionnaire.

Related work

Anthropomorphism plays an important role in the attribution of social skills to robots and gestures are a part of the anthropomorphic cues making robots more social [125]. In [126], authors compared two robots during an interaction with a human: one using gestures when speaking, while the other stayed motionless. The robot using gestures "was more anthropomorphized, participants perceived it as more likable, reported greater shared reality with it and showed increased future contact intentions" compared to the motionless robot. The impact of the non-verbal expressions of emotions and their anthropomorphic incidence was also investigated as in [127], where authors showed that "the robot's socially intelligent behaviour (i.e., the expression of emotional states) affected subsequent user evaluations" of robot's anthropomorphism and "clearly influence both perceptions of the interaction partner as well as the interaction itself" during HRI.

While there exist studies exploring the effect of movements during interactions, the effect of gestures during idle moments (i.e. moments when the robot doesn't perform a task) has not been largely studied, especially in robotics. Some research in this domain concerns avatars. In [128], the authors proposed to "generate subtle head and face movements while a virtual character is in idle mode" and showed that this leads the character to be perceived as more friendly. [129] reports an implementation of adaptor movements for the whole body of virtual agents. The authors designed a module that allowed the generation of subtle movements such as "changing balance because of fatigue, variations in body posture caused by small muscle contractions or eye blinking" to avoid "the lack of movements between different animation clips responsible of the unnatural-looking frozen posture between motions". This research shows that avatars performing adaptor movements were perceived as more human by users [130]. A few instances of recent research present implementations of adaptor movements for humanoid robots. In [117], the authors implemented one hundred interactive behaviours on a Robovie humanoid robot where twenty of them were "idle behaviors such as scratching the head or folding the arms". However, the goal of this research was totally different than ours since the behaviours were implemented to increase the number of interactions between children and the robot. In [131], the authors showed the importance of bodily motion in speechless situations by exploring five distinguished roles of motion used during moments where one does not speak: *mood-setting*, *observing*, *listening*, *expecting* and *idling* (corresponding to our adaptor movements). The study does not explore the anthropomorphical incidence of one particular category taken separately as the aim of the research was to provide a guideline to "help create preferable motion designs of a humanoid robot" in speechless situations.

Method

Design of adaptor modes

In order to evaluate the degree of motion the robot should adopt during idle moments, we designed three different levels of adaptor mode. We created a database containing a set of 60 animations that mimic adaptor movements. These animations were labeled with a mode according to their intensity (namely: low, medium and high). Several factors may influence the intensity of an animation. For instance, the angle swept by joints during motions, the duration of animations, the velocity of joints, or the joints concerned (a movement of hand will have a lower impact compared to a movement of the shoulder for example). The motion intensity can thus be computed in several ways, possibly only loosely related to the intensity perceived by the observed. Since the focus of this study is the impact of robot idle motions, and their intensity, on the human interacting with it we decided to manually sort all the animations according to the perceived intensity.

In order to check the consistency of our manual sorting, we computed the intensity of the animations composing the three groups to verify the manual annotation. the product of three variables was used: the angle swept by joints during the animation, the angular speed of these joints, as well as the duration of the animation. In addition, as some joints have a smaller impact on the motion (hands, fingers) than others (knees, shoulders), a weighting factor was also introduced (see Table 9.1) to take that into account in the intensity calculation. This is based on a method commonly done to compute the quantity of motion [132]. Concretely, the following equation was used to compute the intensity I of a given animation:

$$I = t_{motion} * \sum_{i=1}^{n} c_i * \alpha_i * v_i \tag{9.1}$$

with:

- c = weighting factor depending on the type of joint concerned (see table 9.1)
- α = Angle swept during motion
- v = Mean joint velocity while moving
- t_{motion} = Total duration of motion
- i = The joint id starting from the first one to the last one (the n-th)

Ankle	Elbow	Hand	Head	Hip	Knee	Wrist
6	2	1	3	4	5	1

Table 9.1: Weighting factor assigned to the different joints of Nao to compute the quantity of motion of the different animations.

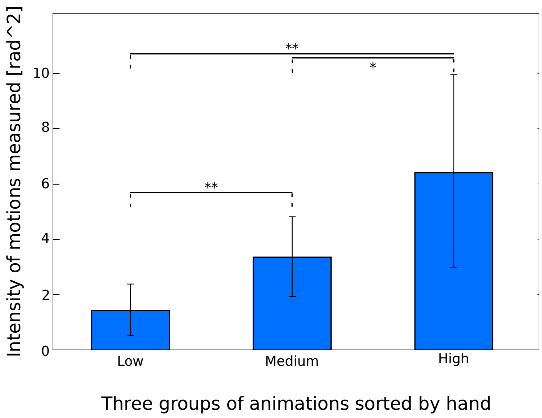

Figure 9.1: The measured intensity for the different groups of animations. Asterisks denote statistically significant differences between the mean of different groups found with a t-test. $^{*}p < 0.05$ and $^{**}p < 0.01$.

Figure 9.1 shows the intensity of motion for each of the three levels sorted manually (calculated with equation 9.1). Even though we observe high variability, a t-test between the means of the groups of animation showed statistically significant differences between each of them. We decided therefore that the manually sorted groups were truly representative of different levels of intensity, both from a human observer perspective and from an intrinsic motion perspective.

In addition, animations were launched with a frequency depending on the level of intensity of adaptor movements: animations were triggered every 25 seconds on the first level of intensity (low), every 18 seconds for the second level (medium) and finally every 12 seconds for the third one (high). For sake of clarity, this scale will be called the *adaptor scale* from now on.

Experimental evaluation of adaptor modes

The memory game

The memory game features a set of matching card pairs that are randomly disposed face-hidden. Players alternatively flip two cards (now face-showing) in an attempt to find and collect matching pairs. The player that wins is one who collected the maximum number of pairs. Since we targeted children younger than 6, we decided to limit the number of cards to 16 (disposed in a 4 by 4 grid). The choice of the memory game was done due to the simplicity of its rules, the short duration of the game and the fact that it is a competitive game (which pushes task engagement). It requires children to pay attention and to memorize previously returned cards in order to defeat the robot players. A very simple artificial intelligence was implemented for the robots' strategy. The robots were forgetting the cards flipped during a given round R_i according to a certain probability depending on the difference between the current round R and R_i. The probability to forget a card on R_i is given by the formula: $P(i) = 1 - 1/(R_i - R)$ (with i being the index of the round). The game was implemented on an Android tablet.

Experimental Setup

The experiment was conducted in a school in France and involved 20 children (12 girls, and 8 boys) aged 5 years old. Children played the memory game against two identical NAO robots placed in front of them (see Figure 9.2, A and B). In addition to being physically identical (to remove any bias, the two robots were of same color and gender (i.e. called Mimi and Clem)), the robots behaved identically whenever actively involved in the game. This means that they acted the same way when returning a card, exhibited similar behaviors when winning and losing and employed the same artificial intelligence based decision-making process throughout the game. The head movements were also implemented on both robots as a lot of studies showed their importance on human perception [128, 133, 134, 135].

The only difference was that while one robot (called static robot) was completely static during idle moments, the other (called adaptor robot) would display specific movements, called adaptor movements, during this period.

During the experiment, each child played three games, each time at a different level of adaptor mode (in a random order).

Measures

Task Performance The number of matching pairs collected during rounds of the memory game by the child has been used as a measure of performance.

Child's Visual Attention Two cameras facing the children were used (see Figure 9.2, C). The first one was simply used to record the experiment (audio and video) while the second one was used to track the child's gaze and head pose.

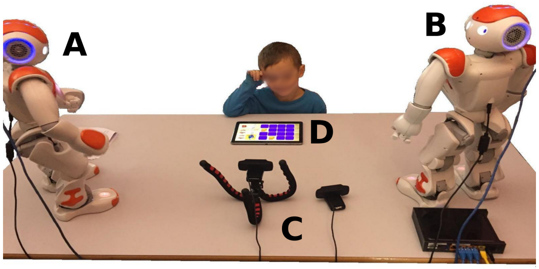

Figure 9.2: A and B: The two Nao robots playing memory. One of them stays static whereas the other performs adaptor movements during idle moments. C: Two cameras are facing the child. One is used to record a video of the experiment, the other is used to detect if the child is either looking at the left robot, right robot or the tablet computer. D: The tablet implementing the memory game.

Perceived Anthropomorphism and Proficiency We adapted some items from the GodSpeed questionnaire [136] to measure the anthropomorphical difference between the two robots[1]. Children evaluated the robots on four different metrics, namely humanity, friendliness, attention and proficiency. Specifically, the questionnaire included 6 items with a 5-levels scale, assessing each robot individually and 4 open questions.

Hypotheses

H1: The robot with adaptor movements will be perceived more human and friendly by the children compared to the other one. We think that this impact will depend on the level of the adaptor mode. We also don't expect to see any correlation of the adaptor movements with the perceived attentiveness of the robot and its perceived proficiency in the task.

H2: The best robot (the one that collects the highest number of pairs during a game) will be perceived as more proficient and attentive by children independently from the idle mode.

H3: The child will be more prone to look at the robot displaying adaptor movements. Once again, we expect to see an increase of this tendency with the level of adaptor mode.

[1] https://github.com/asselbor/Questionnaire

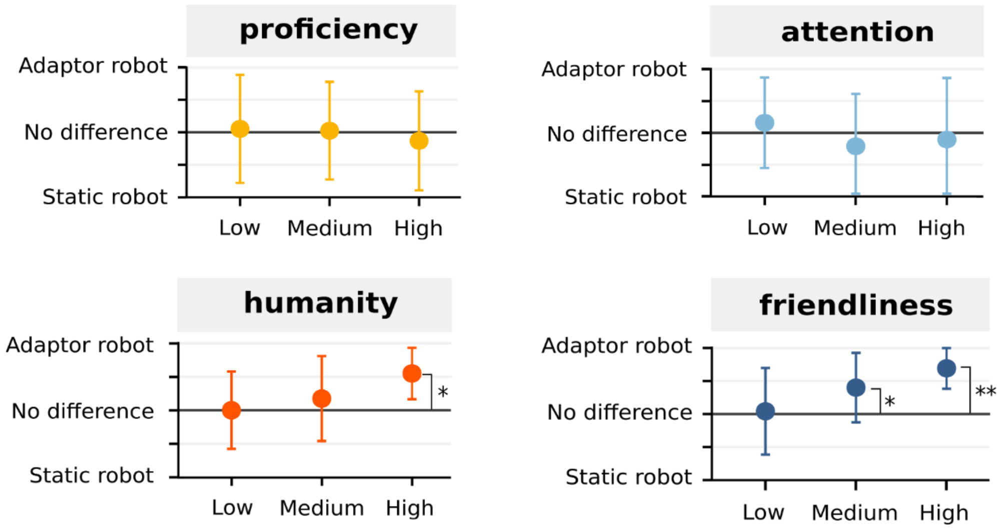

Figure 9.3: Anthropomorphic differences between the robot with adaptor movements (called "Adaptor Robot") and the other one (called "Static Robot") as a function of adaptor mode in terms of friendliness, humanity, proficiency and attention (*$p < 0.05$ and **$p < 0.01$).

Results and Discussion

H1: Perceived Anthropomorphism

Figure 9.3 shows an increase in terms of perceived humanity and friendliness with the level of adaptor mode. As the data were not found to follow a normal distribution, a one-sided Wilcoxon test was used and showed that the robot with adaptor movements was perceived more friendly ($M = 1.38$, $SD = 0.92$, $t(19) = 2.5$, $p < 0.01$) and human-like ($M = 1.1$, $SD = 1.3$, $t(19) = 9.0$, $p < 0.05$) compared to the static one for adaptor mode on level 3. It was also found to be more friendly at adaptor mode on level 2 ($M = 0.8$, $SD = 1.16$, $t(19) = 13.0$, $p < 0.05$). For adaptor mode on level 1, no statistical significance was found. As expected, we didn't find any difference concerning the perceived attentiveness of the robots. No significant difference was found for the proficiency metric for any adaptor mode level, meaning that the static and adaptor robots both looked equally competent.

It was also extracted that the robot with adaptor movements was found to be more human-like and friendly at level 3 than at level 1 ($p < 0.05$ for both).

H2: Proficiency

As expected, it was found that the best robot (the one that collected the highest number of pairs at the end of a game) was perceived more proficient than the other one (Wilcoxon test: $t(49) = 125.0$, $p < 0.01$). No influence of the static or adaptor mode was found on

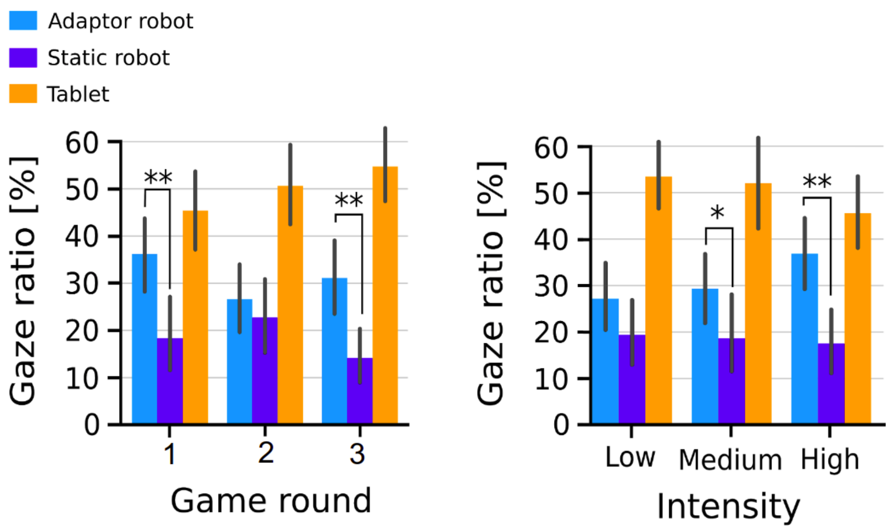

Figure 9.4: Gazing proportion for the robots and the tablet according to (Left): the game session, (Right): the mode of adaptor movements (*$p < 0.05$ and **$p < 0.01$).

the proficiency metric. These results are not surprising but tend to prove the proper functioning of the questionnaire.

H3: Visual Attention

In Figure 9.4, we can see that children tend to look more at the robot with adaptor movements compared to the static one. As expected, this tendency depends on the mode of the adaptor movements (see Figure 9.4-right). The proportion of gaze on the static robot seems to be relatively constant (around 20%) across the three adaptor mode conditions. The gaze proportion to the tablet seems to decrease as the intensity of the adaptor movements increases. We expected the mode of adaptor movements to be correlated with the average number of pairs collected by the children during a game, as they will be more distracted by the robot (an intuition supported by the results in Figure 9.4, suggesting that the extra time dedicated to the adaptor robot is taken from the time spend looking at the tablet). However, no statistical difference in the number of cards collected was found in function of the mode of adaptor movement (see Figure 9.5, right). These results show that adaptor movements are, in addition of making the robot look more natural, not disruptive for performances in this task. Of course, this result is only valid for this specific example. This result will be further developed in the next Section.

Figure 9.4-left showed a significant increase in the proportion of gaze match on the tablet according to the game session (children were playing three games in a row during the experiment). We believe that the children's habituation of the robots is responsible of this tendency. As time goes by, children become less distracted by the robots and more focused on the tablet and therefore on the memory game itself. We would therefore expect

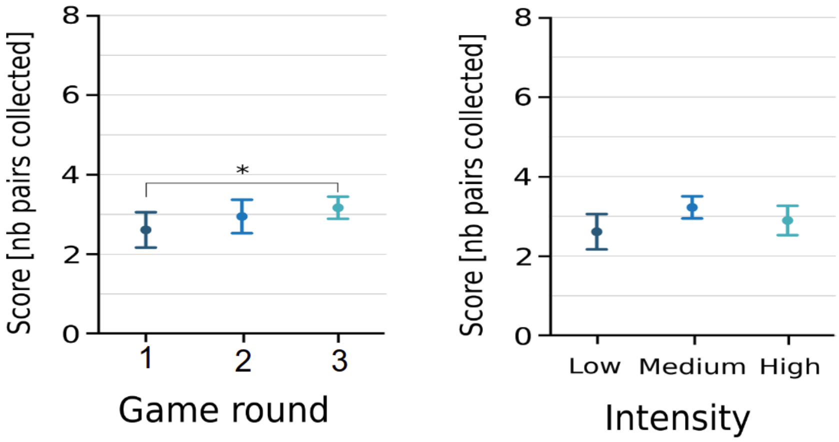

Figure 9.5: Number of matching pairs found by children according to (left): the game session, (Right): the mode of adaptor movements (*$p < 0.05$).

an increase in the children's performance measured by a higher number of matching pairs collected and therefore by a diminution of the game session duration as the game is more likely to be finished quicker. The increase in the number of matching pairs collected as a function of the game session could be extracted as can be seen on Figure 9.5-Left. A one-sided Wilcoxon test showed a statistical increasing number of matching pairs between game round 1 and 3 ($p < 0.05$).

Conclusion and Discussion

In this chapter, we saw that the introduction of adaptor movement during idle moment can enhance the humanity and friendliness of a humanoid robot. Indeed, statistically significant correlations were extracted between the adaptor movements and the humanity and friendliness perception. It was also shown that these correlations are dependent on the intensiveness of the adaptor movements during idle moments. Interestingly and even if the exhibition of adaptor movements was attracting children's gaze depending on their intensity, these movements were not found to be disruptive for the task as the number of pairs collected by children stayed relatively constant across the three levels of adaptor mode.

However, it is important to temper the results extracted through this experiment. Indeed, it would be very naive to affirm that there is a linear relation between the intensiveness of adaptor movements and the anthropomorphic capabilities of the robot. We believe that a threshold (in term of intensiveness of adaptor movements), where the movements are no more beneficial for anthropomorphical perceptions, might exist. In this research, the intensiveness of the movements might just be too low to be disruptive for children. In a future work, it would be interesting to measure the anthropomorphic implications of more intensive adaptor movements and to investigate at which point the intensiveness of adaptor movements becomes more disruptive than useful for the interaction.

Another limitation comes from the fact that this experiment was conducted only on five years old children. In a future work, it would be also interesting to check if our assumptions also stand for other age groups, such as adults, or for different cultures.

In the next Chapter, we explore the use of small mobile robots to support learning the basic handwriting skills, typically acquired in pre-school.

10 Enhancing Letters Representation

Introduction

In this chapter, we aim to exploit the sensory information provided using the tangible, haptic-enabled Cellulo robots [137] to enhance the acquisition of some important aspects of handwriting generally learnt during pre-school including the representation of the letter's grapheme (letter's shape) and ductus (letter's drawing pattern). To this end, we designed innovative activities aiming to provide enhanced learning experience. In a next phase, the activities developed and tested here can possibly go and enrich that portfolio of remediation activities that is the main theme of this part.

This work corresponds to the following publications:

T. Asselborn, A. Guneysu, K. Mrini, E. Yadollahi, A. Ozgur, W. Johal and P. Dillenboug, "Bringing letters to life: handwriting with haptic-enabled tangible robots", *Proceedings of the 17th ACM Conference on Interaction Design and Children (IDC)*, **2018** [9].

A. Guneysu, A. Ozgur, **T. Asselborn**, W. Johal, E. Yadollahi, B. Bruno, M. Skweres and P. Dillenbourg, "Iterative Design and Evaluation of a Tangible Robot-Assisted Handwriting Activity for Special Education", *Frontiers in Robotics and AI*, **2020**.

Method

Pedagogical Design

A letter representation is defined by the letter's grapheme (visual representation), ductus (the direction its writing should follow between starting and end point) and phoneme (its pronunciation).

Our objective is that children enhance the link between the grapheme and the ductus with the phoneme. In other words, they should be able to write the letter with the right dynamic when they hear its pronunciation. It is important to notice that we excluded the fine-motor aspect of handwriting from this study as it is different from the visual perception and visual-motor coordination aspects that we are interested in. Concretely, our aim here is complementary to the one of the remediation activities proposed in Chapter 8: in fact our goal is to enhance the child's letters representation (grapheme and ductus) but not the way these letters are produced on paper.

Using this, the following sub-goals are defined:

1. **Remember Grapheme**: Memorizing the letter's physical representation.
2. **Remember Ductus**: Memorizing the letter's drawing pattern.
3. **Memorizing the Phoneme to Ductus-Grapheme Link**: Memorizing the link between the letter's pronunciation (phoneme) and the corresponding grapheme and ductus.

In order to help children enhance these multi-modal mental representations, several learning activities were designed and are described in the section below.

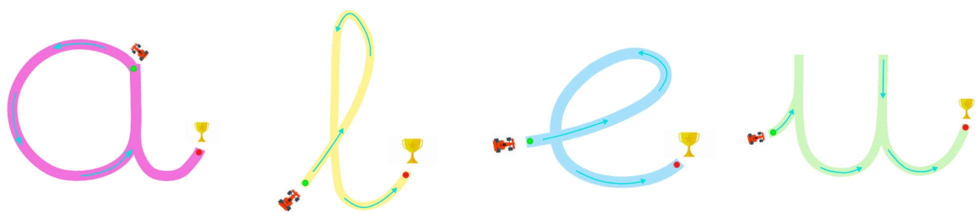

Figure 10.1: Letter maps with direction cues.

Participants

As explained in the introduction, we focused in this study on the visual perception as well as the visual-motor coordination that are two aspects of handwriting that need to be developed early in the handwriting learning process. That is why we targeted in this study children from kindergarten (5.2 years old in average at the moment of this study). We obtained the consent of the 17 children from the kindergarten of this school as well as the one of their legal responsible. All experiments and tests were carried out in a special classroom put at our disposal under the control of one of the school's teacher.

Activity Design

Before designing our activities, we discussed with pre-school teachers how we can position Cellulo in handwriting activities. The following robot's features were found to be particularly useful: haptic information, autonomous motion and synchronized behavior of multiple robots. Haptics allows a child touching the robot to receive individuals, instantaneous information in a form of a force, or vibration, that the robot transmits to the child's hand. This can be used in our case to inform the child of errors and guide him towards the right ductus. Autonomous motion makes the robot capable of autonomously reproducing the ductus, whereas synchronization helps designing collaborative team activities. In order to avoid the split attention effect and an excessive cognitive load for the school-children, we started with passive activities where children do not have to move the robot and then continued with active learning activities where children have to move the robot and get haptic, LED-based and sound feedback.

Since children were taught the letters '*a,l,e,u*' at the moment of this study, we decided to focus on this subset of letters during our teaching activities.

Watch the Robot: Learning letter's Grapheme and Ductus

The goal of the first activity is to help children in learning the letter's ductus by watching the robot writing the letter on a map with the letter's grapheme drawn on top of it (with graphical elements and arrows providing direction cues as can be seen in Figure 10.1). The robot performs the motion dynamics of the letter with its LED turning red when the robot starts writing, and turning green when it finishes. The lights are informative of the writing process, and the movement of the robot on top of the grapheme provides an innovative way to represent the ductus of the letter. Finally, the letter's phoneme is generated at the beginning and the end of the writing.

Feel the Robot: Learning letter's Grapheme and Ductus

In the second activity, we add another representation of the letter's ductus by asking the child to feel the way the letter is drawn. To do so, we ask the child to put his/her hand on top of the robot, while it autonomously writes the letters as in the activity "Watch the Robot".

The child does not actively move the robot, but only follows its automated motion as it draws the letter. This activity can be seen in Figure 10.2-top.

This passive hand-held activity is a prelude for the following activities where the child takes on a more active role.

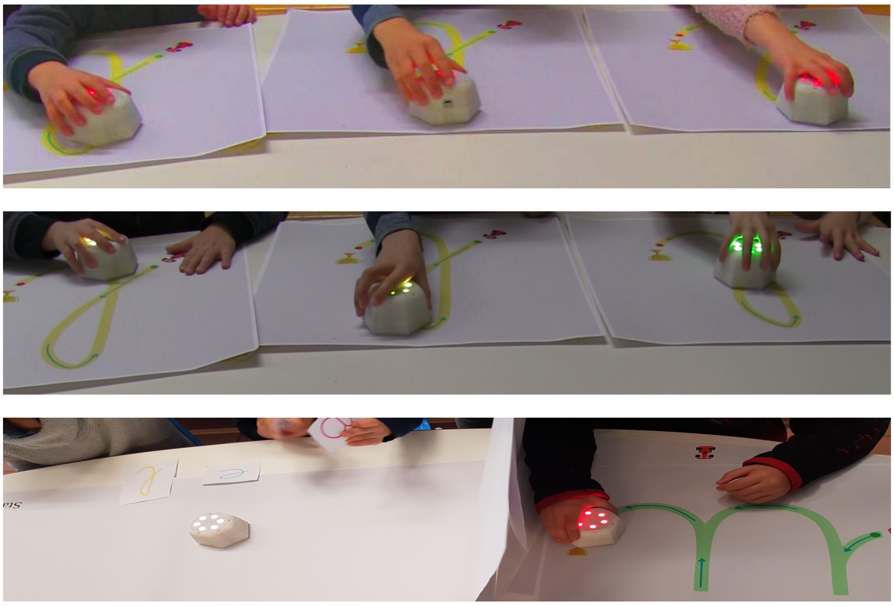

Figure 10.2: **Top**: Feeling the Robot. All robots move simultaneously while each child keeps a hand on top of his/her robot in order to feel the ductus of the letter. **Middle**: Driving the Robot. The child drives his/her own robot with real-time haptic feedback when the robot is out of the letter's path. **Bottom**: Team activity.

Drive the Robot: Memorizing the Ductus of the Letter

In the third activity, the child takes a more proactive role since he/she needs to drive the robot in order to produce the correct ductus of the letter displayed on the map (see Figure 10.2-middle). The robot provides assistive haptic feedback by pushing towards the expected path whenever the child moves away from it.

A visual feedback is also offered since the robot's LEDs turn green when the correct path is followed, and turn red otherwise. These two feedback elements condition the child to recognize errors, and they serve as an extrinsic motivation for drawing correctly.

Team Activity: From Ductus to Grapheme

One after the other, a child plays the "writer" and the other two play the "guessers". A physical barrier separates the writer from the guessers, so that they cannot see each other. At the beginning of the activity, all children are given a Cellulo robot and a map, but while the one of the writer reproduces the grapheme of a letter, the ones for the guessers are blank. Then, the writer is asked to write the letter with his/her robot (as in "Drive the Robot"): the motions of this robot are exactly replicated by the robots of the two guessers, who, by looking at them, have to guess which is the letter written by the writer. An illustration of this activity can be seen in Figure 10.2-down.

Performance Measurement

In order to explore the added value of our system to the handwriting learning process, we want to assess the visual perception (representation of the letter's grapheme) and the visual-motor coordination (representation of the letter's ductus) aspects of the learners in detail. In other words, we want to assess the quality of the letter representation in the child's mind in terms of ductus and grapheme. In all the performance evaluation tests we describe below, 8 letters are investigated (the 4 letters (e, u, l, a) targeted during the sessions (with traditional methods and with robots) and 4 letters not taught (y, m, n, o)). The remaining letters (y, m, n, o) are added to see if transfer learning occurs during our teaching session. As described in the next Section, three different sub-skills are investigated:

- From Phoneme to Grapheme-Ductus: The first sub-skill we want to assess is the link between the sound of the letter (phoneme) and how it is drawn (grapheme and ductus). A software (see Figure 10.3-left) with the following functionality was created: the child hears the phoneme of a letter (when pressing button #1) and is asked to draw the corresponding grapheme on a graphic tablet (Wacom Cintiq Pro) with a pen. As the link between the grapheme and the phoneme of the letter might not yet be fully operational, children have the possibility to see the letter's grapheme for a short period (one second, upon pressing button #2). Throughout the test, we ensured that children do not see the model while they already have a representation of the letter in their mind.

- From Grapheme to Ductus: This test evaluates the link between the grapheme and the ductus. As can be seen in Figure 10.3-middle, the graphemes of the letters are displayed on the tablet's screen. Children are expected to draw the letter directly on top of the model. The ductus of the letter is assessed during this test. The quality of the final grapheme is not taken into account since we do not focus on fine motor skills in this study.

- From Phoneme to Grapheme: The last test evaluates the ability of the child to find the right grapheme after hearing the phoneme of a letter. The child has to press a button to hear the phoneme of a letter and find the associated grapheme among a choice of 8 letters (u, y, l, a, e, m and o) as can be seen in Figure 10.3-right.

Experimental Design

Our aim in this study is to explore how the use of haptic-enabled robots can have a positive contribution in learning the representation of letters (grapheme and ductus). In order to compare this new way of teaching, we decided to take the traditional way of

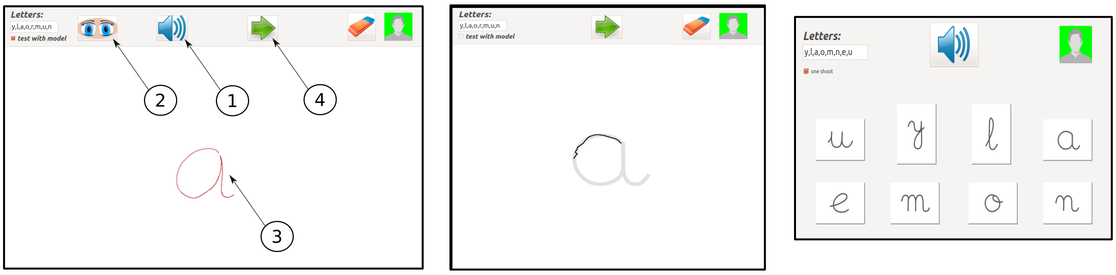

Figure 10.3: **Left**: assessment of the phoneme to grapheme-ductus link. With button 1, the child hears the phoneme of the letter. With button 2, the child has access to the model of the letter (grapheme only) for 1 second. In area 3, the child writes the letter. Once finished, button 4 is used to save the data. **Middle**: assessment of the grapheme to ductus link. The model of the letter is presented on the screen, the child simply draws the letter on top of the model. **Right**: assessment of the phoneme to grapheme link. The child hears the phoneme of a letter and has to select its corresponding grapheme among the shown options

teaching letters' representation (such as the ones presented in [57, 27]) as a baseline for our method.

All participating children were split in two groups based on their pre-test performance (averaged performance was similar in the two groups). The first group, named **"robot group"**, used the robots during their teaching sessions while the other one, named **"control group"**, learned the same letters with traditional methods. The robot group was composed of 4 males and 5 females (mean age 5.36 years old) while the control group was composed of 3 males and 5 females (mean age 5.24 years old).

In Figure 10.4, you can see the schedule of the experiment for both groups. Two teaching sessions (with robots for the robot groups, with traditional methods for the control group) were given between the Pre-test and the Post-test. The session duration for the two groups was controlled to be the same (2 x 40 minutes), and the letters taught during the sessions (e, l, u, a) were the same as well. The comparison between the results of the Pre-test and the Post-test of the two groups allows us to compare the two teaching methods and give us an insight on the possible value that our robots and activities can bring to the teaching of the foundations of handwriting. Finally, in order to have a first insight on what the combined effect of the two teaching methods could be, another teaching session with the robots was given to the control group followed by another post-test called Bonus-test.

Results

In order to assess the test results of within subject studies we used the Wilcoxon signed rank test since our data are not normally distributed. Concerning the between-subject

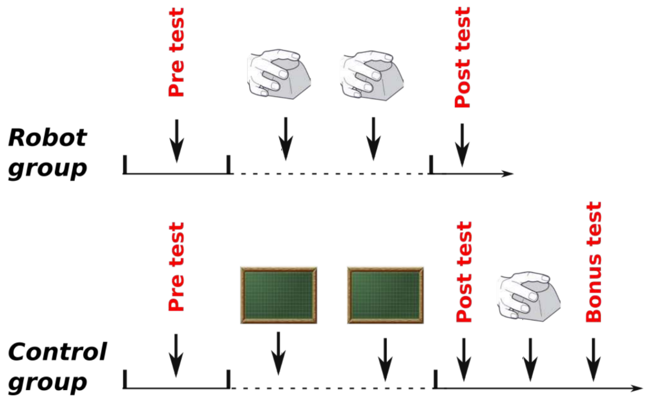

Figure 10.4: Timeline of the experiment.

studies, the Mann-Whitney U test was used.

The majority of the graphs presented in this section displays the results concerning the 4 letters taught during the teaching sessions (u,l,a,e) and omit the results concerning the 4 letters not taught. This is done for clarity reasons as no significant results could be observed for these 4 letters.

From Phoneme to Grapheme-Ductus

In the test assessing the quality of the Phoneme to Grapheme-Ductus link, children were asked to write the letter after hearing the phoneme of it. They also have access to a button showing the model of the letter (grapheme) for 1 second. We hypothesized that the teaching sessions (both traditional and with the robot) would help children to memorize the letters' graphemes, and therefore the number of clicks on the button showing the model would decrease from the pre-test to the post-test.

In Figure 10.5, we can see the mean number of clicks for the letters taught for the two groups. Concerning the robot group, even if no significant difference between the pre-test and the post-test can be extracted, a slight decrease in the number of clicks to see the letter model can be observed. The same test on a larger sample of children would need to be conducted in order to confirm this result.

Concerning the control group, no significant results were found between the pre-test and the post-test. A significant decrease in the number of clicks was found between the post-test and the bonus-test (Wilcoxon signed rank test: $W(36) = 0.0, p = 0.0421$). It might be an added value of the two teaching sessions combined together, but may also be a consequence of getting three sessions of learning, or assessing their knowledge after one sessions instead of two. In order to answer this question, it would be necessary to investigate the impact of 3 teaching sessions in a future work.

The second important aspect we were interested in during this test was related to the quality of the ductus produced. 4 experts were asked to grade every child's letter's ductus between 0 (for totally wrong ductus) and 3 (perfect ductus with proper start and end points and directions). Figure 10.6 shows the mean grade of the 4 letters taught during the teaching sessions for both groups. Concerning the robot group, results show a slight increase but no statistically significant improvement in the letter's ductus after the two teaching sessions with the robots. For the control group, no statistically significant progress was found after the two teaching sessions with traditional methods. However, when the two teaching sessions with traditional methods were combined with a teaching session with the robots, a statistically significant improvement in the quality of the ductus of the letter can be observed (Wilcoxon signed rank test: $W(36) = 1.0$, $p < 0.05$). We find this result promising as it might imply that the combination of the two teaching methods can bring additional value in the learning process of the letter's ductus. In order to confirm this hypothesis, we would need to compare the progress emerging after 3 teaching sessions with traditional methods, with the progress emerging from 3 teaching sessions with traditional methods together with robots, similar to the one done here.

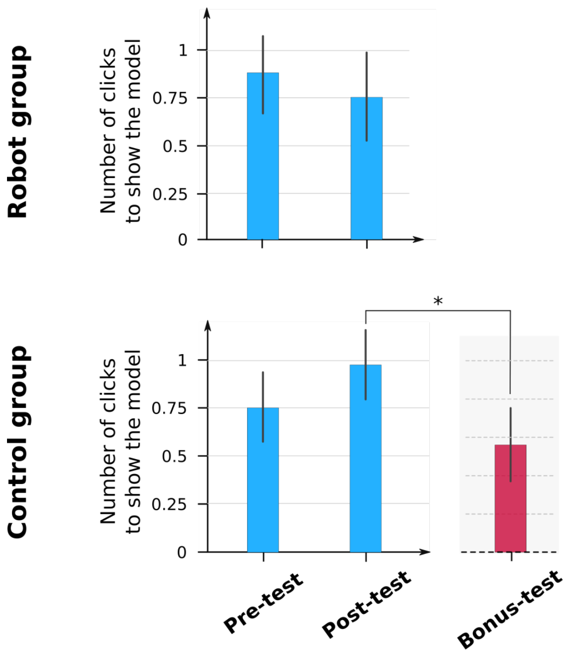

Figure 10.5: Number of clicks to see the letter's grapheme averaged for the letters taught during the test "From Phoneme to Grapheme-Ductus". A significant difference in the number of clicks was observed between the bonus-test and the post-test of the control group (W (36) = 0.0, p = 0.0421)

From Grapheme to Ductus

In order to assess the quality of the link between the grapheme and the ductus, the same expert-based grading method previously described was used on the childrens' production

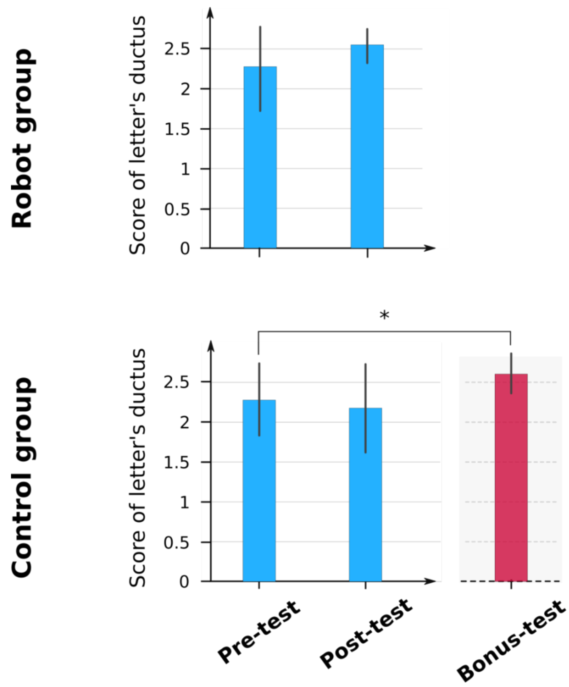

Figure 10.6: Ductus performance of children in the test "From Phoneme to Grapheme-Ductus" for the two groups. Asterisks denote statistically significant differences between the mean rank of population found with a Wilcoxon signed rank test ($*p < 0.05$, $**p < 0.01$)

during this test. In Figure 10.7, the mean score given to the children letters' ductus is presented. In this test, since children were writing directly on top of the model, they needed to remember the ductus of the letter but not necessarily its grapheme.

Concerning the robot group, a statistically significant improvement in the ductus representation of the letters can be observed after the two teaching sessions with the robots (between pre-test and post-test: $W(36) = 0.0, p = 0.0117$). On the other hand, no statistically significant improvement was found after the two sessions (between pre-test and post-test) with traditional teaching methods. In addition, a statistically significant improvement was measured after the last session taught using robots (between the post-test and the bonus-test: $W(32) = 1.0, p = 0.0274$).

A Mann-Whitney U test with independent samples reporting the improvement in term of letter's ductus from the pre-test to the post-test between the two populations (with robots and with traditional methods) was conducted. A statistically significant difference in term of learning gain was observed ($U(32) = 291.5, p = 0.0031$). In other words, children in the robot group were learned more than the children in the control group, in this specific task.

The results presented here show that robots can bring additional value in a typical classroom environment, as we can see progress when robots are used together with traditional teaching methods. Similarly, we see progress when only robots are used, but not when only traditional methods are used. Results should however be taken with a pinch of salt since a possible ceiling effect present in the pre-test of the control group could be biasing the results.

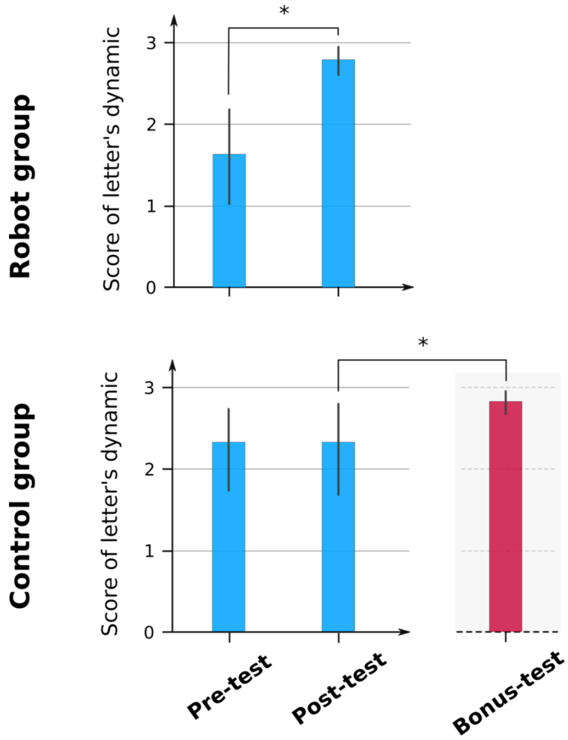

Figure 10.7: Ductus performance of children in the test "From Grapheme to Ductus". Statistically significant progress in the letters' ductus score was observed between the post-test and the pre-test of the robot group ($W(36) = 0.0, p = 0.0117$) as well as between the bonus-test and the post-test of the control group ($W(32) = 1.0, p = 0.0274$)

From Phoneme to Grapheme

In this test, only the first attempt of children while guessing the letter is considered. Indeed, we realized that if children failed to discriminate the expected letter' grapheme among others on their first attempt, they will often randomly selects another letter until finding the right answer.

Figure 10.8 presents the results of this test. The y axis represents the percentage of letters guessed at the first trial (100 % means 4 letters are correctly guessed by the child at the first selection) while the x axis represents the test index.

For the two groups, no statistically significant improvement could be found between the pre-test and the post-test, even though an increasing trend can be observed. However, when the two teaching methods were taken together (between the pre-test and the bonus-test in the control group, with the two traditional sessions and one robot session), a statistically significant improvement can be seen with respect to the pre-test ($W(32) = 0, p = 0.039$). This may indicate the benefits that the combination of the two teaching method can bring for learning visual perception skills and discrimination of the letters. An alternative hypothesis would be that results were not statistically significant after two sessions (both with the robots and with traditional methods) only because of the too small amount of time spent to teach this concept. In this case, adding one teaching session (with robots or with traditional methods) would make the progress statistically significant. In order to get more clarity on these hypotheses, another study with more teaching sessions would need to be conducted.

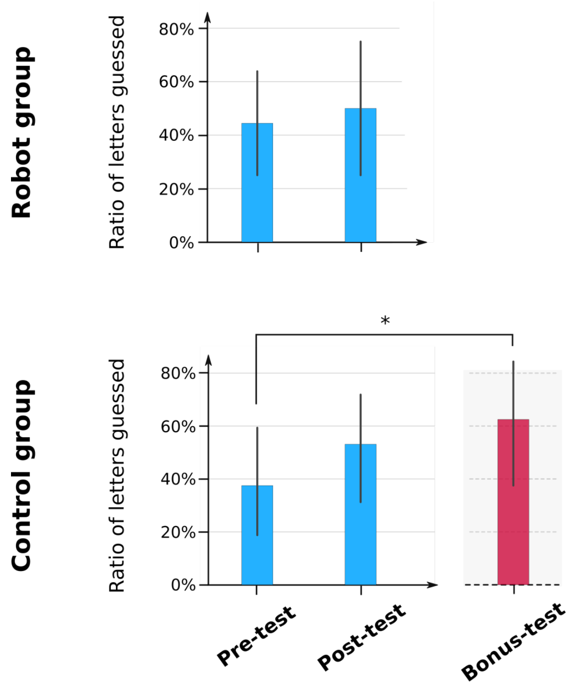

Figure 10.8: Results concerning the link between the letter's phoneme and its associated grapheme, obtained from the test "From Phoneme to Grapheme". The y-axis indicates the percentage of letters guessed at the first attempt for the two groups (control group and robot group). Asterisks denote statistically significant differences between the mean rank of population found with a Wilcoxon signed rank test (*$p < 0.05$)

Additional analysis

The primary goal of this study was to explore how haptic enabled robots could be used to enhance the quality of the letter representation in the learner's mind. Nonetheless, it is also interesting to explore if the same robots could also affects other skills of handwriting; more specifically, whether they could have a positive impact on the fine motor skills. Even though it is difficult to extract an handwriting feature linked with the quality of the motor control skill, previous research suggests that the average handwriting speed as well as the overall letter legibility can be correlated with the handwriting fluency and in particular the fine motor skill controlling the handwriting production [62, 138].

Handwriting Speed

Figure 10.9 presents the evolution of the average handwriting speed for the robot and the control group. We can see a statistically significant increase of this measure ($W(36) = 25.0, p = 0.0261$) at the end of the two teaching sessions (between the pre-test and the post-test) for the robot group. It is also interesting to notice that the average speed also increases for the letters not taught during the learning session ($W(36) = 26.0, p = 0.0299$).

A plausible explanation concerning the progress in the letters not taught could be that the learner might be able to transfer the skill acquired concerning the fluidity of the gesture from one letter to another, as any letter is a recomposition of well-known shapes and strokes that are shared by many of the letters of the alphabet [139].

No significant improvement related to the handwriting speed was found for the control group after the two teaching sessions with traditional methods. A Mann-Whitney U test was run to compare the difference between the learning gain of the two groups (between the post-test and the pre-test). This test reported a marginally significant difference between the learning gain of the two groups ($U(32) = 2390.0, p = 0.0518$).

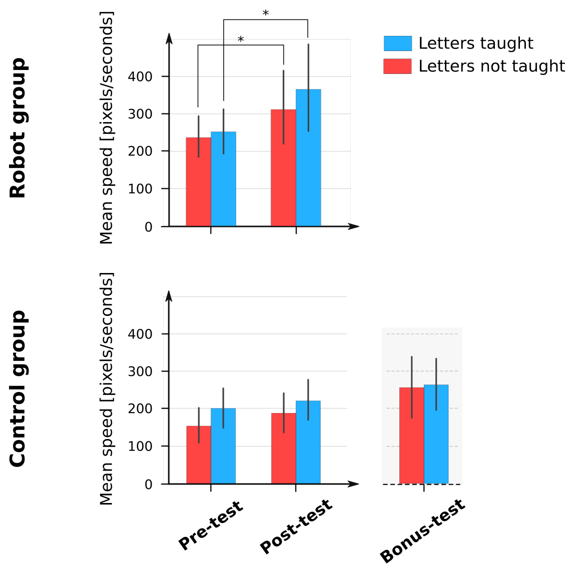

Figure 10.9: The average handwriting speed extracted from the test "From Phoneme to Grapheme-Ductus". Asterisks denote statistically significant differences between the mean of different groups found with a Wilcoxon signed rank test ($*p < 0.05$)

Handwriting legibility

In order to assess the letter's handwriting legibility, 4 experts were asked to grade every child's letter's grapheme between 0 (for totally wrong grapheme) and 3 (perfect grapheme).

Marginally significant progress could be observed for the handwriting legibility of the letters taught in the robot group ($W(36) = 10.5, p = 0.091$) but not for the letters not taught. Concerning the control group, no difference could be observed after the two teaching sessions with the traditional methods. However, after the last teaching session with robots (between the post-test and the bonus-test), marginally significant progress ($W(36) = 5.0, p = 0.067$) could be observed both for the letters taught and not taught.

As in the case of the mean handwriting speed, these results give us reason to believe that our robots may bring an added value when incorporated in a typical school environment. Again, more work will be needed to confirm this hypothesis. We believe that the haptic assistive feedback allowed by the use of the robot is responsible of the results (improvement in the handwriting legibility and speed) observed here and presents a possible betterment compared to traditional methods. This assumption appears to be in line with the literature investigating the contribution of haptic on learning [140, 54].

Conclusion and Discussion

The objective of this study was to explore how the haptic sensory modality activated by tangible robots can be used for the learning of the ductus and grapheme representation of letters, which requires mastering visual perception and visual-motor coordination, two important concept in handwriting acquisition.

Tests evaluating children performance were carefully designed in order to extract the child's knowledge in the specific aspects of their letter representation (phoneme-grapheme, phoneme-ductus and grapheme-ductus link) that we aimed to enhance.

Despite the limited number of trials and participants, improvements concerning the letter representation could be observed. The tests showed that the use of a haptic device allows children to enhance their representation of the ductus of the letters in a fast and effective way (link between grapheme and ductus). We could also see that combining the two teaching methods may be beneficial in the learning of letters' visual perception (link between phoneme and grapheme) as well as in the ductus representation (link from phoneme to grapheme-ductus). However, it would be to premature to claim that similar performance cannot be obtained using only traditional methods. For the moment, we can only say that using robots can be at least as effective as traditional methods to teach the concepts presented in this study. We feel that the use of robots to complement traditional methods helps to cross ceilings/barriers that are encountered by some children and could help overcome stagnation for them. More research will need to be performed in the future to investigate the assumption raised here.

Interestingly, significant results could be observed concerning the fine motor skills and in particular in the mean handwriting speed as well as in the letter legibility. These results were observed to be transfered from the letters taught to the ones not taught, which we hypothesise to be caused by the common strokes shared across letters in the Latin alphabet.

Results of this study are clearly limited by the small number of subjects (17 children) involved in our experiment. In order to confirm the results observed here, another experiment involving a higher number of student would need to be conducted.

Another limitation comes with the design of our experiment. Since we could not run the bonus test with both groups, the control and the robot groups differ in their number of learning sessions and the modality of learning (different teaching methods). This prevents from clearly isolating the factor that is responsible for the learning rate differences. Future work will be done in order to more deeply investigate the causes leading to the progress shown through this study.

Finally, the introduction of these kinds of robots in the classroom is typically affected by the novelty effect, which might result in an increase of engagement from the children which might result in bias the test outcomes. It is certainly interesting to run additional experiments in a long term study to test this effect.

Synthesis Part IV

11 Main findings

Overview

Even with correct training, up to 25% of children never master handwriting like their peers. While research shows a correlation between handwriting difficulties and school failure, these difficulties can also impact children in their self esteem and behavioral development. Since the mastery of handwriting requires a lot of different skills, it is never easy to understand where a given child is facing difficulties, nor how exactly to help him/her overcome them. The two problems, the analysis of handwriting, difficulties and their remediation, albeit related, can be treated independently and correspond respectively to the Part 2 and Part 3 of this thesis.

In Part 2, we tried to modernize the currently adopted handwriting tests (with a specific reference to the de-facto standard BHK), that show clear limitations in the era of digitalization. Indeed, the nature itself of these tests, conducted on paper, restricts them to the analysis of the final static aspect of handwriting. Its dynamics, found to be very important as seen in Chapter 3, is therefore hidden and cannot be taken into consideration. For this reason, we designed in collaboration with therapists several features that describe different aspects of handwriting, which are not limited to static but also capture kinematic, pressure and tilt. Using these features as input of a Random Forest model, we were able to detect dysgraphia with a remarkable accuracy and a considerably shorter time with respect to the BHK test (see Chapter 4). The designed features have the main advantage to describe very low level aspects of handwriting, which makes them quite independent of the writing content. In Chapter 5, we verified this hypothesis by giving the proof of concept that our model for automatic detection of handwriting difficulties can be translated from the latin to the the cyrillic alphabet. In the same way, we demonstrated in Chapter 6 that our model can also, given retraining, be used on paper or directly on digital tablets, like iPads. Finally, in the last Chapter of the analysis part, we introduced our iPad-based test allowing to extract the handwriting multidimensional handwriting profile of the child. This test aims to answer the first of the afore-discussed problems, by allowing for extracting the specific strengths and weaknesses of a child, in less than a minute, on different aspects and at different granularities.

Part 3 of this thesis tackles the problem of designing remediation activities for handwriting difficulties. In Chapter 8 we designed activities specifically targeting the handwriting aspects identified by the model developed in Chapter 7 and obtained a preliminary proof of concept that serious games targeting specific skills of handwriting (e.g. pressure, kinematic, tilt, ...) can have a positive impact on the overall quality of handwriting. The following chapters tackle two corollary, but still crucial questions related to handwriting remediation. As we saw, the remediation activities discussed in Chapter 8 are integrated in a Child-Robot Interaction scenario, in which the child is the teacher of the robot, and the reported study highlighted the importance of the child's perception of the robot and the interaction with it. In Chapter 9 we gave a proof of concept of the importance of the design of robot behaviors towards social acceptance with children, by specifically addressing adaptor movements during idle moments. Finally, in Chapter 10 we investigated whether it is possible to "remediate some handwriting difficulties by preventing them", i.e., by supporting pre-school children in the acquisition of the fundamental visuo-motor coordination skills required by handwriting.

Scientific contributions

The main scientific contribution of this thesis is undeniably the handwriting analysis part. The model presented in Chapter 3 allowed us to show the importance of the dynamics of handwriting in the detection of handwriting difficulties. We showed that its incorporation in machine learning models was associated with a significant increase of performance. In particular, we showed that the dynamics, by bringing insights on "how the letter is produced", help our model to distinguish visually similar letters.

In Chapter 4, we presented a first version of a model allowing to detect handwriting difficulties based on low level features. In addition, this work allowed us to put in light aspects of handwriting previously unexplored and explain them from a clinical point of view that may bring new insight for therapists within their sessions with children. Indeed, one of our goals was to ensure the interpretability of the model, using our machine learning models to reveal the aspects of handwriting that are important to explain handwriting difficulties, in a manner which is understandable by human practitioners. The fact that the kinematic and pressure features turned our to be the most important to explain handwriting difficulties is, in our opinion, an important result that shows that the current tests used to detect handwriting difficulties need to be modernized.

In Chapter 5 and 6, we showed another advantage that the low level features on which our models are built upon can bring. Indeed, since they describe very low level aspects of handwriting, that make them independent of the writing content, we could show through two statistical tests that: (1) while there are absolute differences in the feature values when children are changing alphabet or support, (2) the specificity of handwriting is preserved. In other words, a child one sigma away from average for one feature when writing in a given alphabet will also be one sigma away from average for the same feature when writing in another alphabet (or when switching support). Of course, it would be to premature to extend this result to all alphabets or supports, but we can reasonably believe that this result stands between languages that share the same roots (for example between children writing in French or English). To further investigate this point, we are currently expanding the set of alphabets we tested our model on, namely considering the Chinese alphabet.

In Chapter 7, further enhancing the interpretability of its outcomes, we extend the model presented in Chapter 4, giving more interpretability and becoming totally independent from any handwriting test currently existing. Indeed, by combining the features designed together with therapists with unsupervised learning method, we designed a data-driven approach to find the handwriting aspects best explaining the difference in handwriting between children. Contrary to the analysis conducted in Chapter 4, this model allows us to extract information about handwriting difficulty themselves, abandoning any initial reference to the BHK test. Specifically, the feature importance computed by the PCA analysis of Chapter 7 captures information about handwriting itself, and what creates the differences between one's handwriting and another's, thus allowing for an analysis of a person's handwriting quality which is independent from any human-based handwriting quality assessment. We believe that this is an important contribution since it can be used by the therapists to better understand their patients and also by the scientific community for even broader applications. For instance, our contribution raised significant interest well beyond the world of handwriting difficulties diagnosis and remediation, from researchers interested in using our model to diagnose aphasia, hemiplegia or autism for example. In view of the high interest obtained from therapists, school directors, teachers and parents when our articles were published, we decided to exploit the afore-discussed scientific contribution as a commercial product as presented in the next and last Chapter of this thesis. Finally, it is important to notice that the system described in this part solely focuses on motor-related handwriting difficulties, which represent a portion of all handwriting difficulties. For example, the overlap of dysgraphia and dyslexia (severe problem with reading), i.e., the set of handwriting difficulties which are caused by or related to reading difficulties, is not covered by our system. If we want to extend the type and number of handwriting difficulties detected by our system, additional work will need to be carried out.

Concerning the remediation part, Chapter 8 presents the study in which we tested for the first time our system involving remediation games targeting specific aspect of handwriting (e.g. pressure, kinematic, tilt, ...) within a long term study of 9 months. In addition to giving the proof of concept that the extension of the Co-Writer activity can be successfully used in a long term context, we wanted to test whether working on subskills of handwriting (and not only on handwriting directly), could have a positive impact on handwriting itself. In the same way, this study allowed to test whether an amelioration of handwriting skills on a digital tablet (with a different friction coefficient) brings an amelioration on paper. The results extracted with the BHK tests, conducted (on paper) at regular intervals during the 9 months of the study, show a spectacular improvement of the handwriting both in terms of quality and speed. This result therefore gives us reasons to believe that the two above questions can be answered positively. Of course, this study is exploratory (only one child involved) and was aimed at refining the hypotheses, rather than providing conclusive evidence about them. Additional work will need to be conducted in order to give a more solid answer to the hypotheses raised. As explained in the Introduction, children facing handwriting difficulties are often stuck in a vicious circle where their difficulties generate avoidance of written tasks that will in turn increase handwriting difficulties. For this reason, it can be interesting to use methods aiming to re-motivate the child and make him re-gain confidence towards handwriting like in the Child-Robot Interaction presented in Chapter 8. In Chapter 9, we show the importance of the design of robot behaviors towards social acceptance with children, important condition to increase the quality of the interaction between the child and the robot.

Finally we show in the last Chapter that we can exploit some capabilities that robots can offer in order to teach some handwriting's specific sub-skills (i.e. visuo-motor coordination) at a preschool level, and therefore support children in the first steps of their handwriting acquisition and hopefully prevent handwriting difficulties to arise later in their school curriculum.

12 Current and Future Work

One of our main and ongoing efforts is towards the extension of the proposed model for the automatic detection of handwriting difficulties to other languages and alphabets. Following the analysis conducted in Chapter 5, we are currently exploring the possibility to transfer the model from the French to the German language as well as trying to assess similarities between the Chinese and Latin alphabets.

In line with the "early remediation" principle put forth in Chapter 10, we are trying to extend the model to drawings. This will have mainly two advantages: with the first one being a potential earlier detection of handwriting difficulties since children generally start drawing before writing. This early detection might then allow to remediate handwriting difficulties even before they have an impact on the child's curriculum. The second advantage that drawings bring with respect to letters and words is that children facing handwriting difficulties are often in a situation of such failure that they refuse any kind of handwriting activities. Being able to propose a test based on drawing, where the child won't feel pressured nor judged, would be undeniably a great added value. That being said, it appears that this task might be quite complex since, as seen in Chapter 2, the act of writing and drawing differ in many ways. In particular, drawing does not relate to reading, while current handwriting tests (including the one we propose) rely on a written text that the child has to copy and therefore depend on the child's reading skills. Indeed, it won't be surprising to find a sample of population presenting handwriting difficulties but no drawing difficulties, corresponding to the dyslexic dysgraphic children population according to the classification proposed by Deuel et al. [141].

Another interesting line of research would be to use our model, not to detect handwriting difficulties but other types of problems such as aphasia, emiplegia or autism. As mentioned in the previous chapter, we were lucky to generate an important interest of different research groups and medical specialists after our scientific articles got published. Among the different research lines being set up, machine learning models using our features as inputs can be trained to detect these diseases. In a different way, we could use the model presented in Chapter 7 to project the children under investigation (for example children with emiplegia) into our three (or more) dimensional space (represented by the PCA main components) to explore if these children are clustered together into a certain region of this space. If this is the case, reverse engineering of the model may lead to interesting discoveries about these conditions.

As mentioned in the previous Chapter, we are currently creating a startup to translate the research presented in this thesis into a commercially available product. This startup is developing an application, called Dynamico, exhibiting the following functionality:

- A fully automated, detailed handwriting analysis, integrating the model developed in Chapter 7.
- A set of target remediation activities, such as the ones we saw in Chapter 8
- A monitoring dashboard, to follow the progress of children.

A complete redesign of the activities introduced in Chapter 8 has been conducted and are now being integrated on iPads. In the same way, we are currently designing new activities in order to have a bigger portfolio to propose. Examples of the redesigned activities can be seen in Figure 12.1.

Figure 12.1: Left: the pressure activity, Right: the kinematic activity.

A monitoring dashboard has been designed allowing therapists, school teachers or parents (with a customized view for each of these categories) to analyze the results of children. As an example, Figure 12.2 shows on the left the different scores introduced in Chapter 7, computed for a child at a given moment in time, and, on the right, their evolution as the child interacts with the system for multiple weeks. Of course, the design of this dashboard is meant to be adapted to the public using the application, a school teacher needing different information than a therapist or a parent.

The structure of the application is built upon the integration of the analysis with the remediation activities. With data coming from both sides being available (i.e., the scores

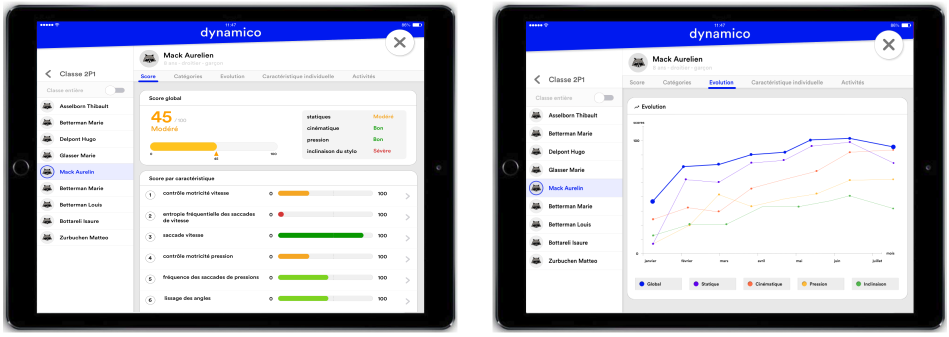

Figure 12.2: Left: the scores dashboard, Right: the scores evolution dashboard.

computed by the analysis and the different performance metrics extracted from the activities), our next objective will be to train a recommendation algorithm suggesting activities to the user. Indeed, if we know the handwriting state of the child (described by the scores) at t0, its new state at t1, as well as all the activities he/she did in the middle, given enough data, we could build a recommendation algorithm that would learn what activities are beneficial for a given handwriting profile. In that sense, it would be necessary at some point to conduct a large-scale validation of the system and its efficacy. Indeed, while the analysis part of the system builds upon solid scientific grounding, the remediation part needs further evaluation before the application can be released on the market. Additionally, future plans also include to integrate in the Dynamico application the research conducted to extend the model to different languages and alphabets in order to increase the impact our solution may have on society. Finally, our work mainly focuses on the motor aspects of handwriting, which constitute only a portion of all handwriting difficulties, which, in turn, constitutes only a small fraction (around 20%) of the general learning difficulties that affect children. While the road towards effectively detecting and remediating general learning difficulties in children is still long and obscure, this work sheds a light on one of them and will hopefully help improve the lives of many children, by making them capable and happy of expressing their thoughts and opinions in a written form.

A Pressure Calibration between two tablets

Two different tablets were used for the database in Chapter 4: A wacom Intuos 4 for the data collection in schools and a wacom Intuos 3 for the data collection in hospitals. Since the difference of hardware might bring variation in the way data were recorded that might bias the model trained, a carefull calibration was made between the two devices. No difference were found in term of absolute position (x, y) as well as in term of tilt. However, small difference were recorded concerning the pressure measurement.

Following the advice of the tablet manufacturer (Wacom Co., Ltd), a pressure calibration was done for the two tablets. A construction was made to position the pen vertically on the surface of the tablets with minimal friction. 15 different weights (called X) from 0g (pen without load) to 400g (saturation of both tablets) were used as an input while the values returned by each tablets (called Y) were logged. We then extracted the relation X/Y for the two tablets which ended up being very similar (the Spearman correlation shows a correlation of 99.15% (p = 5.32e-12), mean square error of 0.6). A 4th degree polynomial fit was then created to model the function describing the X/Y relation of the first tablet and used on the input of the second in order to rectify its output. After this correction, the Spearman correlation was found to be 99.998% (p = 1.81e-21) and the mean squared error was 5.1e-3

B Similarity of age distribution between dysgraphic and non dysgraphic children

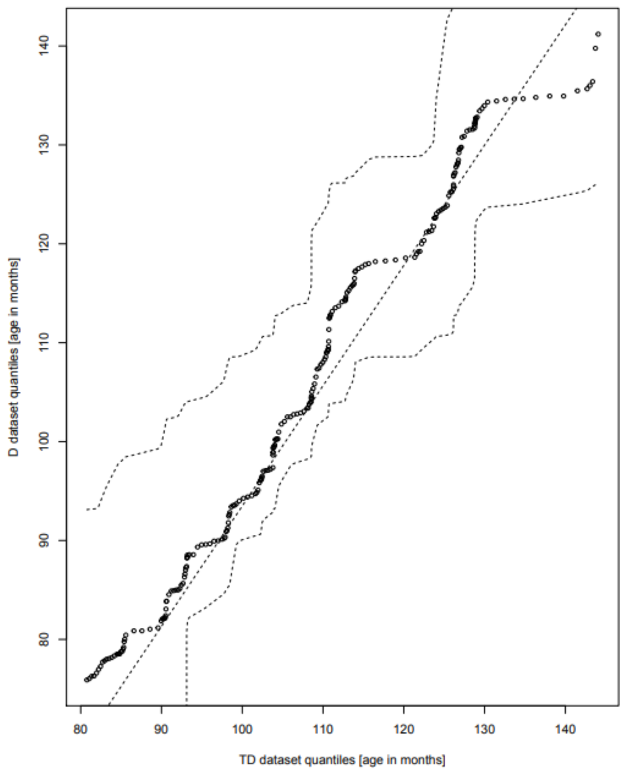

Figure B.1: Quantiles of the TD dataset (x-axis) against the quantiles of the D dataset (y-axis). The points are closed from the diagonal dashed line and included in the confidence bounds, showing the similarity of the two distributions.

C Features Engineering

Every feature used in the analysis are described below:

Static features

(1) The *Handwriting Moment.* To compute this feature, we extracted bins of 300 points (from the same line of text) and computed their barycenters. The distance in the y direction between consecutive barycenters was computed and averaged for all of the points, reflecting the degree of straightness of the line of text.

(2) The *Handwriting Size.* To compute this feature, we extracted bins of 300 points (from the same line of text) and computed the total surface occupied by the box bounding the trace.

(3) The *Space Between Strokes.* This feature refers to the distance between strokes.

(4) The *Handwriting Density.* A grid with 20-pixel cells covering the entire range of the handwriting trace is created, as can be seen in Figure C.1. The number of points recorded by the iPad in each cell, if present, is then stored in an array. The mean of this array is represented by this feature. There is a positive correlation between handwriting density and handwriting quality, as handwriting becomes denser with age.

(5) The *Average Stoke Direction.* For two consecutive points within a stroke, the direction of movement is calculated as the arctan of the ratio dy/dx. We can obtain the average stroke direction by finding the average of these two directions.

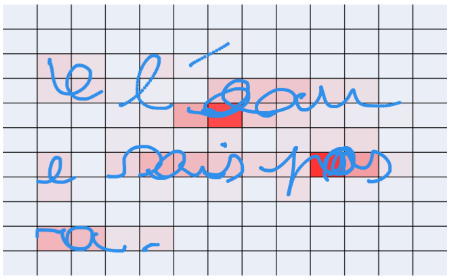

Figure C.1: Illustration of the Handwriting density. The space is split into 20-pixel-wide square cells. Using the number of points per cell, we can then compute a density.

(6) The *Angle Smoothness*. To compute this feature, we first need to calculate the absolute angle between each set of three consecutive points in an handwriting sample. This feature is then obtain by averaging all the angles calculated and measures the smoothness of the handwriting (a high value for this feature relates to abrupt angles).

(7) The *Area of the Text's Convex Hull*. To compute this feature, we first need to compute the convex hull of the text, which is computed using Graham's algorithm [**?**]. We can then compute the area of the convex hull surrounding the text.

(8), (9), (10), (11) & (12) The *Bandwidth of the Power Spectral of Tremor Frequencies*, the *Median of the Power Spectral of Tremor Frequencies*, the *Entropy of Mean of Tremor Frequencies*, the *Correlation of Mean of Tremor Frequencies* and the *Distance of Mean of Tremor Frequencies*. These features describe shaky handwriting. For every child, the following process was followed: the signal was first divided into bins of 600 points (as highlighted in Figure C.2). For each bin, we extracted the deviance from the handwriting path using two types of vectors: the first one was a vector representing the global handwriting direction (computed with 10 points, as can be seen in green in Figure C.2), while the second was a local vector computed with two consecutive points (in blue in Figure C.2). We then computed the cross product of these two vectors to calculate how orthogonal the local vectors were with the global vector. The greater this cross product is, the higher the deviance from the handwriting path. We hypothesized that shaky handwriting would result in local vectors that were poorly aligned with the global vectors. Making it possible to detect shaky handwriting with this method.

For each of these 600 points, we saved the norm of the cross product and computed the Fourier transform on this vector. Then we averaged all the Fourier transforms extracted from all the different 600-point bins (see Figure C.2). A high frequency or a wide bandwidth in the spectral distribution would be an indication of shaky handwriting.

For instance, we computed the range of frequencies covering 90% of the spectral density.

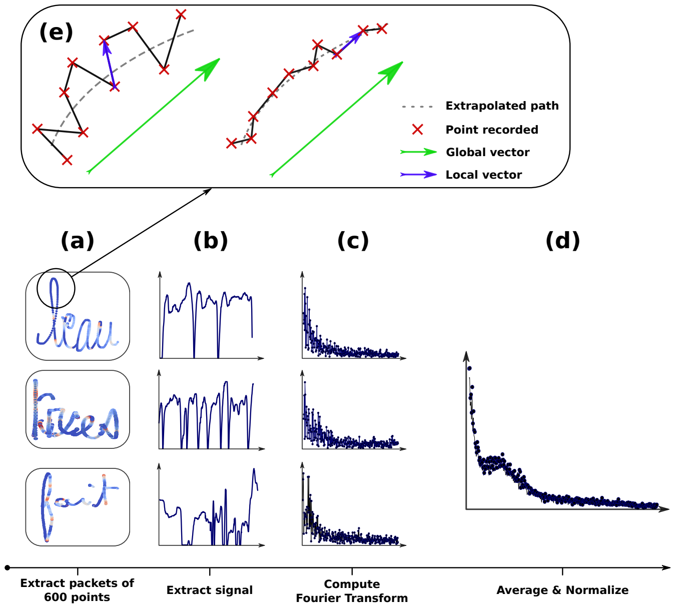

Figure C.2: The whole process used to extract the frequency spectrum of our signal. (a) We first divided the BHK text into bins of 600 points. (b) For each packet, the signal was extracted. (c) We computed the Fourier transform of the signal. (d) We took the average of all signals and performed a normalization. (e) In these **sample** signals extracted from the data, the red dots are the point coordinates recorded by the device during handwriting, the vectors in blue are "local" vectors linking two consecutive points and the vector in green is the "global" vector (average of the nine blue vectors) representing the global direction of the handwriting. The cross product of these two vectors gives us an indication of the smoothness/shakiness of the handwriting. The right side of the figure comes from a writer with smoother/less shaky handwriting than the writer producing the one on the left, and the cross product operation will detect this difference. This image has been adapted from [3].

The smaller this value is (corresponding to a more clustered distribution), the more proficient the writer is. This feature is called the (8) *Bandwidth of the Power Spectral of Tremor Frequencies*. We also extracted the median of the power spectral density. A large value for this feature indicates the presence of high frequencies. This feature is called the (9) *Median of the Power Spectral of Tremor Frequencies*.

The last feature defined in this context is the entropy between the spectral distribution of the writer and the averaged spectral distribution of all the writers in our database. The higher this entropy is, the more different the handwriting of this particular writer is. This feature is called (10) *Entropy of Mean of Tremor Frequencies*. We also computed the correlation between the two signals, called the (11) *Correlation of Mean of Tremor Frequencies*, as well as the distance between them, called the (12) *Distance of Mean of Tremor Frequencies*.

Kinematic Features

(13), (14) & (15) The *Mean Velocity*, *Maximum Velocity* and *Standard Deviation of Velocity*. These features quantify handwriting speed, where the speed is the distance traveled by the pen divided by the time taken to write out the passage. Research shows that children presenting handwriting difficulties have lower mean velocities as well as higher maximum velocities. Furthermore, the mean velocity increases with age.

(16) The *Increase in Velocity*. Using the handwriting speed over time, we performed a linear regression to compute the evolution of this handwriting speed.

(17) The *Nb of Peaks of Velocity Change Per Second*. Motivated by insights from clinicians, we applied a Gaussian filter to the velocity signal over time (using the scipy library, method *general gaussian* with $M = 3, p = 0.5, sigma = 2$) and computed the number of local maximum and minimum that were extracted. We expected that the number of changes would grow with the total duration of the test and, therefore, we normalize this number with time.

(18), (19), (20), (21) & (22) The *Bandwidth of the Power Spectral of Speed Frequencies*, the *Median of the Power Spectral of Speed Frequencies*, the *Entropy of Mean of Speed Frequencies*, the *Correlation of Mean of Speed Frequencies* and the *Distance of Mean of Speed Frequencies*. Handwriting can be interpreted as a two-dimensional time series. As in the (19) *Median of the Power Spectral of Tremor Frequencies*, a Fourier transform can be calculated with the handwriting velocity, median and resulting bandwidth of the spectral distribution. We can observe very rapid changes in speed in the handwriting of children with dysgraphia (some jerks resulting from a low level of handwriting automation). These abnormal changes in speed are translated into high frequencies in the Fourier transform, resulting in a shift in the median towards higher frequencies. Children with a low level of automation change also use variable speeds in their writing. Hence, a writer presenting a high bandwidth will not be fluent, as they are less consistent in their movements. The (20) *Entropy of Mean of Speed Frequencies* is the entropy between the spectral distribution of the writer and the average spectral distribution of all the writers in our database. The higher this entropy is, the more eclectic the handwriting of this particular writer is. We also computed the (21) *Correlation of Mean of Speed Frequencies*, which is the correlation between the spectral distribution of the writer and the average spectral distribution of all the writers in our database, as well as the (22) *Distance of Mean of Speed Frequencies*.

(23) The *In-Air-Time Ratio* represents the proportion of time the writer spends without touching the surface of the tablet. This feature has been shown to be positively correlated

with handwriting problems [3, 100, 83].

Pressure Features

(24), (25) & (26) The *Mean Pressure*, *Maximum Pressure* and *Standard Deviation of Pressure*. These first features concerning pressure are simply the mean, maximum, and standard deviation of the pressure.

(27), (28) & (29) The *Mean Speed of Pressure Change*, *Max Speed of Pressure Change* and *Standard Deviation of Speed of Pressure Change* are extracted by working with averaged bins of 10 recorded points of pressure and dividing the time spent by the difference between two averaged bins of points. These features are then computed by finding the mean, maximum and standard deviation of all measurements.

(30) The *Increase of Speed of Pressure Change*. Using the changes in the Speed of Pressure change over time, we performed a linear regression to track its evolution.

(31) The *Nb of Peaks of Pressure Change Per Second* computes the number of pressure inversions per second.

(32), (33), (34), (35) & (36) The *Bandwidth of the Power Spectral of Speed of Pressure Change Frequencies*, the *Median of the Power Spectral of Speed of Pressure Change Frequencies*, the *Entropy of Mean of Speed of Pressure Change Frequencies*, the *Correlation of Mean of Speed of Pressure Change Frequencies* and the *Distance of Mean of Speed of Pressure Change Frequencies*. The speed of pressure change can be seen as a time series, and frequencies can be extracted using a Fourier transform. The same process as that described in Figure C.2 is followed to extract these five features.

Tilt Features

The iPad system measures the pen tilt with two different angles referred to as the **Tilt-Azimuth** angle and the **Tilt-Altitude** angle (see Figure C.3 for additional information). Both angles are between 0 to 180 degrees.

It has to be noted that the Wacom tablet measures the pen tilt differently. The Wacom system logged the data measuring the pen tilt with two different angles, which we referred to in this paper as the Tilt-x and Tilt-y angles (as can be seen in Figure C.4). Both angles are measured in the range between -60 and 60 degrees. The tilt-x reflects the inclination of the pen in the direction of the written line, and the tilt-y reflects the

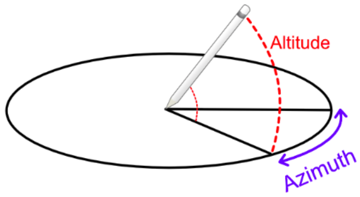

Figure C.3: The two angles (Tilt-Azimuth and Tilt-Altitude) recorded for the pen with the iPad.

inclination of the pen below the written line. Thus, for the chapters were the data were recorded with Wacom tablets (Chapters 3, 4, 5 and 6), the denomination Tilt-Azimuth should be replaced by Tilt-x while the denomination Tilt-Altitude should be replaced by Tilt-y.

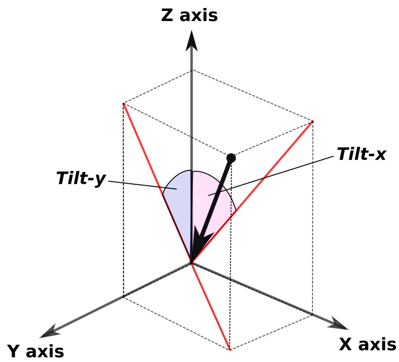

Figure C.4: The two angles (Tilt-x and Tilt-y) recorded for the pen with the Wacom.

(38), (39) & (40) The *Mean Tilt-Azimuth*, *Maximum Tilt-Azimuth* and *Standard Deviation of Tilt-Azimuth*. These first features concerning the Tilt-Azimuth are simply its mean, maximum and standard deviation.

(41), (42) & (43) The *Mean Tilt-Altitude*, *Maximum Tilt-Altitude* and *Standard Deviation of Tilt-Altitude*. These first features concerning the Tilt-Altitude are simply its mean, maximum and standard deviation.

(44), (45) & (46) The *Mean Speed of Tilt-Azimuth Change*, *Max Speed of Tilt-Azimuth Change* and *Standard Deviation of Speed of Tilt-Azimuth Change* are extracted by working with averaged bins of 10 recorded Tilt-Azimuth points and dividing the time spent by the difference between two averaged bins of points. These features are then computed by finding the mean, maximum and standard deviation of all measurements.

(47), (48) & (49) The *Mean Speed of Tilt-Altitude Change*, *Max Speed of Tilt-Altitude*

Change and *Standard deviation of Speed of Tilt-Altitude Change* are extracted by working with averaged bins of 10 recorded points of Tilt-Altitude and dividing the time spent by the difference between two averaged bins of points. These features are then computed by finding the mean, maximum and standard deviation of all measurements.

(50), (51) The *Increase of Speed of Tilt-Azimuth Change*, *Increase of Speed of Tilt-Altitude Change*. Using the changes in the Speed of Tilt-Azimuth (Tilt-Altitude) over time, we performed a linear regression to track its evolution.

(52), (53) The *Nb of Peaks of Tilt-Azimuth Change Per Second* and *Nb of Peaks of Tilt-Altitude Change Per Second*. This feature computes the number of Tilt-Azimuth (Tilt-Altitude) inversions per second.

(54), (55), (56), (57) & (58) The *Bandwidth of the Power Spectral of Speed of Tilt-Azimuth Change Frequencies*, the *Median of the Power Spectral of Speed of Tilt-Azimuth Change Frequencies*, the *Entropy of Mean of Speed of Tilt-Azimuth Change Frequencies*, the *Correlation of Mean of Speed of Tilt-Azimuth Change Frequencies* and the *Distance of Mean of Speed of Tilt-Azimuth Change Frequencies*. The speed of Tilt-Azimuth change can be seen as a time series, and frequencies can be extracted using a Fourier transform. The same process as that described in Figure C.2 is followed to extract these five features.

(59), (60), (61), (62) & (63) The *Bandwidth of the Power Spectral of Speed of Tilt-Altitude Change Frequencies*, the *Median of the Power Spectral of Speed of Tilt-Altitude Change Frequencies*, the *Entropy of Mean of Speed of Tilt-Altitude Change Frequencies*, the *Correlation of Mean of Speed of Tilt-Altitude Change Frequencies* and the *Distance of Mean of Speed of Tilt-Altitude Change Frequencies*. The speed of Tilt-Altitude change can be seen as a time series, and frequencies can be extracted using a Fourier transform. The same process as that described in Figure C.2 is followed to extract these five features.

D Features means and standard deviations for the two alphabets

Feature		Grade 1		Grade 2		Grade 3		Grade 4	
		Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin	Cyrillic	Latin
Bandwidth Tremolo ***1e-2**	Static	2.53±0.07	2.56±0.06	2.53±0.04	2.55±0.05	2.54±0.04	2.56±0.04	2.54±0.03	2.54±0.04
Median Tremolo ***1e3**		2.73±0.12	2.83±0.17	2.68±0.11	2.76±0.13	2.71±0.12	2.76±0.13	2.68±0.12	2.71±0.15
Space Between Words ***1e3**		5.09±4.48	3.54±2.54	5.17±3.64	3.26±2.46	3.99±2.21	2.65±1.46	4.73±3.59	3.68±2.36
Handwriting Moment ***1e4**		5.83±8.26	14.2±12.3	7.89±10.7	12.3±12.3	8.03±10.1	10.2±11.0	7.28±7.05	9.50±8.53
Handwriting Density ***1e-2**		1.33±0.80	1.31±0.95	1.33±0.68	1.34±0.72	1.15±0.59	1.27±0.54	1.04±0.38	1.19±0.60
Mean Velocity	Kinematics	1.98±0.93	1.94±1.07	2.17±1.07	2.00±1.06	2.31±0.78	2.06±0.73	2.41±0.66	2.14±0.75
Max Velocity ***1e-1**		1.98±0.83	1.49±0.88	1.95±0.72	1.44±0.67	1.84±0.61	1.33±0.43	2.01±0.62	1.57±0.56
In-Air-Time Ratio ***1e1**		2.73±1.46	3.87±1.58	3.06±1.46	3.06±1.48	2.61±1.39	2.69±1.51	2.99±1.32	3.23±1.50
Bandwidth Speed ***1e-2**		2.62±0.03	2.63±0.04	2.62±0.02	2.61±0.03	2.62±0.03	2.62±0.03	2.62±0.02	2.61±0.03
Median Speed ***1e3**		2.58±0.13	2.69±0.28	2.56±0.10	2.55±0.19	2.58±0.10	2.59±0.24	2.57±0.09	2.54±0.18
Mean Pressure ***1e1**	Pressure	3.93±1.28	3.73±1.28	3.66±1.07	3.63±0.97	3.65±1.14	3.62±0.90	3.58±1.18	3.48±1.00
Mean Speed of Pressure Change ***1e1**		1.96±0.86	2.42±1.48	1.98±0.87	2.64±1.03	2.17±0.82	2.73±1.07	2.01±0.84	2.70±1.07
Max Speed of Pressure Change		3.78±1.35	3.73±1.76	3.85±1.38	4.36±1.22	4.09±1.26	4.75±1.38	4.32±1.48	4.99±1.60
nb Peaks of Pressure Change per secs		2.34±0.53	1.37±0.52	2.65±0.53	1.79±0.56	2.78±0.68	2.03±0.64	3.06±0.60	2.31±0.66
Median Pressure ***1e3**		2.54±0.09	2.70±0.18	2.56±0.10	2.63±0.11	2.59±0.09	2.66±0.15	2.54±0.11	2.61±0.14
Std tilt-X ***1e2**	Tilt	1.98±6.26	1.75±7.74	1.94±7.35	1.77±7.51	1.97±5.47	1.74±5.38	1.95±6.70	1.68±5.67
Std Speed of tilt-X change ***1e1**		1.98±0.63	1.75±0.77	1.94±0.73	1.77±0.75	1.97±0.55	1.74±0.54	1.95±0.67	1.68±0.57
Median tilt-Y ***1e3**		9.40±0.35	9.31±0.50	9.45±0.29	9.47±0.39	9.56±0.28	9.43±0.37	9.63±0.23	9.54±0.29

Table D.1: Features means and standard deviations for the two alphabets (Cyrillic and Latin). For visibility purposes, some features values have been multiplied by the factor next the feature name.

Bibliography

[1] S. Pinto and M. Sato, *Traité de Neurolinguistique: Du cerveau au langage*. De Boeck Superieur, 2016.

[2] L. P. Ayres, *A scale for measuring the quality of handwriting of school children*. No. 113, Russell Sage Foundation. Dept. of Child Hygiene, 1912.

[3] T. Asselborn, T. Gargot, Ł. Kidziński, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," *npj Digital Medicine*, vol. 1, no. 1, p. 42, 2018.

[4] L. Hamstra-Bletz and A. W. Blöte, "A longitudinal study on dysgraphic handwriting in primary school," *Journal of Learning Disabilities*, vol. 26, no. 10, pp. 689–699, 1993.

[5] R. Karlsdottir and T. Stefansson, "Problems in developing functional handwriting," *Perceptual and motor skills*, vol. 94, no. 2, pp. 623–662, 2002.

[6] A. Overvelde and W. Hulstijn, "Handwriting development in grade 2 and grade 3 primary school children with normal, at risk, or dysgraphic characteristics," *Research in developmental disabilities*, vol. 32, no. 2, pp. 540–548, 2011.

[7] M. U. Bers, I. Ponte, C. Juelich, A. Viera, and J. Schenker, "Teachers as designers: Integrating robotics in early childhood education," *Information technology in childhood education annual*, vol. 2002, no. 1, pp. 123–145, 2002.

[8] A. Sullivan, E. R. Kazakoff, and M. U. Bers, "The wheels on the bot go round and round: Robotics curriculum in pre-kindergarten," *Journal of Information Technology Education*, vol. 12, pp. 203–219, 2013.

[9] T. Asselborn, A. Guneysu, K. Mrini, E. Yadollahi, A. Ozgur, W. Johal, and P. Dillenbourg, "Bringing letters to life: handwriting with haptic-enabled tangible robots," in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, pp. 219–230, ACM, 2018.

[10] A. Sullivan and M. U. Bers, “Gender differences in kindergarteners’ robotics and programming achievement,” *International journal of technology and design education*, vol. 23, no. 3, pp. 691–702, 2013.

[11] M. Kandlhofer, G. Steinbauer, S. Hirschmugl-Gaisch, and H. Eck, “Children discover science: robotics, informatics and artificial intelligence in kindergarten and school,” in *Workshop at International Conference on Robotics in Education*, 2015.

[12] F. Riedo, M. Chevalier, S. Magnenat, and F. Mondada, “Thymio ii, a robot that grows wiser with children,” in *2013 IEEE Workshop on Advanced Robotics and its Social Impacts*, pp. 187–193, IEEE, 2013.

[13] B. S. Barker and J. Ansorge, “Robotics as means to increase achievement scores in an informal learning environment,” *Journal of research on technology in education*, vol. 39, no. 3, pp. 229–243, 2007.

[14] F. R. Sullivan, “Robotics and science literacy: Thinking skills, science process skills and systems understanding,” *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, vol. 45, no. 3, pp. 373–394, 2008.

[15] K. A. Mills, V. Chandra, and J. Y. Park, “The architecture of children’s use of language and tools when problem solving collaboratively with robotics,” *The Australian Educational Researcher*, vol. 40, no. 3, pp. 315–337, 2013.

[16] D. J. Barnes, “Teaching introductory java through lego mindstorms models,” in *Proceedings of the 33rd SIGCSE technical symposium on Computer science education*, pp. 147–151, 2002.

[17] P. B. Lawhead, M. E. Duncan, C. G. Bland, M. Goldweber, M. Schep, D. J. Barnes, and R. G. Hollingsworth, “A road map for teaching introductory programming using lego© mindstorms robots,” *Acm sigcse bulletin*, vol. 35, no. 2, pp. 191–201, 2002.

[18] F. Klassner and S. D. Anderson, “Lego mindstorms: Not just for k-12 anymore,” *IEEE robotics & automation magazine*, vol. 10, no. 2, pp. 12–18, 2003.

[19] T. Balch, J. Summet, D. Blank, D. Kumar, M. Guzdial, K. O’hara, D. Walker, M. Sweat, G. Gupta, S. Tansley, *et al.*, “Designing personal robots for education: Hardware, software, and curriculum,” *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 5–9, 2008.

[20] E. Yadollahi, W. Johal, A. Paiva, and P. Dillenbourg, “When deictic gestures in a robot can harm child-robot collaboration,” in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, pp. 195–206, 2018.

[21] T. Belpaeme, J. Kennedy, P. Baxter, P. Vogt, E. E. Krahmer, S. Kopp, K. Bergmann, P. Leseman, A. C. Küntay, T. Göksun, *et al.*, "L2tor-second language tutoring using social robots," in *Proceedings of the ICSR 2015 WONDER Workshop*, 2015.

[22] A. Chemin, "Handwriting vs typing: is the pen still mightier than the keyboard?," 2014.

[23] H. Russell, "Signing off: Finnish schools phase out handwriting classes," 2015.

[24] K. H. James and L. Engelhardt, "The effects of handwriting experience on functional brain development in pre-literate children," *Trends in neuroscience and education*, vol. 1, no. 1, pp. 32–42, 2012.

[25] P. A. Mueller and D. M. Oppenheimer, "The pen is mightier than the keyboard: Advantages of longhand over laptop note taking," *Psychological science*, vol. 25, no. 6, pp. 1159–1168, 2014.

[26] S. Graham, "The role of production factors in learning disabled students' compositions.," *Journal of educational psychology*, vol. 82, no. 4, p. 781, 1990.

[27] V. W. Berninger, K. B. Vaughan, R. D. Abbott, S. P. Abbott, L. W. Rogan, A. Brooks, E. Reed, and S. Graham, "Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition.," *Journal of Educational Psychology*, vol. 89, no. 4, p. 652, 1997.

[28] B. Bourdin and M. Fayol, "Is written language production more difficult than oral language production? a working memory approach," *International journal of psychology*, vol. 29, no. 5, pp. 591–620, 1994.

[29] K. P. Feder and A. Majnemer, "Handwriting development, competency, and intervention," *Developmental Medicine & Child Neurology*, vol. 49, no. 4, pp. 312–317, 2007.

[30] C. A. Christensen, "The critical role handwriting plays in the ability to produce high-quality written text," *The SAGE handbook of writing development*, pp. 284–299, 2009.

[31] B. Bourdin and M. Fayol, "Is graphic activity cognitively costly? a developmental approach," *Reading and Writing*, vol. 13, no. 3, pp. 183–196, 2000.

[32] D. McCutchen, "From novice to expert: Implications of language skills and writing-relevant knowledge for memory during the development of writing skill," *Journal of Writing Research*, vol. 3, no. 1, pp. 51–68, 2011.

[33] K. Feder and A. Majnemer, "Handwriting development, competency, and intervention.," *Developmental Medicine and Child Neurology*, vol. 49, pp. 312–317, 2007.

[34] E. Chartrel and A. Vinter, "The impact of spatio-temporal constraints on cursive letter handwriting in children.," *Learning and Instruction*, vol. 18, no. 6, pp. 537–547, 2008.

[35] S. Rueckriegel, F. Blankenburg, R. Burghardt, S. Ehrlich, G. Henze, R. Mergl, and P. Hernaiz Driever, "Influence of age and movement complexity on kinematic hand movement parameters in childhood and adolescence.," *International Journal of Developmental Neuroscience*, vol. 26, no. 7, pp. 655–663, 2008.

[36] R. Sassoon, I. Nimmo-Smith, and A. Wing, "Developing efficiency in cursive handwriting: An analysis of "t" crossing behavior in children.," in *Computer recognition and human production of handwriting*, pp. 287–297, Singapore: Plamondon, R., Suen, C.Y., Simner, M.L., world scientific publisher ed., 1989.

[37] A. Vinter and E. Chartrel, "Effects of Different Types of Learning on Handwriting Movements in Young Children.," *Learning and Instruction*, vol. 20, pp. 476–486, 2010.

[38] J. Ziviani and M. Wallen, "The development of graphomotor skills.," in *Hand function in the child: Foundations for remediation (2nd edition)*, pp. 217–236, St. Louis, MO: A. Henderson & C. Pehoski, mosby ed., 2006.

[39] R. Kellogg, "Analyzing children's art. mountain view," 1970.

[40] A. Stetsenko, "The psychological function of children's drawing. a vygotskian perspective," *Drawing and looking. Hemel Hempstead: Prentice Hall/Harvester Wheatsheaf*, 1995.

[41] F. Bonoti, F. Vlachos, and P. Metallidou, "Writing and drawing performance of school age children: Is there any relationship?," *School Psychology International*, vol. 26, no. 2, pp. 243–255, 2005.

[42] G. V. Thomas and A. M. Silk, *An introduction to the psychology of children's drawings.* New York University Press, 1990.

[43] K. Brenneman, C. Massey, S. F. Machado, and R. Gelman, "Young children's plans differ for writing and drawing," *Cognitive Development*, vol. 11, no. 3, pp. 397–419, 1996.

[44] E. Adi-Japha and N. H. Freeman, "Development of differentiation between writing and drawing systems.," *Developmental psychology*, vol. 37, no. 1, p. 101, 2001.

[45] J. Goodnow, S. L. Friedman, M. Bernbaum, and E. B. Lehman, "Direction and sequence in copying: The effect of learning to write in english and hebrew," *Journal of Cross-Cultural Psychology*, vol. 4, no. 3, pp. 263–282, 1973.

[46] K. E. Beery, "Developmental test of visual-motor integration," *Administration, scoring and teaching manual*, 1989.

[47] M. Benbow, B. E. Hanft, and D. Marsh, *Handwriting in the classroom: Improving written communication.* American Occupational Therapy Association, 1992.

[48] A. Kushki, T. Chau, and E. Anagnostou, "Handwriting Difficulties in Children with Autism Spectrum Disorders: A Scoping Review," *Journal of Autism and Developmental Disorders*, vol. 41, pp. 1706–1716, Dec. 2011.

[49] M. H. Tseng and S. M. Chow, "Perceptual-motor function of school-age children with slow handwriting speed," *American Journal of Occupational Therapy*, vol. 54, no. 1, pp. 83–88, 2000.

[50] F. Bara and E. Gentaz, "Haptics in teaching handwriting: The role of perceptual and visuo-motor skills," *Human movement science*, vol. 30, no. 4, pp. 745–759, 2011.

[51] R. Karlsdottir, "Development of cursive handwriting," *Perceptual and Motor Skills*, vol. 82, no. 2, pp. 659–673, 1996.

[52] A. Kushki, T. Chau, and E. Anagnostou, "Handwriting difficulties in children with autism spectrum disorders: A scoping review," *Journal of autism and developmental disorders*, vol. 41, no. 12, pp. 1706–1716, 2011.

[53] M. H. Tseng and S. A. Cermak, "The influence of ergonomic factors and perceptual–motor abilities on handwriting performance," *American Journal of Occupational Therapy*, vol. 47, no. 10, pp. 919–926, 1993.

[54] J. Bluteau, S. Coquillart, Y. Payan, and E. Gentaz, "Haptic guidance improves the visuo-manual tracking of trajectories," *PLoS One*, vol. 3, no. 3, p. e1775, 2008.

[55] J. Danna and J.-L. Velay, "Basic and supplementary sensory feedback in handwriting," *Frontiers in psychology*, vol. 6, 2015.

[56] D. Hayes, "Handwriting practice: The effects of perceptual prompts," *The Journal of Educational Research*, vol. 75, no. 3, pp. 169–172, 1982.

[57] D. Arslan, "Examining first grade teachers' handwriting instruction.," *Educational Sciences: Theory and Practice*, vol. 12, no. 4, pp. 2839–2846, 2012.

[58] S. Graham and N. Weintraub, "A review of handwriting research: Progress and prospects from 1980 to 1994," *Educational psychology review*, vol. 8, no. 1, pp. 7–87, 1996.

[59] A. Accardo, M. Genna, and M. Borean, "Development, maturation and learning influence on handwriting kinematics.," *Human Movement Science*, vol. 32, no. 1, pp. 136–146, 2013.

[60] L. Hamstra-Bletz, J. de Bie, and B. den Brinker, *Concise evaluation scale for children's handwriting.* swets 1 zeitlinger ed., 1987.

[61] M. Charles, R. Soppelsa, and J.-M. Albaret, *BHK : échelle d'évaluation rapide de l'écriture chez l'enfant.* ecpa ed., 2003.

[62] S. Graham, V. Berninger, N. Weintraub, and W. Schafer, "Development of handwriting speed and legibility in grades 1–9," *The Journal of Educational Research*, vol. 92, no. 1, pp. 42–52, 1998.

[63] A. Maeland and R. Karlsdottir, "Development of reading, spelling, and writing skills from third to sixth grade in normal and dysgraphic school children.," in *Development of graphic skills: Research perspectives and educational implications*, pp. 179–189, San Diego, CA, USA: J. Wann, A. M. Wing, & N. Sõvik, academix press ed., 1991.

[64] S. Graham, "The role of production factors in learning disabled students' compositions.," *Journal of Educational Psychology*, vol. 82, no. 4, pp. 781–791, 1990.

[65] V. W. Berninger and S. Graham, "Language by hand: A synthesis of a decade of research on handwriting," *Handwriting Review*, vol. 12, no. 1, pp. 11–25, 1998.

[66] C. I. Chase, "Essay test scoring: Interaction of relevant variables," *Journal of Educational Measurement*, vol. 23, no. 1, pp. 33–41, 1986.

[67] C. O. Sweedler-Brown, "The effect of training on the appearance bias of holistic essay graders.," *Journal of Research and Development in Education*, vol. 26, no. 1, pp. 24–29, 1992.

[68] K. Feder, A. Majnemer, and A. Synnes, "Handwriting: Current trends in occupational therapy practice," *Canadian Journal of Occupational Therapy*, vol. 67, no. 3, pp. 197–204, 2000.

[69] B. Smits-Engelsman, A. Niemeijer, and G. van Galen, "Fine motor deficiencies in children diagnosed as DCD based on poor grapho-motor ability.," *Human Movement Science*, vol. 20, pp. 161–182, 2001.

[70] J. Mekyska, M. Faundez-Zanuy, Z. Mzourek, Z. Galaz, Z. Smekal, and S. Rosenblum, "Identification and rating of developmental dysgraphia by handwriting analysis," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 235–248, 2016.

[71] E. L. Thorndike, *Handwriting.* Teachers College, Columbia University, 1912.

[72] S. Rosenblum, P. L. Weiss, and S. Parush, "Product and process evaluation of handwriting difficulties," *Educational Psychology Review*, vol. 15, no. 1, pp. 41–81, 2003.

[73] R. Bezzi, "A standardized manuscript scale for grades 1, 2, and 3," *The Journal of Educational Research*, vol. 55, no. 7, pp. 339–340, 1962.

[74] F. N. Freeman, "A new handwriting scale," *The Elementary School Journal*, vol. 59, no. 4, pp. 218–221, 1959.

[75] V. E. Herrick and A. Erlebacher, "The evaluation of legibility in handwriting," *New horizons for research in handwriting*, pp. 207–236, 1963.

[76] S. J. Amundson, *Evaluation Tool of Children's Handwriting: ETCH examiner's manual.* OT KIDS, 1995.

[77] N. Erez and S. Parush, *The Hebrew Handwriting Evaluation (2nd ed.).* school of occupational therapy, faculty of medicine, hebrew university of jerusalem ed., 1999.

[78] L. Hamstra-Bletz, J. DeBie, and B. Den Brinker, "Concise evaluation scale for children's handwriting," *Lisse: Swets*, vol. 1, 1987.

[79] R. Soppelsa and J. Albaret, "Evaluation de l'écriture chez l'adolescent. le bhk ado," *Entretiens de Psychomotricité*, pp. 66–76, 2012.

[80] A. L. Barnett, S. E. Henderson, B. Scheib, and J. Schulz, "Development and standardization of a new handwriting speed test: The detailed assessment of speed of handwriting," in *BJEP Monograph Series II, Number 6-Teaching and Learning Writing*, vol. 137, pp. 137–157, British Psychological Society, 2009.

[81] N. Erez and S. Parush, "The hebrew handwriting evaluation," *School of Occupational Therapy. Faculty of Medicine. Hebrew University of Jerusalem, Israel*, 1999.

[82] E. Pagliarini, L. Scocchia, M. Vernice, M. Zoppello, U. Balottin, S. Bouamama, M. T. Guasti, and N. Stucchi, "Children's first handwriting productions show a rhythmic structure," *Scientific reports*, vol. 7, no. 1, p. 5516, 2017.

[83] S. Rosenblum and G. Dror, "Identifying developmental dysgraphia characteristics utilizing handwriting classification methods," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 293–298, 2016.

[84] S. M. Cahill, "Where does handwriting fit in? strategies to support academic achievement," *Intervention in school and clinic*, vol. 44, no. 4, pp. 223–228, 2009.

[85] V. W. Berninger and F. Fuller, "Gender differences in orthographic, verbal, and compositional fluency: Implications for assessing writing disabilities in primary grade children," *Journal of School Psychology*, vol. 30, no. 4, pp. 363–382, 1992.

[86] S. Graham, K. R. Harris, L. Mason, B. Fink-Chorzempa, S. Moran, and B. Saddler, "How do primary grade teachers teach handwriting? a national survey," *Reading and writing*, vol. 21, no. 1-2, pp. 49–69, 2008.

[87] C. E. Oliver, "A sensorimotor program for improving writing readiness skills in elementary-age children," *American Journal of Occupational Therapy*, vol. 44, no. 2, pp. 111–116, 1990.

[88] J. E. Reisman, "Poor handwriting: Who is referred?," *American Journal of Occupational Therapy*, vol. 45, no. 9, pp. 849–852, 1991.

[89] H. Cornhill and J. Case-Smith, "Factors that relate to good and poor handwriting," *American Journal of Occupational Therapy*, vol. 50, no. 9, pp. 732–739, 1996.

[90] L. Stephens and P. Pratt, "School work tasks and vocational readiness," *Occupational therapy for children*, pp. 311–324, 1989.

[91] M.-A. Bonney, "Understanding and assessing handwriting difficulty: Perspectives from the literature," *Australian Occupational Therapy Journal*, vol. 39, no. 3, pp. 7–15, 1992.

[92] K. Zolna, T. Asselborn, C. Jolly, L. Casteran, W. Johal, P. Dillenbourg, *et al.*, "The dynamics of handwriting improves the automated diagnosis of dysgraphia," *arXiv preprint arXiv:1906.07576*, 2019.

[93] D. Alamargot and M.-F. Morin, "Does handwriting on a tablet screen affect students' graphomotor execution? a comparison between grades two and nine," *Human movement science*, vol. 44, pp. 32–41, 2015.

[94] C. Jolly, C. Huron, and E. Gentaz, "A one-year survey of cursive letter handwriting in a french second-grade child with developmental coordination disorder," *L'Année Psychol*, 2014.

[95] C. Jolly and E. Gentaz, "Analysis of cursive letters, syllables, and words handwriting in a french second-grade child with developmental coordination disorder and comparison with typically developing children," *Frontiers in psychology*, vol. 4, p. 1022, 2014.

[96] L. Torrey and J. Shavlik, "Transfer learning," *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, vol. 1, p. 242, 2009.

[97] M. Charles, R. Soppelsa, and J. Albaret, "Échelle d'évaluation rapide de l'écriture chez l'enfant (bhk)," *Paris: Édition du Centre de psychologie appliquée*, 2004.

[98] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, 2001.

[99] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[100] S. Rosenblum, S. Parush, and P. L. Weiss, "The in air phenomenon: Temporal and spatial correlates of the handwriting process," *Perceptual and Motor skills*, vol. 96, no. 3, pp. 933–954, 2003.

[101] E. Altynsarina, "Kazakhstan adopts new version of latin-based kazakh alphabet.," 26 February 2018. [Accessed on 1-July-2019: https://astanatimes.com/2018/02/kazakhstan-adopts-new-version-of-latin-based-kazakh-alphabet/].

[102] C. Matta Abizeid, A. Tabsh Nakib, C. Younès Harb, S. Ghantous Faddoul, and J.-M. Albaret, "Handwriting in lebanese bigraphic children: Standardization of the bhk scale," *Journal of Occupational Therapy, Schools, & Early Intervention*, vol. 10, no. 4, pp. 420–435, 2017.

[103] S. Gerth, T. Dolk, A. Klassert, M. Fliesser, M. H. Fischer, G. Nottbusch, and J. Festman, "Adapting to the surface: a comparison of handwriting measures when writing on a tablet computer and on paper," *Human movement science*, vol. 48, pp. 62–73, 2016.

[104] S. Gerth, A. Klassert, T. Dolk, M. Fliesser, M. H. Fischer, G. Nottbusch, and J. Festman, "Is handwriting performance affected by the writing surface? comparing preschoolers', second graders', and adults' writing performance on a tablet vs. paper," *Frontiers in psychology*, vol. 7, p. 1308, 2016.

[105] V. W. Berninger, K. H. Nielsen, R. D. Abbott, E. Wijsman, and W. Raskind, "Gender differences in severity of writing and reading disabilities," *Journal of School Psychology*, vol. 46, no. 2, pp. 151–172, 2008.

[106] L. Bourke and A.-M. Adams, "Is it differences in language skills and working memory that account for girls being better at writing than boys?," *Journal of Writing Research*, vol. 3, no. 3, 2011.

[107] C. A. Rohrbeck, M. D. Ginsburg-Block, J. W. Fantuzzo, and T. R. Miller, "Peer-assisted learning interventions with elementary school students: A meta-analytic review.," 2003.

[108] C. C. Chase, D. B. Chin, M. A. Oppezzo, and D. L. Schwartz, "Teachable agents and the protégé effect: Increasing the effort towards learning," *Journal of Science Education and Technology*, vol. 18, no. 4, pp. 334–352, 2009.

[109] T. Asselborn, W. Johal, and P. Dillenbourg, "Keep on moving! exploring anthropomorphic effects of motion during idle moments," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 897–902, IEEE, 2017.

[110] S. Shamsuddin, H. Yussof, L. Ismail, F. A. Hanapiah, S. Mohamed, H. A. Piah, and N. I. Zahari, "Initial response of autistic children in human-robot interaction

therapy with humanoid robot NAO," *Proceedings - 2012 IEEE 8th International Colloquium on Signal Processing and Its Applications, CSPA 2012*, pp. 188–193, 2012.

[111] E. S. Kim, L. D. Berkovits, E. P. Bernier, D. Leyzberg, F. Shic, R. Paul, and B. Scassellati, "Social robots as embedded reinforcers of social behavior in children with autism," *Journal of autism and developmental disorders*, vol. 43, no. 5, pp. 1038–1049, 2013.

[112] S. Lemaignan, A. Jacq, D. Hood, F. Garcia, A. Paiva, and P. Dillenbourg, "Learning by Teaching a Robot: The Case of Handwriting," *IEEE Robotics and Automation Magazine*, vol. 23, no. 2, pp. 56–66, 2016.

[113] S. Sturgeon, A. Palmer, J. Blankenburg, and D. Feil-Seifer, "Perception of social intelligence in robots performing false-belief tasks," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1–7, IEEE, 2019.

[114] J. Vroon, G. Englebienne, and V. Evers, "Detecting perceived appropriateness of a robot's social positioning behavior from non-verbal cues," in *2019 IEEE First International Conference on Cognitive Machine Intelligence (CogMI)*, pp. 216–225, IEEE, 2019.

[115] A. S. Ghazali, J. Ham, E. Barakova, and P. Markopoulos, "Persuasive robots acceptance model (pram): roles of social responses within the acceptance model of persuasive robots," *International Journal of Social Robotics*, pp. 1–18, 2020.

[116] M. M. de Graaf, S. Ben Allouch, and J. A. van Dijk, "Why would i use this in my home? a model of domestic social robot acceptance," *Human–Computer Interaction*, vol. 34, no. 2, pp. 115–173, 2019.

[117] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial," *Human-Computer Interaction*, vol. 19, no. 1, pp. 61–84, 2004.

[118] T. Asfour and R. Dillmann, "Human-like motion of a humanoid robot arm based on a closed-form solution of the inverse kinematics problem," in *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 2, pp. 1407—-1412 vol.2, 2003.

[119] H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takariobu, S. Roccella, M. Carrozza, P. Dario, and a. Takanishi, "Effective emotional expressions with emotion expression humanoid robot WE-4RII," *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 3, no. 4, pp. 2203–2208, 2004.

[120] C. Breazeal and B. Scassellati, "How to build robots that make friends and influence people," *Intelligent Robots and Systems*, pp. 858–863, 1999.

[121] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," 1969.

[122] T. G. H. Mark L. Knapp, Judith A. Hall, *Nonverbal Communication in Human Interaction*. Wadsworth Publishing, 8 ed., 2013.

[123] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, "Child-robot interaction in the wild: advice to the aspiring experimenter," in *Proceedings of the 13th international conference on multimodal interfaces*, pp. 335–342, ACM, 2011.

[124] D. Cameron, S. Fernando, E. Collins, A. Millings, R. Moore, A. Sharkey, V. Evers, and T. Prescott, "Presence of life-like robot expressions influences children's enjoyment of human-robot interactions in the field," in *Proceedings of the AISB Convention 2015*, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2015.

[125] T. Ribeiro and A. Paiva, "The illusion of robotic life: Principles and practices of animation for robots," in *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 383–390, March 2012.

[126] M. Salem, F. Eyssel, K. Rohlfing, S. Kopp, and F. Joublin, "Effects of gesture on the perception of psychological anthropomorphism: a case study with a humanoid robot," pp. 31–41, 2011.

[127] F. Eyssel, F. Hegel, G. Horstmann, and C. Wagner, "Anthropomorphic inferences from emotional nonverbal cues: A case study," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, no. October, pp. 646–651, 2010.

[128] M. Kocoń, "Idle motion synthesis of human head and face in virtual reality environment," pp. 299–306, 2013.

[129] A. Egges, T. Molet, and N. Magnenat-Thalmann, "Personalised real-time idle motion synthesis," *Proceedings - Pacific Conf on Computer Graphics and Appl*, pp. 121–130, 2004.

[130] A. Egges, R. Visser, and N. Magnenat-Thalmann, "Example-Based idle motions in a real-time Application," *CAPTECHWorkshop*, no. December, pp. 13–19, 2004.

[131] J. Jung, T. Kanda, and M. S. Kim, "Guidelines for Contextual Motion Design of a Humanoid Robot," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 153–169, 2013.

[132] C. Larboulette and S. Gibet, "A review of computable expressive descriptors of human motion," in *Proceedings of the 2nd International Workshop on Movement and Computing*, pp. 21–28, ACM, 2015.

[133] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 25–32, ACM, 2014.

[134] E. Wang, C. Lignos, A. Vatsal, and B. Scassellati, "Effects of head movement on perceptions of humanoid robot behavior," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 180–185, ACM, 2006.

[135] K. Fischer, K. S. Lohan, and K. Foth, "Levels of embodiment: Linguistic analyses of factors influencing hri," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 463–470, ACM, 2012.

[136] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[137] A. Özgür, S. Lemaignan, W. Johal, M. Beltran, M. Briod, L. Pereyre, F. Mondada, and P. Dillenbourg, "Cellulo: Versatile handheld robots for education," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 119–127, ACM, 2017.

[138] S. Klein, V. Guiltner, P. Sollereder, and Y. Cui, "Relationships between fine-motor, visual-motor, and visual perception scores and handwriting legibility and speed," *Physical & Occupational Therapy in Pediatrics*, vol. 31, no. 1, pp. 103–114, 2011.

[139] M. M. Smyth, "Visual control of movement patterns and the grammar of action," *Acta psychologica*, vol. 70, no. 3, pp. 253–265, 1989.

[140] R. Palluel-Germain, F. Bara, A. H. De Boisferon, B. Hennion, P. Gouagout, and E. Gentaz, "A visuo-haptic device-telemaque-increases kindergarten children's handwriting acquisition," in *EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint*, pp. 72–77, IEEE, 2007.

[141] R. Deuel, "Developmental Dysgraphia and Motor Skills Disorders," *Journal of Child Neurology*, vol. 10, no. suppl., pp. S6–S8, 1995.

Thibault ASSELBORN
21/07/1991
CHILI Lab – Ecole Polytechnique Fédérale de Lausanne (EPFL)
RLC D1 740 (Rolex Learning Center)
Station 20
CH-1015 Lausanne, Switzerland
Phone : +33785597512
E-mail : thibault.asselborn@epfl.ch

FORMATIONS

2020 *(expected)* **PHD in Computer Science, Robotics**
Swiss Federal Institute of technology (EPFL) // ***Lausanne/Switzerland***

2015 **Master of Science MSc in Microengineering, orientation Robotics and Autonomous Systems**
Swiss Federal Institute of technology (EPFL) // ***Lausanne/Switzerland*** / University College London (UCL) // ***London/UK***

2013 **Bachelor of Science BSc in Microengineering**
Swiss Federal Institute of technology (EPFL) // ***Lausanne/Switzerland***

2009 **Scientific Baccalauréat option engineering science, speciality in physics**
Lycée Ana-de-Noailles // ***Evian/France***

PROFESSIONAL EXPERIENCE

07/2016 - present **PhD Student at EPFL/CHILI**
Swiss Federal Institute of Technology (EPFL), Computer-human Interaction in Learning and Instruction (CHILI) // ***Lausanne, Switzerland***
Computer Science, Machine learning and Human Robot Interaction. Teaching math and computer science courses

09/2015 - present **Chief Algorithm Officier (CAO) at Flyerz Applications**
Flyerz Applications // ***Marin, France***
Computer Science and Machine learning

09/2015 - 06/2016 **Intern at University College London (UCL)**
University College London (UCL) – Department of biochemical engineering and departement of computer science // ***London/UK***
Computer Science and Machine learning

03/2015 - 09/2015 **Master Project**
University College London (UCL) – Department of biochemical engineering and departement of computer science // ***London/UK***
Computer Science and machine learning

08/2014 - 02/2015 **Junior engineer at Imina technologies SA**
Imina Technologies SA // ***Lausanne, Switzerland***
Computer Science, robotic and electronic

PUBLICATIONS

-Climbing favours the tripod gait over alternative faster insect gaits. *Pavan Ramdya, Robin Thandiackal, Raphael Cherney,* ***Thibault Asselborn****, Richard Benton, Auke Jan Ijspeert, Dario Floreano*, Nature Communication, 2017

-Automated human-level diagnosis of dysgraphia using a consumer tablet. ***Thibault Asselborn****, Thomas Gargot, Łukasz Kidziński, Wafa Johal, David Cohen, Caroline Jolly, Pierre Dillenbourg*, Nature Digital Medicine, 2018

-Keep on moving! Exploring anthropomorphic effects of motion during idle moments. ***T Asselborn****, W Johal, P Dillenbourg*, 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2017

-Iterative classroom teaching. *Teresa Yeo, Parameswaran Kamalaruban, Adish Singla, Arpit Merchant,* ***Thibault Asselborn****, Louis Faucon, Pierre Dillenbourg, Volkan Cevher*, Proceedings of the AAAI Conference on Artificial Intelligence,2019

-Bringing letters to life: handwriting with haptic-enabled tangible robots. ***Thibault Asselborn****, Arzu Guneysu, Khalil Mrini, Elmira Yadollahi, Ayberk Ozgur, Wafa Johal, Pierre Dillenbourg*, Proceedings of the 17th ACM Conference on Interaction Design and Children, 2018

-The Dynamics of Handwriting Improves the Automated Diagnosis of Dysgraphia. *Konrad Zolna,* ***Thibault Asselborn****, Caroline Jolly, Laurence Casteran, Wafa Johal, Pierre Dillenbourg*, Archivx, 2019

-CoWriting Kazakh: Transitioning to a New Latin Script using Social Robots. *Anton Kim, Meruyert Omarova, Adil Zhaksylyk,* ***Thibault Asselborn****, Wafa Johal, Pierre Dillenbourg, Anara Sandygulova*, 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN),2019

-CoWriting Kazakh: Learning a New Script with a Robot. *Anara Sandygulova, Wafa Johal, Zhanel Zhexenova, Bolat Tleubayev, Aida Zhanatkyzy, Aizada Turarova, Zhansaule Telisheva, Anna CohenMiller,* ***Thibault Asselborn****, Pierre Dillenbourg*, Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 2020

-Iterative Design and Evaluation of a Tangible Robot-Assisted Handwriting Activity for Special Education. *Arzu Guneysu Ozgur, Ayberk Özgür,* ***Thibault Asselborn****, Wafa Johal, Elmira Yadollahi, Barbara Bruno, Melissa Skweres, Pierre Dillenbourg*, Frontiers in Robotics and AI, 2020

-Extending the Spectrum of Dysgraphia: A Data Driven Strategy to estimate Handwriting Quality. ***T Asselborn****, M Chapatte, P Dillenbourg*, Scientific Reports, 2020

-Learning By Collaborative Teaching: An Engaging Multi-Party CoWriter Activity. *Laila El Hamamsy, Wafa Johal,* ***Thibault Asselborn****, Jauwairia Nasir, Pierre Dillenbourg*, 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2019

-Bridging Multilevel Time Scales in HRI: An Analysis Framework. ***Thibault Asselborn****, Kshitij Sharma, Wafa Johal, Pierre Dillenbourg*, ACM Transactions on Human-Robot Interaction (THRI), 2019

-Designing Configurable Arm Rehabilitation Games: How Do Different Game Elements Affect User Motion Trajectories?. *Arzu Guneysu Ozgur, Maximilian J Wessel,* ***Thibault Asselborn****, Jennifer K Olsen, Wafa Johal, Ayberk Özgür, Friedhelm C Hummel, Pierre Dillenbourg*, 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019

-Reply: Limitations in the creation of an automatic diagnosis tool for dysgraphia. ***Thibault Asselborn****, Thomas Gargot, Łukasz Kidziński, Wafa Johal, David Cohen, Caroline Jolly, Pierre Dillenbourg*, Nature Digital Medicine, 2019

-Learning goog handwriting in greek by teaching a humanoid robot. *Christia Ioannou, Christalla Neophytou,* ***Thibault Asselborn****, Wafa Joahal, Thanasis Hadzilacos*, 14th International Technology, Education and Development Conference, 2020

PATENTS

-Method of handwritten character recognition confirmation. *Konrad Zolna,* ***Thibault Asselborn****, Wafa Joahal,* 2019

-Method of analysing handwritten items. ***Thibault Asselborn****, Wafa Joahal, Pierre Dillenbourg, Corinne Lebourgeois,* 2019

SKILLS

Robotic ; Machine Learning ; Artificial Inteligence ; Programming ; Management ; Electronics ; Image and signal processing.

Programming Language: Python, Swift, C/C++, Matlab, Java
Development tools: Pytorch, TensorFlow, Matlab, Pandas, ROS, Qt, Webots, Spice, Altium Designer, Pro Engineer, SolidWorks
Office tools: MS Office (Word, Excel, PowerPoint), Latex
Version Control: Git, SVN