

Virtual reality and analytics to enhance communication skills: a multidisciplinary approach

Dissertation submitted in fulfilment of the requirements
for the degree of Ph.D. in Economics and Management
by


Elodie ETIENNE

Jury:

- Prof. Laurence DESSART, President of the Jury
(HEC Liège, Management School of the University of Liège)
- Prof. Carlos FLAVIÁN
(METODO Lab, University of Zaragoza)
- Prof. Ashwin ITTOO
(HEC Liège, Management School of the University of Liège)
- Prof. Magalie OCHS
(LIS Lab, Aix-Marseille University)
- Prof. Anne-Lise LECLERCQ
(FPLSE, University of Liège)
- Prof. Michaël SCHYNS, Supervisor
(HEC Liège, Management School of the University of Liège)

Université de Liège - Atelier des Presses
Chemin des Amphithéâtres - Bât B7a
4000 Liège (Belgique)

© 2025

 Atelier des Presses

Tous droits de reproduction,
d'adaptation et de traduction
réservés pour tous pays.

Ouvrage mis en page par l'auteur
Imprimé en Belgique

D/2025/13.315/1

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Michaël Schyns, my supervisor and mentor. You have always been there to support me, to push me to give my best, and to listen through the highs and lows over the years. I know this journey was a gamble (a long one), but I'm truly thankful you took that chance on me. You offered me the space to grow, the chance to take part in events I never imagined attending, to enjoy so many unique experiences, and, most importantly, the privilege of having you as a supervisor. I'm especially grateful for your presence beyond the professional sphere during moments of doubt, major life decisions, and future uncertainties. I couldn't have hoped for someone who understood me better, even if I may have exhausted you more than once over these six years. Your guidance, kindness, and belief in me have meant more than words can express.

To Anne-Lise Leclercq, but also to Angélique Remacle, thank you for your unwavering encouragement from the very beginning. Your kind words have meant so much since my first chaotic (from my point of view) experience in Virtual Reality five years ago, when, I admit, I wondered what I had gotten myself into. You have always generously shared your ideas and comments whenever needed, and I deeply appreciated your presence and guidance throughout this journey.

I am also immensely grateful to Laurence Dessart and Ashwin Ittoo, who have accompanied me for many years with smiles, thoughtful advice during thesis committee meetings, and precious suggestions whenever I reached out.

A warm thank you to Magalie Ochs, who accepted to welcome me into her team in Marseille for seven months without really knowing me at that time, yet integrated me fully and allowed me to participate in numerous events. Your constant support, especially for what comes next, means a lot.

To Carlos Flavián, thank you for showing interest in our research from the very beginning, for giving me a chance as a young researcher, and for always having kind and encouraging words. I am especially honoured to have had you as part of my jury.

I would also like to thank someone for whom I have deep admiration: Gentiane Haesbroeck, my Master's thesis supervisor. Her trust and guidance

six years ago were crucial in shaping the path I followed.

I am also profoundly grateful to those who allowed me to explore another passion: teaching. Leading tutorials in Mathematics and Statistics helped me realise that I wanted to continue on this path. I would especially like to thank Marie-Christine Cillis. Her trust and encouragement played a key role in shaping this part of my journey. Thank you to Pascal Dupont and Isabelle Pays for the freedom they gave me in handling their courses. I am especially thankful for the wonderful memories over the years, from the work itself to oral exams where we had a lot of fun and various activities. Thanks as well to Elise Vandomme for trusting me to continue teaching and always involving me in her new ideas. In this same spirit, I cannot forget Sabine Maron and Eddy Flas, who have always been there with a helping hand and a listening ear whenever I needed it. Thank you also to Stéphanie Aerts and Christine Bertrand for trusting me with the responsibility of teaching the Marketing Analytics and Statistics courses. These two opportunities not only deepened my passion for leading a course, but also significantly enhanced my professional profile.

I would also like to express my sincere thanks to HEC Liège for providing a stimulating academic environment throughout these years, and for the opportunity to be part of PRISME, an involvement that allowed me to better understand the research academic landscape. A special thanks to Georges Koidakis and Etienne Angenot, always cheerful in our interactions and always responding to my many logistical requests.

To my wonderful colleagues, especially those from my office, formerly Office 334, now Office 337, it was a pleasure to share a space with you throughout these six years. Over the years, I had the pleasure of meeting many people in this office, and each of you left a positive mark on my journey. You made the day-to-day work life lighter, more supportive, and filled with laughter. Moreover, the days would not have been the same without all the members of QuantOM, with whom I laughed and shared countless conversations about everything and nothing. I would also like to thank Lisa Baiwir, Jordan Fleissig, and Florence Nizette, colleagues from the Marketing team. Discussing with you is always a pleasure, and I truly appreciated the moments we shared.

I am also thankful to the AR/VR SIG Lab, without whom this thesis would not have been possible. Beyond their work, I thank them for all the moments we shared, from lunch breaks to my near-daily visits to chat or complain about one thing or another. Over the years, many faces came and went. A special thanks to Sarah Saufnay for being both an amazing colleague and a trusted friend. Over these two years, we have attended numerous conferences, supported each other through difficult times, and shared a fair bit of gossip too.

Throughout this PhD, I was fortunate to take part in many collaborations. I especially want to thank Lamia Bettahi, Pauline Menjot, Justin Cho,

and Marion Ristorcelli. Our collaborations were not only enriching and constructive, but also built on trust and support. I would also like to express my gratitude to their supervisors, and beyond those already mentioned, I thank Timothy Jung and Rémy Casanova. These projects also gave me the opportunity to take part in research stays, during which I had the pleasure of meeting colleagues and sharing moments I truly appreciated, including the team at Aix-Marseille University, especially Alice Delbosc, Ioana Ivan and Ikram Belmadani.

I want to extend a heartfelt thank you to Emeline Leloup, a colleague, but above all my best friend for the past ten years. I have always been able to count on you, both professionally and personally. I am also deeply grateful to my closest friends, especially Kevin, Elise, Romane, and Tom for their constant support and presence.

Et à ceux sur qui j'ai toujours pu compter : mes parents. Je vous admire de tout cœur pour votre force de caractère, pour avoir surmonté les épreuves de la vie, et pour avoir toujours été là, de manière infaillible, pour ma sœur Justine et moi. À ma sœur, dont la fierté à mon égard m'a toujours touchée, et dont l'humour et les surnoms qu'elle me trouve ne font souvent rire qu'elle, mais qui font d'elle une personne unique que j'ai la chance d'avoir à mes côtés.

And finally, to the person who came into my life during these past few years, Raphaël, known to my colleague and friends as "the Tiler", but who is to me the person who matters most. Your support and love, whether near or far from each other, mean everything to me. I'm deeply grateful for your presence in my life, and I sincerely hope we'll continue walking this path together for many years to come.

Contents

I	Introduction	1
1.1	Analytics	1
1.1.1	The Role of Data-Driven Approaches	2
1.2	New Technologies	2
1.2.1	Virtual Reality	3
1.3	Communication	7
1.4	Virtual Reality to Enhance Communication Skills Using Analytics	8
1.4.1	A Promising Tool for Training	8
1.4.2	Analytics and Human Behaviour Research	9
1.4.3	Limitations and Challenges	10
1.5	Thesis Overview and Research Line	11
1.5.1	Thesis Structure	12
II	Perception of IVAs' behaviour	15
2.1	Introduction	16
2.2	Methodology	18
2.3	Results	20
2.3.1	Emotions	20
2.3.2	Personality	22
2.3.3	Trustworthiness	24
2.3.4	Social Capabilities	26
2.3.5	Believability	27
2.4	Discussion	28
III	Perception of Avatars in VR	33
3.1	Introduction	34
3.2	Theoretical Development	36
3.3	Methodology	40
3.3.1	The Virtual Reality Environment	40
3.3.2	Non-Verbal Behaviours	41
3.3.3	The Virtual Reality Conditions	43
3.3.4	The Experiment Settings	43

CONTENTS

3.4	Results	45
3.4.1	Analysis of Posture, Facial Expression, and Head Movement	45
3.4.2	Library of Animated Avatars	46
3.4.3	Comparison Between Low-end and High-end Headsets	49
3.4.4	Comparison Between Photo-realistic and Cartoon Avatars	49
3.4.5	Sense of Presence	50
3.5	Qualitative Assessment	50
3.6	Discussion	51
3.7	Limitations and Future Research	53
3.8	Appendix	56
3.8.1	Analysis of Postures, Facial Expressions, and Head Movements	56
IV	Implementation of VR in legal education	59
4.1	Introduction	60
4.2	Literature Review	61
4.2.1	Legal Simulation	61
4.2.2	Impacts of Simulated Learning Techniques in Legal Education	61
4.2.3	Virtual Reality Education	62
4.2.4	Impacts of Virtual Reality on Learning	63
4.2.5	Limitations of Virtual Reality in Education	63
4.3	Framework Development	64
4.3.1	Identification of Research Gaps	64
4.3.2	Proposal of Extended Kolb's Experiential Learning Theory	65
4.4	Methodology	67
4.4.1	Virtual Reality Environment Design	67
4.4.2	Refining the Virtual Reality Environment Through Educator Feedback	69
4.4.3	Mixed Methods Strategy	70
4.4.4	Procedure and Participants	71
4.5	Findings	72
4.5.1	Enhanced and Contextualised Learning	72
4.5.2	Technological Feasibility of Virtual Reality for Mock Trials	78
4.5.3	Influence of Immersion on Learning	81
4.6	Discussion	84
4.7	Conclusions	86
4.7.1	Theoretical Contributions	86
4.7.2	Practical Contributions	87
4.7.3	Limitations and Future Research Directions	87
4.8	Appendix	88

V	Multimodal Cues for PST in VR	91
5.1	Introduction	92
5.2	Verbal Features	93
5.2.1	Acoustic Features	93
5.2.2	Textual Features	95
5.3	Non-Verbal Features	96
5.4	Virtual Reality Solution	101
5.5	Discussion	103
VI	EVE	105
6.1	Introduction	106
6.1.1	Emotions and Implications for Affective Computing	106
6.1.2	Existing Emotional Speech Corpora	107
6.2	The EVE (Emotional Validated Expressions) Corpus	108
6.2.1	Corpus Creation: Audiovisual Data Collection	108
6.2.2	Availability of the Corpus	110
6.2.3	Corpus Validation: Perceptual Study	110
6.2.4	Availability of the Perceptual Raw Data	111
6.3	Results of the Perceptual Study	112
6.4	Discussion	115
VII	Corpora of PS in VR	117
7.1	Introduction	118
7.2	Corpora Collection	119
7.2.1	Conditions	119
7.2.2	Public Speaking Tasks	120
7.2.3	Measures	121
7.2.4	Recordings	123
7.2.5	Procedure	123
7.3	Participants	125
7.4	Performance Evaluations	126
7.5	Results	126
7.6	Conclusions	127
VIII	PS Performance Prediction using ML	129
8.1	Introduction	129
8.2	Theoretical Background	130
8.2.1	Prediction	130
8.2.2	Multi-layer Perceptron	130
8.2.3	Cross-Validation	132
8.2.4	Model Evaluation	133
8.3	Data	133
8.3.1	Description of the Dataset	133
8.3.2	Data Segmentation	134

CONTENTS

8.3.3	Dataset Preparation	135
8.3.4	Data Augmentation	136
8.3.5	Model Configuration	136
8.3.6	Finetuning	136
8.4	Results	136
8.4.1	Performance with Seven Classes	137
8.4.2	Performance with Grouped Classes	137
8.4.3	Segmentation	138
8.4.4	Expert Evaluations	139
8.5	Discussion	139
IX	Conclusion	147

Index

AI	Artificial Intelligence
AR	Augmented Reality
BFI	Big Five Inventory
CAVE	Cave Automatic Virtual Environment
CV	Coefficient of Variation
ELT	Experiential Learning Theory
F0	Fundamental Frequency
HMD	Head-Mounted Display
HR	Heart-Rate
IPQ	Igroup Presence Questionnaire
ITC-SOPI	ITC-Sense Of Presence Inventory
ITQ	Immersive Tendencies Questionnaire
IVA	Interactive Virtual Agent
LIWC	Linguistic Inquiry and Word Count
LOOCV	Leave-One-Out Cross-Validation
LSAS-SR	Liebowitz Social Anxiety Scale - Self Report
ML	Machine Learning
MLP	Multi-Layer Perceptron
MR	Mixed Reality
NLP	Natural Language Processing
PAD	Pleasure-Arousal-Dominance
PRPSA	Personal Report of Public Anxiety
PS	Public Speaking
PST	Public Speaking Training
SAM	Self-Assessment Manikin
SER	Speech Emotion Recognition
SSI	Stuttering Severity Instrument
STAI-T	State-Trait Anxiety Inventory – Trait
SUDS	Subjective Units of Distress Scale
TTR	Type-Token Ratio
VE	Virtual Environment
VHI	Voice Handicap Index
VR	Virtual Reality
XR	eXtended Reality

Chapter I

Introduction

This chapter introduces the context and motivation for this thesis, outlining the growing role of data-driven approaches in business research and the transformative potential of emerging technologies. Businesses increasingly rely on analytics and quantitative methodologies to enhance decision-making, optimise performance, and better understand human behaviour in organisational settings. At the same time, advancements in Machine Learning (ML) have opened new possibilities for immersive training and behavioural analysis. Virtual Reality (VR)'s ability to create controlled, interactive, and data-rich environments makes it an effective tool for skill development, including public speaking training (PST). This chapter explores the evolution of VR, detailing its mechanisms, system types and its expanding research applications. Furthermore, the chapter highlights how VR can enhance communication skills, particularly in public speaking (PS), by providing dynamic, interactive, and adaptive training environments. Finally, this chapter sets the stage for the rest of the dissertation, presenting the research questions, objectives, and the structure of the following chapters, each addressing a key aspect of VR-based PST and its evaluation through multimodal analytics.

1.1 Analytics

In today's rapidly evolving digital landscape, businesses increasingly rely on analytics and artificial intelligence (AI) to drive decision-making, optimise operations, and gain deeper insights into consumer behaviour and market trends. The ability to extract meaningful patterns from large volumes of data has revolutionised industries, enabling organisations to improve efficiency, reduce costs, and enhance customer experiences. Analytics serves as the foundation of modern business intelligence, combining statistical methods, ML algorithms to transform raw data into actionable knowledge.

1.1.1 The Role of Data-Driven Approaches

Data-driven decision-making has become a fundamental aspect of business strategy, allowing organisations to base their decisions on empirical evidence rather than intuition. Statistical analysis plays a crucial role in this process, providing methods for identifying trends, making predictions, and quantifying uncertainty. Techniques such as regression analysis, hypothesis testing, and clustering enable businesses to segment markets, optimise pricing strategies, and forecast demand with greater precision [45]. By leveraging structured and unstructured data, businesses can improve performance in areas such as financial analysis, supply chain management, and human resource optimisation.

Furthermore, the rise of big data analytics has enabled companies to process and analyse massive datasets that were previously too complex to handle with traditional statistical techniques. Cloud computing, distributed storage, and real-time data processing have made it possible to capture, store, and interpret data at an unprecedented scale. As a result, organisations can uncover hidden correlations and anticipate market shifts, leading to more agile and informed business strategies [259].

ML extends traditional statistical methods by enabling systems to learn from data and improve performance without explicit programming. For example, supervised learning techniques, such as decision trees, support vector machines, and neural networks, allow businesses to predict customer churn, detect fraud, and recommend personalised products based on historical data [267].

1.2 New Technologies

The rapid advancement of technologies, especially those of new reality formats (XR, with X being a placeholder for any form of extended reality [309]), has significantly impacted various sectors, offering transformative opportunities in business and daily life. XR encompasses a spectrum of immersive technologies, including VR, Augmented Reality (AR), and Mixed Reality (MR). While AR overlays digital elements onto the real world and MR blends physical and virtual environments interactively, this work focuses specifically on VR.

Experts from universities, consultancies, and public authorities recognise that immersive technologies and Virtual Environments (VEs) have the potential to revolutionise society by enhancing education and training. Indeed the *European Commission* emphasises the role of virtual worlds in education, stating that these technologies can increase the efficiency of training at a lower cost and produce better results in areas such as soft skills training [112]. Furthermore, *Deloitte* highlights the benefits of immersive

learning, noting that VR can improve learner engagement and create interactive experiences that enhance knowledge retention and application [89]. Additionally, the *XR Association* points out that extended VR is a powerful tool for workforce development, offering immersive training that replicates real-world scenarios, thereby democratizing and enriching education [385]. These perspectives underscore the consensus among experts regarding the transformative potential of immersive technologies.

Industry reports highlight the substantial economic impact of immersive technologies. For instance, *PwC* estimates that VR could contribute nearly \$300 billion to the global economy by 2030, with companies increasingly adopting VR for employee training, customer engagement, and process optimisation [83]. Moreover, 82% of European businesses that have integrated VR into their workflows consider digital training programs to be a key next step [371].

Specifically, the global VR market is experiencing rapid growth. According to *Grand View Research*, the market size was valued at approximately \$59.96 billion in 2022 and is expected to reach \$435.36 billion by 2030, growing at a compound annual growth rate of 27.5%. Major companies such as *Meta*, *Apple*, and *Nvidia* are heavily investing in VR development, signalling its significance in the market. The metaverse, which is defined in [160] (p. 1) as being a “computer-mediated environment consisting of virtual ‘worlds’, within which users can act and communicate in real-time using virtual people”, is also emerging as a transformative space for businesses, enabling new consumer experiences, digital collaboration, and innovative marketing strategies.

In this context, VE-based education is considered by experts as strategic topics with significant societal impact. VR’s potential extends to various business areas and psychology, offering realistic, safe, and controllable simulations for research and education, as well as new 3D-enriched consumer experiences and services.

1.2.1 Virtual Reality

Definition, Origin and Evolution

Virtual Reality refers to an immersive, computer-generated environment typically experienced through a head-mounted display (HMD), where users can interact with 3D content in real-time and experience a strong sense of presence [341, 260]. In this context, VR relies on technologies such as stereoscopic displays, spatial audio, motion tracking, and haptic feedback to simulate sensory input and respond dynamically to the user’s movements. While the concept encompasses a wide range of systems, this work focuses specifically on immersive VR using HMDs for training and public speaking applications. However, a broader overview of VR technologies

and their historical evolution is presented in the introduction to provide context.

The concept of VR dates back to the 1960s, with pioneers like Morton Heilig, who developed the Sensorama [158], and later Ivan Sutherland, who created the first head-mounted display (HMD) [354]. VR gained momentum in the 1990s, notably with the work of Milgram et al. Reference [260], but technological limitations, such as low processing power and limited tracking accuracy, initially hindered public adoption. However, the field has experienced a rapid evolution over the past decade due to advancements in hardware and software. The resurgence of interest in VR began in the 2010s, driven by the introduction of powerful and affordable headsets such as the Meta Quest and HTC Vive. As explained in [382], in 2016, VR reached Gartner's *slope of enlightenment* in the Gartner Hype Cycle for emerging technologies after being stuck in the *trough of disillusionment* [222]. Thanks to continuous technological progress, VR systems now offer higher-resolution displays, lower latency, and more precise motion tracking, making virtual experiences more immersive and realistic than ever before.

Mechanisms

VR systems operate by combining multiple technologies to create an immersive and interactive experience. HMDs are central to VR setups, featuring dual high-resolution screens positioned close to the user's eyes to create a stereoscopic 3D effect. To ensure a sense of depth and realism, modern HMDs optimise graphical performance by rendering high-resolution details only where the user is looking. Motion tracking plays a crucial role in VR immersion, allowing users to move naturally within a virtual space. This is achieved through inside-out tracking, which uses onboard cameras and sensors to map the user's surroundings, or outside-in tracking, where external sensors (e.g., lighthouse systems in HTC Vive) track movements with high precision.

Beyond visual immersion, spatialised audio enhances the experience by simulating realistic sound propagation and adjusting volume and direction based on the user's head position. Interaction in VR is enabled through hand controllers, full-body motion tracking, or even hand-tracking technology, which allows direct manipulation of virtual objects without physical controllers.

Types of Systems

VR systems can be categorised based on both the hardware used and the type of virtual environment provided. Broadly, VR setups fall into two main categories: CAVE systems and VR headsets. A Cave Automatic Virtual Environment (CAVE) consists of multiple large projection screens arranged

around the user, creating an immersive experience where movement is possible within a controlled physical space, and interaction is facilitated through motion capture, controllers, or hand-tracking technology [81]. On the other hand, VR headsets offer a more personal and flexible immersive experience by placing screens directly in front of the user's eyes. VR headsets can further be classified based on their level of autonomy and required hardware. First, smartphone-based VR (see Figure 1.2a), such as Google Cardboard, which relies on a mobile device inserted into a headset and offers only basic immersion with limited tracking and interaction capabilities [179]. VR headsets can generally operate in two modes: PC-tethered and stand-alone (see Figures 1.2b and 1.2c). In PC-tethered mode, devices such as the *Meta Quest*, *HTC Vive*, *Pico Neo* and *Valve Index* connect to a powerful computer to deliver high-quality graphics and full interactivity with precise motion tracking. These setups sometimes rely on additional sensors, like lighthouses, to enhance environmental tracking and ensure accurate detection of user movements and spatial positioning [342]. In stand-alone mode, headsets function independently with integrated processing and tracking capabilities, offering greater mobility and accessibility while maintaining a high level of immersion. In addition to hardware distinctions, VR experiences can also differ in terms of the virtual environment they provide. The first type, 360° video-based VR, uses pre-recorded spherical videos captured from real-world locations, allowing users to explore the scene visually but without physical interaction beyond changing the viewpoint [343]. The second type consists of fully virtual environments, which are entirely computer-generated and allow users to navigate, interact with objects, and experience dynamic changes in the environment. These environments are typically built using game engines such as Unity or Unreal Engine, enabling high levels of realism and interactivity. The rapid evolution of both VR hardware and software has significantly expanded the possibilities for research and applications. Advancements such as real-time motion tracking, integrated eye-tracking, and physiological sensors have made it possible to collect precise behavioural data within VR environments, allowing for a more detailed analysis of user engagement, cognitive load, and emotional responses. As a result, VR is becoming an increasingly powerful tool not only for immersive experiences but also for scientific research, training, and behavioural analysis, enabling new opportunities for studying human interaction and performance in controlled virtual settings.

Research Opportunities

As previously stated, with the rapid evolution of VR technology, an increasing number of tools are available to record and analyse user behaviour within virtual environments. Eye-tracking technology embedded in modern headsets allows for the precise measurement of gaze patterns, enabling

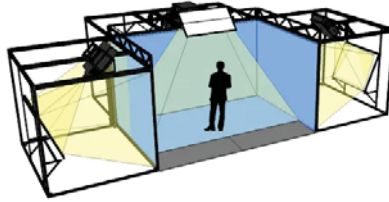


Figure 1.1: CAVE



Figure 1.2: Different types of VR headsets

researchers to study user attention, cognitive load, and engagement in real-time. This is particularly useful in fields such as marketing, psychology, and training simulations. Additionally, facial expression tracking is being integrated into high-end VR systems, providing valuable insights into emotional responses.

Another critical advancement is physiological sensing, which enables VR systems to be upgraded with additional objects to record biometric data such as heart rate, skin conductance, and pupil dilation, offering deeper insights into user reactions. This is particularly relevant for research on stress, emotional engagement, and cognitive load in immersive experiences.

Moreover, the ability to collect motion-tracking data in real-time has opened new possibilities for research. By analysing body posture, gestures, and locomotion patterns, VR can provide objective measures of user behaviour.

These advancements significantly expand the scope of research possible in VR. The integration of AI and ML allows for real-time data analysis, enhancing adaptive training environments that respond dynamically to user behaviours. As a result, VR has become a powerful tool for studying human behaviour, improving training simulations, and developing personalised virtual experiences.

Applications

As digitalisation continues to enhance business performance [313], VR presents new opportunities across various sectors. Major corporations are incorporating VR into their strategies, with Meta's (formerly Facebook) Metaverse project serving as a prominent example [182]. VR is particularly valuable when human behaviour is a crucial element of the business process, leading to increased activity in marketing and psychology research. For instance, reference [227] analysed 150 papers across 115 journals on the use of VR in marketing, highlighting its growing significance. Additionally, VR is now commonly used in sectors such as tourism [24, 181] and retailing [34]. Beyond business, VR has seen significant growth in the health sector [172, 171, 29] and education, particularly in training applications [372, 148].

To ensure effective application, VR environments must be truly immersive and accurately replicate desired situations. However, challenges such as barriers to adoption and acceptance persist [213, 242]. Many research papers focus on applications using simple and well-established approaches like 360-degree videos, which may not fully exploit VR's potential. In VR, the concepts of presence and immersion are crucial, and their measurement is essential for validating environments [381, 289]. Furthermore, research exists on enhancing presence and immersion [176, 175] and on stimulating additional emotions [120, 74, 130, 264].

Since human behaviour is critical in most business processes, designing realistic human interactions in VR is essential. The increasing ability to collect fine-grained behavioural and physiological data within VR environments enables more robust research, paving the way for future innovations and deeper insights into human interaction in virtual spaces.

1.3 Communication

PS is among the most critical competencies in professional and managerial contexts, essential for tasks such as pitching a product to investors, engaging with customers and collaborators, or leading a team. Despite its societal significance, PS often provokes apprehension, affects performance [318], and can be a decisive factor in career progression.

However, PS is not an innate talent but a skill that can be developed through training [320, 361, 173]. Research demonstrates that teachers can learn to manage disruptive behaviour more effectively [32], entrepreneurs can enhance the persuasiveness of their speeches [256], and individuals can reduce their stress during presentations [297]. In all these cases, speakers improve their ability to communicate effectively.

PS is a cornerstone of business activities. Sales representatives must confidently present products to customers, tourist guides must engage

groups with compelling narratives, and managers must defend their projects in front of stakeholders. Yet, many organisations report deficiencies in their employees' PS abilities. One major obstacle is PS anxiety—the fear of delivering a speech or presentation due to concerns about negative evaluation or embarrassment [367]. This is one of the most common fears [297] and a primary cause of poor performance in PS [318, 256].

Crucially, PS anxiety differs from general social anxiety, as noted in [32]. While it is common, it is also manageable through repeated training, exposure, and skill development. Training in front of an audience, even in controlled environments, helps speakers gain confidence and improve both emotional regulation and delivery [320, 361, 173, 223]. Given its fundamental role in professional success, organisations and educators must prioritise the development of PS skills, ensuring individuals are equipped to communicate effectively in diverse settings.

1.4 Virtual Reality to Enhance Communication Skills Using Analytics

1.4.1 A Promising Tool for Training

Learning by doing has many benefits, and it is well known that VR has significant potential in this area, as demonstrated in [123, 171, 172, 103, 350]. In the context of PS, references [153, 186, 115] have demonstrated the benefits of VR-based training. As illustrated in Figure 1.3, active learning approaches such as learning by doing are associated with significantly higher retention rates than passive methods. Although not a substitute for real-world training, VR facilitates experiential learning in a highly accessible and controlled manner.

A virtual environment with an interactive virtual audience—i.e., an audience providing non-verbal feedback—has proven to be highly effective in PST [63, 68, 66, 20, 65, 188, 141, 230, 337]. Such environments allow speakers to practice in realistic conditions similar to those they will face in real-life speaking scenarios.

VR enables learners to practice whenever and wherever needed, reinforcing their sense of competence and readiness to deliver speeches [327, 80]. Furthermore, VR-based training can be designed to be progressive, allowing controlled variations in audience size, behaviour, and level of engagement, which leads to more efficient learning processes and faster progression [103].

Another key strength of VR lies in the high transferability of the acquired skills to real-world contexts, especially when the virtual environment closely replicates real-life situations. This fidelity allows learners to adapt and apply what they have practised more effectively than with traditional theoretical training—although real-life practice remains a final and essential

1.4. VIRTUAL REALITY TO ENHANCE COMMUNICATION SKILLS USING ANALYTICS

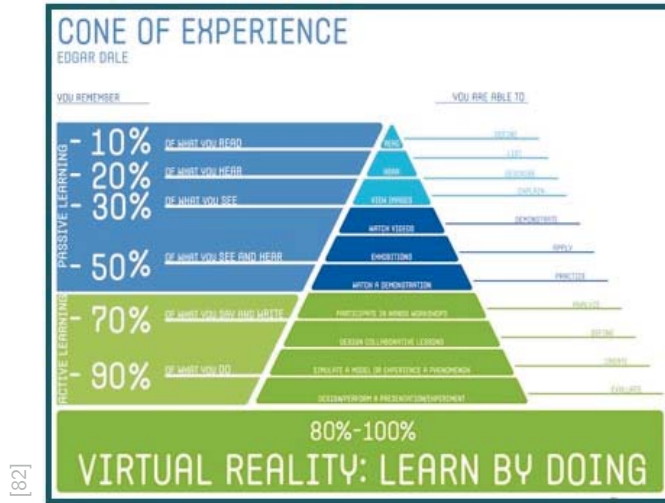


Figure 1.3: Dale's Cone of Experience (adapted), illustrating how active learning methods, such as VR-based training, lead to higher retention and application rates compared to passive methods.

step in skill development.

The interactivity of the virtual audience plays a crucial role in training effectiveness. Indeed, reference [63] demonstrated that VR can improve speaking performances, particularly when the audience reacts dynamically to the speaker's behaviour [384, 143]. Thus, the reactions of the audience significantly impact the speaker's emotions and performance [292, 296, 200]. As shown in [64, 65], interactivity is a key factor in the effectiveness of VR-based PST.

1.4.2 Analytics and Human Behaviour Research

The integration of analytics with VR has opened new avenues for research and training, particularly in fields that require immersive and interactive simulations. By combining VR with AI-driven analytics, researchers can assess human behaviour with a level of precision that was previously unattainable. Real-time tracking of user movements, eye gaze, and physiological responses provides valuable insights into engagement, decision-making processes, and emotional states [380].

In PST, for example, AI-powered speech and gesture analysis can objectively evaluate performance, identifying factors such as vocal intonation, pacing, and body language [65]. Similarly, emotions analysis and Natural

Language Processing (NLP) techniques can assess speaker-audience interactions providing adaptive feedback to improve communication skills. These advancements enable the development of personalised training programs tailored to individual needs, fostering skill acquisition in a controlled and measurable manner.

Additionally, predictive analytics in VR allows businesses to simulate and test different scenarios before implementing them in real-world settings. This is particularly valuable in areas such as product design, user experience testing, and workplace training, where virtual simulations reduce costs and enhance safety.

1.4.3 Limitations and Challenges

While VR presents numerous advantages for PST, it is not without limitations. The effectiveness of VR-based training is still under evaluation, and several aspects require further research.

First, although VR offers a high level of immersion, it does not fully replicate the social dynamics of real-life PS. Factors such as audience unpredictability, emotional reactions, and live interaction remain difficult to simulate accurately.

Furthermore, despite promising results, PST using VR has received relatively little attention in the literature. While some studies confirm that VR can help speakers cope with anxiety and improve their presentation skills, few have thoroughly investigated the impact of virtual audience characteristics on performance evolution. For example, audience realism, responsiveness, and variability are key factors that may influence training outcomes but have not yet been systematically analysed.

Another limitation lies in the erosion effect observed in repeated presentations. In traditional settings, repeated exposure to the same speaking conditions can lead to diminishing engagement and effectiveness. Nevertheless, reference [65] found that VR training helped mitigate this effect, as users adapted their presentation and vocal effort depending on the size of the virtual audience, as they would in real-life situations. Additionally, research suggests that VR training can induce changes in vocal parameters toward a more charismatic voice [65]. However, further studies are needed to confirm the consistency and transferability of these effects to real-world speaking scenarios.

Finally, customisation remains one of VR's greatest advantages but also a challenge. While VR allows for adjustments in audience size, behaviour, and environmental factors, creating fully adaptive and personalised experiences requires advanced AI models and NLP. Currently, many VR-based training systems rely on pre-scripted audience reactions, which limits their effectiveness in simulating spontaneous, real-time interactions.

Despite its limitations, VR remains an exceptional tool for PST. By

immersing users in a controlled and scalable simulation, VR provides a safe environment to practice and refine presentation skills. The flexibility of VR enables personalised training, allowing speakers to gradually increase the difficulty of their sessions and receive tailored feedback. However, to fully unlock its potential, future research should focus on enhancing audience realism, refining real-time interactivity, and ensuring that improvements in VR translate effectively into real-world performance. Addressing these challenges will be crucial for establishing VR as a comprehensive public speaking training tool—this is what this thesis is humbly trying to do.

1.5 Thesis Overview and Research Line

This thesis, conducted within a PhD program in Economics and Management, explores the intersection of VR, analytics, and multimodal communication to enhance PST. Given the increasing reliance on immersive technologies in education, corporate training, and human resources, this research aligns with the multidisciplinary nature of modern business studies. It contributes to both theoretical and applied knowledge by developing methodologies for assessing and improving PS performance using VR and multimodal analytics.

This thesis is structured around a series of research papers, published, submitted, and forthcoming (see Table 1.1 for the list of publications, and 1.2 for the affiliations of co-authors). These papers investigate key multimodal characteristics of effective PS in virtual environments. The research aims to address five critical questions related to VR-based training, including:

1. How are virtual agents' behaviours perceived, and how do they impact the VR-based user?
2. How does VR training compare to traditional methods in terms of effectiveness, engagement, and user experience, and what limitations of current VR training systems can be addressed through technological and pedagogical improvements?
3. What are the key verbal and non-verbal indicators of effective PS in VR?
4. Can emotions be accurately detected based solely on speech?
5. How can ML and multimodal analytics be leveraged to assess and improve PS skills?

To answer these questions, this research employs a combination of systematic literature reviews, empirical studies, and ML techniques. The methodology integrates quantitative and qualitative analyses, experimental

design, and advanced data processing to extract insights from VR-based PS sessions.

This thesis contributes to the broader discussion on the effectiveness of VR for skill development while advancing business analytics methodologies. By focusing on PST, it underscores how emerging technologies can be used to enhance critical workplace competencies, ultimately providing valuable insights for researchers, educators, and industry practitioners.

1.5.1 Thesis Structure

This dissertation is structured into multiple research studies, each addressing a key aspect of VR-based PST. First, Chapter II provides a comprehensive literature review on the perception of Intelligent Virtual Agent (IVA) behaviours, examining how they can influence speaker by emphasizing realism, responsiveness, and interactivity. Next, Chapter III focuses on the role of non-verbal cues in virtual agents, exploring how gestures, facial expressions, and body language contribute to the perception of presence and immersion in VR-based training scenarios. Building on this, Chapter IV investigates the application of VR in legal education, particularly in courtroom simulations, and assesses its effectiveness in developing advocacy and argumentation skills among law students compared to traditional training methods. Furthermore, Chapter V delves into the automatic assessment of PS cues (including verbal, para-verbal, and non-verbal behaviour) in order to provide objective feedback. In addition, Chapter VI introduces a dataset of emotionally validated audiovisual expressions, specifically designed to improve the detection and analysis of speaker emotions, thereby enhancing real-time feedback mechanisms in virtual training environments. Subsequently, Chapter VII presents a set of corpora collected from VR-based PS sessions, offering a rich dataset for studying speech dynamics and the impact of training conditions on speaker performance. Finally, Chapter VIII investigates predictive models for assessing PS performance, exploring how ML techniques can be applied to multimodal features in order to predict the performance of a presentation.

This research advances our understanding of VR as a tool for PST, addressing both its strengths and limitations. This thesis aims to bridge the gap between technological innovation and practical skill development, contributing to the broader fields of business analytics, education, and human-computer interaction.

1.5. THESIS OVERVIEW AND RESEARCH LINE

Chapter	Title	Authors	Type	Number of pages	Status	Venue	Ranking
II	A Systematic Review on the Socio-affective Perception of IVAs' Multimodal behaviour	ETIENNE Elodie RISTORCELLI Marion SAUFNAY Sarah QUILEZ Aurélien CASANOVA Rémy SCHYNS Michaël OCHS Magalie	Conference paper	11	Published in 2024	Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents	B (ERA)
III	Perception of Avatars Non-verbal Behaviours in Virtual Reality	ETIENNE Elodie LECLERCQ Anne-Lise REMACLE Angélique DESSART Laurence SCHYNS Michaël	Journal Paper	18	Published in 2023	Psychology & Marketing	B (HEC)
IV	A framework for the implementation of virtual reality in legal education: A mixed methods and multiple case study investigation	CHO Justin ⁸ ETIENNE Elodie ¹ SCHYNS Michaël ¹ JUNG Timothy ⁹	Journal paper	20	Submitted in 2025	Studies in Higher Education	B (HEC)
V	Automatic Assessment of Multimodal Cues for Public Speaking Training in Virtual Reality	ETIENNE Elodie RISTORCELLI Marion PERGANDI Jean-Marie CASANOVA Rémy SCHYNS Michaël OCHS Magalie	Conference paper	9	Submitted in 2025	17th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS2025)	A1 (Qualis)
VI	EVE: Emotional Validated Expressions, an acted audiovisual corpus	ETIENNE Elodie REMACLE Angélique LECLERCQ Anne-Lise SCHYNS Michaël	Conference paper	5	Submitted in 2025	INTERSPEECH2025	A1 (Qualis)
VII	Corpora of Public Speaking in Virtual Reality	BETTAHI Lamia REMACLE Angélique SCHYNS Michaël ETIENNE Elodie LECLERCQ Anne-Lise (Uliège) RISTORCELLI Marion PERGANDI Jean-Marie ETIENNE Elodie SCHYNS Michaël CASANOVA Rémy OCHS Magalie (AMU)	Journal Papers	20 (Uliège) Unknown (AMU)	To be submitted in 2025	Computer in Human Behaviour (Uliège) Unknown (AMU)	B (HEC) Unknown (AMU)
VIII	Public Speaking Performance Prediction Using Machine Learning	ETIENNE Elodie RISTORCELLI Marion SCHYNS Michaël OCHS Magalie	Journal paper	Unknown	To be submitted in 2025	Unknown	Unknown

Table 1.1: Summary of publications and contributions in peer-reviewed venues

Co-author	Affiliation(s)
BETTAHI Lamia	RUCHE, University of Liège, Belgium
CASANOVA Rémy	ISM and CNRS, Aix-Marseille University, France
CHO Justin	Business School, Manchester Metropolitan University, United-Kingdom
DESSART Laurence	Marketing and Service Innovation, University of Liège, Belgium
JUNG Timothy	Business School, Manchester Metropolitan University, United-Kingdom School of Management, KyungHee University, South Korea
LECLERCQ Anne-Lise	RUCHE, University of Liège; Belgium
OCHS Magalie	LIS and CNRS, Aix-Marseille University, France
PERGANDI Jean-Marie	ISM and CNRS, Aix-Marseille University, France
QUILEZ Aurélien	LIS and CNRS, Aix-Marseille University, France
REMACLE Angélique	RUCHE, University of Liège, Belgium
RISTORCELLI Marion	LIS and CNRS, Aix-Marseille University, France
SAUFNAY Sarah	QuantOM, University of Liège, Belgium
SCHYNS Michaël	QuantOM, University of Liège, Belgium

Table 1.2: Affiliation of co-authors

Chapter II

A Systematic Review on the Socio-affective Perception of IVAs' Multi-modal behaviour

This chapter presents a collaboration done from February 2024 to April 2024 to prepare for my upcoming (at that time) research stay at the Aix-Marseille University. The co-authors are Marion Ristorcelli, Sarah Saufnay, Aurélien Quilez, Rémy Casanova, Michaël Schyns, and Magalie Ochs. It is now a paper, for which I am the first author, that has been published in *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents* in 2024. In this thesis, the paper is presented as originally published, and I apologise for any outdated information, repetitions, or formatting inconsistencies with other chapters.

Abstract

The multimodal behaviour of IVAs may convey different socio-affective dimensions, such as emotions, personality, or social capabilities. Several research works show that factors may impact the perception of the IVA's behaviour. This paper proposes a systematic review, based on the PRISMA method, to investigate how the multimodal behaviour of IVAs is perceived with respect to socio-affective dimensions. To compare the results of different research works, a socio-emotional framework is proposed, considering the dimensions commonly employed in the studies. The conducted analysis of a wide array of studies ensures a comprehensive and transparent review, providing guidelines on the design of socio-affective IVAs.

2.1 Introduction

Intelligent Virtual Agents (IVAs) are able to display a wide range of multimodal behaviours to interact naturally with the users in a virtual environment, whether in 2D or 3D. Depending on the role endowed by the IVAs (e.g., virtual guide [31], virtual recruiter [51], or virtual patient [277]), they may be required to convey different socio-affective states such as a dominant attitude, positive emotions, or a behaviour that inspires confidence. Several research works show that both the verbal and the non-verbal behaviour of the IVA strongly impact the users' perception of the IVA's socio-affective state. For instance, IVA may display different facial expressions such as smiles, frowns and raised eyebrows to convey a specific emotional state during the interaction [31, 299, 300, 207, 140, 51, 170, 169, 67, 90, 111, 244, 323, 215, 275, 351, 374]. The IVA can also direct or avert its gaze from the interlocutor to display different levels of engagement [92, 93, 207, 140, 51, 67, 90, 111, 244, 272, 323, 351]. The other modalities of non-verbal behaviours are the head movements (shaking, nodding, or tilting) [67, 111], the orientations of the head (downward, upward, inclined or straight) [51, 395], the torso positions (leaning forwards, backwards or straight) [244], the arms positions (e.g., crossed arms, arms behind the head, hands on the hips), and the body position (e.g., standing or sitting on a chair) [31, 316, 351, 272]. Furthermore, the breadth of movements (from small to large radius), their intensity (light, moderate, or forceful) [31], and the directionality of gestures (pointing with a finger or an open palm) further enrich this non-verbal vocabulary to convey socio-emotional state [51, 395]. Hand gestures also play a crucial role, ranging from central positioning to peripheral actions, and alternating between behaviours like fiddling with hands, and shrugging [31, 272], or hands clasped on a table.

Regarding the verbal modalities, there exists a variety of linguistic and paralinguistic signals. Linguistic signals pertain to the elements of language and communication that can be explicitly observed and analysed. These include the choice of words, the structure of sentences, grammar, and syntax. Some specific examples are the use of subjective pronouns, verbs, nouns, the level of formality of the language, the use of self-references, the variation of vocabulary, the length of the sentences, the use of positive or negative contents [31, 25, 51], and the use of commanding or suggesting sentences [395]. Paralinguistic signals, although closely related to linguistic ones, extend beyond the basic grammar and syntax to include aspects of meaning that are not directly encoded in the linguistic elements. These include acoustic signals such as the pitch and the speech rate [121], the behaviour impacting the flow of the conversation, as for example the overlapping of speech when the participant interrupts the IVA [132]. Another important aspect is the alignment or coordination of the behaviours of the two interlocutors, considering, for instance, the sequentiality and temporality of signals [92,

275] to develop IVAs capable to display different attitude variations or to adapt to the user's perception, as proposed in [31].

Furthermore, several factors can have an impact on the socio-affective perception of the IVA's behaviour, some related to the IVA (e.g., its appearance [85, 93, 170, 244]), others related to the user (e.g., age) [170].

Although there are many studies on this subject, it remains difficult to have a clear vision of the different socio-affective behaviours that IVAs can convey and how they can express them. The objective of the paper is precisely to *provide a systematic review, based on the PRISMA method [378, 249], of the research works investigating the users' perception of the socio-affective dimensions of IVAs conveyed through their multimodal behaviours.*

Researchers have explored a wide range of socio-affective dimensions that IVAs may convey through their behaviour. In order to compare the results of different studies, the socio-affective dimensions have been gathered into five categories: emotions, personality, trustworthiness, social capabilities, and believability. The *emotions* category groups the perceptive studies on the users' perception of the emotions or the moods expressed by the IVAs. The *personality* category includes the studies on the users' perception of the IVAs' personality traits such as friendliness, dominance, and extroversion. The *trustworthiness* category gathers perceptive studies on the impact of IVAs' behaviour on the perception of trust, including the perception of competence, intelligence and cooperativeness. The *social capabilities* category covers studies on the perception of a virtual relationship for the creation of a social connection with the user or between IVAs [145]. The last category is dedicated to the perceived *believability* of the IVAs. Of course, the proposed socio-emotional framework is subject to scrutiny. The categories are closely linked to each other, e.g., emotions are influenced by personality [251, 97], some personality traits are sometimes considered as emotional dimensions [323]. However, this socio-emotional framework - constructed by grouping the socio-affective dimensions mentioned in the papers according to their proximity in terms of definition and use - enables a comprehensive and transparent review of the existing works presented in this paper.

This systematic review is guided by a central research question: *How are the emotions, personality, trustworthiness, social capabilities, and believability of IVAs perceived by users through their multimodal behaviour?* Indeed, the objective is to identify more precisely the signals that IVAs can use to convey these socio-affective dimensions, but also the importance of each modality depending on the considered socio-affective dimension, and the effects of the combinations of signals on perception. Moreover, the aim is to highlight the different factors - related to the IVA, to the user, and to the interactive device - that may influence the perception of the user.

This paper is structured as follows. The next section (Section 2.2) outlines the approach, based on the PRISMA method [378, 249], to perform the systematic review. Subsequently, Section 2.3 describes the results of the

papers included in the systematic review on users' perception of the five socio-affective categories introduced above through the IVAs' multimodal behaviour. Finally, Section 2.4 delves into the discussion of the variability in the perception of multimodal behaviour influenced by various factors.

2.2 Methodology

The PRISMA method [378, 249] is used to guide the systematic review process. This method ensures transparency and rigour by systematically identifying, assessing, and evaluating relevant literature. The focus is on the perceptual review of how the verbal and non-verbal behaviours of IVAs influence the perception of emotions, personality, trustworthiness, social capabilities, and believability. The criteria applied in the selection process include studies on the perception of the socio-affective dimensions of at least one IVA depicted in virtual environments based on monitor or VR.

Thus, the literature search conducted in the Scopus database¹ uses the following request:

```
(virtual AND (audience* OR avatar* OR agent* OR listener*  
OR character*))  
AND (perception OR perceive*)  
AND (behaviour* OR behavior* OR "body language" OR  
arousal  
OR valence OR stance OR attitude*)  
AND (nonverbal OR non-verbal OR verbal)  
NOT patholog*  
NOT autism.
```

This search aims to identify papers exploring the perception and behaviours of IVA, excluding those related to pathology and autism.

The search process retrieves a total of 162 papers in Scopus. Exclusion criteria are defined to refine the search results, ensuring relevance and focus on the research question. The criteria encompass the exclusion of papers with accessibility issues, duplicates entries, non full papers or papers without any DOI, papers that are not a perceptive study. Furthermore, as the focus is on the perception of IVA using multimodal behaviour, papers in which the IVA does not have a face or a humanoid body are removed. The same happened for papers that present neither verbal nor non-verbal behaviour for the IVA. With the focus on virtual environments using monitor or VR, papers not using monitor or VR are excluded. Finally, the emphasis is placed on adult Occidental participants. Consequently, all studies involving participants with a mean age below 18 years old or non-Occidental

¹In fact, the search was also performed using Web of Science, resulting in 98 papers. However, since these papers were already included in Scopus, they were not utilised.

Table 2.1: Reasons and numbers of papers rejected during the PRISMA process.

Reason	Step 1	Step 2
Accessibility issue	0	4
Duplicate entry	1	1
Abstract or poster	2	0
Paper without DOI	11	0
No perception task	25	18
Perception of human agents	0	1
Perception of embodied avatars	9	2
Perception of robots	7	2
Faceless agents	0	5
Non-humanoid agents	2	2
No non-verbal or verbal behaviours	0	6
No monitor or VR	4	0
Participants with mean age under 18 years old	1	2
Non-Occidental participants	3	14
Pathological participants	8	0
Total	73	57

participants are also excluded².

The first step of the PRISMA method consists of selecting articles by reading their title and abstract. From the 162 articles identified using the query presented above, the initial screening leads to the exclusion of 73 papers. Subsequently, a more thorough examination of the remaining papers is conducted, involving reading through each in its entirety, which constitutes step 2 of the screening process. After this comprehensive review, an additional 57 papers are excluded. See Table 2.1 for more details.

This methodical filtration underscores the rigorous selection and exclusion criteria inherent in the PRISMA approach, ensuring that the review focuses on the studies most relevant to the research question. Table 2.2 proposes a summary of the dimensions, sub-dimensions, and signals involved as dependent variables for each selected paper. In the next section, the results of the paper for each socio-affective category are described in more detail.

²For studies conducted in Occidental universities or research groups, the absence of such participants cannot be definitively confirmed when full details were not given.

2.3 Results

2.3.1 Emotions

In this section, the research works investigating the perception of IVA's emotions through their multimodal behaviour are reported. Initially, the literature reveals that emotional dimensions are explored through different affective concepts such as mood and emotions [7]. To compare the research results, the analysis of the research works on the emotional perception of IVA behaviour is proposed in light of the *valence* and *arousal* dimensions.

These dimensions are expressed and perceived in various ways, for example, through posture and facial expressions [189]. According to the authors, the SAM questionnaire, which studies the dimensions of valence, arousal, and dominance, is a good tool to assess the perception of these dimensions. However, dominance, which could also be considered as a personality trait, is discussed in Section 2.3.2.

According to [67], the notion of valence refers to the IVA's opinion, *i.e.*, the positive or negative feelings it has towards the user. As shown in several research works [275, 215, 90, 67, 140, 111], the emotional state, and more specifically valence, is conveyed by facial expressions to display basic emotions such as anger, happiness, and sadness [215], but also more subtle emotional states such as stress [90], amusement, or politeness [275]. Even if facial expressions are a strong cue of emotional state, the way they are displayed in IVAs can vary considerably from one study to another, leading to misinterpretation depending on the considered representation, the intensity of the expression, and the combination of modalities [111, 215]. Although it seems that the frown conveys a negative valence [106, 90], the way it is represented in the IVA could vary, particularly in terms of intensity and of the parameters used to display facial expression. For example, in the study of [111], a smiling facial expression is perceived as neutral, while it is generally a positive valence signal, indicating a potential confusion between a fake smile and a genuine smile [275, 111]. Valence is mainly conveyed by the combination of facial expressions and head movements. Indeed, researchers consistently associate smiling and nodding with positive valence, and frowning and shaking the head with negative valence [67, 140, 111, 90]. However, some signals may be predominant in the perception of valence. Indeed, as shown in [111], head shake is the most negative modality identified and is always judged negatively, regardless of the other modality it is associated with. Regarding the other signals, the literature is not consensual. For example, posture, such as crossed arms, may convey a negative valence as shown in [67, 111], while [140] does not find no impact of posture on the assessment of valence. To express the valence, some modalities may be more important than others. For instance, as highlighted in [67], to assess the IVA's valence, the users generally first consider the

head movements, then the posture, the gaze direction, and finally the facial expressions, as this signal is more subtle in VR. Additionally, research on the perception of behaviour on the valence highlights the importance of combining non-verbal behavioral modalities. For example, nodding is mainly a sign of positive valence but is sometimes considered as neutral based on the associated signals [111]. In the same way, an IVA with the head tilted, its elbow on the table and its torso leaning forward appears to be positive, while separately, these signals convey a neutral or negative valence [111]. The perception of the non-verbal modality may also vary according to the associated verbal behaviour, as highlighted in [275]. In addition, vocal non-verbal immediacy, corresponding to the pitch level and speech rate, may also have an impact on the assessment of participants' affect towards the content or the virtual model. It also impacts the likelihood of following the same virtual instructor again for other similar videos in the future [121]. Indeed, stronger vocal immediacy, characterised by an average pitch of 260 Hz and a speech rate of 133 words per minute (wpm), enhance affective learning compared with a virtual model that uses weaker vocal immediacy, with an average pitch of 115 Hz and a speech rate of 119 wpm [121]. Some factors may influence the perception of valence, such as the appearance of the IVA itself. Although few studies examine this factor, some research shows that a female IVA is perceived as more positive than a male IVA when smiling [275]. Another study also shows that for a given facial expression, such as raising eyebrows, a virtual female IVA appears to be more stressed if it uses a frontal body with direct gaze than if it used an averted body with an averted gaze, whereas no difference is found between these behaviours for the male IVA [90].

The dimension of arousal describes the excitement of the event for a person and ranges from low to high alertness [393]. It takes the form of an IVA interested or attentive to what people are saying and is characterised by two types of non-verbal behaviour, namely proximity and body movements [393]. Consistently, all the authors who study this dimension find a relationship between the evaluation of arousal and the direction of gaze, the frequency of movements, and the proximity of posture [67, 140, 111]. A high level of arousal or engagement is associated with an IVA looking at the speaker, with a closer posture, *i.e.* a torso leaning forward and frequent head movements and facial expressions. On the contrary, an IVA that looks away with a more distant and relaxed posture is perceived as disconnected [67, 140]. It is also interesting that eyebrow raising is judged to be neutral in terms of arousal and that head shaking is always associated with high arousal, regardless of other signals [111]. Another interesting point is the relationship between perception of valence and arousal. The combination of the valence-arousal pair can convey a specific attitude, and, as indicated by [140], users are able to perceive different social attitudes based on these dimensions (indifferent, critical, and enthusiastic). However, several au-

thors report that users do not distinguish between different levels of valence (positive and negative) for low arousal [67, 140, 111].

2.3.2 Personality

In this section, three personality traits are considered: the *dominance*, the *extroversion*, and the *friendliness*. The dominance is part of the PAD (Pleasure-Arousal-Dominance) model, described by [319]. The dominance corresponds to a scale ranging from the absence of control or impact on the event to the feeling of influence or control on the situation [393]. Extroversion and friendliness are two components of the "Big Five" personality traits model [167] that identifies Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. As explained in [93], agreeableness is an indicator of friendliness. Warmth may also refer to friendliness [208]. Similarly, the concept of politeness involves the dimension of friendliness, as explained in [396, 395]. For the sake of clarity, only the term *friendliness* is used in this section. According to [7], extroversion corresponds to the sociability of IVA perceived by users.

The choice of only keeping dominance, extroversion and friendliness to express personality is explained by the interrelationships between them. Indeed, reference [25] shows that dominance and extroversion are positively correlated, while dominance and friendliness are negatively correlated. In addition, friendliness and dominance are two dimensions that appear in the Interpersonal Circumplex model [208], which further indicates their strong relationship.

To express dominance, an IVA can look at the user in various ways: by looking from below, which is perceived as less dominant, from above, or by aligning its eyes close to those of the participant [25]. Furthermore, as shown by [316, 351], dominant non-verbal behaviours can be displayed using akimbo posture, crossing arms, sagittal head up, gesture with large radius. On the contrary, submissive non-verbal behaviour includes neck-adapter (self-touch), arms open, sagittal head down, gesture with small radius.

Regarding non-verbal signals, the more dominant position is adopted, the more dominant the IVA is perceived [316]. Furthermore, reference [351] shows that, concerning the perception of dominance, crossing arms and the akimbo position are the most effective gestures, but taking up more space (gesture with a large radius) is not perceived as more dominant than the other gestures. Furthermore, references [93] and [92] demonstrate that by using more (or less) dominant cues, an increase (or decrease) in perceived dominance can indeed be implied. As shown by [25], the more linguistic friendly cues are used, the less dominant is the IVA perceived. It is further established by [51] when using verbal modalities of friendliness. Moreover, regarding vocal modalities, reference [132] shows that the longer

the interruption handling time, the more dominant the IVA was perceived.

To express extroversion, IVAs can exhibit positive emotional states and medium arousal levels (see Section 2.3.1), and high dominance behaviours (closer interaction distances, expansive gestures, rapid gesture execution, sustained eye contact, and prolonged gaze duration). In contrast, introverted agents display negative emotional states, low arousal, more distant interactions, reduced spatial extent in gestures, slower gesture speeds, and averted gazes [85, 323].

As shown by [25], the more dominant gaze cues are used, the less extroverted the IVA is perceived. Generally, it is possible to distinguish extroversion from introversion by voice or facial expression alone [348]. However, if the voice is combined with body movements, it is the most informative signal to judge the extroversion of the virtual agent [348]. More specifically, reference [323] shows that a virtual agent is considered to be extroverted if it has a positive emotional state, a medium level of arousal (see Section 2.3.1), and a positive dominance value. On the contrary, an introverted virtual agent has a negative emotional state, a low level of arousal, and a low dominance value. At the behavioural level, extroverted personality translates into a closer distance during the interaction (e.g., the torso leaning forward), a higher value of spatial extent during the execution of gestures, a higher speed of execution gestures with greater eye contact, and a longer duration of gaze for an extroverted virtual agent. On the contrary, an introverted virtual agent is more distant during interaction (leaning back) and adopts a lower value of spatial extent and speed of gesture execution, with an averted gaze [85].

To express friendliness, regarding verbal modalities, an IVA can use fewer synonyms and negations, shorter sentences, more pronouns, verbs, informal language, positive content, and references to the speaker [31].

Regarding non-verbal modalities, the more dominant positions are adopted, the less friendly the IVA is perceived [316]. As expected, IVA friendliness can be perceived through positive facial expression (e.g., smile [275]) but specific gestures, such as commanding the user through arm gestures (using finger pointing) can have quite the reverse effect [396]. However, reference [51] shows that non-verbal behaviour or verbal modalities of friendliness alone are not sufficient to render friendliness, suggesting that verbal modalities must be added to ensure the right perception of it. Furthermore, reference [396] shows that, to maintain a certain degree of friendliness in the IVA, it is preferable to use suggestion rather than command. In addition, one may prefer stronger vocal immediacy (*i.e.* higher pitch and faster speech rate) to depict a more friendly IVA [25]. Lastly, regarding vocal modalities, reference [132] shows that the longer the interruption handling time, the less friendly the IVA is perceived. Moreover, reference [92] and [93] show that using more (respectively less) dominant cues and/or less (respectively more) friendly cues imply a decrease (respectively an increase)

of perceived friendliness. Similarly, reference [31] shows that the model that uses adaptive algorithm (Reinforcement Learning) to adapt to user impressions can indeed increase the perceived degree of friendliness compared to when IVA does not adapt its behaviour to user reactions.

2.3.3 Trustworthiness

Trustworthiness is a complex dimension, closely related to underlying concepts, such as the performance of the IVA and its inclination to collaborate [150]. In this literature review, the perception of several sub-dimensions of trust is investigated, such as the IVA's *competence*, *intelligence*, *autonomy* and *helpfulness*, reflecting its performance, and its cooperativeness and persuasiveness, displaying a certain predisposition for collaboration with the user.

In the literature, research highlights the impact of IVA's emotional non-verbal behaviour on its attributed *trust* level. In [170], the IVA displays an emotional non-verbal behaviour, either positive (*i.e.* smile, head nod, head nod plus smile) or negative (*i.e.* sad face, head down, dropping the arms plus sad face), while respectively announcing good or bad news. Compared to an IVA that remains neutral in its behaviour, its perceived trustworthiness increases under emotional conditions. This relationship is further studied in [351], more specifically by investigating the impact of emotional behaviours on perceived cooperativeness. The hypothesis assumed by the authors, justifying the adoption of such behaviours to depict cooperativeness, states that this sub-dimension is closely related to IVAs' expressiveness. This hypothesis is confirmed, therefore strengthening the already established link between an IVA's emotions and trustworthiness. The results of the considered study include the preeminence of behaviours combining expressive gestures and mimic to provide the IVA with a cooperative attitude. The impact of lateral head tilts, whatever the side, is positive as well. Naturally, opposite results are found for neutral and non-expressive behaviours, which are associated to lower cooperativeness perception levels, however not as low as averting gazes.

The personality of the IVA, either depicted by its friendly or dominant behaviour towards the user, is also recognised as primordial in shaping perceived competence, trustworthiness, and cooperativeness [272, 316, 31, 92, 351]. A model commonly used in the literature is the *Warmth and Competence* model [118], which links the IVA's friendliness (see 2.3.2) with its competence. This model is adopted by [31] and [272] to assign specific non-verbal behaviours to IVAs depending on the desired competence level. High competence perception is successfully obtained by synchronising the IVA's gestures with the semantic content of the speech [272]. In contrast, IVAs convey low competency through desynchronised gestures [272]. This manipulation effectively transmits the expected competence signal, regardless

of the IVA warmth level, which is characterised by an open (high warmth) or closed gesture (low warmth). However, it is worth mentioning that IVA's warmth, and by extension its friendliness, also has an impact on competence perception, with a positive correlation between these two variables. The accuracy of the warmth and Competence model to design IVAs perceived as such is confirmed in a second study [31].

As previously stated, studies also investigate the interrelation between dominant behaviours and trust sub-dimensions, more specifically considering the impact on competence, intelligence and persuasion [316], as well as on helpfulness [92] and cooperativeness [316, 92, 351]. The effect of dominance is, however, not confirmed for all sub-dimensions. In [316], classical dominant non-verbal signals (*i.e.* akimbo posture, arms crossed, gestures with large radius, sagittal head up) have no effect on competence, autonomy, intelligence, and persuasion ratings. On the contrary, the helpfulness of the IVA appears to be negatively impacted [92]. When it comes to cooperativeness perception, contradictory results can, however, be observed. Despite [316] additionally confirms that no relationship can be established between these two variables, reference [351] finds a negative effect of dominance on cooperativeness, although its effect remains quite small. The latter study additionally identifies specific dominant signals, particularly perceived as uncooperative, such as keeping the arms crossed, which corresponded to the lowest cooperativeness rating among several dominant behaviours, including keeping the head up or making large radius movements. Conversely, the authors identify IVA' behaviours, associated to submission, interpreted as cooperative signals. More specifically, an IVA adopting small radius gestures or keeping its arms open improves its attributed cooperativeness level.

The persuasiveness of IVAs' multimodal behaviour is also investigated. Two studies firstly compare the effectiveness of several strategies in order to persuade the user to join a group of IVAs, either using monitor [396] or a VR headset [395]. In both conditions, similar results are obtained. Strategies adopting a direct approach, explicitly formalising the request, are the most persuasive ones. Actually, when the IVA directly commands the user to move to a specific location, while pointing it with its index, the level of associated persuasiveness is the highest among all possible strategies [396]. In comparison, less commanding approaches, such as politely asking or proposing to the user to join the group, are less persuasive but yet still effective [396]. The authors further highlights the primordial importance of clear formulation of the demand to maximise IVAs' force of persuasion. In addition to the influence of the strategy adopted, communication modalities also play a role. Indeed, multimodal modalities, combining verbal (*i.e.* proposition or command) and non-verbal communication (*i.e.* gaze and hand movements), reach the highest impact levels on persuasiveness [395]. Nevertheless, in [285], results indicate that verbal modality only should be

preferred to maximise the IVA's force of persuasion, reducing distractions for the user, who can therefore be focused on the command itself. An important difference between this study and previous ones that should be considered is that the non-verbal behaviour in this case is not directly related to the command asked of the user, thereby justifying its perceived uselessness.

Generally speaking, the IVA's verbal behaviour also plays a significant role in trust perception. The importance of a realistic voice is highlighted, with IVAs eliciting higher level of trust when endowed with a human voice rather than a synthetic one [285]. Verbal behaviour also turns out to be important in eliciting competence. An interesting signal used to reflect high competence and proved to be relevant is the formulation of sentences using "We" or "You" pronouns, rather than sentences formulated in the first person singular [31]. Similarly, IVAs disclosing personal information are not perceived as more competent [301], further reinforcing the previous statement regarding the irrelevance of "I" pronouns to elicit competence. Although the words used are important, the number of words used has no influence on competence perception [301].

Valuable insights for the design of trustworthy IVAs are provided, but the design of their behaviours should still be carefully considered. Indeed, specific aspects can considerably impact perception, leading to undesired effects. Such negative effects are observed in [374], with mimicking IVAs, evoking a low level of trust and of helpfulness among users when reproducing their behaviour, being even perceived as creepy. The immediacy of the reaction further reinforces the identified negative effect. Another surprising result from the considered papers. Reference [316] identifies that guiding behaviours reduce the perceived level of competence. Such behaviours are represented with deistic gestures and gaze, meaning that IVAs pointing to specific elements are not seen as competent.

2.3.4 Social Capabilities

In this section, the research works exploring users' perception of IVAs' social capabilities are presented. The articles considered in the systematic review identify the following main dimensions related to social capabilities: the *mutual understanding* or *comprehension*, the *mutual agreement*, the *intimacy* or *self-disclosure*, the *interpersonal and dyadic stances* and finally, the *social status*. These dimensions are strongly related to the notion of virtual rapport [145].

The *mutual understanding*, *attention*, *agreement*, *interest* and *pleasantness* are explored in [300] in terms of users' perception of an interaction between two IVAs. In social interaction, intimacy corresponds to "a reciprocal expression of personal or emotional contents, and the perception of positive feelings and comprehension." [298]. In [190], intimacy refers to the perceived self-disclosure of the IVA, which corresponds to the capacity to share

personal information with the user to create social connections. The interpersonal stances in [275] correspond to the perceived relationship expressed by the IVA toward its interlocutor, as for instance cold or warmth. Finally, the dyadic stance results from a behaviour alignment or nonalignment between the two agents reflecting for instance an agreement or hostility [275].

The perception of these social capabilities dimensions described above can be significantly influenced by the multimodal behaviour of the IVA. For instance, as shown in [300] and [275], the mutual reinforcement of smiles between two IVAs or between an IVA and a user, enhances the perception of the *mutual understanding* and impacts the perception of the interpersonal and dyadic stances [275].

Concerning *intimacy*, in [299, 298], the authors show that a high level of perceived intimacy leads to a longer interaction with the user. Moreover, the behaviour of the agent may influence its perceived honesty and genuineness: users rated the IVA as more honest and authentic when it displays behaviours associated with intimacy such as emotional facial expressions (e.g., smiling), open-arm gestures, self-directed motions, head nods and tilts, and eye contact. Moreover, reference [301] investigates the perception of the talkative and self-disclosure dimensions. Results show that an IVA is perceived as more talkative and open, when it uses more words and shares personal information with the user.

A last sub-dimension is the influence of its non-verbal behaviours on its perceived social status. Among the selected papers in this literature review (see Table 2.2), a unique study, reference [274] aims to investigate the effects of different gaze behaviours, reflective of either high or low social status, during a job interview scenario with an IVA. The study finds that an IVA perceived as belonging to a higher social status displays longer duration of eye contact and prolongs stares following the user's responses. Furthermore, an IVA raising his head posture and with a body leaned toward the user is also associated with a higher social status, highlighting the significance of non-verbal modalities in shaping perceptions of social hierarchies within virtual interactions.

2.3.5 Believability

The user's perception may be strongly impacted by the believability of the IVAs influenced by the quality of the animations and of the voice. In this literature review, believability is assessed across several dimensions. In [121], the authors refer to the dimensions of *animacy*, *anthropomorphism* and *liveliness* which correspond to dimensions of the Godspeed questionnaire [18]. The notion of believability corresponds also to the perception of *credibility*, for instance in [51] to explore how credible the IVA is in its role of recruiter. In [207], the believability of the IVA's behaviour is evaluated through the notion of *plausibility* and *naturalness* focusing on both the

behaviour and the appearance.

Several articles in our literature review have shown the influence of IVAs' behaviours on the users' perception of believability. First of all, concerning the animations, references [285] and [207] emphasise that an animated IVA, regardless of the animation mode, is perceived as more natural and more plausible than a static one, indicating that the movements are a key factor in the perception of believability. Moreover, reference [285] shows that a more varied and nuanced behaviour leads to a better perception of the realism. Reference [207] also focuses on the naturalness and plausibility of facial animation behaviour, showing that synthesised expressions (*i.e.* facial expressions generated from data such as audio, head movements, and tagged gaze targets to correspond with expected facial expressions in specific situations) are evaluated as more natural and plausible than tracked expressions (*i.e.* facial expressions captured in real-time from facial movements), both for verbal and non-verbal behaviour. In [216], the authors highlight the importance of the appropriateness of the expressed signals showing that inappropriate nods are perceived as less natural than those adhering to established norms.

In [121], the authors investigate the impact of the verbal behaviour and in particular the vocal immediacy on the animacy, anthropomorphism and liveliness. They manipulate the virtual agent's vocal parameters, such as pitch and speech rate. Results indicate that participants exposed to the condition with stronger vocal immediacy (high pitch and fast speech rate) perceive the agent as more anthropomorphic, animated, and likeable compared to those in the weaker vocal immediacy condition.

2.4 Discussion

In this paper, the interplay of socio-affective perceptions of IVAs by users is delved into, specifically focusing on how emotions, personality, trustworthiness, social capabilities, and believability are discerned through the IVAs' multimodal behaviour. To achieve this goal, the PRISMA method is employed, ensuring a systematic and transparent review process that underpins the analysis.

This exploration reveals a rich tapestry of dimensions used across studies to discuss socio-affective perception. However, a detailed analysis of the dimensions described in the papers shows that the diverse terminology are used to describe similar or overlapping concepts. This diversity poses challenges but also offers an opportunity for synthesis. For example, as explained in Section 2.3.2, several terms can be used to describe friendliness (*e.g.*, agreeableness, politeness, warmth). In this paper, a novel comparison of papers included in this study is proposed by aligning compatible sub-dimensions that reference analogous notions or concepts. This categorisation

is grounded in the definitions of the sub-dimensions, the behavioural models they reference, and the contexts within which they are employed. However, this categorisation can suffer of limitations because of the interplay between dimensions and sub-dimensions. Indeed, for example, the concepts of extroversion and dominance are intimately linked to the valence and arousal, as encapsulated in the PAD model, suggesting that dominance could be considered as a component of extroversion. Similarly, trustworthiness is deeply intertwined with valence and arousal, and warmth significantly influences the competence (warmth and competence model).

In Sections 2.3.1 to 2.3.5, for each dimension separately, the influences of modalities and signals influence on these dimensions are explained. Thus, it is now possible to highlight the main effect of a specific signal on a dimension. For instance, a smile, generally associated with positive valence [111], can lead to ambiguity if not clearly genuine [275], and its impact on friendliness can be reversed by contradictory gestures like authoritative arm movements [396]. Head movements such as nodding, typically signalling agreement [67, 140, 111, 90], can be perceived differently based on accompanying signals [111], highlighting the importance of signal congruence. Similarly, head tilts and orientation underscore the nuanced interpretation of emotional states, where a combination with other positive signals or specific contexts (like a forward-leaning posture) can significantly alter perceptions from negative to positive valence and arousal [67]. For verbal modalities, the linguistic choices made by IVAs, such as the use of inclusive pronouns and the level of formality, play pivotal roles in shaping perceived competence, friendliness and dominance [301]. Sentence length and content tone further influence perceptions of clarity, engagement, and emotions, affecting the user's interaction experience. Paralinguistic vocal immediacy such as pitch and speech rate stands out as a critical factor, with higher immediacy enhancing the IVA's anthropomorphism and friendliness [25]. Interruption handling times also markedly affect perceived friendliness and dominance, underscoring the delicate balance between responsiveness and assertiveness in verbal exchanges [132]. Overall, the synchronisation of verbal and non-verbal modalities, alongside the realism of vocal expressions, emerges as paramount in amplifying the IVA's perceived competence, trustworthiness, and naturalness in communication.

Throughout this paper, it is shown how the interplay and combination of various signals and modalities significantly influence the perception of IVAs, altering and amplifying user interpretations. For instance, a head shake stands out as a strongly negative cue [111], consistently associated with high arousal and negativity, irrespective of other modalities it accompanies. Contrarily, nodding typically signifies positive valence but can be perceived as neutral if conflicting signals accompany it [111]. Similarly, an IVA with a forward-leaning posture and head tilt can project positivity, a perception that might shift to neutrality or negativity when these signals

are isolated. Interestingly, while a smile generally conveys friendliness, commanding gestures like finger-pointing can negate this effect [396]. The synchronisation of an IVA's gestures with its speech notably enhances perceived competence, highlighting the importance of congruence between verbal and non-verbal modalities. Moreover, the combination of voice and facial expressions, especially when aligned with body movements, serves as a powerful indicator of extroversion, underscoring the amplifying effect of multimodal communication. This synchronisation not only boosts the perception of competence [272] but also the persuasive power of the IVA, demonstrating the significant impact of integrated verbal and non-verbal modalities on user perceptions.

The interdependence between socio-affective perception and user characteristics should not be disregarded. Users' age has an influence on trust [170, 285] and on autonomy, persuasiveness and cooperativeness perception [316], with older users who are more likely to put their trust in the IVA. Divergent results are obtained in [351] regarding the impact of age on cooperativeness perception, since no correlation is established between these variables. With regard to competence [272], intelligence [316] and helpfulness [170], no impact of age is identified. The gender of participants also plays a crucial role in the interaction dynamics with IVAs. Indeed, reference [85] shows that women tend to engage more closely and attentively with IVAs than men, who prefer a closer interpersonal distance particularly with female IVAs. Additionally, in [51], the authors show that the IVA, playing the role of a recruiter, is perceived significantly more believable by women than by men. Moreover, the gender of participants affects the perception of an IVA's friendliness, with female participants often perceiving IVAs as less friendly, regardless of the non-verbal behaviours displayed by the IVA [51, 85, 244, 351]. A last factor influencing the perception is the realism. Indeed, both animations quality [216, 207] and voice quality [121] crucially affect user perceptions of VA, with effective verbal and non-verbal behaviours enhancing believability.

In the discussion of the systematic review, it should be noted that only a small fraction of the studies specifically address immersive VR headsets. Out of the 32 papers analysed, merely five focus on immersive VR technologies. This observation underscores a significant gap in the literature, as immersive VR environments offer unique opportunities and challenges for the study of IVAs. These environments can potentially provide a more controlled and immersive context for examining the nuances of user interactions with IVAs, which might differ significantly from interactions in less immersive or monitor-based setups.

In addition, there is a significant body of research that is not included in this literature review. The strict adherence to the PRISMA method, while ensuring rigour, may have limited the scope of included studies and research groups considered and overlooked relevant research that falls outside the

specified criteria. For example, the impact of cultural and demographic factors on socio-affective perceptions, though noted, warrants further exploration to understand how different cultural backgrounds influence user experiences with IVAs. It should be noted that several research works investigate these differences between cultures, comparing the perception of participants from Europe and Asia. Perception differs effectively when it comes to personality (*i.e.*, friendliness) [170, 169, 244] and trustworthiness aspects [169]. The same statement can be made when it comes to social capabilities of IVAs. Indeed, users are more likely to take part in a conversation with agents when they belong to the same culture [228], highlighting a preference between users for signals attributed to their culture [229]. In [88], a comparison is established between Individualistic and Collectivist cultures. A significant impact of this cultural aspect is identified, influencing the perceived appropriateness of a discussion between virtual agents, either adopting a warmth and friendly behaviour, or a more aggressive and competitive conduct.

In conclusion, this systematic review sheds light on the intricate relationship between the multimodal behaviour of IVAs and the socio-affective perceptions of users, offering valuable insights and guidelines for the design of socio-affective IVAs. The findings underscore the need for a nuanced approach to IVA design that considers the full spectrum of non-verbal modalities and their interplay with verbal communication. As technology evolves and user expectations change, the field must continue to explore these dynamics, ensuring that IVAs remain effective, engaging, and capable of meeting the diverse needs of their users.

Table 2.2: Research articles considered in the systematic review

Reference	Dimensions	Sub-dimensions	Device	Non-verbal signals	Verbal signals	Nbr. of VAs
Boe et al. [25]	Personality	Dominance, extroversion, friendliness	Monitor	Facial expressions, gaze, lip movements	Linguistic	1 (M)
Burrows et al. [31]	Personality	Friendliness	Monitor	Arm movements, facial expressions	Linguistic	1 (F)
Calais et al. [34]	Believability	Credibility	Monitor	Arm movements, facial expressions, gaze, head movements	Linguistic	1 (F)
Chen and Scherer [97]	Personality	Dominance, friendliness	Monitor	Facial expressions, gaze, gestures, head movements	/	1 (F/M) + 1 (F/M) + 2 (F/M)
Darwin et al. [80]	Personality	Extroversion, dominance, friendliness	Monitor	Facial expressions, gaze, head movements	/	1 (F/M)
Demary et al. [80]	Emotions	Valence	Monitor	Facial expressions, gaze, body direction	/	1 (F/M)
Demouris and Kiehlbas [92]	Personality	Dominance, friendliness	Monitor	Arm movements, gaze, head movements, postures	/	1 (F)
Demouris and Kiehlbas [93]	Personality	Dominance, friendliness	Monitor	Arm movements, facial expressions, gaze, head movements, postures	/	3 (F/M)
Burrows et al. [111]	Emotions	Dominance, friendliness	VR	Facial expressions, head movements, gestures	1 (F/M)	1 (M)
Fornalander et al. [121]	Emotions	Attention, social approach, frustration	Monitor	/	Paralinguistic	1 (M)
Griffiths et al. [118]	Personality	Dominance, friendliness, dominance	Monitor	/	Paralinguistic	1 (M)
Grosz et al. [116]	Personality	Dominance, friendliness, extroversion	VR	Facial expressions, gaze, head movements, postures	1 (F/M)	10 (F/M)
Hessing et al. [136]	Personality	Friendliness	Monitor	Arm movements, facial expressions, head movements	/	1 (F/M)
Hessing et al. [137]	Personality	Helpfulness, indulgence, trustworthiness	Monitor	Arm movements, facial expressions, head movements	/	1 (M)
Kane et al. [100]	Believability	Naturalness, plausibility	VR	Gaze, head movements	/	1 (M)
Kane et al. [102]	Believability	Attention, valence	VR	Gaze, facial expressions, postures	/	1 (M)
Lazar et al. [213]	Emotions	Amused, valence	Monitor	Facial expressions	Linguistic, Paralinguistic	1 (F)
Lee et al. [103]	Personality	Believability	Monitor	Gaze, facial expressions, postures	/	1 (F/M)
Macrae, Anstey and Isler, Kahn-Herder and Jackson, Philip L. [244]	Personality	Friendliness	Monitor	Gaze, gestures, facial expressions, postures	/	1 (F)
Nasser, D'Arca and Bressan [294]	Personality	Competence	Monitor	Gaze, postures	/	1 (M)
Ochs, Feldsuss and Prepin [278]	Emotions	Valence	Monitor	Facial expressions	Linguistic	1 (M)
Pernat et al. [283]	Personality	Friendliness	Monitor	Facial expressions	Linguistic	1 (F/M) or 2 (F)
Pernat et al. [285]	Personality	Friendliness	Monitor	Facial expressions	Linguistic	1 (F)
Pernat et al. [286]	Personality	Friendliness	Monitor	Facial expressions, postures	Linguistic, Paralinguistic	2 (F)
Pernat et al. [287]	Social capabilities	Intimacy, mutual comprehension	Monitor	Facial expressions, gaze, gestures, head movements, postures	Linguistic, paralinguistic	2 (F)
Pernat et al. [288]	Social capabilities	Intimacy, mutual comprehension	Monitor	Facial expressions, gaze, gestures, head movements, postures	Linguistic, paralinguistic	2 (F)
Pernat et al. [289]	Social capabilities	Mutual agreement, comprehension, understanding	Monitor	Facial expressions	/	1 (M)
Pernat et al. [290]	Personality	Dominance, friendliness	Monitor	Gestures, gaze, sentence formulation	/	1 (M)
Pernat et al. [291]	Social capabilities	Mutual comprehension	Monitor	/	/	1 (M)
Pernat et al. [292]	Social capabilities	Mutual comprehension, agreement, persuasion	Monitor	/	/	1 (M)
Pernat et al. [293]	Personality	Empathy, extroversion	VR	Facial expressions, gaze, gestures, postures	/	1 (F)
Pernat et al. [294]	Personality	Dominance	Monitor	Arm movements, facial expressions, gaze, gestures, head movements, postures	/	1 (M)
Pernat et al. [295]	Personality	Extroversion, dominance	Monitor	/	Linguistic, paralinguistic	1 (M)
Pillon et al. [301]	Social capabilities	Tolerance, self-disclosure	Monitor	/	Linguistic, paralinguistic	1 (M)
Pillon et al. [301]	Personality	Competence	Monitor	/	Linguistic, paralinguistic	1 (M)
Wang et al. [374]	Trustworthiness	Helpfulness, trustworthiness	Monitor	Facial expressions, gaze, head movements	/	1 (M)
Zajac, Peters and McKelvey [364]	Trustworthiness	Extroversion, friendliness	Monitor	Arm movements, gaze	Linguistic, paralinguistic	8 (M), 4 (F)
Zajac, Cernot and Peters [365]	Trustworthiness	Dominance, friendliness	VR	Arm movements, gestures	Linguistic	8 (M), 4 (F)

This table presents the selected papers after the second step of the PRISMA method (see Section 2.2). For these papers, the dimensions and associated sub-dimensions presented in Section 3 are described. The device, the verbal and non-verbal signals, the number of IVAs involved in these studies, and their genders are reported.

Note: M means Male and F means Female.

Chapter III

Perception of Avatars Non-verbal Behaviours in Virtual Reality

This chapter is based on a collaboration with the Faculty of Psychology, Speech Therapy, and Education of the University of Liège and HEC Liège. This chapter is a paper, for which I am the first author, that has been published in the Journal *Psychology & Marketing* in 2023. The co-authors are Anne-Lise Leclercq, Angélique Remacle, Laurence Dessart, and Michaël Schyns. In this thesis, the paper is presented in its original published form, and I apologise for any outdated information, repetitions, or formatting inconsistencies, including the use of numbered citations instead of author names, without modifying the original text.

Abstract

Virtual reality has shown great potential in many fields, especially in business and psychology. By immersing someone in a new computer-generated reality, it is possible to create realistic, safe, and controllable simulations for research and training, as well as new three-dimensional-enriched consumer experiences and services. Most of these environments, especially in the metaverse, rely on virtual representations of people called “avatars”. The design and non-verbal behaviours of these avatars must be carefully crafted to provide a realistic and truly immersive experience. This paper aims to understand how avatar non-verbal behaviours (i.e., body posture, facial expression, and head movement) are perceived by users immersed in a virtual reality context, a very common situation encountered in many simulations and especially during training. Therefore, the first objective of this study is to validate, through an experiment with 125 participants, how the

audience's levels of emotional valence and arousal are perceived in virtual reality. Based on these results, a library of audience non-verbal behaviours corresponding to different arousal and valence levels is now available for future applications. The experiment also examines the benefits of using low-end versus high-end virtual reality headsets, and photo-realistic versus cartoon avatars. The results have implications for the design of realistic, challenging, and interactive virtual audiences.

3.1 Introduction

Virtual reality offers tremendous opportunities in business practice due to its ability to provide realistic, safe, and controllable experiences and services. By immersing individuals in a life-like environment and replicating real-life situations [311], virtual reality is particularly valuable when human behaviour is a critical element of the business process under study. Not surprisingly, virtual reality research is especially prevalent in the fields of marketing and psychology. Reference [227] studied no less than 150 papers in 115 journals on the use of virtual reality in marketing. In a special issue on virtual reality in marketing, reference [39] provide essential insights into various aspects of virtual reality and its implications for this field. Virtual reality, but also augmented reality, are now widely used in many business areas, such as tourism [24, 181], retail [34], and health [172, 171, 29]. Virtual reality is also widely used in education through training [372, 148, 312].

A concept related to virtual reality is the *metaverse*, which reference [160, p. 1] defines as a “computer-mediated environment consisting of virtual worlds, within which users can act and communicate in real-time using virtual people”. In these virtual worlds, if virtual people (also called *avatars*) are present, their design and non-verbal behaviours must be realistic to enhance the learning experience. Avatars can be defined as “digital entities with anthropomorphic appearance, controlled by a human or software, that have an ability to interact” [258, p. 67]. Accordingly, the term avatar is used in this study to refer to interactive representations of people, whether realistic or cartoons. The metaverse and avatars are becoming increasingly popular concepts, especially in marketing. For instance, in their recent study, reference [102] argues that the metaverse has the potential to revolutionise the way businesses interact with consumers. The authors discuss various applications of the metaverse in marketing, including brand experiences, customer engagement, and immersive advertising, highlighting the challenges and ethical considerations associated with using the metaverse in marketing, such as privacy concerns and potential biases in avatar design.

The present study focuses on public speaking situations where an avatar listens to a user speaking. Public speaking is a widespread activity in everyday life. Whether in front of a small or large audience, individuals

often need to mobilise their public speaking skills. Psychology research has investigated different activities, such as presenting research findings, teaching and lecturing in classrooms or professional development workshops [311, 312]. There are also more complex situations that involve public speaking, such as providing psycho-education to patients in individual or group therapy settings, advocating for mental health policies and initiatives at public events or rallies, leading support groups for individuals with mental health challenges, and so forth [126].

Focusing on public speaking tasks in various fields and a virtual environment with an interactive virtual audience (i.e., providing non-verbal feedback) has been the focus of several studies [63, 66, 68, 20, 65, 188, 141, 230, 337, 337]. Virtual reality allows public speakers to practice in a situation very similar to what they will face in real life. People can practice where, when, and as often as they need to build their confidence and be ready to give a speech [327, 80]. In addition, the training can be progressive and controlled in terms of the number of participants, their non-verbal behaviours, and so on, resulting in much more efficient and faster learning [103].

There are also a number of tasks in marketing that involve public speaking: pitching a product or service to potential customers or investors, giving a progress report to a team or supervisor, presenting a proposal or idea to a group or board of directors, and interacting with clients. The practical application of this study is improving the public speaking skills of frontline employees who play a critical role in fostering positive consumer perceptions through effective communications in face-to-face interactions. Frontline employees play a crucial role in delivering high-quality services that increase customer satisfaction and loyalty [245]. In such customer-facing situations, frontline employees would benefit from public speaking training. However, providing effective training for these employees can be a challenging task, particularly in the context of rapidly evolving service technologies and complex service encounters [211, 315, 102]. Virtual reality has also emerged as a promising technology for employee training, providing an immersive and interactive learning experience that can improve their skills, knowledge, and confidence in dealing with customers. Indeed, virtual reality technology has become increasingly popular in frontline employee training across a range of industries [87] due to its ability to simulate real-life situations, allowing trainees to develop and practice their skills in a safe and controlled environment.

Whilst the benefits of virtual reality for training purposes have been demonstrated, blind spots remain in the precise application and design of this technology to maximise training effectiveness and foster positive perceptions among trainees. Specifically, the trainee-avatar interaction is relevant to understanding the impact that the listening avatar in the simulated environment may have on the trainee practising public speaking. Indeed, the

immersed person must perceive these listening avatars as realistic in order to provide a truly immersive and effective experience. However, the impact of the avatar's perceived responses on the trainee and the resulting sense of presence in the virtual reality environment remain unclear. Furthermore, the development of accurate and authentic non-verbal avatar behaviour is still in its infancy. Therefore, the first objective of this study is to investigate how virtual reality users perceive the levels of emotional valence and arousal of avatars. The second and third objectives concern a more in-depth analysis of the sense of presence using different types of headsets and different avatar qualities. In sum, this study aims to investigate how avatars are perceived and the benefits of using different headsets or graphics.

3.2 Theoretical Development

One of the key benefits of virtual reality in training frontline employees is its ability to simulate real-life service encounters and practice situational skills. According to [225], interaction behaviours that lead to feeling comfortable in the service encounter are crucial to creating a positive customer experience. Virtual reality can help trainees develop these behaviours by creating realistic simulations of different service encounters, allowing them to practice their skills in a safe and controlled environment. This can help build confidence and reduce anxiety in real-life service encounters. In addition, virtual reality can provide a platform for role-playing exercises that are effective in developing the skills and behaviours required for successful service encounters. Reference [347] suggests that role theory can be used to understand the dynamics of service encounters, with service employees and customers playing specific roles in the interaction. Virtual reality can provide a platform for employees to practice different roles and scenarios, allowing them to develop a better understanding of the service encounter dynamics and how to respond effectively. Furthermore, virtual reality can help improve the quality of service encounters by enhancing employees' non-verbal communication skills. According to [353], non-verbal communication plays a critical role in service encounters, with cues such as facial expressions, gestures, and tone of voice conveying important information to customers. Virtual reality can also be used to simulate various non-verbal cues, allowing employees to practice their communication skills and become more effective in conveying information to customers. As the question of how frontline employees respond to consumer emotions in service encounters is still a black hole [303], virtual reality can help address this question.

In addition, virtual reality can provide a cost-effective and scalable solution for training frontline staff [163]. Reference [114] suggests that conceptualizing service encounters is important for understanding employee

behaviour and how it affects customer perceptions of service quality. However, traditional training methods, such as classroom teaching and on-the-job training, can be time consuming and costly. Virtual reality can provide a cost-effective solution for training large numbers of employees, allowing organizations to scale their training programs without incurring high costs. Reference [91] highlights the importance of technology infusion in frontline service delivery. They argue that technological advances have changed the nature of service encounters and created new opportunities for service providers to improve the customer experience. However, they also note that effectively integrating technology in service encounters requires adequate training and support for frontline staff. Virtual reality can be an effective tool for such training by simulating realistic service encounters in a controlled environment. However, research on the appropriate design of a virtual reality environment for frontline employee training is still lacking [196].

In the context of public speaking in general, virtual reality can improve communication self-efficacy [312], and speaking performance [63], especially when the audience (the listening avatar) is interactive [384, 143]. The avatar's reactions can have a significant impact on the speaker's emotions and performance [292, 296, 200, 64, 65], interactivity in virtual reality is also an important part of the training process. Therefore, it is essential to know if the user perceives the interactions between the avatar and her/himself in the virtual environment as representative of reality and how each interaction is interpreted. Two dimensions related to emotion and affect are essential: the level of arousal and emotional valence. According to [63], *arousal* can be understood as the avatar's level of alertness, while *valence* corresponds to how positively or negatively the avatar feels toward the speaker or the presentation. Specifically, body postures, facial expressions, and head movements are found to express some degrees of arousal and valence [9, 201].

In the context of service encounters, marketers face the challenge of creating positive experiences that meet or exceed customer expectations. To achieve this goal, marketers can use the concepts of valence and arousal, which are critical in shaping customer perceptions and behaviours during service encounters. A positive emotional response to a service encounter can lead to more favourable perceptions of the service provider and the service itself, resulting in higher customer satisfaction and loyalty [11, 42]. However, negative emotional responses and excessive arousal can have the opposite effect, resulting in lower customer satisfaction and loyalty. Negative emotional responses can lead to negative word-of-mouth, which can have a significant impact on a service provider's reputation [345]. In addition, excessive arousal can lead to anxiety and discomfort, which can negatively impact the overall experience and result in lower customer satisfaction [220]. Valence and arousal are important concepts in marketing, particularly in the context of service encounters. They play a critical role in

shaping customer perceptions and behaviours, thereby influencing their satisfaction and loyalty. By designing service encounters that promote positive emotional experiences and appropriate levels of arousal, service providers can increase customer satisfaction, loyalty, and ultimately business success. It is therefore important to gain a better understanding of how frontline employees perceive the emotional valence and arousal of avatars in virtual reality.

This leads to two hypotheses linking emotional valence and arousal to body postures, facial expressions, and head movements.

H1: Avatar smiles and nods are associated with positive valence by users, frowns and head shakes with negative valence, and raised eyebrows are mostly neutral.

H2: Higher levels of avatar facial expressions, head movements, and forward postures are positively related to higher levels of arousal perceived by participants.

Following [341, p. 1]’s terminology, *immersion* is what the technology delivers from an objective point of view, or the objective level of sensory fidelity a virtual reality system provides. *Presence* is the human response to immersion, i.e., the participant’s subjective sense of being in the virtual place. It is the “observer’s sense of psychologically leaving their real location and feeling as if transported to a virtual environment” [379, p. 2]. It has long been known that presence can affect user performance [268]. For a virtual reality environment and its associated goals to be effective (in this case, training frontline employees to adequately respond to the emotions of avatarised consumers), the presence and the immersion components of virtual reality must be as high as possible.

To study variations in presence due to the immersive quality of the virtual reality environment, the device used to display the virtual reality is of paramount importance [119, 219]. Indeed, low-end virtual reality headsets, i.e., smartphones in Samsung Gear VR Cardboards (see Figure 3.1), and high-end virtual reality headsets, i.e., Meta Rift S headsets (see Figure 3.2) are expected to lead to significant differences in the sense of presence. Due to hardware limitations, the low-end headset is not expected to perform better than the high-end headset (higher display resolution, wider field of view, more accurate tracking, amongst others), and the latter is often the preferred choice in terms of sense of presence (e.g., [336]). The reasons for this are cost, ease of operation, and portability. As [6] explain in comparing immersion in the Cardboard and Oculus Rift headsets, a high-end headset is expensive, requires professional technical operation, and is less mobile than Cardboard with its affordability (only a smartphone is needed in addition to the Cardboard) and ease of use. A high-end headset can thus make virtual reality inaccessible to everyone or enable mass adoption (a hundred

people at a time). Reference [6]’s results suggest that cardboard is capable of providing an acceptable level of immersion. Similar results may be derived when assessing valence and arousal between low-end and high-end headsets. Should this be the case, it may be possible to address the challenges of affordability and widespread adoption at a lower cost. This leads to the third hypothesis regarding the device used.



Figure 3.1: Low-end headset:
Samsung Gear Virtual Reality
headset



Figure 3.2: High-end headset:
Oculus Rift S
headset

H3:

- a. Both headsets, including the low-end, are capable enough to accurately evaluate emotional valence and arousal with confidence.
- b. High-end headsets provide a higher sense of presence than low-end headsets.

The third objective concerns the quality of the avatars used. In the event that virtual reality and the metaverse come to fruition, avatars will assume the role of our digital representations or embodiments [197]. Most often the virtual audience is represented by “simplified” cartoon avatars (in terms of the method used to create them), but interactive three-dimensional avatars have the highest level of the digital object continuum, as [203] explain. Several studies examine the importance to users of the creation of their avatars [340, 332, 187, 338]. For example, reference [338] examines whether participants have different affinity, trustworthiness¹, and preferences for avatars with two levels of realism (one almost human-realistic and one a cartoon). Another study shows that participants who interacted with the human-realistic avatar in a two-dimensional video display had a positive experience, rated the avatar as more trustworthy, had more affinity, and preferred it [339]. The level of confidence in their answers was therefore higher when they dealt with human-realistic avatars. When the same questions were asked of participants immersed in virtual reality, the affinities and preferences were even stronger. Similar results were found for robots and their anthropomorphism. For example, reference [218] found in their work

¹Trust can be applied to information systems as explained in [375]

that consumers prefer a higher level of humanness and a moderate to high level of sociability. Other studies emphasise the importance of anthropomorphism in promoting trust and positive responses [199, 365, 266, 125, 26, 388]. Regarding the comparison of avatar non-verbal behaviours, reference [206] shows a high degree of correspondence between the effects produced by synthetic and human smiles. In addition, reference [50] highlights that first impressions may determine important relationship decisions. The final question therefore examines whether fully rigged three-dimensional photo-realistic models can significantly improve participants' perceptions of the avatar's arousal and valence or their confidence level. The desire for this comparison lies in the ease of avatar creation. In the context of a virtual audience, it would be possible to create a virtual environment with avatars that look like the people supposedly present in the real context [332]. The last hypothesis is therefore related to the method used to create the avatars.

H4: Compared to cartoon avatars, photo-realistic avatars improve the level of confidence of users.

3.3 Methodology

3.3.1 The Virtual Reality Environment

The avatars and the environment were created by three-dimensional animation specialists, technical artists, and three-dimensional game artists working in a research lab at HEC Liège (University of Liège). The technical artists used the Unity three-dimensional engine. The three-dimensional artists used the *Blender* and *Maya* software. The photo-realistic avatars were created and animated using the *Reallusion* suite.

The virtual reality environment (see Figure 3.3) depicts an office with the avatar behind a desk, animated (see Section 3.3.2 for the possible animations) as he listens to the immersed participant for at least 15 seconds. The environment was designed to represent a common interaction between a frontline employee and a customer, such as in the context of a banking, insurance, or real estate agency appointment. The environment was optimised to support the highest quality virtual reality experience possible using an autonomous low-end Samsung S7 smartphone and a high-end Oculus headset connected to a computer (HP Omen) with a virtual reality-compatible graphic power unit (NVIDIA GeForce GTX1070 graphics card) and processor (Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99 GHz). Due to the relative simplicity of the scene, it was not necessary to drastically reduce the complexity of the three-dimensional avatars.

Eight avatars were created for this experiment, represented in Figure 3.4 in their neutral state in the Unity software. Thus, in the virtual reality environment, these images represent the quality of the avatars, not their



Figure 3.3: Screenshot of the virtual reality environment used for the experiment

non-verbal behaviours. To determine whether the method used to create the avatars influenced participants' perceptions of the avatar's arousal and valence or their interpretation of the audience's confidence levels, half were cartoon avatars and half were photo-realistic avatars based on real people. Photo-realistic avatars are created from photos of real people (women and men) who gave informed consent. The software used is *Character Creator* with the plugin *HeadShot*, which is part of the *Reallusion* suite. The avatars were then enhanced by three-dimensional artists and animated by three-dimensional animators. For the sake of diversity and to avoid gender bias, half of the avatars are female and half male. In terms of origin, four avatars are European, two are Indian with a dark complexion, and two are African.

3.3.2 Non-Verbal Behaviours

The avatars and their animations were created by three-dimensional and technical artists. Several animations were created for different types of avatars (male or female avatars of European, African, or Indian origin). To define the sets of animations, the results of [63] were used as a basis. The parameters considered in their study include: posture (forward, backward, neutral), amount of averted gaze (0, 25, 50, 75, or 100%), direction of averted gaze (sideways, down, or up), type of facial expression (smile, frown, or raised eyebrows), facial expression frequency, head movements (nod or shake), and head movement frequency. Considering all these parameters for a true immersion virtual reality experiment is excessive and unnecessary, since the will is not to study the valence and level of arousal for each possible combination. In fact, only the most common non-verbal behaviours of an audience are needed. Therefore, the results of [63] were used to define



Figure 3.4: Avatars created using two different methods. The first row represents cartoon avatars and the second row photo-realistic avatars.

appropriate combinations that are expected to represent positive, neutral, and negative valence with a low or high level of arousal. For example, a head nod is expected to have positive valence in contrast to a head shake, and a forward posture is expected to have a higher level of arousal. Table 3.1 displays the parameters used. It leads to 144 possible combinations (9 postures x 4 facial expressions x 4 head movements).

Table 3.1: Parameters of non-verbal behaviour

Postures (P)	Facial expressions (F)	Head (H)
1: Backward posture – Arms crossed	1: None	1: None
2: Backward posture –Arms stand (elbows on the table with hands crossed)	2: Smiling	2: Nod
3: Backward posture – Arms behind head	3: Frowning	3: Shake
4: Upright posture – Hand on hand (hands on the table, one on top of the other)	4: Eyebrows raised	4: Questioning
5: Upright posture – Hands together (hands crossed on the table)		
6: Upright posture – Hands separated		
7: Forward posture – Hands together		
8: Forward posture – Arms stand (elbows on the table with hands crossed)		
9: Froward posture – Arms crossed		

In order to reduce the number of manageable combinations in the experiment, a set of 40 non-verbal behaviours (as described in the next section) was randomly selected. In this selection process, a priori *clear* animations (e.g., smile and nod), neutral animations, and more ambiguous

animations (e.g., smile and shake) were intentionally kept to broadly cover the set of possibilities.

3.3.3 The Virtual Reality Conditions

An experiment was conducted to determine whether a set of predefined animated avatars were interpreted in virtual reality to express different levels of valence and arousal. Each participant had to rate the level of valence and arousal for 20 animated 15-second sequences. Across participants, 40 sequences (i.e., combination of postures, facial expressions, and head movements) were tested. Table 3.2 shows these animations.

Table 3.2: Sequences

Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6	Seq. 7	Seq. 8	Seq. 9	Seq. 10
P1F3H4	P3F4H2	P7F2H2	P7F2H4	P7F3H3	P7F4H3	P7F4H4	P9F1H2	P9F2H3	P9F3H3
Seq. 11	Seq. 12	Seq. 13	Seq. 14	Seq. 15	Seq. 16	Seq. 17	Seq. 18	Seq. 19	Seq. 20
P2F2H3	P2F4H2	P3F3H3	P3F4H3	P4F4H1	P4F4H3	P5F3H1	P7F3H1	P8F4H1	P9F1H3
Seq. 21	Seq. 22	Seq. 23	Seq. 24	Seq. 25	Seq. 26	Seq. 27	Seq. 28	Seq. 29	Seq. 30
P2F1H2	P2F4H4	P3F1H2	P3F3H2	P3F4H4	P4F1H2	P5F1H4	P6F3H4	P7F1H1	P7F1H2
Seq. 31	Seq. 32	Seq. 33	Seq. 34	Seq. 35	Seq. 36	Seq. 37	Seq. 38	Seq. 39	Seq. 40
P1F3H1	P2F4H3	P3F2H3	P3F4H1	P4F1H3	P4F2H2	P6F3H3	P7F1H4	P7F4H3	P8F3H4

The avatars and animations were divided into four sets². Each set contains one male and one female avatar, one photo-realistic version and one cartoon version. Ten animations were randomly attached to each of these two avatars.

Each participant in the experiment viewed 20 sequences from one of the four sets. The sequences were randomly presented. For the first three sets, all the experiments were conducted using an autonomous low-end virtual reality headset based on a Samsung S7 smartphone and the Samsung Gear virtual reality adapter, an improved version of the Google Cardboard. The experiment was repeated for the Set 3 sequences using a high-end Oculus headset connected to a computer with an Nvidia Geforce GTX1070 graphics card. The experiment for Set 4 was also conducted using a high-end Oculus headset connected to a computer with an Nvidia Geforce GTX1070 graphics card.

3.3.4 The Experiment Settings

The present study was approved by the Ethics Committee of the Faculty of Psychology, Speech Therapy, and Educational Sciences of the University of Liège (file number: 1920-81). The participants were recruited voluntarily through a targeted advertisement sent via the official university e-mail. A website was created and participants had to register and provide

²Throughout this paper, a distinction is made between the animations (sequences), which are divided into *sets*, and participants divided into *groups*.

informed consent after receiving a full description of the study. Once registered, they were given an anonymous identifier that allowed them to access the online questionnaire once the experiment began. Based on a power test, 24 participants were required to rate each animation in order to draw valid conclusions. 125 participants took part in the experiment. The participants were immersed in a virtual reality environment to rate the valence and arousal of a single avatar. They were divided into 5 groups by order of registration³. To test the third hypothesis (a and b), the same animation settings were used for the third and fourth groups, but with different headsets.

After being given the definitions of valence and arousal (written at the beginning of their online questionnaire) as defined in Section 3.1, participants were asked to wear the provided headset (Cardboard and Samsung S7, or Rift S headset). The virtual reality application contained the list of animations to be viewed. The 20 animated sequences were presented in random order. Each participant had to watch each animation for at least 15 seconds (at the end of the 15 seconds, the sequence repeated endlessly until the participant moved on), and each animation was played in its entirety before moving on to the next one. In addition, the sequence was restarted as often as necessary. For each of the 20 animations, using their computer or smartphone, they responded to an online questionnaire set up for this experiment. Participants had to remove the headset after each sequence to complete the questionnaire. They were asked to rate their level of arousal (7-point Likert-type scale from very low to very high) and valence (7-point Likert-type scale from very negative to very positive), as well as their level of confidence for each of their answers (7-point Likert-type scale from very unconfident to very confident). They could also write down open-ended comments about the video they had watched.

After watching all the sequences (post-immersion), participants answered some questions (in the online questionnaire) about the sense of presence they experienced in the virtual reality environment. The French-Canadian version of The Gatieneau Presence Questionnaire [209], was used [36]. It is a 4-item questionnaire scored on a percentage scale. As [209, p. 4] explain, Gatieneau Presence Questionnaire consists of four items: “1) the feeling of being there, 2) the perception of the experience as real, 3) the awareness of the virtual environment as artificial, and 4) the feeling of being in the physical office instead of a virtual environment”. The last two items were scored in reverse, and the average percentage was computed to obtain the global score for the Gatieneau Presence Questionnaire and thus the sense of presence in general. In the following, these items will be summarised as

³25 participants for the first group with the Samsung Gear, 25 in the second group with the Samsung Gear, 25 in the third group with the Samsung Gear, 25 in the third group with the Oculus, 25 for the fourth group with the Oculus. The first three groups used the Samsung Gear, i.e., the low-end headset, and the two other groups the Oculus, i.e., the high-end headset.

1) presence in virtual reality, 2) level of realism, 3) level of artificiality, and 4) spatial awareness. At the end of the experiment, participants completed the online sociodemographic questionnaire, reporting any recent events that might have altered their perception, such as an upsetting or joyful event.

A total of 125 people (64 men and 58 women) participated in the study, with an age range between 17 and 62 (with an average of 27.25), and most are Belgian (with French as mother tongue).

3.4 Results

This section analyses emotional valence and level of arousal for the parameters of the non-verbal behaviours separately (posture, facial expression, and head movement). Then, the sequences presented in Table 3.2 are analysed for emotional valence and level of arousal. The two subsequent sections compare the use of the low-end and high-end headsets, and photo-realistic versus cartoon avatars. Finally, the results of the Gatineau Presence Questionnaire are presented.

3.4.1 Analysis of Posture, Facial Expression, and Head Movement

This section first presents the levels of arousal and valence attributed by participants for each parameter (posture, facial expression, and head movement). The global goal is to identify in the list of possibilities the behaviours that are clearly associated with specific levels of valence (positive, neutral, negative) and arousal (high, neutral, low). An association is *clear* when all participants, or at least the vast majority, interpret a behaviour in the same way. If the interpretations differ significantly among participants, special care must be taken when using them. A Chi-square test was performed to see if the participants' responses were equally distributed between negative interpretations (responses 1 to 3 in the Likert-type scale) and positive interpretations (responses 5 to 7 in the Likert-type scale)⁴. The aim is to determine whether the difference between the proportions of positive and negative interpretations is statistically significant. For the sake of brevity, Table 3.3 summarises the results of the statistical tests. The full analysis can be found in the Appendix (see Tables 3.5 and 3.6).

In order to determine whether some postures are more significant than others in assessing valence or arousal, ANOVA tests were conducted. Only significant results are presented in this section. Among the backward postures, the crossed arms posture is considered the most negative (p -value

⁴Throughout this paper, and for the sake of clarity, all responses from 1 to 3 in the Likert-type scale are referred to as "negative" and all responses from 5 to 7 in the Likert-type scale are referred to as "positive". However, the words "negative" and "positive" need to be used with care. For valence, it can be understood as negative and positive levels, for arousal, as low and high levels.

Table 3.3: Interpretations of the Chi-square tests

	Interpretation valence	Interpretation arousal
Postures		
P1 Backward posture – Arms crossed	-	+
P2 Backward posture – Arms down	-	+
P3 Backward posture – Arms behind the head	-	+
P4 Upright posture – Hand on hand	/	+
P5 Upright posture – Hands together	-	/
P6 Upright posture – Hands separated in front	-	+
P7 Forward posture – Hands together	/	+
P8 Forward posture – Arms down	-	/
P9 Forward posture – Arms crossed	-	+
Facial expressions		
F1 Neutral	/	+
F2 Smiling	/	+
F3 Frowning	-	+
F4 Eyebrows raised	-	+
Head movements		
H1 Neutral	-	+
H2 Nod	+	+
H3 Shake	-	+
H4 Questioning	-	+

= 0.041). Similarly, among the head movements associated with negative valence, head shaking expresses the most negative valence. Furthermore, although arousal is almost always perceived as positive, the degree of positivity differs between postures, facial expressions, and head movements (p values < 0.001).

3.4.2 Library of Animated Avatars

A particularly interesting question is how the combination of the parameters (postures, facial expressions, and head movements) is perceived. Some combinations may enhance the arousal or valence perception, or conversely, blur the results. The perception of the different sequences presented in Table 3.2 is next analysed. Figures 3.5 and 3.6 show the results for emotional valence and arousal. Each row in the figure corresponds to a particular sequence, where the distribution of responses is shown according to the legend below the figure (responses from 1 to 7 represent the 7 points of the Likert-type scale). The three percentages represent the proportion of negative responses (1 to 3), neutral responses (4), and positive responses (5 to 7). The sequences are ordered by the percentage of positive responses (decreasing) and then by the percentage of negative responses (increasing). The first row is thus the most positive sequence for the parameter assessed, and the last is the most negative. For example, for emotional valence, Sequence

36 (P4F2H2: an avatar with an upright posture, hands on top of each other, a neutral facial expression, and nodding his head) is perceived as having the most positive emotional valence, while Sequence 14 (P3F4H3: an avatar with a backward posture, hands behind his head, frowning eyes, and shaking his head) is perceived as having the most negative emotional valence (see Figure 3.5). For level of arousal, Sequence 11 (P2F2H3: an avatar with a backward posture, elbows on the table, a smiling face, and shaking his head) is perceived as having the highest level of arousal, whereas Sequence 30 (P7F1H2: an avatar with a forward posture, hands together, a neutral face, and nodding the head) is perceived as having the lowest level of arousal (see Figure 3.6).

To determine which sequence is associated with a particular level of valence and arousal, further tests were conducted (using the same method as explained in Section 3.4.1), and three groups emerged (see Figures 3.5 and 3.6, and Table 3.4):

- Sequences for which the difference between the proportion of negative and positive responses is statistically significant and for which there are more positive than negative responses. These are the sequences above the green line in the figures.
- Sequences for which the distribution of the Likert-type scale responses are equally distributed between negative responses and positive responses. These are the sequences between the green and the red line in the figures.
- Sequences for which the difference between the proportion of negative and positive responses is statistically significant and for which there are more negative than positive responses. These are the sequences below the red line in the figures.

Comparing the results from the previous section (see Table 3.3) and this section (see Figures 3.5 and 3.6) shows that sometimes non-verbal behaviours are separately perceived in same way, but their combination leads to a different perception. For example, regarding emotional valence, in Sequence 4 (i.e., P7F2H4: an avatar with a forward posture, with elbows on the table, head tilted), all the non-verbal behaviours were separately perceived as neutral or negative, but their combination led to positive valence. In this case, a smile seems to be perceived as positive and outperforms the evaluation of other non-verbal behaviours. Another example of the level of arousal is Sequence 29 (i.e., P7F1H1: an avatar with a forward posture, without any facial expression or head movement). In this sequence, all the non-verbal behaviours separately were perceived as representing a high level of arousal. However, their combination led to a perception of low arousal.

Table 3.4: Sequences per level of valence and arousal
(Please refer to Tables 3.1 and 3.2 for the meaning of the parameters and sequences)

	Negative valence	Neutral valence	Positive valence
Low level of arousal	∅	Seq. 27: P5F1H4 Seq. 29: P7F1H1 Seq. 30: P7F1H2	∅
Neutral arousal	Seq. 07: P7H4H4 Seq. 19: P8F4H1	Seq. 15: P4F4H1 Seq. 22: P2F4H4 Seq. 25: P3F4H4 Seq. 34: P3F4H1	Seq. 02: P3F4H2 Seq. 26: P4F1H2
High level of arousal	Seq. 01: P1F3H4 Seq. 05: P7F3H3 Seq. 06: P7F4H3 Seq. 09: P9F2H3 Seq. 10: P9F3H3 Seq. 11: P2F2H3 Seq. 13: P3F3H3 Seq. 14: P3F4H3 Seq. 16: P4F4H3 Seq. 17: P5F3H1 Seq. 18: P7F3H1 Seq. 20: P9F1H3 Seq. 28: P6F3H4 Seq. 31: P1F3H1 Seq. 32: P2F4H3 Seq. 33: P2F2H3 Seq. 35: P4F1H3 Seq. 37: P6F3H3 Seq. 39: P7F4H3 Seq. 40: P8F3H4	Seq. 23: P3F1H2 Seq. 24: P3F3H2 Seq. 38: P7F1H4	Seq. 03: P7F2H2 Seq. 04: P7F2H4 Seq. 08: P9F1H2 Seq. 12: P2F4H2 Seq. 21: P2F1H2 Seq. 36: P4F2H2

A closer look at the categories in Table 3.4 shows that some non-verbal behaviours dominate over others. For example, every time an avatar shakes his head (sequences associated with the third head movement, i.e., P*F*H3), it is perceived as having negative emotional valence and a high level of arousal, regardless of the associated posture or facial expression. Similarly, when the avatar nods his head (sequences associated with the second head movement, i.e., P*F*H2), it is mainly perceived as having positive emotional valence. Sometimes, depending on the other parameters (posture and facial expression), it is perceived as neutral, but never as negative, as expected. This confirms the first hypothesis.

It also seems that when the avatar raises his eyebrows (sequences asso-

ciated with the fourth facial expression, i.e., P*F4H*), the level of arousal is perceived as neutral, except if shaking his head at the same time (sequences associated with the fourth facial expression and the third head movement, i.e., P*F4H3). In this case, it is perceived as having a high level of arousal.

Some links between emotional valence and the level of arousal can be inferred. If the avatar's emotional valence is perceived as positive or negative, the associated level of arousal is never low.

This confirms the second hypothesis.

3.4.3 Comparison Between Low-end and High-end Headsets

This section investigates whether there is a difference in the confidence of assessing emotional valence and level of arousal when different headsets are used. T-tests were used to answer these questions. For consistency, only the results for the third set of non-verbal behaviours are compared. As explained in Section 3.3.3, both low-end and high-end headsets were used for the sequences of the third set. Therefore, only Sequences 21 to 40 will be considered in this section. The results of the comparison between these two headsets in terms of sense of presence are presented in Section 3.4.5).

The confidence level for the valence rating does not differ between the low-end headset ($mean_{low} = 5.972$) and high-end headset ($mean_{high} = 5.96$) (p -value=0.863). The confidence level for the arousal rating does not differ between the low-end headset ($mean_{low} = 6.252$) and the high-end headset ($mean_{high} = 6.266$) (p -value=0.820). For both tests, the headset used does not influence the confidence level, and confirms Hypothesis 3a.

3.4.4 Comparison Between Photo-realistic and Cartoon Avatars

Valence does not differ between the cartoon and the photo-realistic avatars ($mean_{cartoon} = 3.476$, $mean_{photo} = 3.519$, and p -value=0.427), nor does arousal ($mean_{cartoon} = 4.681$, $mean_{photo} = 4.687$, and p -value=0.427). For both tests, the non-verbal behaviours of the avatars are perceived in the same way for valence and arousal, regardless of the quality of the graphics used to represent them.

Finally, the seek was to determine whether the confidence level improves when assessing the level of valence and arousal for photo-realistic avatars. The confidence level for the valence rating is statistically different between the cartoon avatars and photo-realistic avatars ($mean_{cartoon} = 5.729$, $mean_{photo} = 6.228$, and p -value < 0.001), as is the confidence level for the arousal rating ($mean_{cartoon} = 6.067$, $mean_{photo} = 6.577$, and p -value < 0.001). Both tests confirmed Hypothesis 4.

3.4.5 Sense of Presence

Overall, participants rated presence in virtual reality as 66.76% on average based on the Gatineau Presence Questionnaire. Participants found their experience half realistic (average of 51.1%), and the virtual reality environment highly artificial (average of 84.16%). They forgot about their presence in the actual room (average of 58.32% for spatial awareness). The average of these four items is 65.085%.

T-tests, on the results of the Gatineau Presence Questionnaire for both the low-end and high-end headsets, were performed. On average, presence in virtual reality does not differ between headsets ($mean_{low} = 64.78\%$, $mean_{high} = 65.35\%$, and $p\text{-value} = 0.962$). Apart from the level of artificiality ($mean_{low} = 83.33\%$, $mean_{high} = 85.58\%$, and $p\text{-value} = 0.498$), using the high-end headset compared to the low-end headset increases presence in virtual reality ($mean_{low} = 62.46\%$, $mean_{high} = 73.14\%$, and $p\text{-value} = 0.002$), the level of realism ($mean_{low} = 46.44\%$, $mean_{high} = 57.3\%$, and $p\text{-value} = 0.018$), and decreases spatial awareness⁵ ($mean_{low} = 66.9\%$, $mean_{high} = 45.4\%$, and $p\text{-value} < 0.001$). Thus, there is a significant difference for presence, realism, and spatial awareness. Only artificiality does not seem to have changed, which is confirmed by a t-test. As a result, using the high-end headset improves the sense of presence in general⁶. This supports Hypothesis 3b.

3.5 Qualitative Assessment

In this study, participants were given the opportunity to provide written comments on the avatars as they watched the sequences, allowing them to nuance their responses or provide valuable insights into their experiences. This section is not intended to be exhaustive, but rather to highlight certain aspects mentioned in the previous sections or add further insights.

First, participants felt able to imagine realistic situations that went beyond the sense of presence. As one participant noted, "This is an attitude that my boss could have."

Second, participants found it more difficult to evaluate emotions due to the absence of micro-expressions. One stated, "There are not enough micro facial expressions". On the other hand, the high-end headset provided a more detailed avatar, which interestingly made participants perplexed. As one participant mentioned, "The corner of the mouth changes everything. It looks like he is mocking or if we agree on a sensitive issue. Never sure about the level of alertness."

⁵This means that with the high-end headset, participants are more likely to forget they are in the actual room.

⁶In the Gatineau Presence Questionnaire, spatial awareness is rated in reverse order to presence in virtual reality and level of realism.

Third, some comments related to the potential difficulty of interpreting some expressions. Three different participants who watched an avatar smiling and shaking her head mentioned “The smile doesn’t go with the head movement, it looks like mockery”, “Her smile looks fake”, and “I don’t like her wry smile”. Cultural specificities also played a role in participants’ interpretations. Without context, it was sometimes difficult to determine the meaning behind certain behaviours. As one participant pointed out, “If I had a European person in front of me, they would say yes, but if it was an Indian person, they would say no”, when referring to a shake of the head.

In addition, participants sometimes expressed their own emotional reactions to the avatar’s emotions, which was not a specific aspect of the study. As one participant noted, “The sighs suggest an indifferent behaviour, which annoys me” or “He’s scary, he doesn’t look happy, I want to run away”.

Finally, participants also expressed their own perceptions of the avatar’s emotions: “She looks quite angry” or “He looks happy and joyful”.

3.6 Discussion

This study examines how people perceive a virtual audience provided by virtual reality technology and how to select non-verbal behaviours to faithfully represent a range of audience reactions. The main contributions can be summarised as follows.

First, the study of perceived emotional valence and arousal for pre-defined parameters individually and in combination provides insights into how different combinations of non-verbal behaviours can be used to express specific levels of valence and arousal. Second, the comparison of the valence and arousal results when using low-end versus high-end virtual reality headsets shows that the sense of presence is improved with high-end headsets. Third, the comparison of the valence and arousal results when using cartoon avatars versus photo-realistic avatars shows that photo-realistic avatars improve the confidence level of participants’ judgments without changing their assessment of valence and arousal. The results provide a typology of perceptions of valence and arousal, while clearly identifying the sequences associated with positive, neutral, and negative valence, and a low, neutral, or high level of arousal.

This study builds on the results of [63]’s work focused on participants creating combinations of non-verbal behaviours from a list (body postures, facial expressions, and head movements) to express the desired level of arousal and valence. In the present study, the focus is on immersed users’ perceptions of some of the resulting non-verbal combinations in virtual reality rather than only on how creators might build such combinations in a two/three-dimensional setting (as in [63]). Moreover, their experiment was

conducted on a flat screen via the web, rather than in a full virtual reality setting (with a head-mounted display), so it is unknown whether their results can be generalised to virtual reality. The aim of this present study is to determine whether the experiment medium influences the results. Another area of interest is the measurement of valence and arousal for specific non-verbal behaviours separately.

An interesting observation from these findings is the perception of smiles. In this study, smiles are primarily neutral, and raised eyebrows are associated with negative valence. This finding is notable because smiles are typically perceived as positive. However, based on participants' comments, this may be due to the fact that participants cannot distinguish between genuine and fake smiles. This ambiguity may be influenced by the combination of the smile with other non-verbal behaviours in the sequence, such as head shaking. This analysis is consistent with [206]'s work on smiles.

Furthermore, this study shows that photo-realistic avatars can increase confidence levels by making it easier for participants to associate the avatar's non-verbal behaviours with real-life experiences (as also highlighted in the qualitative assessment and in [87]). This idea is consistent with [339] who find increased trustworthiness and affinity with human-realistic avatars. Despite the uncanny valley phenomenon implying an aversion to near-realistic avatars, our findings are consistent with [339]. Indeed, they conclude that it is now possible to overcome the uncanny valley using real-time-rendered human realistic avatars, and we believe our similar results support this notion.

In addition, this study shows that emotional valence is easily perceived regardless of the headset used, although high-end headsets provide more accurate details. Valence may not be directly affected when non-verbal behaviour combinations are easily visible. However, arousal levels can be influenced by facial expressions, with high-end headsets providing clearer details, particularly around the eyes, resulting in higher perceived arousal. This is consistent with [279], as high-end headsets enhance the participant's sense of presence, making it easier to associate the avatar's non-verbal behaviours with real-life experiences. Surprisingly, high-end headsets do not increase confidence levels, possibly because their effect is already evident in the arousal assessment. Another explanation can be linked with the uncanny valley effect, as explained in [95]. The comparison between low-end and high-end headsets highlights the effectiveness of future virtual reality training environments using simple Cardboard, addressing issues of affordability and mass diffusion.

In this study, the Gatineau Presence Questionnaire was used to measure the sense of presence in virtual reality. The average score obtained is 65.085%, consistent with the studies of [209] and [37] who report scores between 48.11% and 65.73%. The lower than expected presence score in this study may be due to the time and disorientation of leaving and re-entering

the virtual reality environment, as explained in [334]. Moreover, asking participants about their sense of presence can lead to biased results and the so-called break-in-presence, as discussed in [180] and [344].

Overall, this study demonstrates the complexity of interpreting non-verbal cues and emotions in virtual environments. The participants' comments shed light on the subjective experiences and the impact of various factors on their perceptions.

These findings have significant implications for the use of virtual reality technology in training frontline staff and improving their interactions with consumers. By identifying specific non-verbal behaviours associated with certain levels of emotional valence and arousal, this study offers guidance on designing avatars in virtual reality for training purposes. For example, using photo-realistic avatars can improve participants' confidence in their valence and arousal judgements, potentially leading to more effective training outcomes. Similarly, using high-end virtual reality headsets can enhance the sense of presence, resulting in a more realistic training experience. Understanding the link between non-verbal behaviours and emotional valence and arousal can help tailor training content to specific situations and interactions that employees may encounter with customers. This can lead to more targeted and effective training programs, ultimately resulting in improved employee behaviours and better customer experiences.

3.7 Limitations and Future Research

This study has a number of limitations that should be taken into account when interpreting the results. One such limitation is the fact that certain sequences were performed exclusively by either female or male avatars, despite attempts to control for gender-related factors. This limitation may have influenced the results, and future research could address this issue by using a more balanced design that includes an equal number of male and female avatars executing all sequences. In addition, future research could explore the complex ways in which gender may interact with other factors, such as emotional valence and arousal levels, to better understand the impact of gender on avatar non-verbal behaviours. Another limitation relates to the need for caution when interpreting the results and their generalizability to different populations. For instance, the interpretation of nods varies across cultures, and this factor may have influenced the present results. Moreover, the number of combinations of animations that we were able to test in this study was limited due to the nature of the experiment based on true virtual reality immersion. Finally, the methodology is limited by the need for participants to leave and re-enter the virtual reality environment to complete the questionnaire, which may also have influenced the results.

The participants' comments also highlight the complexity of interpreting emotions in virtual environments and the need to consider cultural influences when analysing virtual interactions. This unexpected finding provides insights into how participants perceive and internalise emotions, offering potential avenues for future research.

In conclusion, this study provides valuable insights into the emotional design of avatar non-verbal behaviours in the virtual reality context. Furthermore, thanks to the typology developed in this paper, it is now clear which sequences to choose in order to demonstrate specific levels of arousal (low, neutral, and high) and emotional valence (negative, neutral, and positive). This can be very useful when designing new virtual reality environments for training to determine the non-verbal behaviour to employ in response to a particular situation. As well as contributing to understanding virtual interactions, this research opens up new possibilities for applications, such as public speaking training, highlighting the virtual reality potential to simulate future scenarios.

Future work will therefore aim to use these avatar non-verbal behaviours to elicit appropriate responses from the avatars during a user's training in virtual reality, such as gradually increasing the induced anxiety or rewarding good performance. There will be two types of avatars: other participants (trainees, experts, teachers, etc.) and artificial intelligence driven avatars [48]. A study on automatic methods based on statistical, machine learning, and natural language processing methods to implement real-time feedback from the audience to the speaker's presentation is under consideration. While the first application being considered next is the training in front of multiple avatars (public speaking in general), the present results could be useful in many other contexts.

3.7. LIMITATIONS AND FUTURE RESEARCH

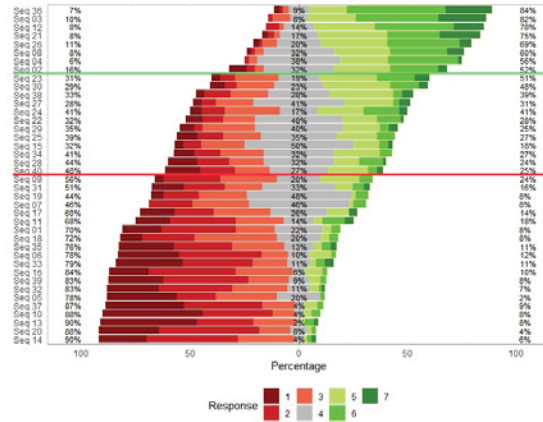


Figure 3.5: Valence per sequence evaluated on a seven-point Likert scale. All sequences above the green line are perceived as positive, and all sequences under the red line are perceived as negative.

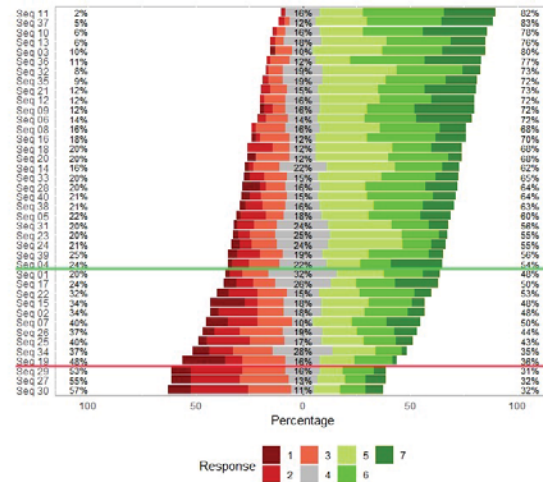


Figure 3.6: Arousal per sequence evaluated on a seven-point Likert scale. All sequences above the green line are perceived as positive, and all sequences under the red line are perceived as negative.

Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6	Seq. 7	Seq. 8	Seq. 9	Seq. 10
P1F3H4	P3F4H2	P7F2H2	P7F2H4	P7F3H3	P7F4H3	P7F4H4	P9F1H2	P9F2H3	P9F3H3
Seq. 11	Seq. 12	Seq. 13	Seq. 14	Seq. 15	Seq. 16	Seq. 17	Seq. 18	Seq. 19	Seq. 20
P2F2H3	P2F4H2	P3F3H3	P3F4H3	P4F4H1	P4F4H3	P5F3H1	P7F3H1	P8F4H1	P9F1H3
Seq. 21	Seq. 22	Seq. 23	Seq. 24	Seq. 25	Seq. 26	Seq. 27	Seq. 28	Seq. 29	Seq. 30
P2F1H2	P2F4H4	P3F1H2	P3F3H2	P3F4H4	P4F1H2	P5F1H4	P6F3H4	P7F1H1	P7F1H2
Seq. 31	Seq. 32	Seq. 33	Seq. 34	Seq. 35	Seq. 36	Seq. 37	Seq. 38	Seq. 39	Seq. 40
P1F3H1	P2F4H3	P3F2H3	P3F4H1	P4F1H3	P4F2H2	P6F3H3	P7F1H4	P7F4H3	P8F3H4

Postures	Facial expressions	Head
1: Backward posture – Arms crossed	1: None	1: None
2: Backward posture – Arms down	2: Smiling	2: Nod
3: Backward posture – Arms behind head	3: Frowning	3: Shake
4: Upright posture – Hand on hand	4: Eyebrows raised	4: Questioning
5: Upright posture – Hands together		
6: Upright posture – Hands separated		
7: Forward posture – Hands together		
8: Forward posture – Arms stand		
9: Forward posture – Arms crossed		

Recall of Tables 1 and 2 for the sake of readability

3.8 Appendix

3.8.1 Analysis of Postures, Facial Expressions, and Head Movements

A Chi-square test was conducted to see if the participants' responses were evenly distributed between negative interpretations (answers 1 to 3 in the Likert-type scale) and positive interpretations (answers 5 to 7 in Likert-type scale)⁷; first for valence (Table 3.5) and then for the arousal (Table 3.6). The aim is to determine whether the difference between the proportions of positive (denoted with p_+) and negative interpretations (denoted with p_-) are statistically significant. The p-values obtained for this test are presented in the two tables. If the difference between the proportions of positive and negative interpretations is not statistically significant, it is denoted with $"/"$. If not, and if there are more positive than negative interpretations, it is denoted with $"+"$. If there were more negative than positive interpretations, it is denoted with $"-"$. The two percentages correspond to the proportion of participants with a negative interpretation (1, 2, or 3) and a positive interpretation (5, 6, or 7). The proportion of neutral answers (4) is thus not in the table but can be easily derived⁸.

⁷Throughout this paper, and for clarity, all responses from 1 to 3 in the Likert-type scale will be referred to as "negative" and those from 5 to 7 in the Likert-type scale as "positive". The words "negative" and "positive" must be used with caution. For valence, it can be understood as having negative and positive levels. For arousal, it can be understood as low and high levels.

⁸100 % minus the two other percentages.

Table 3.5: P-values and interpretations for valence per non-verbal behaviour. (Please refer to Table 3.1 and Section 3.4.1 for the notations)

		Proportions		P-value	Interpretation
		p_-	p_+		
Postures					
P1	Backward posture – Arms crossed	48%	32%	< 0.001	-
P2	Backward posture – Arms down	49%	30%	< 0.001	-
P3	Backward posture – Arms behind the head	55%	15%	< 0.001	-
P4	Upright posture – Hand on hand	40%	39%	0.887	/
P5	Upright posture – Hands together	56%	22%	< 0.001	-
P6	Upright posture – Hands separated in front	64%	21%	< 0.001	-
P7	Forward posture – Hands together	38%	38%	0.918	/
P8	Forward posture – Arms down	52%	21%	< 0.001	-
P9	Forward posture – Arms crossed	50%	30%	0.002	-
Facial expressions					
F1	None	36%	41%	0.185	/
F2	Smiling	38%	43%	0.305	/
F3	Frowning	57%	26%	< 0.001	-
F4	Eyebrows raised	50%	23%	< 0.001	-
Head movements					
H1	None	50%	28%	< 0.001	-
H2	Nod	22%	60%	< 0.001	+
H3	Shake	70%	14%	< 0.001	-
H4	Questioning	38%	29%	0.028	-

Table 3.6: P-values and interpretations of the test where the null hypothesis is $p_- = p_+$ for arousal per non-verbal behaviour. (Please refer to Table 3.1 and Section 3.4.1 for the notations)

		Proportions		P-value	Interpretation
		p_-	p_+		
Postures					
P1	Backward posture – Arms crossed	21%	52%	0.0006	+
P2	Backward posture – Arms down	18%	68%	< 0.001	+
P3	Backward posture – Arms behind the head	26%	51%	< 0.001	+
P4	Upright posture – Hand on hand	23%	60%	< 0.001	+
P5	Upright posture – Hands together	43%	40%	0.8262	/
P6	Upright posture – Hands separated in front	16%	70%	< 0.001	+
P7	Forward posture – Hands together	29%	56%	< 0.001	+
P8	Forward posture – Arms down	34%	53%	0.0536	/
P9	Forward posture – Arms crossed	24%	60%	< 0.001	+
Facial expressions					
F1	None	31%	52%	< 0.001	+
F2	Smiling	18%	66%	< 0.001	+
F3	Frowning	22%	58%	< 0.001	+
F4	Eyebrows raised	29%	55%	< 0.001	+
Head movements					
H1	None	33%	50%	0.0004	+
H2	Nod	22%	60%	< 0.001	+
H3	Shake	18%	65%	< 0.001	+
H4	Questioning	35%	47%	0.0049	+

Chapter IV

A framework for the implementation of virtual reality in legal education: A mixed methods and multiple case study investigation

This chapter presents a collaboration done from January 2023 to December 2024 between the Manchester Metropolitan University (MMU) and the University of Liège based on Justin Cho's subject thesis. Justin Cho is a Ph.D. candidate from MMU in the Faculty of Business and Law under the supervision of Timothy Jung. The present chapter is now a paper, for which I am the second author, that has been submitted to the peer-reviewed Journal *Studies in Higher Education* in 2025. The co-authors are thus Justin Cho, Michaël Schyns, and Timothy Jung. Although the co-authors themselves acknowledged that my contribution was equal to that of the first author, we agreed that Justin would be listed first, as the project is based on his main thesis topic. In this thesis, the paper is presented in its original submitted form, and I apologise for any outdated information, repetitions, or formatting inconsistencies, including the use of numbered citations instead of author names, without modifying the original text.

Abstract

This study explores the implementation of Virtual Reality in higher legal education, focusing on mock trials as a pedagogical tool. Building on the limitations of traditional simulation methods, a high-fidelity VR environment was designed to enhance immersion and engagement. The study

integrates insights from legal education literature and empirical feedback from 60 participants across two universities in the UK and Belgium, using a mixed-method approach combining qualitative interviews and quantitative questionnaires. Results highlight VR's potential to foster knowledge acquisition, skill development, and confidence through immersive and contextualised learning. The study proposes the Immersive Experiential Learning Theory, an extension of Kolb's framework, tailored for VR learning contexts, offering theoretical and practical contributions to the design of VR environments for legal education.

4.1 Introduction

The Carnegie Report on educating US lawyers identifies in 2007 three key dimensions: legal knowledge, practical skills, and ethical awareness [352]. Effective practice requires proficiency in all three [235, 134]. However, the UK Legal Education and Training Review (LETR) in 2013 highlighted gaps in the UK higher education system in training future lawyers. Simulated learning has traditionally been used to teach various skills [84]. Unlike traditional methods, simulations provide a contextualised and active learning environment [377, 293, 84]. However, simulations have limitations, such as lack realism [377]. Technologies are now being used to enhance legal simulations to increase immersion and engagement [270, 377]. For example, reference [239] developed a virtual environment for role-playing as solicitors, actively engaging students in a professional context. Virtual reality (VR) is increasingly used in various higher education learning contexts [194]. Its visual and auditory characteristics enhance student immersion and engagement [389, 194]. VR creates rich, contextualised learning experiences, closely replicating real-life scenarios to help students develop skills [194, 65, 111]. Evidence supports VR as a useful tool for simulated learning in legal education. Although VR has been explored in areas such as healthcare [183] and soft skills training [111] its use in legal education is limited. This study addresses three main gaps in the literature. Firstly, it integrates learning design and technical feasibility to explore VR's potential in higher legal education, addressing the lack of empirical evidence. Secondly, existing learning theories are unsuitable for immersive contexts [149]. This study proposes a novel conceptual theory, the Immersive Experiential Learning Theory, for immersive learning. Thirdly, despite the importance of environment fidelity in technology-enhanced learning, few studies use high-quality environments. This study develops and proposes a high-quality virtual learning environment tailored to legal education. Based on the research questions, proposed conceptual theory, and novel immersive learning environment, this study aims to propose a theoretical framework for implementing VR in legal education. Section 2 explores key themes in legal simulation and VR

education studies. Section 3 examines theoretical foundations in existing research and extends experiential learning theory to immersive education for further empirical investigation. Section 4 discusses methodological choices. Section 5 analyses findings and proposes the novel immersive experiential learning theory for legal education. Section 6 discusses contributions, future research directions, and limitations.

4.2 Literature Review

4.2.1 Legal Simulation

Various forms of legal simulation facilitate different learning objectives, including clinical teaching, mock trials, and case studies [8]. Clinical teaching involves students acting as legal advisors in law clinics, supporting professional clinicians and performing tasks like legal research, client interviews, and document drafting [362]. Mock trials require students to prepare and present legal arguments in a courtroom setting, enhancing their research and advocacy skills [392, 76]. Case studies present problem scenarios for students to apply their legal knowledge to practical situations [127, 38].

4.2.2 Impacts of Simulated Learning Techniques in Legal Education

Previous studies have extensively investigated the impacts of simulated learning. Simulations facilitate knowledge acquisition [174, 221, 248]. The active nature of simulations helps students visualise and apply legal concepts, enhancing their understanding [38, 392, 261]. Reference [269] highlights that effective learning environments, like simulations, should emphasise active, problem-centred approaches, aligning with benefits in legal education. [177] suggest that this approach can help students identify gaps in their knowledge. Besides legal knowledge, simulations aid in acquiring practical skills [49] such as advocacy [75], legal research [56], case management [27], client interviewing [22], and document drafting [75]. More generally, simulations can enhance time management [377], communication [346], teamwork [265], and public speaking skills [392]. They also improve professionalism and ethical awareness [174, 346], fostering a “professional identity” where students take ownership of their work and act responsibly [177, 139].

Immersion plays a crucial role in simulations [221]. Contextualising legal concepts leads to deep learning [346, 127], and increased interactivity enhances understanding and knowledge retention [261]. The novel and active nature of simulations increases enjoyment [49, 108, 265], leading to greater engagement and motivation [116, 255]. Problem-centred learning approaches [269], support these findings by showing how constructivist

principles guide effective learning tool design, enabling the VR courtroom to contextualise legal concepts and foster critical thinking and skill development. Reference [349] notes that the practical relevance of simulations boosts student motivation, and [49] state that the active nature helps students focus on the learning experience.

Limitations of Simulated Learning in Legal Education

In contrast, studies have identified numerous limitations of simulated learning. Although simulations can improve legal knowledge acquisition, a lack of existing legal knowledge may hinder their effectiveness, as students have nothing to put into practice [56, 356]. Insufficiently prepared students are unlikely to benefit from simulations [168]. Simulations are resource-demanding, requiring extensive preparation from both staff and students [265, 100, 56]. Educators find simulations time-consuming, making it difficult to provide sufficient individual feedback, which is essential in legal education [100, 168, 116].

Furthermore, increased immersion is a significant advantage of simulated learning, but inadequate immersion or realism can discourage students and hinder learning [377]. For example, poor acting skills in simulations involving real actors can reduce immersion [56]. Additionally, accurately replicating the complex realities of a lawyer's role is challenging, limiting scenario authenticity [168]. Conversely, too much detail can result in cognitive overload, making simulations physically and emotionally draining and distracting students from the learning content [127, 346].

Scholars are increasingly incorporating technologies in simulated learning [358]. Reference [248] used a blended learning approach for case study simulations, employing online methods to handle large student numbers and provide personalised feedback. Reference [261] used a video game to create a narrative, increasing immersion and engagement. However, reference [270] argues that using technology without careful consideration can distract students and hinder learning. This theme is common in VR education literature, which will be explored further in the following subsection.

4.2.3 Virtual Reality Education

Technological innovations such as the advancement of internet speeds, computing power, and high-quality hardware have paved the way for immersive technologies [309, 119]. More generally, students in higher education are now accustomed to using digital technologies in their learning, also known as digital natives [159]. In particular, the use of VR in educational contexts has greatly increased due to its novelty and engaging nature [194]. This subsection explores the ways in which VR has been used in educational contexts and the impacts that the technology has had. Before exploring VR's

impacts in education, it is crucial to define virtual reality. There is confusion about the boundaries between immersive technologies [309, 119]. Reference [260] describes VR as computer-generated environments that fully immerse users. Reference [119] defines VR by high embodiment, presence, and interactivity. Reference [309] highlights VR's ability to create telepresence, fully immersing users by blocking out the physical environment. This study defines VR as "a computer-generated 3D fully immersive environment using sensory-enhancing and interactive devices to create telepresence."

4.2.4 Impacts of Virtual Reality on Learning

VR has numerous positive impacts in education. Studies show VR enhances both substantive and skills-based knowledge. For example, reference [185] found VR helps students visualise concepts, deepening their understanding of human anatomy. Reference [135] noted VR improves problem-solving skills. VR also enhances psychomotor skills like surgical [226] and resuscitation skills [290], which are transferable to real scenarios. VR can improve empathy [307, 151] and interpersonal skills through social interactions [161]. For instance, reference [307] created a VR simulation for users to experience life as a wheelchair user. VR learning experiences boost confidence, providing a safe context for making mistakes [322, 221].

VR's design flexibility also suits various learning contexts. In classroom management, reflection is key [5]. To facilitate this, the authors embedded a recording function in VR for performance reflection. In healthcare, VR can be used to enhance communication and teamwork skills [368]. Immersion and presence are crucial in VR learning. Increased immersion enhances learning [5, 1, 262]. Reference [310] stated VR provides detailed 3D visuals for understanding complex concepts. VR also immerses learners in emotionally engaging contexts, fostering empathy [307]. Greater immersion boosts motivation [165], engagement [94], and enjoyment [35]. Reference [94] noted immersion leads to deeper cognitive learning and engagement. Furthermore, enhanced engagement increases interest, motivation, and learning effectiveness [290].

4.2.5 Limitations of Virtual Reality in Education

Although many studies have found a positive impact of VR on learning outcomes, some studies found no impact [321, 240]. Both [321] and [185] used VR in human anatomy, but only [185] found a positive effect, suggesting other factors influence VR's effectiveness. Some studies found VR provided no additional benefit compared to desktop simulations and conventional learning [241, 305].

While VR's design flexibility can tailor experiences to learning objectives, inadequate design can hinder learning. For example, reference [193]

found conventional radiography training more effective than VR due to the lack of patient interactions within VR. Technical issues also affect VR use. Some studies found VR easy to use [243, 290], but others noted difficulties in navigation and limited physical space [210, 165]. Lower-quality VR applications also led to cybersickness [246, 165], hindering enjoyment [321]. However, the use of higher-end headsets reduced discomfort and cybersickness [243, 35], though [321] found these issues did not significantly impact learning outcomes.

Lastly, increased sensory stimulation in VR often leads to cognitive overload. Reference [355] noted the human mind's limited capacity to process information, and excessive input can hinder memory retention [78, 17, 119]. Compared to other learning methods, VR can cause higher cognitive overload, distracting learners from main tasks and reducing learning effectiveness [129, 241].

4.3 Framework Development

4.3.1 Identification of Research Gaps

There is theoretical evidence supporting the use of VR in legal education. Inadequate immersion in simulations can hinder learning [377]. In this regard, VR's novelty and sensory stimulation can enhance immersion [310]. However, cognitive load may make VR unsuitable, as too much detail can distract students [168, 129].

Despite VR's potential, empirical studies in legal education are limited. Two key studies should be noted. Reference [253] used mobile VR to improve law students' presentation skills, measuring engagement and perceptions of the usefulness of VR. They found low engagement because using VR was time-consuming to learn and poorly aligned with their curriculum. Reference [98] used 360 video VR to teach pleading skills, measuring scenario authenticity and perceived learning. Students found the VR experience anxious but useful.

[253] provided insights into VR's technological feasibility and suitability, while [98] focused on realistic learning scenarios. However, neither study holistically explores VR's impact on both the learning experience and technological feasibility. Effective VR educational experiences must be tailored to the learning objectives [270]. Indeed, existing VR studies highlight this need (e.g., [5, 368]). In this way, both VR's impact on learning and its technological suitability must be investigated to design effective VR educational experiences.

4.3.2 Proposal of Extended Kolb's Experiential Learning Theory

Many studies draw from learning theories to design and facilitate teaching methods. Academics in legal simulation and VR education literature emphasise the importance of strong pedagogical foundations to achieve desired learning outcomes [19, 302]. Key theories include experiential learning theory (ELT) [75, 290], problem-based learning theory [390, 185], and constructivism [151, 247].

ELT [202] is the most commonly used. It posits that learning occurs through experience, involving four stages: concrete experience, reflective observation, abstract conceptualisation, and active experimentation. Learners participate in tasks, reflect on their actions, conceptualise their learning, and apply new knowledge in subsequent tasks, creating a continuous learning cycle. ELT is widely used in practical and reflective learning contexts. For instance, reference [75] used ELT to develop a negotiation simulation, while [377] applied it in a mediation exercise, highlighting the role of reflection in deep learning. In VR education, reference [290] used ELT for resuscitation skills training, and [287] explored its effects on healthcare training effectiveness.

Despite their benefits, learning theories are underused in both legal simulation and VR education [238, 302]. Reference [149] suggested that existing theories may not suit immersive education contexts. Literature indicates that VR's technological suitability and immersion levels impact learning experiences. VR experiences should be designed with appropriate functions to create realistic simulations (e.g., [5]. Additionally, the increased interactivity and sensory inputs of VR require careful design to avoid cognitive overload, which can arise when excessive sensory details or overly complex tasks are presented, distracting learners from the main objectives and making it harder for them to retain information [129].

This concern has also been previously highlighted in the works of [86], where Kolb's theory [202] was extended to account for the rise of digital pedagogy. The authors noted that the increase in e-learning tools provided learners with more autonomy over their learning and suggested the need for learning theories to account for this new "exploratory" step to learning [86]. As a result, the authors propose a 5-stage exploratory learning model shown in Figure 4.1.

In a similar vein, a conceptual framework is proposed based on Kolb's experiential learning cycle. This framework extends Kolb's theory to include the influence of VR technological suitability and immersion to account for the new factors that can contribute to the overall immersive learning experience (see Figure 4.2).

Using this proposed conceptual framework, and based on the gaps identified in the literature, the following research questions are proposed for the present study.



Figure 4.1: The Exploratory Learning Model

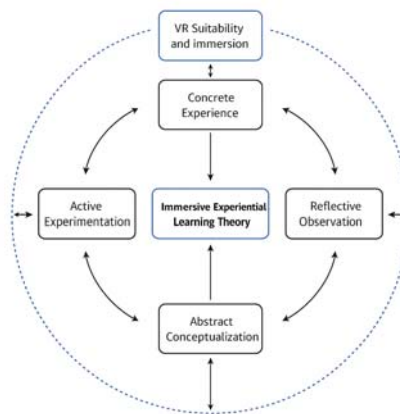


Figure 4.2: The Immersive Experiential Learning Theory

1. How does VR technology impact the simulated learning experience in legal education?
2. How suitable is VR technology for facilitating simulated learning in legal education?
3. How does immersion of VR influence the learning experience in a legal education context?

4.4 Methodology

4.4.1 Virtual Reality Environment Design

A high-quality VR environment, based on [324] and [111], was developed to address the lack of advanced virtual mootting settings. Unlike existing simulated learning tools in legal education [8], this innovative VR setup supports a mock trial format to help students cultivate essential legal skills [392].



Figure 4.3: Virtual Reality mock trial courtroom environment. User is positioned at the lectern. The opposing counsel is seated to the left.

To enhance immersion, the VR environment was developed in Unity 3D, modeled on an actual courtroom, and accessed via Meta Quest 2 headsets. High-fidelity design, shown to improve real-world skill transfer [304] and task performance [4], included photorealistic avatars displaying realistic nonverbal cues, such as posture adjustments for valence and arousal [111], addressing VR limitations noted by [338, 339]. Recognising the importance of supporting diverse student groups, the design also reflects insights from [55], leveraging technology to foster equitable interaction (e.g., by offering diverse agents representations and creating neutral spaces that minimise the impact of race, gender, or social background) and reduce barriers to communication. Avatars were pivotal in enhancing users' sense of presence [333, p.445], with behavioural fidelity proving more critical than visual fidelity for achieving realism and effective immersion [333].

The VR environment was designed to meet educational goals, following research on effective technological learning tools [270, 302] and incorporating principles of active, problem-centered learning to foster en-

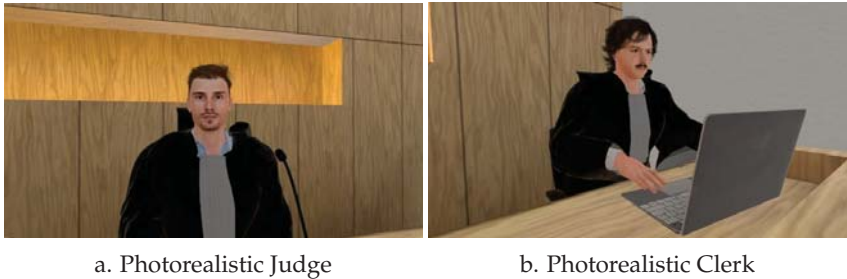


Figure 4.4: Avatars

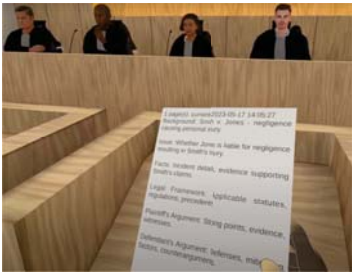
gement and skill development [269]. Key features include a recording tool [324] for replaying presentations via avatars (see Figure 4.5) and reflecting on performance with verbal (voice and transcript) and nonverbal (eye tracking, body, head movements) feedback as preconised by [49] (see Figure 4.6). Additional features, like a timer and a notes function (see Figure 4.7b), enhance realism, with notes prepared through a companion website (see Figure 4.7a) before the simulation [349, 324].



Figure 4.5: Replay scene. An avatar is placed at the user’s position and replicates his speech and movements.



Figure 4.6: Verbal and nonverbal feedback given using a companion website



- a. Website interface used for note encoding prior to immersion.
- b. View of the notes in VR.

Figure 4.7: Notes

4.4.2 Refining the Virtual Reality Environment Through Educator Feedback

Before student implementation, the VR courtroom was rigorously tested by legal education experts from UK and Belgian universities, who provided valuable feedback. Positive remarks highlighted its immersive quality and anxiety-reducing potential. A Belgian professor noted,

“VR offers unique advantages for students who may struggle with anxiety or shyness, providing a controlled environment that allows them to build confidence without the immediate pressures of a real audience” [PB1].

UK professors praised its repeatable practice benefits:

“You are able to repeat the experience and personalise it as well... it’s about practicing the moot multiple times without actually needing to go into a mock trial” [PA1].

Critiques led to significant enhancements. A UK professor suggested,

“There needs to be some sort of responsiveness from the court... nodding or occasional affirmation” [PA2],

prompting the addition of nonverbal cues. Lip synchronization was introduced after feedback stating,

“Initially I didn’t know who was talking... the lips weren’t moving as they would normally” [PA4].

Usability concerns with note-taking were addressed by creating a virtual note-holding system after a professor remarked,

“There was no way I could hold [the paper] in a way that I could actually read it” [PA6].

Accessibility improvements included a step-by-step guide based on the suggestion,

“Adding a detailed introduction at the start would help” [PA9].

These expert-driven adjustments preserved the VR tool’s strengths while addressing key limitations, creating an immersive and effective training environment for law students.

4.4.3 Mixed Methods Strategy

The study employed two comparative case studies at universities in the UK and Belgium [391, 105], chosen for their contrasting educational and cultural contexts where societal and institutional norms interact with individual assumptions to influence design and implementation (Heaton, 1998). It provided valuable insights into learning tool expectations and theoretical development. While the UK’s common law system emphasises case law, and Belgium’s civil law system is codified, these differences were deemed irrelevant as the study focused on practical skills and legal knowledge application [377, 233]. A mixed-method approach was used, with participants completing a quantitative questionnaire and a subset participating in semi-structured interviews. Initially, students familiarised themselves with VR in a waiting room before delivering their courtroom presentations. This period allowed exploration of VR functionalities. Qualitative data

were gathered from 9 UK participants (A1–A9) and 14 Belgian participants (B1–B14), with recruitment continuing until data saturation. To ensure adequate sample size for quantitative analysis, 21 additional UK participants and 16 Belgian participants were recruited, totaling 30 from each university. Ethical approval was received for this study, and informed consent from all participants was received.

4.4.4 Procedure and Participants

All participating students had a mock trial or exam scheduled in the days or weeks following the study. Before starting the study, students provided informed consent to participate and to be recorded by the VR system. They then encoded their initial presentation notes using a dedicated website (see Figure 4.7a). Next, they completed a pre-immersion questionnaire assessing their anxiety about presenting in a real courtroom for their actual moot, not in VR (measured on a percentage scale) and their initial opinion about VR as a learning tool (measured on a 7-point Likert scale, where 1 means not at all useful and 7 means totally useful).

After completing the questionnaire, students entered the VR environment and delivered their speech in the virtual courtroom.

Immediately after the immersion, they completed the 24-item Presence Questionnaire, in either English [381] or French [314], assessing dimensions such as Realism, Interaction, Examination, Performance, Sounds, Haptics, and Interface Quality, using a 7-point Likert scale. Presence scores were calculated following established formulas [314] and normalised in this study to provide a comprehensive view of participants' immersive experiences. Following this, participants assessed the extent to which they felt they mastered the content of the case they were assigned to defend (mastery of the presentation, using a percentage scale) and completed a post-immersion questionnaire measuring their anxiety about presenting their case in an actual mock trial or exam setting. They then received feedback via the website and had the opportunity to replay their performance in VR and through the website. After having received the feedback, participants reassessed their mastery of the presentation and answered an updated questionnaire about their opinion on VR. Finally, participants were asked about their willingness to reuse VR in the future, measured on a 7-point Likert scale (with 1 meaning not at all likely and 7 meaning extremely likely).

A subset of participants subsequently took part in a qualitative interview conducted using a semi-structured interview guide (see Table 4.3 in the Appendix section). Semi-structured interviews explored the learning experience, VR's suitability, and the role of immersion. The questions were aligned with Kolb's experiential learning cycle [202] (concrete experience, reflective observation, abstract conceptualisation, and active experimentation) to ensure that the concept of *learning experience* was clearly understood by

participants and to capture perceptions and impacts at each stage. Thematic analysis [43] was employed to analyse transcripts, codify data, and identify key themes, with cross-case synthesis used to compare findings across participants [391]. The study involved 60 participants, with a balanced gender distribution and a diversity of native languages. Among them, 23 participated in qualitative interviews. Regarding prior VR experience, 63% of participants had no previous exposure to VR, while those with prior experience mainly used it for entertainment purposes. Full demographic details are provided in Appendix (see Section 4.8).

4.5 Findings

A total of 3 themes and 8 sub-themes were identified through the analysis of the qualitative data, in direct relation to the research questions.

4.5.1 Enhanced and Contextualised Learning

The first theme focuses on how VR can enhance the overall learning experience through a contextualised environment and additional tools tailored to the learning objectives. Under the overarching theme of learning, 4 main sub-themes related to the learning experience were identified: Knowledge and skills acquisition, Accuracy of learning context, Reflection and feedback, and Novelty of learning method.

Knowledge and Skills Acquisition

Overall, participants found the VR mock trial experience beneficial for knowledge and skills acquisition. Some viewed VR as an excellent supplementary tool, allowing them to practice what they had learned. Furthermore, applying knowledge to practical situations strengthened their substantive knowledge and improved confidence in their legal skills.

“It gives you a backbone on what you already know” [A8]

“Good to learn to speak” [B3]

The difference in participants’ pre- and post-immersion mastery scores was calculated, yielding values from -100 to 100. Positive scores indicate improved mastery, while negative scores suggest a decrease. These results highlight VR’s role in skill development: improvement reflects confidence-building, while decreases emphasise self-critical awareness and opportunities for targeted practice. In this study, scores ranged from -45 to 40, with an average change of -4.61, suggesting VR increases self-awareness by highlighting areas for improvement (see the sub-theme *Reflection and Feedback* for related quotes).

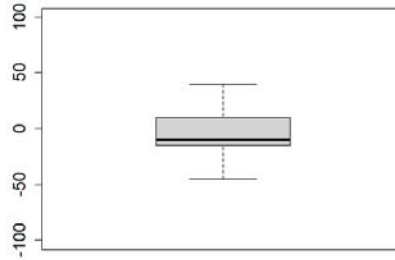


Figure 4.8: Boxplot illustrating the change in mastery of presentation.

Differences were noted between the two case studies. UK participants placed more emphasis on knowledge and skills acquisition than Belgian participants. This may be due to differences in how mock trials are used at each institution. At the UK institution, mock trials are used mainly for practising advocacy skills and preparing for competitions, whereas at the Belgian institution, they are used for assessments. These differences are reflected in participants' comments. For example, one UK participant seemed to implicitly divide their university degree from mock trials. Furthermore, one Belgian participant talked about a "d-day", discussing how their behaviour during the VR experience may differ from their examination.

"Solidify the knowledge that you've learnt in your degree" [A2]

"I think that there are things that I wouldn't do on the d-day (for example, throwing notes in the courtroom)." [B10]

Quantitative analysis revealed a statistically significant difference in self-assessed presentation mastery between Belgian and UK participants ($T=2.9063$, $df=22.282$, $p=0.0081$). Belgian students rated themselves lower post-VR (see Figure 4.9), indicating increased self-awareness of weaknesses, consistent with qualitative findings where they viewed VR as an assessment tool for critical self-evaluation. In contrast, UK participants maintained their initial assessment and saw VR primarily as a skill enhancement tool. This dual perspective underscores VR's role in promoting both self-awareness and skill development.

Accuracy of Learning Context

Many participants commented on the contextualization that VR provides, allowing them to get a feel for the courtroom setting. Linked to the

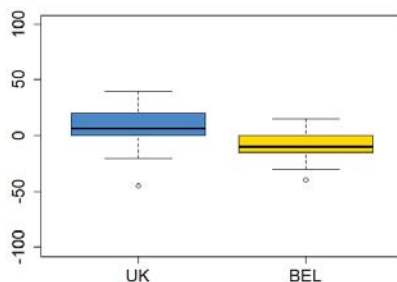


Figure 4.9: Boxplot illustrating the change in mastery levels pre- and post-immersion by case study.

theme of immersion, participants noted that their interactions with the environment and avatars helped them understand what it would be like to be a lawyer. One participant stated that this experience could influence students' career considerations.

“Give students the opportunity to see what other career options they might have” [A1]

Despite these benefits, participants highlighted limitations regarding interactivity. In an actual mock trial, participants would expect to be interrupted or questioned by the judge. The lack of these interactions in the VR environment changed the atmosphere and impacted learning outcomes.

“Regular mootings is more spontaneous and virtual mootings is more planned and straightforward” [A8]

“[It lacks] the ability to cut you off whilst you're talking” [A2]

Differences were noted between the two case studies. The UK study focused more on the learning context, providing feedback on expected interactions, whereas the Belgian study placed less importance on this factor. This may be due to institutional curriculum differences. For example, one Belgian participant stated,

“at the university, we never had a mock trial” [B2].

Reflection and Feedback

Participants stated that replaying their performance helped them identify unnoticed mistakes. The 3D representation of hand and head

movements made them more aware of their body language compared to simple audio recordings, allowing for critical reflection and improvement. Findings suggest that VR reflections provide richer insights into learner performances by going beyond audio.

“[The recording] makes you more aware of your body language... not just verbally but physically too” [A4]

“[It allows for] critical element of improvement” [A6]

“Nice to see it, otherwise I wouldn’t have noticed” [B1]

VR also facilitates enhanced feedback. One participant noted that VR recordings with 3D visualizations might be better for feedback than videoconferencing software, as

“things (for example, body language) [that] might be missed online, contrastingly might be noticed in virtual reality” [A6].

Another suggested that individual performance recordings would allow for more personalised feedback.

“Things [that] might be missed online, contrastingly might be noticed in virtual reality” [A6].

“Feedback, I think that would help... that is catered to you specifically” [A8].

As seen previously, change of mastery scores support these findings: for some students, increased self-awareness led to lower self-assessed mastery (See Figure 4.8), while others gained confidence through the visualisation of their performance, illustrating VR’s varying impact on participants.

Results also indicate a general reduction in anxiety levels following the VR experience. In the Belgian study, where mock trials are used for assessment, one participant noted that the lack of immediate feedback made the VR experience less anxious. Another valued the opportunity for self-feedback rather than receiving it from an educator.

“Not rated... less stressful” [B1]

“Good to have a neutral point of view, thanks to virtual reality” [B4]

The difference in pre- and post-immersion anxiety scores showed an average decrease of -18.87, ranging from -80 to 40, with a median of -20 (See Figure 4.10). A t-test confirmed this reduction was significant ($T=3.666$, $df=85.114$, $p=0.0004$). This suggests the controlled VR environment fostered a less anxious, more relaxed atmosphere, aligning with participants’ descriptions of VR as a “safe space” for skill practice in a non-judgmental setting.

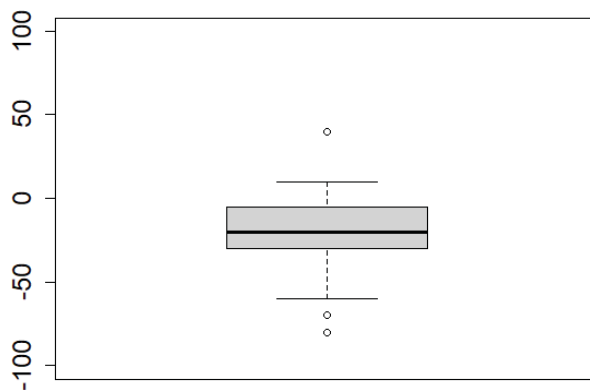


Figure 4.10: Boxplot illustrating the change in anxiety levels pre- and post-immersion, calculated as the difference between post- and pre-immersion percentage scores (ranging from -100 to 100). Positive values indicate an increase in anxiety, while negative values represent a decrease in anxiety following immersion.

A key difference between the case studies lies in their focus. The UK study emphasised VR as a platform for educator feedback, reflecting its use of mootings to prepare for competitions. This likely led UK students to seek educator input and compare VR with videoconferencing tools. Conversely, the Belgian study valued VR for self-criticism and individual learning, aligning with its use of mock trials for assessment. Belgian students appreciated VR as a risk- and anxiety-free environment for preparation.

Quantitative findings showed Belgian students rated their mastery lower post-task (See Figure 4.9), supporting qualitative insights that VR fosters self-reflection, especially in assessment contexts. In contrast, UK students, viewing VR as a preparatory tool, exhibited less change in self-assessed mastery. Although no significant difference in anxiety reduction between groups was found via t-test, Belgian students reported feeling less anxiety in VR due to the absence of real-time assessment. This underscores VR's role as a "safe space" for skill practice, reducing anxiety through the lack of immediate judgment.

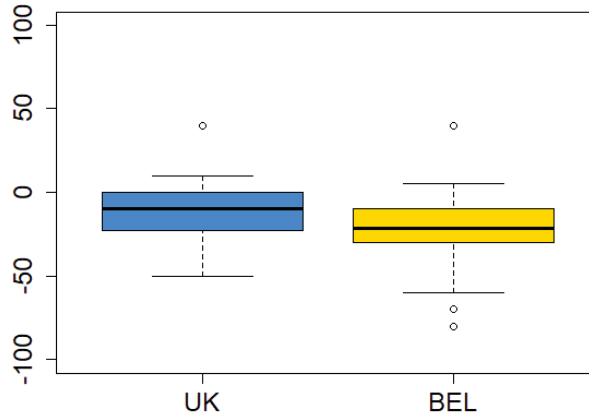


Figure 4.11: Boxplot illustrating the change in anxiety levels pre- and post-immersion by case study

Novelty of learning method

Participants felt that the new learning method increased their engagement and provided opportunities to experience courtroom settings they had not previously seen. They found the VR mock trial more engaging than conventional methods (e.g., simulations in a regular classroom, or online mock trials via videoconferencing, and PowerPoint-based case discussions), which are usually more passive. Furthermore, increased engagement made students likely to use the VR tool again.

“[It’s a] new thing for me.. Very engaging... take a different approach” [A1]

“I haven’t had the opportunity to use [a real courtroom]” [A6]

“Actually experience it for myself rather than just listening” [A1]

“It generates an interest and curiosity in everyone... may be inclined to return” [B11]

Participants also stated that VR provided a safe environment where mistakes would not have great consequences.

“[It’s a] safe environment. You can make a mistake without consequence” [A5]

Quantitative results confirm an overall reduction in anxiety, with decreases up to -80 points (See Figure 4.10), attributed to the VR environment’s

low-pressure nature. However, some participants experienced increased anxiety, likely due to the novelty of VR. No significant difference in anxiety reduction was found between groups (See Figure 4.11), consistent with qualitative insights that VR offers an engaging experience across educational contexts. Post-immersion opinions on VR as a learning tool showed a statistically significant improvement compared to pre-immersion opinions (See Figure 4.12, $T=-6.7771$, $df=109.38$, $p<0.001$). This aligns with qualitative findings, indicating participants recognised VR's value for enhancing mock trial learning despite usability challenges. No significant difference in opinions was observed between UK and Belgian participants, reflecting consistently positive attitudes across both groups.

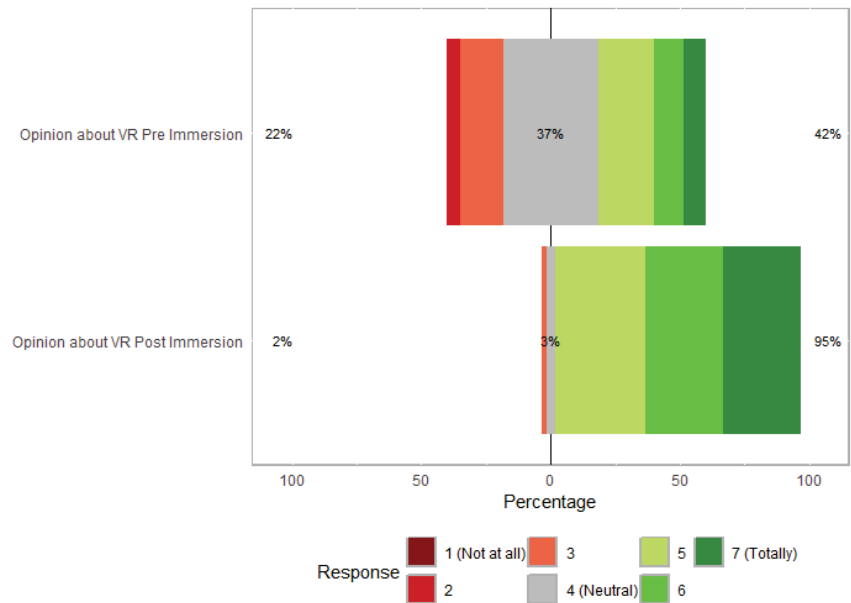


Figure 4.12: Opinions on the usefulness of VR as a learning tool, measured on a 7-point Likert scale ranging from 'not at all' to 'totally.' Percentages represent responses grouped as 1-3 (left), 4 (middle), and 5-7 (right).

4.5.2 Technological Feasibility of Virtual Reality for Mock Trials

The second theme focuses on the suitability of VR technology and its characteristics to facilitate mock trials. Importantly, the theme focuses on how these characteristics improve or limit the learning experience. Two sub-themes were identified: Usability/functionality, and accessibility.

Usability and Functionality

Many participants found the VR environment difficult to use, particularly with reading notes. However, some felt that the note function closely replicated real mock trials.

“It was quite difficult to read [the notes]” [A4]

“The notes were good... very similar to reality” [A2]

Quantitative findings from the Presence Questionnaire support qualitative impressions, with the Examination category achieving a mean score of 5.18, reflecting a positive experience in exploring the VR environment (See Table 4.1 and Figure 4.14).

Categories	Our findings (1–7)	[314] (1–7)
Realism	4.84 (0.8)	4.21 (1.72)
Interaction	4.92 (1.05)	5.19 (1.5)
Examination	5.18 (0.92)	5.12 (1.71)
Performance	5.16 (1.12)	5.50 (1.44)
Sounds	5.23 (1.29)	/
Haptics	5.04 (1.04)	/
Quality	2.98 (1.09)	5.12 (1.71)

Table 4.1: Presence Questionnaire results. Normalised means and standard deviations (in brackets) for the categories found in our study are shown, alongside the standard reference values established by [314] for the Presence Questionnaire. Note that sounds and haptics were not part of the French questionnaire and are thus not compared. Except for the quality of immersion category, the higher the better.

Some participants found the headset to be quite uncomfortable to use and noted that they felt sick after use, confirming the findings of [246].

“Made me feel a little sick.” [A3]

“Physically, it’s a bit annoying, it’s the headset.” [B7]

Similarly to [210], participants suggested adding more guidance to navigate the VR experience, though opinions varied.

“Adding a detailed introduction at the start [would help].” [A9]

The Quality of Immersion category scored a mean of 2.98, indicating good VR usability (see Figure 4.14 and Table 4.1, and note that for this item, lower scores are better). Self-assessed Performance in VR had a mean score of 5.16, with English participants rating their performance significantly higher than French participants ($T=3.44$; $df=56.501$, $p=0.0011$), possibly due to differing educational expectations or familiarity with VR.

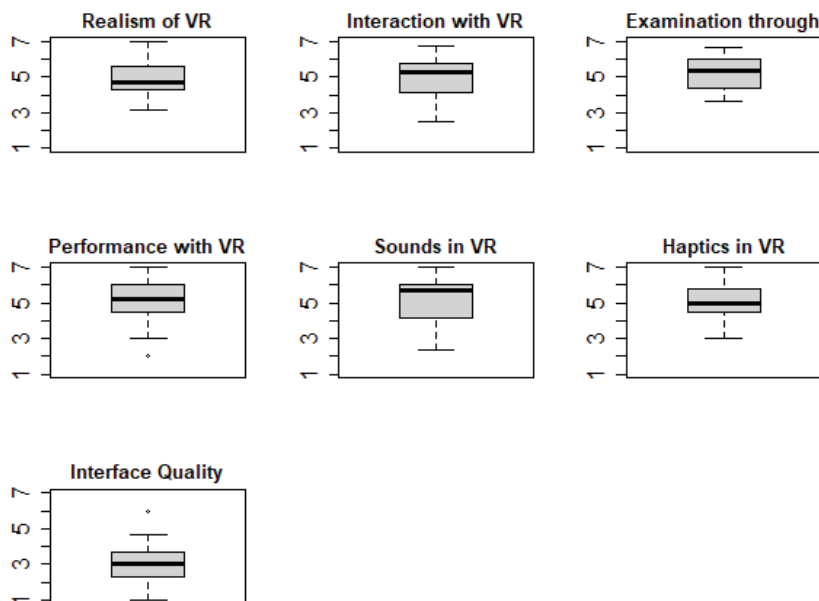


Figure 4.13: Boxplots illustrating the normalised Presence Questionnaire scores on a 7-point Likert scale for each category. Scores range from 1 (lowest) to 7 (highest), except for the quality category, where the scale is reversed.

Accessibility

Participants stated that VR as a learning tool could enable them to practice in their own time, making mock trials more accessible. VR allows access to the courtroom environment from any physical space, providing flexibility in when and where students practice. This accessibility may encourage more participation in mock trials outside of VR, as VR helps students build confidence and better understand the trial process, reducing common barriers to engagement.

“[The VR] would be really helpful to just practice in my own time” [A8]

“[I could] have that at home... I would have all day to train” [B1]

Participants also noted that VR could facilitate unique experiences, as gaining access to real courtrooms can be difficult.

“[Mooting] requires a lot of time and resources. . . [when using VR], people are more inclined to squeeze things in” [A7]

Quantitative results show that 95% of participants expressed willingness to reuse VR for legal training, with 5% remaining neutral (See Figure 4.13). This strong endorsement aligns with qualitative insights on VR’s accessibility benefits, highlighting its potential to make legal education more flexible and inclusive. No significant difference was observed between UK and Belgian participants, reflecting consistently positive attitudes across both educational contexts.

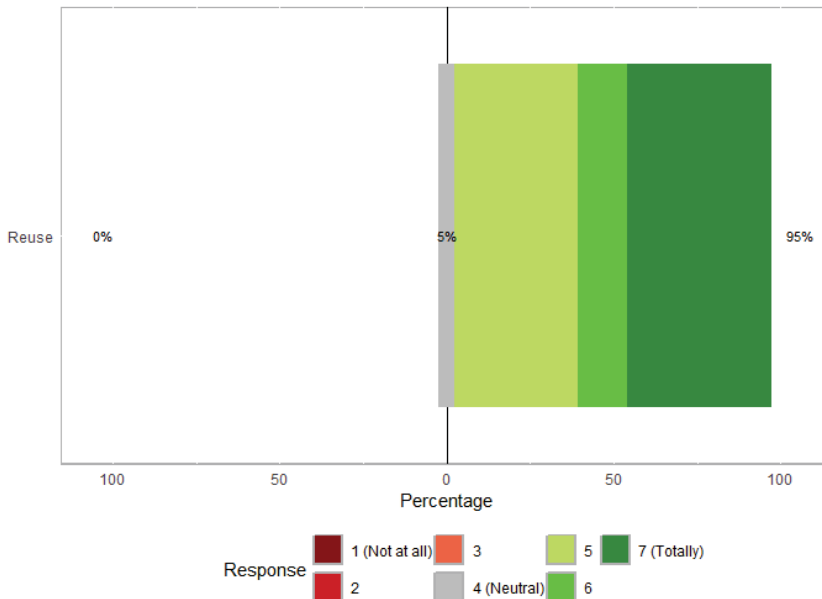


Figure 4.14: Willingness to reuse VR for learning, measured on a 7-point Likert scale ranging from ‘not at all’ to ‘totally.’ Percentages represent responses grouped as 1-3 (left), 4 (middle), and 5-7 (right).

4.5.3 Influence of Immersion on Learning

The third theme focuses more on the role of immersion into the VR environment in the learning experience. Specifically, this theme focuses on how the immersive qualities of VR contribute to or limit the learning experience. 2 sub-themes were identified: Realism and sense of presence/engagement using qualitative analysis.

Realism

Opinions on the realism of the VR environment were divided. Many found it unnatural and unrealistic, criticising the avatars' movements. Some felt that knowing the avatars weren't real reduced their immersion and learning, as it lacked the pressure of real people watching.

"The breathing was quite forced" [A3]

"Not real people and therefore the pressure won't be the same" [B5]

Similarly, participants noted that the environment didn't feel like reality, impacting their behaviour and the transferability of their learning.

"It is not completely adequate... you're not going to evaluate something based on an environment that isn't adequate for reality" [B8]

"I think that there are things that I wouldn't do on the d-day (for example, throwing notes in the courtroom)." [B10]

Conversely, some participants found the avatars' movements realistic, with their reactions and appearance creating a sense of immersion. Also, the courtroom environment accurately recreated courtroom etiquette, adding to the realism.

"There is an etiquette... a way to talk and present yourself in court" [A5]

"You could see that they reacted, they weren't robotic people who were static" [B5]

"We capture his attention with his head nodding... I have the impression that he is shaking his head and passing judgement" [B2]

Quantitative findings support these observations, with the Realism category scoring an average of 4.84 (Table 1, Figure 4.14). Participants noted that the courtroom setting did not accurately represent an English and Welsh court, which was expected as it was based on a Belgian court. This was not necessarily a problem for learning.

"[It] doesn't look like an English and Welsh court." [A2]

Figure 4.15 reveals that Belgian participants rated realism significantly lower than English participants ($T=5.8444$, $df=54$, $p<0.001$). This is likely due to the Belgian VR courtroom design reflecting their real-life context, prompting Belgian participants to critique the realism more harshly.

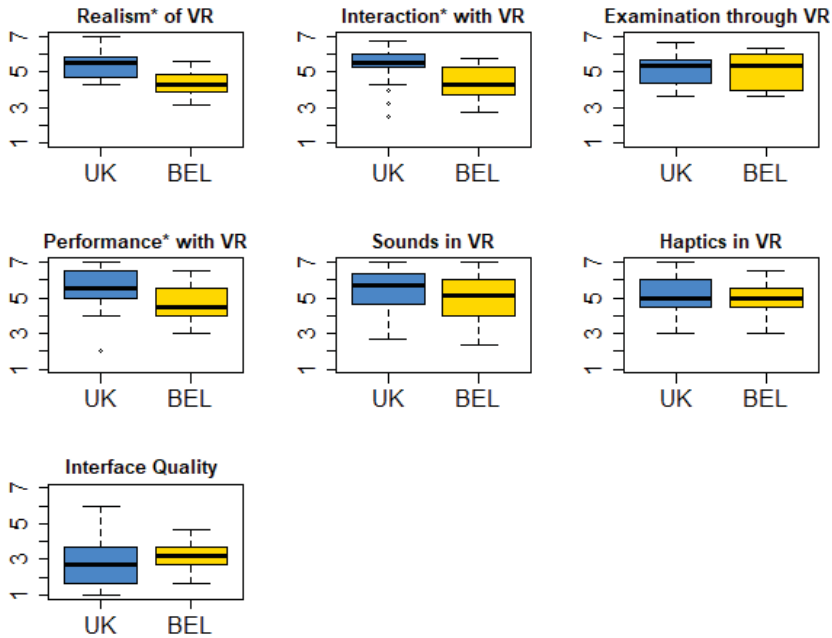


Figure 4.15: Boxplots illustrating, by case study, the normalised Presence Questionnaire scores on a 7-point Likert scale for each category. Scores range from 1 (lowest) to 7 (highest), except for the quality category, where the scale is reversed.

Sense of presence and engagement

Closely linked to realism is the sense of presence. Commenting on the avatars, one participant reported that the avatars' actions helped them feel as if they were actually present in the courtroom. Participants discussed how VR transported them to a realistic courtroom regardless of their physical location.

"[The avatars] actually looked at you and followed you" [A1]

"[It] felt like I was a real person in this virtual reality" [A4]

"[I can use it] in my living room, it really makes for a real environment" [B1]

The Interaction category scored a mean of 4.92, indicating generally positive engagement (Table 1, Figure 4.14). However, English participants reported significantly higher Interaction scores than Belgian participants

($T=4.2036$, $df=57.828$, $p<0.0001$; Figure 4.15), in VR familiarity and the novelty effect of using the technology (see Figures 4.18a and 4.18b). Being transported to a different environment helped students focus and stay engaged, separated from real-world distractions. This contrasts with concerns of cognitive overload highlighted in previous studies in VR [129, 241]. No comments were made about the VR environment being cognitively strenuous.

“Being able to put yourself into that mind space to optimis my learning...
from anywhere” [A6]

“Switch off from the distractions and just focus on what matters” [A6]

4.6 Discussion

Using the findings, the immersive experiential learning theory extended to the legal education context is proposed and discussed below.

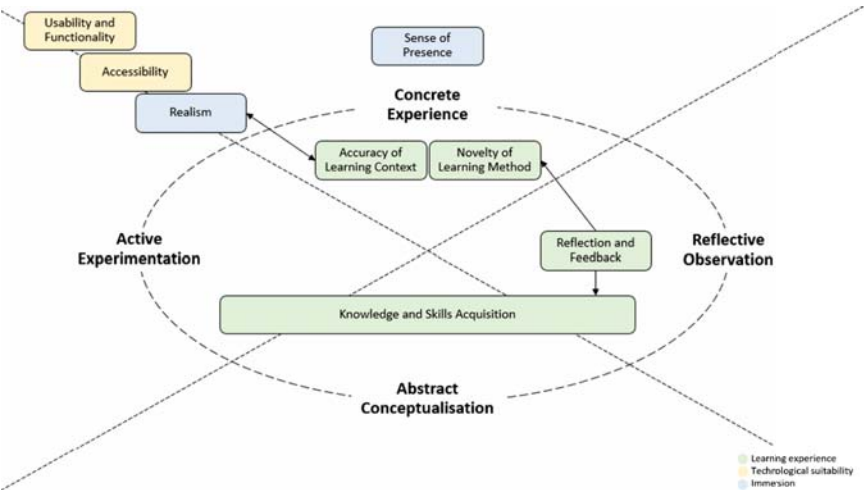


Figure 4.16: The Immersive Experiential Learning Theory in Legal Education

Findings show VR significantly enhances knowledge and skills acquisition in legal education (RQ1). Students found VR useful for applying substantive knowledge and boosting confidence, aligning with prior research [185, 226, 322]. However, preparation is crucial [56, 356], as simulations are less effective without prior substantive knowledge [168]. Notably, VR education literature overlooks this necessity, highlighting its particular relevance in legal education.

Qualitative findings confirmed VR’s benefits, while quantitative results showed differences between UK and Belgian students. UK students,

focused on competition skills, valued VR for future moots, whereas Belgian students, seeing it as an assessment tool, prioritised performance evaluation. These differences reflect cultural influences [156] and prior experiences with teaching methods [94]. Knowledge and skills acquisition shape abstract conceptualization (where learning takes place), active experimentation (application of knowledge in future scenarios), and reflective observation (self-evaluation), following the stages of Kolb's experiential learning cycle.

The accuracy of the learning context is a key factor. VR created a realistic environment, allowing students to assume the lawyer's role. Interactivity enhanced immersion [261, 377, 194], yet limited interaction with judges hindered the experience, supporting findings that restricted VR functions reduce learning effectiveness [193].

Reflection is crucial in legal education [377]. This study confirmed that VR feedback helps students identify mistakes and become more aware of body language. Its 3D visualization aids understanding [185], fostering critical reflection. UK students valued VR's enhanced feedback [248], while Belgian students found post-experience feedback less anxious. These factors influence reflective observation, abstract conceptualization (learning from feedback), and concrete experience (reduced anxiety).

VR's novelty boosts engagement through active participation, supporting prior findings [290, 165, 94]. It also provides a safe environment where mistakes carry no severe consequences, unlike traditional mock trials [221, 322], fostering confidence in the concrete experience itself. Overall, VR positively impacts legal simulations by creating a safe, contextualised practice environment that boosts confidence and knowledge application. It enhances learning through reflection, though its effectiveness varies with students' familiarity. However, VR's limited interactivity and reliance on prior knowledge (i.e., the learning context and the scenario) may hinder its ability to fully replicate real mock trials and serve as a stand-alone teaching tool.

Regarding the suitability of VR for facilitating simulated learning (RQ2), participants struggled with some embedded functions and experienced sickness, confirming previous findings [246]. Some suggested further guidance during the experience [210]. However, others had no issues and felt the notes function increased realism. Quantitative data showed a positive change in perceived usefulness of VR pre and post-experience, indicating that difficulties and sickness did not significantly hinder VR's usefulness. Reference [321] also found cybersickness had no significant effect on learning outcomes and technology acceptance. VR increases accessibility as there is no need for physical space, and students can practice in their own time, encouraging participation in mock trials, which are often difficult to schedule [265, 100, 56]. For students without courtroom experience, VR provides valuable exposure, impacting both the concrete experience (unique exposure and fewer space/time restrictions) and active experimentation

(willingness to participate in future VR trials).

To conclude RQ2, VR technology can be challenging for some students due to function use and sickness. However, these issues do not necessarily hinder perceptions of VR's usefulness. VR creates a learning space accessible from anywhere at any time, increasing opportunities for students to access a contextualised learning experience. This is beneficial given the difficulty of organising and participating in resource-demanding mock trials. Immersion significantly impacts learning (RQ3), with realism playing a key role. Some participants found the avatars and environment unrealistic, negatively affecting learning performance and skill transferability, aligning with research on visual fidelity [304, 4, 338]. Others felt the avatars and environment accurately represented the courtroom and mock trial, supporting [333] claim that avatars enhance presence, increasing immersion and situated learning, as confirmed by [98]. This sub-theme influences concrete experience (learning performance, contextualization) and active experimentation (future application). However, realism is subjective; Belgian students were more critical of the courtroom's design than UK students due to its Belgian style. VR also created a strong sense of presence by immersing students in realistic surroundings, supporting [309] and [282]. This sense of presence helped contextualis learning without concerns of cognitive overload [129, 241].

To address RQ3, immersion strongly influences learning. While realism is subjective, higher realism enhances learning and skill transferability. VR fosters sense of presence, allowing students to focus on content without cognitive strain.

4.7 Conclusions

Technology has already solidified its place in many areas of education [159]. In doing so, it has significantly influenced pedagogies and teaching methods. However, technology should not be used as an end goal in learning but rather must be designed to accommodate and enhance the existing learning objectives of the curriculum [270]. Therefore, it is vital that pedagogies are adapted to ensure that technologies are implemented properly and effectively.

4.7.1 Theoretical Contributions

This study has provided two theoretical contributions. Firstly, although learning theories can provide a solid foundation for the design of excellent pedagogical tools [358, 302, 149] identified that existing learning theories may not be suitable for the immersive learning context. By exploring the literature on how VR has been used in education, a conceptual framework to extend Kolb's [202] experiential learning theory to the im-

mersive context was proposed, similar to the works of [86]. Specifically, the literature review presented in this paper has shown that the technological suitability of VR and the qualities of immersion that it creates are deeply connected to and influence the learning experience, which confirms suggestions made by [270] and [5]. This framework can be further employed to explore the uses of VR in other areas of education. Secondly, through empirical investigation, the immersive experiential learning theory has been extended to the legal education context. As a result, the study has proposed a theoretical framework to guide the implementation of VR in legal education. While developed for legal education, this framework could also be applied more broadly to other educational contexts involving experiential and skills-based learning. The framework reveals 8 influential sub-themes under the 3 overarching themes of learning, technological suitability, and immersion. In this way, the framework holistically brings together the elements of legal simulation and VR technology, extending the studies conducted by [253] and [98].

4.7.2 Practical Contributions

In terms of practical contributions, this study has designed a tailored, high-quality VR learning environment for mock trials, following recommendations from the literature [270, 302]. By incorporating context-relevant tools such as recording and note-taking features, it provides practical insights into how VR environments can enhance conventional mock trial methods, offering a realistic and effective courtroom experience. To the authors' knowledge, this is the first VR mock trial application with these functionalities.

Furthermore, the study brings important implications and guidelines for educators when designing VR experiences in legal education. As discussed, VR learning experiences should be designed carefully based on solid learning theory to effectively facilitate a high quality of learning [270, 358, 302]. The framework provides insight into the various considerations that educators must consider developing and implement immersive experiences.

Lastly, the study has shown that educators must ensure that the characteristics of VR technology must be carefully and smoothly integrated with the learning content. This will ensure that the learning experience is enhanced, and that the students are not distracted by the technology [270]. Indeed, cognitive overload was not identified as a concern in the findings of this study.

4.7.3 Limitations and Future Research Directions

It is important to acknowledge key limitations. First, although the study employed a comparative case study method to explore themes, its

findings are limited by the sample size and contextual differences between institutions. Future research could include larger, longitudinal studies with control groups to quantitatively test the proposed framework and compare VR with non-VR learning approaches.

A significant limitation was the use of the same virtual courtroom for UK and Belgian students, despite their distinct legal systems. The courtroom was designed based on the Belgian context, which may not have aligned with some UK students' expectations, potentially affecting their immersion and engagement. While the mock trial focused on transferable skills like advocacy and public speaking, the mismatch in environmental fidelity could have impacted students' ability to fully contextualise their learning. This highlights the need to tailor VR environments to specific legal systems to ensure equitable benefits for all students.

Another limitation was the restricted interactivity within the VR environment due to the structured nature of the task. The absence of real-time interactions, such as interruptions or questioning by a virtual judge, reduced the authenticity of the experience. These dynamic elements are crucial for developing adaptability and quick thinking in legal practice. Future iterations should incorporate such features to enhance realism and interactivity. Additionally, the VR environment limited participants' ability to use notes effectively. In traditional mock trials, students rely on notes to structure arguments and reference key points, but the VR setup restricted access, making some feel less prepared. Difficulties in navigating virtual note-taking tools further added to the challenge. Since legal practitioners rely on documents in real-world courtrooms, improving note functionality or integrating physical notes into VR could enhance the learning experience and realism.

4.8 Appendix

Group	Nb. of participants	Nb. of men	Nb. of women	Mean Age (SD)
Overall	60	26	34	25.95 (11.42)
UK (EN)	30	12	18	26.13 (15.66)
Belgium (FR)	30	14	16	25.77 (4.47)

Table 4.2: Participant demographics.

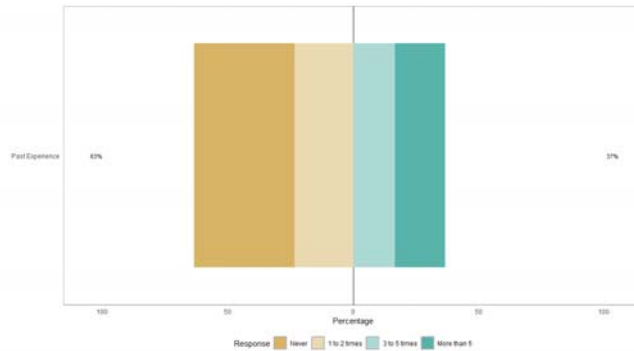
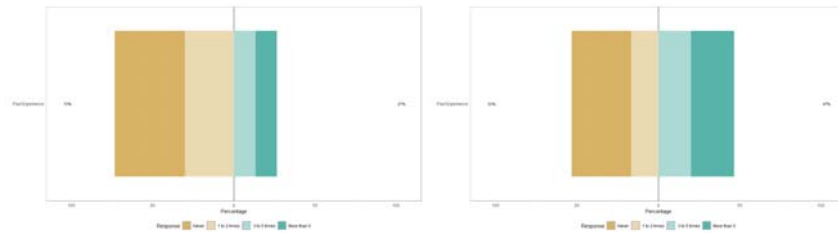


Figure 4.17: Past experience in VR of participants rated as a 7-point Likert scale from 1 (=no experience) to 7(=expert).



a. Past experience in VR of UK participants

b. Past experience in VR of Belgian participants

Figure 4.18: Comparison of past experience of participants.

Interview Question	
1	How was your learning experience with the Virtual Reality moot trial?
2	What was good and what could be improved? Why do you think so?
3	Could you tell me a bit more about the trial itself?
4	What characteristics of the virtual trial stood out the most, and how did they affect your learning?
5	Tell me more about [X] (e.g., realism, engagement, sense of presence, use of skills, interactivity, active participation)
6	How did the virtual trial make you feel?
7	As part of the experience, you were given time to look over your recording.
8	Do you think that having this time contributed to your learning? How?
9	Tell me more about [X] (e.g., curiosity, structure, critical reflection)
10	How effective was the experience as a learning tool? Or: How did this experience help your learning?
11	Tell me more about [X] (e.g., applying previous knowledge, acquiring new knowledge, personal achievement, real-life applicability)
12	How did you feel overall during the learning experience?
13	Do you feel that this overall experience has better equipped you for future moot scenarios? Why?
14	Tell me more about [X] (e.g., learning from mistakes, application of knowledge, motivation, learning new skills)
15	Would you be willing to use the VR moot trial again for your studies? Why?
16	Have you participated in a non-VR moot exercise before?
17	If so, what are your general thoughts on the pros and cons of mooted in terms of your learning?
18	What role has VR played in your learning experience? What differences has it made, if any?
19	In your opinion, how suitable is VR technology for use in mooted?
20	What characteristics of VR in particular make it suitable or not suitable?
21	How did the virtual reality technology make you feel overall?
22	What did you think about the design of the virtual environment?
23	Are there any further comments you would like to make about the overall experience or about any particular points?

Table 4.3: Interview guide used to explore students' learning experience with the VR moot trial.

CHAPTER IV. IMPLEMENTATION OF VR IN LEGAL EDUCATION

Participant ID	University	Level of study	Age	Previous VR experience	Previous mootng experience
A1	UK	3rd year undergraduate	20	No	Yes
A2	UK	3rd year undergraduate	20	Yes (1 time)	Yes
A3	UK	1st year postgraduate	26	Yes (A few times)	Yes
A4	UK	3rd year undergraduate	21	No	No
A5	UK	1st year postgraduate	23	Yes (Many times)	Yes
A6	UK	3rd year undergraduate	21	No	Yes
A7	UK	3rd year undergraduate	20	No	Yes
A8	UK	1st year postgraduate	24	No	Yes
A9	UK	1st year postgraduate	61	No	Yes
B1	Belgium	1st year postgraduate	22	No	No
B2	Belgium	1st year postgraduate	25	No	Yes
B3	Belgium	2nd year postgraduate	25	No	Yes
B4	Belgium	1st year postgraduate	25	Yes (Many times)	Yes
B5	Belgium	1st year postgraduate	22	No	Yes
B6	Belgium	1st year postgraduate	22	Yes (A few times)	No
B7	Belgium	1st year postgraduate	22	Yes (Many times)	No
B8	Belgium	1st year postgraduate	25	No	No
B9	Belgium	1st year postgraduate	25	No	Yes
B10	Belgium	1st year postgraduate	24	No	No
B11	Belgium	1st year postgraduate	26	Yes (1 time)	No
B12	Belgium	1st year postgraduate	23	Yes (1 time)	No
B13	Belgium	1st year postgraduate	32	No	No
B14	Belgium	1st year postgraduate	30	Yes (A few times)	No

Table 4.4: Participant demographics for the qualitative interviews.

Chapter V

Automatic Assessment of Multimodal Cues for Public Speaking Training in Virtual Reality

This chapter presents part of what I worked on during my research stay at Aix-Marseille University. This work was done between July 2024 and November 2024. The co-authors are Marion Ristorcelli, Jean-Marie Pergandi, Rémy Casanova, Michaël Schyns, and Magalie Ochs. It is now a paper, for which I am the first author, that has been submitted to the *17th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS2025)* in 2025. In this thesis, the paper is presented as originally submitted, and I apologise for any outdated information, repetitions, or formatting inconsistencies with other chapters.

Abstract

The quality of public speaking depends heavily on the multimodal behaviour of the speaker. Recent advances in virtual reality and automated measurement tools have enabled the computation of various multimodal cues (verbal and non-verbal) that contribute to assess public speaking performance. However, the challenge remains to determine which among the myriad of available cues are the most relevant for training purposes. This paper addresses this question by (1) defining criteria for characterising the relevance of behavioural cues in public speaking training, (2) identifying a set of key multimodal cues to consider for this training context, and (3) outlining methods for their computation, from raw data extraction to index calculation. These multimodal cues are integrated into a VR-based public

speaking tool, offering actionable feedback to trainees and supporting skill enhancement.

5.1 Introduction

Public speaking is a skill that, while invaluable, does not come naturally to everyone. Effective public speaking requires training and practice to refine one's ability to engage, inform, and persuade an audience. Even if the appropriate behaviour of a public speaking may depend on the context, several research have shown that a set of generic multimodal cues characterising a good performance of public speaking can be defined, whatever the context [384, 284]. For instance, high loudness [99], hands with palms upward [71], and eye contact toward the audience [147] play critical roles, contributing to the speaker's overall effectiveness.

This paper aims to identify multimodal cues relevant for assessing public speaking performance, with the goal of providing feedback enabling users to train and improve their skills. A cue is defined as *relevant* if it meets four key criteria: (1) the cue has an impact on the perceived quality of public speaking as shown by empirical and theoretical research works, (2) the cue can be measured automatically, (3) the cue does not depend on the context of the public speaking (i.e. use of slides, standing or sitting position, etc.), and (4) the cue is explainable and interpretable by the users, allowing for targeted self-improvement. These criteria are based on a synthesis of current gaps in existing Virtual Reality (VR) public speaking training (PST) tools presented in the next paragraph.

Despite advancements in VR training tools, no comprehensive list of relevant multimodal cues for PST in VR or clear guidance on calculating them automatically exists, and current tools have yet to address this gap. Indeed, commercial tools in VR like Chiara (<https://bechiara.com/en/>), Ovation (ovationvr.com), Virtual Orator (virtualorator.com), and VirtualSpeech (virtualspeech.com) use VR and sometimes AI to offer feedback on public speaking elements such as volume, pacing, body language, eye contact, and speech clarity. However, while these tools provide tips on these key performance aspects, the methods for calculating metrics remain largely undisclosed, reducing transparency and potentially limiting users' ability to fully understand the reasons behind the proposed feedback. Available research training tools also suffer of limitations. In the research, many VR tools for PST exist [20, 64, 281, 142, 111, 324] but they also face significant limitations. While some tools focus primarily on eye gaze and time spent on specific components [142], others include feedback on clarity of speech but lack comprehensive multimodal feedback [20, 64]. Efforts have been made to address multimodal cues [280, 324]; however, the feedback relies on raw metrics (e.g. number of words per minute, volume in dB), limiting both the

interpretability and practical utility for users.

The main objective of the paper is to provide a framework to integrate automatic multimodal cues assessments for PST in Virtual Reality. Thus, its first contribution is the presentation of empirically and theoretically grounded multimodal cues identified as critical for effective public speaking. Then, its second contribution is a practical methodology to automatically compute these cues to provide interpretable feedback, helping users enhance specific aspects of their performance.

The paper is organised as follows. The next two sections cover the theoretical background of multimodal cues in public speaking, focusing on verbal (Section 5.2) and non-verbal (Section 5.3) features. Each feature is developed based on its theoretical foundation, interpretability, calculation method, and, if available, expected values for effective versus ineffective speaking. Then, Section 5.4 introduces our integrated tool, which combines VR PST with these feedback mechanisms. The last section (Section 5.5) discusses future work and limitations.

5.2 Verbal Features

5.2.1 Acoustic Features

Acoustic features are crucial for assessing vocal delivery in public speaking, as they provide insights into elements that greatly influence audience engagement, such as pitch, loudness, and rhythm [44, 143, 387, 99].

Despite the wide range of acoustic features available, this section is dedicated to interpretable and straightforward features, intentionally excluding more complex metrics (e.g., formants or harmonic differences) and low-interpretability metrics often used in sustained vowel exercises in contexts such as speech therapy (e.g., jitter, shimmer).

The methods of calculating these features vary among researchers, with some relying on formulas and others using different software, often without clarity on the exact settings used for the extraction, which complicates direct comparisons [113]. To ensure metric consistency, the GeMAPS set [113] from the OpenSMILE toolkit is used wherever possible, as it was specifically designed for this purpose. Otherwise, the widely accepted for acoustic analysis software Praat (<https://www.fon.hum.uva.nl/praat/>) is used.

Pitch

It is “the degree of highness or lowness of a tone, which is determined by the vibration of the vocal folds (i.e., the faster the vibration per second (Hz), the higher the pitch).” [122].

Average pitch and coefficient of variation: The *mean* and the *coefficient of variation* (CV) of the pitch are extracted using GeMAPS.

Indeed different pitch significantly enhances the expressiveness, charisma, and emotional tone conveyed by a speaker [44, 387] and is crucial in perceived speaker engagement and effectiveness [164]. Pitch during speech varies among individuals of the same sex, with men's average ranging from about 80 to 180 Hz and women's from 125 to 300 Hz [295].

Furthermore, effective public speakers likely display a moderate pitch CV, balancing consistency with expressive variation to engage the audience. High CV can signal joy [44] and confidence [326], while low CV may convey sadness [44]; however, extreme values in either direction could suggest either nervousness or monotony [143].

Loudness

It serves as an indicator of vocal projection and intensity, essential for ensuring audibility and emphasising key points in public speaking.

Average loudness and coefficient of variation: The *mean* and the *CV* of loudness are extracted using GeMAPS.

Effective speakers adjust their loudness to highlight important content, enhancing audience engagement. Research shows that managing loudness is critical in communication contexts, where modulation can influence perceptions of confidence and control [143]. For example, weaker loudness is associated with shy people and higher loudness to confident people [99].

Rate of loudness peaks: The frequency of changes in loudness per second, indicating emphasis and expressiveness. It is extracted using GeMAPS. Controlled loudness peaks help draw attention to key points, while a lack of variation can make speech monotonous [143].

Silent Pauses

They are marked by a complete absence of sound and serve to reduce cognitive load on listeners, giving them time to mentally organise and absorb the content, adding rhythm and clarity to speech [234]. To capture these nuances, additional parameters are included in our analysis.

Silent pause count, duration and speech rate: The total number of silent pauses within a speech segment. This value will be calculated using Praat software and the TextGrid method. Then the length of silent pauses, and the frequency of silent pauses by speaking time is extracted. Campione and Véronis (2002) categorised silent pauses into brief (less than 200 ms), medium (200-1000 ms), and long (over 1000 ms). They suggest that while brief and medium pauses aid in structuring speech and enhancing clarity, long pauses (over 1000 ms) should generally be avoided as they may disrupt the flow and reduce engagement.

5.2.2 Textual Features

Textual features play a crucial role in assessing speech quality by capturing the nuances of language use, structure, and content.

This analysis uses the Linguistic Inquiry and Word Count (LIWC) [40], a dictionary-based text analysis tool that quantifies psychological, emotional, and cognitive elements in written or spoken language by matching words against predefined categories representing psychological constructs.

LIWC Linguistic Metrics

These metrics are used to retain the following parameters due to their relevance to PST.

Analytical thinking: The logical and structured thinking in language [288]. A high score is well perceived in academic settings and is correlated with things like grades and reasoning skills. However, a low score is perceived as more familiar [40].

Clout: The speaker's confidence or authority level. Higher scores suggest that the speaker projects confidence and social dominance, whereas lower scores may indicate a more humble or tentative tone [184].

Authenticity: The honesty and openness [271]. Higher scores are associated with spontaneous conversations between friendly people, while lower scores might suggest a more prepared (in advance) speech [40].

Emotional tone: The overall positivity or negativity of language [73]. Numbers below 50 indicate a negative emotional tone, while those above 50 suggest a positive tone [40].

LIWC Percentage Words

The percentages encompass the following parameters retained for their relevance to PST.

Anxiety: Words like "worry, fear, nervous" that may risk distancing the audience if overused or translates discomfort.

Swear words: Profanity or strong language (e.g., "damn," "hell"). These should be avoided as they can detract from professionalism.

Netspeak: Internet slang or abbreviations like "lol," "fyi". These terms are generally discouraged in professional public speaking to maintain formality.

Assent: Words that indicate agreement, such as "yes," "okay," or "alright." These can help foster a sense of rapport with the audience.

Non fluencies and fillers: Non fluencies are words like "uh," "um," or "ah" that indicate hesitation or thought pauses. Fillers are words that add no substantive meaning, like "you know", "like", or "I mean". Fillers can influence listeners' perceptions of confidence [96], as frequent fillers

often signal hesitation and affect the listener's impression of the speaker's confidence Cholle et al. Furthermore, minimising fillers improves perceived speech fluency, leading to smoother and more engaging public speaking performances [64].

Flow of the Speech

It can be disrupted by the excessive use of fillers. As noted in the previous section, minimising the use of non fluencies and fillers, also known as *disfluencies*, is essential for effective speech [64]. To efficiently compute disfluencies, in addition to the LIWC dictionary presented earlier, one should examine the most frequently used words within the specific speech. Words that appear excessively may act as unintended fillers.

Disfluencies and speech flow metrics: The total count of disfluencies, the number of times disfluencies occur at the beginning of sentences, or within sentences, and the overall speech rate, measured as the frequency of words over time [96]. Excessive disfluencies can negatively impact speech rate, which ideally should not be too slow or too fast to maintain audience engagement.

Diversity Metric

It includes vocabulary richness features and captures the depth and sophistication of language use, making them essential for assessing verbal mastery and complexity.

Type-token ratio (TTR): The total number of unique words (types) by the total number of words (tokens) in a text segment [363]. It provides insight into how varied the vocabulary is within a text. A higher TTR indicates more diverse vocabulary while a lower TTR suggests repetition or limited vocabulary. TTR obviously varies with text length. A threshold near 0.7 effectively distinguishes lexical diversity, as seen in the *MTLD* (measure of textual lexical diversity), which measures the average length of continuous text sequences that maintain a TTR above 0.72 [250, 15]. This metric can help assess the richness of vocabulary in spoken presentations, where a diverse lexicon enhances engagement and clarity.

5.3 Non-Verbal Features

Non-verbal behavioural cues, such as gaze, postures, gestures, facial expressions, are widely recognised in the literature as indicators of a speaker's confidence, anxiety, and engagement, all of which impact presentation performance [147, 329, 330]. Given the lack of standardised VR measures for these cues, this paper adapts existing features and introduces new metrics tailored for automated feedback in VR, emphasising relevant literature to support these innovations.

All the non-verbal behavioural measurements described below can be computed using the Meta Quest Pro headset and a Unity game engine, by using the advanced tracking capabilities of gaze and facial expressions (e.g., the OVR Camera Rig offered for 3D visualization and tracking management).

The Gaze Behaviour

The gaze plays a crucial role in presentations. For example, a positive performance assessment requires maintaining at least 80% audience eye contact [69] and a successful presentation involves the speaker keeping their head and eyes directed toward the audience [306, 147] rather than toward slides [70] or notes [360]. Experts also agree that eye contact should alternate rather than fixate on a single person [330]. Effective speakers sustain continuous eye contact by sweeping their gaze across the audience, adjusting their focus according to audience size [280]. To ensure audience-directed gaze, most studies approximate eye gaze direction through face or head orientation in both VR and live contexts [59, 70, 306, 280, 394]. Based on this literature, the paper proposes the following features linked to the gaze behaviour.

Score of characters viewed: A weighted average of the numbers of audience members viewed across set time segments, with each unique audience member observed contributing to the total. A score closer to the total number of audience members N indicates more balanced gaze distribution over time, reflecting consistent engagement with all audience members and suggesting effective gaze behaviour.

Gaze fixation time: The duration of uninterrupted gaze toward each virtual audience character. The *median* and *interquartile range* of this duration time are used to assess if the gaze duration is excessively long or too brief, with uninterrupted gazes beyond a threshold, to be determined, considered inappropriate.

Gaze duration metrics: The total time spent looking at characters is first assessed across the *whole audience*, then further examined by *specific position within the audience* and then, *by gender*. The overall proportion of time spent focusing on the audience versus other elements serves as a measure of attentiveness, with effective presentations maintaining at least 80% audience focus [69]. For smaller audiences, an even distribution of gaze across individual positions demonstrates spatial attentiveness, suggesting strong engagement. Separately summing gaze duration toward male and female characters, normalised by the presentation's total time, indicates whether any gender bias influences gaze patterns.

Hesitation time: The speaker's hesitation time, indicated by looking at the ceiling or floor when searching for words or ideas. It represents the proportion of time spent in reflection during a presentation, calculated as time

spent gazing downward or upward, normalised by the total presentation duration. If this feature reflects hesitation [23], a higher value may correlate with a more negative assessment of presentation quality.

Entropy of gaze direction: This feature assesses the dynamics of gaze direction towards the virtual audience, measuring the uncertainty or diversity in gaze distribution across targets. Entropy helps determine how evenly the speaker's attention is spread across these targets, with each target's gaze time calculated as a proportion of total gaze time. Various entropy measures can be used to assess distribution balance, allowing flexibility in selecting a formula that best fits specific analysis needs. A higher entropy score suggests a well-distributed gaze across all targets, while a lower score indicates a focus on fewer targets.

Body Movements and Postures

They are powerful indicators of a speaker's attitude and state of mind [394]. Fidgeting or crossing legs may suggest low self-confidence and a negative attitude [3, 330, 147]. Experts identify ineffective stage practices, such as standing stationary behind a screen and moving purposelessly across the stage, while recommending purposeful movements that align with the presentation's structure and timing [330].

Studies consistently indicate that gestures like crossing arms, hiding hands, touching the face or hair, or fidgeting should be avoided during presentations [330, 394]. Arm-crossing and holding the opposite wrist or arm, known as arm barrier gestures, signal a defensive or nervous attitude, implying uncertainty [3]. An open posture, with arms open, hands near the body, palms facing the audience, or hands held above the belt without interlacing, is recommended. Effective speakers use smooth upper-body gestures to emphasize points, while less effective speakers keep their arms down or hands at the abdomen [147].

Drawing on this literature, this paper proposes the following features related to body movements, derived from tracking data captured within the virtual environment.

Amplitude of the horizontal displacement: The speaker's use of the stage. By calculating horizontal movement, amplitude is determined as the absolute difference between the maximum and minimum x-axis values during the presentation. Depending on the context, one can expect to see some horizontal movement, but not too much, because, according to the experts, movement on stage must be linked to a particular section of the presentation and have a purpose [330].

Distance covered by the speaker: The stage use and potential fidgetiness. The Euclidean distance covered (chosen for reasons of interpretability of this metric) is calculated and normalised by the presentation duration, yielding a distance per unit of time. Combining it with the previous feature gives

interesting information about the speaker, for instance, it is expected that if a speaker covers a large distance with low amplitude, it may indicate agitation.

Entropy of body movements: This feature captures the dynamics of the trainee's body movements, indicating movement or immobility during the presentation. To align with previous entropy calculations, the space is divided into N zones, and for each zone, the proportion of being in that zone is computed. The entropy formula is then applied. High entropy reflects frequent movement or varied positions, while low entropy suggests a dominant, static posture.

Gestures

They provide valuable insight into a speaker's state of mind and message. Palm-up gestures convey confidence and have a positive impact on listeners [71], while open palms signify qualities like truth, honesty, and openness [3]. Conversely, palm-down gestures, such as pointing, may communicate authority or even intimidation [254]. Mixed palm gestures can indicate knowledgeable power, which may either build trust or come across as intimidating. Gestures such as clasping hands, rubbing hands, or self-touching (e.g., touching the body or head) are typically associated with anxiety and are best avoided [3, 21, 71].

Building on the insights from this literature, this paper suggests the following features associated with gestures. The analysis employs advanced hand tracking technology to capture each hand's 3D position and rotation (without controllers). Hand orientation is categorised based on rotational values along the vertical axis, with positions classified as *Palm Down* for rotations above some chosen threshold α , *Palm Up* for rotations below α , and *Neutral* for rotations between $-\alpha$ and α . Hand openness is similarly classified: *open* when all fingers are extended, *closed* when all fingers are flexed, and *neutral* when finger positions vary. This approach combines continuous and categorical features to provide a comprehensive analysis of hand gestures.

Frequency of hand positioning and openness: The frequency of each hand's positioning (neutral, palm-up, palm-down) and openness (neutral, open, closed) states is calculated as the count of each state divided by the total presentation duration. Effective speakers are generally expected to display a higher frequency of open states and palm-up positions, indicating confidence and openness, and a lower frequency of closed states and palm-down positions.

Proportion of hand detection: Hands are detected by the headset's integrated camera when they are within its field of view, about 10 cm from the torso and between the headset and abdomen. The proportion of right and left hand detections relative to total measurements during the presentation

is calculated, with skilled speakers expected to show a high proportion of detections in this space [54].

Duration of palm positions: The duration of uninterrupted palm specific posture. The *median* and *interquartile* range assess if the speaker maintains an open posture for extended or brief periods. For example, beyond a certain threshold, to be determined, an uninterrupted palm-up posture is considered ideal for oral presentations.

Duration of a closed and open posture: The proportion of open posture relative to total hand measurements. Indeed, posture openness is also indicated by the distance between hands, and distances below a threshold, to be determined, suggest clasped or overlapped hands, while negative values indicate crossed arms, reflecting a closed posture. Effective speakers are expected to maintain a higher proportion of open posture.

Entropy of hand openness and direction: The variability in hand openness and direction states (neutral, open, closed; palm-up, palm-down, neutral) is analysed using entropy to assess the balance in hand positioning dynamics. Proportion of being in each state are thus calculate. High entropy indicates a balanced variation in hand states, while low entropy suggests a preference for a particular hand position or openness.

Facial Expressions

They are crucial for assessing presentation quality [386]. Experts advise against maintaining a blank expression; instead, speakers should appear lively, using varied expressions and smiles, even in professional settings. Expressions should align with the presentation's content, and positive emotions should be shown over 20% of the time, with negative expressions kept below 20% [69].

Blend shapes, available through the Meta data, are numbers representing specific muscle movements. They vary between 0 (no activation) and 1 (maximum activation) and align with the *Facial Action Coding System (FACS)* and provide a framework for describing facial expressions through muscle activations (Action Units), either positive or negative. In this paper, the left and right sides of the face expressions are extracted and are averaged for each facial expression.

Frequency of positive or negative facial expressions: The frequency of facial expressions either positive or negative. Only blend shapes that deviate by more than one standard deviation from the subject's mean blend shape value are considered. The frequency is calculated by counting the number of blend shapes associated to positive or negative AU and dividing it by the presentation time. To reflect engagement, a speaker should use varied facial expressions. Ideally, positive expressions should appear more frequently, while negative expressions should be rare, indicating higher frequencies for positive blend shapes and lower frequencies for negative ones.

*Entropy of blend-shapes: The dynamics of the speaker's facial expressions by measuring expression variability is analysed using *entropy*. The interval of values between $[0,1]$ are divided into N categories and for each category, the proportion of the blend shapes belonging to this category is calculated. Then the formula of entropy is then applied. A high entropy value is expected for effective speakers, indicating diverse facial expressions throughout the presentation.*

5.4 Virtual Reality Solution

The goal of this section is to demonstrate the feasibility of automatically computing all the multimodal cues identified in this paper within the context of a VR public speaking training environment.

The VR environment used in this project (see Fig. 5.1) supports various configurations, including audience size, demographic composition, and behavioural styles, to tailor training experiences to individual needs. Additionally, the system employs full-body tracking to make presentations feel natural and lifelike, accurately capturing users' gestures and postures without the need for restrictive devices.



Figure 5.1: The virtual environment used for public speaking training.

To complement this VR environment, based on the automatic assessment of multimodal cues identified in this paper, a companion website (see Fig. 5.2) serves as the central hub for feedback and analytics. After each training session, users can access a detailed performance report, which integrates verbal and non-verbal cues. The website further provides visualizations and replay features of the VR scene (see Fig. 5.3) enabling users to review their presentations and track their improvement over time.



Figure 5.2: Combined view of the website for feedback and analytics.



Figure 5.3: Replay screen

The integration of this VR environment with its companion website creates a user-friendly solution for public speaking training, addressing both skill development and self-assessment. This dual-platform approach

ensures that users receive actionable insights and maintain engagement throughout their training journey.

5.5 Discussion

This paper makes several key contributions to advancing VR-based PST by defining and detailing a comprehensive set of multimodal features that provide insights into effective presentation skills. In the verbal domain, metrics related to acoustic (pitch, loudness, pauses) and textual (linguistic metrics, word percentages, speech flow, and vocabulary diversity) features are carefully selected and quantified using established methodologies. These metrics are designed to be applicable beyond VR contexts. On the non-verbal side, methods were developed to calculate critical metrics such as gaze distribution, body movement patterns, hand gestures, and facial expressions. These metrics are derived from VR tracking data, enabling an in-depth analysis of speaker non-verbal behaviour.

The presented tool incorporates open-access software and toolkits. The only exception is the LIWC, a widely used software for textual analysis but that requires a license. However, as its selected features are based primarily on linguistic metrics and word percentages, users interested in a totally open-access alternative could generate similar metrics by creating custom dictionaries.

Despite these contributions, some limitations remain. The tool's context-independent design means it does not account for factors that may impact performance. For instance, while the identified LIWC categories capture verbal features, they provide no insight into specific topics addressed in the speech. For non-verbal behaviour, the tool lacks indicators for contextual elements, such as using a computer, slides, or podium, which may influence movement or gaze direction (Schneider et al., 2017).

Through this paper, specific choices for computing certain metrics (e.g. palm orientation thresholds, spatial zone splits, or entropy functions), along with ideal values for effective public speaking metrics are not yet fully established. Although a set of multimodal cues has been provided, optimal ranges for these indicators have not been determined. Future work could leverage machine learning techniques on a corpus of PST sessions to establish these thresholds, enhancing the tool's precision and making the feedback even more actionable.

While this paper evaluates a set of multimodal indicators, each modality is presented individually, allowing users to focus on specific cues depending on their interest. However, public speaking is inherently multimodal, with dynamic interactions across modalities in real-life scenarios. For example, emphasising words using higher pitch and large gestures, coordinating movement on stage with the presentation structure, or synchronising pauses

with gestures are key combined features not considered here. Future iterations could integrate these multimodal interactions to provide a more holistic view, yielding richer insights and more contextually relevant feedback.

Another limitation not yet discussed is the challenge of data visualization for users. Presenting feedback in an interpretable and user-friendly format is essential for effective training. Decisions must be made on whether metrics should be presented as numerical values, graphs, or interactive visual elements. For example, metrics such as gaze distribution could be represented with heat maps to show focus areas, while more abstract metrics like entropy could benefit from graphical representations to illustrate variability and balance in gaze or movement patterns. Developing an intuitive visualization approach that allows users to understand and act on these insights is crucial for enhancing the tool's usability and impact on PST.

In conclusion, the VR-based tool presented in this paper delivers on critical verbal and non-verbal features to support public speaking skills improvement. While the tool is effective in its current form, addressing the limitations highlighted would make it an essential resource in PST.

Chapter VI

EVE: Emotional Validated Expressions, an acted audiovisual corpus

This chapter is a paper, for which I am the first author, that has been submitted to the *International Conference INTERSPEECH25* in 2025. It is based on a collaboration with the Faculty of Psychology, Speech Therapy, and Education of the University of Liège and HEC Liège. The co-authors are Anne-Lise Leclercq, Angélique Remacle, and Michaël Schyns. In this thesis, the paper is presented as originally submitted, and I apologise for any outdated information, repetitions, or formatting inconsistencies with other chapters.

Abstract

This paper presents the creation and perceptual validation of the EVE corpus, a resource for speech emotion recognition in English and French audio and audiovisual content. For each language, ten native or near-native actors per language performed 10 linguistically and semantically neutral sentences with different emotions: six basic (fear, anger, happiness, sadness, disgust, surprise), and four complex (confidence, confusion, contempt, empathy). Each was expressed at two arousal levels, with two trials per level. Additionally, each sentence was also produced in a neutral condition, leading to a total of 4,100 recordings. The emotional content of the corpus was perceptually validated by 600 participants per language. The audiovisual recordings were made in high-definition quality following a rigorous methodology, positioning the EVE corpus as a valuable tool for academic and development purposes. The corpus is available under an open license to facilitate research.

6.1 Introduction

6.1.1 Emotions and Implications for Affective Computing

As Ekman said, emotions are “a particular kind of automatic appraisal influenced by our evolutionary and personal past, in which it is sensed that something important to one’s welfare is occurring, and a set of psychological changes and emotional behaviours begins to address the situation.”[286]. These emotions play a crucial role in daily human interactions, manifesting themselves through speech, written text, and facial expressions.

Two models are usually used for emotion analysis: categorical and dimensional. The first model categorizes basic emotions into discrete states [107], such as fear, anger, happiness, sadness, disgust, and surprise (see Table 6.1). In contrast, the second model, illustrated by the Valence-Arousal-Dominance framework [369], views emotions as continuous variables, characterized by dimensions of valence, arousal, and dominance. The present paper focuses on the categorical model. This choice reflects the model’s compatibility with actor-based data collection and practicality for annotation by external raters.

The six basic emotions fail to capture the full spectrum of traditional interactions, where more nuanced emotions like self-confidence, confusion, and empathy often play a crucial role in communication. Self-confidence [291] is an important emotion in professional contexts, influencing leadership and decision-making. Similarly, confusion [162] is pivotal in decision-making scenarios, where individuals often experience uncertainty when faced with ambiguous or complex information. Additionally, emotions like contempt [152] are critical in social dynamics, especially in power-related interactions, and empathy [133] is fundamental for building compassionate, supportive relationships. Expanding the range of emotions to include these complex states can enhance the effectiveness and relevance of emotional expression analysis in diverse social and interactive contexts. While other complex emotions could have been considered, only a limited set was included to maintain clarity and reliability, as too many similar emotions can hinder their distinction [79].

As human-computer interactions expand, accurately detecting emotions across multiple modalities (such as speech, text, and facial expressions) becomes crucial, yet recognizing emotions remains a significant challenge, particularly when relying solely on speech. Audiovisual speech databases are vital for emotion recognition, as they support feature extraction and classifier training. These databases rely on diverse speakers and multilingual data to ensure robust performance.

Speech corpora are generated through various methods, including acted or induced modes, each offering unique advantages. The first method requires speakers to express specific emotions according to set guidelines,

while the latter method captures spontaneous speech in real-life scenarios.

Table 6.1: Definitions of the 6 basic emotions by Ekman* and 4 complex** emotions

Emotion	Definition
Fear	A primal emotion that is important to survival and triggers a fight or flight response.
Anger	An emotional state leading to feelings of hostility and frustration.
Happiness	A pleasant emotional state that elicits feelings of joy, contentment, and satisfaction.
Sadness	An emotional state characterized by feelings of disappointment, grief, or hopelessness.
Disgust	A strong emotion that results in feeling repulsed.
Surprise	A brief emotional state, either positive or negative, following something unexpected.
(Self)-Confidence	A person's belief that he or she can succeed.
Confusion	Emotion where the individual is unsure about how to interpret certain stimuli.
Contempt	A response that devalues its objects to the point of nullifying them and their capabilities.
Empathy	The ability to perceive another person's point-of-view, experience the emotions of another, and behave compassionately.
* https://online.uwa.edu/infographics/basic-emotions/	
** [291, 162, 152, 133]	

6.1.2 Existing Emotional Speech Corpora

The creation of the EVE (Emotional Validated Expressions) corpus was motivated by the scarcity or limited accessibility of publicly available high-quality databases for Speech Emotion Recognition (SER) in both French and English, a gap that became apparent with the advancement of emotion-based software applications. Indeed, many existing speech audiovisual corpora are limited in scope or prohibitively expensive. Affective computing has evolved into a multidisciplinary field focused on recognising human emotions to improve work conditions, entertainment, and services through artificial intelligence [369]. However, the challenges with existing SER databases are manifold due to their limitations. For instance, the Hume AI database [79], despite its extensive data volume (400,000 recordings), may be inaccessible to many researchers due to its cost or usage conditions. Additionally, its restriction to only five different sentences results in an unbalanced distribution of expressions, limiting the range available for study. Similarly, databases commonly used in SER research in English (see Table 6.2), often do not simultaneously fulfil various crucial criteria such as emotional diversity, phonetic balance, and perceptual validation of intended emotions by a statistically significant number of participants. The challenge intensifies for French databases which face similar issues in terms of emotional diversity, affordability, recording quality (e.g., control of microphone types, background noise, or studio vs. in situ recordings),

phonetic balance, and validity through perceptual studies. These databases have various limitations, underscoring the need for more inclusive and freely available datasets for effective SER.

Table 6.2: Most popular open-access datasets.

Language	Name of the corpus	Modality	Generation method	Nb. of speakers	Nb. of emotions	Phonetically balanced	Nb. of arousal levels	Perceptual study
EN	CREMA-D [53]	AV	A	91	5+1	No	4	Yes
EN	EMOVOX [325]	A	I & A	16	2	No	1	No
EN	IEMOCAP [47]	AV	A	10	8+1	No	1	No
EN	RAVDESS [224]	AV	A	24	7+1	No	2	Yes
EN	SAVEE [178]	AV	A	4	6+1	Yes	1	Yes
EN	TESS [294]	AV	A	2	6+1	No	1	No
FR	CaFE [144]	A	A	12	6+1	No	2	No
FR	EMOVOX [325]	A	I & A	54	2	No	1	No
FR	GEMEP [14]	AV	A	10	15	Yes	3	No
FR	Oréau [195]	A	A	32	6+1	No	1	Yes

In this table, for the modality, *A* (respectively *AV*) refers to audio (respectively audiovisual). Regarding the generation method, *A* (respectively *I*) refers to acted (respectively induced) Regarding emotions, $n + 1 = n$ refers to n different emotions and a neutral condition.

6.2 The EVE (Emotional Validated Expressions) Corpus

This study was approved by the Ethics Committee of the Faculty of Psychology, Speech Therapy, and Educational Sciences of a European University (file number: 2223-087). Actors were recruited voluntarily through targeted advertisements shared via the university’s official email and social media channels, and provided informed written consent after receiving a comprehensive description of the study. Similarly, participants for the perceptual study were recruited via the Prolific platform and also provided informed consent before participating.

6.2.1 Corpus Creation: Audiovisual Data Collection

Beyond the neutral condition, the emotions selected for the corpus were chosen to encompass a broad spectrum, including six basic emotions and four complex emotions, ensuring both foundational and nuanced emotional states are represented for diverse research applications (see Table 6.1).

Each emotion was expressed at two levels of arousal (low and high). For example, for sadness, low arousal was like having a lump in the throat, while strong arousal resembled almost bursting into tears [144].

The sentences of this corpus were carefully chosen to ensure neutral linguistic and semantic content, avoiding any emotional connotations. From a phonemic perspective, the goal was to achieve phonetic completeness and balance. To this end, a phonetically balanced list of sentences was selected from established corpora: the first ten sentences of the Harvard Sentences for English [317] and the FHarvard corpus for French [10].

The EVE corpus was created with the participation of ten actors per language, consisting, for each language, of five males and five females, all possessing expertise in the dramatic arts. These individuals had at least three years of professional acting experience or a background that spans at least ten years in the cinema, television, or theatre as amateur actors. Moreover, all actors were native speakers or near-native speakers of English or French respectively.

The recording sessions were conducted individually in a professional soundproof room to ensure optimal audio quality. The actors stood approximately 20 cm away from a green wall. They wore a headset microphone (AKG C 54) positioned around 5 cm from their mouths and connected to a microphone preamp (Focusrite iTrack Solo). This setup was linked to an Apple MacBook Pro laptop (2.3 GHz Intel Core i5 Dual CoreN) running Camtasia software (version 22.5.4) to enable recording at the high definition (1280x720). In addition, a tracking camera (Obsbot Tiny OWB-2004-CE) was placed in front of them to capture their faces and upper torsos.

Each recording session was structured to last two hours per actor. The actors were instructed to perform each sentence with every emotion at both arousal levels, with two trials for each level, incorporating techniques from the Meisner acting method (also known as the repetition technique [72]) to enhance emotional authenticity and spontaneity. Indeed, it was observed that emotions tended to be more accurately conveyed on the second trial, possibly due to the emphasis on repetition and response [328]. A random sequence of sentences and emotions was assigned for each actor to prevent any order effects. Rigorous quality control measures were implemented, focusing on correcting mispronunciations, reducing hesitations and background noise, and ensuring a consistent recording environment.

In total, for each language, 4,100 utterances were recorded: 100 neutral utterances (10 actors \times 10 sentences), and 4000 emotional utterances (10 actors \times 10 sentences \times 10 emotions \times 2 arousal levels \times 2 trials)

The 2-hour raw video was segmented in Camtasia, with each clip including 300 ms before and after the sentence based on the waveform plot.

For each sentence in the corpus, two files were exported: one stereo video containing both audio and visual cues (format H.264, image resolution of 1,920 \times 1,080, 16:9 aspect ratio, 60 fps, with a MP4 extension), and one

mono audio file (format Waveform Audio, 16-bit, 44.1 kHz, with a WAV extension).

6.2.2 Availability of the Corpus

The EVE corpus is available under the CC BY-NC-SA 4.0 license, accessible through this link: [ANONYMOUS_LINK](#). It comprises 8,200 high-quality recordings, evenly split between English and French. The English corpus totals 3 hours, 46 minutes, 50 seconds, with individual file durations ranging from 2 to 8.12 seconds. The French corpus totals 4 hours, 3 minutes, 45 seconds, with durations from 2.06 to 11.6 seconds.

In addition to the full dataset, the corpus offers subsets organized by arousal level (low or high), modality (audio-only or audiovisual), and individual actors, enabling researchers to focus on specific aspects of SER.

The availability of these structured datasets ensures that researchers can select the most relevant data for their specific purposes, offering maximum flexibility and utility across a wide range of research applications.

6.2.3 Corpus Validation: Perceptual Study

An online perceptual experiment was carried out on the corpus, to assess the presence of the portrayed emotions in the audio recordings. For the perceptual study, 2000 recordings were selected for each language. They correspond to the second trial of each utterance. The goal was to assess the perceived emotions using first the audio and then the audiovisual stimuli. Thus, for each audio, participants were asked to identify the emotion being depicted, and rate their confidence in their identification. Additionally, they were asked whether they would have preferred to select “no idea” instead of choosing an emotion and whether they hesitated between several possible emotions. In the latter case, they were prompted to specify the emotions they considered, ranking them from the most to the least probable. This process was then immediately repeated for the corresponding audiovisual recording, with participants’ previous choices from the audio-only evaluation preselected to help them recall their initial responses. However, they were free to modify their selections if the addition of visual information altered their perception of the emotion or confidence level. Figure 6.1 depicts the perceptual task for both audio and audiovisual modalities

Participants for the perceptual study were recruited through Prolific, a widely used platform for online participant sourcing, ensuring a diverse and representative sample. Participants were compensated £6 for an estimated 1-hour study, although the median completion time for this study was 40 minutes. To ensure data quality, participants were rigorously screened using platform-based filters based on their commitment in previous studies. Random responses were filtered out through automated and manual

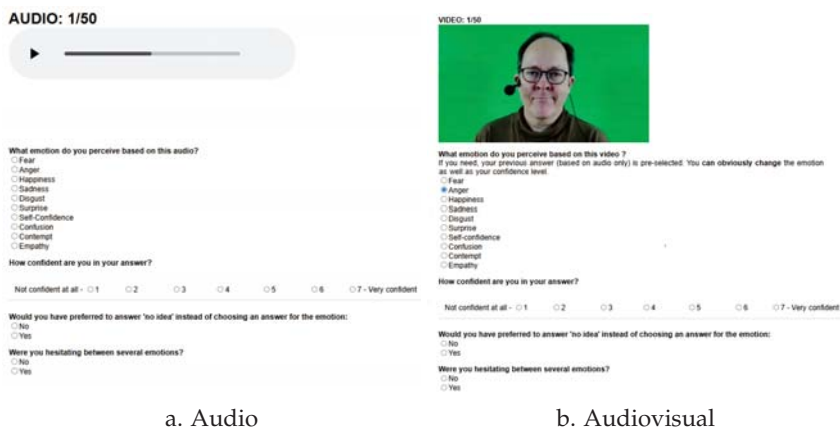


Figure 6.1: Screenshots of the online platform for both audio (a) and audiovisual (b) modalities of the perceptual study.

validation checks. The platform implemented controls, with hidden mechanisms flagging inattentive behaviour, ensuring high-quality annotations throughout.

The study involved 1,200 participants, evenly divided between those fluent in English and French, who were screened for language proficiency (using filters available in Prolific). Participants were also required to reside in countries where the respective language is an official language. Ages ranged from 18 to 78 years for the English group (median age: 34 years), comprising 55.5% females, 43.3% males, and 1.2% who did not specify their gender. For the French group, ages ranged from 18 to 75 years (median age: 29 years), with 49% females, 50% males, and 1% preferring not to disclose their gender.

6.2.4 Availability of the Perceptual Raw Data

Trough the link `ANONYMOUS_LINK`, comprehensive online perceptual experiment results are provided in CSV format, where each row corresponds to an evaluation by a participant for a specific file. Another CSV file summarizes perceived emotions for each recording, with rows representing individual recordings and columns indicating the proportions of perceived emotions. CSV files specific to each parameter are also provided, ensuring a detailed analysis and offering all the necessary information of this corpus.

6.3 Results of the Perceptual Study

In terms of recognition rates per emotion, Figures 6.2 and 6.3 show that visual cues significantly enhance emotion recognition rates in both English and French. Even with audio-only data, all emotions are recognized above the random recognition rate (i.e., 0.1), with some emotions being more accurately identified than others. Specifically, sadness, anger, self-confidence, and surprise tend to be recognized more reliably in English, while in French, sadness, anger, surprise, self-confidence, and confusion show higher recognition rates compared to other emotions.

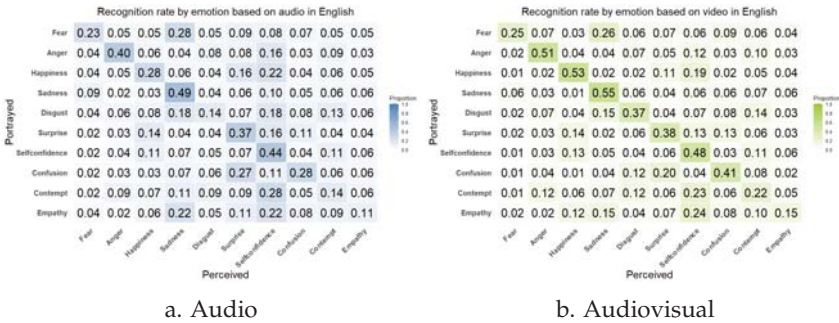


Figure 6.2: Confusion matrices of recognition rate based on audio (a) and audiovisual (b) data in English.

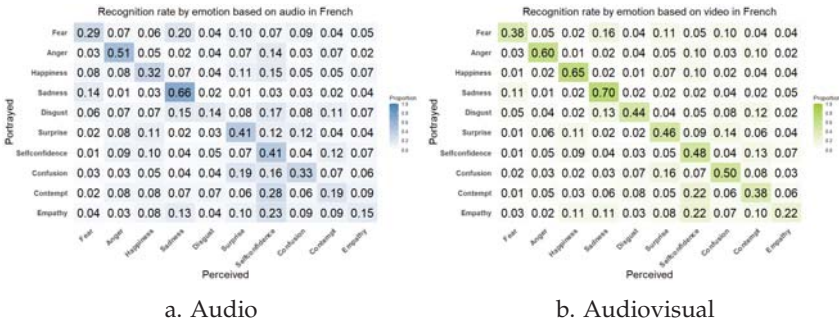


Figure 6.3: Confusion matrices of recognition rate based on audio (a) and audiovisual (b) data in French.

For confidence levels per modality, Figure 6.4 highlights the significant improvement brought by visual cues. Paired t-tests confirm this statistical difference for both English, ($T = -65.64, df = 29999, p < 0.001$) and French ($T = -74.96, df = 29999, p < 0.001$), suggesting that visual information not only boosts recognition rates (Figure 6.2 and 6.3) but also increases

confidence in emotion identification. When analysing emotions individually, all tested emotions exhibit significant differences. Disgust and happiness exhibit the largest confidence gains in both languages. In English, confusion, anger, and self-confidence follow, while in French, contempt, confusion, and self-confidence follow with similar effects.

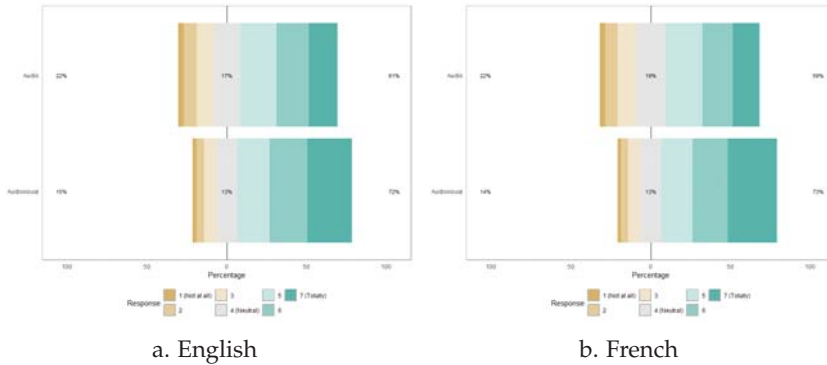


Figure 6.4: Comparison of confidence level based on modality in English (a) and in French (b). Participants answered on a 7-point Likert scale ranging from "Not at all" to "Totally".

When examining recognition rates across arousal levels for all emotions combined, Figure 6.5 illustrates a trend where emotions expressed with higher arousal are generally recognized more accurately. However, even at lower arousal levels, recognition rates remain consistently above the random choice threshold (0.1), indicating that emotions can still be reliably identified regardless of the arousal level. Paired *t*-tests suggest a tendency for improved recognition at higher arousal levels, though the statistical difference is not strongly significant for either English ($T = -1.9762, df = 9, p = 0.07955$) or French ($T = -2.1722, df = 9, p = 0.05791$). In fact, Chi-squared tests comparing recognition rates for each emotion individually further indicate that, in English, only sadness, self-confidence and empathy do not show a significant difference (at a level of significance $\alpha = 0.05$) in recognition rates between arousal levels. In French, with the same level of significance, only confusion and empathy are non-significant, while all other emotions exhibit a significant effect of arousal levels.

The ranking of recognition rates per actor differs between audio-only and audiovisual conditions (see Figures 6.6 and 6.7), as the actors who are the best in conveying emotions through voice alone are not necessarily the best when visual information are included. These results highlight the impact of individual differences in actors' ability to portray emotions.

Lastly, for each language, the Marascuilo Procedure was applied to test for statistical differences between the recognition rates by sentence.

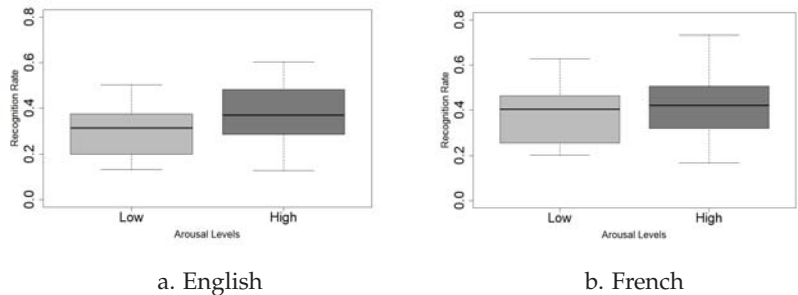


Figure 6.5: Comparison of recognition rate based on emotion arousal level.

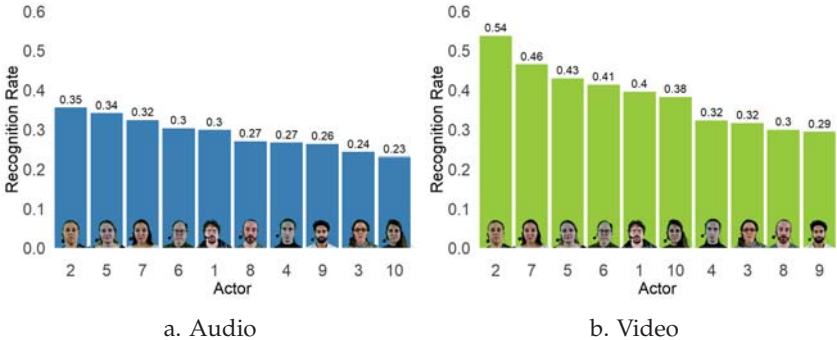


Figure 6.6: Comparison of recognition rate for each actor based on audio (a) and audiovisual (b) data in English.

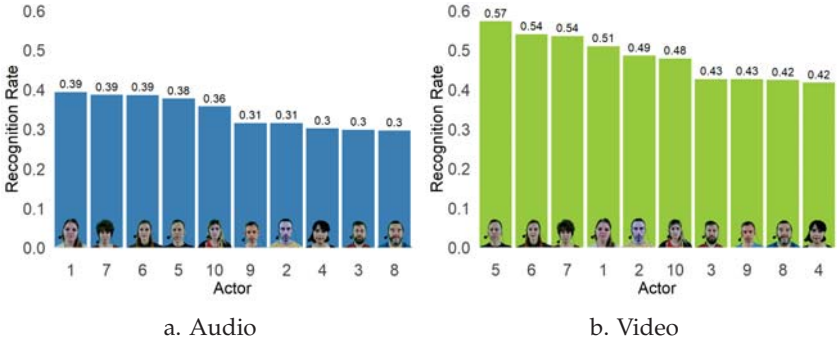


Figure 6.7: Comparison of recognition rate for each actor based on audio (a) and audiovisual (b) data in French.

As expected, no difference was found, as the sentences were designed to be emotionally neutral. Yet, this is an important result that validates our

methodological procedure.

6.4 Discussion

This paper presents a significant advancement in SER research through the creation and validation of the EVE corpus, focusing on emotional audio and audiovisual speech in English and French. While the corpus provides a valuable resource, it carries some limitations or constraints that highlight opportunities for future improvements.

First, the dataset is relatively small. However, research has demonstrated that high-quality small datasets can outperform low-quality large datasets in deep learning applications [308]. Expanding the dataset size could nonetheless enhance its robustness.

Additionally, the corpus is limited to English and French. Although these two languages yielded consistent results, suggesting shared perceptual biases or acoustic and visual patterns, incorporating a wider array of languages and cultural contexts would increase the corpus's global applicability.

Furthermore, the dependence on acted emotional expressions within a single sentence, while ensuring consistency, may not fully capture the richness of natural or spontaneous emotional nuances or how they are perceived [128]. In line with this, research has shown that sentence-length stimuli yield higher recognition accuracy and increased variability compared to isolated words [263], highlighting the importance of stimulus selection in auditory emotion recognition studies. A notable example is surprise, which could benefit from more contextual exploration, as it can range from curiosity and engagement to speechless disbelief marked by incredulity [278]. Similarly, confusion can be split into internal confusion, marked by introspective doubt, and external confusion, shown through hesitation or visible disorientation, with its perception varying depending on factors such as emotion-regulation strategies and cultural context [16]. For these emotions, longer speech segments could offer a richer representation, enhancing recognition accuracy.

In addition, basic emotions (e.g., sadness, anger, happiness) were recognised more accurately than complex ones like empathy or contempt, which showed greater overlap with other emotions. This supports findings by [79], who showed that closely related emotions are harder to distinguish.

Moreover, while some patterns in recognition accuracy were consistent, differences in arousal levels, especially for high-arousal emotions, did not always reach statistical significance. This may be due to perceptual ambiguity or limited expressive cues in short utterances, highlighting the need for further investigation into how arousal level is perceived in emotional speech.

Finally, the variability in emotion perception presents a challenge. Native and fluent participants may perceive emotions differently from non-fluent individuals or those unfamiliar with the languages [104, 359]. Conducting a perceptual study with participants of varying language proficiency levels and different cultural backgrounds could provide valuable insights.

Expanding linguistic and cultural diversity, exploring more nuanced emotional expressions, and incorporating a broader range of participant profiles would significantly enhance the corpus's utility and applicability to real-world scenarios.

Chapter VII

Corpora of Public Speaking in Virtual Reality

This chapter presents part of two ongoing research projects led by two Ph.D. candidates, Lamia Bettahi (University of Liège) and Marion Ristorcelli (Aix-Marseille University). As these works have not yet been published, the detailed results will not be included here to prevent any publication conflicts. Although I was not the lead investigator, I contributed to various aspects of the project and will be a co-author of some papers involving this research, which justifies the inclusion of the data presentation in my thesis. My contributions involved assisting with the study design (*ULiège*), contributing to the statistical analysis (*ULiège* and *AMU*), and extracting voice and speech parameters (*ULiège* and *AMU*). Beyond the ones already cited, the co-authors for *ULiège* are Angélique Remacle, Michaël Schyns, and Anne-Lise Leclercq, while the co-authors for *AMU* are Jean-Marie Pergandi, Rémy Casanova, and Magalie Ochs.

Specifically, this chapter details the development of two corpora designed for research on PS in VR. It aims to provide an overview of their structure, data collection process, and intended applications. The first corpus was created at the University of Liège, and the second at Aix-Marseille University, referred to as the *ULiège corpus* and the *AMU corpus*, respectively, throughout this chapter. It is worth mentioning that these corpora were developed for different research purposes and objectives, and neither this chapter nor any future research intends to compare them directly or use them simultaneously. However, since their overall structure and data collection procedures are similar, they are presented together in a single chapter for clarity and coherence.

7.1 Introduction

PS anxiety can manifest across cognitive, physiological, and behavioural dimensions [214]. Cognitively, it can lead to catastrophic thoughts stemming from the fear of negative evaluation. Physiologically, PS is an acute psychosocial stressor that activates the autonomic stress response, increasing heart rate (HR), sweat production, and pupil dilation within minutes [357]. HR is particularly relevant as an objective measure of stress, given its well-established use in neurobiological research [198]. Elevated HR values correlate with heightened anxiety, with increases of as little as 4.7 beats per minute (bpm) being indicative of significant stress responses in PS contexts [331]. Behaviourally, PS anxiety affects voice and speech production. It leads to an increase in fundamental frequency (F0), resulting in a higher-pitched voice, along with a reduction in F0 variability, making speech sound more monotonous [46, 136, 146, 366]. Speech fluency is also impacted, with an increase in disfluencies such as filled pauses (e.g., "uh," "um" [101]), and prolonged silent pauses [46, 143, 166, 212, 257]. These changes in speech patterns can negatively influence audience perception, affecting the speaker's credibility, career progression, and social reputation [273, 384]. Beyond verbal and vocal aspects, PS anxiety also alters non-verbal behaviour, including posture, gestures, and facial expressions [67, 140].

Repeated exposure to anxiety-inducing situations in a safe environment can help reduce PS anxiety by breaking the association between the situation and the expected negative outcomes [80]. However, real-world exposure is often impractical due to logistical constraints. VR offers a controlled, immersive, and customisable solution, allowing individuals to practice PS in simulated environments that replicate real-world conditions [80, 191, 327]. Despite its potential, only a few studies have directly compared oral communication performance in front of real and virtual audiences. To address this gap, the University of Liège and Aix-Marseille University developed two corpora dedicated to the study of PS in VR. These corpora aim to provide a richer dataset for analysing both non-verbal and/or verbal behavioural indicators, as well as their relationship with anxiety and performance in virtual environments.

The primary objective of these corpora is to analyse PS behaviour in VR, thereby improving the acceptability and usability of virtual environments for training and assessment purposes. The *ULiège corpus* was designed to validate a virtual audience environment by comparing three conditions: PS in VR without an audience, VR with an audience, and in a real-world setting. The main goal was to assess how the presence of a real versus virtual audience influences anxiety, voice, and speech parameters. In contrast, the *AMU corpus* focuses on the impact of virtual audience characteristics, such as gender composition and social attitude, on speakers' perceived difficulty and performance. By examining multimodal behavioural indicators, this

corpus provides valuable insights into how different audience configurations affect PS performance in VR. Indeed, [64, 142, 12, 30] highlighted the audience as a key factor influencing speaker behaviour, and thus a central focus for improving VR training design.

In the context of this thesis, these corpora serve as a critical resource for addressing the scarcity of PS datasets in VR. By leveraging these corpora, the aim is to evaluate objective performance metrics, which will later enable the assessment of training strategies in VR-based PS interventions. Establishing reliable performance measures is a necessary step before exploring the effectiveness of training methodologies, making these corpora foundational to advancing PST in immersive environments.

7.2 Corpora Collection

Both corpus creations received approval from the Ethics Committees of their respective universities.

7.2.1 Conditions

In designing the experimental conditions for both the ULiège and AMU corpora, methodological choices were guided by the need to systematically examine how different audience types and settings influence PS anxiety (for exact measurement details, see Subsection 7.2.3) and performance (for exact measurement details, see Section 7.4). The ULiège corpus included a real audience, a virtual audience, and an empty virtual room to control for anxiety induced by VR immersion itself. Similarly, the AMU corpus explored variations in virtual audience composition and attitude (gender-balanced versus all-male or all-female audiences and positive versus negative attitudes). These decisions align with literature recommendations to control for both environmental and social factors impacting anxiety [205, 204, 373]. Including these varied conditions enables a comprehensive understanding of how virtual and real audiences influence PS behaviour and stress responses.



Figure 7.1: Different conditions for the ULiège Corpus

The real audience consisted of eight individuals (four women and four men) seated around a rectangular table in the meeting room. They were instructed to remain neutral, neither explicitly distracted nor supportive. The virtual environments and virtual agents were created by the *SIG AR/VR Lab* at HEC Liège using Unity platform. In the virtual audience condition, eight virtual agents (four women and four men) were seated around a rectangular table, with the participant standing in front of a screen displaying slides and a timer. A projection screen behind the participant also showed the slides. Virtual agents' posture, facial expressions, and head movements were calibrated to express medium levels of valence and arousal, ensuring a realistic but neutral audience. The perception of these non-verbal behaviours was validated in Chapter III.

In the AMU corpus, participants were asked to deliver their speech under six different conditions. The virtual room was developed by the *Centre de Réalité Virtuelle de la Méditerranée* (CRVM) using Unity platform, featured four agents arranged in a half-circle, varying in gender composition (all-male, all-female, or mixed-gender) and social attitude (positive versus negative, according to [276]) (see Figure 7.2).



Figure 7.2: Different conditions for the AMU Corpus

7.2.2 Public Speaking Tasks

The design of PS tasks in both corpora aimed to ensure ecological validity while controlling for task complexity and content familiarity. This approach ensures methodological consistency and aligns with recommendations to minimise cognitive overload [129].

In the ULiège corpus, each condition featured a different theme related to sports, ecology, or university cultural organisations, ensuring minimal prior knowledge differences among participants. Before their speech, participants had 10 minutes to prepare using a provided set of slides and a one-page informative text. They were instructed to deliver a five-minute presentation.

In the AMU corpus, for each condition, participants were assigned one of several presentation topics, including introducing themselves, describing a city, recounting the story of a film or book, discussing a life project, talking

about their passions, or narrating a typical day. These topics were designed to be accessible and engaging while allowing for variation in speech content. Participants had to prepare in advance notes to assist them and were also given 10 minutes of preparation.

7.2.3 Measures

The selection of subjective and objective measures in both corpora was driven by the need to capture a holistic view of PS performance and anxiety. Subjective measures offered valuable insights into personal experiences and perceptions, while objective measures provided robust, quantifiable data, ensuring a comprehensive and balanced understanding of the findings. This combination aligns with established methodologies for measuring PS anxiety and performance in immersive environments [28, 33].

Subjective Measures

In the ULiège corpus, subjective measures were collected to assess the impact of real and virtual audiences on PS performance. Participants provided self-reported anxiety ratings using the Subjective Units of Distress Scale (SUDS [28]) at multiple time points: before the task, immediately after, and as a retrospective evaluation of anxiety during the task. Anxiety was further assessed using the Liebowitz Social Anxiety Scale - Self Report (LSAS-SR [157]), the State-Trait Anxiety Inventory – Trait subscale (STAI-T [131]) and the Personal Report of Confidence as a Speaker (PRCS [138]). Additionally, the Voice Handicap Index (VHI [383]) was used to screen for voice disorders, while speech fluency was assessed using the Stuttering Severity Instrument - 4 (SSI-4 Riley, 2009). To evaluate immersive tendencies and the sense of presence in VR, participants completed the French version [314] of the Immersive Tendencies Questionnaire (ITQ [381]) before immersion and the ITC-Sense of Presence Inventory (ITC-SOPI [217]) afterwards.

Similarly, in the AMU corpus, subjective measures were collected throughout different stages of the experiment. Upon arrival, participants completed the Personal Report of Public Speaking Anxiety (PRPSA [252]) to assess their PS-related anxiety, the Big Five Inventory (BFI [77]) to measure personality traits. Before their PS task, participants rated their predisposition to feeling immersed in virtual environments using ITQ, then their ease of PS using the PRCS, and the self-reported emotional states were tracked using the Self-Assessment Manikin (SAM [41]), capturing dimensions of valence, arousal, and dominance. After each presentation, they were asked to evaluate their performance (for exact performance measurement details, see Section 7.4), rate the audience attitude and then answer the PRCS and SAM questionnaires again. After completing all the tasks, participants assessed their sense of presence in the virtual environment using the Igroup Pres-

ence Questionnaire (IPQ [334]) and their perceived social presence with the Co-Presence Questionnaire [13]. Finally, an end-of-experience questionnaire was administered to gather qualitative feedback on participants' perceptions of the VR environment and their overall experience.

Objective Measures

The recording methodology for both corpora was carefully designed to ensure high-quality data while maintaining consistency across conditions. Both studies prioritised minimising external distractions and technical biases, aligning with best practices in VR research to enhance immersion and reduce cybersickness [155, 117]. These protocols ensured reliable and valid data collection across all experimental conditions.

In the ULiège corpus, physiological anxiety was tracked via HR recorded every second using a Polar Verity Sense wrist-worn device [137], with HR values analysed at different phases of the task to capture variations in physiological stress. Speech fluency was manually transcribed and analysed using CLAN software [236], measuring the total percentage of disfluencies and the percentage of filled pauses. Regarding speech production, the Praat software and Python code were used to automatically retrieve most important acoustic features decided by the Speech and Therapy team. These included the fundamental frequency (F0) (mean, median, inter-percentile range, inter-quartile range, standard deviation, minimum, and maximum). Additional spectral characteristics were assessed, including the cepstral peak prominence, centre of gravity, spectral slope, and alpha ratio. Voice quality parameters, such as the L0/L1 ratio, noise-to-harmonics ratio, and harmonics-to-noise ratio were also examined. Furthermore, temporal aspects of speech were analysed by measuring silent pause count, total silent duration, and the number of pauses per minute.

In the AMU corpus, objective measures were collected to assess speech, physiological, and behavioural responses during PS in VR. Physiological data were obtained using electrodes and AcqKnowledge software to capture variations in stress responses. Automatic speech-to-text transcription was performed using Whisper, based on pause detection, to facilitate further linguistic analysis. Then verbal, para-verbal, and non-verbal features were extracted through automated scripts developed specifically for this dataset. These scripts, based on the framework outlined in Chapter V, enabled precise analysis of multimodal behavioural indicators. The combination of physiological recordings and automated feature extraction allows for a detailed evaluation of speaker performance in different virtual audience conditions.

In this research, my specific contribution centred on enhancing the transcription process and the automatic extraction of verbal features. By designing and implementing customised Python scripts, I streamlined the

integration of multiple software outputs into a unified processing pipeline. This approach significantly reduced the time and complexity typically associated with managing diverse tools for speech analysis. The development of these scripts required a deep understanding of various software systems, allowing for efficient and accurate extraction of key verbal features. This not only facilitated a more consistent and reliable analysis but also contributed to the overall methodological robustness of the study.

7.2.4 Recordings

For the ULiège corpus, each of the two recording sessions, lasting 1.5 hours, took place in a dedicated 10 × 8 m quiet room. In the real audience condition, participants stood in front of a rectangular table, with the audience seated at a distance of 2 meters. The VR setup featured an HTC Vive Pro Eye Office which was connected to high-performance workstation, providing real-time rendering and low-latency tracking. Two HTC Vive Pro controllers were used for interaction, while external SteamVR 2.0 base stations ensured precise spatial tracking. The researcher monitored the participant's perspective on an external screen. To maintain consistency between conditions, the real meeting room served as the reference for designing the virtual environment. No auditory stimuli were incorporated into the VR simulations.

For the AMU corpus, recording sessions lasted approximately two hours, during which participants delivered their presentations sequentially within a controlled virtual environment. The setup used a Meta Quest Pro head-mounted display connected to a high-performance workstation, providing real-time rendering and low-latency tracking. The virtual environment was developed under Unity. The room was arranged to minimise external distractions, ensuring a controlled experimental setting. Similarly, no auditory stimuli were included in the virtual simulation.

7.2.5 Procedure

The procedures for both corpora were carefully structured to control for confounding variables, such as order effects (a common experimental bias) and adaptation to VR, as unfamiliarity with VR interfaces can increase cognitive load and impact performance [129, 241]. Both studies used counterbalancing to mitigate sequence effects and included calibration phases to ensure participants' comfort within the VR environment. Pre- and post-task measures were applied to track anxiety and presence. By adhering to these standardised procedures, the studies ensured consistent data collection and minimised variability arising from procedural differences, reflecting methodological rigour consistent with previous VR research.

In the ULiège corpus, the experiment was conducted over two sessions scheduled 1 to 14 days apart. One session was dedicated to the real meeting room condition, where participants delivered their speech in front of a live audience, while the other session focused on the VR conditions, where they spoke in a virtual meeting room with either a virtual audience or no audience. The order of exposure was counterbalanced across participants to control for sequence effects. Upon arrival at their first session, participants provided demographic information and completed initial assessments. Each PS task was preceded by a preparation phase, where participants were given a topic and supporting materials. They then delivered their speech while physiological and behavioural data were recorded. In the VR sessions, participants first underwent headset calibration and a brief acclimatisation period prior to immersion. After completing each speech, participants provided self-reported measures and answered questionnaires. The sessions took place in a controlled, quiet environment to minimise external distractions. Figure 7.3 describes the procedure for ULiège.

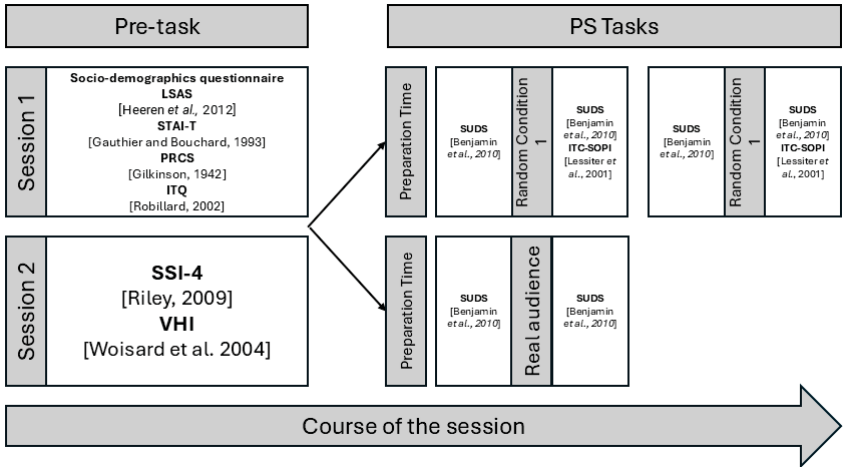


Figure 7.3: ULiège Procedure

In the AMU corpus, the experiment was conducted in a single two-hour session, where participants successively delivered the six presentations in a VR environment. After providing demographic information, participants completed preliminary assessments before beginning the PS tasks. Each participant was immersed in a VR setting where they performed their speech in front of a virtual audience varying in gender composition and social attitude. The sequence of audience conditions was randomised to avoid order effects. The VR headset was calibrated before immersion, and participants were monitored for potential cybersickness symptoms. Dur-

ing each speech, physiological and behavioural data were recorded while the researcher observed the participant’s perspective on an external screen. After each presentations, participants provided post-task self-reports and quantitative questionnaires on their experience. Finally, they answer two presence questionnaires. Figure 7.4 describes the process for AMU.

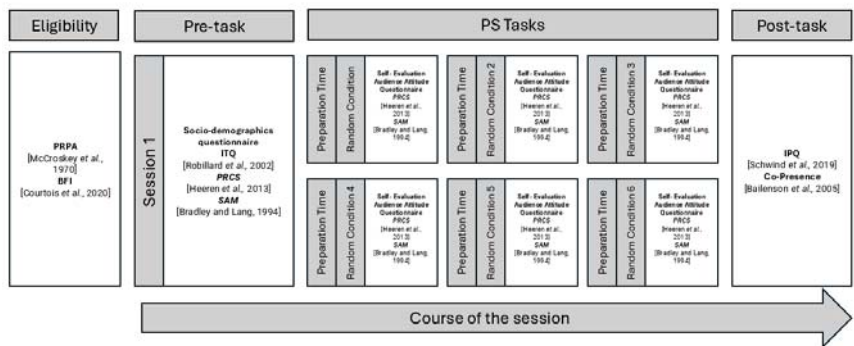


Figure 7.4: AMU Procedure

7.3 Participants

To ensure adequate statistical power and the reliability of findings, a sample size of 60 participants was determined to be sufficient for both studies. This number was selected based on standard practices in behavioural research, allowing for robust comparisons and meaningful statistical analyses across experimental conditions.

The sample for the ULiège corpus included university students (34 females and 26 males) with a mean age of 21.03 years (SD = 6.4). Inclusion criteria required participants to be native French speakers and first- or second-year university students. Recruiting native speakers ensured that variations in anxiety and oral communication (voice and speech) were not influenced by language proficiency. Additionally, selecting newly enrolled students minimised variability in PS experience related to academic exposure. As the study focused on non-pathological individuals, exclusion criteria included self-reported voice, fluency, or hearing disorders at the time of the experiment. The absence of voice and fluency disorders was further verified through specific screening tasks.

For the Aix-Marseille University corpus, participants were recruited based on similar criteria, with an additional selection criterion related to anxiety levels (having a PRPSA score below 120, consistent with [252]’s norms and prior research [110]). Participants were required to be within the

target age range and have no professional training in PS to ensure a natural variation in speech fluency and anxiety responses. The sample consisted of 27 females, 30 males, and 3 individuals who preferred not to disclose their gender, with a mean age of 31 years ($SD = 13$).

In total, the ULiège corpus comprises 180 PS recordings, each lasting five minutes, while the AMU corpus consists of 360 presentations (60 participants delivering six speeches each), with a duration of three minutes per speech.

7.4 Performance Evaluations

Performance evaluation combined subjective self-assessments and/or external perceptive judgments to ensure robust and balanced insights. Both corpora included a subset of recordings reviewed for inter-annotator reliability. This dual approach is supported by research that advocates for diverse rater perspectives to ensure comprehensive and unbiased performance evaluations [273, 384]. The integration of structured rating criteria and qualitative feedback ensures a deeper understanding of PS effectiveness across different conditions.

In the ULiège corpus, each speech recording was evaluated by four independent judges: two PS experts and two Speech Therapy experts raters with no specific expertise in PS. This approach ensures a balanced assessment, incorporating both expert-level evaluation and general audience perception. In addition to providing ratings on a 5-point Likert scale, the experts answered specific questions about key voice and speech parameters (simplified for accessibility) to establish a connection between their subjective assessments.

In the AMU corpus, self-assessment was conducted using a 7-point Likert scale, allowing participants to rate their perceived effectiveness. Then, speeches will be evaluated by one PS expert, with 20% of the recordings additionally assigned to a second expert to measure inter-annotator agreement. While expert evaluations remain subjective, they offer a more informed and consistent assessment than self-evaluations. These external evaluations will enhance the analysis of speech quality and effectiveness by providing a more reliable reference point for performance measurement.

7.5 Results

The ULiège corpus demonstrated that the virtual audience elicited notable changes in anticipatory anxiety, voice (median fundamental frequency), and speech (frequency of filled pauses) when compared to the control condition without an audience. Additionally, participants reported a strong sense of presence in the virtual environment, with minimal side

effects such as cybersickness, thereby supporting the acceptability and usability of the VR setup for PS tasks. Furthermore, distinct VR immersion profiles were identified among participants, suggesting variability in how individuals experience and adapt to virtual environments.

For the AMU corpus, statistical analyses are still ongoing to assess the specific effects of audience gender composition and social attitude (positive versus negative) on anxiety, voice, and speech parameters. These analyses aim to provide deeper insights into how variations in virtual audience characteristics influence PS performance and participant experience in immersive environments.

As the findings from both the ULiège and AMU corpora are part of ongoing publications, detailed results and statistical analyses are not presented in this chapter to avoid any conflict with future dissemination.

7.6 Conclusions

This chapter presented the development and structure of two corpora designed to investigate PS in VR environments: the ULiège and AMU corpora. These corpora contribute to advancing research on PS anxiety, performance, and behavioural indicators within immersive contexts.

My contributions to these projects encompassed key aspects of the research process, including supporting the design of the experimental methodology, assisting with the setup and execution of VR recordings, and contributing to the statistical analysis. These contributions were integral to ensuring methodological rigour and the reliable collection of data.

While this chapter does not detail specific results, the corpora provide a valuable foundation for future research. They will be used to explore objective and subjective performance metrics, investigate individual differences in VR immersion, and develop more adaptive and effective VR-based PST interventions. As previously mentioned, detailed findings will be disseminated through forthcoming publications.

Overall, these corpora lay the groundwork for advancing research on VR-based PS interventions by offering empirical data on behavioural, verbal, and physiological responses, contributing to the development of more effective and adaptive training environments.

Chapter VIII

Public Speaking Performance Prediction Using Machine Learning

This chapter presents part of what I worked on during my research stay at Aix-Marseille University. This work began in November 2024. The co-authors are Marion Ristorcelli, Michaël Schyns, and Magalie Ochs. Specifically, this chapter aims to predict the self-evaluation of participants performances using the AMU corpus, presented in chapter VII, based on the multimodal features presented in Chapter V. As this research is still ongoing, the current analysis is limited to self-evaluations (excluding expert evaluations) and focuses solely on one type of neural network, namely multi-layer perceptron.

8.1 Introduction

VR environments offer immersive platforms for simulating PS scenarios, making the evaluation and enhancement of presentation performance a critical task. Accurately assessing presentation quality is challenging, particularly when it involves capturing and interpreting subtle multimodal cues, including vocal characteristics, non-verbal gestures, and behavioural patterns [384, 20].

To address this, the chapter explores the application of multi-layer perceptron (MLP), a type of Machine Learning (ML) model known for effectively capturing complex, non-linear relationships within data. The classification task focuses on predicting self-evaluation scores reported on a 7-point Likert scale after each presentation based on the multimodal features.

However, given the central goal of providing useful feedback to

speakers, several labelling strategies were tested to improve learning performance and interpretability. These include three-class groupings (e.g., low–neutral–high) and two-class groupings (e.g., bad vs. good).

8.2 Theoretical Background

8.2.1 Prediction

ML develops algorithms capable of learning patterns from data and making predictions without being explicitly programmed for each specific task. In the context of this chapter, the aim is to model complex relationships between multimodal features (input) and the self-evaluated quality of participant’s presentation (output).

This chapter focuses on *supervised learning*, a subset of ML where the model is trained on a dataset containing input-output pairs. The goal is to learn a function that can accurately predict the output for unseen inputs. A classification model, also known as a classifier, aims to categorise inputs into one of several predefined classes. In this study, the classes represent levels of self-evaluation scores (grouped or not) provided by participants.

8.2.2 Multi-layer Perceptron

An artificial neural network is a computational model inspired by the human brain’s network of neurons. It consists of layers of interconnected neurons that process input data and generate an output.

At the most fundamental level, a single neuron takes multiple inputs, each associated with a weight, and computes their weighted sum. This sum is passed through an activation function, which determines whether the neuron should be activated (see Figure 8.1). A commonly used activation function is the *sigmoid* function. In the case of multi-class classification, the *softmax* function is used instead to convert the outputs into a probability distribution over multiple classes. This simple model can be extended to networks with multiple layers, allowing the representation of complex relationships in data.

A Multi-Layer Perceptron (MLP) extends this concept by introducing multiple layers of neurons. An MLP consists of an input layer, one or more hidden layers, and an output layer (see Figure 8.2). Each neuron in a layer is connected to every neuron in the next layer, forming a fully connected network. The neurons in each layer apply a weighted sum and an activation function before passing their outputs to the next layer. This architecture enables the network to model complex, non-linear relationships. The intelligence of the MLP is thus encoded in its weights, which determine how inputs influence the final prediction.

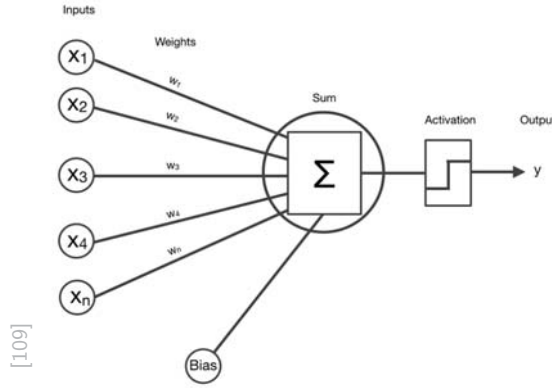


Figure 8.1: A single neuron computing a weighted sum ($z = \sum_{i=1}^n w_i X_i + b$) and applying an activation function to predict a single output neuron y .

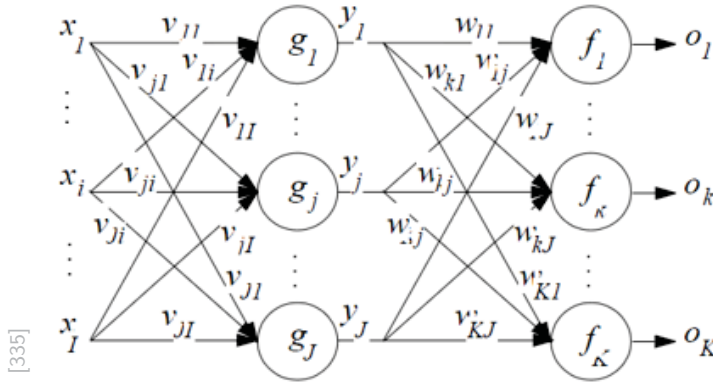


Figure 8.2: Multi-Layer Perceptron architecture for classification, including forward and backward propagation. The network consists of an input layer with I inputs, a hidden layer with J neurons, and an output layer with K neurons. Each hidden neuron g_j computes a weighted sum of the inputs using weights v_{ji} , and then applies an activation function, and outputs y_j . Then, each output neuron f_k computes a weighted sum of the y_j using weights w_{kj} and finally applies activation functions to compute the output o_k . During training, errors between predicted and target outputs propagate backward through the network, and all weights v_{ji} and w_{kj} are adjusted to minimise the classification loss.

In classification tasks, the structure of the output layer depends on the number of target classes. For binary classification, a single output neuron is typically used, while for multi-class classification, the output layer contains one neuron per class.

The effectiveness of an MLP depends not only on hyper-parameters such as the number of hidden layers, neurons, and the choice of activation functions, but also on the training process, which aims to minimise the error between predicted and actual outputs. This is achieved by iteratively adjusting the weights using backpropagation, an algorithm that computes the gradient of the loss function with respect to each weight. The model continuously refines its parameters until the loss converges to an optimal value.

To achieve optimal performance, an MLP relies on several key hyper-parameters that influence its learning capacity and efficiency. One of the most critical parameters is the learning rate, which controls how much the model adjusts its weights during training. A learning rate that is too high may prevent convergence, while one that is too low can lead to slow learning. The batch size also plays a significant role, determining how many samples are processed before updating the weights. Training is typically performed in mini-batches, striking a balance between computational efficiency and convergence stability. The number of epochs, i.e., the complete passes through the dataset, must be carefully chosen to avoid both underfitting and overfitting. Another essential hyper-parameter is the optimiser, which determines how the model updates its weights. For example, *Adam* is a widely used optimiser that combines adaptive learning rates with estimates of first and second moments of the gradients, allowing for faster and more stable convergence during training. The activation functions applied in hidden and output layers, further influence how the network learns complex patterns.

8.2.3 Cross-Validation

The training and validation procedure is essential to ensure that the model learns meaningful patterns while maintaining generalisability to unseen data. During training, the model iteratively updates its weights based on a training dataset by minimising the difference between predicted and actual outputs. This optimisation process enables the model to progressively refine its internal parameters. Validation is used to assess performance on a separate dataset, helping to detect overfitting, where the model performs well on training data but generalises poorly to new data. By evaluating the model on a validation set, adjustments can be made to hyper-parameters and the overall architecture to improve predictive capability.

To address the small dataset size, Leave-One-Out Cross-Validation (LOOCV) was used to evaluate model performance. This approach max-

imises data usage but is computationally demanding (see Figure 8.3 for full description).

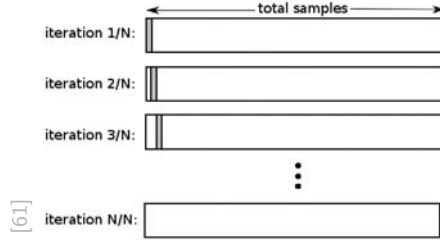


Figure 8.3: Leave-One-Out Cross-Validation process. Each iteration selects a single sample as the validation set (highlighted), while the remaining samples are used for training. This process is repeated N times (N being the total number of observations), ensuring that each sample is used once as validation.

8.2.4 Model Evaluation

The performance of classification models is evaluated using standard metrics, including the confusion matrix, accuracy, precision, recall, and F1-score.

In K -class classification problem, the *confusion matrix* is a $K \times K$ table. Each entry (i, j) represents the number of instances whose true label is class i but were predicted as class j . The diagonal elements correspond to correct predictions (true positives for each class), while off-diagonal elements indicate misclassifications. *Accuracy* measures the proportion of correct predictions, *precision* reflects the proportion of true positives among all predicted positives, and *recall* indicates the proportion of true positives among all actual positives. The *F1-score*, the harmonic mean of precision and recall, balances the two. In the case of imbalanced datasets, macro and weighted averages are often used: the *macro average* gives equal weight to each class, while the *weighted average* accounts for class frequency.

8.3 Data

8.3.1 Description of the Dataset

The data used in this study is sourced from the AMU corpus (see Chapter VII), comprising multimodal features (verbal and non-verbal) fully described in Chapter V. Specifically, the features used in this study include

verbal features, with acoustic cues such as pitch (mean and variation), loudness (mean and variation), rate of loudness peaks, silent pause count (distinguishing small, medium, and long pauses), total duration of silent pauses, and speech rate. Linguistic features and complementary metrics, include the percentage of words related to analytical thinking, clout, authenticity, emotional tone, anxiety, swear words, netspeak, assent, non-fluencies, fillers, and overall disfluencies count. Non-verbal features include gaze-related metrics (such as the score of characters viewed, gaze fixation time, gaze duration by audience member, hesitation time, and gaze direction entropy), body posture and movement features (including horizontal displacement amplitude, distance covered, and movement entropy), gesture-related metrics (such as the frequency and duration of palm-up, palm-down, neutral, open, and closed hand positions, hand detection rate, posture openness, and entropy of hand openness and direction), and facial expression features (including the frequency of positive and negative expressions and the entropy of blend-shapes).

The self-evaluation scores, provided by participants after their presentations, serve as the target variable for prediction. Figure 8.4a shows the distribution of scores across the 7-point Likert scale. In order to provide more interpretable and meaningful feedback to the speaker, grouping strategies are explored. Three-class strategies aim to preserve distinctions between weak, average, and strong performances: one uses the midpoint of the scale (1–3 as low, 4 as neutral, 5–7 as high), and the other follows thresholds used in prior work (1–2 as low, 3–5 as moderate, 6–7 as high) [60, 283]. In addition, two binary grouping strategies were tested: one grouping 1–4 as low and 5–7 as high, and the other considering 1–3 as low and 4–7 as high. These configurations allow for different levels of granularity in feedback, depending on whether the goal is to distinguish finer differences or provide broader, actionable guidance. Figures 8.4b to 8.4d show the distributions of these different groupings.

8.3.2 Data Segmentation

The dataset originally comprised 360 full-length presentation recordings. To increase data volume and capture temporal dynamics, each presentation was segmented into three equal parts (beginning T_1 , middle T_2 , and end T_3) following [30], resulting in 1,080 segments. This approach enables finer-grained analysis of performance over time and aligns with socio-cognitive theories such as the primacy and recency effects, which suggest that the beginning and end of a presentation may disproportionately influence audience perception [30]. All features were recomputed separately for each segment to capture these temporal variations in behaviour and delivery.

8.3.3 Dataset Preparation

A fundamental step in any data-driven study is ensuring that the dataset is clean, consistent, and properly formatted before analysis. The quality of the data directly impacts the validity of the results, making data cleaning and preprocessing essential in avoiding misleading interpretations. Preparing the data involves handling missing values, ensuring uniform feature representation, and applying transformations to standardise the input variables.

Handling missing values is particularly crucial to preserve the integrity of the dataset while maintaining as much information as possible. In this study, feature extraction follows predefined criteria based on Chapter V, allowing for the computation of relevant features. However, due to the nature of the task, certain missing values arise that require careful handling. Missing values may result from technical issues during recording. When a specific feature is partially missing but available in other segments for the same subject and condition, the missing value is imputed using the mean of that feature within the corresponding subject-condition combination. In cases where no data is available for a specific condition, the mean is instead computed using observations from other conditions of the same subject. This approach ensures that missing values are filled in a way that maintains individual consistency while leveraging existing data to approximate missing information. Another case of missing data occurs when features are inherently undefined due to the experimental setup. For instance, the duration of gaze towards female members is undefined when the audience consists solely of male participants (see Chapter VII). This situation is different from a case where female members were present but not looked at. To address this, such undefined values are assigned a value of -1 to maintain consistency in feature representation.

Beyond handling missing data, standardisation and normalisation are critical steps in ensuring that features are on a comparable scale. Differences in magnitude across variables can introduce biases in machine learning models, where features with larger values may dominate those with smaller values. To prevent this, all variables are standardised per participant, ensuring that internal variations do not introduce unintended differences between subjects. This step is particularly important for studies involving human performance, where individual baselines can vary significantly.

The choices made in this preprocessing stage, from handling missing values to applying standardisation, have a direct impact on the downstream analysis. Any inconsistency or bias introduced at this step can propagate through the modelling process, influencing both the interpretability and the performance of predictive models. Ensuring a robust data preparation is therefore essential in achieving reliable and meaningful results.

8.3.4 Data Augmentation

The dataset presents class imbalance (see Figure 8.4), which can lead to biased predictions favouring majority classes. To address this, the *Synthetic Minority Over-sampling Technique (SMOTE)* was applied [58]. It generates synthetic minority class samples by interpolating between similar observations [60]. SMOTE was applied to verbal and non-verbal features of each training set to preserve the internal structure of each modality and avoid unrealistic combinations. This approach was preferred over class weighting or undersampling, as it better maintains multimodal diversity and supports the model’s ability to capture nuanced behaviours in PS performances.

8.3.5 Model Configuration

To ensure transparency and reproducibility, Table 8.1 summarises the final configuration of the MLP model used throughout the experiments.

Parameter	Value / Description
Number of hidden layers	2
Neurons per layer	16
Activation function	ReLU
Output layer activation	Softmax (multi-class) / Sigmoid (binary)
Loss function	Categorical Cross-Entropy / Binary Cross-Entropy
Optimizer	Adam
Learning rate	5.75×10^{-4}
Batch size	16
Epochs	300

Table 8.1: Final MLP model configuration

8.3.6 Finetuning

Hyper-parameter tuning was performed using *Optuna* [2] on the seven-class classification task to optimise model performance. The search covered the number of hidden layers (1 or 2), neurons per layer (8 or 16), learning rate ($1e^{-5}$ to $1e^{-3}$), batch size (16, 32, or 64), and number of epochs (100 to 300). The final configuration, selected after 100 trials, used 2 hidden layers with 16 neurons each, a learning rate of 5.75×10^{-4} , batch size of 16, and 300 epochs. These hyper-parameters were then fixed for all label groupings to ensure comparability.

8.4 Results

All results presented in this section are based on the MLP model configuration described in Sections 8.2.2 and 8.3.6. Due to the limited size

of the dataset, performance was evaluated using LOOCV, and reported metrics represent the mean across all iterations. This approach maximises data usage and provides an almost unbiased estimate of generalisation performance [57, 232]. Thus, the class distributions used for comparison correspond to those shown in the distribution figures.

8.4.1 Performance with Seven Classes

The results of the MLP model on seven classes (see Figure 8.5) show limited performance on the seven-class classification task. Accuracy reached 21.6%, which is above the random baseline of 14.3% but still below the naive strategy of always predicting the most frequent class (28.7%, see Figure 8.4a). The weighted F1-score was 21.8%. Class 6 achieved the highest recall (29.4%), while classes 1, 2, and 7 were classified worse than random, with recall scores below the random baseline. The confusion matrix revealed frequent misclassification between adjacent classes from classes 3 to 6. These results highlight the difficulty of generalising in a seven-class setting and support the use of alternative groupings to improve prediction reliability and feedback relevance.

8.4.2 Performance with Grouped Classes

In this section, the results of the MLP model are presented, now focusing on a grouping strategy with fewer classes.

Grouping in Three Classes

Figures 8.6 and 8.7 show the results for the 3-class groupings. The 123–4–567 strategy yielded 38.7% accuracy, which is above the random baseline of 33.33% but substantially below the naive strategy of always predicting the most frequent class (57.5%, see Figure 8.4b) and a weighted F1-score of 41.1%. Class 123 was the easiest to identify (recall: 43.3%), while class 4 was the most difficult (recall: 34.7%, score just above the random baseline), frequently misclassified into adjacent groups.

The 12–345–67 grouping achieved better results, with 48% accuracy (which is, once again, above the random baseline of 33.33% but substantially below the naive strategy of always predicting the most frequent class (62.1%, see Figure 8.4c) and a score of 49% for weighted F1-score. Class 345 showed the highest recall (51%). Classes 12 and 567 were often misclassified as 345. These results suggest that broader groupings improve classification, though distinguishing between low and high performance levels, especially relative to the neutral class, remains challenging. This difficulty is likely due to overlapping features and subjective self-evaluations, and since the second

class was dominant in this grouping (see Figure 8.4c), SMOTE may have had limited impact on improving performance.

Grouping in Two Classes

Figures 8.9 and 8.8 present the results for the two binary groupings.

The 1234–567 strategy achieved 53.4% accuracy, which is above the random baseline of 50% but slightly below the naive strategy of always predicting the most frequent class (57.5%, see Figure 8.4d). The weighted F1-score is similar, since this grouping was quite balanced. This grouping allows the model to identify the lower class 1234 relatively well, but performs close to random for the higher class 567, which is not even predicted correctly in most cases.

Using 123–4567 strategy achieved 56.9% accuracy, which is above the random baseline of 50% but substantially below the naive strategy of always predicting the most frequent class (77.9%, see Figure 8.4e). The weighted F1-score is 57.1%. This grouping allows the model to identify the lower class 123 relatively well, and the model manages to distinguish the higher class 4567, correctly predicting it 54% of the time.

8.4.3 Segmentation

This section evaluates the MLP model on different segments of the presentations—overall (full 3 minutes), T_1 (first minute), T_2 (second minute), and T_3 (third minute)—using binary classification with the 123–4567 grouping, which showed the best performance (see Figure 8.9). Unlike previous analyses based on 1,080 segments, this comparison uses 360 full presentations, analysing each time interval independently.

The classification performance without segmentation, using the entire 3-minute presentations, achieved an accuracy of 56%, with a similar F1-score. The results align with results using segmentation (see Figure 8.10a) and suggests a reasonable ability to distinguish between the two groups when considering the full presentation.

Figures 8.11, 8.12, and 8.13 present results obtained using only segments T_1 , T_2 , and T_3 respectively.

The first minute (T_1) achieved the highest accuracy (58.6%) and has a similar confusion matrix than when using the whole presentation (see Figure 8.10b). Accuracy dropped in T_2 (52.3%) and further in T_3 (50.3%), indicating increasing difficulty in distinguishing performance levels as the presentation progresses (see Figures 8.12 and 8.13).

8.4.4 Expert Evaluations

Until now, the model has aimed to predict participants' self-evaluation scores, introducing two levels of approximation: one from the model itself, and another from the variability and subjectivity inherent in self-assessment. The latter (stemming from participants' personal biases) can be mitigated by relying on external expert ratings, which provide a more consistent and standardised reference across participants. Thus, this section briefly explores the impact of using expert evaluations for 7-class classification. While the distribution (Figure 8.14) resembles that of self-evaluations (Figure 8.4a), individual ratings often differ, highlighting the added value of more objective assessments.

Figure 8.15 shows the 7-class model results using expert evaluations. Overall performance remains similar to self-evaluations, with nearly identical macro F1-scores (16.0%, and 16.1%). However, class 1 recall improves markedly to 44.4%, suggesting that the identified multimodal cues are better able to predict expert low ratings than low self-assessments. While this highlights the potential of expert-driven labels for more reliable classification, further analysis is beyond the scope of this chapter.

8.5 Discussion

This study examined the impact of class grouping strategies on model performance, highlighting the trade-offs between classification accuracy and feedback granularity in VR PST. The results confirm that simpler class structures improve classification accuracy but at the cost of reduced interpretability. Reducing intra-class variance helped the model generalise better, as reflected in lower misclassification rates within the confusion matrices of grouping strategies. Depending on the intended goal, among the three-level grouping strategies presented in this chapter, the approach proposed by [60, 283], i.e., 12-345-67 grouping, appears to be the most suitable choice at this stage. For binary classification, using the preliminary results of this chapter, the neutral score should be considered a high performance rather than a low one, leading to the selection of the 123-4567 grouping.

However, broader groupings introduced bias by merging distinct performance levels, potentially masking subtle differences in PS skills. The difficulty in classifying intermediate categories, suggests that the features extracted for these performance levels were not sufficiently distinct. This challenge may be due to overlap in verbal and non-verbal characteristics or subjective variability in self-assessment, both of which contribute to ambiguity in classification.

Another important consideration is the impact of timing segmentation on classification. Segmenting the presentation into temporal phases allows

for finer-grained analysis and may uncover when specific behaviours are most predictive of performance. Preliminary results suggest that early segments are more strongly correlated with self-evaluation scores, indicating a possible primacy effect. However, additional testing, particularly with expert ratings, is needed to validate this finding or to corroborate [30]’s findings.

The choice of an MLP architecture enabled the model to capture complex relationships between multimodal features. However, its performance was constrained by the complexity of multimodal feature interactions and the inherent variability introduced by subjective self-evaluations. It was intentionally chosen for its simplicity, to demonstrate that the data, preprocessing, and modelling pipeline were ready and functional for predicting performance scores based on the multimodal features identified. Hyperparameter tuning using Optuna improved model generalisation by selecting optimal configurations for learning rate, network depth, and regularisation. Despite these improvements, the results indicate that more advanced architectures, such as Recurrent Neural Networks (specifically Long Short-Term Memory or Gated Recurrent Unit models) or Transformer based approaches, may be better suited for capturing long-range dependencies and subtle variations in multimodal PS data.

While the current experiments establish the foundation for automated performance prediction using multimodal features, they represent an initial step. Further experiments are needed to systematically evaluate the impact of alternative model architectures, input representations, and training strategies. This includes testing different neural network depths, activation functions, and optimisers, as well as exploring feature selection or dimensionality reduction methods. Such investigations would help refine the modelling approach and identify configurations that better capture the complexity of public speaking behaviour.

Data augmentation and SMOTE were applied to mitigate class imbalance, improving data diversity and model robustness. However, SMOTE may not have fully captured the complexity of multimodal features, particularly for speech and behavioural patterns in PS. While generating synthetic samples helped balance class distributions, it did not entirely resolve the difficulty in classifying under-represented or intermediate classes. This suggests that augmentation techniques need to be refined to better reflect the temporal and expressive variations inherent in PS performances. Further work could explore alternative strategies such as few-shot learning [376] or Generative Adversarial Network (GAN)-based [192] data generation to better address the natural imbalance in performance classes.

Additionally, the absence of feature attribution analysis means that it is unclear which specific verbal or non-verbal cues most influenced the model’s decisions. Once the model achieves sufficient performance, a deeper interpretability analysis using SHAPLEY values [231] could provide insights

into the importance of different multimodal features in classification decisions.

In addition, future work should consider setting aside a small test set for model selection or fine-tuning. While LOOCV offers an almost unbiased estimate of generalisation, it may exhibit higher variance compared to other cross-validation methods, such as k-fold cross-validation [154]. Using a dedicated test set would help avoid overfitting during hyper-parameter optimisation and ensure better assessment of model performance when comparing different architectures or preprocessing strategies.

Furthermore, the large number of extracted features may introduce redundancy and noise, potentially hindering the model's ability to focus on the most relevant predictors. Another important consideration is the potential benefit of selecting only participants with high variability in presentation scores for training, as they may exhibit specific patterns that are not captured in participants with less variation.

Future research should aim to enhance data reliability by incorporating objective performance measures, such as expert evaluations, in addition to self-assessments. Expanding the dataset to include more diverse and representative samples (self-assessments or expert ratings), especially from under-represented classes, will also be crucial for improving model generalisability and robustness.

From an applied standpoint, the preliminary results suggest that binary and three-class classifications have the potential to provide consistent and interpretable feedback within VR-based training system, but these findings are based on limited data and exploratory analyses and require further validation to confirm their reliability and effectiveness across broader contexts. Anyway, this broader categorisation comes at the cost of reduced granularity, limiting the ability to capture specific strengths and weaknesses. To address this, future VR training tools should strive for a balance between classification accuracy and detailed, personalised feedback—ideally by combining automated predictions with expert input to deliver a more comprehensive and actionable evaluation of public speaking performance.

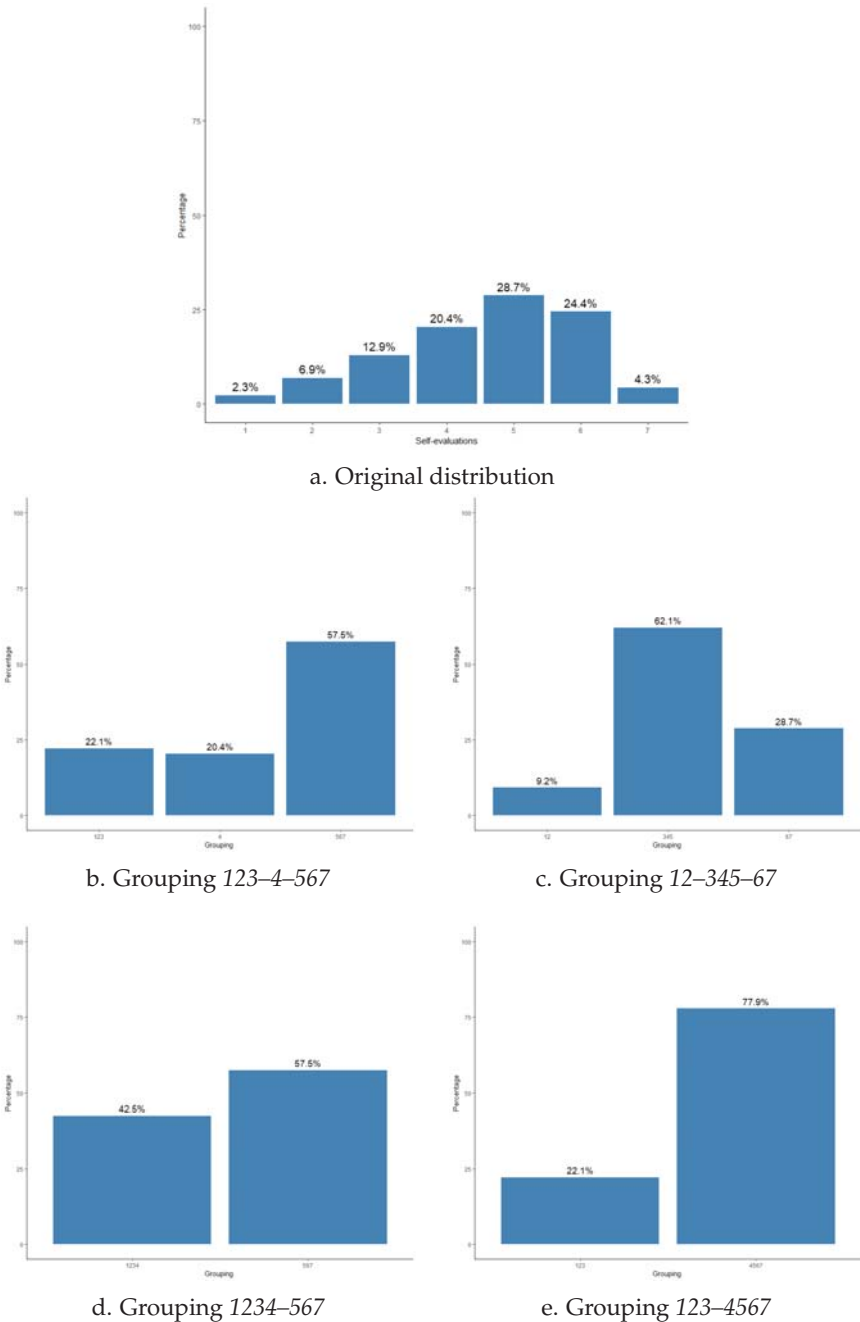
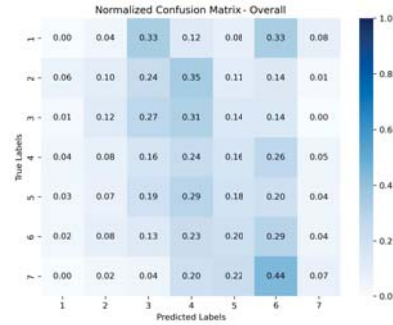


Figure 8.4: Distribution of self-evaluation scores across all grouping strategies

Class	Precision	Recall	F1-Score
1	0.000	0.000	0.000
2	0.083	0.097	0.090
3	0.196	0.274	0.228
4	0.185	0.239	0.209
5	0.298	0.177	0.222
6	0.302	0.294	0.298
7	0.079	0.067	0.072
Accuracy	-	-	0.216
Macro Average	0.163	0.164	0.160
Weighted Average	0.232	0.216	0.218

a. Metrics summary



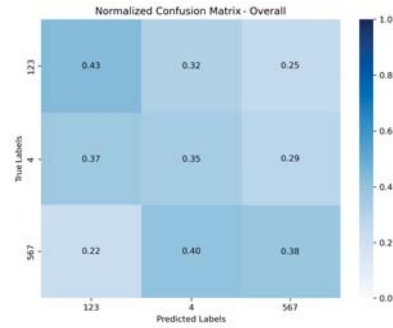
b. Confusion matrix

Figure 8.5: Overall results with 7 classes

Class	Precision	Recall	F1-Score
123	0.324	0.433	0.370
4	0.192	0.347	0.247
567	0.659	0.383	0.485
Accuracy	-	-	0.387
Macro Average	0.391	0.388	0.367
Weighted Average	0.489	0.387	0.411

a. Metrics summary

Grouping 123-4-567



b. Confusion matrix

Grouping 123-4-567

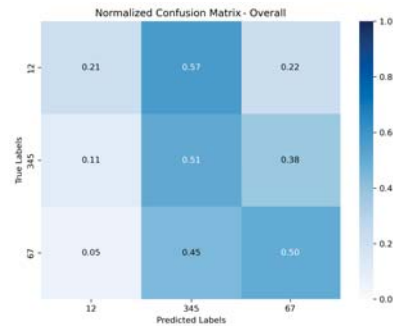
Figure 8.6: Overall results with 3 classes

Grouping 123-4-567

Class	Precision	Recall	F1-Score
12	0.183	0.208	0.195
345	0.637	0.511	0.567
67	0.361	0.500	0.420
Accuracy	-	-	0.48
Macro Average	0.394	0.406	0.394
Weighted Average	0.516	0.480	0.490

a. Metrics summary

Grouping 12-345-67



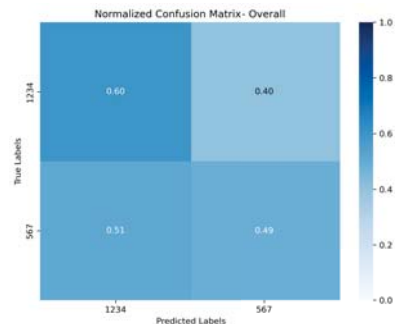
b. Confusion matrix 12-345-67

Figure 8.7: Overall results with 3 classes

Grouping 12-345-67

Class	Precision	Recall	F1-score
1234	0.463	0.599	0.523
567	0.621	0.487	0.546
Accuracy	-	-	0.534
Macro Average	0.542	0.543	0.534
Weighted Average	0.554	0.534	0.536

a. Metrics summary

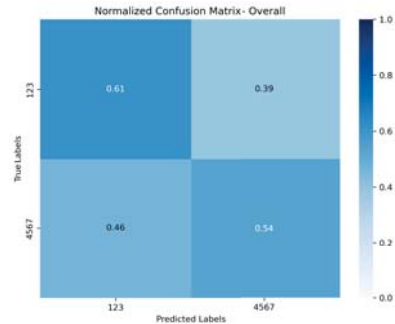


b. Confusion matrix

Figure 8.8: Overall results with 2 classes
Grouping 1234-567

Class	Precision	Recall	F1-Score
123	0.495	0.610	0.546
4567	0.651	0.538	0.589
Accuracy	-	-	0.569
Macro Average	0.573	0.574	0.568
Weighted Average	0.585	0.569	0.571

a. Metrics summary

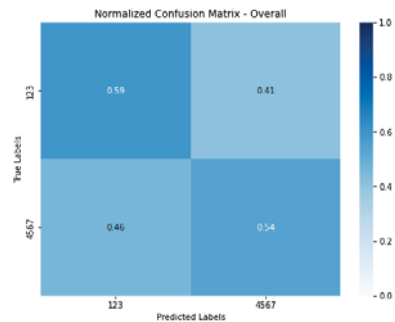


b. Confusion matrix

Figure 8.9: Overall results with 2 classes
Grouping 123-4567

Class	Precision	Recall	F1-Score
123	0.486	0.588	0.532
4567	0.639	0.540	0.585
Accuracy	-	-	0.560
Macro Average	0.563	0.564	0.559
Weighted Average	0.574	0.560	0.563

a. Metrics summary

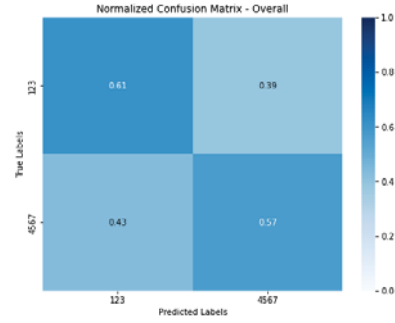


b. Confusion matrix

Figure 8.10: Confusion matrix without segmentation
Grouping 1234-567

Class	Precision	Recall	F1-Score
123	0.511	0.615	0.558
4567	0.665	0.565	0.611
Accuracy	-	-	0.586
Macro Average	0.588	0.590	0.585
Weighted Average	0.599	0.586	0.588

a. Metrics summary

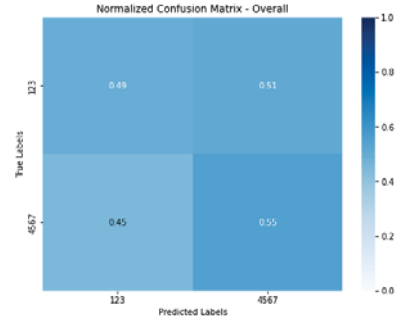


b. Confusion matrix

Figure 8.11: Overall Results with only T_1 segments
Grouping 1234-567

Class	Precision	Recall	F1-Score
123	0.444	0.486	0.465
4567	0.591	0.550	0.570
Accuracy	-	-	0.523
Macro Average	0.518	0.518	0.517
Weighted Average	0.529	0.523	0.525

a. Metrics summary

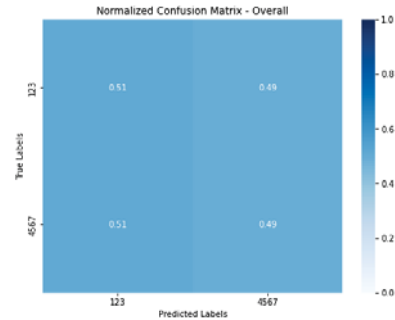


b. Confusion matrix

Figure 8.12: Overall Results with only T_2 segments
Grouping 1234-567

Class	Precision	Recall	F1-Score
123	0.429	0.514	0.468
4567	0.579	0.495	0.534
Accuracy	-	-	0.503
Macro Average	0.504	0.504	0.501
Weighted Average	0.515	0.503	0.506

a. Metrics summary



b. Confusion matrix

Figure 8.13: Overall Results with only T_3 segments
Grouping 1234-567

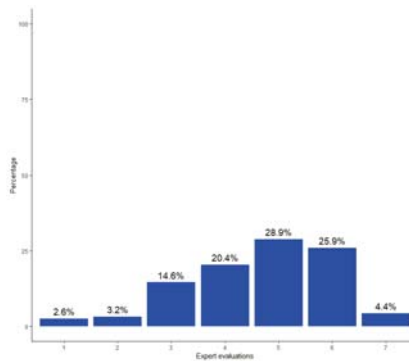
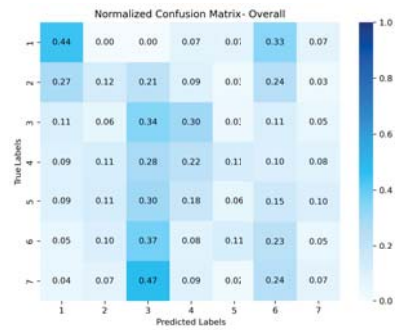


Figure 8.14: Distribution of expert performance scores

Class	Precision	Recall	F1-Score
1	0.121	0.444	0.190
2	0.040	0.121	0.060
3	0.156	0.340	0.214
4	0.266	0.219	0.240
5	0.222	0.061	0.095
6	0.353	0.228	0.277
7	0.040	0.067	0.050
Accuracy	-	-	0.190
Macro Average	0.171	0.211	0.161
Weighted Average	0.239	0.190	0.189

a. Metrics summary



b. Confusion matrix

Figure 8.15: Overall results with 7 classes using expert evaluations

Chapter IX

Conclusion

This thesis has explored the development of a VR-based PST system, integrating machine learning and multimodal analytics to provide structured, data-driven feedback for speakers. Over six years, this tool has been developed from scratch, refining its capabilities to assess and enhance PS skills effectively. This research was guided by five core questions, each of which has been addressed across different chapters of this thesis.

First, the perception of virtual agents' behaviours and their impact on VR-based users was investigated through a systematic literature review and an empirical study on agent's perception in VR. This research highlighted the importance of non-verbal cues in shaping user experience and engagement. Second, the comparison of VR training with traditional methods (e.g., simulations in a regular classroom, or online mock trials via video-conferencing) was explored in the context of legal education, where results demonstrated that VR-based training provides a more immersive and structured learning experience, improving both engagement and knowledge retention. Third, the identification of key verbal and non-verbal indicators of effective PS was examined through multimodal analysis, emphasising the role of vocal, gestural, and facial dynamics. Fourth, the question of whether emotions can be accurately detected based solely on speech was addressed through the EVE perceptive study, suggesting that acoustic parameters only is a promising approach to recognise speaker's emotions, though further validation, such as using ML to predict emotions from the EVE audio corpus, is needed to confirm its robustness. Finally, the potential of machine learning and multimodal analytics for assessing and improving PS skills was examined through the development of performance prediction models, demonstrating *preliminary* directions for automated speaker evaluation and feedback.

The contributions of this work extend beyond empirical findings. The literature review has consolidated knowledge from human-computer interaction, PS pedagogy, and affective computing, providing a robust theoretical

foundation for VR-based training research. The development of the EVE corpus has contributed to emotional speech recognition, while the creation of a PS corpus has enabled data-driven analysis of performance assessment in VR. These corpora provide essential datasets for further advancements in speech analytics and interactive training systems.

Beyond its technical contributions, this research also addresses key challenges in management and business education, where data-driven decision-making and soft skills development are increasingly interlinked. As organisations turn to immersive technologies and learning analytics to train future professionals, this work illustrates how VR, combined with machine learning and multimodal analysis, can enhance communication skills in a measurable, scalable, and ethical way. It aligns with the growing trend in management science to integrate behavioural insights, digital tools, and interpretable AI systems to improve human performance and support evidence-based training strategies.

As part of this research, several VR environments were designed and implemented to support PST. These environments, depicted in the figure below (see Figure 9.1), include realistic settings such as a meeting room, an auditorium, a courtroom, a classroom, a boardroom, and an office.

As part of this thesis, three VR environments were developed in collaboration with other research groups to support specific research goals: the office environment (see Chapter III), the virtual courtroom (see Chapter IV), and the meeting room used in the ULiège corpus (see Chapter VII). Beyond the scope of this thesis, the *AR/VR SIG* lab was subsequently approached to design additional environments (see Figure 9.1) for other collaborative projects. This includes the boardroom to train manager, the classroom, created in partnership with the Faculty of Education, and the auditorium, developed for the Faculty of Political Sciences. All of them that are now part of another doctoral thesis. These ongoing developments reflect the broader applicability of VR for public speaking training across varied academic contexts. Each environment was developed to provide immersive and contextually relevant training experiences tailored to a range of professional and academic speaking scenarios. Moreover, these environments were refined and enhanced based on insights gained from the findings presented in this dissertation, ensuring their effectiveness in addressing key aspects of PS performance. Specifically, all environments were equipped with validated audience attitudes based on the findings from Chapter III, as well as the integrated note-taking and replay functionalities described in Chapter IV. The feedback modules were enriched to include the multimodal cues identified in Chapter V. The *AR/VR SIG lab* is currently working on integrating more advanced prediction models (both for performance and emotion recognition) by deepening machine learning techniques primarily in this thesis.

Despite its numerous advantages, the widespread adoption of ana-



Figure 9.1: VR environment for Public Speaking Training

lytics and AI presents several challenges, including data privacy concerns (e.g., protecting sensitive audio-visual and behavioural data collected during training), algorithmic bias, and the need for transparency in decision-making. The increasing reliance on automated systems necessitates rigorous validation of machine learning models to ensure fairness and reliability, particularly in high-impact applications like PST.

This research aligns with the principles outlined in the AI Act, which emphasises transparency, fairness, and accountability in AI-driven systems [237]. By integrating explainability into every analytical process, this study ensures that users and practitioners can understand how performance evaluations are generated, fostering trust in AI-assisted training. Furthermore, insights from this dissertation have contributed to refining AI-driven feedback systems, enhancing the interpretability and ethical application of automated PS assessments.

Reflecting on six years of development, this project has grown from an initial concept into a functional and promising VR-based training system. The work has progressed from theoretical exploration to the implementation and validation of multimodal analytics, culminating in a comprehensive framework for PS assessment. The refinement of AI-driven models, the creation of corpora, and the execution of user studies have shaped a tool that aims to provide meaningful feedback and support skill development in an immersive and controlled environment.

As this research moves forward, the focus should be on enhancing adaptive learning mechanisms, incorporating real-time physiological signals to deepen insights into speaker engagement, and expanding datasets to ensure the system's applicability across diverse speaking contexts. Another critical challenge lies in simulating realistic audience reactions in VR, requiring advanced AI models, robust datasets, and carefully designed virtual agents that respond dynamically to speaker performance. Furthermore,

ensuring that feedback remains both interpretable and actionable is also crucial for user engagement and learning outcomes.

Additionally, bias in data presents a significant challenge [62]. Stereotypes related to gender, culture, or professional hierarchies can shape both AI-generated feedback and virtual audience reactions [364, 124], potentially reinforcing societal inequalities [370]. This, in turn, can introduce further bias into human interactions and the design of virtual agents, affecting the authenticity and fairness of the training experience. To mitigate these biases, it is essential to diversify training data, refine bias-detection mechanisms, and continuously validate both AI models and virtual agent behaviours against fairness benchmarks.

Beyond its academic scope, this thesis contributes to managerial practice by illustrating how immersive technologies can support strategic skills development in organisations. The proposed system addresses real-world challenges such as scalable soft-skills training, standardised performance evaluation, and data-informed coaching. By enabling structured, explainable, and ethical feedback, it provides a model for integrating AI-driven tools into leadership development, employee assessment, and communication-intensive roles, where effective speaking directly impacts managerial decision-making and organisational success.

In conclusion, this thesis has demonstrated the potential of VR, AI, and multimodal analytics to transform PST by offering structured and objective feedback in a virtual setting. By embedding fairness-driven methodologies into AI development, virtual agent design, and interaction modelling, this research contributes to the creation of ethical and inclusive VR-based training systems. With further advancements, these systems will continue to evolve, providing increasingly effective, fair, and interpretable solutions for skill development in PS.

Bibliography

- [1] Friday Joseph Agbo et al. "Examining the relationships between students' perceptions of technology, pedagogy, and cognition: the case of immersive virtual reality mini games to foster computational thinking in higher education". In: *Smart Learning Environments* 10.1 (2023), p. 16.
- [2] Takuya Akiba et al. "Optuna: A next-generation hyperparameter optimization framework". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 2623–2631.
- [3] Pease Allan. *Body Language, How to read others' thoughts by their gestures*. 1995.
- [4] John A Allen, Robert T Hays, and Louis C Buffardi. "Maintenance training simulator fidelity and individual differences in transfer of training". In: *Human factors* 28.5 (1986), pp. 497–509.
- [5] Ibis M Álvarez et al. "Immersive Virtual Reality to improve competence to manage classroom climate in secondary schools". In: *Educación XX1* 26.1 (2023), pp. 249–272.
- [6] Ashfaq Amin et al. "Immersion in cardboard VR compared to a traditional head-mounted display". In: *International Conference on Virtual, Augmented and Mixed Reality*. Springer. 2016, pp. 269–276.
- [7] Elisabeth André et al. "Integrating Models of Personality and Emotions into Lifelike Characters". In: *Affective Interactions*. IWA1 1999. Springer, 1999, pp. 150–165.
- [8] Rigmor Argren. "Teaching Law of Armed Conflict with Virtual Reality". In: *Teaching International Law*. Routledge, 2024, pp. 166–182.
- [9] Michael Argyle and Roger Ingham. "Gaze, mutual gaze, and proximity". In: *Semiotica* 6.1 (1972), pp. 32–49.
- [10] Vincent Aubanel et al. "The Fharvard corpus: A phonemically-balanced French sentence resource for audiology and intelligibility research". In: *Speech Communication* 124 (2020), pp. 68–74.

- [11] Barry J Babin and Jill S Attaway. "Atmospheric affect as a tool for creating value and gaining share of customer". In: *Journal of Business Research* 49.2 (2000), pp. 91–99.
- [12] Manuel Bachmann et al. "Virtual reality public speaking training: effectiveness and user technology acceptance". In: *Frontiers in virtual reality* 4 (2023), p. 1242544.
- [13] Jeremy N Bailenson et al. "The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments". In: *Presence* 14.4 (2005), pp. 379–393.
- [14] Tanja Bänziger, Hannes Pirker, and K Scherer. "GEMEP-Geneva Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions". In: *Proceedings of LREC*. Vol. 6. 2006, pp. 15–019.
- [15] Alisa Barkar et al. "Insights Into the Importance of Linguistic Textual Features on the Persuasiveness of Public Speaking". In: *Companion Publication of the 25th International Conference on Multimodal Interaction*. 2023, pp. 51–55.
- [16] Lisa Feldman Barrett et al. "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements". In: *Psychological science in the public interest* 20.1 (2019), pp. 1–68.
- [17] Pierre Barrouillet et al. "Time and cognitive load in working memory." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33.3 (2007), p. 570.
- [18] Christoph Bartneck et al. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International Journal of Social Robotics* 1.1 (2009), pp. 71–81. ISSN: 18754791.
- [19] Karen Barton and Paul Maharg. "E-Simulations in the Wild: Interdisciplinary Research, Design and Implementation". In: *Games and simulations in online learning: Research and development frameworks*. IGI Global, 2007, pp. 115–149.
- [20] Ligia Batrinca et al. "Cicero-towards a multimodal virtual audience platform for public speaking training". In: *International workshop on intelligent virtual agents*. Springer. 2013, pp. 116–128.
- [21] Tobias Baur et al. "Nova: Automated analysis of nonverbal signals in social interactions". In: *Human Behavior Understanding: 4th International Workshop, HBU 2013, Barcelona, Spain, October 22, 2013. Proceedings* 4. Springer. 2013, pp. 160–171.

- [22] Teresa Rivera Bean. "Learning beyond the Classroom: The Case for Establishing an Undergraduate Pre-Law Clinic". In: *U. Balt. LF* 52 (2021), p. 147.
- [23] Geoffrey W Beattie. "Planning units in spontaneous speech: Some evidence from hesitation in speech and speaker gaze direction in conversation". In: *Linguistics* (1979).
- [24] Julia Beck, Mattia Rainoldi, and Roman Egger. "Virtual reality in tourism: A state-of-the-art review". In: *Tourism Review* (2019).
- [25] Nikolaus Bee et al. "Bossy or wimpy: Expressing social dominance by combining gaze and linguistic behaviors". In: *10th International Conference on Intelligent Virtual Agents*. IVA 2010. Springer, 2010, pp. 265–271.
- [26] Daniel Belanche et al. "Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The Humanness-Value-Loyalty model". In: *Psychology & Marketing* 38.12 (2021), pp. 2357–2376.
- [27] Lyndsey Bengtsson. "The Law in the Community Model of Clinical Legal Education: Assessing the Impact on Key Stakeholders". In: *Int'l J. Clinical Legal Educ.* 30 (2023), p. 54.
- [28] Courtney L Benjamin et al. "Patterns and predictors of subjective units of distress in anxious youth". In: *Behavioural and cognitive psychotherapy* 38.4 (2010), pp. 497–504.
- [29] Martine J van Bennekom, Pelle P de Koning, and Damiaan Denys. "Virtual reality objectifies the diagnosis of psychiatric disorders: a literature review". In: *Frontiers in Psychiatry* 8 (2017), p. 163.
- [30] Beatrice Biancardi, Mathieu Chollet, and Chloé Clavel. "Introducing the 3MT_French Dataset to investigate the timing of public speaking judgements". In: *Language Resources and Evaluation* (2024), pp. 1–20.
- [31] Béatrice Biancardi et al. "A Computational Model for Managing Impressions of an Embodied Conversational Agent in Real-Time". In: *8th International Conference on Affective Computing and Intelligent Interaction*. ACII 2019. Cambridge, UK: Institute of Electrical and Electronics Engineers Inc., 2019, pp. 234–240. ISBN: 978-1-72813-888-6.
- [32] Anke W Blöte et al. "The relation between public speaking anxiety and social anxiety: A review". In: *Journal of anxiety disorders* 23.3 (2009), pp. 305–313.
- [33] Josien Boetje and Stan van Ginkel. "The added benefit of an extra practice session in virtual reality on the development of presentation skills: A randomized control trial". In: *Journal of Computer Assisted Learning* 37.1 (2021), pp. 253–264.

BIBLIOGRAPHY

- [34] Francesca Bonetti, Gary Warnaby, and Lee Quinn. "Augmented reality and virtual reality in physical and online retailing: A review, synthesis and research agenda". In: *Augmented reality and virtual reality* (2018), pp. 119–132.
- [35] Benjamin Stephanus Botha, Lizette de Wet, and Yvonne Botma. "Undergraduate nursing student experiences in using immersive virtual reality to manage a patient with a foreign object in the right lung". In: *Clinical Simulation in Nursing* 56 (2021), pp. 76–83.
- [36] S. Bouchard and G. Robillard. "Validation canadienne-française du Gatineau Presence Questionnaire auprès d'adultes immergés en réalité virtuelle". In: *87e Congrès de l'ACFAS, Québec, mai 2019* (May 2019).
- [37] S. Bouchard et al. "Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial". In: *The British Journal of Psychiatry* 210.4 (2017), pp. 276–283.
- [38] Robin Bowley. "Enabling law students to understand business concepts: reflections on developing a business case study for corporate law". In: *The Law Teacher* 54.2 (2020), pp. 169–193.
- [39] D. Eric Boyd and Bernadett Koles. "An Introduction to the Special Issue "Virtual Reality in Marketing": Definition, Theory and Practice". In: *Journal of Business Research* 100 (2019), pp. 441–444. ISSN: 0148-2963.
- [40] Ryan L Boyd et al. "The development and psychometric properties of LIWC-22". In: *Austin, TX: University of Texas at Austin* 10 (2022).
- [41] MARGARET M BRADLEY. "Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavioral Therapy and Experimental". In: *Psychiatry* 25 (1994), pp. 49–59.
- [42] Michael K Brady, Christopher J Robertson, and J Joseph Cronin. "Managing behavioral intentions in diverse cultural environments: An investigation of service quality, service value, and satisfaction for American and Ecuadorian fast-food customers". In: *Journal of International Management* 7.2 (2001), pp. 129–149.
- [43] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2 (2006), pp. 77–101.
- [44] Caterina Breitenstein, Diana Van Lancker, and Irene Daum. "The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample". In: *Cognition & Emotion* 15.1 (2001), pp. 57–79.
- [45] Erik Brynjolfsson, Lorin M Hitt, and Heekyung Hellen Kim. "Strength in numbers: How does data-driven decisionmaking affect firm performance?" In: *SSRN 1819486* (2011).

- [46] Tony W Buchanan, Jacqueline S Laures-Gore, and Melissa C Duff. "Acute stress reduces speech fluency". In: *Biological psychology* 97 (2014), pp. 60–66.
- [47] Carlos Busso et al. "IEMOCAP: interactive emotional dyadic motion capture database". In: *Lang Resources & Evaluation* 42.4 (), pp. 335–359. ISSN: 1574-020X, 1574-0218.
- [48] Asad H Butt et al. "Let's play: Me and my AI-powered avatar as one team". In: *Psychology & Marketing* 38.6 (2021), pp. 1014–1025.
- [49] Rebecca Byrnes and Peter Lawrence. "Bringing diplomacy into the classroom: Stimulating student engagement through a simulated treaty negotiation". In: *Legal Education Review* 26.1/2 (2016), pp. 19–45.
- [50] Angelo Cafaro, Hannes Högni Vilhjálmsson, and Timothy Bickmore. "First impressions in human-agent virtual encounters". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 23.4 (2016), pp. 1–40.
- [51] Zoraida Callejas et al. "A Computational model of Social Attitudes for a Virtual Recruiter". In: *13th International Conference on Autonomous Agents and Multiagent Systems*. AAMAS 2014. Paris, France: International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 93–100.
- [52] Estelle Campione and Jean Véronis. "A large-scale multilingual study of silent pause duration". In: *Speech prosody 2002, international conference*. 2002.
- [53] Houwei Cao et al. "Crema-d: Crowd-sourced emotional multimodal actors dataset". In: *IEEE transactions on affective computing* 5.4 (2014), pp. 377–390.
- [54] Adelia Carstens. "Advice on the use of gestures in presentation skills manuals: alignment between theory, research and instruction". In: *Image & Text* 33 (2019), pp. 1–34.
- [55] Traci Carte and Laku Chidambaram. "A capabilities-based theory of technology deployment in diverse teams: Leapfrogging the pitfalls of diversity and leveraging its potential with collaborative technology". In: *Journal of the Association for Information Systems* 5.11 (2004), p. 4.
- [56] James L Cavallaro and Meghna Sridhar. "Reducing bias in human rights fact-finding: The potential of the clinical simulation model to overcome ethical, practical, and cultural tensions in" foreign" contexts". In: *Human Rights Quarterly* 42.2 (2020), pp. 488–512.
- [57] Gavin C Cawley and Nicola LC Talbot. "Fast exact leave-one-out cross-validation of sparse least-squares support vector machines". In: *Neural networks* 17.10 (2004), pp. 1467–1475.

- [58] Nitesh V Chawla et al. "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16 (2002), pp. 321–357.
- [59] Lei Chen et al. "Utilizing multimodal cues to automatically evaluate public speaking performance". In: *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE. 2015, pp. 394–400.
- [60] Afef Cherni, Roxane Bertrand, and Magalie Ochs. "From neutral human face to persuasive virtual face, a new automatic tool to generate a persuasive attitude". In: *Advances in Signal Processing and Artificial Intelligence (ASPAI 2022)*. 2022.
- [61] Nikolaos-Kosmas Chlis. "Comparison of statistical methods for genomic signature extraction". In: *Tech. Univ. Crete, Chania, Greece, Tech. Rep* (2013).
- [62] Kristy Choi et al. "Fair generative modeling via weak supervision". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 1887–1898.
- [63] M. Chollet and S. Scherer. "Perception of virtual audiences". In: *IEEE computer graphics and applications* 37.4 (2017), pp. 50–59.
- [64] M. Chollet et al. "Exploring feedback strategies to improve public speaking: an interactive virtual audience framework". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2015, pp. 1143–1154.
- [65] Mathieu Chollet, Stacy Marsella, and Stefan Scherer. "Training public speaking with virtual social interactions: effectiveness of real-time feedback and delayed feedback". In: *Journal on Multimodal User Interfaces* (2021), pp. 1–13.
- [66] Mathieu Chollet, Talie Massachi, and Stefan Scherer. "Investigating the Physiological Responses to Virtual Audience Behavioral Changes A Stress-Aware Audience for Public Speaking Training". In: *IVA 2017 Workshop on Physiologically-Aware Virtual Agents*. 2016.
- [67] Mathieu Chollet and Stefan Scherer. "Perception of virtual audiences". In: *IEEE Computer Graphics and Applications* 37.4 (2017), pp. 50–59. ISSN: 0272-1716.
- [68] Mathieu Chollet et al. "A multimodal corpus for the assessment of public speaking ability and anxiety". In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 2016, pp. 488–495.

- [69] Mathieu Chollet et al. "Influence of individual differences when training public speaking with virtual audiences". In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 2018, pp. 1–7.
- [70] Mathieu Chollet et al. "Public speaking training with a multimodal interactive virtual audience framework". In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 2015, pp. 367–368.
- [71] Bianca Malaika Ciuffani. "Non-verbal Communication and Leadership: the impact of hand gestures used by leaders on follower job satisfaction". B.S. thesis. University of Twente, 2017.
- [72] Drama Classes and Performing Arts School. *Meisner Technique*. Accessed: 31-Jan-2025. 2025. URL: <https://www.%20dramaclasses.biz/meisner-technique>.
- [73] Michael A Cohn, Matthias R Mehl, and James W Pennebaker. "Linguistic markers of psychological change surrounding September 11, 2001". In: *Psychological science* 15.10 (2004), pp. 687–693.
- [74] Julie Collange and Jérôme Guegan. "Using virtual reality to induce gratitude through virtual social interaction". In: *Computers in Human Behavior* 113 (2020), p. 106473.
- [75] Karl S Coplan. "Teaching Substantive Environmental Law and Practice Skills Through Interest Group Role-Playing". In: *Vermont Journal of Environmental Law* 18.2 (2016), pp. 194–221.
- [76] Teresa Nesbitt Cosby. "To the head of the class: Quantifying the relationship between participation in undergraduate mock trial programs and student performance in law school". In: *John's L. Rev.* 92 (2018), p. 797.
- [77] Robert Courtois et al. "Validation française du Big Five Inventory à 10 items (BFI-10)". In: *L'Encéphale* 46.6 (2020), pp. 455–462.
- [78] Nelson Cowan. "The magical number 4 in short-term memory: A reconsideration of mental storage capacity". In: *Behavioral and brain sciences* 24.1 (2001), pp. 87–114.
- [79] Alan S Cowen et al. "The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures". In: *Nature human behaviour* 3.4 (2019), pp. 369–382.
- [80] Michelle G Craske et al. "Maximizing exposure therapy: An inhibitory learning approach". In: *Behaviour research and therapy* 58 (2014), pp. 10–23.

- [81] Carolina Cruz-Neira et al. "The CAVE: audio visual experience automatic virtual environment". In: *Communications of the ACM* 35.6 (1992), pp. 64–72.
- [82] Edgar Dale. "Audiovisual methods in teaching". In: (1969).
- [83] J Dalton and J Gillham. "Seeing is believing: How virtual reality and augmented reality are transforming business and the economy". In: *PWC, techreport* (2019).
- [84] Yvonne Marie Daly and Noelle Higgins. "The place and efficacy of simulations in legal education: A preliminary examination". In: *All Ireland Journal of Higher Education* 3.2 (2011).
- [85] Ionut Damian et al. "Individualized agent interactions". In: *4th International Conference on Motion in Games*. MIG 2011. 2011, pp. 15–26.
- [86] Sara De Freitas and Tim Neumann. "The use of 'exploratory learning' for supporting immersive learning in virtual environments". In: *Computers & Education* 52.2 (2009), pp. 343–352.
- [87] Arne De Keyser et al. "Frontline service technology infusion: Conceptual archetypes and future research directions". In: *Journal of Service Management* 30.1 (2019), pp. 156–183.
- [88] Nick Degens et al. "'What I see is not what you get': why culture-specific behaviours for virtual characters should be user-tested across cultures". In: *AI and Society* 32.1 (2017), pp. 37–49. ISSN: 0951-5666.
- [89] Deloitte Insights. *The Future of Learning: Unleashing the Potential of VR*. Accessed: 2025-03-22. 2023. URL: <https://action.deloitte.com/insight/3844/the-future-of-learning-unleashing-the-potential-of-vr>.
- [90] Guillaume Demary et al. "How do Leaders Perceive Stress and Followership from Nonverbal Behaviors Displayed by Virtual Followers?" In: *19th ACM International Conference on Intelligent Virtual Agents*. IVA 2019. Paris, France: Association for Computing Machinery, 2019, pp. 56–61. ISBN: 978-145036672-4.
- [91] Xiaoyan Deng, H Rao Unnava, and Hyojin Lee. "'Too true to be good?' when virtual reality decreases interest in actual reality". In: *Journal of Business Research* 100 (2019), pp. 561–570.
- [92] Soumia Dermouche and Catherine Pelachaud. "Attitude modeling for virtual character based on temporal sequence mining: Extraction and evaluation". In: *5th International Conference on Movement and Computing*. MOCO 2018. Genoa, Italy: Association for Computing Machinery, 2018. ISBN: 978-1-4503-6504-8.

- [93] Soumia Dermouche and Catherine Pelachaud. "Leveraging the Dynamics of Non-Verbal Behaviors for Social Attitude Modeling". In: *IEEE Transactions on Affective Computing* 13.2 (2022), pp. 1072–1085. ISSN: 1949-3045.
- [94] Michael Detyna and Margaret Kadiri. "Virtual reality in the HE classroom: feasibility, and the potential to embed in the curriculum". In: *Journal of Geography in Higher Education* 44.3 (2020), pp. 474–485.
- [95] Anna Flavia Di Natale et al. "Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research". In: *Computers in Human Behavior Reports* (2023), p. 100288.
- [96] Tanvi Dinkar et al. "How confident are you? Exploring the role of fillers in the automatic prediction of a speaker's confidence". In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2020, pp. 8104–8108.
- [97] Tahirou Djara, Abdoul Matine Ousmane, and Antoine Vianou. "Mood and personality influence on emotion". In: *2nd International EAI Conference on Emerging Technologies for Developing Countries*. AFRICATEK 2018. Cotonou, Benin: Springer Verlag, 2019, pp. 166–174.
- [98] Emanuel van Dongen. "Pleading in the Virtual Courtroom: Exploring Experiential Learning in Law through Virtual-Reality-Based Exercises and Student Feedback". In: *European Journal of Legal Education* 5.1 (2024), pp. 157–190.
- [99] Kariane Pereira Dos Santos et al. "Does shyness influence the self-perception of vocal symptoms, public speaking, and daily communication?" In: *Journal of voice* 36.1 (2022), pp. 54–58.
- [100] Kathy Douglas. "The role of ADR in developing lawyers' practice: lessons from Australian legal education". In: *Legal Education at the Crossroads*. Routledge, 2018, pp. 69–84.
- [101] Danielle Duez. "Acoustico-phonetic characteristics of filled pauses in spontaneous French speech: preliminary results". In: *Proc. DiSS 2001*. 2001, pp. 41–44.
- [102] Yogesh K Dwivedi et al. "Metaverse marketing: How the metaverse will shape the future of consumer research and practice". In: *Psychology & Marketing* 40.4 (2023), pp. 750–776.
- [103] Daniel Eckert and Andrea Mower. "The effectiveness of virtual reality soft skills training in the enterprise: a study". In: *PwC Public Report* (2020). URL: <https://www.5discovery.com/wp-content/uploads/2020/09/pwc-understanding-the-effectiveness-of-soft-skills-training-in-the-enterprise-a-study.pdf>.

BIBLIOGRAPHY

- [104] D. Efron. *Gesture, Race and Culture: A Tentative Study of the Spatio-temporal and "linguistic" Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living Under Similar as Well as Different Environmental Conditions*. Approaches to semiotics. Mouton, 1972.
- [105] Kathleen M Eisenhardt. "Building theories from case study research". In: *Academy of management review* 14.4 (1989), pp. 532–550.
- [106] Paul Ekman and Wallace V Friesen. *Facial action coding system: Investigator's guide*. Consulting Psychologists Press, 1978.
- [107] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. *Emotion in the human face: Guidelines for research and an integration of findings*. Vol. 11. Elsevier, 2013.
- [108] Tigran W Eldred. "Insights from psychology: Teaching behavioral legal ethics as a core element of professional responsibility". In: *Mich. St. L. Rev.* (2016), p. 757.
- [109] Merouane Elazami Elhassani et al. "Deep Learning concepts for genomics: an overview". In: *EMBnet. journal* 27 (2022), e990.
- [110] James WB Elsey et al. "Reconsolidation-based treatment for fear of public speaking: a systematic pilot study using propranolol". In: *Translational Psychiatry* 10.1 (2020), p. 179.
- [111] Elodie Etienne et al. "Perception of avatars nonverbal behaviors in virtual reality". In: *Psychology & Marketing* 40.11 (2023), pp. 2464–2481.
- [112] European Commission. *Virtual Worlds (Web4.0) - European Commission*. Accessed: 2025-03-22. 2023. URL: <https://digital-strategy.ec.europa.eu/en/policies/virtual-worlds>.
- [113] Florian Eyben et al. "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing". In: *IEEE transactions on affective computing* 7.2 (2015), pp. 190–202.
- [114] Andrew M Farrell, Anne L Souchon, and Geoffrey R Durden. "Service encounter conceptualisation: Employees' service behaviours and customers' service quality perceptions". In: *Journal of Marketing Management* 17.5-6 (2001), pp. 577–593.
- [115] Wendy A Farrell. "Learning becomes doing: Applying augmented and virtual reality to improve performance". In: *Performance Improvement* 57.4 (2018), pp. 19–28.
- [116] Ronán Feehily. "Problem-based learning and international commercial dispute resolution in the Indian Ocean". In: *The Law Teacher* 52.1 (2018), pp. 17–37.

- [117] William M Felton and Russell E Jackson. "Presence: A review". In: *International Journal of Human-Computer Interaction* 38.1 (2022), pp. 1–18.
- [118] Susan T. Fiske, Amy J.C. Cuddy, and Peter Glick. "Universal dimensions of social cognition: warmth and competence". In: *Trends in Cognitive Sciences* 11.2 (2007), pp. 77–83. ISSN: 1364-6613.
- [119] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. "The impact of virtual, augmented and mixed reality technologies on the customer experience". In: *Journal of business research* 100 (2019), pp. 547–560.
- [120] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. "The influence of scent on virtual reality experiences: the role of aroma-content congruence". In: *Journal of Business Research* 123 (2021), pp. 289–301.
- [121] Sofia Fountoukidou et al. "Effects of a Virtual Model's Pitch and Speech Rate on Affective and Cognitive Learning". In: *14th International Conference on Persuasive Technology*. PERSUASIVE 2019. Limassol, Cyprus: Springer Verlag, 2019, pp. 16–27.
- [122] Sofia Fountoukidou et al. "Effects of a virtual model's pitch and speech rate on affective and cognitive learning". In: *Persuasive Technology: Development of Persuasive and Behavior Change Support Systems: 14th International Conference, PERSUASIVE 2019, Limassol, Cyprus, April 9–11, 2019, Proceedings 14*. Springer. 2019, pp. 16–27.
- [123] Katherine Franceschi et al. "Engaging group e-learning in virtual worlds". In: *Journal of Management Information Systems* 26.1 (2009), pp. 73–100.
- [124] Eric Frankel and Edward Vendrow. "Fair generation through prior modification". In: *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*. 2020.
- [125] Marlena R Fraune et al. "Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact". In: *Computers in Human Behavior* 105 (2020), p. 106220.
- [126] Daniel Freeman et al. "Virtual reality in the assessment, understanding, and treatment of mental health disorders". In: *Psychological Medicine* 47.14 (2017), pp. 2393–2400.
- [127] Brett Freudenberg and Anna Mortimore. "The firm: Re-thinking tutorials to provide greater authenticity for future tax professionals". In: *Journal of Australian Taxation* 21.1 (2019), pp. 53–73.
- [128] Steffi Frigo. "The relationship between acted and naturalistic emotional corpora". In: *Workshop" Corpora for research on emotion and affect"*. *5th International Conference on Language Resources and Evaluation (LREC'2006)*. 2006, pp. 34–36.

- [129] Jennifer Fromm, Stefan Stieglitz, and Milad Mirbabaie. "Virtual Reality in Digital Education: An Affordance Network Perspective on Effective Use Behavior". In: *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 55.2 (2024), pp. 14–41.
- [130] Enora Gabory and Mathieu Chollet. "Investigating the Influence of Sound Design for Inducing Anxiety in Virtual Public Speaking". In: *Companion Publication of the 2020 International Conference on Multimodal Interaction*. 2020, pp. 492–496.
- [131] Janel Gauthier and Stéphane Bouchard. "Adaptation canadienne-française de la forme révisée du State-Trait Anxiety Inventory de Spielberg". In: *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 25.4 (1993), p. 559.
- [132] Patrick Gebhard et al. "Designing the Impression of Social Agents' Real-time Interruption Handling". In: *19th ACM International Conference on Intelligent Virtual Agents*. IVA 2019. Paris, France: Association for Computing Machinery, 2019, pp. 19–21. ISBN: 978-1-4503-6672-4.
- [133] James H Geer, Laura A Estupinan, and Gina M Manguno-Mire. "Empathy, social skills, and other relevant cognitive processes in rapists and child molesters". In: *Aggression and violent behavior* 5.1 (2000), pp. 99–126.
- [134] Alli Gerkman and Logan Cornett. "Foundations for practice: The whole lawyer and the character quotient". In: *AccessLex Institute Research Paper* 16-04 (2016).
- [135] Anith Khairunnisa Ghazali et al. "The usage of virtual reality in engineering education". In: *Cogent Education* 11.1 (2024), p. 2319441.
- [136] Cheryl L Giddens et al. "Vocal indices of stress: a review". In: *Journal of voice* 27.3 (2013), 390–e21.
- [137] Damian Gil et al. "Validity of average heart rate and energy expenditure in Polar OH1 and Verity Sense while self-paced running". In: *International Journal of Exercise Science: Conference Proceedings*. Vol. 14. 1. 2021, p. 27.
- [138] Howard Gilkinson. "Social fears as reported by students in college speech classes". In: *Communications Monographs* 9.1 (1942), pp. 141–160.
- [139] Sonia M Gipson Rankin. "Creating Lightbulb Moments: Developing Higher-Order Thinking in Family Law Classrooms Through Court Observations". In: *JL & Educ.* 51 (2022), p. 13.
- [140] Yann Glemarec et al. "Indifferent or Enthusiastic? Virtual Audiences Animation and Perception in Virtual Reality". In: *Frontiers in Virtual Reality* 2 (2021). ISSN: 2673-4192.

- [141] Y. Glémarec et al. "A Scalability Benchmark for a Virtual Audience Perception Model in Virtual Reality". In: *25th ACM Symposium on Virtual Reality Software and Technology*. 2019, pp. 1–1.
- [142] Yann Glémarec et al. "Controlling the stage: a high-level control system for virtual audiences in Virtual Reality". In: *Frontiers in Virtual Reality* 3 (2022), p. 876433.
- [143] Alexander M Goberman, Stephanie Hughes, and Todd Haydock. "Acoustic characteristics of public speaking: Anxiety and practice effects". In: *Speech communication* 53.6 (2011), pp. 867–876.
- [144] Philippe Gournay, Olivier Lahaie, and Roch Lefebvre. "A canadian french emotional speech dataset". In: *Proceedings of the 9th ACM Multimedia Systems Conference* (2018).
- [145] Jonathan Gratch et al. "Virtual rapport". In: *6th International Conference on Intelligent Virtual Agents*. IVA 2006. Springer Verlag, 2006, pp. 14–27. ISBN: 978-354037593-7.
- [146] Muriel A Hagensaars and Agnes van Minnen. "The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia". In: *Journal of Anxiety Disorders* 19.5 (2005), pp. 521–537.
- [147] Fasih Haider et al. "Presentation quality assessment using acoustic information and hand movements". In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2016, pp. 2812–2816.
- [148] Thomas Hainey et al. "A systematic literature review of games-based learning empirical evidence in primary education". In: *Computers & Education* 102 (2016), pp. 202–223.
- [149] David Hamilton et al. "Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design". In: *Journal of Computers in Education* 8.1 (2021), pp. 1–32.
- [150] Peter A. Hancock et al. "A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction". In: *Human Factors* 53.5 (2011), pp. 517–527.
- [151] Philip Hardie et al. "Nursing & Midwifery students' experience of immersive virtual reality storytelling: an evaluative study". In: *BMC nursing* 19.1 (2020), p. 78.
- [152] Shlomo Hareli, Mano Halhal, and Ursula Hess. "Dyadic dynamics: The impact of emotional responses to facial expressions on the perception of power". In: *Frontiers in psychology* 9 (2018), p. 1993.

BIBLIOGRAPHY

- [153] Sandra R Harris, Robert L Kemmerling, and Max M North. "Brief virtual reality therapy for public speaking anxiety". In: *Cyberpsychology & behavior* 5.6 (2002), pp. 543–550.
- [154] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.
- [155] Carrie Heater. "Being there: The subjective experience of presence." In: *Presence Teleoperators Virtual Environ.* 1.2 (1992), pp. 262–271.
- [156] Lorna Heaton. "Talking heads vs. virtual workspaces: A comparison of design across cultures". In: *Journal of Information Technology* 13.4 (1998), pp. 259–272.
- [157] Alexandre Heeren et al. "Self-report version of the Liebowitz Social Anxiety Scale: Psychometric properties of the French version." In: *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 44.2 (2012), p. 99.
- [158] Morton L Heilig. *Sensorama simulator*. US Patent 3,050,870. Aug. 1962.
- [159] Michael Henderson, Neil Selwyn, and Rachel Aston. "What works and why? Student perceptions of 'useful' digital technology in university teaching and learning". In: *Studies in higher education* 42.8 (2017), pp. 1567–1579.
- [160] Thorsten Hennig-Thurau et al. "Social interactions in the metaverse: Framework, initial evidence, and research roadmap". In: *Journal of the Academy of Marketing Science* (2022), pp. 1–25.
- [161] Rachel Herbst et al. "A virtual reality resident training curriculum on behavioral health anticipatory guidance: development and usability study". In: *JMIR Pediatrics and Parenting* 4.2 (2021), e29518.
- [162] Ursula Hess. "Now you see it, now you don't—the confusing case of confusion as an emotion: Commentary on Rozin and Cohen (2003)." In: (2003).
- [163] Tim Hilken et al. "Exploring the frontiers in reality-enhanced service communication: From augmented and virtual reality to neuro-enhanced reality". In: *Journal of Service Management* 33.4/5 (2022), pp. 657–674.
- [164] Rebecca Hincks. "Measuring liveliness in presentation speech." In: *INTERSPEECH*. 2005, pp. 765–768.
- [165] Li-Hsing Ho, Hung Sun, and Tsun-Hung Tsai. "Research on 3D painting in virtual reality to improve students' motivation of 3D animation learning". In: *Sustainability* 11.6 (2019), p. 1605.
- [166] Stefan G Hofmann et al. "Speech disturbances and gaze behavior during public speaking in subtypes of social phobia". In: *Journal of Anxiety Disorders* 11.6 (1997), pp. 573–585.

- [167] Willem K. B. Hofstee, Boele de Raad, and Lewis R. Goldberg. "Integration of the Big Five and Circumplex Approaches to Trait Structure". In: *Journal of Personality and Social Psychology* 63.1 (1992), pp. 146–163.
- [168] Steven K Homer. "From Langdell to Lab: The Opportunities and Challenges of Experiential Learning in the First Semester". In: *Mitchell Hamline L. Rev.* 48 (2022), p. 265.
- [169] Adineh Hosseinpanah and Nicole C. Krämer. "Lost in Interpretation? The Role of Culture on Rating the Emotional Nonverbal Behaviors of a Virtual Agent". In: *Cross-Cultural Design. Applications in Cultural Heritage, Tourism, Autonomous Vehicles, and Intelligent Agents*. Springer International Publishing, 2021, pp. 350–368. ISBN: 978-3-030-77080-8.
- [170] Adineh Hosseinpanah, Nicole C. Krämer, and Carolin Straßmann. "Empathy for Everyone? The Effect of Age When Evaluating a Virtual Agent". In: *Proceedings of the 6th International Conference on Human-Agent Interaction*. HAI '18. Southampton, United Kingdom: Association for Computing Machinery, 2018, pp. 184–190. ISBN: 9781450359535.
- [171] Matt C Howard. "A meta-analysis and systematic literature review of virtual reality rehabilitation programs". In: *Computers in Human Behavior* 70 (2017), pp. 317–327.
- [172] Matt C Howard, Melissa B Gutworth, and Rick R Jacobs. "A meta-analysis of virtual reality training programs". In: *Computers in Human Behavior* 121 (2021), p. 106808.
- [173] William T Howe and Ioana A Cionea. "Exploring the associations between debate participation, communication competence, communication apprehension, and argumentativeness with a global sample". In: *Argumentation and Advocacy* 57.2 (2021), pp. 103–122.
- [174] Jill Howieson and Shane Rogers. "Using the role-play at the lectern: developing "work-ready" and confident professionals". In: *The Law Teacher* 52.2 (2018), pp. 190–200.
- [175] Sarah Hudson et al. "With or without you? Interaction and immersion in a virtual reality experience". In: *Journal of Business Research* 100 (2019), pp. 459–468.
- [176] Martin Yongho Hyun and Robert Martin O'Keefe. "Virtual destination image: Testing a telepresence model". In: *Journal of Business Research* 65.1 (2012), pp. 29–35.
- [177] Nicole G Iannarone, Benjamin P Edwards, and Kevin Conboy. "Identifying and Teaching Non-Traditional Transactional Skills". In: *Transactions: Tenn. J. Bus. L.* 18 (2016), p. 381.

- [178] Philip Jackson and SJUoSG Haq. "Surrey audio-visual expressed emotion (savee) database". In: *University of Surrey: Guildford, UK* (2014).
- [179] Ana Javornik. "'It's an illusion, but it looks real!' Consumer affective, cognitive and behavioural responses to augmented reality applications". In: *Journal of Marketing Management* 32.9-10 (2016), pp. 987–1011.
- [180] Jason Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, 2015.
- [181] Lena Jingen Liang and Statia Elliot. "A systematic review of augmented reality tourism research: What is now and what is next?" In: *Tourism and Hospitality Research* 21.1 (2021), pp. 15–30.
- [182] Timothy Jung and Jeremy Dalton. *XR Case Studies*. Tech. rep. Springer, 2021.
- [183] Timothy Jung et al. "A virtual reality-supported intervention for pulmonary rehabilitation of patients with chronic obstructive pulmonary disease: mixed methods study". In: *Journal of medical Internet research* 22.7 (2020), e14178.
- [184] Ewa Kacewicz et al. "Pronoun use reflects standings in social hierarchies". In: *Journal of Language and Social Psychology* 33.2 (2014), pp. 125–143.
- [185] Mohammed Kadri et al. "IVAL: Immersive Virtual Anatomy Laboratory for enhancing medical education based on virtual reality and serious games, design, implementation, and evaluation". In: *Entertainment Computing* 49 (2024), p. 100624.
- [186] Smiti Kahlon, Philip Lindner, and Tine Nordgreen. "Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study". In: *Child and adolescent psychiatry and mental health* 13.1 (2019), p. 47.
- [187] Hyunjin Kang and Hye Kyung Kim. "My avatar and the affirmed self: Psychological and persuasive implications of avatar customization". In: *Computers in Human Behavior* 112 (2020), p. 106446.
- [188] N. Kang et al. "The design of virtual audiences: noticeable and recognizable behavioral styles". In: *Computers in Human Behavior* 55 (2016), pp. 680–694.
- [189] Ni Kang et al. "An expressive virtual audience with flexible behavioral styles". In: *IEEE Transactions on Affective Computing* 4.4 (2013), pp. 326–340.

- [190] Sin-Hwa Kang et al. "Towards building a virtual counselor: Modeling nonverbal behavior during intimate self-disclosure". In: *Proceedings of the 11th International Conference on Autonomous Agents and Multi-agent Systems*. Vol. 1. AAMAS 2012. Valencia, Spain: International Foundation for Autonomous Agents and Multiagent Systems, 2012, pp. 63–70. ISBN: 0981738117.
- [191] Alexandra D Kaplan et al. "The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis". In: *Human factors* 63.4 (2021), pp. 706–726.
- [192] Tero Karras et al. "Training generative adversarial networks with limited data". In: *Advances in neural information processing systems* 33 (2020), pp. 12104–12114.
- [193] Kengo Kato et al. "Radiography education with VR using head mounted display: proficiency evaluation by rubric method". In: *BMC Medical Education* 22.1 (2022), p. 579.
- [194] Sam Kavanagh et al. "A systematic review of virtual reality in education". In: *Themes in science and technology education* 10.2 (2017), pp. 85–119.
- [195] Leila Kerkeni et al. *French emotional speech database-oréau*. 2020.
- [196] Komal Khandelwal and Ashwani Kumar Upadhyay. "Virtual reality interventions in developing and managing human resources". In: *Human Resource Development International* 24.2 (2021), pp. 219–233.
- [197] Do Yuen Kim, Ha Kyung Lee, and Kyunghwa Chung. "Avatar-mediated experience in the metaverse: The impact of avatar realism on user-avatar relationship". In: *Journal of Retailing and Consumer Services* 73 (2023), p. 103382.
- [198] Hye-Geum Kim et al. "Stress and heart rate variability: a meta-analysis and review of the literature". In: *Psychiatry investigation* 15.3 (2018), p. 235.
- [199] Seo Young Kim, Bernd H Schmitt, and Nadia M Thalmann. "Eliza in the uncanny valley: Anthropomorphizing consumer robots increases their perceived warmth but decreases liking". In: *Marketing letters* 30.1 (2019), pp. 1–12.
- [200] Andrea Kleinsmith et al. "Understanding empathy training with virtual patients". In: *Computers in Human Behavior* 52 (2015), pp. 151–158.
- [201] Mark L Knapp, Judith A Hall, and Terrence G Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.
- [202] David A Kolb. *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

- [203] Bernadett Koles and Peter Nagy. "Digital object attachment". In: *Current Opinion in Psychology* 39 (2021), pp. 60–65.
- [204] Oswald D Kothgassner et al. "Habituation of salivary cortisol and cardiovascular reactivity to a repeated real-life and virtual reality Trier Social Stress Test". In: *Physiology & behavior* 242 (2021), p. 113618.
- [205] Oswald D Kothgassner et al. "Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment". In: *Computers in human behavior* 62 (2016), pp. 124–135.
- [206] Eva Krumhuber et al. "Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting". In: *Journal of Nonverbal Behavior* 33.1 (2009), pp. 1–15.
- [207] Peter Kullmann et al. "An Evaluation of Other-Avatar Facial Animation Methods for Social VR". In: *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI EA '23. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394222.
- [208] Yunna Kwan et al. "Development of a structured interview to explore interpersonal schema of older adults living alone based on autobiographical memory". In: *International Journal of Environmental Research and Public Health* 18.5 (2021), p. 2316.
- [209] M. Laforest et al. "Inducing an anxiety response using a contaminated Virtual environment: Validation of a therapeutic tool for obsessive-compulsive disorder". In: *Frontiers in ICT* 3 (2016), p. 18.
- [210] Joakim Laine et al. "Immersive virtual reality for complex skills training: content analysis of experienced challenges". In: *Virtual Reality* 28.1 (2024), p. 61.
- [211] Bart Larivière et al. "'Service Encounter 2.0': An investigation into the roles of technology, employees and customers". In: *Journal of Business Research* 79 (2017), pp. 238–246.
- [212] Petri Laukka et al. "In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech". In: *Journal of Nonverbal Behavior* 32 (2008), pp. 195–214.
- [213] Christofer Laurell et al. "Exploring barriers to adoption of Virtual Reality through Social Media Analytics and Machine Learning—An assessment of technology, network, price and trialability". In: *Journal of Business Research* 100 (2019), pp. 469–474.
- [214] Richard S Lazarus. "The study of psychological stress: A summary of theoretical formulations and experimental findings". In: *Anxiety and behavior* (1966).

- [215] Nicole Lazzeri et al. "The influence of dynamics and speech on understanding humanoid facial expressions". In: *International Journal of Advanced Robotic Systems* 15.4 (2018). ISSN: 1729–8806.
- [216] Jina Lee, Zhiyang Wang, and Stacy Marsella. "Evaluating models of speaker head nods for virtual agents". In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 1. AAMAS 2010. Toronto, Canada: International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 1257–1263. ISBN: 9780982657119.
- [217] Jane Lessiter et al. "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory". In: *Presence: Teleoperators & Virtual Environments* 10.3 (2001), pp. 282–297.
- [218] Kate Letheren et al. "Robots should be seen and not heard... sometimes: Anthropomorphism and AI service robot interactions". In: *Psychology & Marketing* 38.12 (2021), pp. 2393–2406.
- [219] Pierre-Henry Leveau and Sandra Camus. "Embodiment, immersion, and enjoyment in virtual reality marketing experiences". In: *Psychology & Marketing* 40.7 (2023), pp. 1329–1343.
- [220] Hui Liao and Aichia Chuang. "A multilevel investigation of factors influencing employee service performance and customer outcomes". In: *Academy of Management Journal* 47.1 (2004), pp. 41–58.
- [221] Silje Stangeland Lie et al. "Developing a virtual reality educational tool to stimulate emotions for learning: focus group study". In: *JMIR Formative Research* 7 (2023), e41829.
- [222] Alexander Linden and Jackie Fenn. "Understanding Gartner's hype cycles". In: *Strategic Analysis Report N° R-20-1971. Gartner, Inc* 88 (2003), p. 1423.
- [223] Philip Lindner et al. "Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial". In: *Journal of anxiety disorders* 61 (2019), pp. 45–54.
- [224] Steven R. Livingstone and Frank A. Russo. "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English". In: *PLOS ONE* 13.5 (2018), e0196391–e0196391.
- [225] Alison E Lloyd and Sherriff TK Luk. "Interaction behaviors leading to comfort in the service encounter". In: *Journal of Services marketing* 25.3 (2011), pp. 176–189.

- [226] Ryan Lohre et al. "Effectiveness of immersive virtual reality on orthopedic surgical skills and knowledge acquisition among senior surgical residents: a randomized clinical trial". In: *JAMA network open* 3.12 (2020), e2031217–e2031217.
- [227] Sandra Maria Correia Loureiro et al. "Understanding the use of Virtual Reality in Marketing: A text mining-based review". In: *Journal of Business Research* 100 (2019), pp. 514–530.
- [228] Birgit Lugin, Julian Frommel, and Elisabeth André. "Combining a data-driven and a theory-based approach to generate culture-dependent behaviours for virtual characters". In: *Advances in culturally - aware intelligent systems and in cross - cultural psychological studies*. Springer, 2017, pp. 111–142.
- [229] Birgit Lugin et al. "Culture-related differences in aspects of behavior for virtual characters across Germany and Japan". In: *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*. Vol. 1-3. AAMAS 2011. Taipei, Taiwan: Association for Computing Machinery, 2011, pp. 441–448. ISBN: 0982657161.
- [230] Jean-Luc Lugin et al. "Breaking bad behaviors: A new tool for learning classroom management using virtual reality". In: *Frontiers in ICT* 3 (2016), p. 26.
- [231] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions". In: *Advances in neural information processing systems* 30 (2017).
- [232] Aleksandr Luntz. "On estimation of characters obtained in statistical procedure of recognition". In: *Technicheskaya Kibernetika* (1969).
- [233] Andrew Lynch. "Why Do We Moot-Exploring the Role of Mooting in Legal Education". In: *Legal Educ. Rev.* 7 (1996), p. 67.
- [234] Howard Maclay and Charles E Osgood. "Hesitation phenomena in spontaneous English speech". In: *Word* 15.1 (1959), pp. 19–44.
- [235] Gordon A MacLeod. "Creative problem-solving—for lawyers?!" In: *Journal of Legal Education* 16.2 (1963), pp. 198–202.
- [236] Brian MacWhinney and Johannes Wagner. "Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository". In: *Gesprächsforschung: Online-Zeitschrift zur verbalen Interaktion* 11 (2010), p. 154.
- [237] Tambiama Madiega. *Artificial intelligence act*. 2021.
- [238] Paul Maharg, Emma Nicol, et al. "Simulation and technology in legal education: A systematic review". In: *Legal education: Simulation in theory and practice* (2014), pp. 17–42.

- [239] Paul Maharg, Martin Owen, et al. "Simulations, learning and the metaverse: changing cultures in legal education". In: *Journal of Information, Law, Technology* 1 (2007), pp. 1–28.
- [240] Ania A Majewska and Ethell Vereen. "Using immersive virtual reality in an online biology course". In: *Journal for STEM Education Research* 6.3 (2023), pp. 480–495.
- [241] G. Makransky, Thomas S Terkildsen, and Richard E Mayer. "Adding immersive virtual reality to a science lab simulation causes more presence but less learning". In: *Learning and instruction* 60 (2019), pp. 225–236.
- [242] Kerry T Manis and Danny Choi. "The virtual reality hardware acceptance model (VR-HAM): Extending and individuating the technology acceptance model (TAM) for virtual reality hardware". In: *Journal of Business Research* 100 (2019), pp. 503–513.
- [243] Vaia Maragkou et al. "Educational seismology through an immersive virtual reality game: design, development and pilot evaluation of user experience". In: *Education Sciences* 13.11 (2023), p. 1088.
- [244] Marcoux, Audrey and Tessier, Marie-Hélène and Jackson, Philip L. "Nonverbal Markers of Empathy in Virtual Healthcare Professionals". In: *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents. IVA '23*. Würzburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450399944.
- [245] Ruth Maria Stock, Ad de Jong, and Nicolas A Zacharias. "Frontline employees' innovative service behavior as key to customer loyalty: Insights into FLEs' resource gain spiral". In: *Journal of Product Innovation Management* 34.2 (2017), pp. 223–245.
- [246] Benjy Marks and Jacqueline Thomas. "Adoption of virtual reality technology in higher education: An evaluation of five teaching semesters in a purpose-designed laboratory". In: *Education and information technologies* 27.1 (2022), pp. 1287–1305.
- [247] Andreas Maroungkas et al. "Virtual reality in education: a review of learning theories, approaches and methodologies for the last decade". In: *Electronics* 12.13 (2023), p. 2832.
- [248] Fiona Martin and Margaret Connor. "Using blended learning to aid law and business students' understanding of taxation law problems". In: *Journal of the Australasian Tax Teachers Association* 12.1 (2017), pp. 53–76.
- [249] Sébastien Mateo. "A procedure for conduction of a successful literature review using the PRISMA method". In: *Kinesitherapie* 20.226 (2020), pp. 29–37.

BIBLIOGRAPHY

- [250] Philip M McCarthy and Scott Jarvis. "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment". In: *Behavior research methods* 42.2 (2010), pp. 381–392.
- [251] Robert R. McCrae and Oliver P John. "An introduction to the five-factor model and its applications". In: *Journal of Personality* 60.2 (1992), pp. 175–215.
- [252] James C McCroskey. "Measures of communication-bound anxiety". In: *Speech Monographs* 3 (1970), pp. 269–277.
- [253] Hugh McFaul and Elizabeth FitzGerald. "A realist evaluation of student use of a virtual reality smartphone application in undergraduate legal education". In: *British Journal of Educational Technology* 51.2 (2020), pp. 572–589.
- [254] David McNeill. *Language and gesture*. Vol. 2. Cambridge University Press, 2000.
- [255] Nicky McWilliam, Tracey Yeung, and Annabelle Green. "Law students' experiences in an experiential law and research program in Australia". In: *Legal Education Review* 28 (2018), pp. 1–23.
- [256] Kent E Menzel and Lori J Carrell. "The relationship between preparation and performance in public speaking". In: *Communication Education* 43.1 (1994), pp. 17–26.
- [257] Marissa J Metz and Lori E James. "Specific effects of the Trier Social Stress Test on speech fluency in young and older adults". In: *Aging, Neuropsychology, and Cognition* 26.4 (2019), pp. 558–576.
- [258] Fred Miao et al. "An emerging theory of avatar marketing". In: *Journal of Marketing* 86.1 (2022), pp. 67–90.
- [259] Patrick Mikalef et al. "Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment". In: *British journal of management* 30.2 (2019), pp. 272–298.
- [260] Paul Milgram and Fumio Kishino. "A taxonomy of mixed reality visual displays". In: *IEICE TRANSACTIONS on Information and Systems* 77.12 (1994), pp. 1321–1329.
- [261] Luke Moffett, Dug Cubie, and Andrew Godden. "Bringing the battlefield into the classroom: using video games to teach and assess international humanitarian law". In: *The Law Teacher* 51.4 (2017), pp. 499–514.
- [262] Noah Moonen et al. "Immersion or social presence? Investigating the effect of virtual reality immersive environments on sommelier learning experiences". In: *Journal of Wine Research* 35.2 (2024), pp. 101–118.

- [263] Shae D Morgan and Bailey LaPaugh. "Methodological Stimulus Considerations for Auditory Emotion Recognition Test Design". In: *Journal of Speech, Language, and Hearing Research* (2025), pp. 1–16.
- [264] F. Mostajeran et al. "The effects of virtual audience size on social anxiety during public speaking". In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2020, pp. 303–312.
- [265] Nellie Munin and Yael Efron. "Role-playing brings theory to life in a multicultural learning environment". In: *Journal of Legal Education* 66.2 (2017), pp. 309–331.
- [266] Jamie Murphy, Ulrike Gretzel, and Juho Pesonen. "Marketing robot services in hospitality and tourism: the role of anthropomorphism". In: *Journal of Travel & Tourism Marketing* 36.7 (2019), pp. 784–795.
- [267] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [268] Eric B Nash et al. "A review of presence and performance in virtual environments". In: *International Journal of human-computer Interaction* 12.1 (2000), pp. 1–41.
- [269] Karen Neville, Ciara Heavin, and Eoin Walsh. "A case in customizing e-learning". In: *Journal of Information Technology* 20.2 (2005), pp. 117–129.
- [270] Craig John Newbery-Jones. "Trying to do the right thing: experiential learning, e-learning and employability skills in modern legal education". In: *European Journal of Law and Technology* 6.1 (2015).
- [271] Matthew L Newman et al. "Lying words: Predicting deception from linguistic styles". In: *Personality and social psychology bulletin* 29.5 (2003), pp. 665–675.
- [272] Truong-Huy D. Nguyen et al. "Modeling warmth and competence in virtual characters". In: *Proceedings of the 15th International Conference on Intelligent Virtual Agents*. IVA '15. Delft, Netherlands, 2015, pp. 167–180.
- [273] Oliver Niebuhr, Radek Skarnitzl, and Lea Tylečková. "The acoustic fingerprint of a charismatic voice-Initial evidence from correlations between long-term spectral features and listener ratings". In: *9th International Conference on Speech Prosody 2018*. International Speech Communication Association (ISCA). 2018, pp. 359–363.
- [274] Michael Nixon, Steve DiPaola, and Ulysses Bernardet. "An Eye Gaze Model for Controlling the Display of Social Status in Believable Virtual Humans". In: *IEEE Conference on Computational Intelligence and Games*. CIG 2018. Maastricht, Netherlands: IEEE, 2018.

- [275] Magalie Ochs, Catherine Pelachaud, and Ken Prepin. "Social stances by virtual smiles". In: *14th International Workshop on Image Analysis for Multimedia Interactive Services*. WIAMIS 2013. Paris, France: IEEE, 2013, pp. 1–4.
- [276] Magalie Ochs et al. "REVITALISE: viRtual bEhaVioral skills TrAining for pubLIc SpEaking". In: *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*. 2024, pp. 1–3.
- [277] Magalie Ochs et al. "Training doctors' social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence". In: *Journal on Multimodal User Interfaces* 13 (2019), pp. 41–51.
- [278] Andrew Ortony. "Are all "basic emotions" emotions? A problem for the (basic) emotions construct". In: *Perspectives on psychological science* 17.1 (2022), pp. 41–61.
- [279] Carlos Orús, Sergio Ibáñez-Sánchez, and Carlos Flavián. "Enhancing the customer experience with virtual and augmented reality: The impact of content and device type". In: *International Journal of Hospitality Management* 98 (2021), p. 103019.
- [280] Fabrizio Palmas et al. "Acceptance and effectiveness of a virtual reality public speaking training". In: *2019 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE. 2019, pp. 363–371.
- [281] Fabrizio Palmas et al. "Virtual reality public speaking training: Experimental evaluation of direct feedback technology acceptance". In: *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2021, pp. 463–472.
- [282] Veronica S Pantelidis. "Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality". In: *Themes in science and technology education* 2.1-2 (2010), pp. 59–70.
- [283] Sunghyun Park et al. "Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach". In: *Proceedings of the 16th international conference on multimodal interaction*. 2014, pp. 50–57.
- [284] Sunghyun Park et al. "Multimodal analysis and prediction of persuasiveness in online social multimedia". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 6.3 (2016), pp. 1–25.
- [285] Dhaval Parmar et al. "Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents". In: *Autonomous Agents and Multi-Agent Systems* 36 (2022). ISSN: 1387-2532.

-
- [286] Paul Ekman Group. *Universal Emotions*. Accessed: 31-Jan-2025. 2025. URL: <https://www.paulekman.com/universal-emotions/>.
- [287] Shiva Pedram, Grace Kennedy, and Sal Sanzone. "Assessing the validity of VR as a training tool for medical students". In: *Virtual Reality* 28.1 (2024), p. 15.
- [288] James W Pennebaker et al. "When small words foretell academic success: The case of college admissions essays". In: *PloS one* 9.12 (2014), e115844.
- [289] Andrew Perkis et al. "QUALINET white paper on definitions of immersive media experience (IMEx)". In: *arXiv preprint arXiv:2007.07032* (2020).
- [290] Janaya Elizabeth Perron et al. "Resuscitating cardiopulmonary resuscitation training in a virtual reality: prospective interventional study". In: *Journal of medical Internet research* 23.7 (2021), e22920.
- [291] Patricia Perry. "Concept analysis: Confidence/self-confidence". In: *Nursing forum*. Vol. 46. 4. Wiley Online Library. 2011, pp. 218–230.
- [292] David-Paul Pertaub, Mel Slater, and Chris Barker. "An experiment on public speaking anxiety in response to three different types of virtual audience". In: *Presence* 11.1 (2002), pp. 68–78.
- [293] Edward Phillips. "Law games–role play and simulation in teaching legal application and practical skills: a case study". In: *Compass: Journal of Learning and Teaching in Higher Education* 3.5 (2012).
- [294] M. Kathleen Pichora-Fuller and Kate Dupuis. *Toronto emotional speech set (TESS)*. Version DRAFT VERSION. 2020.
- [295] Katarzyna Pisanski, Agata Groyecka-Bernard, and Piotr Sorokowski. "Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications". In: *Biology letters* 17.9 (2021), p. 20210356.
- [296] S. Poeschl. "Virtual reality training for public speaking — A QUEST-VR framework validation". In: *Frontiers in ICT* 4 (2017), p. 13.
- [297] C Alec Pollard and J Gibson Henderson. "Four types of social phobia in a community sample." In: *Journal of Nervous and Mental Disease* 176(7) (1988), pp. 440–445.
- [298] Delphine Potdevin, Celine Clavel, and Nicolas Sabouret. "A virtual tourist counselor expressing intimacy behaviors: A new perspective to create emotion in visitors and offer them a better user experience?" In: *International Journal of Human-Computer Studies* 150 (2021), p. 102612. ISSN: 1071-5819.

- [299] Delphine Potdevin, Céline Clavel, and Nicolas Sabouret. "Virtual intimacy in human-embodied conversational agent interactions: the influence of multimodality on its perception". In: *Journal on Multimodal User Interfaces* 15 (2021), pp. 25–43. ISSN: 1783-7677.
- [300] Ken Prepin, Magalie Ochs, and Catherine Pelachaud. "Beyond back-channels: co - construction of dyadic stance by reciprocal reinforcement of smiles between virtual agents". In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2013, pp. 1163–1168. ISBN: 978-0-9768318-9-1.
- [301] Astrid M. von der Pütten et al. "Quid pro quo? Reciprocal self-disclosure and communicative accomodation towards a virtual interviewer". In: *Proceedings of the 10th International Conference on Intelligent Virtual Agents*. IVA '11. Reykjavik, Iceland: Springer-Verlag, 2011, pp. 183–194. ISBN: 978-3-642-23974-8.
- [302] Jaziar Radianti et al. "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda". In: *Computers & education* 147 (2020), p. 103778.
- [303] Anat Rafaeli et al. "The future of frontline research: Invited commentaries". In: *Journal of Service Research* 20.1 (2017), pp. 91–99.
- [304] Eric D Ragan et al. "Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task". In: *IEEE transactions on visualization and computer graphics* 21.7 (2015), pp. 794–807.
- [305] Bina Rai, Hui Shin Tan, and Chen Huei Leo. "Bringing play back into the biology classroom with the use of gamified virtual lab simulations". In: *Journal of Applied Learning and Teaching* 2.2 (2019), pp. 48–55.
- [306] Vikram Ramanarayanan et al. "Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring". In: *Proceedings of the 2015 acm on international conference on multimodal interaction*. 2015, pp. 23–30.
- [307] Rui Raposo et al. "Increasing awareness and empathy among university students through immersive exercises—testing of the virtual reality application: A pilot study". In: *Medycyna Pracy. Workers' Health and Safety* 74.3 (2023), pp. 187–197.
- [308] Ishfaq Hussain Rather, Sushil Kumar, and Amir H Gandomi. "Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets". In: *Artificial Intelligence Review* 57.9 (2024), p. 226.

- [309] Philipp A Rauschnabel et al. "What is XR? Towards a framework for augmented and virtual reality". In: *Computers in human behavior* 133 (2022), p. 107289.
- [310] F Jerry Reen et al. "Developing student codesigned immersive virtual reality simulations for teaching of challenging concepts in molecular and cellular biology". In: *FEMS microbiology letters* 369.1 (2022), fnac051.
- [311] Angélique Remacle et al. "A virtual classroom can elicit teachers' speech characteristics: Evidence from acoustic measurements during in vivo and in virtuo lessons, compared to a free speech control situation". In: *Virtual Reality* 25 (2021), pp. 935–944.
- [312] Anne-Françoise Remacle, Simon Bouchard, and Dominique Mor-somme. "Can teaching simulations in a virtual classroom help trainee teachers to develop oral communication skills and self-efficacy? A randomized controlled trial". In: *Computers & Education* 200 (2023), p. 104810.
- [313] Samuel Ribeiro-Navarrete et al. "The effect of digitalization on business performance: An applied study of KIBS". In: *Journal of Business Research* 126 (2021), pp. 319–326.
- [314] Genevieve Robillard et al. "Validation canadienne-française de deux mesures importantes en réalité virtuelle: l'Immersive Tendancies Questionnaire et le Presence Questionnaire". In: *Poster presented at the 25e congrès annuel de la Société Québécoise pour la Recherche en Psychologie (SQRP), Trois-Rivières* (2002).
- [315] Stacey Robinson et al. "Frontline encounters of the AI kind: An evolved service encounter framework". In: *Journal of Business Research* 116 (2020), pp. 366–376.
- [316] Astrid M. Rosenthal-von der Pütten et al. "Dominant and submissive nonverbal behavior of virtual agents and its effects on evaluation and negotiation outcome in different age groups". In: *Computers in Human Behavior* 90 (2019), pp. 397–409. ISSN: 0747-5632.
- [317] EH Rothauser. "IEEE recommended practice for speech quality measurements". In: *IEEE Transactions on Audio and Electroacoustics* 17.3 (1969), pp. 225–246.
- [318] J Dan Rothwell. *In the company of others: An introduction to communication*. Oxford University Press New York, 2010.
- [319] James A. Russell and Albert Mehrabian. "Evidence for a three-factor theory of emotions". In: *Journal of Research in Personality* 11.3 (1977), pp. 273–294. ISSN: 0092-6566.

- [320] Connie Rust, William M Gentry, and Heath Ford. "Assessment of the effect of communication skills training on communication apprehension in first year pharmacy students—A two-year study". In: *Currents in Pharmacy Teaching and Learning* 12.2 (2020), pp. 142–146.
- [321] Grace Ryan et al. "Virtual reality in midwifery education: A mixed methods study to assess learning and understanding". In: *Nurse education today* 119 (2022), p. 105573.
- [322] Mohamad M Saab et al. "Incorporating virtual reality in nurse education: a qualitative study of nursing students' perspectives". In: *Nurse Education Today* 105 (2021), p. 105045.
- [323] Pejman Sajjadi et al. "On the Effect of a Personality-Driven ECA on Perceived Social Presence and Game Experience in VR". In: *10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. Würzburg, Germany: IEEE, 2018, pp. 1–8. ISBN: 978-1-5386-7123-8.
- [324] Sarah Saufnay, Elodie Etienne, and Michael Schyns. "Improvement of Public Speaking Skills using Virtual Reality: Development of a Training System". In: *12th International conference on affective computing and intelligent interaction (ACII)*. IEEE. Institute of Electrical and Electronics Engineers, New-York, United States. 2024.
- [325] Klaus R Scherer. "Vocal markers of emotion: Comparing induction and acting elicitation". In: *Computer Speech & Language* 27.1 (2013), pp. 40–58.
- [326] Klaus R Scherer, Harvey London, and Jared J Wolf. "The voice of confidence: Paralinguistic cues and audience evaluation". In: *Journal of Research in Personality* 7.1 (1973), pp. 31–44.
- [327] Sara Scheveneels et al. "Virtually unexpected: No role for expectancy violation in virtual reality exposure for public speaking anxiety". In: *Frontiers in psychology* 10 (2019), p. 2849.
- [328] Sheldon Schiffer. "How actors can animate game characters: integrating performance theory in the emotion model of a game character". In: *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*. Vol. 15. 1. 2019, pp. 227–229.
- [329] Jan Schneider et al. "Enhancing public speaking skills—an evaluation of the Presentation Trainer in the wild". In: *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, Lyon, France, September 13–16, 2016, Proceedings 11*. Springer. 2016, pp. 263–276.
- [330] Jan Schneider et al. "Presentation Trainer: what experts and computers can tell about your nonverbal communication". In: *Journal of Computer Assisted Learning* 33.2 (2017), pp. 164–177.

- [331] Martha Schneider et al. "Life events are associated with elevated heart rate and reduced heart complexity to acute psychological stress". In: *Biological Psychology* 163 (2021), p. 108116.
- [332] Claudia Schrader. "Creating avatars for technology usage: Context matters". In: *Computers in Human Behavior* 93 (2019), pp. 219–225.
- [333] Ulrike Schultze. "Embodiment and presence in virtual worlds: a review". In: *Journal of Information Technology* 25.4 (2010), pp. 434–449.
- [334] Valentin Schwind et al. "Using presence questionnaires in virtual reality". In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–12.
- [335] M. Schyns. *Artificial Neural Networks*. Lecture notes, Business Analytics, HEC Liège. 2025.
- [336] Matias N Selzer, Nicolas F Gazcon, and Martin L Larrea. "Effects of virtual presence and learning outcome using low-end virtual reality systems". In: *Displays* 59 (2019), pp. 9–15.
- [337] Christian Seufert et al. "Classroom management competency enhancement for student teachers using a fully immersive virtual classroom". In: *Computers & Education* 179 (2022), p. 104410.
- [338] Mike Seymour, Kai Riemer, and Judy Kay. "Actors, avatars and agents: Potentials and implications of natural face technology for the creation of realistic visual presence". In: *Journal of the association for Information Systems* 19.10 (2018), p. 4.
- [339] Mike Seymour et al. "Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments". In: *Journal of the Association for Information Systems* 22.3 (2021), p. 9.
- [340] Federica Sibilla and Tiziana Mancini. "I am (not) my avatar: A review of the user-avatar relationships in massively multiplayer online worlds". In: *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 12.3 (2018), article 4.
- [341] M. Slater. "A note on presence terminology". In: *Presence connect* 3.3 (2003), pp. 1–5.
- [342] Mel Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557.
- [343] Mel Slater and Maria V Sanchez-Vives. "Enhancing our lives with immersive virtual reality". In: *Frontiers in Robotics and AI* 3 (2016), p. 74.

BIBLIOGRAPHY

- [344] Mel Slater and Anthony Steed. "A virtual presence counter". In: *Presence* 9.5 (2000), pp. 413–434.
- [345] Shaun Smith and Joe Wheeler. *Managing the customer experience: Turning customers into advocates*. Pearson Education, 2002.
- [346] Gemma Smyth, Samantha Hale, and Neil Gold. "Clinical and experiential learning in Canadian law schools: Current perspectives". In: *Can. B. Rev.* 95 (2017), p. 151.
- [347] Michael R Solomon et al. "A role theory perspective on dyadic interactions: The service encounter". In: *Journal of Marketing* 49.1 (1985), pp. 99–111.
- [348] Sinan Sonlu, Ugur Gudukbay, and Funda Durupinar. "A Conversational Agent Framework with Multi-modal Personality Expression". In: *ACM Transactions on Graphics* 40.1 (2021), pp. 1–16. ISSN: 0730-0301.
- [349] Stephan Sonnenberg. "The Law Clinics at SNU School of Law: A Laboratory for Pedagogical Entrepreneurialism". In: *J. Korean L.* 20 (2021), p. 89.
- [350] Jacob H Steffen et al. "Framework of affordances for virtual reality and augmented reality". In: *Journal of Management Information Systems* 36.3 (2019), pp. 683–729.
- [351] Carolin Straßmann et al. "The effect of an intelligent virtual agent's nonverbal behavior with regard to dominance and cooperativity". In: *Proceedings of the 16th International Conference on Intelligent Virtual Agents*. IVA '16. Los Angeles, USA, 2016, pp. 15–28. ISBN: 978-3-319-47664-3.
- [352] William M Sullivan et al. *Educating lawyers: Preparation for the profession of law*. Vol. 2. John Wiley & Sons, 2007.
- [353] Dilip S Sundaram and Cynthia Webster. "The role of nonverbal communication in service encounters". In: *Journal of Services Marketing* 14.5 (2000), pp. 378–391.
- [354] Ivan E Sutherland et al. "The ultimate display". In: *Proceedings of the IFIP Congress*. Vol. 2. 506–508. New York. 1965, pp. 506–508.
- [355] John Sweller. "Cognitive load during problem solving: Effects on learning". In: *Cognitive science* 12.2 (1988), pp. 257–285.
- [356] Monica Taylor and Tamara Walsh. "Perceptions of competence and well-being in clinical legal education". In: *Australian Journal of Clinical Education* 3.1 (2018), pp. 1–19.
- [357] Shelley E Taylor et al. "Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight." In: *Psychological review* 107.3 (2000), p. 411.

- [358] Ann Thanaraj. "Evaluating the potential of virtual simulations to facilitate professional learning in law: a literature review". In: *World Journal of Education* 6.6 (2016), pp. 89–100.
- [359] Stella Ting-Toomey and Tenzin Dorjee. *Communicating across cultures*. Guilford Publications, 2018.
- [360] Ha Trinh et al. "Robocop: A robotic coach for oral presentations". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.2 (2017), pp. 1–24.
- [361] Art Tsang. "The relationship between tertiary-level students' self-perceived presentation delivery and public speaking anxiety: A mixed-methods study". In: *Assessment & Evaluation in Higher Education* 45.7 (2020), pp. 1060–1072.
- [362] Andy Unger et al. "Evaluating the Academic Benefits of Clinical Legal Education: An Analysis of the Final Average Marks of Five Cohorts of LSBU LLB Graduating Students, 2011-2015". In: *Int'l J. Clinical Legal Educ.* 31 (2024), p. 206.
- [363] Sowmya Vajjala. "Automated assessment of non-native learner essays: Investigating the role of linguistic features". In: *International Journal of Artificial Intelligence in Education* 28 (2018), pp. 79–105.
- [364] Daniel Van Niekerk et al. "Challenging Systematic Prejudices: An Investigation into Bias Against Women and Girls". In: (2024).
- [365] Michelle ME Van Pinxteren et al. "Trust in humanoid robots: implications for services marketing". In: *Journal of Services Marketing* (2019).
- [366] Martine Van Puyvelde et al. "Voice stress analysis: A new framework for voice and effort in human performance". In: *Frontiers in psychology* 9 (2018), p. 1994.
- [367] Gary R VandenBos. *APA dictionary of psychology*. American Psychological Association, 2007.
- [368] Pablo Buitron de la Vega et al. "Virtual reality simulated learning environments: a strategy to teach interprofessional students about social determinants of health". In: *Academic Medicine* 97.12 (2022), pp. 1799–1803.
- [369] Gyanendra K Verma and Uma Shanker Tiwary. "Affect representation and recognition in 3D continuous valence–arousal–dominance space". In: *Multimedia Tools and Applications* 76 (2017), pp. 2159–2183.
- [370] Lucía Vicente and Helena Matute. "Humans inherit artificial intelligence biases". In: *Scientific reports* 13.1 (2023), p. 15737.
- [371] Alexandros Vigkos et al. "XR and its potential for Europe". In: *Ecorys, Brussels* (2021).

- [372] Dimitrios Vlachopoulos and Agoritsa Makri. "The effect of games and simulations on higher education: a systematic literature review". In: *International Journal of Educational Technology in Higher Education* 14.1 (2017), pp. 1–33.
- [373] Astrid M Von der Pütten et al. ""It doesn't matter what you are!" Explaining social effects of agents and avatars". In: *Computers in Human Behavior* 26.6 (2010), pp. 1641–1650.
- [374] Isaac Wang et al. "Stop Copying Me: Evaluating nonverbal mimicry in embodied motivational agents". In: *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents. IVA '23*. Würzburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450 - 399944.
- [375] Wei-quan Wang and Izak Benbasat. "Trust in and adoption of online recommendation agents". In: *Journal of the association for information systems* 6.3 (2005), p. 4.
- [376] Yaqing Wang et al. "Generalizing from a few examples: A survey on few-shot learning". In: *ACM computing surveys (csur)* 53.3 (2020), pp. 1–34.
- [377] Ben Waters. ""A part to play": the value of role-play simulation in undergraduate legal education". In: *The Law Teacher* 50.2 (2016), pp. 172–194.
- [378] Jane Webster and Richard T. Watson. "Analyzing the Past to Prepare for the Future: Writing a Literature Review". In: *MIS quarterly* 26.2 (2002), pp. xiii–xxiii. ISSN: 02767783.
- [379] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. "Presence and cybersickness in virtual reality are negatively related: A review". In: *Frontiers in psychology* 10 (2019), p. 158.
- [380] Nane Winkler et al. "Lose yourself in VR: exploring the effects of virtual reality on individuals' immersion". In: *Proceedings of the 53rd hawaii international conference on system sciences*. 2020.
- [381] Bob G Witmer and Michael J Singer. "Measuring presence in virtual environments: A presence questionnaire". In: *Presence* 7.3 (1998), pp. 225–240.
- [382] Isabell Wohlgenannt, Alexander Simons, and Stefan Stieglitz. "Virtual reality". In: *Business & Information Systems Engineering* 62.5 (2020), pp. 455–461.
- [383] Virginie Woisard, S Bodin, and M Puech. "The Voice Handicap Index: impact of the translation in French on the validation". In: *Revue de laryngologie-otologie-rhinologie* 125.5 (2004), pp. 307–312.

- [384] Torsten Wörtwein et al. "Multimodal public speaking performance assessment". In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. 2015, pp. 43–50.
- [385] XR Association. *XRA Response to EU Consultation on Virtual Worlds*. Accessed: 2025-03-22. 2024. URL: <https://xra.org/wp-content/uploads/2024/03/XRA-Response-to-EU-Consultation-on-Virtual-Worlds.pdf>.
- [386] Yutaro Yagi et al. "Predicting multimodal presentation skills based on instance weighting domain adaptation". In: *Journal on Multimodal User Interfaces* 16.1 (2022), pp. 1–16.
- [387] Zixiaofan Yang et al. "What makes a speaker charismatic? Producing and perceiving charismatic speech". In: *Proc. 10th International Conference on Speech Prosody*. Vol. 2020. 2020, pp. 685–689.
- [388] Qi Yao, Ling Kuai, and Lan Jiang. "Effects of the anthropomorphic image of intelligent customer service avatars on consumers' willingness to interact after service failures". In: *Journal of Research in Interactive Marketing* ahead-of-print (2023).
- [389] Ellen Yeh and Guofang Wan. "The use of virtual worlds in foreign language teaching and learning". In: *Emerging Tools and Applications of Virtual Reality in Education*. IGI Global Scientific Publishing, 2016, pp. 145–167.
- [390] Elaine HJ Yew and Karen Goh. "Problem-based learning: An overview of its process and impact on learning". In: *Health professions education* 2.2 (2016), pp. 75–79.
- [391] Robert K Yin. *Case study research: Design and methods*. Vol. 5. sage, 2009.
- [392] Julie Furr Youngman. "From Remembering to Analyzing: Using Mini Mock Arguments to Deepen Understanding and Increase Engagement". In: *J. Legal Stud. Educ.* 37 (2020), p. 53.
- [393] Zuhair Zafar, Ashita Ashok, and Karsten Berns. "Personality Traits Assessment using P.A.D. Emotional Space in Human-robot Interaction". In: *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. Vol. 2. VISIGRAPP 2021. Vienna, Austria: SciTePress, 2021, pp. 111–118. ISBN: 978-989-758-488-6.
- [394] Hangyu Zhou et al. "Virtual reality as a reflection technique for public speaking training". In: *Applied Sciences* 11.9 (2021), p. 3988.

- [395] Sahba Zojaji, Adam Červeň, and Christopher Peters. “Impact of Multimodal Communication on Persuasiveness and Perceived Politeness of Virtual Agents in Small Groups”. In: *Proceedings of the 23th ACM International Conference on Intelligent Virtual Agents*. IVA '23. Würzburg, Germany: Association for Computing Machinery, 2023. ISBN: 978-145039994-4.
- [396] Sahba Zojaji, Christopher Peters, and Catherine Pelachaud. “Influence of virtual agent politeness behaviors on how users join small conversational groups”. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. IVA '20. Virtual Event, Scotland, UK: Association for Computing Machinery, 2020, pp. 1–8. ISBN: 978-145037586-3.