UNIVERSITY OF CALIFORNIA SAN DIEGO

Expressive, Interactive Robotic Patient Simulators for Clinical Education

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Computer Science (Computer Engineering)

by

Maryam Pourebadi khotbesara

Committee in charge:

Professor Laurel D. Riek, Chair
Professor Virginia de Sa
Professor Tania Morimoto
Professor Kristen Vaccaro

2023

The Dissertation of Maryam Pourebadi khotbesara is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my cherished mother, Dr. Saltanat Ravaee,

In the realm of unwavering dedication and remarkable passion, you have set the bar impossibly high. As a devoted mother raising three brilliant children, you seamlessly balanced the demands of a full-time career while tirelessly pursuing the highest academic degrees. Your remarkable journey stands as an unbeatable testament to your strength, resilience, and unwavering commitment. With boundless love, you nurtured our growth, shaping our minds with profound care, and providing us with exceptional education opportunities. With each sleepless night and every long day, you ensured that we received the best education, and forged a path of inspiration that I am privileged to walk upon today.

Maman joonam, as I stand here today, ready to embark on the next chapter of my academic journey, I am deeply humbled and profoundly grateful for the tremendous support you have afforded me. The love you have poured into my life, the sacrifices you have made, and the remarkable achievements you have attained have laid the foundation for my own success. Thank you for being my rock and my role model.

To my wonderful sister, Mahta,

Though physically separated by distance, our souls have remained forever intertwined. You stood by my side in every triumph and tribulation, a beacon of strength and love. Your unwavering belief in me and constant support have given me the courage to overcome obstacles. Thank you for being my confidant and a source of inspiration.

To my fantastic brother, Mahdi,

Your love, kindness, and support have illuminated my path throughout this journey. In your presence, I have found solace, encouragement, and a shoulder to lean on. Your unwavering faith in my abilities has fortified my resolve. Thank you for being my cheerleader, constant source of encouragement, and the driving force behind my growth.

And lastly, to my faithful feline companion, Pashmak,

In the midst of busy deadlines, your comforting presence has brought solace to my soul, and your purrs and playful antics provided a much-needed break. Thank you for reminding me to pause, appreciate life’s little joys, and find comfort in your affectionate companionship.

My beloved family, you have been the catalysts for my strength and success. With each step I have taken, I carried your hopes and dreams within my heart, a reminder of the profound impact we have on one another. Together, we have forged an unbreakable bond, characterized by love, support, and shared experiences. Thank you for continuously inspiring me to aim for greatness and reminding me of the immeasurable power of family.

With immense gratitude, an overflowing heart, and a smile on my face, I dedicate this dissertation to each and every one of you. May this work stand as a testament to the power of resilience, the strength of family, and the unwavering pursuit of knowledge.

This is for you.

EPIGRAPH

When it comes to robots,
looks matter!

Marie Pourebadi

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude to a number of people who played a significant role and provided their direct and indirect influence and support throughout my Ph.D. journey. I sincerely appreciate Prof. Laurel D. Riek for her unwavering support and guidance. Her expertise, patience, and dedication have played a crucial role in shaping my research and academic development. I also acknowledge Dr. Laurel D. Riek for her support as the chair of my committee, providing invaluable guidance and constructive feedback during the countless hours we spent working together. I also thank my dissertation committee members, Dr. Virginia De Sa, Dr. Tania Morimoto, and Dr. Kristen Vaccaro, for their valuable comments and suggestions toward my dissertation. I also thank Dr. Benjamin Lok for his valuable feedback.

I thank my collaborators from the School of Medicine at UC San Diego, Dr. Jamie Labuzetta, Dr. Steven Yang, Dr. Preetham Suresh, Dr. Sean Evans, Dr. Leena Kansal, and Dr. Dominic Ferrey, for their support during my Ph.D. journey as discussed herein. I thank Suhas Pai and Rainee Pei for their assistance in this work described in the following content. I also thank my past and present labmates, Dr. Angelique Taylor, Dr. Darren Chan, Dr. Hee Rin Lee, Dr. Dago Cruz-Sandoval, Alyssa Kubota, Sachiko Matsumoto, Sandhya Jayaraman, Anya Bouzida, Rabeya Jamshad, and Pratyusha Ghosh for their unwavering support and encouragement during my journey.

I would like to take the opportunity to thank my mother, Dr. Saltanat Ravaee, and my siblings, Mahta Pourebadi and Mahdi Pourebadi. I am deeply grateful to my loving family for their unwavering support, endless encouragement, and boundless love throughout my journey.

Chapter 2 contains material from “Expressive Robotic Patient Simulators for Clinical Education”, by M. Pourebadi, and L. D. Riek, which appears in the Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI), Robots 4 Learning workshop, and material from “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future” by M. Pourebadi, and L. D. Riek., which appears in In the Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM

HEALTH), 3(2), 1-32. I was the primary investigator and author of these papers.

Chapter 3 contains material from “Expressive Robotic Patient Simulators for Clinical Education”, by M. Pourebadi, and L. D. Riek, which appears in the Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI), Robots 4 Learning workshop, and material from “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future” by M. Pourebadi, and L. D. Riek., which appears in In the Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32. I was the primary investigator and author of these papers.

Chapter 4 contains material from “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future” by M. Pourebadi, and L. D. Riek., which appears in the Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32. I was the primary investigator and author of this paper. This chapter also contains material from part of “Modeling and synthesizing idiopathic facial paralysis” by M. Moosaei, M. Pourebadi, and L. D. Riek, which appears in the Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition. Maryam Moosaei and I were the primary authors of this work.

For the research in Chapter 5, I thank Drs. Jamie Labuzetta and Steven Yang for facilitating data collection from patients with neurological conditions, and providing feedback on my data analysis efforts. I thank Dr. Preetham Suresh, who gave me insights into designing interactive expressive RPSs for the clinical simulation. I also thank Suhas Pai for his assistance with the FPM framework implementation. This chapter contains material from “Mimicking acute stroke findings with a digital avatar” by M. Pourebadi, J. LaBuzetta, C. Gonzalez, P. Suresh, and L. D. Riek., which appears in STROKE, Vol.51, and “Modeling and Synthesizing Stroke on Expressive Patient Simulator Robots” by M. Pourebadi, and L. D. Riek, which appears in Proceedings of the AAAI Artificial Intelligence for Human-Robot Interaction (AAAI AI-HRI), and “Modeling and Synthesizing Stroke on Robotic Patient Simulators” by M. Pourebadi, J., Labuzetta, and L. D. Riek, which will be submitted to the ACM Transactions on Human-Robot

Interaction (THRI) I was the primary investigator and author of these papers.

For the research in Chapter 6, I thank Dr. Jamie Labuzetta for helping me design and evaluate the clinical scenarios for ROSE. I thank Drs. Sean Evans, Leena Kansal, and Dominic Ferrey for their assistance with prototype evaluation and study design. I thank Suhas Pai for helping me implement and report the MMC framework development during the prototype design phase. I also thank Rainee Pei for her assistant with the data analysis and results reporting. This chapter contains material from “ROSE: An Interactive Social Robot for Medical Education” by M. Pourebadi, S. Pai, R. Pei, and L. D. Riek, which will be submitted to the ACM/IEEE International Conference on Human-Robot Interaction (HRI). I was the primary investigator and author of this paper.

My research was made possible by funding support from the National Science Foundation (NSF).

VITA

2014	Bachelor of Science in Computer Engineering, University of Alzahra, Tehran
2017	Masters of Science in Computer Science, Kent State University, Ohio
2023	Doctor of Philosophy in Computer Science (Computer Engineering), University of California San Diego, California

PUBLICATIONS

1. **Pourebadi, M.**, Pai, S., Pei, R., Riek, L.D. (2024) “ROSE: An Interactive Social Robot for Medical Education”, Submission Pending, The ACM/IEEE International Conference on Human-Robot Interaction (HRI).

2. **Pourebadi, M.**, LaBuzetta, J. N., Riek, L.D. (2023) “Modeling and Synthesizing Stroke on Expressive Patient Simulator Robots”, Submission Pending, The ACM Transactions on Human-Robot Interaction (THRI).

3. **Pourebadi, M.**, Riek, L.D. (2022) “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future”, In Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32.

4. Kubota, A., **Pourebadi, M.**, Banh, S., Kim, S., Riek, L.D. (2021) “Somebody That I Used to Know: The Risks of Personalizing Robots for Dementia Care”, In Proceedings of We Robot 2021. [Acceptance rate: 15%]

5. **Pourebadi, M.**, and Riek, L.D. (2020). “Stroke Modeling and Synthesis for Robotic and Virtual Patient Simulators”, In Proceedings of the AAAI Fall Symposium on Artificial Intelligence in Human-Robot Interaction: Trust and Explainability in Artificial Intelligence for Human-Robot Interaction (AAAI AI-HRI).

6. **Pourebadi, M.**, Gonzalez, C. G., LaBuzetta, J. N., Meyer, B. C., Suresh, P., Riek, L. D. (2020). “Mimicking Acute Stroke Findings With a Digital Agent”, Stroke, Proceedings of the American Heart Association Journal (AHA).

7. Ghayoumi, M., **Pourebadi, M.** (2019). “Fuzzy Knowledge-Based Architecture for Learning and Interaction in Social Robots”, Proceedings of the AAAI Fall Symposium on Artificial Intelligence and Human-Robot Interaction: Service Robots in Human Environments (AAAI AI-HRI).

8. Moosaei, M., **Pourebadi, M.**, and Riek, L.D. (2019). “Modeling and Synthesizing Idiopathic Facial Paralysis”, Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG). [Acceptance rate: 20%]

9. **Pourebadi, M.**, and Riek, L.D. (2018). “Expressive Robotic Patient Simulators for Clinical Education”, Robots 4 Learning workshop at the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI).

10. **Pourebadi, M.**, Pourebadi, M. (2016). “MLP Neural Network Based Approach for Facial Expression Analysis”, In Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). [Acceptance rate: 24%]

11. Ghayoumi, M., Khan, J., **Pourebadi, M.**, Bauer, E., Hossain, A. (2016). “Follower Robot with an Optimized Gesture Recognition System”, Proceedings of Socially and Physically Assistive Robotics For Humanity workshop at Robotics: Science and Systems (RSS).

ABSTRACT OF THE DISSERTATION

Expressive, Interactive Robotic Patient Simulators for Clinical Education

by

Maryam Pourebadi khotbesara

Doctor of Philosophy in Computer Science (Computer Engineering)

University of California San Diego, 2023

Professor Laurel D. Riek, Chair

Preventable patient harm is the root cause of many adverse events in healthcare, and is a leading cause of mortality and morbidity worldwide. One way to address this is through career-long clinical education, often via the use of robotic patient simulator (RPS) systems. These highly realistic human-like physical or virtual platforms enable clinical learners to safely practice their diagnostic, procedural, and social interaction skills without harming real patients. However, most commercial RPS systems lack a realistic depiction of non-verbal facial cues, limiting learner engagement and immersion, which can ultimately lead to incorrect skill transfer and patient harm.

In my PhD, I have worked to address this gap by building new interactive and expressive RPS systems, whose faces are based entirely on real patients, and the system's expressions are realistic. In this dissertation, I will describe the main contributions of my work, including 1) Developing an end-to-end analysis-modeling-synthesis framework that can easily and robustly recognize, model, and synthesize patient-driven facial expressions and clinical cues on the faces of virtual and physical RPS systems, 2) Developing new methods to create accurate computational models of multiple pathologies, including stroke and Bell's Palsy, 3) successfully synthesizing these models on RPS systems, and, 4) designing the RPS as a clinical educational tool tested with clinicians.

My research opens new avenues of exploration in robotics, human-robot interaction, and health technology, and may trigger a new round of relevant technological innovations by creating the next generation of RPS technology. My work will enable roboticists to discover platform-independent methods to control the facial expressions of both robots and virtual agents, yielding new modalities for interaction. Furthermore, disseminating the results of this work to the research community will help both the broader robotics and healthcare communities employ these novel systems in their own application domains. This work serves as a bridge between robotics and healthcare research and practice, and offers promising opportunities to reduce misdiagnoses and bias in healthcare, ultimately reducing preventable patient harm.

Chapter 1

Introduction

For several decades, researchers in the field of human-robot interaction (HRI) have been studying how humanlike robots can collaborate with humans, support their work, and assist them in their daily lives [207, 131, 117, 81]. Their design and deployment are influenced by a range of socio-technical, economic, and contextual factors, which, in return, inspires research areas, especially in the realm of clinical applications [109]. For example, recent research topics include cognitively assistive robots, providing social engagement for older adults, and supporting telemedical care delivery in hospitals [148, 163, 162, 183, 82].

There is emerging interest in using robotics technology to address key challenges in healthcare, particularly those related to the quality, safety, and cost of care delivery. However, there are several key contextual challenges to realizing this vision, including the increasing cost of healthcare services [63], the dynamic nature of clinical environments [118], and a global shortfall in professional healthcare workers with sufficient clinical education and skills [260].

Providing healthcare systems with robots may help address these gaps. For example, robots may help reduce "non-value added" physical tasks like transportation and inventory management [207], thus minimizing errors and freeing up more time for patient care [237]. Robots could mitigate the considerable workload and cognitive strain experienced by healthcare workers, who are grappling with labor shortages, hazardous environments, and significant personal risk [107].

In the space of healthcare education and training (HET), robots can provide clinical educators with new learning opportunities for their students.

Simulation-based clinical learning is an important component of HET already incorporated into several subspecialties that intersect with diagnosing and treating patients [207, 196]. Simulation-based clinical learning offers healthcare professionals with clinical education and training in safe, realistic environments that replicate a range of scenarios, allowing them to practice their clinical and procedural skills without risking harm to actual patients [174].

There are many benefits to this learning method. Research suggests using patient simulators may reduce preventable patient harm [215], which causes death and serious injury to nearly 25% of medicare patients, representing millions of lives affected [121]. Furthermore, researchers found that when compared with non-digital educational methods, patient simulator systems are superior in terms of improving knowledge and skill-building [146]. (See Chapter. chapter 2 for a detailed description of simulation-based clinical learning).

Robotic patient simulator (RPS) systems are one of the common simulators incorporated into simulation-based clinical learning. RPS systems are humanlike physical or virtual platforms that serve as a conducive medium for clinical learners (CLs) to safely practice their diagnostic, procedural, and social interaction proficiencies. RPS systems can also benefit clinical educators (CEs) by enabling them to perform diverse simulated clinical scenarios tailored to the needs of CLs, rather than relying on the availability of real patients. Research demonstrated that simulation systems might enhance CLs' clinical skills, understanding, comprehension, and keenness for education more efficiently than traditional text-based learning approaches [146, 137]. Career-long clinical education with RPS systems may reduce the incidence of preventable patient harm [195, 197, 192, 196, 172, 136].

Despite the many benefits of using patient simulators, there are several challenges with existing systems that may impede how effective they are at supporting HET.

Most current commercial RPS systems suffer from a major design flaw: they completely lack facial expressions (FEs) and, thus, the ability to convey key diagnostic features of different disorders and social cues, which can eventually cause problems with learner immersion and skill transfer. Lack of facial expressiveness is problematic because FEs serve as an important social and clinical cue in patients; thus, the lack of expressions in simulators may adversely affect CLs' learning performance. In order to successfully support clinicians in real-world HET settings, RPS systems will need to be able to exhibit humanlike actions and behaviors.

The anthropomorphization of social robots gives rise to concerns about the robot's impact on users' emotions, expectations, and interactions [199, 183, 114]. The limited expressiveness of humanlike social robots can lead to reduced emotional engagement with users, hindering the development of empathy and social presence in HRI [185]. Even for the RPS systems with expressive faces, their appearance and characteristics may not be widely customizable to address CEs' needs. These gaps call for researchers to be mindful of humanlike social robots' ethical and societal impacts.

While enabling RPS systems with an expressive face can address this challenge, it still creates a bigger challenge with designing expressive systems: facial expressions are very person-dependent and can vary from person to person [275]. It is challenging to analyze, model, and synthesize FEs of a small subgroup of patients on simulators' faces and develop generalized expressive simulator systems that are capable of representing a diverse group of patients (including but not limited to different ages, genders, and ethnicities who are affected by different diseases and conditions) [275].

Another challenge is that incorrectly (or not) exhibiting symptoms on a simulator's face may reinforce incorrect skills in CLs, and could lead to future patient harm. Furthermore, developers may face physical limitations preventing them from advancing the state-of-the-art. Other challenges include the simulator's usability, controllability, high costs, physical limitations, and the need to recruit experts with various skills.

Tackling these technical challenges to advance the state-of-the-art needs work on several fronts. These include the creation of capable and usable RPS systems, new techniques for recognizing and synthesizing facial expressions on simulators, novel computational methods for developing humanlike face models for them, and new means for evaluating these systems.

My work is situated in this problem domain of enabling RPS systems to realistically exhibit humanlike actions and behaviors similar to human patients to support learners in dynamic, real-world educational environments. My work examines state-of-the-art technical approaches for human facial expression analysis, facial action modeling, and facial expression synthesis on RPS systems. More specifically, my research addresses the need for new training tools in HET. In my work, I develop expressive RPSs capable of realistically synthesizing non-verbal asymmetric facial cues that are important for the rapid diagnosis of neurological emergencies, such as stroke. I contextualize this work within the field of HRI, as ultimately, I am interested in how this technology can be leveraged to improve immersion, engagement, and educational outcomes for learners.

1.1 Motivation and scope

Current RPS systems do not promote humanlike interaction with CLs, nor can they operate autonomously. The existing RPS systems rarely provide CLs with an interactive platform to adequately engage them in automatic interaction, which may result in transferring poor social interaction skills to CLs, leading them to perform poorly on patient exams. This can be problematic for diagnosing and interacting with real patients, for example, those presenting with neurological impairments such as stroke. Additionally, our clinical collaborators have shared that CLs often fail to master the neurological examination on simulated patients. This may result in inadequately performing the exam on real patients [116]. Even if a CL performs the exam well, they may have little confidence in the accuracy of their findings. Given the subjective nature of the interpretation of these findings, low confidence in the neurological exam, irrespective of how well it is performed, may lead to an uncertain interpretation of the results. This uncertainty can lead to missed opportunities for acute interventions, prompt treatments, and prevention of serious harm [176, 47].

As the purpose of existing RPS systems is mainly informative rather than interactive, they lack various communication modalities, which may limit the range and quality of interactions between the RPS robot and users, affecting the overall training effectiveness. The lack of immersion and engagement in the interaction can also result in reduced motivation, interest, and retention of the training content in the context of HRI.

Existing RPS systems may offer tools to practice basic clinical skills (e.g., taking vital signs, and performing physical exams); however, they only partially replicate interactive clinical scenarios that replicate real-life medical situations with evolving and changing conditions. This can lead to limited opportunities for effective clinical skill acquisition and knowledge transfer, potentially resulting in missed opportunities for acute interventions, tools, prompt treatment, and prevention of serious harm.

Designing a clinical training tool with an interactive, expressive RPS to address these gaps, can also introduce design and technical challenges. For example, if poorly designed, interacting with the system can heavily rely on advanced technology or complex interfaces, limiting clinicians' perceived ease of use. The perception of robots' ease of use may significantly influence the clinicians' acceptance of new technology in their professional life [109]. Moreover, the limited usability of robots can make it challenging for users to effectively work with the robot and access the training content, ultimately resulting in frustration and lower learning outcomes.

These challenges necessitate a new interactive clinical training intervention to enhance the learning experience of CLs by providing a realistic and immersive environment for practicing dynamic clinical scenarios.

Thus, **the research goal of my work is to create highly lifelike and interactive robots, capable of accurately replicating patient symptoms and autonomously engaging socially with clinical learners.** My work focuses on supporting CLs for diagnosis and treatment of neurological emergencies, especially stroke. This dissertation discusses research at the intersection of HRI, robotics, automatic facial and gesture recognition (FG), computer vision, and health technology, to enable socially interactive robots to simulate human-patient-like expressions and interaction. While there are many dimensions to this problem, this dissertation explores the following aspects:

- How can one generate data-driven statistical models of patients' facial expressions, representative of clinical conditions, such as facial paralysis.
- How might robotic faces automatically synthesize these models.
- How RPS systems simulate existing dynamic clinical scenarios, such as neurological exams, while providing interactive and engaging learning experiences.
- How can robots automatically engage in interactions and effectively communicate with CLs in real-time?
- How using social robots as an educational tool can create an immersive and realistic environment for practicing diagnosis and treatment skills.
- How can robots provide end-users with lower levels of technology literacy intuitive control features to support the systems' ease of use
- What are the key design requirements for building socially interactive robots for CLs.
- How to consider the ethical implications of the use of humanlike robots in HET.

1.2 Contributions

The contributions of this work are as follows:

Presented the potential for humanlike robotic patient simulators in transforming HET [195]. My work identified the gaps and opportunities in existing learning modalities in order to recognize the potential of humanlike RPS in the context of HET. I outlined the root causes of preventable patient harm in clinical settings, and how simulation-based clinical education is one of the best defenses to reduce the incidence of patient harm. Second, I examined the benefits and challenges that accompany common learning modalities of HET, including virtual and robotic patient simulators. Finally, I presented major gaps in introducing the use of humanlike learning modalities into clinical education. This work established the foundations of designing and deploying expressive RPS systems in HET.

Created new virtual and physical faces for robots in HET [195, 196]. I investigated the effect of expressive mechanical and rendered faces in RPS design and presented my work on building new expressive faces. First, I discussed the role of humanlike behaviors in social interactions and outlined the benefits and key challenges of enabling virtual and robotic embodiments to depict verbal and non-verbal behaviors. Second, I explored techniques for detecting, modeling, and synthesizing humanlike FEs in robotic faces. Finally, I discussed my research on virtual and robot patient simulator faces, enhanced with the capacity to exhibit nuanced verbal and non-verbal behaviors and cues, while displaying diverse appearances and backgrounds. This work stands as a potential instrument in HET, opening new frontiers in developing expressive RPS systems. Moreover, this work provides valuable insights to researchers by examining methods for detecting, modeling, and synthesizing FEs, with potential applications in enhancing social interactions, and clinical education.

Built an analysis-mask-synthesis (AMS) framework and developed a general facial paralysis masks (FPM) framework to generate accurate representations of FEs for RPSs based on real patient's facial characteristics [172, 195]. This work had two main goals. First, it aimed to enable people to easily synthesize human facial movements on any robotic and virtual faces in real time. Second, it aimed to understand how robots can accurately and realistically depict asymmetric facial expressions. In this work, we designed and developed the end-to-end AMS control framework which robustly recognizes the facial movements of an operator, masks their movements with computational models of FP, and automatically, easily, and robustly synthesizes the masked facial movements across a range of RPS embodiments. Moreover, we developed the FPM framework, a system designed to automatically generate high-precision computational models that realistically depict FEs associated with patient pathologies, and are constructible in real time. Furthermore, we integrated these two frameworks, enabling the overall system to elevate the authenticity and accuracy of FE representations on RPS faces, based on real patients' facial characteristics. Finally, we reported the results from an expert-based user study, highlighting that experts perceived our expressive virtual patient simulator to be realistic and comparable to humans with Bell's Palsy. This study fosters technological innovations in HET and provides platform-independent methods for controlling the FE of robots and virtual agents, leading to novel modalities for interaction.

Introduced RPSwS for modeling and synthesizing acute stroke [192, 197, 191]. The core objective of this work was to architect a comprehensive system that adapts our general FPM and AMS frameworks in order to create data-generated models of real humans, and overlay them on a robot to enable it to depict realistic verbal and non-verbal cues. This enables RPSs to accurately and effectively depict stroke symptoms, thereby advancing the landscape of HET for stroke diagnosis and treatment. Thus, I introduced robotic patient simulators with stroke (RPSwS): a new expressive clinical training tool capable of realistically depicting non-verbal, asymmetric FP cues representing acute stroke. First, I developed the Stroke FPM framework, comprising a machine learning method for accurate facial landmark tracking in a newly collected dataset of PwS, and a statistical modeling approach to use tracked facial point values to automatically represent the visually significant features of stroke. Second, I created a RPSwS by developing an end-to-end Stroke AMS control framework which uses the generated stroke models to automatically depict stroke on the face of RPS systems. Third, I identified key features for enhancing realism and similarity between synthesized stroke faces of RPSwS and those of PwS through a study with clinical experts.

To my understanding, the Stroke FPM framework pioneers a computational modeling approach that can capture stroke-associated FP cues across the upper and lower facial regions, while extracting various representations of neurological effects through systematic analysis of facial movement patterns. Thus, the Stroke FPM provides researchers with a reliable tool for modeling and analyzing facial asymmetry and movements in both upper and lower facial regions.

In addition, the Stroke AMS framework allows for the accurate rendering of stroke characteristics on various RPS systems, enabling the creation of highly realistic FP simulations, thereby offering robotics researchers a means to develop empirically derived facial representations for robots in a HET environment [197]. By producing a set of 75 facial models that represent stroke in various facial regions, our research yielded insights into the best representations of stroke in each facial region based on professional expertise, enhancing the precision and reliability of stroke representations for different facial regions. This work has significant cross-disciplinary impacts in clinical education, health informatics, FG, robotics, and HRI, as it pioneers a statistical modeling methodology for comprehensive stroke-associated FP representation, facilitates realistic FP simulations on various RPS systems, and provides insights for asymmetric FE analysis, social robot design, and understanding the effects of facial asymmetry on social interactions.

Designed and developed ROSE: an interactive social robot for medical education [193].

The core objective of this work was to understand how to enable a robot with social intelligence to autonomously exhibit realistic behaviors and effectively engage in interactions within real clinical education settings. This work spearheads the design and development of a clinical training tool employing an interactive, socially adept RPSwS, ROSE, to enhance the learning experience for CLs. ROSE provides a diverse, inclusive, and customizable platform that enhances the realism and accessibility of HET. Through close collaboration with neurologists and CEs, we identified user-centered design requirements for building ROSE. Second, we co-designed and implemented a new automatic multi-modal communication (MMC) framework for the robot, supporting the automation of clinical scenario simulations, interaction, and engagement in RPSwS. Third, we presented and evaluated ROSE as a clinical training tool that leverages the MMC framework to simulate clinical scenarios and autonomously engage in user interactions, providing a realistic and immersive learning environment for CLs to hone their diagnostic skills.

This work is important for HET, as well as the broader healthcare, FG, and HRI communities. To our knowledge, ROSE is the first patient-data-driven, interactive clinical training tool accessible to clinicians to practice diagnosis and treatment of neurological disorders such as stroke. To our knowledge, ROSE is the first patient-data-driven, interactive clinical training tool accessible to clinicians to practice the diagnosis and treatment of stroke. Our work lays the foundation for extending the accessibility of educational interventions to the healthcare domain, enabling humanlike social robots to support HET through an automated interactive, expressive robot. Moreover, our work provides a framework for researchers to explore HRI in new experiential learning settings (e.g., build RPS systems to enable CLs to avoid forming biased impressions) and broader domains (e.g., explore methods for designing social robots to enhance people's perception of individuals with FP).

1.3 Publications

The work presented in this dissertation is based on the following papers that are either in the process of being submitted for publication or have already been published.

1. **Pourebadi, M.**, Pai, S., Pei, R., Riek, L.D. (2024) "ROSE: An Interactive Social Robot for Medical Education", Submission Pending, The ACM/IEEE International Conference on Human-Robot Interaction (HRI).

2. **Pourebadi, M.**, LaBuzetta, J. N., Riek, L.D. (2023) "Modeling and Synthesizing Stroke on Expressive Patient Simulator Robots", Submission Pending, The ACM Transactions on Human-Robot Interaction (THRI).

3. **Pourebadi, M.**, Riek, L.D. (2022) "Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future", In Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32.

4. **Pourebadi, M.**, and Riek, L.D. (2020). "Stroke Modeling and Synthesis for Robotic and Virtual Patient Simulators", In Proceedings of the AAAI Fall Symposium on Artificial

Intelligence in Human-Robot Interaction: Trust and Explainability in Artificial Intelligence for Human-Robot Interaction (AAAI AI-HRI).

5. **Pourebadi, M.**, Gonzalez, C. G., LaBuzetta, J. N., Meyer, B. C., Suresh, P., Riek, L. D. (2020). “Mimicking Acute Stroke Findings With a Digital Agent”, International Stroke Conference (ISC), Proceedings of the American Heart Association Journal (AHA).

6. Moosaei, M., **Pourebadi, M.**, and Riek, L.D. (2019). “Modeling and Synthesizing Idiopathic Facial Paralysis”, Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG). [Acceptance rate: 20%]

7. **Pourebadi, M.**, and Riek, L.D. (2018). “Expressive Robotic Patient Simulators for Clinical Education”, Robots 4 Learning workshop at the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI).

1.4 Ethical procedures

The Institutional Review Board at the University of California San Diego has formally reviewed human subject experiments described in this dissertation. Participants provided informed consent to participate in experimental research in all human subjects experiments. The researcher appropriately anonymized and securely stored all collected data.

1.5 Dissertation overview

The dissertation is organized as follows:

- **Chapter 2** provides an overview of related work in the areas of HET, learning modalities, and robotic patient simulators.

- **Chapter 3** presents my work on developing physical and virtual faces as a communication modality for robots, and explores common methods for facial expression analysis (FEA), facial action modeling (FAM), and facial expression synthesis and animation (FSA).

- **Chapter 4** introduces my work on creating new embodiments for robots and virtual agents, developing an end-to-end analysis-modeling-synthesis (AMS) control framework that can robustly recognize, model, and synthesize FEs across a range of robotic embodiments, and building the Facial Paralysis Mask (FPM) framework that generates accurate computational models of multiple patient-driven pathologies that can be synthesized onto robotics faces.

- **Chapter 5** describes the design, implementation, and evaluation of the stroke AMS and stroke FPM: two frameworks for generating statistical modeling approaches representing facial characteristics of stroke, and applying the generated models onto the face of an RPS system to automatically display stroke.

- **Chapter 6** introduces the design, development, and results of evaluating ROSE, an interactive social robot for clinical education.

- **Chapter 7** concludes by summarizing the primary contributions of this dissertation, deliberating on prospects for future research, presenting open questions for the HRI and FG communities, and delivering concluding remarks.

Chapter 2

Background

2.1 Introduction

For more than five decades, researchers in the field of HRI have been building and studying how robots can collaborate with humans, support them with their work, and assist them in their daily lives [207, 131, 117]. For example, autonomous mobile robots work side-by-side with skilled human workers in factories and retail sectors [214]. Social robots inform and guide passengers in large and busy airports [242]. In both clinical and home settings, robots have been used to assist healthcare workers, clean rooms, ferry supplies, and support people with disabilities and older adults in rehabilitation and task assistance [207].

There is emerging interest in using robotics technology to address key challenges in healthcare, particularly those related to the quality, safety, and cost of care delivery. However, there are several key contextual challenges to realizing this vision. One big concern is the rapidly increasing costs of healthcare. For example, in the United States, healthcare is expensive across a range of services including administrative costs, pharmaceutical spending, individual services, and the use of high-income trained healthcare workers [63]. Another challenge is the dynamic nature of clinical environments with occupational hazards that put health care workers at risk of injury and disability [118, 237, 238]. Additionally, the global shortfall in professional healthcare workers with sufficient clinical education and skills is challenging [260].

Providing healthcare systems with robots may help address these gaps. For example,

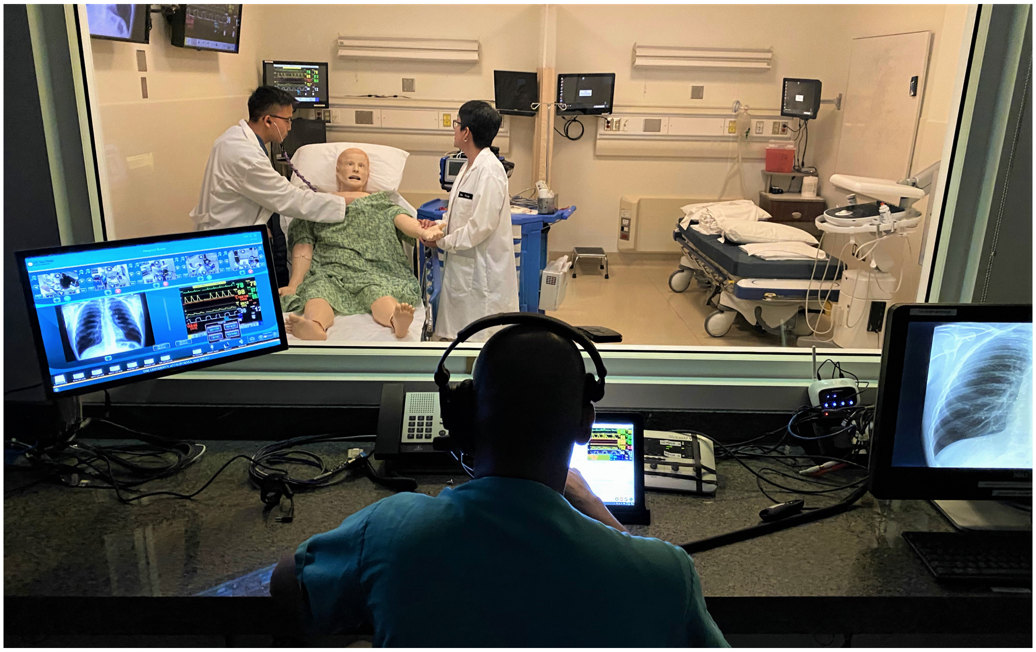


Figure 2.1. A typical patient simulation center setup. Clinical learners treat a non-expressive robotic patient simulator. Its physiology is controlled by a clinical educator.

robots can support the independence of people with disabilities by enabling transitions to home based care. Robots can also help clinicians and caregivers with care tasks including physical, cognitive, and manipulation tasks [207, 107, 18, 254, 120], as well as healthcare worker education (See Figure 2.1).

Robots can potentially enable healthcare workers to spend more time with patients and less time engaging in "non-value added" physical tasks, and reduce the errors caused by the overburden of these tasks [237, 107]. These physical tasks include transportation, inventory, and spending time searching and waiting [207]. For example, Tug robots [11] are medical transportation robots that autonomously move through hospitals, delivering supplies, meals, and medication to patients.

Moreover, robots can assist in clinical learning. For example, humanoid patient simulators can mimic human function (physiology) or anatomy (biology). Some of these simulators are engineered systems that model information integration and flow to help clinical learners study human physiology. Others present models of human patient biology and cognition to provide clinicians with a platform to practice different skills including task execution, testing and validation, diagnosis and prognosis, training, and social and cognitive interaction.

Robotic patient simulators, virtual patient simulators (VPS) and augmented reality patient simulators (APS) are three main technologies used to represent realistic, expressive patients within the context of clinical education. Clinical educators (CE) can use them to convey realistic scenarios, and clinical learners (CL) can practice different procedural and communication skills without harming real patients.

Although there are many benefits associated with using RPS, VPS, and APS systems, their designs suffer from a lack of FE, which are both a key social function and clinical cue conveyed by real patients. While enabling RPS and VPS systems with an expressive face can address this challenge, still it creates a bigger challenge with designing expressive systems: facial expressions are very person-dependent and can vary from person to person [275]. It is challenging to analyze, model, and synthesize FEs of a small subgroup of patients on simulators' faces and develop generalized expressive simulator systems that are capable of representing a diverse group of patients (including but not limited to different ages, genders, and ethnicities who are affected by different diseases and conditions) [275].

Another challenge is that incorrectly (or not) exhibiting symptoms on a simulator's face may reinforce incorrect skills in CLs, and could lead to future patient harm. Furthermore, developers may face physical limitations preventing them from advancing the state-of-the-art. For example, VPSs are limited by flat 2D display mediums, making them unable to represent a physical 3D human-shape volume which clinicians can palpate in order to perform clinical assessments. Other challenges include the simulator's usability, controllability, high costs, and physical limitations, as well as the need of recruiting experts with various skills.

Tackling these technical challenges to advance the state-of-the art needs work on several fronts. These include the creation of capable and usable RPS and VPS systems, new techniques for recognizing and synthesizing facial expressions on simulators, novel computational methods for developing humanlike face model for them, and new means for evaluating these systems.

Ultimately addressing these gaps can provide healthcare education with realistic, expressive simulators capable of mimicking patient-like expressions. This has the potential to positively affect CLs' retention, and eventually, revolutionize healthcare education.

In this review, we discuss research at the intersection of robotics, computer vision, and clinical education, to enable socially interactive robots and virtual agents to simulate human-patient-like expressions and interact with real humans. In this work we provide an overview of the root causes of preventable patient harm, and contextualize clinical education as a means for addressing it. We outline common learning modalities, including VPS and RPS systems, and outline key opportunities to improve them.

2.2 Background

2.2.1 Patient Safety and Healthcare Education

The World Health Organization defines patient safety as "the absence of preventable harm to a patient during the process of health care and reduction of risk of unnecessary harm associated with health care to an acceptable minimum" [33]. Taking an action (errors of omission) or inaction (errors of commission) by healthcare workers, system failures, or a combination of these two factors may cause or lead to preventable patient harm [128].

Preventable patient harm represents the root cause of many adverse events experienced in healthcare departments including intensive care units, and is a leading cause of mortality and morbidity in the world. Conservative estimates suggest preventable patient harm causes over 400,000 preventable deaths per year in the US hospitals alone [136], and 4-8 million experience serious harm and injury. It is estimated between 27-33% of patients experience an adverse event as a result of their care [245, 230, 84, 1].

While better designed healthcare systems, services, and processes, as well as new technologies, can help reduce the incidence of patient harm, in the short term one of the best approaches is high-quality clinical education. Recent work shows that healthcare education and training is the most effective mechanism to reduce the incidence of patient harm and improve patient safety [215].

One way to advance the state-of-the-art of healthcare education is through the development intelligent learning modalities, such as simulation systems. Simulators provide CLs the chance to safely study the causes and effects of errors, while avoiding harm to real patients. Using simulators also improves CLs' comprehension, confidence, efficiency, and enthusiasm for learning [137]. When compared with non-digital learning methods, using patient simulators can more effectively improve CLs' skills, and at least as effectively improve knowledge [146].

CEs may also benefit from using simulation systems to run a variety of desired clinical simulation scenarios on realistic patients based on a learner's need, instead of patients' availability. Examples of these scenarios include nursing simulation scenarios [9], physician scenarios [30], and surgical simulation scenarios [32]. Studies also indicate that using simulation improves the performance of learner evaluation and educational needs diagnosis by CEs [45]. This work, and others, are encouraging, and suggest that augmenting existing healthcare simulation systems with emerging AI-based technologies offers promising opportunities to substantially reduce preventable patient harm, as well as risks to clinicians.

2.2.2 Patient Simulator Types, Benefits, and Challenges

There are four types of simulated patients used in simulation-based clinical learning: standardized human patients, augmented reality patient simulators, virtual patient simulators, and robotic patient simulators. Table 2.1 illustrates the structure, functionality, and controlability for each type of patient simulator. This is further discussed below.

SHP are live actors who assume the roles of patients. They convey a series of symptoms and/or a scenario defined by CEs [57]. SHPs are beneficial as they provide CLs with a real-human case study to practice their history-taking and clinical assessment skills. As a result, SHPs enable the learning process to sometimes deviate from a predefined senario, as this type of simulator can adapt to unexpected changes on-the-fly.

Table 2.1. Simulators: the structure, functionality, and controlability.

Type	Platform	Physiological variables	Visual appearance	Control	Scheduling time
Standardized human patients (SHPs)	Real: 3D real-human body	Can present some of the variables.	Can display dynamic facial expressions (FEs), gestures, and some of the abnormal visual findings.	Controlled by a human.	High
Augmented reality patient simulators (APSs)	Hybrid: Visual appearance projected to a 3D physical surface.	Can easily present all the variables.	Can be programmed to richly display dynamic FEs, gestures, and all abnormal visual findings.	Ranges from fully automated to teleoperated to pre-recorded mode.	Low
Virtual patient simulators (VPSs)	Virtual: 2D monitor or TV or Tablet	Can only present the visual physiological variables due to 2D display limitations.	Can be programmed to richly display dynamic FEs, gestures, and all abnormal visual findings.	Ranges from fully automated to teleoperated to pre-recorded mode.	Low
Robotic patient simulators (RPSs)	Mchanical: 3D human-like physical robot	Can exhibit 5000+ physiology changes on it. Verbal responses controlled using a live operator.	Mostly have a static face. They can be programmed to display some of dynamic FEs, gestures, and abnormal visual findings.	Ranges from fully automated to teleoperated to pre-recorded mode.	Low

However, SHPs cannot accurately exhibit many symptoms of real patients, such as facial paralysis or physiological changes. Furthermore, recruiting SHPs can be difficult and expensive, especially ones at younger ages because of child labor laws and scheduling difficulties [57, 44, 243, 115].

APSs, also known as physical-virtual simulators, use augmented reality (AR) techniques to combine physical human-shaped surfaces with dynamic visual imagery projected on its surface [89]. APSs combine the benefits of two worlds: its physicality can convey a realistic, embodied similarity to people, while its virtual component can display dynamic appearances and FEs without being limited by hardware infrastructure.

However, it is still challenging to display an accurate representation of naturalistic symptoms even in an AR environment. APSs also present some challenges depending on the AR modalities and techniques used. Recent work [74] suggests to avoid the use of commercially available head-mounted displays for AR surgical interventions, because perceptual issues can affect user performance. In front-projected imagery [212], the shadow of users can hover over the projection [258] and cause the CEs fail to display desired scenarios. Rear-projected imagery [88]

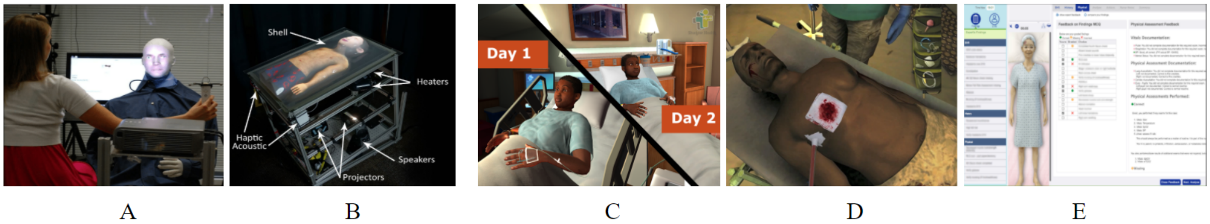


Figure 2.2. Robotic patient simulators with virtual faces. **Left:** Augmented reality patient simulators (APSs) with rendered faces based on projector placement. **A)** An APS system with front-projected imagery [212], and **B)** An APS system with rear-projected imagery [88]. **Right:** Examples of virtual patient simulators (VPSs) software with rendered faces. **C)** Shadow Health [34], **D)** CliniSpace [8], and **E)** i-Human [28].

can solve both multi-user and projection occlusion problems; however, it requires a sufficient physical space behind the augmented platform to place the projectors [89] See Fig. 2.2, left).

VPSs are interactive digital simulations of real patients in clinical settings displayed on a screen See Fig. 2.2, left). For example, the Shadow Health VPS keeps CLs engaged with digital patients, and lets them practice communication skills, assessing virtual patients, and documenting their findings [34]. CliniSpace offers both a stand alone healthcare education system and a fully immersive game [8]. i-Human VPS agents are capable of presenting human physiology and pathophysiology, as well as 3D anatomy of the human body [28]. Gabby is a VPS system which provides support to African-American women to decrease their preconception health risks and eliminate racial and ethnic disparities in maternal and child health [249, 62].

VPSs benefit from virtually portraying physiological variables (e.g., heart rate) without being limited by hardware infrastructure. The virtual display also provides the opportunity to richly and quickly display changes in the appearance, symptoms, behavior, or body language. Furthermore, Kononowicz et al. [146] found that VPS systems can help improve knowledge and skill-building (e.g., clinical reasoning, procedural, and teamwork skills) when compared with non-digital educational methods, including didactic-learning modalities (e.g. lectures, reading exercises, group discussion in the classroom), and non-digital models such as SHPs. Another advantage to VPSs is that they make clinical education more accessible to CLs in low resource settings, which Kononowicz et al. [146] discuss as being effective in a range of countries

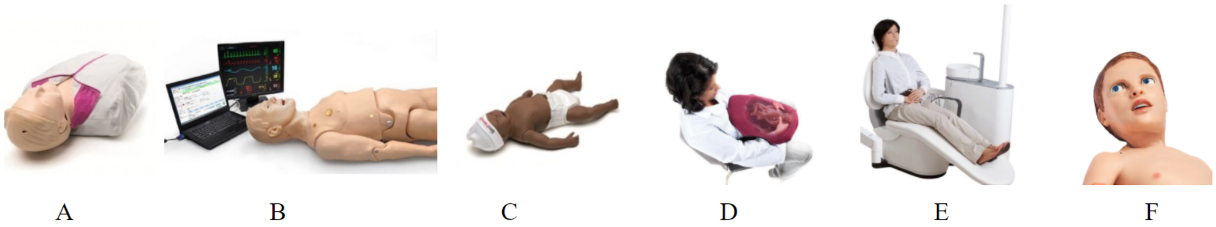


Figure 2.3. Examples of robotic patient simulators (RPSs) with physical faces. **A)** Laerdal's Little Resusci Anne [76], **B)** Code Blue III by Gaumard Scientific [24], **C)** Laerdal's SimNewB [35], **D)** Laerdal's Mama Natalie [29], **E)** Simroid by Morita Corp [36], and **F)** Gaumard's Pediatric HAL [31].

worldwide.

RPSs are lifelike physical robots that can simulate realistic patient physiologies and pathologies (See Fig. 2.3) [196]. The use of physical simulators originated with *Resusci Anne*, a static mannequin created to teach cardiopulmonary resuscitation in 1960. It was used to train more than half a billion people in life-saving skills [182, 15]. Later in the 1960s, in an effort to train anesthesiologists, researchers developed a physical RPS called SimOne, able to show palpable pulses, heart sounds, and movement. Its software provided several pre-programmed events, such as different changes in heart rate or blood pressure [76]. Since then, many companies have built more advanced RPS systems to support a range of clinical scenarios, including Gaumard Scientific and Laerdal.

Recent RPSs benefit from the ability to interactively convey thousands of physiological signals. Their high fidelity physical bodies are comparable to the bodies of real patients, affording CLs a practice platform for physical examinations and procedures.

2.2.3 Open Problems in Simulation-Based Education

Despite the many benefits of using patient simulators, there are several challenges with existing systems that may impede how effective they are at supporting CL education, particularly with regard to skill transfer (how well skills map from simulated patients to real patients).

One main challenge with existing RPS and VPS systems is low usability and controllability, which can cause delay and distraction. These simulators are very complicated and difficult for CEs to control, particularly when running complex simulations in a dynamic learning environment. Running clinical scenarios on these simulators has several time-consuming tasks and requires scheduling. As a result, CEs often cannot run the necessary simulations to support effective learning strategies. Furthermore, clinicians tend to have fairly low technology literacy, so a poorly designed system along with poor socio-technical integration can adversely affect skill learning performance [207]. Finally, using robots in healthcare settings can potentially add disruption and delay to the simulation process, which will change the clinical workflow in unforeseen directions [207, 236].

The other main challenge is that most current commercial VPS and RPS systems suffer from a major design flaw: they completely lack FEs, and thus the ability to convey key diagnostic features of different disorders and social cues, which can eventually cause problems with learner immersion and skill transfer. This is critical for scenarios that require dynamic changes in appearance (e.g., abnormal visual findings such as drooping, which cannot be easily portrayed on a mannequin). Therefore, this lack of expression limits the extent to which a CL will become engaged with and immersed in a simulation, which may adversely affect their learning performance [168]. Consequently, CLs may be learning to incorrectly read patient social cues and signals, and may need to be retrained. Due to the importance of FEs as a key social function and clinical cue in patients, it is essential to study the synthesis of expressions (both symmetric and asymmetric) in simulators.

While RPS, APS, and VPS systems with expressive faces can address the previous challenge, they introduce several technical challenges and opportunities with designing expressive systems. First, because facial expressions and their intensities are very person-dependent and can vary greatly from person to person [275], it can be challenging to develop one generalized system to recognize, model, and display facial expressions of a wide range of different individuals and cultures. Furthermore, some of the simulators, such as VPS systems, are limited by a flat 2D display medium, making them unable to convey a physical 3D human-shape which clinicians can palpate in order to perform clinical assessments. Inaccurately exhibiting symptoms on a simulator's face may reinforce incorrect skills in CLs and eventually lead to incorrect diagnoses in their future career [77].

Other challenges with creating expressive simulators include the need to recruit experts with various skills for development, high development costs, and systematic physical limitations.

Therefore, in order to design robots and virtual agents with human-like expressive faces capable of accurately exhibiting patient-like symptoms, it is beneficial to examine the effect of expressive mechanical or rendered faces. To do this, roboticists and engineers need to closely co-design systems with developers and designers with a range of expertise, and also include a diverse set of stakeholders, including CLs, CEs, and patients [207, 196, 210].

Adopting an interface to a physical or virtual robotic face similar to a human-patient's face to mimic real FEs and symptoms requires knowledge on building and controlling physically-embodied robots and/or animating virtual systems. It also requires having the knowledge on the nature of human facial expressions, and the existing methods of analyzing (recognizing, detecting, and tracking) human facial features. Moreover, it requires knowledge of the existing methods on developing models of human-like facial expressions, and techniques to incorporate and synthesize patient-like FEs onto the simulator's face.

2.3 Chapter Summary

The unique application of facially expressive robots in patient simulation offers a wealth of research opportunities for advancements in medical training and healthcare. This chapter provided a comprehensive overview of the research's objectives, and described the motivation behind my proposed work. It began by identifying the gaps and opportunities in existing learning modalities within the context of healthcare education and training. The benefits and challenges associated with virtual and robotic patient simulators are examined, highlighting the need for alternative learning approaches. The next chapter will introduce the concept of utilizing RPS systems with expressive faces as learning modalities to address major gaps in current healthcare education and training, and will present my efforts in the creation of diversified, expressive physical and virtual faces to be integrated into RPS design.

2.4 Acknowledgments

This chapter contains material from “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future” by M. Pourebadi, and L. D. Riek., which appears in In the Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32, 2022 [195]. I was the primary investigator and author of this paper.

Chapter 3

Expressive Faces For Robots

3.1 Introduction

The human face is a key expressive modality for communicating with others and understanding their intentions and expressions. Facial expressions are a form of visual communication that help to enhance other modalities of communication, such as spoken or gestural language, and enable people to spontaneously communicate important information [171, 71]. In clinical settings, healthcare workers use other non-verbal cues to infer patient physiological states, such as pallor, blinking, eye gaze, blushing, and sweating.

RPS and VPSs with expressive faces also can benefit from this human-like ability to create better connections and interactions with users, and be more favorably perceived [80]. This is why many roboticists develop physical or virtual embodiments capable of displaying facial expressions. Sometimes these expressions are conveyed physically (e.g., with mechanically moving parts), sometimes they are conveyed virtually (e.g., using 2D displays) (See Figure 3.1).

While building accurate physical and virtual platforms for robots can enhance interaction, poorly designed faces can adversely affect the interaction and create distractions [80]. In the 1970s, Mori introduced the uncanny valley concept which explains people's negative reaction to certain lifelike robots [173]. The idea is that as robots become more human-like, they become more attractive until they reach a certain point, after which, people perceive the robots as being creepy and/or immoral. This effect has since been validated across multiple experimental studies

[139, 246].

It is important to consider the variability of facial expressions while designing robotic platforms capable of generating humanlike expressions. For many years, facial expressions were considered a universal language to express internal emotional states across all cultures[133]. However, recent cross-cultural studies suggest that culture is a well-documented source of variance in facial expressions. Studies by Jack et al. [133, 134] and Elfenbein et al. [99] suggest that humans across different cultures communicate emotions using different sets of facial expressions, and therefore, the notion of "universal" facial expressions proposed by Ekman [96] is now refuted in the light of demonstrated cultural nuances.

Another important consideration is the source videos and models used to create expressions on VPS or RPS systems. Many of these systems are trained on datasets of actors, presenting exaggerated facial expressions with little variance or cultural nuances, and tend to propagate the now unfavored Ekman "universal" framing of facial expressions with action units (AU) based models. This can lead to bias and errors in both facial expression analysis and synthesis systems. (See Section 3.2.6).

These studies raise an awareness that the impact of including different facial expressions, features, and functionalities in designing virtual and physical faces requires meticulous attention while designing realistic appearance and performance for human faces. Furthermore, the results suggest to carefully study the effects of using different facial analysis methods before, during, and after the realistic face design process.

In this work, I explore common methods for facial expression analysis (FEA), facial action modeling (FAM), and facial expression synthesis and animation (FSA), present my work on developing physical and virtual faces as a communication modality for robots. In Section 3.2, I explore common methods for detecting and tracking human-like expressions to contextualize facial expression analysis in HRI. In Section 3.3, I describe different facial action modeling techniques while considering various information processing and knowledge modeling methods. Section 3.4, I examine technical approaches to synthesizing dynamic FEs on virtual agents and

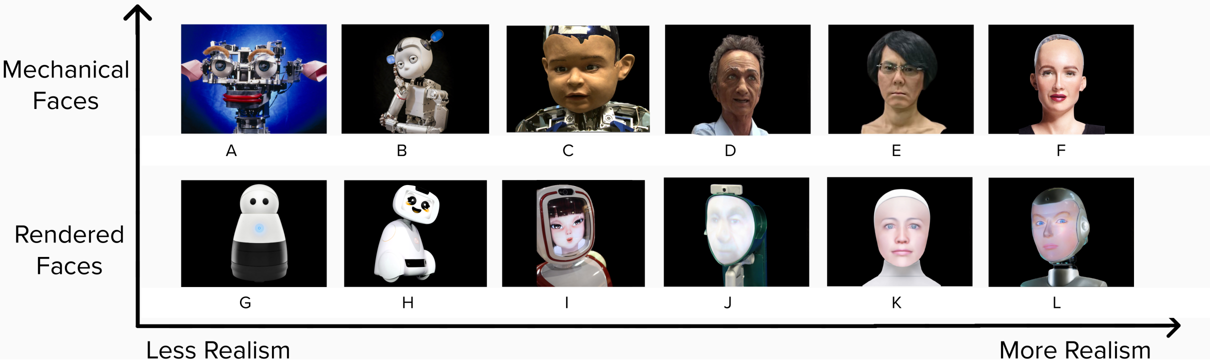


Figure 3.1. Top: Physical robots with mechanical faces: **A)** Kismet [5], **B)** Simon [93], **C)** Diego-San [7], **D)** Charles [149], **E)** Geminoid HI-5 [123], **F)** Sophia [38]. Bottom: virtual and hybrid robots with rendered faces: **G)** Kuri [19], **H)** BUDDY [17], **I)** FURo-D [13], **J)** Mask-Bot 2i [188], **K)** Furhat [80], **L)** Socibot [37].

robots. In Section 3.6, I present robotic faces with diverse expressivity and diversity created by our team.

3.2 Automatic Facial Expression Analysis

In order to build physical and virtual robotic faces that can replicate realistic, understandable, human-like expressions, it is necessary to be able to recognize how people express FEs. This section discusses common methods for manually and automatically detecting, locating, and analyzing human-like expressions in the presence of noise and clutter. First, we list a few key concepts.

Facial landmarks (FL), also known as facial feature points or facial fiducial points, are visually highlighted points in the facial area, mainly located around facial components and contours such as the eyes, mouth, nose and chin.

Facial action units (AUs) are individual components representing the movements of one or several specific facial muscles in each facial component surrounded with specific FLs [12]. Researchers introduced 46 main facial AUs [239] and others have added 8 head movement AUs and 4 eye movement AUs [12]. Examples include AU6-Cheek Riser, AU12-Lip Corner Puller, 5-Upper Lid Raiser, or AU-26 Jaw Drop. In order to express each specific facial expression, people need to move a specific subset of AUs in different facial components of their face. For example, researchers have identified AU6 and AU10 are associated with the expression of pain, and AU 10 with the expression of disgust [168].

Facial action coding system (FACS) is a system for manually describing facial actions according to their appearance, first published in 1978 and later updated in 2002 [96]. The main focus of FACS systems is to recognize facial expression *configuration*, which refers to the combination of AUs. This means that the system associates facial expression changes into a set of facial AUs (out of 46 uniquely defined AUs) that produce them. This system also characterizes the variation of AU *intensity*, which represents the degree of difference between the current state of facial expression and neutral face. [184]. FACS provides a 5-point intensity scale (A-E) for representing the AU intensity (A weakest intensity, and E strongest intensity).

Manual FACS are based on annotations done by trained FACS coders who manually recognize both configuration and intensity of AUs in video recordings of an individual according to AUs described by FACS [96]. However, manual FACS rating requires extensive training, and is subjective and time consuming. Thus, it is impractical for real-time applications [124].

Nowadays, many researchers work on automating FACS systems to analyze AUs [111]. Using automatic FACS instead of a manual approach can be beneficial, because training experts and manually scoring videos is time consuming. Furthermore, studies suggest using automatic FACS can enhance reliability, accuracy, and temporal resolution of facial measurements [161]. In developing these systems, in addition to *configuration* and *intensity* variation, researchers also analyze facial expression *dynamics* (i.e., the timing and the duration of different AUs). Dynamics can be important for human facial movement interpretation [111]. For example, facial expression dynamics can be beneficial for learning complex physiological behavioral states such as different types of pain [257, 266, 265].

The rest of this section briefly describes the main stages involved in automatic FEA, as suggested in a recent survey by Martinez et al. [161], which include: face detection and tracking, facial point detection and tracking, facial feature selection and extraction, AU classification based on extracted features, and new approaches on jointly estimating landmark detection and AU Intensity. Finally, we include a list of facial expression analysis software used by the community.

3.2.1 Face Detection and Tracking

In order to engage in facial expression analysis, systems need to be able to engage in “face localization”, which Deng et al. define as including face detection, alignment, parsing, and dense face localization [92]. Deng et al. introduced RetinaFace [92, 91], “a robust, single-stage, multi-level face detector”. It performs face localization on different scales of the image plane using joint extra-supervised and self-supervised multi-task learning. Many acknowledge that RetinaFace provides one of the most robust and strongest approaches to face detection. Others have made strides on related problems, for example, Hu et al. [130] explored a new approach of training separate detectors for face images with different scales. Their result reduced error by a factor of two compared to prior state-of-the-art methods.

In general, most current methods for face detection employ deep learning techniques, including Cascade-Convolutional Neural Network (CNN) Based Models, region-based Convolutional Neural Network (R-CNN) and Faster Regions with Convolutional Neural Network Features– (Faster-R-CNN) based models, Single Shot Detector Models, and Feature Pyramid Network Based Models, see [112] for a recent survey.

3.2.2 Facial Feature Point Detection and Face Alignment

Facial feature point detection (FFPD) (also known as landmark localization) generally refers to a supervised or semi-supervised process of detecting the locations of FLs. FFPD algorithms are sensitive to facial deformations that can be due to either rigid deformations (e.g., scale, rotation, and translation) or non-rigid deformations (e.g., facial expression variation, head poses, illuminations, noise, clutter, or occlusion) [250, 190]. Enabling FFPD methods to align faces in an input image can lower the effect of changes in face scale as well as in-plane rotation.

Cascaded regression-based methods are one type of FFPD method that recognize either local patches or global facial appearance variations, and directly learn a regression function to map facial appearance to the FL locations of the target image [262]. These methods do not explicitly build any global shape model, but they may implicitly embed the information regarding the global shape constraints (i.e, estimate the shape directly from the appearance without learning any shape model or appearance model).

Deep learning regression-based methods combine deep learning models, such as CNN, with global shape models to enhance performance. Early work in this field employed Cascaded CNNs [232], which predict landmarks in a cascaded way. Researchers then presented Multi-task CNNs [271] to further benefit from multi-task learning to increase the performance rate. Studies show the cascade regression with deep learning (DL) performs better than cascade regression, and cascade regression better than direct regression [262].

In terms of facial feature point detection and face alignment, the Face Alignment Network (FAN) proposed by Bulat and Tzimiropoulos [69]is considered to be the state-of-the-art. They constructed FAN by combining landmark localization with a residual block. They then trained the network on a 2D facial landmark dataset and evaluated it for large-scale 2D and 3D face alignment experiments. Researchers have proposed different follow up methods in order to reduce the complexity of the original approach. For example, MobileNets is a class of efficient models that uses light weight deep neural networks (DNN) to improve the performance [129].

3.2.3 Facial Feature Selection and Extraction

If the number of facial features becomes relatively large in comparison to the number of observations in a dataset, some algorithms may not be able to train models effectively. High dimensional vectors may cause two problems for classifiers: one, data may become sparser in high-dimensional space, and two, too many extracted features may cause overfitting [132]. Traditionally, this was addressed by employing techniques such as PCA or LDA.

Li and Deng [153] provide a recent comprehensive survey on deep facial expression recognition, and include discussion of feature learning and feature extraction techniques. A few examples are briefly discussed below. *CNNs* have been widely employed for the purpose of feature extraction, due to their ability to being robust when encountering facial location changes and variations [106]. For example, researchers in [231] used a region-based CNN (R-CNN) to combine multi-modal texture features for facial expression recognition in the wild. Moreover, researchers [152] proposed a Faster Regions with CNN Features (Faster R-CNN) technique to prevent from the explicit feature extraction step by producing region proposals.

Deep autoencoders (DAE) and their variations have also been used for feature extraction. For example, researchers [150] used the deep sparse autoencoder network (DSAE) on a large dataset of images to prune learned features and develop high-level feature detectors using unlabeled data. The proposed DSAE-based detector is robust to different transformations, including translation, scaling, and rotation. As another example, researchers [211] employed contractive Autoencoder network (CAE) that adds a penalty term to induce locally invariant features, leading to a set of robust features.

3.2.4 Facial Feature Classification

In the classification step, the classifier predicts expressions by categorizing the facial features into different categories. Similar to the facial feature extraction stage, classification performance directly affects the performance of the facial expression recognition system.

Early facial feature classification work used techniques such as Naive Bayes [159, 234], multi-layer perceptrons [194, 65], and SVMs [206], however, these have fallen out of favor given newer deep learning methods. While traditional facial expression analysis approaches usually perform the feature extraction step and the feature classification step independently, deep facial expression analysis approaches are able to perform both steps in an end-to-end training manner by adding a loss layer as the final layer to the DNN to adjust the error, and then directly estimating the probability distribution over a set of classes [153].

For this purpose, many researchers have adapted CNN techniques for expression detection and classification [67, 155, 274]. The results of work done by Zeng et al. [155] shows that CNN classifiers trained faster and performed well. Another study indicates CNN classifiers also provide better accuracy compared to other neural network-based classifiers [206]. One main challenge to some of CNN classifiers is that they are sensitive to occlusion [155].

In addition to using deep neural networks for end-to-end training, other researchers [94, 219, 181, 43] have used DNNs for feature extraction and then added independent classifiers to the system for expression classification.

3.2.5 Jointly Estimating Landmark Detection and Action Unit Intensity

Early FEA work often included a computationally intensive and laborious process (e.g., face and facial landmark detection, hand-crafted feature extraction, and limited classification methods). Nowadays, researchers benefit from having access to comprehensive, large-scale facial data sets, as well as advanced computing resources to develop more efficient facial analysis methods [83, 103, 102, 144, 154, 156].

One line of research is the work done on jointly estimating landmark and action unit intensity. For example, Wu et al. [261] proposed a constrained joint cascade regression framework to simultaneously perform landmark detection and AU intensity measurement. This method learns a constraint to model the correlation between AUs and face shapes. Next, they use the learnt constraint as well as the proposed framework to estimate the landmark location and recognize AUs. The results of the study suggests the connection between these two parameters can improve the performance for both tasks.

Furthermore, many researchers consider the work done by Ntinou et al. [180] as the state-of-the-art method for jointly estimating landmark localization and AU intensity. In this work, researchers employed heatmap regression to model the the existence of an AU at specific location. For this purpose, they used a transfer learning technique between the face alignment network and the AU network.

It is worth mentioning that the newer directions for estimating AU intensity seek learning models with little or no supervision, including work done by Sanchez et al. [217], Wang and Peng [252], Wang et al. [251], and Zhang et al. [273].

One of the applications for AU intensity estimation is to further analyze and synthesize facial expressions representing specific feelings, such as pain. Many researchers have already conducted studies that indicate there is a relationship between a combination of AUs and pain, including work done by Xu et al. [266, 265, 267, 264], Kaltwang et al. [140] and Werner et al. [256]. Furthermore, it is worth mentioning that a fully functional automatic pain estimation system requires enough representative data, and for that purpose, there are some pain datasets publicly available (c.f. [158]).

3.2.6 Facial Expression Analysis Software

Dynamic facial expression analysis (FEA) systems integrate automatic FACS to assess human expressions. Several commercial and open-source FEA software packages are available, including iMotions, AFFDEX, FaceReader, IntraFace, and OpenFace 2.0.

iMotions developed a commercial tool for FEA that offers assessing FEs in combination with EEG, GSR, EMG, ECG, and eye tracking [22]. This tool lets users record videos with a mobile phone camera or laptop webcam, and then detects changes in FLs. The researcher can set the tool to apply either the AFFDEX algorithm by Affectiva Inc. [98] or the Computer Expression Recognition Toolbox (CERT) algorithm used by FaceReader tool [157] to classify expressions. Different classifier algorithms such as CERT and AFFDEX employ various facial datasets, FLs, and statistical models to train the ML system to perform the classification task [22].

Affectiva's AFFDEX software developer kit (SDK) [166] is a commercially available real-time facial expression coding toolkit which is able to simultaneously recognize the expressions of several people, and is available across different platforms (IOS, Windows, Android). AFFDEX algorithm uses Viola-Jones [248] for detecting a face and identifying 34 landmarks, Histogram of Oriented Gradient (HOG) to extract facial textures, SVM classifiers to classify facial action and finally code seven facial expressions based on combinations of facial according to FACS [22]. AffdexMe is the name of the IOS-based AFFDEX SDK which enables developers to emotion-enable their own apps and digital experiences. The tests we performed on the trial version of this SDK show that the app can efficiently analyze and respond to seven basic emotions in real-time.

FaceReader [21] is a commercially available automated expression analysis system developed by Noldus. It enables developers to integrate expression recognition software with eye tracking data and physiology data. This tool provides an assessment of seven expressions, head orientation, gaze direction, AUs, heart rate, valence and arousal, and person characteristics.

FaceReader's algorithm uses the Viola-Jones algorithm [248] to find a face, then makes a 3D face model using facial points and face texture. It then analyzes the face using deep learning methods, and classifies the expressions using an ANN. Studies show that FaceReader is more robust than AFFDEX [226].

IntraFace is a software package developed by De La Torres et. al. [90] for automated facial feature tracking, head pose estimation, facial attribute recognition, and facial expression analysis. This package also includes an unsupervised technique for synchrony detection that supports the function of discovering correlated facial behavior between two people.

IntraFace uses the SDM method to extract and track facial feature landmarks, and normalize the image with respect to scale and rotation [90]. They then extract HoG features at each landmark and perform a linear SVM for classifying facial attributes. Finally, they use the Selective Transfer Machine (STM) learning approach to classify facial expressions and AUs.

OpenFace 2.0 is an open source and cross-platform tool for facial behavior analysis released by the Multimodal Communication and Machine Learning Laboratory (MultiComp Lab) at Carnegie Mellon University in 2018 [10]. OpenFace 2.0 is capable of performing facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation in real-time [52].

OpenFace 2.0 uses a newly developed Convolutional Experts Constrained Local Model

[268] and optimized FFPD algorithm for facial landmark detection and tracking which enables real-time performance [52]. Using this approach also enables OpenFace 2.0 to cope with challenges such as non-frontal or occluded faces and low illumination conditions. The algorithm of this tool is able to operate on recorded video files, image sequences, individual images, and real-time video data from a webcam without any specialist hardware. GANimation [200, 201] is an anatomically-aware facial synthesis method that automatically generates anatomical facial expression movements from a single image. This method provides the opportunity to control the magnitude of activation of each AU and combine several of them.

Latent-pose-reenactment [72] uses latent pose descriptors for neural head reenactment. This system can use videos of a random person and map their expressions to generate realistic reenactments of random talking heads.

3.3 Facial Action Modeling for Synthesis

In many robotics and AI applications, in addition to recognizing FEs in people, we also need the ability to synthesize them on robotic and virtual characters. We discuss this further in Section 3.4, however, it is first important to discuss facial modeling.

Facial action modeling (FAM) builds a bridge between facial analysis (recognizing and tracking facial movements) and facial expression synthesis (translating modeled FEs onto an embodied face and animating its facial components) [216]. Thus, technology developers need to incorporate two key ideas in the design of face models: 1) Patterns that model the human face (e.g, shape, appearance), both in its neutral state and the way facial movements (i.e., AUs) change to display different expressions. 2) Patterns of the temporal aspects of facial deformation (e.g., acceleration, peak, amplitude).

The complexity of facial modeling can vary based on the degrees of freedom (DoF) of the embodiment (e.g., a mechanical robot or virtual face). It is less complex to build face models for more machine-like robots with very simple faces, such as Jibo [26] which only has one eye with varying properties and details. The complexity of designing a face model increases as the face becomes more realistic and detailed. For both robots with hyper realistic faces (e.g., Charles [210], Geminoid HI-2 [177]), or a human-like computer-generated virtual face (e.g., Furhat [23]), developers need to design highly accurate models in order to engage in synthesis.

There are two groups of information processing strategies for face modeling: theory-driven modeling and data-driven modeling [135].

3.3.1 Theory-Driven Modeling Methods

Ekman and Friesen's FACS theory [96] describes the facial movements through observing the effect of each facial muscle on facial appearance, and decomposes the visible movements of the face in the form of 46 AUs. Formerly, many researchers adopted FACS theory for facial modeling and embedded FEs derived from this theory constrained into their social robots [126, 66]. In this approach, programmers selected a small set of (static) FEs (e.g., tightening and slightly raising the corner of the lip unilaterally to express contempt) [97]. They then asked actors to contract *k* different combinations of muscle AUs to display the selected FEs to generate *k* different face images, and score the face with FACS to verify muscle AUs depicted in each image. Finally they asked observers to select which image better mimics each specific FEs, and therefore identify which combinations of muscle AUs are signals for each specific FE.

However, there are several challenges with the theory-driven modeling methods. For one, these models are based on FEs that precisely met criteria selected and specified by researchers [97]. Moreover, since these models are based on static FEs, they lack dynamical data including the temporal order of FE movements (e.g., acceleration, peak, amplitude) [135], resulting in less realistic facial models and ultimately less human-lik simulators. Furthermore, even in studies on cross-cultural FE analysis where subjects pose cultural-specific expressions, still most subjects are identified as Westerners [175], leading to less diverse face models. Finally, people may have asymmetric facial expressions, such as people who have facial paralysis or deformities are rarely included, thus also limiting the diversity of facial models [172]. As a result, expressive physical and virtual robotic faces developed using theory-driven modeling methods lack the ability to generate a wide range of FEs. Therefore, these embodiments are not able to adequately communicate and interact with users.

3.3.2 Data-driven Modeling Methods

To address the gaps associated with theory-driven methods, researchers have proposed data-driven modeling methods (or, example-based deformation models) to computationally model (dynamic) FEs based on real data. Data-driven modeling methods usually consist of three main steps: data collection, facial expression and intensity data labeling, and facial expression model creation [135].

Data Collection

Data is generally collected in one of two ways: via recordings of human participants, and through the use of artificial data creation.

One way of collecting data is to capture videos of facial expressions of human subjects (e.g., via an actor or layperson performing facial movements, or use of existing datasets). In this method, a researcher can use any statistical analysis method or facial expression analysis software package (See Section 3.2.6) to derive a parametric representation of facial deformations and identify the AUs correlated to each frame of a video. For example, Wang et al [253] created a new FE dataset of over 200 thousand images with 119 persons, 4 poses and 54 expressions, which is about enough to evaluate the effects of unbalanced poses, expressions on the performance of the FE tasks.

Another way of collecting data is by generating artificial data through artificial data creation methods. In this method, developers usually use facial movement generators to randomly-generate an enormous range of artificial dynamic facial expression videos. For example, Jack et al. use a facial movement generator, which randomly selects a subset of AUs, assigns a random movement to each AU by setting random values for each temporal parameter, combines randomly activated AUs, and finally projects them to a robotic face to create random facial animation videos [79].

Facial Expression and Intensity Data Labeling

Researchers have used different techniques for labeling FE data correlated to each frame of videos and their intensities, including manual labeling by both lay participants and domain experts, and unsupervised data labeling via use of machine learning.

For instance, Jack et al. [79] recruited participants to watch videos of facial expressions. If the projected video formed a pattern that correlated with the perceivers' prior knowledge of one of six expressions, they manually assigned a label to identify the expression and its intensity rating accordingly. Other researchers working on labeling FEs use domain experts (e.g., clinicians) to manually label data [172]. Other researchers develop facial expression datasets that use different semi-supervised or unsupervised techniques to label the data [253].

Facial Expression Model Creation

The next step is the learning phase, where the system uses the shape and texture variations of several sample images in datasets to build a face model and generate its appearance parameters. The parameters of the face model are reversible, meaning that they represent the shape and the texture of all images in the dataset, and therefore, are able to regenerate realistic images similar to each of the learned sample images. Thus, researchers can reverse-engineer specific dynamic FE patterns. This helps to derive the unique patterns of correlated AUs that are activated over time, which are correlated with human perception of each expression. For example, Chen et al. [79] developed their models by calculating a 41-dimensional binary vector per emotion detailing all AUs, and also seven values detailing the temporal parameters of each AU .

Using these three steps, developers can learn and build mathematical models of the dynamic FEs within a video stream that make it possible to reconstruct these FEs on a robot or virtual agent's face and animate them later [79].

3.4 Facial Expression Synthesis and Animation

Facial expression synthesis and animation (FSA) refers to techniques used to animate dynamic expressions on the faces of virtual agents or robots using previously developed face models. FSA techniques provide the facial movement vocabulary that maps the developed model of AU movements and densities into the mesh topology of the social robot or virtual agent heads [189]. Using this technique makes the simulated face able to display AU movements corresponding to developed facial expression models. Concerning FSA, many articles have reviewed state-of-the-art methods and techniques, including [101, 189, 216].

3.4.1 FSA Technical Approaches

Existing surveys in facial expression synthesis and animation include [151], [216], and [101]. The surveys suggest there are three primary categories of techniques for synthesis purposes: skeletal-based, shape blend-based, and performance-driven approaches. Table 3.1 provides a summary of common approaches, which are further discussed below.

Skeletal-based approach (also known as the key-framing approach) works by rigging a skeletal model using an interactive tool to mimic the contraction of facial muscles and generate synthetic facial movements[101, 27]. For this purpose, animators use the 3D rigging tool first to construct a rig of bones and joints based on an estimation of the locations of facial muscles. They manually define the combinations of muscles representing each and every facial expression, and associate each bone into different parts of the virtual agent's visual presentation accordingly. Using this mapping, animators can automatically animate the virtual face using skeletal motion data.

Animating a virtual model using this approach is less labor-intensive, as animators only need to manipulate a set of vertices (bones) instead of each individual vertex. However, the downside of the skeletal-based approach is that generating the accurate mapping between the bones with facial parts is labor- and time-consuming. Furthermore, because it is difficult to accurately model facial movements based on bone movements, using this method can generate unrealistic artificial-looking animations and lead to inaccurate synthesis and unrealistic FEs on a virtual robot [216].

Blend-shape approach works by creating a number of main mesh topologies of the expressions and poses examples collected from the face of a real subject (one for each main expression), and then using an automatic interpolation function to linearly blending these topologies to create a smooth transition between them [216]. In order to achieve smooth animations, animation developers need to generate hundreds of blended topologies.

This approach is commonly used to animate virtual faces as it benefits from low computational time and is easy to implement. However, the performance of this approach greatly depends on the existing examples of different expressions [27]. Furthermore, this method only provides synthetic FEs in between the existing examples [101]. Furthermore, manually designing the main mesh topologies and manipulating each vertex to create animations is labor-intensive and time-consuming, making it an inconvenient modeling technique for creating real-time, long animations [216].

Parameter-based approach (also known as Motion Capture or Performance-driven approach) uses a system of sensors and cameras to record motions and FE movements of a subject [216]. It then learns the face and deformation parameters from the captured data (including visual or physical effects of muscle actions) and finally transfers synthetic FEs onto the virtual robot's face.

In comparison with the other two methods, the performance-driven approach has the potential to be more realistic [27]. The use of parameter-based models makes it possible to create a wide range of deformations. These techniques also support creating interactive animations by incorporating text, audio, or video data in the model developing process [216]. However, in order to get the best and most accurate simulation using this method, it is necessary to use lots of high-quality motion capture equipment. Although this method is greatly used by major film making companies, it is not a convenient approach for technology developers and animators

[101].

3.4.2 Advanced FSA Methods

Recently, researchers have performed more research-oriented studies of facial expression generation, that reflect ongoing attempts to address several of the challenges with respect to the expressivity of a facial expression synthesis system. More specifically, recent studies have focused on automatically synthesizing facial expressions from a few or single images using the newest advances in Generative and Adversarial Networks (GAN).

For example, Pumarola et al. [200, 201] introduced GANimation to automatically generate facial expressions in a continuous domain, without using any facial landmarks. They conditioned the network on a one-dimensional vector that represents the existence and the magnitude of each AU. This provides the opportunity to control the magnitude of activation of each AU and combine several of them. Additionally, they trained the network in a fully unsupervised manner, only requiring images annotated with their activated AUs, leading to an approach that is robust to changing backgrounds and lighting conditions.

In addition, other recent work addresses face reenactment and synthesis in a landmark-driven way. For instance, Burkov et al. [72] recently proposed a “neural head reenactment system” which uses a latent pose representation, based solely on image reconstruction losses. This system can use videos of a random person and maps their expressions to generate realistic reenactments of random talking heads.

Another recent work in this field is by Zakharov et al.[269], who developed a system that can generate plausible video sequences of speech expressions and mimicry of a particular person. They use a deep network that combines adversarial fine-tuning into a meta-learning framework to train lifelike digital speaking heads based on only a few photos of a person (e.g., a few-shot approach). This model can generate photorealistic animations of both random people and portrait paintings.

Gecer et al. [110] proposed a novel multi-branch GAN architecture that synthesizes

Table 3.1. An overview of technical approaches for the purpose of facial expression synthesis and animation [151, 216, 101, 27].

Categories	Process	Benefits	Drawbacks
Skeletal-based	Associates each bone and joint to various facial parts via a rigged skeletal model, animated with skeletal motion data.	Reduced labor as animators only need to manipulate a set of vertices (bones) instead of each individual vertex.	Time-consuming to create accurate bone to facial part mappings. May lead to artificial-looking animations and inaccurate synthesis of FEs.
Blend-shapes	Creates key mesh facial topologies and uses interpolation for smooth transition among them.	Low computational time and easy implementation.	Requires a large number of key topologies of different expressions. Only provides synthetic FEs in between the existing examples. Mesh design and animation creation is labor and time-consuming. Not suitable for real-time applications.
Parametric-based	Uses a parameter system for creating the face and deformation models, based on visual or physical effect of muscle actions.	Creates realistic animations, capable of creating various deformations. Enables creation of interactive animations with text, audio, or video data.	Requires high-quality motion capture equipment. Real-time performance is usually not feasible.

photo-realistic expressions. It adopts a multimodal approach by including multiple 3D features (e.g., shape, texture, normals, etc). They then trained the network to generate all modalities in a local and global correspondence, and condition the GAN by expression labels to create 3D faces with various expressions.

OpenPose, proposed by Cao et al. [73], is a open-source, real-time system that detects the 2D pose (including the face) of multiple people in a single image. It employs a non-parametric representation in order to learn which body or facial parts is related to which person in the image. The system achieves high accuracy and real-time performance, regardless of the number of people.

3.4.3 FSA Exemplar

Researchers have mapped the synthesized motions to the face of different embodiments using FSA software packages (See Fig. 3.2). For example, Faceposer SDK [20] for the Steam Source engine [39] is a virtual platform that uses synthesis framework to transfer facial expressions and skeletal animations to a virtual character's control points for animation. After generating facial movements and transformation parameters from a source video using one of the methods described in Section 3.3, Faceposer's synthesis framework converts the parameters into

21 control points (Flex sliders)r. The system saves the values of the Flex sliders in a .VCD scene file consisting of a header section with date, simulator, and timescale information; a variable definition section; and a value change section. Finally, after importing the .VCD file to the Faceposer SDK as the input, the SDK transfers the FEs on a virtual agent's face accordingly and animates the virtual agent.

Moreover, Pelachaud [187, 186] introduced Greta, which is a conversing socio-emotional virtual agent. This agent's software provides users with a real-time platform to control socio-emotional virtual characters and develop natural interaction with humans. Greta animation engine receives body animation parameters and facial animation parameters as inputs, and synthesizes the expressions on a virtual character using Ogre3D or Unity3D [25].

Furthermore, Chen et al. [79] introduced a social physical-virtual agent displayed on a Furhat robot [23], which is capable of re-displaying facial expression using state-of-the-art 3D animation techniques. The introduced agent's algorithm provides full control over face designs, and includes realistic lip movements, as well as high-level control over the eyes and other facial movements [23]. It also provides the user with the opportunity to change the projected face's ethnicity, gender, language, and even its species. In order to measure the humanlikeness of their synthesis approach, they performed an experiment to compare two FE synthesis methods (one generated through their reverse-engineering and synthesizing method, and one manually pre-programmed on their social robot). Their results suggest that users perceived their reverse-engineered expressions as more humanlike than the existing expressions of the robot [79].

Charles is a humanoid, hyper-realistic robot head from Hanson Robotics [210, 208]. Charles is able to display lifelike human expressions as it has wrinkles on the skin and 22 degrees of freedom (DOF) in the face and neck. The robot has microcontrollers to control the motors that move the brow, eyes, midface, lips, mouth, jaw, head, and neck. Its control system generates motions using a direct AU-to-motor mapping system to synthesize expressions.

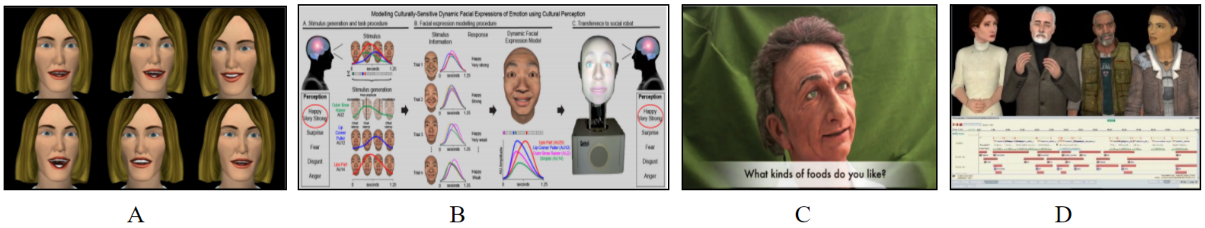


Figure 3.2. A) Figure of the Greta virtual agent [25, 186]. B) Figure of synthesizing dynamic facial expressions onto the Furhat robot [80]. C) Figure of the Charles robot mimicking a human [210]. D) Figure of the Faceposer software interface [20].

3.5 Ethical Considerations

Using FEA and FSA technologies to develop new RPS and VPS systems and integrating them within clinical learning contexts presents a number of ethical and social challenges that require specific attention. It is important researchers and technology developers carefully consider these challenges, and work to design inclusive technologies to avoid unintended consequences. While this is by no means an exhaustive list, a few key challenges are highlighted herein.

3.5.1 Racial and Ethnic Bias in FEA technologies

There are many concerns regarding racial, ethnic, misogynistic, and ableist biases in FEA technologies, which can perpetuate social and fiscal oppression [179, 178, 61]. For example, many studies show high rate of misidentifying blacks by recognition systems, which can be due using FEA algorithms trained on a racially biased datasets, as well as systemic biases embedded within the systems themselves [70]. Such biased models can then affect FSA, and further perpetuating biases in clinical education [221]. Moreover, there are challenges regarding distancing and dividing effects caused by using FEA systems for controlling patient simulators. For example, an operator of an expressive robot sometimes need to adjust their feelings to express exaggerated facial expressions (e.g., intense smile) or fake facial expressions (e.g., reflecting different feeling than what they genuinely feel at the moment), so the FEA algorithm can detect and/or track the expression. Although some researchers think these adjustments may only cause minor problems or difficulties, others think using these technologies can distance and dehumanise people [46].

3.5.2 Privacy

Another concern is on privacy and the extensive use of data in FEA and FSA systems. Widespread use of these systems in healthcare settings can lead to the collection of large amounts of patients' and clinical workers' actions, locations, personal, physiological, and behavioral information. This can raise many concerns about the ways of protecting the privacy of collected personal data, as well as the ways simulator developers use the data.

3.5.3 Uncanny Valley

Another concern that often arises with highly humanlike RPS and VPS systems is a phenomenon called the Uncanny Valley [173]. This is a theory that suggests that as robots become more humanlike they are more attractive, until they reach a certain point, where people's affinity for these humanlike robots descends into a feeling of strangeness and unease [173, 139]. This is reflected in both their appearance and their behavior [218]. While CLs require highly humanlike RPS and VPS systems to learn proper clinical skills, ones that miss the mark can cause learner distress, and adversely affect their learning, Thus, RPS/VPS designers should carefully consider learners' perceptions as part of their design process.

3.5.4 Risks and Benefits of Diverse FSA

Just like humans, human-like patient simulators that resemble a certain gender, race, or culture in their design can face judgement and aggression based on the biases towards such social identities. Designing human-like robots with diverse appearance and behavior has numerous benefits. For example, building a human-like robot resembling a patient who has had a stroke for healthcare education application provides the clinical learners with a great opportunity to practice their communication and procedural skills on these robots, preparing them for treating real human patients with stroke in their future careers [197, 192].

However, diversifying the appearance and behavior for simulators also introduces risks. For example, roboticists may implicitly or explicitly reinforce gender biases by assigning a specific gender to the robot during the design process, and CLs and CEs might as well during simulation sessions [221].

People also more readily dehumanize robots racialized in the likeness of marginalized social identities than those racialized White [227]. As such, people with racist behavioral biases represented similar racist biases while interacting with human-like RPS or VPS systems of a similar race.

3.6 Creating New Embodiments for Robots and Virtual Agents

In dynamic, real-world environments such as HET, social robots need to have realistic human-like faces capable of accurately exhibiting verbal and non-verbal cues. However, many existing RPS systems have limited to no capabilities for human-like expressiveness in their faces, which may impede emotional engagement, empathy, and social presence, leading CLs to experience reduced motivation, interest, and retention of training content [185]. Thus, our research concentrated on developing nuanced patient simulator faces for virtual agents and robots, characterized by diverse appearances, backgrounds, and the capacity to exhibit sophisticated verbal and non-verbal cues. The core aim was to leverage the expressivity and diversification of these embodiments, enhancing their potential to improve outcomes in their respective application domains [196, 172, 195, 191, 193] (See Fig. 4.1).

3.6.1 Approach

For this purpose, we created different virtual agents with diverse ethnic backgrounds and genders using the aforementioned Source SDK tool [172] and Furhat virtual SDK [191]. Furthermore, we supported the redesign of our team's bespoke robotic head and a low-cost expressive face[209] by increasing its DOF to 21 and performing iterative experimentation to

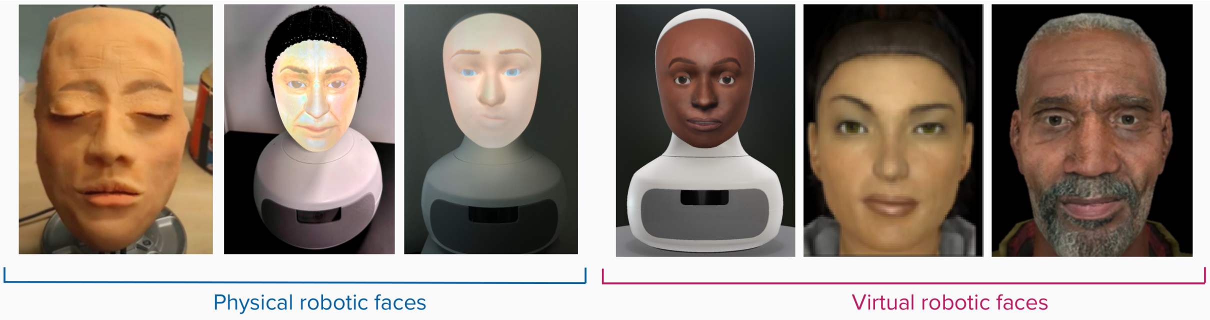


Figure 3.3. Some examples of physical and virtual robotic faces with diverse expressivity and diversity created by our team.

increase their realism and efficacy [196]. Moreover, as per our final study, we used the Furhat robot from Furhat Robotics [3] as the physical platform, enabling the system to perform real time rendering of dynamic facial expressions, head movements, and speech [3, 191].

Ultimately, we expanded the diversity of these expressive faces, crafting designs that represent a wide array of backgrounds and accurately portray different age groups, genders, and races. Some examples of our creations are displayed in Figure 3.3.

3.6.2 Results

The use of virtual agents and expressive robotic faces has shown promising results in terms of diversification and expressivity. Expressive robotic faces have been developed to showcase a range of facial expressions in order to foster more natural conversations. These features are key components in improving the user experience and promoting dialogue between humans and machines.

The redesign of the robotic head, with an increased degree of freedom (DOF) to 21, has significantly improved the robot's expressivity. The iterative experimentation resulted in a more realistic and effective robotic embodiment capable of conveying dynamic and nuanced human-like expressions. These developments have led our work to meet, and in certain areas exceed, the current state of the art in expressive robotics.

Additionally, the representation of various age groups, genders, and races in the expressive faces of the robots has added a layer of inclusivity and diversity to the project. The ability to emulate diverse backgrounds allows the robots to connect with users on a deeper level, improving their efficiency in health and safety applications.

This work stands as a potential transformative instrument in HET, opening new frontiers in developing expressive RPS systems. Moreover, this work provides valuable insights to researchers by examining methods for detecting, modeling, and synthesizing FEs, with potential applications in enhancing social interactions, knowledge modeling, and education.

3.6.3 Discussion

The utilization of expressive robotic and virtual faces has been demonstrated to be effective in many applications, yet areas still require more thorough investigation. For instance, real-time performance and integration into complex systems must be further evaluated, especially within high-stress scenarios. To ensure user satisfaction, long-term interaction studies should be conducted to analyze user preferences and assess design improvements. Furthermore, exploring creative functions and incorporating machine-learning techniques may expand the capabilities of these embodiments.

In conclusion, this work has demonstrated the benefits of using diversified embodiments to improve the expressivity and inclusivity of physical and virtual robots. This success indicates that further research into such approaches could result in improved robotic face designs, providing greater user satisfaction and efficacy while engaging in human-robot interactions. The potential benefits of this research are substantial, with implications for increased efficacy and usability across a wide variety of domains.

3.7 Chapter Summary

This chapter investigated the effect of expressive mechanical and rendered faces in RPS design, introduced the concept of human-like RPS learning modalities as a potential solution to address major gaps in current practices, and presented my work on building new expressive faces. The next chapter will present the foundations and frameworks for designing expressive RPS systems to support researchers in developing and deploying systems capable of effectively depicting clinical conditions.

3.8 Acknowledgments

This chapter contains material from “Expressive Robotic Patient Simulators for Clinical Education”, by M. Pourebadi, and L. D. Riek, which appears in the Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI), Robots 4 Learning workshop, 2018 [195], and material from “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future” by M. Pourebadi, and L. D. Riek., which appears in In the Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32, 2022 [195]. I was the primary investigator and author of these papers.

Chapter 4

Frameworks Development

4.1 Introduction

Every year, millions of individuals experience conditions such as stroke, Parkinson's disease, Moebius syndrome, and Bell's palsy, leading to facial paralysis and A-FEs. People's misperceptions and biased impressions can make it challenging for them to interact socially with and understand the emotions of people with A-FEs These misperceptions in clinical settings can adversely impact the quality of care provided to FP patients. This highlights the need for new training tools to enhance clinicians' interaction skills and improve care for individuals with facial paralysis. The lack of exploration in using FP patient simulators highlights the need for researchers to develop training tools that consider individuals with FP, aiming to enhance clinician skills in avoiding biased impressions, improve clinical communication, and deliver better care for this population.

For the past decade, our team has had numerous projects that address crucial aspects of facial expression modeling and synthesis in expressive simulation technologies for healthcare education (See Figure 4.1). This chapter presents the foundations and frameworks required to design expressive RPS systems, capable of depicting A-FE on their faces based on real patient's facial characteristics. Particularly, this chapter introduces 1) a new end-to-end control framework for integrating three systems presented in Chapter 3 (FEA, FAM, and FSA) to more robustly transfer human-like expressions from a subject's face onto an expressive robotic face, 2) a new

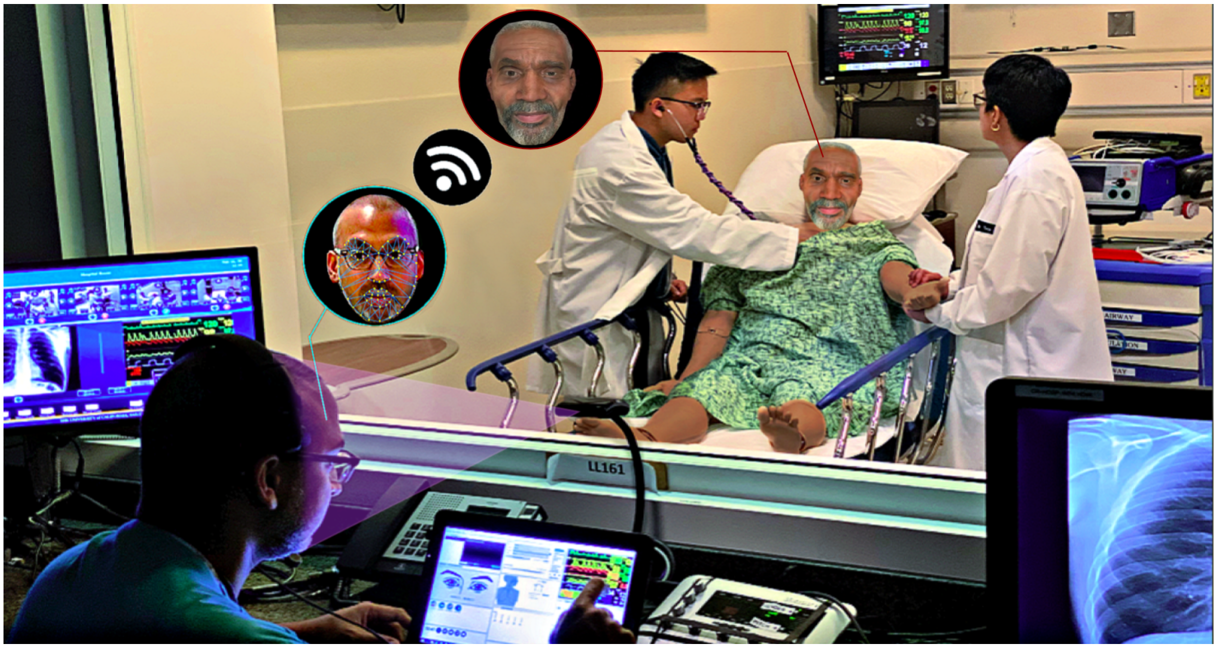


Figure 4.1. Robotic patient simulators are tele-operated, life-size mannequins that can exhibit thousands of physiological signals, and can breathe, bleed, and respond to medications. However, they are largely inexpressive, leading to poor training outcomes for CLs, and possibly poor clinical outcomes for patients. Our work addresses this gap by introducing patient simulator systems with a much wider range of expressivity, including the ability to express pain, neurological impairment (e.g. stroke, Bell's Palsy), and other clinically-relevant expressions, via simulators with diverse genders, races, and ages.

computational framework for modeling a range of clinical conditions, and 3) the connection between these two frameworks [210, 169, 168, 172, 192, 209, 171, 196, 197]. We briefly discuss this work below.

4.2 Analysis-Masking-Synthesis Framework Development

In addition to building robotic and virtual embodiments, we designed and developed an end-to-end Analysis-Masking-Synthesis (AMS) framework, which included: FEA, FAM, and FSA systems (See Figure 4.2). Modularizing the AMS framework into three components allowed for a more organized and encapsulated structure of the framework. The FEA component enables the AMS framework to more robustly detect and track FE movements in real time. The FAM component overlays a computational representation of a clinical condition onto the tracked

FE movements. Finally, the FSA component automatically synthesizes facial movements onto the face of robotic and virtual simulators with different ages, races, and genders, and animates their facial components.

The contributions of this work are as follows. First, we extended an FEA system previously developed by our team [171] to improve automatic FACS ratings of facial AUs. The extended FEA system benefits from preprocessing techniques such as noise reduction and facial alignment techniques to diminish the effects of facial deformations, including translation, rotation, and distance to the camera. Next, a CLM-based tracker [87] is used in the FEA system, as it is robust to illumination and occlusion. This tracker robustly locates the FL locations on an input frame based on the global statistical shape models and the independent local appearance information around each landmark.

Second, we proposed a novel data-driven FAM system developed in three steps: first, we collected real dynamic facial expression data, second, we labeled the FE data correlated to each frame of videos and their intensities using manual outsourcing technique, and third, we generated reversible appearance parameters by calculating a 46-dimensional binary vector detailing all AUs. This FAM system makes it possible to computationally model dynamic FEs tracked by the FEA system based on real human facial expression data, which ultimately can make it easier for developers to generate a diverse set of realistic face models derived from real patients.

Third, we extended an FSA system previously developed by our team [171] for synthesizing realistic, patient-like FEs on both our bespoke RPS head and virtual agent faces. The method is based on data-driven synthesis, which maps motion from video of an operator/CE onto the face of an embodiment (e.g., virtual agent or physical robot). This platform-independent software makes it possible for SMs to easily and robustly synthesize and animate realistic expressions on the faces of a range of embodiments, and makes it easy for CEs to perform simulation.

This end-to-end AMS framework models and synthesizes patient- data-driven facial expressions, and can easily and robustly map these expressions onto both simulated and robotic faces. By leveraging this work, other roboticists and engineers will be able to discover platform-

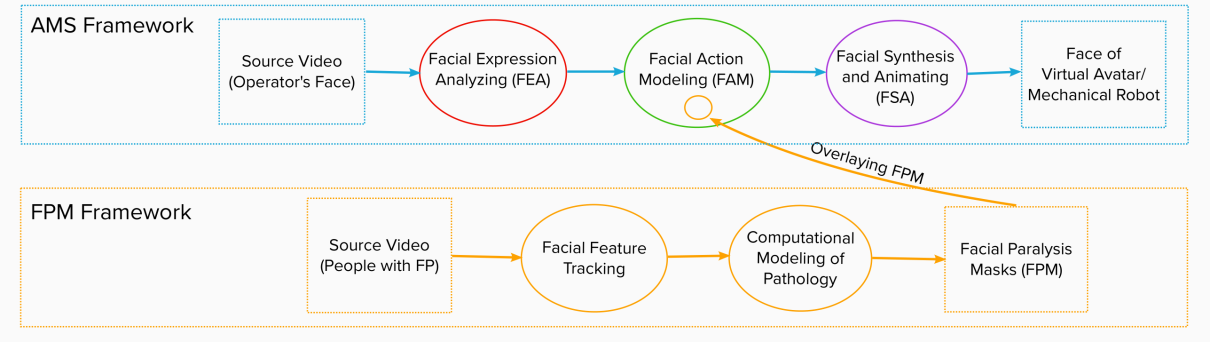

AMS Framework
Source Video (Operator's Face)
Facial Expression Analyzing (FEA)
Facial Action Modeling (FAM)
Facial Synthesis and Animating (FSA)
Face of Virtual Avatar/ Mechanical Robot
Overlaying FPM
FPM Framework
Source Video (People with FP)
Facial Feature Tracking
Computational Modeling of Pathology
Facial Paralysis Masks (FPM)

Figure 4.2. In our work, we have developed of an end-to-end Analysis-Masking-Synthesis (AMS) framework to recognize, model, and synthesize facial expressions of real humans to the face of a physical or virtual robotic head. The AMS framework integrates three systems presented in Chapter 3 (FEA, FAM, and FSA). Furthermore, we developed a novel Facial Paralysis Masks (FPM) framework to build accurate computational models of people with Bell's Palsy that are constructible in real time.

independent methods to control the FEs of both robots and virtual agents. This can also help improve how clinicians interact with patients, and increase their cultural competence when interacting with patients from diverse backgrounds.

4.3 Facial Paralysis Mask Framework Development

Every year, 22 million people experience Bell's Palsy, stroke, Parkinson's disease, and Moebius syndrome [6, 244, 160], leading to facial palsy (FP). FP is the inability to move one's facial muscles on the affected side of the face, leading to asymmetric facial expressions (A-FEs) [58]. Studies show observers perceive the emotions of a person with FP differently from their actual emotional states [233]. For example, people with severe FP are perceived as less happy than people with mild FP [64]. People's misperceptions and biased impressions of FP can make it challenging for them to interact socially with and understand the emotions of people with A-FEs.

In clinical contexts, these misperceptions can lead to poor care delivery. Healthcare providers frequently have negatively biased impressions of patients with facial nerve paralysis [240], which may adversely affect the quality of care they receive [213, 210]. If a patient and a healthcare provider do not communicate effectively, there is a higher chance that their treatment will be unsuccessful [233, 48]. This calls for the development of new training tools to enable CLs to practice their interaction with FP patients, and improve how clinicians calibrate their perception of asymmetric expressions.

However, prior development of facially expressive RPS systems was based on the assumption that human faces are structurally symmetric, and thus have not accounted for expressing A-FEs. Due to the large number of people affected by FP, it is important to also explore synthesizing A-FEs in clinical contexts. To our knowledge, FP patient simulators have not been explored in this way.

For this purpose, we introduced the concept of facial paralysis mask (FPM) framework to provide a platform to generate accurate A-FEs representations for patient simulators based on real patients' facial characteristics, situated within a clinical education context. FPMs are computational models of different pathologies derived from recognized expressions of real people with FP.

Finally, we integrated these two frameworks by overlaying pre-built FPMs on the facial model of the AMS framework described in Section 4.2 to recreate A-FEs on RPS faces. Tthe AMS framework utilizes the results of the FPM framework and enables the system to robustly recognize the facial movements of a human operator, mask the generated model on tracked movements, and automatically synthesize the generated models of FEs across a range of RPS embodiments, thereby animating their facial components.

For the past decade, our team has developed new methods for modeling a range of clinically-relevant conditions, including dystonia, pain, Bell's Palsy, and stroke (See Figure 4.3) [210, 169, 168, 172, 192, 209, 171, 196, 197, 195, 191, 193]. In the remainder of this chapter, I briefly summarize several of these projects below (dystonia and pain), then discuss the results of our work on Bell's Palsy modeling and synthesis project in more detail. I will explain the detail of our work on stroke modeling and synthesis in Chapter 5 and Chapter 6.

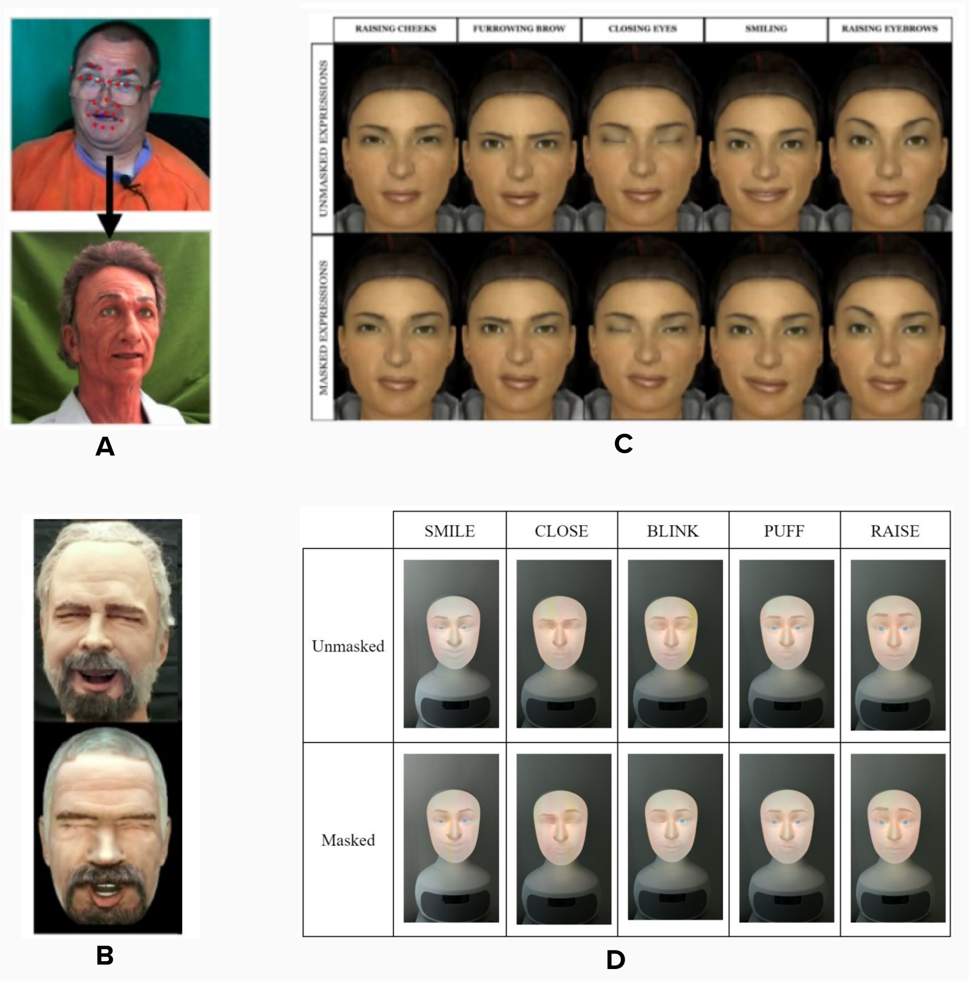


Figure 4.3. Three examples of the expressive patient simulator systems our team has built, with clinically relevant-expressions: **A)** Dystonia [210], **B)** Pain [168], **C)** Bell's Palsy [172], and **D)** stroke [191].

4.3.1 Dystonia

Dystonia is a movement disorder characterized by involuntary motions, often in the head and neck. People with dystonia often struggle during interaction due to the biases of others, raising an possibility to explore if a robot conveying dystonia could serve as a facilitator to help improve human-human communication. Our team interviewed four people with head and facial movement disorders and synthesized their movements on a physical robot, and experimentally explored using these robots as social facilitators to improve communication between people with and without disabilities. The results suggest that a robot may be useful for this purpose [210].

The results also indicate a significant relationship between people who hold negative attitudes toward robots and negative attitudes toward people with disabilities.

4.3.2 Pain

Our team modeled and synthesized both acute and chronic pain, on both virtual agents and physical robots [169, 168]. This study explored people's perceptions of pain, both on a humanoid robot and comparable virtual agent, using autonomous facial expression synthesis techniques.

Our team conducted an experiment with clinicians and laypersons to explore differences in pain perception across the two groups, and also to study the effects of embodiment (physical robot or virtual agent) on pain perception. The results of this study indicated that clinicians have lower overall accuracy in detecting synthesized pain in comparison to lay participants. It also suggested that all participants are overall less accurate detecting pain from a humanoid robot in comparison to a comparable virtual agent [168].

4.3.3 Bell's Palsy

In our work, we focused on a particular type of FP, Bell's Palsy (BP). We presented an FPM framework personalized to model characteristics of BP, and utilized the AMS framework to synthesize it on a virtual RPS. This work explored two research questions. First, *how does one computationally model the facial characteristics of BP, and synthesize them on a patient simulator to help support clinical engagement of those affected?* To address this question, the first step was to collect self recorded, publically-available videos from people with BP conveying four expressions (raising eyebrow, furrowing brow, smiling, and closing the eye).

Next, we presented a novel algorithm for the FPM framework to build accurate computational masks that can model facial characteristics of people with BP and are constructible in real time. (See Figure 4.2) This algorithm tracks faces in each source video and uses the 2D coordinates of the 34 facial features of the unaffected side of the face to calculate the 2D coordinates of the other part of the face, assuming that the person did not have A-FEs. Dividing the actual coordinates of the affected side by the calculated coordinates of the affected side gave us the scaling parameters $\beta_{i,x}$ and $\beta_{i,y}$ for x and y of each of the facial points. A 68-bit array consisting of the scaling parameters of all 68 tracked feature points is the calculated mask for the patient with BP.

Our second research question was *How realistically do these masks convey signs of BP when applied to a virtual patient?* To address this question, we conducted a qualitative, expert-based perceptual experiment to evaluate the realism of the synthesized expressions in comparison to actual patients and get feedback for further refinement. This is a common method for evaluating synthesized FEs [59, 165].

To perform this validation, after collecting videos from a performer without BP, we inputted the videos into the AMS framework (See Section 4.2), and overlaid three pre-built masks of BPs to recreate the AFE. (See Figure 4.2). Next, the generated asymmetric expressions of BP were transferred to the face of a VPS system to create stimuli videos (See Figure 4.3-C).

The results of this study suggest that two of the developed BP masks realistically display signs of BP. Furthermore, clinicians' perceptions of the synthesized expressions were comparable to their perceptions of the expressions of real people with BP. Therefore, the models described in this work have the potential to provide a practical training tool for CLs to better understand the emotions of people with this facial paralysis.

4.4 Future Work

There are several opportunities to advance the state-of-the-art of expressive RPS and VPS systems within the context of clinical learning, as well as in the broader context of robotics and HRI. These include technical advancements, such as new methods for FEA, FMA, and FSA, as well as socio-technical considerations, such as stakeholder-centered design and ethical questions. We briefly outline these below.

4.4.1 Advancing Expression Recognition and Synthesis Systems

As discussed, there are many methods for recognizing and synthesizing facial expressions. However, they have their drawbacks. Many commercially-available systems are unable to perform the tasks necessary for FE analysis or synthesis (e.g., FaceReader is not able to provide head pose estimation). Furthermore, systems may may lack state-of-the-art performance, rendering them impractical for clinical applications.

Thus, there are many opportunities to advance the state-of-the-art. For example, some regression-based methods such as CNNs are successful for FL detection and tracking. Furthermore, Gabor features showed promising results for feature extraction, and CNN and SVM methods improved classification performance. Integrating these approaches into facial expression FEA and FSA systems may improve analysis and/or synthesis of dynamic FEs in individuals with and without facial disorders.

4.4.2 Combining Domain Knowledge with Model Development

As part of the design process, engaging in stakeholder-centered design with CEs and CLs, as well as conducting observations of live simulations is important. For example, neurologists can help validate if neurological impairment models created by the system are realstic, and also ensure the patient simulator's appearance and expressiveness is well-aligned with their clinical education goals.

4.4.3 Real-world, Spontaneous Data Collection

It is important for developers to release systems that are designed and built using enough real-world, spontaneous facial expression data [119]. The number of facial expressions used for training and developing FEA, FAM, and FSA systems should be much higher to lead to more realistic results. In case of having a low number of images for training, it is challenging to choose the best approaches to enlarge the dataset while developing the system. Expressive robot developers also need to make sure the system includes a continuous adoption process that learns each user's expressions over time and adds them to its knowledge base [119]. It is also important to pay close attention to include the variability of the facial data in terms of subjects, by including data from subjects well-represented in gender and ethnicity, as well as diversity in terms of lighting, head position, and face resolution [119]. Given that patient simulators are designed to mimic humans and are designed for use by humans, we added a discussion on the importance of having designs that are informed by human sensory systems and behavioral outputs. Finally, it is important that datasets are labeled and analyzed in concert with domain experts, but to our knowledge little work has been done in this area. One potential solution can be to create a large training set of photorealistic facial expressions generated using existing face generation platforms labeled by human observers.

There are several existing facial expression datasets and Action Unit datasets that tackle some of the data collection challenges, including DISFA [164], BP4D-spontaneous [272], Aff-Wild 2 [145], and SEWA DB [147]. Furthermore, some of the recent facial expression synthesis methods, such as those mentioned in Section 3.4, are also intended to address these challenges. However, more work can be done in this field to tackle all the afordmentioned problems.

Moreover, newer directions also seek learning models with little or no supervision, both for facial landmarks (unsupervised landmark detection) and for Action Units which can help to address these challenges.

In terms of identifying databases of images or videos that reflect real facial expressions, it is important to consider the relationship between internal states and external facial cues. Work done by Benedek et al. [60] indicates people perceive the appearance of the face, especially the eyes of others, to understand both their external goals or actions, and their internal thoughts and feelings. Voluntary facial expressions are sometimes made in the absence of internal states. On the other hand, it is difficult to detect internal states in case attention is not presented externally. Therefore, it is critical to identify datasets of real data to better infer the external facial cues and more accurately interpret internal states.

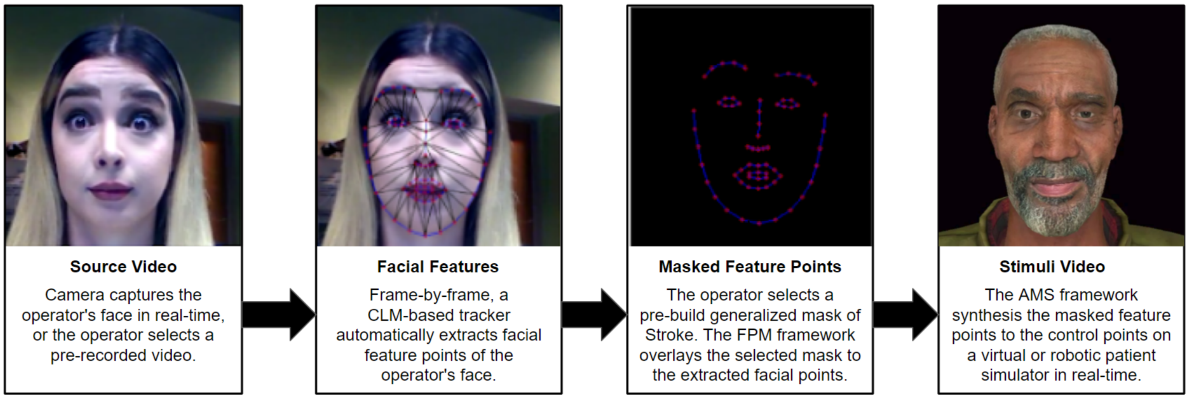
It is worth mentioning that there is the potential of having a pattern of confusions (false alarms and misses) in detected facial expressions. False alarms is the errors of describing a facial expression being present when it was absent. Misses is the errors of describing a facial expression as being absent when it was present. Studies indicate that the pattern of confusion becomes worse when some other challenges occur at the same time, such as illumination or occlusion in an image [198].

4.4.4 Cultural Considerations

Researchers have also explored the caveats associated with cultural variance in the way observers infer internal experiences from external displays of facial expressions. For example, Engelmann et al [100] argues that culture influences expression perception in different ways. For one, people from different cultures may perceive the intensity of external facial expressions differently. For example, American participants rated the intensity of same expressions of happiness, sadness, and surprise higher that Japanese participants. Moreover, depending on cultural contexts, there is a difference in the way people infer internal states from external facial cues of expressions. For example, researchers ran an experiment to ask two groups of American and Japanese participants to rate the intensity of internal and external state of a person expressing certain emotions. American participants gave higher rates to external facial cues of emotions, while Japanese participants gave a higher ratings to internal state of emotions. Therefore, it is important to consider these cross-cultural differences in inferring internal states and external expressions.

4.4.5 Universal Model Generation for Pathologies

In order to generally represent all patients with specific pathologies, one can create a universal model for each that encompasses its predominant features. This can be done by leveraging our previous findings in Section 4.3 to further extend the FPM framework in two directions: 1) Extend the FPM framework to encompass the predominant features of a specific pathology (e.g., stroke), and 2) transfer the framework from being an individual mask generator


Source Video

Camera captures the operator's face in real-time, or the operator selects a pre-recorded video.

Facial Features

Frame-by-frame, a CLM-based tracker automatically extracts facial feature points of the operator's face.

Masked Feature Points

The operator selects a pre-build generalized mask of Stroke. The FPM framework overlays the selected mask to the extracted facial points.

Stimuli Video

The AMS framework synthesis the masked feature points to the control points on a virtual or robotic patient simulator in real-time.

Figure 4.4. The context for performing masked synthesis using the FPM framework and the AMS framework. This can be performed on either a VPS or RPS system.

to a universal model generator. This can be done by using enough source videos of people with the specific pathology, extracting its common features, and creating a general model. (See Figure 4.4). By leveraging this work, CLs will have the potential to more accurately diagnose people with diverse backgrounds, and to be better able to interact with them.

4.4.6 Shared Control System for Expressive Robots

Nowadays, autonomy level is one of the most effective aspects to consider in developing more efficient HRI management systems in different contexts [204]. For example, using a shared control approach in a structured environment enabled the operator to benefit from human intelligence, while benefiting the robot accuracy and precision to improve the performance of grasping and handling of specific objects [104]. Another interesting example of using a shared control system is in flexible industrial automation, where it can help to improve the safety of users who work in dangerous areas, and increase accuracy, reliability and flexibility [125]. Therefore, it will be an interesting context to explore and to exploit the shared control system approach in the area of patient simulators.

Considering how to share autonomy between a human and robot is an important aspect to ensuring effective HRI [204]. It can help to reduce an operator's workload, allow both inexperienced and professional operators to control the system [204, 104].

As such, it is important to focus on interaction between the control system and human users in the context of expressive simulator systems. Thus, researchers can design and validate a customizable, shared autonomy system for expressive RPS systems to leverage the advantages of automation while also having users as "active supervisors". For example, in our work, we are designing a shared autonomy system that can support a range of adjustable control modalities, including direct tele-operation (e.g., puppeteering), pre-recorded modes (e.g., hemifacial paralysis during a stroke), and reactive modes (e.g., wincing in pain given certain physiological signals) [196]. It also can help overcome common control challenges, including the operator being overwhelmed, having high workload, and lack of autonomy in robotic simulator systems. This system can help make robots adjustable to different control paradigms, so that they reliably support CEs' workload in dynamic, safety-critical settings and improve the operator's ability to focus on their educational goals rather than robot control.

4.5 Discussion

The overall system presented in this work, seamlessly creates a comprehensive solution that accurately portrays FEs similar to real patients' facial characteristics on RPS faces, situated within a HET context. The technologies and methods discussed in this review can cultivate a bridge between robotics and healthcare research, and improve existing clinical training practices, by enabling VPS and RPS systems to become more diverse, interactive, and immersive for CLs and CEs. This will enable CLs to further engage during training sessions, will help them to significantly improve their communication and procedural skills, and ultimately save more lives. Building on these approaches will lead to systems with a much wider range of expressivity, such as the ability to express clinically-relevant facial expressions. Through studies with stakeholders, including patients, clinicians, and clinical learners, technologists can improve the expressiveness of simulator robots, and improve the interactions between humans and robots for expressive patient simulators and beyond. Ultimately, this work may help clinicians deliver better clinical care, by both improving their diagnostic skills and by providing new educational opportunities for reducing racial disparities [210].

Furthermore, disseminating the results of this work (and software) to the research community will help both the broader robotics and healthcare communities employ these novel systems in their own application domains. This may trigger a new round of relevant technological innovations by creating the next generation of patient simulator robot technology to support clinicians in healthcare education settings. Furthermore, the results of this study will enable roboticists to discover platform-independent methods to control the FEs of both robots and virtual agents, and yield new modalities for interaction.

4.6 Chapter Summary

This chapter presented the foundations and two frameworks for creating expressive RPS systems, with the aim to support researchers in developing and deploying systems capable of depicting clinical conditions effectively. The next chapter will introduce a new clinical training tool with an expressive face capable of realistically depicting non-verbal, asymmetric FP cues representing acute stroke.

4.7 Acknowledgments

This chapter contains material from “Facial Expression Modeling and Synthesis for Patient Simulator Systems: Past, Present, and Future” by M. Pourebadi, and L. D. Riek., which appears in the Proceedings of the ACM Transactions on Computing for Healthcare Journal (ACM HEALTH), 3(2), 1-32, 2022 [195]. I was the primary investigator and author of this paper. This chapter also contains material from part of “Modeling and synthesizing idiopathic facial paralysis” by M. Moosaei, M. Pourebadi, and L. D. Riek, which appears in the Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2019 [172]. Maryam Moosaei and I were the primary authors of this work.

Chapter 5

Modeling and Synthesizing Stroke on Expressive Patient Simulator Robots

5.1 Introduction

In the previous Chapter chapter 4, I presented our proposed techniques for creating FPM and AMS frameworks. In this chapter, I will describe our work developing a new FPM framework for creating computational models of the predominant features of stroke, and an AMS framework for depicting patient-like stroke characteristics on the face of a physical robot.

Stroke is a substantial contributor to the global economic burden [228], the second leading cause of mortality, and the third leading cause of disability-adjusted life years worldwide [14, 138]. Stroke causes premature death or permanent disability in 10 million people worldwide, of the 15 million people it affects each year [220, 95].

Patients with stroke (PwS) usually experience FP, which is the inability to fully move some or all facial muscles on the affected side of the face, usually caused by weakness or damaged nerves [58]. FP in PwS can present in different ways, including asymmetric facial expressions (A-FE), gaze deviation, loss of blinking control, drooping of the mouth on the affected side, and slurred speech [58, 167].

In clinical contexts, there are several challenges to having clinicians assess PwS properly. Stroke is regularly misdiagnosed in one out of ten cases [223] and is the fourth most common misdiagnosis reported by clinicians [235]. Clinicians may misdiagnose acute stroke and BP, as they have shared symptoms (such as FP), motivating the need for clear language in clinical practice to avoid diagnostic error and patient harm [51].

Traditional healthcare education and training tools rarely provide CLs with adequate training for performing neurological assessment tests (NATs), which may contribute to stroke misdiagnoses for actual patients [116]. Even if a CL performs exams well, they may not be familiar enough with stroke to adequately make a proper diagnosis. Additionally, they may have low confidence in the accuracy of their diagnosis. Considering the subjective nature of stroke diagnosis, this uncertainty can cause missing opportunities for critical interventions, accurate diagnosis, proper treatment plans, and prevention of severe harm [176, 47].

These challenges necessitate new clinical training tools for CLs to practice assessing and treating stroke.

As discussed in Chapters chapter 1 and chapter 2, advances in robotics offer new opportunities for healthcare education and training by means of the creation of RPS [49, 259, 88].

Despite the benefits of existing RPS systems, they have one major challenge: they incorporate facial designs that presuppose a symmetrical human face. This assumption makes RPS systems unable to naturalistically express non-verbal facial cues important for rapid diagnosis of neurological emergencies, such as stroke [197]. Being able to depict asymmetric facial cues is crucial for simulating clinical scenarios on these systems that demand dynamic changes in appearance (e.g., simulating a PwS with abnormal visual symptoms such as A-FE and facial droop). The lack of A-FE in simulators can yield problems with learner immersion, skill transfer, and learning performance [195]. As a result, CLs may be incorrectly learning to read social cues from and diagnose symptoms of people presenting with FP [195].

Although existing RPS systems capable of conveying A-FE can address the previous challenge, developing such systems introduces other technical and design challenges. First, as FEs and their intensities exhibit significant inter-individual variability and dynamicity[275], the development of a universal RPS system capable of accurately modeling and presenting neurological impairments across diverse cultural and demographic spectrums poses a daunting challenge

[195]. Doing this would require access to a large corpora of data from PwS representing a diverse set of characteristics associated with acute neurological disorders and facial impairments, which is both time and labor-intensive.

Bandini et al. recently presented the first publicly available dataset comprising videos featuring individuals with neurological disorders [55]. However, their work primarily focused on exploring impairments within the lower facial region, specifically concerning oro-motor abilities assessed in PwS. Achieving an accurate and comprehensive presentation of acute stroke conditions necessitates the inclusion of data enclosing a broader spectrum of impairments within both the upper and lower facial regions. This entails collecting data from PwS who engage in a more extensive array of speech and non-speech diagnostically relevant facial movements.

Second, it can be challenging to analyze the data collected from a restricted cohort of PwS and extrapolate it to construct stroke models that depict a more extensive population of PwS. Nonetheless, it is important to develop such universal models to design versatile RPS systems with synthesized faces encompassing a diverse patient group. This diverse assortment includes but is not restricted to individuals of varied ages, genders, and ethnicities suffering from various health afflictions [275]. Addressing these challenges can lead to the design and development of diverse, expressive RPS systems, capable of mimicking realistic stroke symptoms and representing a diverse group of people with FP. This may prevent CLs from initiating misperceptions of people with FP, improve clinical diagnosis, enhance clinical communication, and, consequently, improve care delivery for people with FP.

This chapter explores how to address this issue for a particular type of FP, acute stroke, by developing robotic patient simulators with stroke (RPSwS): an expressive training tool capable of realistically depicting non-verbal, asymmetric facial cues. In this research, we collected realistic data from PwS performing a wide range of tasks required for assessing diagnostically relevant facial movements, developed a modeling framework to create mathematical representations for naturalistic facial characteristics of stroke, and synthesized them on a robotic platform.

There has been prior work on stroke recognition and detection [53, 55] and virtual expressive patient simulator development capable of representing other FP pathologies [172]. However, to our knowledge, we are the first to introduce a data-driven, statistical modeling approach representing facial characteristics of stroke, and then use these models to synthesize stroke on RPS systems. Using such systems can help educate and assess CLs' stroke diagnosis skills, particularly in the context of real time RPS interaction [192, 197]. Our system has the potential to help calibrate clinicians' perception of acute stroke and support the health of people who have a stroke.

The contributions of this chapter are threefold.

First, we introduce the stroke facial palsy mask framework (Stroke FPM): a new framework for generating statistical models representing stroke. This consists of three parts: 1) a deep learning face detection and alignment method to automatically extract the region of interest (ROI) around the PwS's face, 2) a facial landmark localization technique to accurately identify and anonymously track the location of specific facial landmark points of interest within the ROI, and 3) a statistical modeling approach to use tracked location values in order to automatically extract stroke statistical measurement (SSM) features. The landmark tracker can accurately identify and automatically track specific points in videos of PwS that are crucial for analyzing A-FE movements. The Stroke FPM can successfully extract a diverse set of visual features that represent stroke-related asymmetric movements in each facial region. We validated the Stroke FPM on a new dataset of PwS we collected, all of whom experienced acute ischemic stroke resulting in neurological findings. This provided a systematic and objective way to analyze and interpret facial movement patterns associated with stroke, contributing to a better understanding of the neurological effects of the condition. (See Section 5.4).

Second, we present the end-to-end stroke analysis-modeling-synthesis (Stroke AMS) framework: which applies the generated models onto the face of an RPS system to automatically display FP [197, 195]. To our knowledge, this is the first data-driven statistical modeling approach to represent the facial characteristics associated with stroke, encompassing both the lower and upper regions of the face, being used to synthesize stroke effects on a range of RPS systems, thereby enabling the generation of realistic FP simulations. The AMS framework enables robot developers to generate diverse data-driven FP faces for patient simulators, situated within a clinical education context [197] (See Section5.5).

Third, we report the results from a perceptual study with seven clinicians to investigate the efficacy of our system for modeling and synthesizing stroke (See Section 5.6). This study explored the visual differences in realism and similarity between the synthesized stroke robot faces and those of stroke patients. The results of these measurements enabled the identification of features that can make the stroke robot look more realistic (See Section 5.8). Overall, participants perceived the stroke models used to mask all three facial regions of the robot and the overall face as very realistic. They also reported the stroke models were moderately similar to real PwS. Participants also reported positive comments with regard to the usefulness of the robot, and gave some suggestions for improvement (See Section 5.8).

This work has impacts on multiple research fields, including clinical education, health informatics, automatic face and gesture (FG), and human-robot interaction (HRI). Our RPS system can depict PwS, with the aim to help train and assess future generations of neurologists in rapid diagnosis of acute neurological injury. Employing simulators in this way may help improve clinicians' diagnostic skills, which will help improve care delivery for people with FP. Our work can also help researchers in the FG community to explore new methods for asymmetric facial expression analysis, modeling, and synthesis. Moreover, our study enables HRI researchers to explore methods for designing social robots to enhance people's perception of individuals with FP and understand the effects of facial asymmetry on social interactions. We discuss the implications of these findings in Section 5.9).

5.2 Background

Many researchers have explored the use of automatic facial analysis for clinical applications. This includes developing virtual and robotic patient simulators cable for expressing a range of pathologies, such as pain and Bell's Palsy [169, 168, 170, 171], and automatic analysis of facial movements in individuals with neurological disorders such as stroke, amyotrophic lateral sclerosis, and Parkinson's [53, 54, 122]. In the development of such systems, researchers use facial landmark localization to extract features that affect the characteristics of each specific disorder, and measure the presence or severity of signs. The common automatic facial landmark localization methods used for clinical applications include active appearance models (AAM) [86], supervised descent method (SDM) [263], constrained local model (CLM) [87], ERT [142], and the deep learning-based face alignment network (FAN) [69].

Bandini et al. [55] compared the accuracy of these facial landmark localization methods for detecting speech and orofacial impairment in PwS. Their work indicated that FAN had the lowest localization error; however, they also identified the existence of bias in the face alignment accuracy when oro-facial impairment was present in a facial video. Furthermore, they studied the effect of fine-tuning the FAN algorithm with data from the target populations (e.g., individuals with stroke) on landmark localization accuracy. Compared to the pre-trained FAN, their work demonstrated a lower bias and improved landmark localization accuracy when using the fine-tuned FAN [55]. Therefore, the fine-tuned FAN could be a better candidate for tracking landmark locations when working with data from PwS.

5.3 Data Collection and Analysis

The main step for developing expressive patient simulators capable of realistically representing facial characteristics of stroke is to model them using real-world data collected and extracted from individuals with stroke. Toronto NeuroFace released by Bandini et al. [55], is the first public dataset that includes videos of oro-facial movements performed by individuals with ALS and post-stroke. However, to the best of our knowledge, there are no publicly available datasets specifically focused on diverse tests to assess PwS. These tests are essential for the purposes of our research, which aims to assess FP in both the lower and upper regions of the face. Additionally, the dataset should include exam findings consistent with acute neurological injuries, anonymously tracked facial landmark features, and relevant clinical metadata. Thus, this required us to collect a new dataset of videos of facial movements performed by individuals with facial impairments.

This section presents our efforts to collect and analyze stroke data. We include details about participants, task selection, data collection procedure, clinical assessment of the recorded videos, and acquired facial landmark localization method. The study was approved under Institutional Review Board number 191488X by the Human Research Protections Program at our institution.

5.3.1 Participants

Unlike other stroke datasets collected from post-stroke and amyotrophic lateral sclerosis patients [53], we focus on people with acute stroke. We recorded patients recently admitted to a neurological unit at an urban academic medical center to ensure the videos represent existing acute neurological features, which are a crucial source of information for diagnostic purposes.

We recruited 16 participants for this study: 14 PwS admitted to the Intensive Care Unit (8 female, 6 male), and two participants without stroke (PwoS) (1 female, 1 male). The PwS had experienced acute neurological injury resulting in neurological findings such as facial droop, eyelid apraxia, dysarthria, and coma. All participants provided verbal informed consent and HIPAA authorization. We collected videos of PwS to generate data-driven stroke models, while videos of PwoS were collected to serve as source videos for controlling the robot's facial expressions.

5.3.2 Selection of Neurological Assessment Tests

The cranial nerves (CN) are components of the peripheral nervous system, some of which transmit instructions to and from the brain, and others which send and receive information to and from the brain [16].

Table 5.1. This table presents a list of cranial nerves (CNs) and their corresponding neurological functions. The CNs required to be assessed for diagnosing impairments in the patient's facial movements, eye movements, and speech are highlighted in gray.

Cranial Nerves	**Corresponding Neurological Function**
CN_I	Smell
CN_II	Vision acuity, blink to threat
CN_III	Horizontal eye movements (adduction)
CN_IV	Vertical eye movements
CN_V	Facial sensation assessment, corneal reflex
CN_VI	Horizontal eye movements (abduction)
CN_VII	Facial muscle strength and movement
CN_VIII	Hearing
CN_IX	Taste and gag reflex
CN_X	guttural sounds and gag reflex
CN_XI	Neck muscles
CN_XII	Tongue and lateral movements

CN_I carries the ability to smell. CN_II takes visual impulses from the eye to the brain by means of the optic nerve, while CN_III, CN_IV, and CN_VI present eye movements in different directions. CN_VII is responsible for the strength of many muscles of facial expression, including the left and right muscles of the upper region of the face (i.e., forehead) and the left and right muscles of the lower region (i.e., cheeks and mouth). CN_VIII is responsible for taking sound impulses from the cochlea to the brain. CN_IX and CN_X are responsible for raising the soft palate of the mouth and the gag reflex. CN_XI innervates the muscles responsible for shoulder shrugging and lateral head movement. Finally, CN_XII is responsible for enabling the tongue muscles to function properly. This study emphasizes the assessment of six CNs necessary for diagnosing impairments related to facial movements, eye movements, and speech in patients. (In Table 5.1, the six selected CNs are highlighted in gray.)

In order to assess the functionality of the selected CNs in PwS, many clinicians, including neurologists at our medical center, perform various neurological assessments [116]. These neurological assessments encompass a series of structured evaluations and examinations designed to assess the neurological status of patients and aid in identifying the presence and characteristics of stroke symptoms in their faces and bodies. By employing these assessments, clinicians can gather valuable diagnostic information, enabling them to manage stroke appropriately.

In this chapter, we aim to develop robots capable of depicting speech and non-speech neurological criteria in both the lower and upper regions of the robot's face. Building upon the findings reported by Banditi et al. [53, 55] and drawing on standard measurement criteria used by our clinical collaborators, we decided to video record PwS performing ten tests from a list of existing neurological assessments (NAT) to examine the six selected CNs [116]. They include:

- following a finger that moves in an "H" shape in front of their face with their eyes while keeping their head fixed (FOLLOW),
- closing and opening the eyes (BLINK),
- closing their eyes tightly while wrinkling their inner eyebrows (CLOSE),
- raising the eyebrows while trying to induce the wrinkling of the forehead muscles (RAISE),
- making the cheeks larger and rounder by filling them with air while having lips tightly closed (PUFF),
- making a big smile with lips wide open (SMILE),
- repetitions of the syllable /ma/ at a comfortable speaking rate and loudness (MAMAMA),
- repetitions of the syllable /la/ at a comfortable speaking rate and loudness (LALALA),
- repetitions of the syllable /ga/ at a comfortable speaking rate and loudness (GAGAGA), and
- rest position with teeth in normal bite and neutral facial expression (REST).

Table 5.2. This table presents a list of selected cranial nerves (CNs) and their corresponding Neurological Assessment Tasks (NAT), including FOLLOW, BLINK, CLOSE, RAISE, PUFF, SMILE, MAMAMA, LALALA, GAGAGA, and REST.

Selected CNs	**Corresponding neurological assessment tests (NAT)**
CN_III, CN_IV, CN_VI	FOLLOW, REST
CN_VII	BLINK, CLOSE, RAISE, PUFF, SMILE
CN_X, CN_XII	MAMAMA, LALALA, GAGAGA

(See Table 5.2 for the list of selected CNs and their corresponding NATs.)

5.3.3 Experimental Setup and Data Collection

Other works recorded their participants in a controlled environment where patients sat in front of the camera and were asked to perform each task for a specific number of times or duration [53]. In contrast, we recorded each patient during their standard neurological assessment procedure, and thus, we could only record one repetition per task from each patient.

We asked each participant to perform the ten selected NATs listed in Table 5.2, resulting in 160 video recordings (140 from PwS, and 20 from PwoS). The average duration of the videos was 5.5 seconds. We recorded participants' faces using the GoPro HERO8 Black camera. During the tasks, participants lay on their beds in front of the camera, with a face-camera distance between 30 and 35 inches. A continuous light source with a color temperature of 5700K and a brightness of 200 lumens, 200 lux @1m was attached on top of the HERO8 to illuminate the face uniformly. A separate video recording for each performed task was stored at approximately 30 frames per second and 1920 x 1080 pixels of image resolution, in Codec H.264 / MPEG-4 AVC.

5.4 Facial Palsy Model Framework Development

This section, introduces the Stroke FPM, a novel framework for generating computational models representing the A-FE characteristics of stroke. (See Figure 5.1 for an overview.) Developing Stroke FPM included several steps. First, we took our previously collected data (see

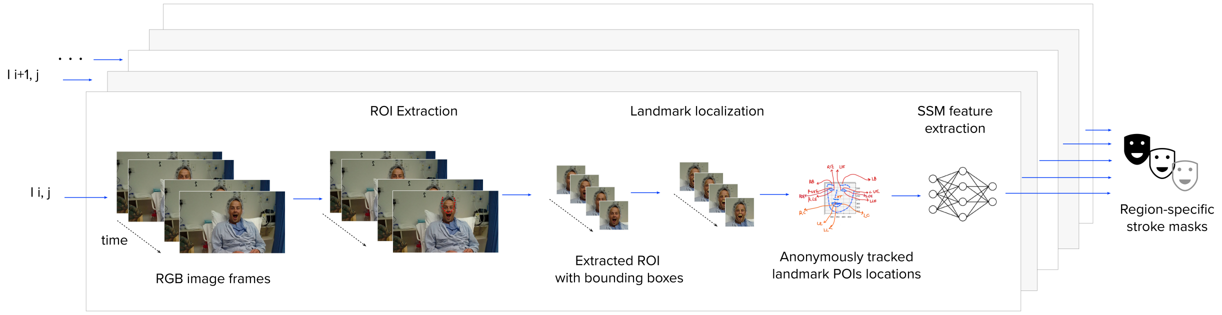


Figure 5.1. Stroke facial palsy mask (Stroke FSM) framework consists of three main components. First, we develop a deep learning face detection and alignment method to automatically extract the region of interest around the PwS's face. Next, we perform landmark localization to accurately identify and anonymously track specific landmark locations within the region of interest that are crucial for analyzing A-FE movements. Next, we perform landmark localization to anonymously track landmark locations. Next, we engage in SSM feature extraction to create 15 SSM features. Finally, we collect SSM features over all NAT videos to create 75 region-specific stroke masks.

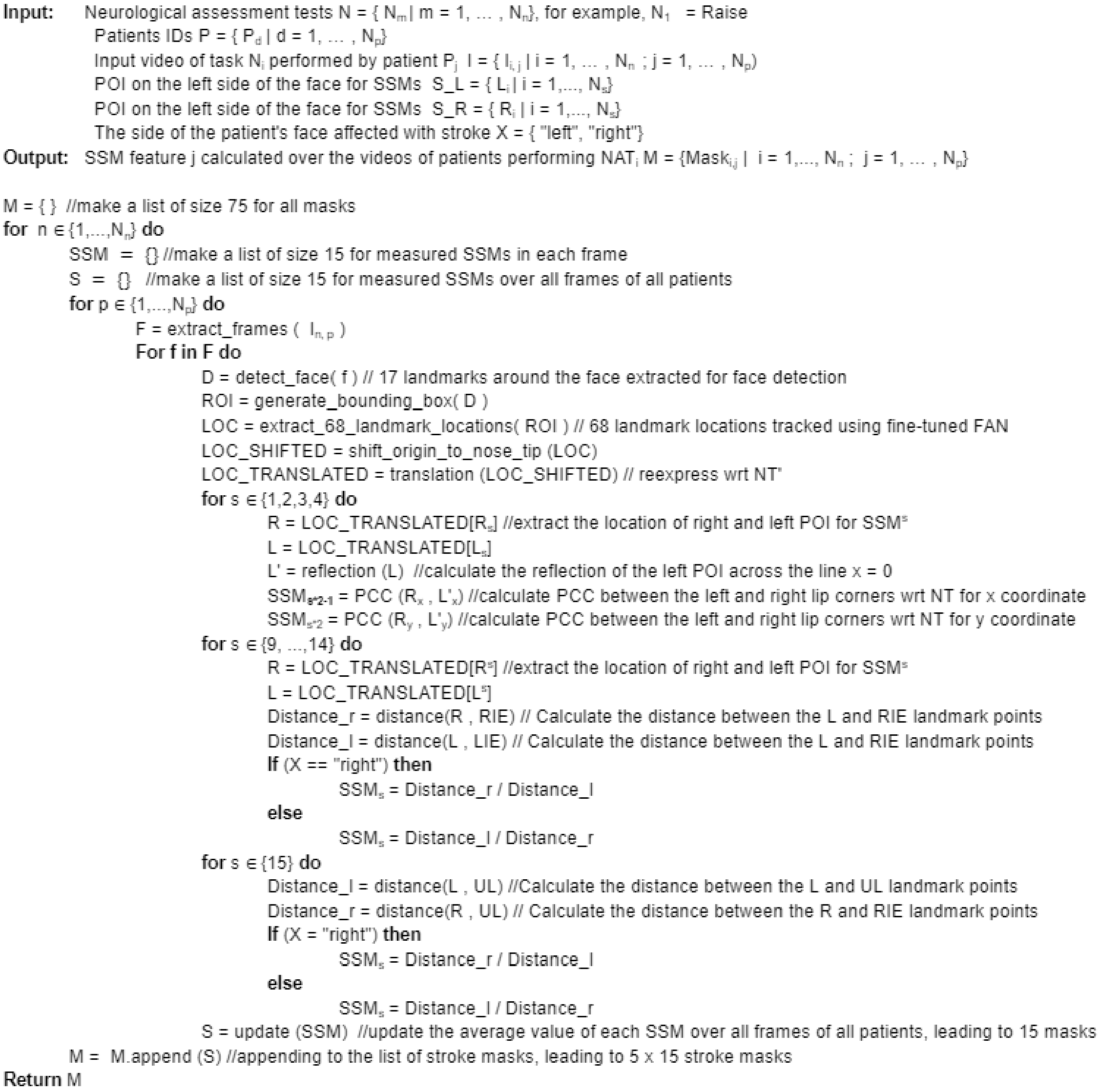
Section 5.3), then engaged in data processing and annotation. We then tracked facial feature landmarks in the videos (Section. 5.4.1) and then extracted 15 SSM features (Section. 5.4.3). Next, we used the average SSM features to create region-specific stroke masks (Section. 5.4.3).

5.4.1 Region Of Interest Extraction

We introduce an automatic region of interest (ROI) extraction technique. We implemented a deep-learning-based face detection and alignment method based on fine-tuned FAN [55] to detect 17 landmarks in proximity to the face in each image frame. This method then uses the extracted landmarks to produce bounding boxes around the face with an expansion factor of 0, a height of 600 px, and a width of 600 px. The proposed method actively defines the aforementioned bounding boxes and identifies an ROI for the accurate localization of the face.

5.4.2 Facial Landmark Localization

We employed the fine-tuned FAN technique [55] to track the location of 68 facial landmarks within the ROI via heatmap regression from the RGB frame image. This technique yielded anonymously tracked landmark features within the ROI for each video, making it possible to prevent participants from being identified by their faces.



```
Input:    Neurological assessment tests $N = \{ N_m | m = 1, \dots , N_n\}$, for example, $N_1$ = Raise
          Patients IDs $P = \{ P_d | d = 1, \dots , N_p\}$
          Input video of task $N_i$ performed by patient $P_j$  $I = \{ I_{i,j} | i = 1, \dots , N_n ; j = 1, \dots , N_p)$
          POI on the left side of the face for SSMs  $S\_L = \{ L_i | i = 1,\dots, N_s\}$
          POI on the left side of the face for SSMs  $S\_R = \{ R_i | i = 1,\dots, N_s\}$
          The side of the patient's face affected with stroke X = { "left", "right"}
Output:   SSM feature j calculated over the videos of patients performing $NAT_i$ $M = \{Mask_{i,j} | i = 1,\dots, N_n ; j = 1, \dots , N_p\}$

M = { }  //make a list of size 75 for all masks
for  $n \in \{1,\dots,N_n\}$ do
    SSM  =  {} //make a list of size 15 for measured SSMs in each frame
    S  =  {}   //make a list of size 15 for measured SSMs over all frames of all patients
    for $p \in \{1,\dots,N_p\}$ do
        F = extract_frames (  $I_{n, p}$ )
        For f in F do
            D = detect_face( f ) // 17 landmarks around the face extracted for face detection
            ROI = generate_bounding_box( D )
            LOC = extract_68_landmark_locations( ROI ) // 68 landmark locations tracked using fine-tuned FAN
            LOC_SHIFTED = shift_origin_to_nose_tip (LOC)
            LOC_TRANSLATED = translation (LOC_SHIFTED) // reexpress wrt NT'
            for $s \in \{1,2,3,4\}$ do
                R = LOC_TRANSLATED[$R_s$] //extract the location of right and left POI for $SSM^s$
                L = LOC_TRANSLATED[$L_s$]
                L' = reflection (L)  //calculate the reflection of the left POI across the line x = 0
                $SSM_{s*2-1}$ = PCC ($R_x$ , $L'_x$) //calculate PCC between the left and right lip corners wrt NT for x coordinate
                $SSM_{s*2}$ = PCC ($R_y$ , $L'_y$) //calculate PCC between the left and right lip corners wrt NT for y coordinate
            for $s \in \{9, \dots,14\}$ do
                R = LOC_TRANSLATED[$R^s$] //extract the location of right and left POI for $SSM^s$
                L = LOC_TRANSLATED[$L^s$]
                Distance_r = distance(R , RIE) // Calculate the distance between the L and RIE landmark points
                Distance_l = distance(L , LIE) // Calculate the distance between the L and RIE landmark points
                If (X == "right") then
                    $SSM_s$ = Distance_r / Distance_l
                else
                    $SSM_s$ = Distance_l / Distance_r
            for $s \in \{15\}$ do
                Distance_l = distance(L , UL) //Calculate the distance between the L and UL landmark points
                Distance_r = distance(R , UL) // Calculate the distance between the R and RIE landmark points
                If (X = "right") then
                    $SSM_s$ = Distance_r / Distance_l
                else
                    $SSM_s$ = Distance_l / Distance_r
            S = update (SSM)  //update the average value of each SSM over all frames of all patients, leading to 15 masks
    M =  M.append (S) //appending to the list of stroke masks, leading to 5 x 15 stroke masks
Return M
```

Figure 5.2. Stroke Model Extraction Algorithm.

After an initial evaluation, we included 15 video recordings from 3 PwS (2 female, 1 male) performing five NATs (BLINK, CLOSE, RAISE, PUFF, and SMILE) in our final study. Some PwS were excluded due to cognitive impairment, facial occlusion, or excessive movements. Among all recorded NATs, we excluded five NATs; four of them did not have clear visual effects on facial asymmetry and movements (e.g., REST, MAMAMA, LALALA, and GAGAGA), and for one of them, the face tracker was not able to track the movements (e.g., the FAN tracker was not able to track gaze movement during FOLLOW task).

For each video recording, two human judges labelled each NAT performance. Furthermore, in order to prune the dataset, we only included the frames of each video where the algorithm accurately tracked the face. For this purpose, we instructed the human judges to label the frames where the face was not accurately tracked across that frame. We then only used the well-tracked frames for the next steps.

5.4.3 Feature Extraction and Selection

Points Of Interest

The second component of Stroke FPM is a new method that uses the values of tracked facial landmark locations within the ROI in each frame to automatically extract a set of SSM features. SSM features characterize the A-FE movements of facial points in different facial regions, illustrating stroke characteristics in each region. Other research on analyzing facial movements in individuals with stroke [54, 55] usually focuses on motion in the lower region of the face. However, in this work, we aim to measure asymmetric movements and gestures on the upper and lower regions of the face. Thus, we defined 15 facial landmark points of interest (POI) for feature extraction purposes, which are described in Figure 5.3.

SSM Feature Formulation

Building upon our findings from a literature review and our interviews conducted with clinicians, we created a list of 15 SSM features for measuring stroke-related clinical criteria. This list of SSM features enables us to assess the asymmetry movement between corresponding POI in the affected and unaffected sides of the face (see Table 5.3). For each NAT video, we used the POIs extracted frame-by-frame by the fine-tuned FAN method to extract the SSM features.

Building upon the work of Bandini et al. [53], we introduced our first two SSM features, which includes:

- Pearson's correlation coefficient (PCC) between the LC and RC landmarks, to extract two SSM features for each x and y coordinate (SSM_1 and SSM_2). In order to remove the

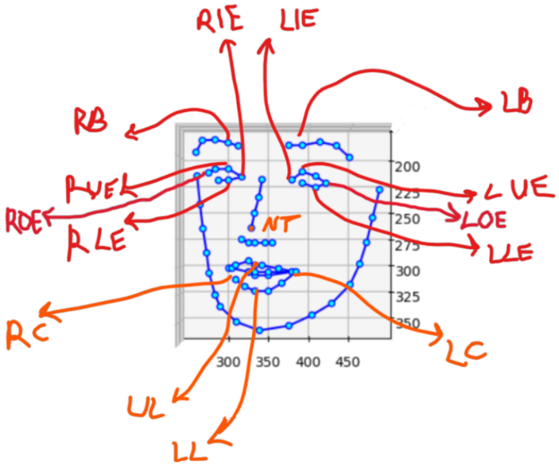


Figure 5.3. 15 facial landmark points of interest (POI) for feature extraction purposes: central points of lower and upper lips (LL and UL), left and right lip corners (LC and RC), left and right outer eye corners (LOE and ROE), left and right inner eye corners (LIE and RIE), central points of the left and right lower eyelids (LLE and RLE), central points of the left and right upper eyelids (LUE and RUE), and left and right brows (LB and RB), and nose tip (NT).

effects of head rotation, we shifted the center of the coordinate (0 , 0) to nose tip (NT), and re-expressed the landmark points with respect to NT.

This feature only measures movement coordination between the mouth's left and right sides in the face's lower region.

We expanded our list of SSM features and added six SSM features to measure asymmetry movement coordination in the upper region of the face. This included:

- Pearson's correlation coefficient between the LB and RB for each x and y coordinate (SSM_3 and SSM_4) to measure asymmetric y movement coordination in the brow area,
- Pearson's correlation coefficient between the LUE and RUE for each x and y coordinate (SSM_5 and SSM_6) to measure movements in the eye area,
- and Pearson's correlation coefficient between the LLE and RLE for each x and y coordinate (SSM_7 and SSM_8) to measure movements in the eye area.
- We measured the distance between LIE and LB and the distance between RIE and RB, and

then calculated the ratio of the measured distance on the affected side over the measured distance on the unaffected side to extract the *SSM*_9 feature.

We used a similar approach for extracting the SSM features listed below.

- the ratio of measured distance between LIE and LC and measured distance between the RIE and RC to extract *SSM*_10,
- the ratio of measured distance between LIE and UL and measured distance between the RIE and UL to extract *SSM*_11,
- the ratio of measured distance between LC and NT and measured distance between the RC and NT to extract *SSM*_12,
- the ratio of measured distance between LB and NT and measured distance between the RB and NT to extract *SSM*_13,
- calculate the ratio of measured distance between LUE and NT and measured distance between the RUE and NT to extract *SSM*_14,
- and calculate the ratio of measured distance between LLE and NT and measured distance between the RLE and NT to extract *SSM*_15.

The first set of features (from *SSM*_1 to *SSM*_8) represents how to coordinate the asymmetry movement between two sides are, whereas the second set of features (from *SSM*_9 to *SSM*_15) represents the difference of the range of movements between the affected and unaffected sides of the face.

Build Region-Specific FPM

We aimed to determine the SSM features that yielded the most salient visual contribution to conveying stroke for each facial region. Thus, we extracted 15 SSM features for five NATs performed by three patients. We then calculated the average value of each SSM feature for

Table 5.3. A list of 15 stroke statistical measurement (SSM) asymmetry features for assessing stroke clinical criteria. Each feature measures asymmetry movements of the landmark points of interest in the corresponding face region.

Region	SSM	Description	POIs
mouth	SSM1	Pearson”s correlation coefficient between the left and right lip corners wrt NT for x coordinate	LC and RC
mouth	SSM2	Pearson”s correlation coefficient between the left and right lip corners wrt NT for y coordinate	LC and RC
brows	SSM3	Pearson”s correlation coefficient between the left and right brows wrt NT for x coordinate	LB and RB
brows	SSM4	Pearson”s correlation coefficient between the left and right brows wrt NT for y coordinate	LB and RB
eyes	SSM5	Pearson”s correlation coefficient between the left and right upper eyes wrt NT for x coordinate	LUE and RUE
eyes	SSM6	Pearson”s correlation coefficient between the left and right upper eyes wrt NT for y coordinate	LUE and RUE
eyes	SSM7	Pearson”s correlation coefficient between the left and right lower eye points wrt NT for x coordinate	LLE and RLE
eyes	SSM8	Pearson”s correlation coefficient between the left and right lower eye points wrt NT for y coordinate	LLE and RLE
brows	SSM9	Ratio between the measured distance between inner canthus of the eyes and brow points affected side over unaffected side	LIE_LB and RIE_RB
mouth	SSM10	Ratio between the measured distance between inner canthus of the eyes and mouth corners affected side over unaffected side	LIE_LC and RIE_RC
mouth	SSM11	Ratio between the measured distance between inner canthus of the eyes and UL affected side over unaffected side	LIE_UL and RIE_UL
mouth	SSM12	Ratio between the measured distance between the left and right lip corners wrt NT affected side over the unaffected side	LC_NT and RC_NT
brows	SSM13	Ratio between the measured distance between the left and right brows wrt NT affected side over the unaffected side	LB_NT and RB_NT
eyes	SSM14	Ratio between the measured distance between the left and right upper eyes wrt NT affected side over the unaffected side	LUE_NT and RUE_NT
eyes	SSM15	Ratio between the measured distance between the lower eye points wrt NT affected side over unaffected side	LLE_NT and RLE_NT

each NAT video over all patients to generate 75 region-specific Stroke FPM models that best characterized acute stroke in different facial areas. This included 30 stroke models to mask the area around the eyes, 25 to mask the mouth region, and 20 to mask the area around the brows. Although the FP models developed in our earlier work [172] are generated so that each mask represents characteristics extracted from one patient, each of our stroke models was developed using facial characteristics of stroke extracted from data collected from a diverse group of patients. Thus, each stroke mask can represent diverse characteristics.

We analyze the features for separate facial regions rather than the entire face for two reasons. One, performing each NAT involves asymmetric movements on a subset of the facial parts but not whole facial parts, and thus, averaging over all facial parts to calculate models for

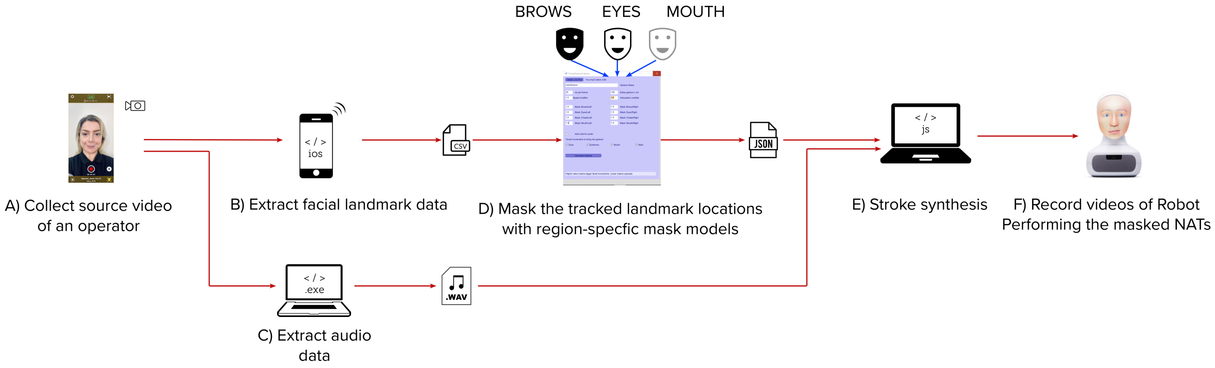

BROWS
EYES
MOUTH
< / >
ios
CSV
JSON
< / >
js
< / >
.exe
.WAV
A) Collect source video of an operator
B) Extract facial landmark data
C) Extract audio data
D) Mask the tracked landmark locations with region-specfic mask models
E) Stroke synthesis
F) Record videos of Robot Performing the masked NATs

Figure 5.4. The end-to-end stroke analysis-modeling-synthesis (Stroke AMS) framework. First, the system collects source video of an operator performing an NAT. Next, the framework tracks the facial landmarks from the video. Next, the system overlays the stroke models over the tracked landmark values with respect to the facial regions to which the landmark belongs. Finally, the framework synthesizes stroke on the robot's face synchronized with the audio extracted from the video.

each task may add values from symmetrical movements of facial parts that are not moving on that task. Second, during our interviews with stakeholders, they indicated that they would prefer to have the flexibility to apply different stroke models to different robots' facial parts of their choice. Unlike previous models [172] that mask the entire face, our system employs multiple stroke models to mask each specific facial region separately, allowing for the representation of varying levels of FP in each region. This approach results in a more realistic stroke facial appearance and movements, enhancing the fidelity of our system.

5.5 Acute Stroke Synthesis System Development

In order to study how clinicians perceived the synthesized stroke faces created using the generated models, we aimed to create robotic faces representing the AFE characteristics of stroke for healthcare education purposes. Thus, we developed an end-to-end stroke analysis-modeling-synthesis (AMS) framework, which extracts facial landmarks and audio from a video of an operator performing NATs, masks the landmark locations, and then streams the audio and masked landmarks to the robot's face to display signs of stroke. See Figure 5.4 for an overview.

For each source video, we used Live Link Face (LLF) (developed by Unreal Engine) to track the facial landmark locations of the operator's face and extract a file that stores the tracked values.

Next, we automatically masked the operator's facial parts in each source video via the generated models. To do this, we developed a Java program that analyzes the LLF generated file, and can incorporate a region-specific Stroke FPM model generated in Section 5.4.3to depict stroke in each facial region (eyes, brows, or mouth) for each facial side (left and right). For each specific facial region and facial side, the system multiplies the region-specific model into the movement values of facial points within that facial region for the selected facial side. A model value of less than 1 means the movements of facial points within that region will be dampened, and a value of higher than 1 means the movements will be more attenuated.

The system translates the streamed masked facial movements file into parameters that are readable for the robot in real time, to display masked stroke facial movements on the robot's face. We used the Furhat robot for this task, because it allows for real time rendering of dynamic facial expressions, head movements, and speech [3]. Figure 5.5 shows sample frames from this part of the system.

Finally, we followed best practices in the literature for evaluating synthesized facial expressions [59, 165](See Section 5.2).

5.6 Evaluation

In order to assess Stroke FPM and AMS, we conducted an expert-based user study with clinicians.

Specifically, we sought to address the following research questions:

RQ 1: How realistic are the region-specific FPM models applied to a robot for synthesizing signs of stroke?

RQ 2: How similarly does the AMS framework display the synthesized expressions on the robot's face compared to those of real patients with stroke?

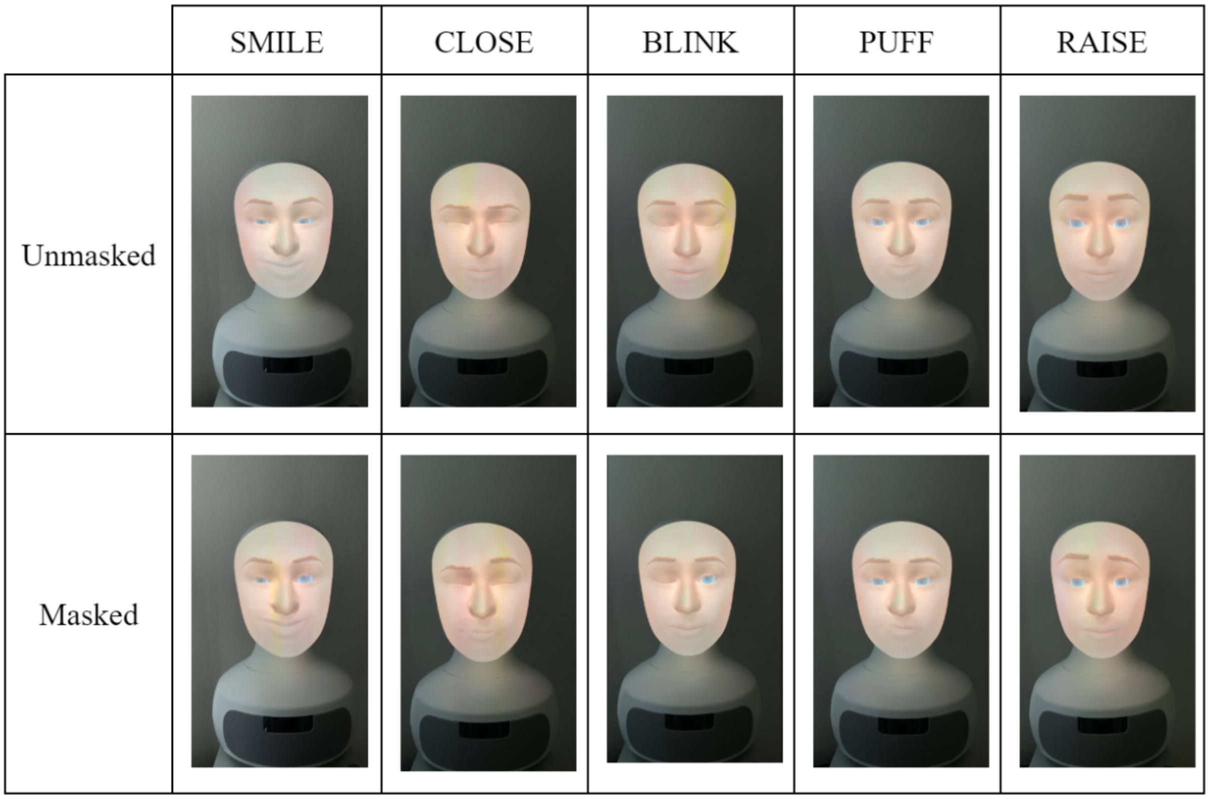


Figure 5.5. Sample frames from our robot. The first row shows unmasked expressions, and the second row shows the robot's face masked by a stroke model. From left to right: SMILE, CLOSE, BLINK, PUFF, and RAISE.

RQ 3: Which model more reliably represents acute neurological injuries in each facial region?

RQ 4: According to the impression of clinical educators, what can be changed to make interacting with the robot closer to the experience of interacting with a human?

5.6.1 Stimuli creation

We video-recorded an operator without stroke[1] performing five NATs (RAISE, CLOSE, BLINK, PUFF, and SMILE) required for assessing stroke facial paralysis. This resulted in five source videos, each five seconds long.

[1] Because eventually, we would like CEs to operate the robot using this stroke synthesis approach in a clinical simulation context, it was essential to study the likely expressions clinicians would make, and how their faces might appear when masked by our stroke models.

For each source video of the operator performing NAT_T, $1 \leq T \leq 5$, I extracted the facial points (See Section 5.5).

Without loss of generality, we assumed the operator wanted to create stroke-like facial movements on the left side of the face.

Therefore, for each source video of the operator performing NAT_T, we applied 15 prebuilt stroke models (SSM_i-$NAT_{T'}$) to the facial point values on the left side of the face, where T' is an NAT task performed by patients, $1 \leq T' \leq 5$, $T' = T$, and $SSM_1 \leq SSM_i \leq SSM_{15}$. For example, we masked the video of the operator performing the task SMILE using 15 different models obtained from patients performing the smile task: SSM_1-$SMILE$, SSM_2-$SMILE$, $\ldots$, SSM_{15}-$SMILE$.

This process led to generating 75 masked facial movement files.

For each masked facial movements file, we ran our automatic synthesis framework to synthesize the masked expressions to the Furhat robot's control points, and video recorded the robot's performance. At the end of this step, we had 75 stimuli videos of the RPSwS, where each video was 5 seconds long.

5.6.2 Participants

We conducted a study with seven physicians to assess the system's usability, specifically for CEs training in neurological diagnosis and treatment skills. We only report data from four participants who fully completed the survey. Three participants were between 34 and 44 years old, and one was between 24 and 34. They had, on average, four years of face-to-face interaction with patients, and all had encountered PwS in their careers. Additionally, all had completed a US-based medical education and had a medical specialty in Neurology.

5.6.3 Procedure

Participants completed a structured online questionnaire via Qualtrics, which probed their impression of the robotic stroke synthesis system across several dimensions. At the beginning of the study, participants received a summary of the project and instructions on completing the survey. They also completed a brief practice of the task, where they watched a test video of the robot with a neutral non-stroke face performing an NAT and answered some questions. This helped contextualize the robot and its functionality, which was important because physicians may be unfamiliar with robotic technology and may therefore have difficulty imagining how people might interact with it. Then, participants viewed the 75 stimuli videos (See Section 5.6.1) in random order, split between five blocks (one block for each NAT), with a short break in between. (See Appendix B.1).

5.6.4 Measures

After viewing each video, participants completed the similarity and realism measures, described below. Then, at the end of the study, they responded to several qualitative questions.

Similarity rating: Participants were asked "*Compared to a real stroke patient, how similar does this video look?*".

Realism rating: Participants were asked "*How realistic does this video look?*".

The participant provided the similarity and realism ratings on a 4-point Discrete Visual Analogue Scale (DVAS). A one on the scale corresponded to "not at all similar/realistic to real patients", and a four on the scale corresponded to "very similar/realistic to real patients".

Qualitative Feedback: At the end of the study, participants responded to several open ended questions asking how useful the system could be for clinical education, and also asking for suggestions for improvement of the robot.

5.7 Analysis

We wanted to explore the masks with the best similarity ratings and realism ratings and identify the best mask for each facial region. Thus, our dependent variables included similarity rating and realism rating, measured at the ordinal level, each has four categories, and the odds of falling into a higher or lower category are the same across categories. The independent variables included SSM and NAT, which are nominal. The filtering variables included Region and the two dependent variables. The reason is that the Region and SSM variables are highly correlated, and thus, by filtering data based on the Region variable, we avoid multicollinearity.

After exporting the report data from the survey, we generated a Python script to parse the output data based on the filtering variables, breaking it down into six dataframes. Then, we used an ordinal regression model in SPSS to analyze the data in each dataframe.

Realism (RQ1): We explored the relationship between participants' 4-DVAS realism ratings and their respective facial regions to examine how realistically the AMS framework displays the synthesized expressions on the robot's face as compared to those of real PwS for each facial region. We analyzed the marginal percentage for each realism rating category from the Case Processing Summary table from the SPSS reports.

Similarity (RQ2): We examined the relationship between participants' 4DVAS similarity ratings and the facial regions to measure how similar the synthesized expressions on the RPSwS are perceived to be to those of real PwS for each facial region. Here, we analyzed the marginal percentage for each similarity rating category from the Case Processing Summary table from the SPSS reports.

Model Selection based on Similarity Ratings (RQ3): We studied the relation between the SSM_NAT models used for masking each facial part and participants' 4DVAS similarity ratings to identify the model that more similarly presents acute neurological injuries on the corresponding facial region of the RPSwS compared to the ones on PwS' face. To do this, I interpreted the estimated value for each case of independent variable NAT and SSM as linear regression. This enabled me to identify the likelihood of each case falling into the higher category of the dependent variable (similarity rating), leading to detect the NAT case and SSM case with the higher likelihood of being ranked with a higher similarity rating category.

Qualitative Findings (RQ4): One researcher performed thematic coding on open-ended question responses, in which they reviewed users' answers and rendered high-level themes to represent key ideas in the data. This enabled us to identify which parts of the FPM and AMS

Table 5.4. This table reports the marginal percentage values of realism ratings for each facial region and overall face.

Realism rating	**Mouth**	**Eyes**	**Brows**	**Overall**
Not at all realistic	11%	10%	10%	10.3%
Slightly realistic	12%	16.7%	18.8%	15.7%
Moderately realistic	36%	31.7%	26.3%	31.7%
Very realistic	**41%**	**41.7%**	**45%**	**42.3%**

frameworks need improvement before running larger studies in the future.

5.8 Results

We provide descriptive statistics of the relevant variables and metrics, including the estimate and significance values.

5.8.1 RQ1: Clinicians' Realism Perception

In order to compare users' realism rating with facial regions, we provide descriptive statistics of the relevant variables, summarized in Table 5.4. From Table 5.4, for each facial region and overall, the highest marginal percentage statistic implies the perceived realism rating that the majority of participants rated the corresponding region. The results indicate that the majority of expert human judges perceived our stroke models used to mask all three facial regions of the robot and its face overall as very realistic.

5.8.2 RQ2: Clinicians' Similarity Perception

In order to compare participants' similarity rating with facial regions, we report descriptive statistics of the relevant variables, summarized in Table 5.5. Based on Table 5.5, clinicians perceive the models used to mask the overall face as moderately similar to real PwS. They

Table 5.5. This table reports the marginal percentage values of similarity ratings for each facial region and overall face.

Similarity rating	Mouth	Eyes	Brows	Overall
Not at all similar	15%	14.2%	17.5%	15.3%
Slightly similar	32%	36.7%	**42.5%**	36.7%
Moderately similar	**41%**	**41.7%**	27.5%	**37.7%**
Very similar	12%	7.5%	12.5%	10.3%

indicated that the stroke synthesized expressions displayed on the mouth and eyes regions of the robot are moderately similar to real PwS. They also found expressions in the brow region of the RPSwS as slightly similar compared to those of actual people with stroke, which suggests that the models for this region can benefit from further improvement.

5.8.3 RQ3: Model Selection based on Similarity Ratings

As for the mouth region, the stroke model associated with NAT_1 (exp(B) $= 1.98$, 95

For the eyes region, the stroke model created using NAT_1 (exp(B) $= 2.14$, 95Finally, for the brows region, the model built using NAT_3 (exp(B) $= 3.77$, 95

Thus, the masks that more reliably represent stroke include: SSM_1-NAT_1 FPM to mask the mouth region, SSM_7-NAT_1 FPM to mask the eyes region, and SSM_9-NAT_3 FPM to mask the brows region.

The masks that more reliably represent stroke include: SSM_1-NAT_1 FPM to mask the mouth region, SSM_7-NAT_1 FPM to mask the eyes region, and the SSM_9-NAT_3 FPM to mask the brows region.

5.8.4 Qualitative Findings

We asked participants two open-ended questions (see Section 6.6.4). One main goal was to gauge their impressions of and preferences for operating and interacting with RPSwS in clinical education scenarios. Another goal was to identify parts to improve to make interacting with the RPSwS closer to the experience of interacting with PwS based on experts' impressions.

The overarching theme across all responses to the first question was that clinicians found the RPSwS "useful" in helping CLs learn the symptoms of acute stroke. For the second question, responses were more detailed, and to analyze them, we employed grounded theory [78], and found emerging themes through an inductive coding process. We then compared codes and identified five overarching themes among the participants, specifically relating to visual presentation, patient similarity, robot behavior, the selection of patient vignettes, and the importance of training.

Visual Presentation. Having a realistic visual presentation and embodiment can significantly impact the user's impression of the robot. Participants perceived faces on the RPSwS as very realistic for each facial part and for the face overall (See Section 5.8.2). However, one participant noted: "At times, it is difficult to distinguish the different facial muscles; consider adjusting the contrast." This was likely due to complications with the video recording of the robot with a bright face projector; however, this challenge will be mitigated once participants interact with the physical robot.

Patient Similarity. Being able to display accurate facial cues of stroke similar to real patients is critical for stroke diagnosis. Overall, participants perceived the faces on the RPSwS as moderately similar to real PwS (See Section 5.8.1), and no participant had difficulty identifying that the robot was displaying signs of stroke. One participant suggested: "the facial droop could be more pronounced". Incorporating facial droop around the mouth and cheeks area can enable the robot to show signs of stroke more similar to PwS.

Robot Behavior. For the purpose of this study, and to assess region-based SSM features, we employed a technique of manipulating one region of the robot's face at a time for each video. One participant suggested: "consider each mask doing all of the movements to better assess the pattern of weakness. For example, have one video where someone smiles, blinks, raises eyebrows, etc." While displaying stroke in each specific region of the face enabled me to identify the most reliable SSM feature for each region (See Section 5.8.3), simultaneously depicting visual signs of stroke on the entire face will make the robot's behavior more realistic.

Patient Vignettes. Our participants were mindful of suggesting providing supplementary materials to end users in future studies. One clinician suggested "consider adding a sentence or two of clinical background." In healthcare education and training, the patient's medical history is an important factor to interpret, and clinicians can learn some extrinsic factors from it.

Training It is essential to provide adequate training on how to use technology, particularly to users who lack sufficient experience in interacting with robotic systems. Without adequate training, users may misunderstand the robot's functionality, overestimate or underestimate its capabilities, or use the system incorrectly, leading to unreliable evaluations. One participant noted that I "may need to show a video of the AI model doing all the commands without deficits prior to [the] session for the participants to know how it looks baseline." Therefore, providing training sessions that cover the system's functionality and modes of interaction are necessary to ensure users can provide reliable and informed assessments.

5.9 Discussion

This work offers multiple implications to several key research and practice communities, including clinical education, automatic face and gesture, and human robot interaction, which we discuss below.

5.9.1 Implications for Clinical Education and the Broader Healthcare Community

Our work provides crucial insights into developing robotic patient simulators capable of representing asymmetric facial expressions, providing opportunities for prompt diagnosis, and, thus, preventing serious harm. Overall, our study reveals that a significant majority of physicians expressed a desire for robots capable of replicating facial characteristics associated with stroke. This finding underscores the value of employing expressive RPS systems capable of depicting signs of stroke as educational tools for training healthcare professionals. The physicians' desire for such technology suggests its potential to enhance clinical education by providing expressive and realistic learning experiences.

Moreover, our research yielded insights into the best models to represent stroke in each facial region based on professional expertise. Our study has produced a set of 75 masks that represent stroke in various facial regions. This diverse set of options allows clinicians to evaluate and select masks that accurately represent stroke based on their professional expertise. Thus, this approach enhances the precision and reliability of stroke representations for different facial regions.

Additionally, identifying the best models provides researchers with the opportunity to create a novel universal stroke mask to model stroke for the entire face. By overlaying the universal mask on a robot's face, researchers can create robotic faces that represent more realistic and reliable representations of stroke symptoms for clinical education. This technology enables clinical educators to design scenarios for the robot that accurately simulate real-world clinical scenarios, providing a platform to effectively train and assess clinical learners' stroke diagnosis and treatment skills. The use of robots capable of representing stroke-related cues can potentially support enhancing the accuracy and effectiveness of learners' stroke diagnosing and treating skills.

The clinicians' interest in and abilities using this technology holds promise for the future of stroke care. The integration of expressive RPS with stroke into clinical education settings has the potential to revolutionize stroke care by offering innovative training opportunities. These RPS systems can provide healthcare professionals with hands-on experiences in diagnosing and treating stroke-related conditions, with the aim of improving their skills and knowledge in a realistic and controlled environment.

The analyses of signs of stroke on PwS can help advance clinicians' understanding of the effects of stroke on the face and improve the development of new stroke treatment methods that are specific to the facial muscles and expressions affected by stroke. Since PwS can often have complex medical needs [205], using a RPSwS can help clinicians focus on specific aspects of stroke treatment (such as facial exercises or neuromuscular stimulation to improve facial muscle strength) without the added complications (such as speech therapy or medication management).

Furthermore, clinical researchers can use an RPS with a stroke face to evaluate the effectiveness of various stroke treatments, such as medication or rehabilitation, on stroke patients' facial muscles and expressions. This can provide a safer, more controlled environment for clinical researchers to develop and test new stroke treatment methods on RPSwS, without the risks associated with working with human patients.

By better understanding the effects of stroke on the face, clinicians can design more adequate care plans for patients with stroke to target explicit facial muscles and expressions affected by stroke. Such a system can also support stroke patient recovery by enabling professional healthcare providers to develop personalized rehabilitation programs for the patient that target specific facial muscles and expressions, based on the patient's needs and conditions.

Researchers can extend our data-driven approach for stroke modeling to collect data and develop statistical models of A-FE caused by other conditions, such as Bell's Palsy and Parkinson's disease. They can achieve this by modeling their method after our FPM framework to develop a framework for generating computational models representing the A-FE characteristics of their specific condition. They can then use a synthesis system similar to the end-to-end stroke AMS framework to stream out their models to the robot's face to display the condition. This can enable the development of more effective interventions for diagnosing, treating, and monitoring these conditions.

Overall, using patient simulator robots that can facially convey stroke can help provide appropriate and timely stroke diagnosis and treatment, and ultimately improve patient outcomes.

5.9.2 Impacts for the FG Community

Furthermore, this work can have significant implications for researchers in the FG community.

We introduced a diverse set of 15 statistical measurement features for quantifying facial asymmetry and asymmetric facial movements in both the upper and lower facial regions. This set of measurement methods can provide researchers with a comprehensive and reliable framework for analyzing facial movements and expressions. This can provide the researchers with the opportunity to perform a more precise and accurate analysis of facial asymmetry in those affected by FP.

Furthermore, researchers in the FG community can use the presented methods in this work to create facial recognition technologies specifically designed to detect subtle asymmetries in facial features' locations and movements. Such technologies can enable more effective recognition and monitoring of neurological disorders that affect facial expressions, such as stroke, Bell's Palsy, or Parkinson's disease.

This work can facilitate cross-disciplinary collaborations between researchers in the fields of computer science and neurology, leading to the development of innovative techniques for future FG research.

5.9.3 Implications for the HRI Community

The development of an expressive robot capable of displaying asymmetric facial expressions has numerous potential applications in HRI research, ranging from enhancing people's perception of individuals with FP to understanding the effects of facial asymmetry on social interactions.

Many existing facial models are limited to symmetric movements. However, almost all human faces display some degree of asymmetry in their features and movements, due to variations in muscle strength on either side of the face, differences in bone structure, and genetic, aging, or medical conditions [255]. Using the modeling approach presented in this study to understand facial asymmetry can advance the design and development of robots that have more realistic and effective facial expressions. Overall, in the field of HRI, understanding and presenting asymmetric facial expressions is critical to improving the expressivity of the robot, enhancing its ability to convey real human-like emotions, and interacting with humans more naturally.

In social contexts, studies show many of those with FP do not seek or receive accurate information, treatment, support, and services because they have inaccurate perceptions of themselves or their conditions [40]. For example, their facial condition sometimes is perceived as a cosmetic condition instead of a medical condition, holding people with FP back from asking for proper treatment or service [40]. HRI researchers can use our techniques to develop robots with asymmetric facial expressions for normalizing A-FE for laypeople, enabling them to accurately perceive the facial characteristics of individuals with FP [210].

Additionally, HRI researchers can use this robot with asymmetric faces to better understand how laypeople perceive and respond to asymmetric facial expressions. Using this technology, they can study the effects of facial asymmetry on the perception of emotions and social interactions between humans and robots.

Our work establishes a versatile platform that facilitates the study of HRI in the context of facial asymmetry. Our system enables the robots to present a wide range of appearances and characters, thereby providing a platform to study users' perceptions of people with diverse backgrounds. This may also help normalize users' real-world communication and social skills, especially when engaging with individuals from diverse populations or those with facial asymmetry. Ultimately, my platform illustrates a valuable tool for advancing research in the field of

HRI.

5.9.4 Future Work

In future work, we will use the selected region-specific models to enable the robot to simultaneously show the characteristics of stroke in all facial parts, leading to a more realistic robot appearance and behavior. As we continue to improve the robot to more accurately represent signs of stroke, we will explore additional features, such as enabling the robot to display loss of blinking control, gaze deviation, and facial droop around the corner of the mouth.

5.9.5 Conclusion

Ultimately, the introduction of this system can significantly advance the fields of education, healthcare, HRI, and FG, with implications for both research and practical applications in healthcare and technology. This study can help advance medical education and training, leading to improvements in the diagnosis, treatment, monitoring, and rehabilitation of various clinical conditions. Designing and deploying advanced healthcare technology can enhance patient care, reducing healthcare costs, and eventually, improving patient outcomes. This study can lead to new insights and technologies to advance research in border HRI and FG communities and promote innovation by encouraging the development of new technologies in these fields.

5.10 Chapter Summary

This chapter presented my proposed FPM and AMS frameworks to model and synthesize stroke on the face of robots. Robotic systems depicting stroke faces can have implications in many domains, including human robot interaction, automatic face and gesture, and healthcare. The next chapter will present the development of an interactive, expressive RPS system capable of depicting stroke, with the aim of improving healthcare education and training.

5.11 Acknowledgments

I thank Drs. Jamie Labuzetta and Steven Yang for facilitating data collection from patients with neurological conditions, and providing feedback on my data analysis efforts. I thank Dr. Preetham Suresh who gave me insights into designing interactive expressive RPSs for the clinical simulation. I also thank Suhas Pai for his assistance with the FPM framework implementation and data annotation. This chapter contains material from “Mimicking acute stroke findings with a digital avatar” by M. Pourebadi, J. LaBuzetta, C. Gonzalez, P. Suresh, and L. Riek., which appears in STROKE, Vol.51, 2020 [192], and “Modeling and Synthesizing Stroke on Expressive Patient Simulator Robots” by M. Pourebadi, and L. D. Riek, which appears in Proceedings of the AAAI Artificial Intelligence for Human-Robot Interaction (AAAI AI-HRI), 2020 [197], and “Modeling and Synthesizing Stroke on Robotic Patient Simulators” by M. Pourebadi, J., Labuzetta, and L. D. Riek, which will be submitted to the ACM Transactions on Human-Robot Interaction (THRI), [191], and I was the primary investigator and author of these papers.

Chapter 6

ROSE: An Interactive Social Robot for Medical Education

6.1 Overview

In the previous Chapter chapter 5, I discussed the development of a new FPM framework for creating computational models of stroke, and an AMS framework for depicting patient-like stroke characteristics on the face of a physical robot. In this chapter, we will discuss our system engineering efforts to create ROSE: an interactive social robot for medical education, which allows CLs to practice their stroke diagnosis skills.

6.2 Introduction

Humanlike virtual and physical robots are increasingly employed in many settings, including hospitals and schools [183]. Their design and deployment are influenced by a range of socio-technical, economic, and contextual factors, which, in return, inspires research areas, especially in the realm of clinical applications [109]. Some recent examples of research topics here include cognitively assistive robots, providing social engagement for older adults, supporting telemedical care delivery in hospitals, assisting healthcare professionals in various tasks related to patient care, and RPS to support healthcare education and training (HET) [148, 163, 162, 113, 191, 195].

Humanlike robots, particularly RPS systems, are frequently deployed in clinical educational environments. See Chapter chapter 2 for more information about current RPS systems and their benefits and limitations.

However, existing RPS systems have limited capabilities for humanlike expressiveness, accurate representation of FP, and autonomous interaction [197]. The limited expressiveness of robots can impede emotional engagement, empathy, and social presence, leading users (such as CLs) to experience reduced motivation, interest, and retention of training content [185]. Moreover, the inability of robots to accurately depict FP and dynamic, realistic clinical scenarios can hinder clinical skill transfer and diminish the robot's utility, potentially compromising patient exams. Thus, it is essential for RPS systems, and humanlike social robots in general, to portray a wide range of expressions and conditions.

With these design paradigms in mind, we explore how to create an interactive, expressive RPSwS system to be used to teach CLs in clinical training contexts. In our prior work, we developed an expressive RPSwS system which reflected data-driven models of stroke. In this work, we take this system and make it interactive, to be used to teach CLs how to diagnose and treat neurological emergencies such as stroke. We have worked with neurologists and clinical educators to co-design an interactive robot that could depict neurological impairments and engage in autonomous interactions with CLs, with the aim of improving CLs' diagnostic and social skills. This robot allows CLs to have access to the repeated clinical practice of stroke diagnostic skills in a realistic clinician-simulated environment.

Most current RPS systems have a face with limited or no humanlike naturalistic expressivity. The anthropomorphization of social robots gives rise to concerns about the robot's impact on users' emotions, expectations, and interactions [199, 183, 114]. The limited expressiveness of humanlike social robots can lead to reduced emotional engagement with users, hindering the development of empathy and social presence in HRI [185]. Even for the RPS systems with expressive faces, their appearance and characteristics may not be widely customizable to address CEs' needs. These gaps call for researchers to be mindful of humanlike social robots' ethical and societal impacts.

Finally, current RPS systems may not promote autonomous interaction and engagement with CLs in a humanlike manner. As the purpose of existing RPS systems is mainly informative rather than interactive, they lack various communication modalities, which may limit the range and quality of interactions between the RPS robot and users, affecting the overall training effectiveness.

Existing RPS systems may offer tools to practice basic clinical skills (e.g., taking vital signs, and performing physical exams); however, they only partially simulate realstic dynamic clinical scenarios that replicate real-life medical situations with evolving and changing conditions. This can lead to limited opportunities for effective clinical skill acquisition and knowledge transfer, potentially resulting in missed opportunities for acute interventions, tools, prompt treatment, and prevention of serious harm.

Designing a clinical training tool with an interactive, expressive RPS to address these gaps, can also introduce design and technical challenges. For example, if poorly designed, interacting with the system can heavily rely on advanced technology or complex interfaces, limiting clinicians' perceived ease of use. The perception of robots' ease of use may significantly influence the clinicians' acceptance of new technology in their professional life [109]. Moreover, the limited usability of robots can make it challenging for users to effectively work with the robot and access the training content, ultimately resulting in frustration and lower learning outcomes.

These challenges necessitate a new interactive clinical training tool to enhance the learning experience of CLs, which provide a realistic and immersive environment for practicing dynamic clinical scenarios. In addition, such a tool must facilitate customization of the robot's expressions, appearance, and clinical scenarios, to encourage its usability, adoption, and accessibility.

To address this need, we introduce ROSE: an interactive social robot for medical education (See Figure 6.1). ROSE consists of two parts. First, it leverages the stroke AMS and FPM frameworks to generate computational mask representations of stroke, which are fully patient-

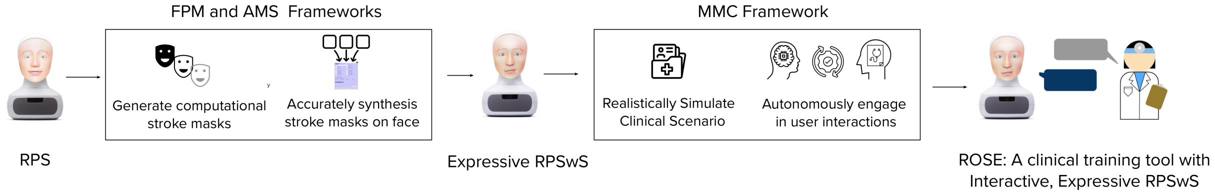

FPM and AMS Frameworks
Generate computational stroke masks
Accurately synthesis stroke masks on face
MMC Framework
Realistically Simulate Clinical Scenario
Autonomously engage in user interactions
RPS
Expressive RPSwS
ROSE: A clinical training tool with Interactive, Expressive RPSwS

Figure 6.1. ROSE: An interactive social robot for medical education. ROSE consists of two parts: Part A, which is an expressive RPSwS, and Part B, which is a framework to transform the robot into an interactive tool for clinical education.

data-driven (See Chapter 5) Next, it uses our newly developed multi-modal communication (MMC) framework to simulate clinical scenarios and autonomously engage in user interactions.

Using the Stroke FPM and AMS frameworks, we built an expressive face for ROSE with customizable characters and appearances, which can realistically depict patients with acute stroke. To demonstrate our system, we implemented it on a Furhat robot [23] in the context of developing stroke diagnosis training tools for CLs.

To build the MMC framework and create communication capabilities for the robot, we teamed up with neurologists interested in developing new technologies to support CLs. We co-designed adaptable clinician-defined controls that enable CEs to define specifications for ROSE and incorporate the types of real-world dynamic scenarios to the robot they view as clinically relevant. We then developed the MMC framework and a keyword-based control mechanism for our system, enabling ROSE to autonomously interact and engage with CLs in real time. These capabilities may empower CLs to practice their diagnostic acumen and communication proficiencies by engaging in hands-on practice with the robot.

We evaluated ROSE with engineers and clinicians. Overall, all participants with and without prior clinical experience successfully interacted with ROSE, and were able to perform neurological assessments required to assess signs of stroke. Moreover, they reported positive comments with regard to the robot's appearance, conversational dialogue, interaction, and usability. Furthermore, they gave detailed suggestions for improvement, which we discuss in Section 6.7.

The contributions of this chapter are as follows: First, we present the collaborative user-centered design requirements for building ROSE (See Section 6.4), offering insights for HRI researchers and developers of interactive, expressive robots, and encouraging them to adapt these design requirements to their own applications.

Second, we introduce a multi-modal communication framework to automatically simulate clinical scenarios, and enable autonomous interaction and engagement on ROSE (See Section 6.5). We developed an interactive, expressive RPSwS as the platform of the robot, with customizable expressions, appearance, and characters, to enable the robot to be more humanlike and realistically portray PwS. We then incorporated a real-world dynamic scenario into our system to suit the users' needs better, and developed a keyword-marching control mechanism to enable the robot to interact with CLs automatically. This framework enables the robotics community to leverage our approach to customize a robot's autonomous behavior, adapt to user needs, and promote effective HRI within their own application domains.

Third, we present ROSE to create a realistic and immersive environment for practicing diagnosis and treatment skills, offering opportunities for repeated clinical practice, and promising avenues for advancing the capabilities of clinical learning. We report the results from pilot studies and interviews with robotics engineers and clinicians to investigate the efficacy of our tool for depicting dynamic clinical scenarios (See Section 6.6), and report the results revealing how they envision using ROSE for stroke diagnosis (See Section 6.7).

To our knowledge, ROSE is the first of its kind, representing an exciting new area of research. This work has implications for HET, as well as the broader healthcare and HRI communities. Employing robots in this way may improve CLs' communication and diagnosis skills, and ultimately improve patient outcomes. Moreover, our work provides a framework for researchers to explore HRI in new experiential learning settings (e.g., build RPS systems to enable CLs to avoid forming biased impressions) and broader domains (e.g., explore methods for designing social robots for enhancing people's perception of individuals with FP, understand the effects of facial asymmetry on social interactions). We discuss the implications of these findings and future work directions in Section 6.8).

6.3 Background

Chapter 5 discussed the development of FPM and AMS frameworks to model and synthesize stroke on the face of robots, and Chapter 2 described the importance of using expressive robots for HET. To our knowledge, there appears to be only one other notable expressive RPSwS system designed apart from our own, which was presented by Daher et al. [88].

This system was comprised of a physical-virtual agent head, that can display some stroke-related cues on its face, from which we take inspiration [88]. Using a human-in-the-loop control system, this system can change facial appearance, listen, speak, and react physiologically in response to CL's behavior. Notwithstanding the advancements made by the system, the robot could only perform a limited number of facial movements representing NATs (e.g., smile, frown, raise eyebrows). The robot's responses were limited to specific information and simple responses (such as yes or no). The robot also used manual control for verbal responses and thus lacked autonomous interaction capability, which may adversely affect perceived realisticness and engagement. Consequently, the existing system had limitations in adequately simulating the automatic and interactive scenarios necessary for comprehensive stroke diagnosis and treatment training. These limitations may impede the development of effective training programs for CLs, eventually reducing patient outcomes and healthcare efficiency.

The above mentioned challenges underscore the need for novel and reliable clinical training tools to train CLs in evaluating and managing stroke. The intricacies of clinical care in the context of stroke necessitate the development of HET tools for maximizing the applicability and versatility of robotic systems in stroke diagnosis and treatment training.

Table 6.1. Our proposed design requirements for situating and supporting ROSE in HET and broader HRI.

Component	Robot Design Considerations
Dynamic Clinical Scenario Simulation	Robots should enable active learner engagement throughout simulating standard clinical scenarios of assessing diverse NATs in a flexible order and repetition, tailored to the learner's behavior.
Expressivity and Accurate Representation of FP	Robots should be able to synthesize realistic human-driven depictions of non-verbal cues on its face in different degrees of severity and variations of FP.
Adaptability	Robot's character, appearance, and features should facilitate adaptability based on experts' opinions and users' educational needs.
Accessibility and Equality	Robots should provide learners with inclusivity and equal repeated access to similar comprehensive educational experiences, promoting fairness within simulation-based learning and performance assessment.
Multimodality of Communication	Robots should be proficient in interacting with individuals through various humanlike communication modalities, to foster effective information exchange and communication.
Increase Engagement and Immersion	Robots should have social abilities to enhance engagement and immersion in clinical learning.
Automation	Robots' control systems should support automatic operation and interaction with the end-users.
Perceived Ease of Use	Robots should be easy to use for learning purposes, with intuitive controls enabling CLs to obtain the necessary behavior responses from the robot without necessitating extensive technical knowledge.
Robot's Perceived Usefulness	Robots' design should consider the subjective appraisal of the system's usefulness by end-users.

6.4 Design Requirements for ROSE

In this section, we present the design requirements which informed the development of ROSE, through close collaboration with clinicians and experts in the field. We teamed with neurologists and clinical educators from other specialties who shared our interest in designing and developing expressive interactive robots to support HET. Figure 6.2 (presents our collaborators). We actively participated in in-depth interviews with them, to gain a deeper understanding of and invaluable insights into the intricate landscape of HET, particularly in the realm of stroke diagnosis and treatment. Additionally, we reviewed relevant scientific literature, guidelines, stroke assessment protocols, and existing clinical training sessions to refine our comprehension. We also held numerous meetings and participated in iterative design discussions with stakeholders to incorporate their feedback.

During our interactions with neurologists, we had the opportunity to explore various aspects of stroke diagnosis, including the methodologies they employ when diagnosing and treating real PwS in hospital settings. We also identified the challenges they encountered in this context and gained a deeper understanding of the role patient simulators play in training healthcare professionals in stroke diagnosis skills within educational environments. By delving into the conversations with clinical educators, we explored the functionality and workings of these simulation tools, while also identifying areas for improvement and unmet needs within this field.

This collaborative and iterative process enabled us to create comprehensive design requirements based on scientific knowledge and tailored to the needs of clinicians. Our primary design requirements encompassed the need for the robot to represent clinical scenarios while exuding expressiveness, adaptability, accessibility, user-friendliness, autonomous interaction, sociability, captivating engagement, and benefits for the intended objective (See Table 6.1 for an overview).

6.4.1 Co-designing Clinical Scenarios

Neurologists and clinical educators discussed the importance of simulating realistic clinical scenarios in HET, necessitating the co-design of dynamic clinical scenarios to simulate them on ROSE. Dynamic clinical scenarios are situations that involve ongoing changes, interactions, and complexities of the patient's condition. The scenarios also require real time decision-making

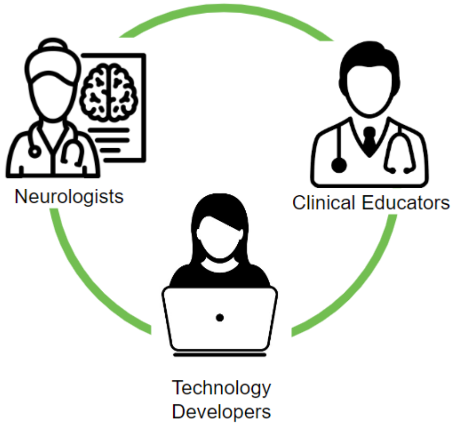


Figure 6.2. We collaborated with neurologists and clinical educators to identify contextual information of real clinical scenarios used for training and assessing CLs' communication and diagnostic skills, and co-created learning scenarios for ROSE.

and adaptability. The goal of co-designing dynamic clinical scenarios with neurologists is to enable ROSE to simulate real-world scenarios CLs will experience while assessing human patients. This allows the CL to actively engage in skill practice with the robot. The design of the scenario should allow learners to practice a diverse set of NATs required for assessing stroke. It should offer the flexibility to alter the order of NATs within the scenario and to repeat performing an NAT, according to the learner's behavior.

Leveraging our collective expertise, we co-designed an interactive neurological assessment scenario for ROSE, modeled after real-world clinical scenarios commonly employed for assessing PwS. The scenario is a dialogue script between a CL, defined here as the user, and an admitted stroke patient embodied as the robot. The dialogue script aims to enable the robot to establish a meaningful interaction with a CL, enabling them to conduct neurological exams. To enhance the authenticity of the interaction, we incorporated non-verbal behaviors into the scenario, thereby heightening its realism and providing a more lifelike experience for the user.

6.4.2 Support Expressivity and Accurately Represent Facial Paralysis

Because we are interested in demonstrating asymmetric facial movements characterizing FP, our collaborators emphasized the importance of enabling the robot to synthesize a realistic patient-driven depiction of non-verbal cues on its face. Ideally, the robot should also present different degrees of severity (e.g., normal symmetry, minor paralysis, partial paralysis, unilateral complete paralysis, and bilateral complete paralysis) and variations of FP (e.g., the ability to change the left or right range of FP, asymmetric blinking, facial droopiness, and gaze deviation).

6.4.3 Adaptability, Accessibility, and Equity

Our clinical collaborators stressed the importance of how the robot's character, appearance, and features should facilitate adaptability based on CEs' expert opinions and CLs' educational needs. The robot should offer customization options to convey diverse clinical conditions and visual representations (such as a range of genders, complexions, and ages), and features (such as different facial features, voices, and names).

The robot's adaptability provides an opportunity for CLs to practice on robots representing patients with diverse backgrounds and appearances, which will help further fairness and equity within simulation-based learning.

6.4.4 Multimodal Communication

Our collaborators discussed the importance of incorporating effective communication capabilities for establishing connection and trust between the robot and CLs, improving the overall user experience. Thus, it is important the robot can employ multiple modes of communication to accommodate each user's preferences and needs. Doing so will lead leading to more adequate comprehension and engagement. Supporting multimodal communication also enables the robot to adapt to diverse social and cultural contexts, further supporting inclusivity. Moreover, proficient interaction can foster effective information exchange between the robot and users, resulting in improved outcomes and user satisfaction.

Thus, the RPS system should be proficient in interacting with individuals through both nonverbal and verbal humanlike communication modalities. For example, the robot's face should be expressive and humanlike, in order to establish meaningful social interactions and engage with users. Additionally, the robot should support non-verbal backchannel cues to make for more naturalistic interaction, such as blinking, gaze, head nods, etc., as these are essential in building rapport.

Ideally, the robot should employ voice recognition and natural language processing to understand the spoken input from the user to be able to participate in the dialogue actively.

Furthermore, designing robust speech synthesis capabilities for the social robot is crucial, as speech serves as a vital source of information exchange (e.g., providing instructions and explanations) and communication (e.g., engaging in conversation).

6.4.5 Interaction to Increase Engagement and Immersion

Based on a review of the literature and conversions with our clinical collaborators, it is essential to develop social abilities in the robot to improve engagement and immersion in HET. Research suggests that social robots can improve learners' engagement [222] and lead to smoother and more positive social interaction [270]. In the HET context, a significant correlation exists between engagement and the improvement of teaching effectiveness [225].

Therefore, providing the robot with social abilities is critical to enhancing engagement and immersion in clinical learning. Achieving this objective involves the design of a robot with social presence and embodiment, characterized by attributes such as having a humanlike appearance and behavior. Moreover, it requires equipping the robot with advanced natural language understanding and generation, and incorporating interactive activities such as gamification into the learning process. Additionally, it is essential to enable the robot to personalize its behavior and content to meet CLs' needs and interests, and provide timely and constructive feedback, such as personalized feedback on CLs' progress.

6.4.6 Automation

Clinical educators discussed the importance of enabling automatic operation for the robot to support CEs' workload. This includes depicting FP, being able to respond to CL-initiated neurological assessment tests, and autonomous interaction with the end-users (e.g., attending to the user, and socially engaging in conversation). Thus, the robot's control system should support it by automatically adapting its behavior and performance based on CLs' responses, in order to make the interaction more realistic, engaging, and enjoyable for CLs. Automating elements of the control system has the potential to provide CEs more flexibility, reduce their workload, and free up their time, allowing them to focus on more critical matters such as scenario planning and CLs' performance assessment.

In the design of ROSE, most interactions are programmed and managed by researchers, and automatically performed by the robot, compared to the traditional training methods in which CEs must fully manage and perform all aspects of interaction.

6.4.7 Robot's Perceived Usefulness

Our collaborators indicated that the robot's design should consider the subjective appraisal of the system's usefulness by CLs and CEs. Enhancing learning outcomes, such as the accuracy of CLs' learning outcomes, the impact of CLs' expectations, and their social interaction skills and confidence, is vital in developing RPS systems. However, it is essential to consider the perceptions of CLs and CEs regarding how similar to real patients the system is and how realistic the interaction is. Ensuring the attainment of these design assumptions will culminate in a system that realistically replicates patients' behaviors and focuses on the needs of CLs and CEs.

6.5 Stroke Robotic Patient Simulator Prototype Design

Building on our design guidelines developed with clinicians, we developed a platform for ROSE using an RPSwS robot that can simulate facial movements, gestures, gaze, and dialogue similar to PwS in stroke evaluation scenarios. In this section, we discuss our prototype and its hardware and software components.

Since a key focus of ROSE is to convey facial expressions and engage in dialogue, we used the Furhat robot from Furhat Robotics as our robot platform. Furhat allows for real time

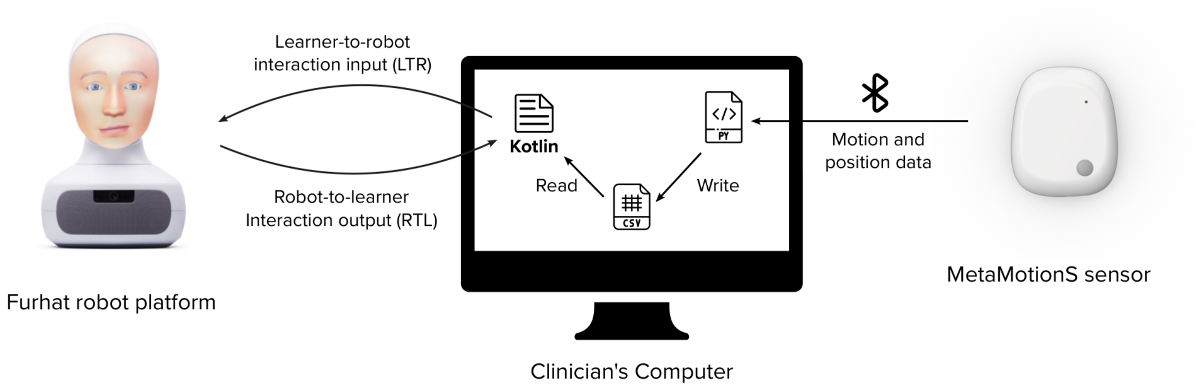


Figure 6.3. A high-level overview of our system's hardware components and interaction diagram of our software.

rendering of dynamic facial expressions, head movements, and speech [3]. Along with the robot, we used the MetaMotionS, an inertial measurement unit (IMU) device from MbientLab [2], which enables the robot to track the user's hand movements while performing the NATs required for assessing stroke. Figure 6.3 demonstrates a high-level overview of our system's hardware components and interaction diagram of our software.

6.5.1 Hardware

Robot platform

The Furhat robot has an organic unibody design [3], consisting of a head and neck. The robot's base contains a built camera, speaker, and microphone, and is situated below the neck of the robot, as shown in Figure 6.4.

Head and face. The size of the head is similar to most human heads, and the face can emulate a wide variety of facial expressions [3]. The outer surface of the head consists of two main parts: the face, made from a proprietary polymer with a smooth surface reflecting the geometry and textures of a human face, and the external shell, which makes up the cranium of the head, covering the internal components [3].

The internal structure of the head contains the projection system, which portrays facial features such as eyes, eyebrows, nose, and mouth on the face. This feature enables ROSE to display facial expressions, playing a crucial role in the development of our tool.

Neck. The lower portion of the head makes up the robot's neck, which encases a mobile platform, enabling the robot to move like a human's head (e.g., nodding and side-to-side motion). This platform contains three high-speed servos with active feedback, and metallic gears to provide three degrees of freedom (pan, tilt, and roll) [3], which enables the robot to move its neck [3].

Camera. The onboard RGB camera can record video, and the vision module can then process the feed in real time to detect human faces. Face detection allows the Furhat robot to "attend to" the nearest human, meaning that it rotates its head toward them, and its eye gaze will follow them, thus enabling more naturalistic HRI.

Speaker and Microphone. It also has a built-in speaker and a microphone that enable vocal communication between people and the robot. The robot uses a 30W power, built-in dual speaker system [3] to play all audio outputs and communicate with humans. For audio input, the robot uses one of two microphone systems. One is a built-in set of two omnidirectional digital microphones placed 180 mm apart on the robot's shoulders [3]. The other is a set of four omnidirectional digital microphones that can pick up voices from up to 5 meters away in all directions. This bundled microphone system has features like echo cancellation and background noise suppression that enhance the quality of voice recorded during interactions.

Wearable Motion Sensor

Per one of the NATs commonly used for stroke assessment, CLs may ask the PwS to visually follow their hand movements with their eyes, while keeping the head still. Similarly, to enable ROSE to track CLs' hand movements, we used the MetaMotionS, which is a wearable sensor that tracks motion continuously and in real time. The sensor is a 10-axis IMU that contains a gyroscope, accelerometer, and magnetometer. Using the Bosch Kalman filter sensor fusion algorithm, it combines inertial and magnetic readings from all three devices to determine real time sensor orientation. We developed a wireless connection script using the MetaMotionS

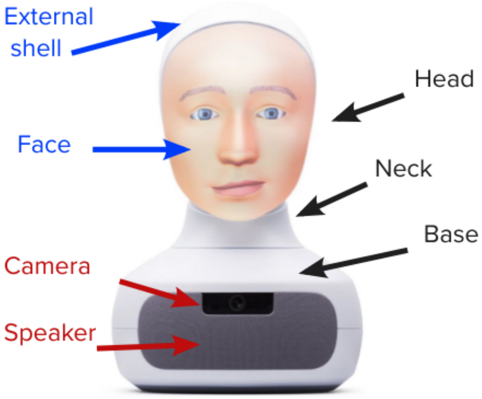


Figure 6.4. The Furhat Robot, and its components. The head consists of the face in the front, and the external shell at the back. The base of the robot contains the camera, speaker, and microphones.

open-source API in Python to enable the robot to communicate with the sensor through Bluetooth. Its light weight (0.2 oz) and miniature form factor make it a suitable choice as a wearable that can be worn by the CL on their wrist, with a band encasing it.

6.5.2 Software Architecture

We built the interaction software stack of ROSE using *FurhatOS*, a Linux-based system used as the Furhat Robotics' operating system for social robotics. This includes the entire set of subsystems required to run the robot and handles functions like facial animation, motion, audio processing, vision, etc. [3]. We wrote this in Kotlin, and it was connected to the Furhat platform and executed on the robot. We discuss some important aspects of our interaction software below.

Clinician-defined Controls

Upon running the interaction software on the robot to simulate a patient, the system can set the parameters for the clinician-defined controls defined by expert CEs, enabling them to define various aspects of the clinical scenario and the patient's characteristics. (See Figure 6.5 for an overview). We discuss these clinician-defined controls below.

Visual Appearance. Our interactive software allows us to portray any character with a

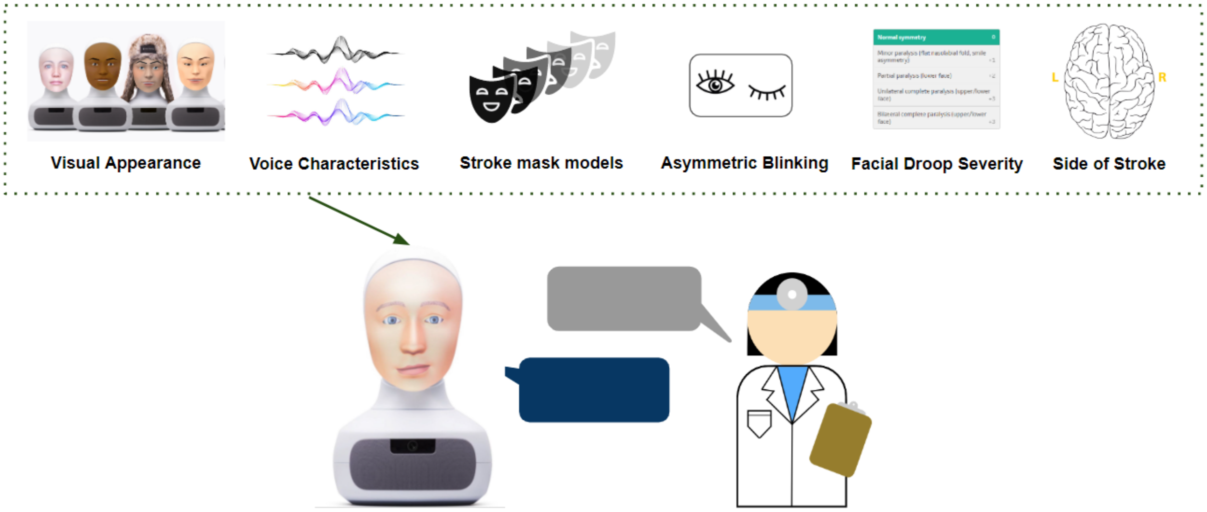


Figure 6.5. ROSE provides the flexibility to customize the parameters for the clinician-defined controls defined by expert CEs, enabling stakeholders to define various aspects of the clinical scenario and the patient's characteristics.

diverse set of apparent gender identities, complexions, ages, and degrees of humanlikeness, and display it on the face of the Furhat robot. The robot's operating system provides a standard set of face models consisting of diverse textures [3]. Developers may use similar configurations for aspects of the robot's visual appearance, or simply vary them from one scenario to another.

The use of these variations facilitates the creation of a greater range of intervention scenarios, leading to diversifying the robots and mitigating inherent biases that may creep into the learning process.

Voice Characteristics. In tandem with visual features, the Furhat platform provides the flexibility to change the robot's voice characteristics. FurhatOS provides support for speech synthesis in over forty spoken languages, in adult and child voices [3]. The system supports both built-in voices (Acapela [4] and Cereproc [50]) and cloud-based voices (Amazon Polly [41] and Microsoft Azure Speech [42]), while also providing pluggability support to easily extend to additional cloud-based voices as needed [3]. Similar to ROSE' visual appearance, developers may use a fixed parameter for the robot's voice, or vary it across scenarios.

Stroke mask models. Our interaction software allows for the selection and utilization of stroke mask models (described in Chapter 5), which are statistical models representing facial characteristics of stroke, developed through our prior research employing our stroke FPM and AMS frameworks. *FurhatOS* allows us to define and display custom actions and behaviors, by controlling different regions of the robot's face (e.g., eyes, eyebrows, cheeks, lower jaw, lips, etc.), thus enabling the robot to synthesize stroke mask models onto the robot's facial regions. At a high-level, clinicians will be able to configure which representation of the stroke to mask each part of the robot's face for their desired simulated clinical scenario.

Stroke clinical characteristics. The symptoms associated with stroke may manifest in different ways on a PwS' face. Clinicians can specify and simulate these manifestations on ROSE's face using the following settings:

1. *Side of the face with stroke:* Stroke can occur in either the left or right hemisphere of the brain. In most cases, a stroke on one side of the brain will lead to weakness or paralysis on the opposite side of the body and face. This phenomenon is known as contralateral motor deficits. Our system enables customizing the side of paralysis on the face.

2. *Blinking asymmetry:* PwS may experience loss of blink control on the affected side. The clinician can specify whether or not they want ROSE to display blinking asymmetry. If they do, they can define the paralyzed side of the face on ROSE (left or right), and the robot will exhibit blinking asymmetry accordingly.

3. *Facial droop severity:* Another symptom associated with stroke can be the presence of drooping on the edge of the lip on the affected side of the face. This can vary in its degree of severity (e.g., none, minor, partial, and complete). Based on the CEs' selections, the robot can exhibit facial droop depending on the specified side (left or right), and the identified level of droopiness.

4. *Gaze deviation:* Some PwS display a "deviated gaze" whereby their eyes will not be able to pan completely to one side of the face. Our robot supports three possible values of this

setting - intact gaze, in which there is no gaze deviation; fixed gaze, in which there is gaze deviation at rest (when the patient is not performing any NAT) in one direction; and gaze preference, in which there is constant gaze deviation in one direction, both at rest and while performing NATs. CEs can identify one of the three values for the deviated gaze and specify the direction of deviation (left or right).

Robot Autonomy (Finite State Machine)

Based on the co-designed dialogue script (See Section 6.4.1), we designed a finite state machine (FSM) (see Figure 6.6 to make interaction autonomous on ROSE. The FSM consists of well-defined states of the robot's behaviors and transitions between the states, which come together to simulate the aforementioned clinical scenarios. The robot uses a keyword-matching technique to identify the user's intent and transit between states. We classified the FSM states into two categories: NAT and Non-Assessment.

NAT States. The NAT states in the FSM refer to specific states that trigger corresponding NAT action performed by the robot upon entering those states.

1. *Rest:* The robot's face is relaxed, it looks at the CLs.

2. *Raise:* The robot raises its eyebrows.

3. *Puff:* The robot puffs its cheeks and holds for a few seconds.

4. *Blink:* The robot blinks its eyes.

5. *Close*: The robot closes its eyes very tightly.

6. *Smile:* The robot smiles wide.

7. *Gaze-Follow:* The robot makes its eyes follow the hand movement of the CLs.

The transition between the NAT states usually occurs as follows: the robot begins in the Rest state as the baseline. Receiving an input with an NAT performance request from the user

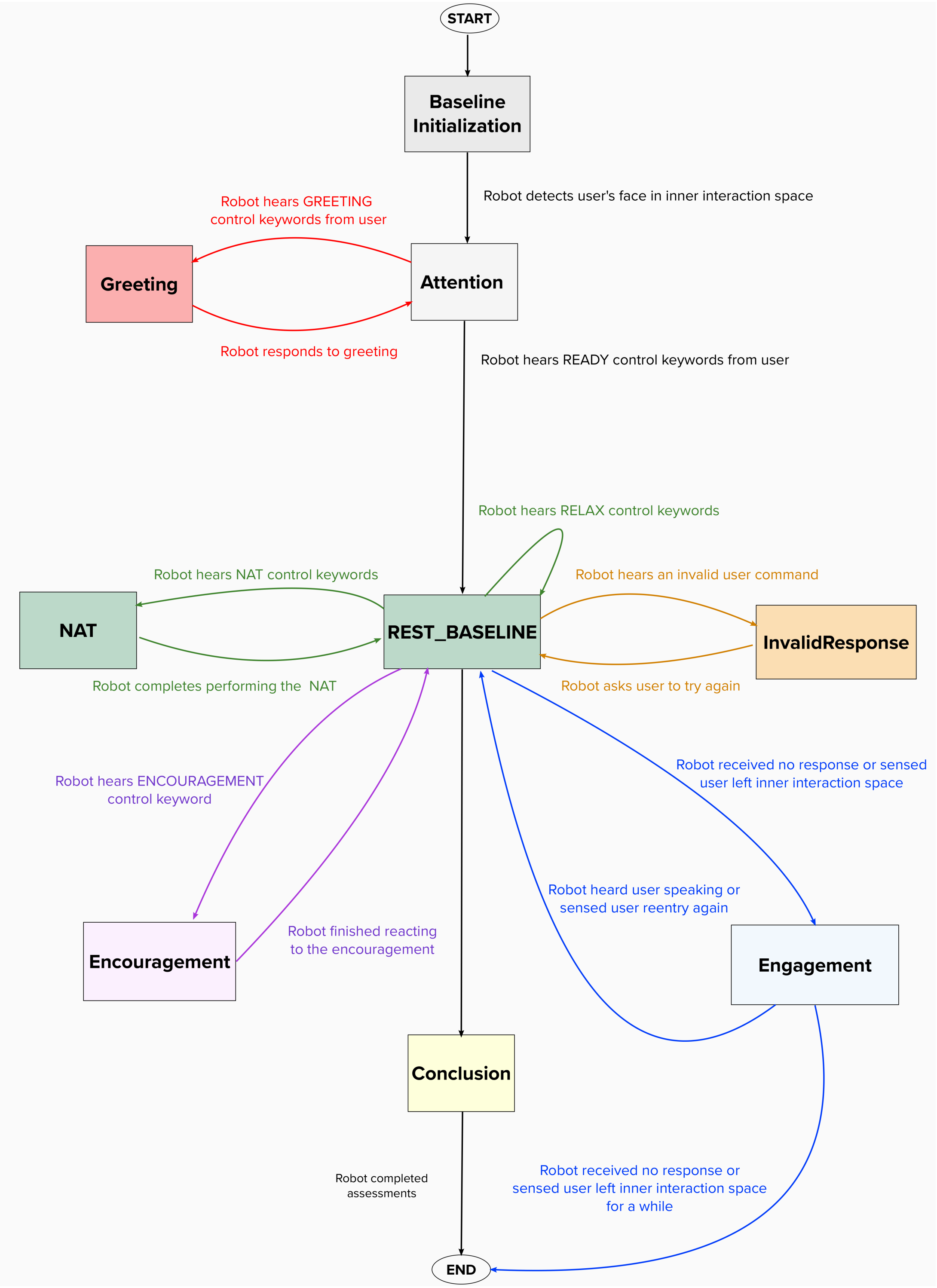

START
Baseline Initialization
Robot detects user's face in inner interaction space
Robot hears GREETING control keywords from user
Greeting
Attention
Robot responds to greeting
Robot hears READY control keywords from user
Robot hears RELAX control keywords
Robot hears NAT control keywords
Robot hears an invalid user command
NAT
REST_BASELINE
InvalidResponse
Robot completes performing the NAT
Robot asks user to try again
Robot received no response or sensed user left inner interaction space
Robot hears ENCOURAGEMENT control keyword
Robot heard user speaking or sensed user reentry again
Encouragement
Robot finished reacting to the encouragement
Engagement
Conclusion
Robot completed assessments
Robot received no response or sensed user left inner interaction space for a while
END

Figure 6.6. The Finite State Machine model of ROSE. The “NAT” state refers to any one of the six NAT States Raise, Puff, Blink, Close, Smile, and Gaze Follow.

can trigger the transition from the Rest state to any of the other six NAT states (e.g., Rest → Raise, Rest → Puff, etc.) After completion of a non-Rest NAT, the robot returns to the Rest state. Such transitions can continue until the CL ends the scenario or ROSE does not see or hear the CL in a while.

Non-Assessment States. The non-assessment states are not directly related to stroke assessment, but help improve the quality of human-robot interaction, to ease the learning process for CLs. These states include:

1. *Attention:* The robot detects the nearest human in its interaction space and automatically rotates its head to attend to them.

2. *Greeting:* The robot automatically responds to a greeting uttered by the CL. The robot randomly chooses a response from a set of predefined replies.

3. *Encouragement:* The robot automatically acknowledges an encouragement comment uttered by the CL and reacts by either nodding or smiling.

4. *Engagement:* The robot attempts to actively and attentively be involved in the interaction. Thus, if the user either stops talking or exits the interaction space of the robot, the robot requests the user's involvement in the interaction. For example, when a user leaves the inner circle of the robot, the robot requests the user to come closer. Or when the robot detects no responses in a while, it requests the user to talk.

5. *Conclusion:* When the user wants to end the conversation, the robot will transition to this state and end the interaction.

6.5.3 MMC Framework

We designed an automatic MMC framework that allows CLs to interact with ROSE as they would with a PwS. We integrated the clinician-defined controls and the FSM model together to produce a robot that can initiate a conversation with a person, listen to what the person says, respond appropriately, and perform NATs just like a PwS would. To enable the robot to participate in learner-to-robot interaction input (LTR), we leveraged the Furhat API to develop capabilities for speech recognition and face tracking, and developed our own system to support hand motion tracking. To enable the robot to establish robot-to-learner interaction output (RTL) effectively, we developed custom action and behavior synthesis and gaze movement synthesis capabilities, while adding non-verbal backchannel cues to further improve the quality of human-robot interaction. We also leveraged the Furhat API to support speech synthesis. (See Figure 6.3.)

Learner-to-robot interaction input (LTR). The CL interacts with the robot like they would interact with a PwS - through verbal conversation and eye contact. When the CL talks to ROSE, the robot records the audio through its microphone and uses speech recognition to understand the words. When the CL is within the interaction space of ROSE, it captures the video using its camera and uses face tracking to detect where to look to establish eye contact. When a CL asks ROSE to follow their hand movements, it tracks these movements and synchronizes its eye movements with the hand.

Robot-to-learner interaction output (RTL). We designed ROSE to automatically interact with the CLs based on the FSM and clinician-defined controls discussed earlier. Once the robot is in one of the NAT states, it will automatically display the corresponding action and behavior on its face. In any state (NAT or otherwise), the robot also responds to the utterances of the CLs through speech synthesis. At times, the robot uses non-verbal backchannel behaviors to convey attentiveness through attention and look alive through animations.

We describe each of these aspects of interaction below.

Speech recognition

The platform provides the recorded audio of the learner's utterance to the speech-to-text engine, and sends the output into a context-sensitive natural language understanding (NLU) component to detect a user's intentions through their utterances [3]. We defined custom categories

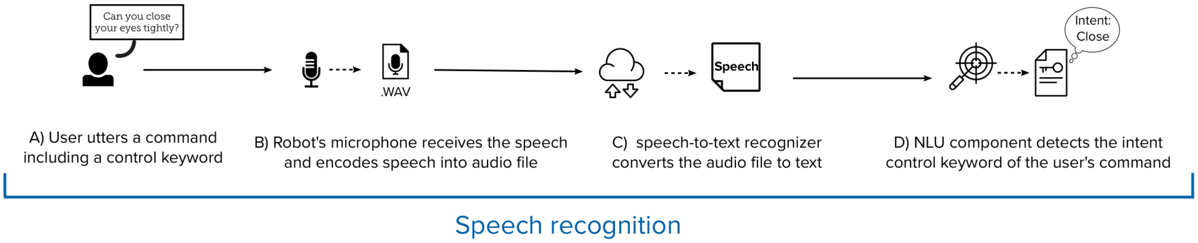

Can you close your eyes tightly?
.WAV
Speech
Intent: Close
A) User utters a command including a control keyword
B) Robot's microphone receives the speech and encodes speech into audio file
C) speech-to-text recognizer converts the audio file to text
D) NLU component detects the intent control keyword of the user's command
Speech recognition

Figure 6.7. LTR speech recognition using the robot's NLU component.

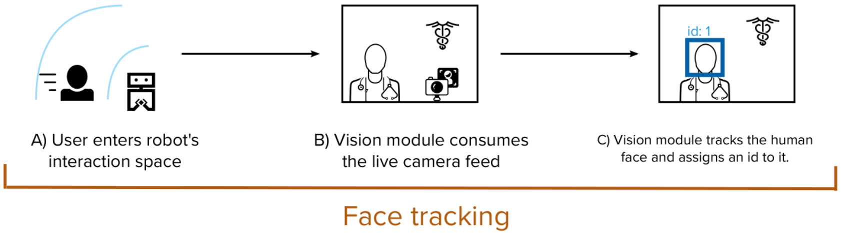

id: 1
A) User enters robot's interaction space
B) Vision module consumes the live camera feed
C) Vision module tracks the human face and assigns an id to it.
Face tracking

Figure 6.8. LTR face tracking using robots vision module.

of intent in our code, and used the NLU component of the robot to classify the utterance into one of the categories, enabling the robot to transition between states. For example, if the CL says, "Can you please close your eyes tightly?", the keyword *close* will trigger a custom intent that initiates a state transition from the Rest to the testitClose state (see Figure 6.7). In this manner, different utterances will lead to different, appropriate reactions and responses synthesized by ROSE.

User's face detection and attention identification

Our system uses the visual perception capabilities in FurhatOS to detect a user's face in a defined interaction space in front of ROSE and identify whether the user is attending to the robot. For this purpose, the live camera feed to the vision module automatically detects the CL's face (see Figure 6.8). We configured the interaction space, which in turn determined how close the users should be to the robot so that the robot starts interacting with them. The robot is also able to perform attention identification, where it recognizes whether a user is directly looking at the robot (i.e., attending to the robot).

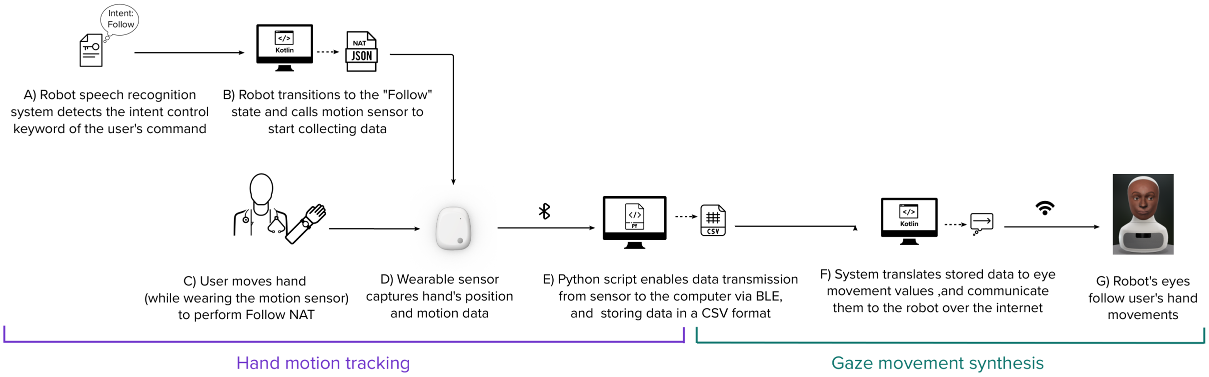

Intent: Follow
Kotlin
NAT
JSON
A) Robot speech recognition system detects the intent control keyword of the user's command
B) Robot transitions to the "Follow" state and calls motion sensor to start collecting data
C) User moves hand (while wearing the motion sensor) to perform Follow NAT
D) Wearable sensor captures hand's position and motion data
E) Python script enables data transmission from sensor to the computer via BLE, and storing data in a CSV format
CSV
Kotlin
F) System translates stored data to eye movement values ,and communicate them to the robot over the internet
G) Robot's eyes follow user's hand movements
Hand motion tracking
Gaze movement synthesis

Figure 6.9. Left: LTR hand motion tracking using the MetaMotionS sensor. Right: RTL gaze movement synthesis by the robot in response to the user's "Follow" command, where the robot's gaze follows the user's hand movements in real time.

Hand motion tracking

To track hand motion, we ask CLs to wear the MetaMotionS sensor on their wrist, and move their hand in desired directions as part of stroke assessment, while the wearable sensor captures their hand's position and motion data. We programmed this functionality in Python, using the Metawear API (developed by MbientLab Inc. [2]). Our program enables data transmission from the sensor to the computer via bluetooth, and stores data on a file in a CSV format on the user's computer. (See Figure 6.9: left). This enables ROSE to follow the CL's hand movement with its eyes.

Speech synthesis

In our MMC framework, we implemented the FSM states in a way that, upon transitioning to a state, it identifies the proper predefined speech response (if any). It then utilizes the text-to-speech engine to convert the response to the speech, encode the speech into an audio format, and play audio using its speakers. Our framework allows developers to customize parameters like language, voice, and accent in code for each scenario as desired.

Besides the audio being played on the speakers, our software will animate the face for visual speech movements such as lip, jaw, and tongue motion in synchrony with the corresponding audio. This gives the impression that ROSE is responding to the CL, enhancing HRI. (See

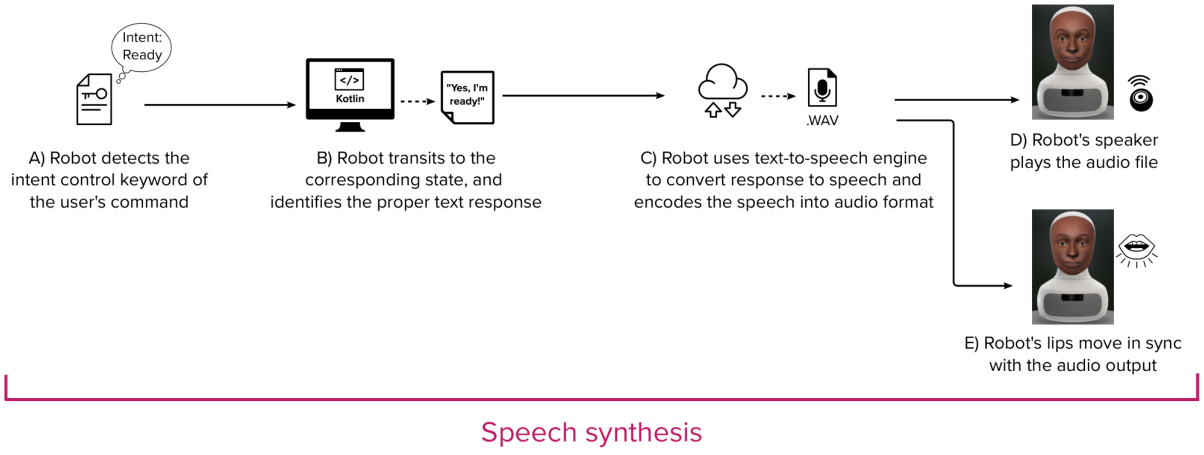

Intent: Ready
Kotlin
"Yes, I'm ready!"
.WAV
A) Robot detects the intent control keyword of the user's command
B) Robot transits to the corresponding state, and identifies the proper text response
C) Robot uses text-to-speech engine to convert response to speech and encodes the speech into audio format
D) Robot's speaker plays the audio file
E) Robot's lips move in sync with the audio output
Speech synthesis

Figure 6.10. RTL speech synthesis by the robot in response to the user's command

Figure 6.10).

6.5.4 Action and Behavior Synthesis

In NAT states, ROSE can interact with CLs by performing actions or behaviors such as raising its eyebrows, blinking, tightly closing its eyes, and more. While *FurhatOS* supports many of these as built-in actions and behaviors, the existing implementation does not reflect the manifestations of stroke on the robot's face. Hence, we defined our own custom actions and behaviors within the MMC framework based on the stroke mask models presented in Chapter 5. Using this functionality, upon identifying an NAT-related intent by the robot, the MMC framework synthesized the corresponding NAT assessment on the robot's face. (see Figure 6.11).

Non-verbal backchannel cues To make the interaction more intuitive and humanlike, we added non-verbal backchannel cues as part of the FSM logic that governs robot behavior.

We discuss these as two main categories of cues, attention and animated momentary behaviors.

Attention: We enabled the robot to change attention from time to time to make the interaction more lifelike.

Earlier in this section, we explained how the robot is able to detect the users' faces

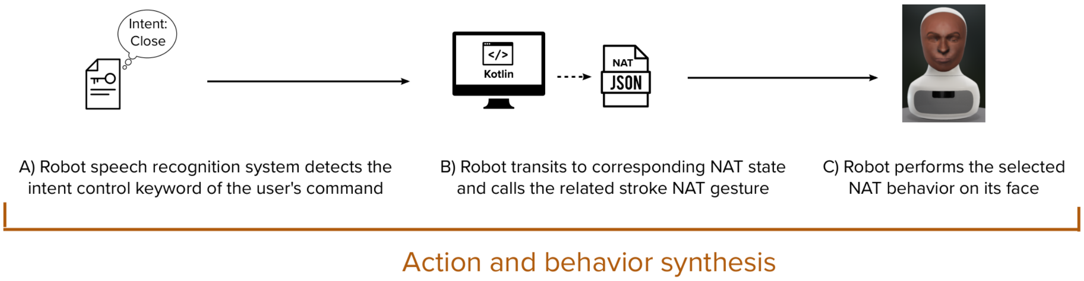

Intent: Close
Kotlin
NAT
JSON
A) Robot speech recognition system detects the intent control keyword of the user's command
B) Robot transits to corresponding NAT state and calls the related stroke NAT gesture
C) Robot performs the selected NAT behavior on its face
Action and behavior synthesis

Figure 6.11. RTL action and behavior synthesis by the robot in response to the user's intent.

close to, and within the interaction space of, the robot and identify their attention (whether they are looking at the robot). FurhatOS also supports estimating head pose and attention. Thus, we programmed the robot in such a way that once it detects a CL within its interaction space attending to the robot, the robot will convey its attention back to the user by looking directly into their eyes. If the CL changes position slightly, but stays within the interaction space of the robot, the robot will rotate its head and orient its face such that it can "attend to" the CL.

Animated Momentary behaviors): In addition to paying attention to the CL, the robot will respond to their utterances not only with speech, but also with momentary, animated behaviors (including small eye movements, head nods, and micro-expressions) in sync with speech, just like a person would. We added this capability to further enhance the quality of HRI.

It is essential the robot moves its eyes during the interaction, because keeping its eyes fixed makes it seem unnatural. Humans tend to get distracted, look away, and slightly shift their gaze when interacting with another person. We encoded these momentary eye movements into the FSM logic, enabling the robot to execute the "LookAway" built-in behavior or shift its eye gaze slightly at different intervals.

Head nod is another important momentary behavior. The MMC framework of the robot enables it to acknowledge the CL's spoken words during an interaction, often responding with affirmatives like "Okay" or "Sure". We enabled the robot to nod its head instead of saying anything, or nod in sync with the affirmatives.

Micro-expressions are slight facial movements around the nose, eyes, and mouth (such as mild twitching, slightly closing of eyelids, etc.), which further increase the quality of human-robot interaction. Within our MMC framework, we provided the setting to enable or disable the micro-expressions rendered from time to time.

Gaze Movement Synthesis

While co-designing the clinical scenario, we deduced that adding gaze tracking as a component of the system would enable ROSE to track the movements of the CL's hands with its eyes. The development of the hand tracking script using data collected from a wearable MetaMotionS sensor allowed us to measure the motion and position of the CL's hand in real time. Using this data, we automated the robot's eye movements in order to follow the CL's hand movements. We developed the MMC framework in a way to convert data into precise eye movement values for the robot (e.g., up, down, left, and right), and subsequently transmit this information to the robot via internet connectivity. Once the CL finishes assessing the eye tracking, the robot's gaze returns to the rest position while attending to the user. (See Figure 6.9: right).

6.6 Evaluation

In order to assess the robot's performance and identify areas for improvement, we carried out pilot studies with engineers and clinician participants. In this study, participants engaged in an autonomous interaction with the robot, pretending that the robot was a real patient, whom they assessed using neurological assessment tests. Through this study, we assessed the perceived usability of the system and its impact on user experience, while collecting feedback on how to improve the system.

Engaging clinicians in this pilot study helped to promote a user-centered approach, and allowed for their valuable input, validation, and expertise in shaping the development of ROSE. Including robotics engineers allowed us to receive expert technical feedback, rigorous evaluation, and collaborative problem-solving. Our study was approved by our institution's Institutional

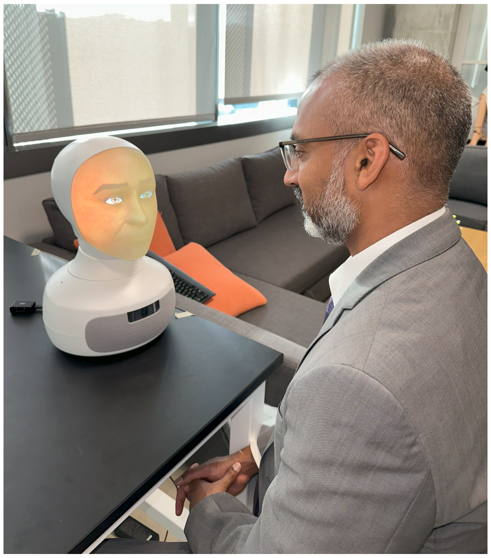


Figure 6.12. A clinician interacting with ROSE.

Review Board, under protocol number 191488X.

6.6.1 Participants

We evaluated the prototype and determined how to improve it in two ways. First, we engaged in extensive discussions with three neurologists through the development of the robot to gather valuable insights and feedback, but we did not conduct a formal interview with them due to constraints on their schedules. Among them, one individual identified as female, while the remaining two identified as male.

Second, we conducted pilot studies with ROSE with eight participants: one physician with a specialty in neurocritical care (N-MD), one clinical professor of medicine who works in medical education simulation (S-MD), and six graduate engineering students (GE). Among them, five participants self-identified as female, while the remaining two participants self-identified as male. The age range of the participants spanned from 18 to 54 years old. Both clinicians had completed a US-based medical education and on average had 19 years of face-to-face interaction with patients. Those who had a robotics-related educational background had no experience in patient-physician interactions, nor in stroke diagnoses.

6.6.2 Measures

We used a mixed methods approach in our data collection and analysis. To understand basic system usability, we administered the System Usability Scale (SUS), a well-validated measure of perceived usability [56] (See Appendix B.4). We were particularly interested to understand how clinicians without a technology background perceived the robot.

Qualitative measures included post-study interviews and researcher observations of participants' interactions with the robot during the study (See Appendix B.2 and Appendix B.3). We assessed the robot's humanlikeness (appearance and behavior), similarity to a patient, clinical scenario realism, multi-modal communication, autonomous interaction, impact on user experience, and usefulness. Finally, we collected their suggestions on future features they would like implemented. We also asked participants about their experience and impression of how the robot appeared, interacted, and conversed. For participants with a clinical background, we also asked questions to evaluate the degree of realism and similarity of ROSE's face to someone experiencing acute stroke (See Appendix B.2). We recorded and transcribed all interviews.

Two researchers employed a grounded theory [78] approach, and individually coded the audio transcripts to find emerging themes through an inductive coding process. We then compared codes and identified three overarching themes among the participants, specifically: increased humanlikeness (Section 6.7.2), means to enable realistic communication capabilities (Section 6.7.3), and ways to improve user experience (Section 6.7.4).

6.6.3 Robotic Patient Simulator Creation

Clinician-defined Controls for Noah

We co-designed and deployed a clinical training scenario with our neurologist collaborators. In this scenario, a male patient, named Noah, who has a fair complexion and is in his 50s, is admitted to the hospital with left-side upper and lower paralysis. The patient had visual signs of one-sided droopiness and presented asymmetric blinking cues. Noah can perform facial

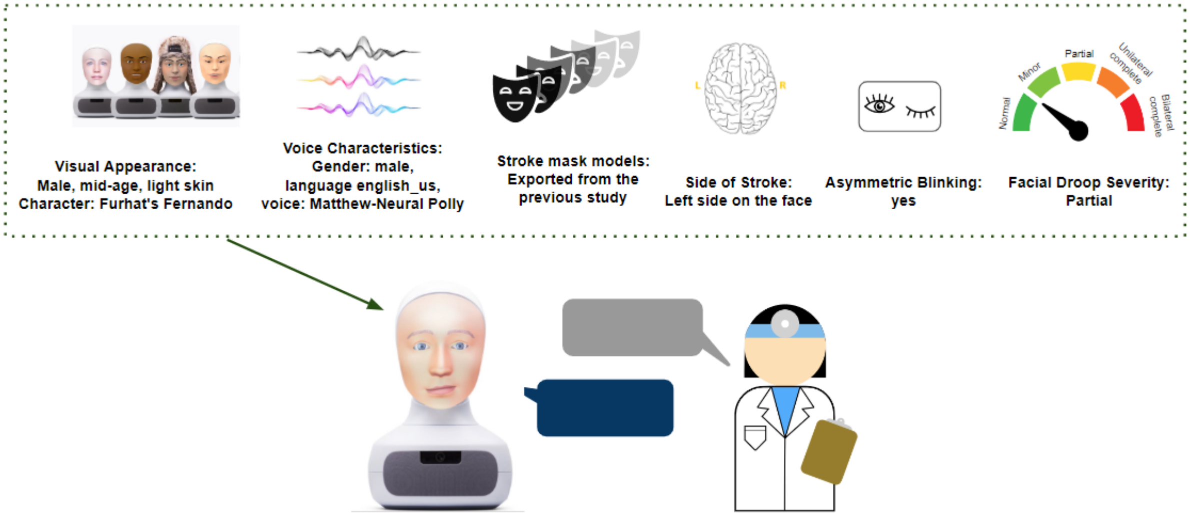


Figure 6.13. The initiated values of the clinical-defined controls for Noah.

actions commonly used to assess acute stroke, including raising eyebrows, closing eyes tight, blinking, puffing cheeks, smiling, and relaxing the face. In terms of personality, Noah's character is polite and neutral. The goal is for CLs to conduct a neurological exam, and diagnose the signs of stroke. (See Figure 6.13).

Control Word Dictionary

We designed a comprehensive control word dictionary for the speech recognition system on the robot to respond to (See Table 6.2). To create this list, we first analyzed ROSE's intended tasks and context to generate the initial list of keywords. We then collaborated with domain experts to understand clinical terminology, and adjusted the keywords to match the language clinicians use during real-world stroke assessment.

We continuously iterated and improved upon the dictionary over time by conducting user research to shape the list based on user preferences and feedback. After several iterations, we finalized the dictionary through testing and evaluation with clinical experts. Thus, this dictionary list aligns with the operational context of the robot, user needs, and domain expertise. Users can include any of these control words in their utterances to interact with the robot.

Table 6.2. This is a dictionary list of control words and their corresponding keywords that the speech recognition system on the robot will understand. Using the *greeting* keywords will start the conversation with the robot. Using the *ready* keywords can initiate an assessment session with the robot. Using the *NAT* keywords will ask the robot to perform any NATs. Users can ask the robot to perform any tasks in any order they like. Using the *repeat* keywords will enable the robot to repeat a task as often as the user likes. Using the *encouragement* keywords inspires the robot after correctly performing a task. Using the *conclusion* keywords can end the conversation at any time the user wants.

	Category	**Main Keywords**
	Initiate	"Hello", "Hey", "Hi"
	Name	"Name", "Your name", "Address", "Address you", "Call you"
	Greetings	"How is it going?", "How are you", "How are things going?", "How is everything"
	Ready	"Ready", "Start", "Set", "Begin", "Test", "NAT", "Assessment", "Perform", "Exam"
NATs	Rest	"Relax","Rest", "Look straight"
	Raise	"Raise", "Lift", "Eyebrow"
	Puff	"Fill", "Cheek", "Air", "Puff", "Enlarge", "Inflate"
	Close	"Close", "Tight", "Tight Tight Tight"
	Blink	"Blink", "Blinking"
	Smile	"Grin", "Mouth", "Teeth", "Smile"
	Repeat	"Repeat" ,"Again", "One more time", "Once more"
	Encouragement	"Good job", "Well done", "Nice", "Perfect", "Alright", "Good", "Great"
	Conclusion	"End", "Finish", "Goodbye", "Bye"

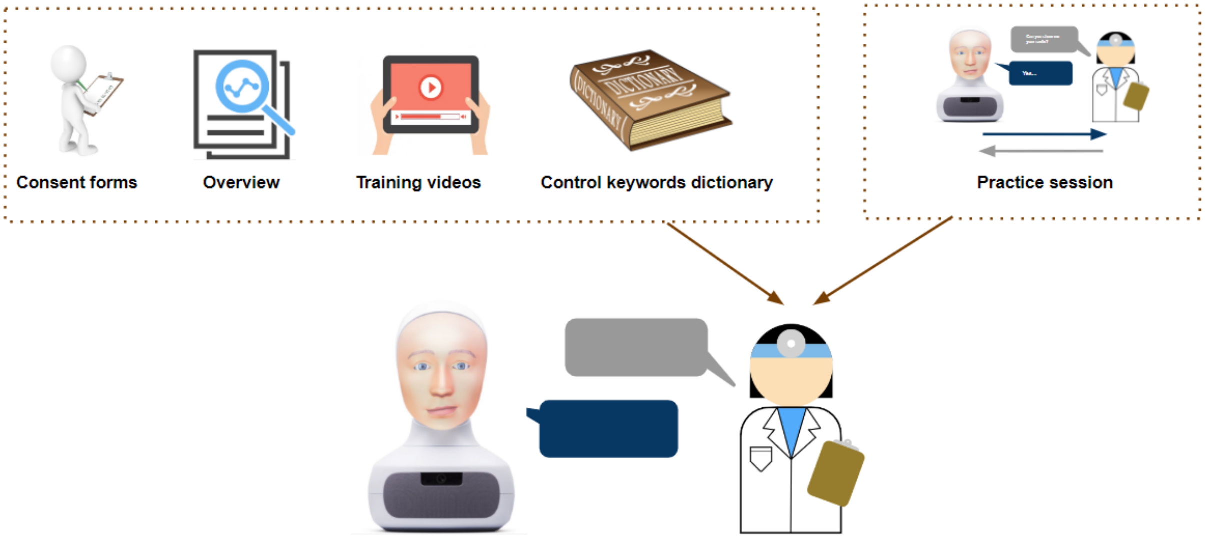


Figure 6.14. A demonstration of information about the procedure of the study design shared by clinical learners prior to the testing session.

6.6.4 Procedure

After giving written consent, we asked participants to sit in front of the robot, introduced them to ROSE, and gave an overview of the study (See Figure 6.12. To give participants without medical backgrounds some context, we showed them a video of someone with a stroke to demonstrate their appearance. We then showed them a video of a clinician assessing a patient with stroke in the real world to show them what a neurological exam looks like.

As all participants did not have experience with this robot, we showed them a demonstrative video of ROSE simulating a real-world clinical scenario, and a user interacting with the robot. Next, we shared the control word dictionary with them, explained what actions the robot can perform, and how they can use the keywords to interact with the robot. We also informed them that if the robot doesn't hear them well, it will ask them to repeat themselves louder, and if the robot doesn't see them in its proximity, it will ask them to come closer or sit in front of the robot.

We then began the practice session. We asked participants to interact with ROSE by asking it to perform commonly used NATs, similar to what they watched in the third video.

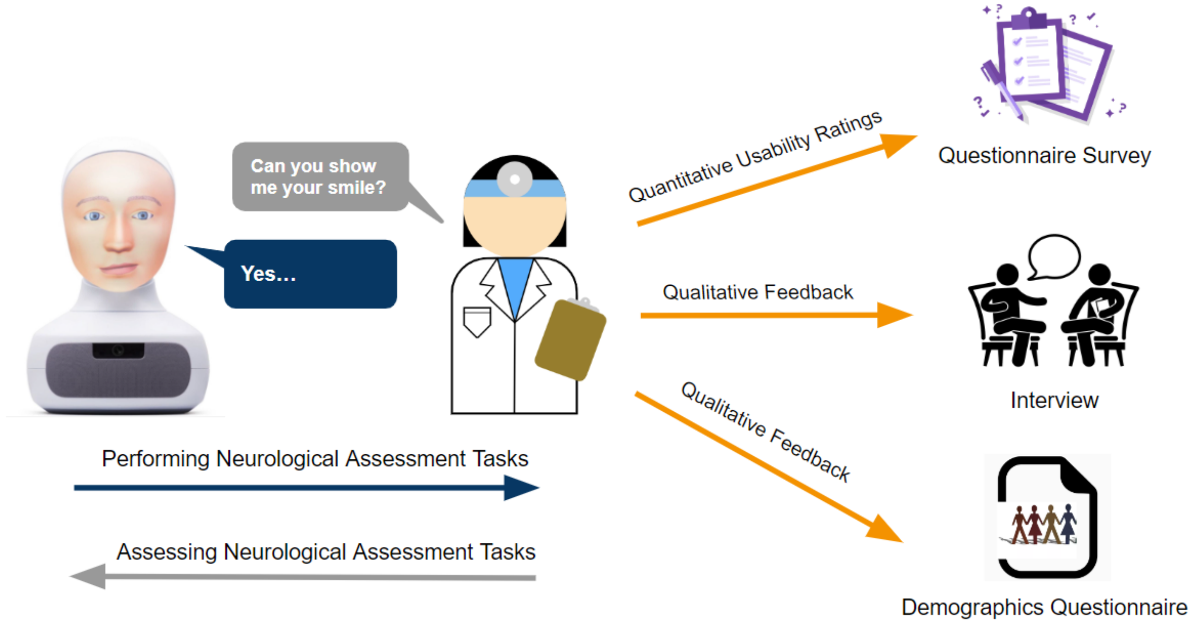

Can you show me your smile?
Yes...
Quantitative Usability Ratings
Questionnaire Survey
Qualitative Feedback
Interview
Qualitative Feedback
Demographics Questionnaire
Performing Neurological Assessment Tasks
Assessing Neurological Assessment Tasks

Figure 6.15. A demonstration of the robot and learners interaction, followed by engaging the participants in short quantitative and qualitative surveys and interviews to evaluate the system.

We informed them that they could refer to the control word dictionary during the interaction as needed to socialize with ROSE, ask it to perform NAT tasks, provide encouragement, or end the interaction. Participants could ask questions before and after the practice session. (See Figure 6.14 for an overview).

Next, we started the testing session (Figure 6.15). We asked participants to interact with the robot just like before and end the interaction at any time. To conclude the testing session, we administered the SUS survey, conducted an open semi-structured interview to receive qualitative feedback on our system, and administered the SUS survey and demographic questions.

6.7 Results

Due to technical difficulties with the network connection, one of the GEs was unable to fully complete the study. Thus, this section reports the results based on the data gathered from the seven participants who successfully completed the study. (See Table 6.3 for an overview).

6.7.1 Usability

On SUS, which represents a composite measure of the overall usability of the system [68], participants scored ROSE an average of 84.64 (SD = 9.83), which is above average compared to other systems [56]. During the interview, participants described using the system as “easy to use”, “easy to interact with”, and “straightforward”. One participant said that the “robot is smooth and polite”. Although participants expressed a few frustrations with the speech recognition component of the system (as discussed in Section 6.7.3), overall, all participants were able to successfully operate ROSE for its intended use.

Participants also provided suggestions for supporting the robot’s interaction and operation in the HET domain, discussed below. Several of these suggestions have already been integrated into the system, while others are potential directions for future research (See Section 6.8).

6.7.2 Increased Humanlikeness

The humanlikeness of Noah is critical for exposing CLs to real-world patient-physician interactions when performing stroke diagnoses. Participants expressed that the realism of ROSE allowed them to feel as if they were truly interacting with another person. They also indicated that the level of realism exhibited by ROSE contributed to a less disruptive experience during their interactions, setting it apart from other social counterparts. The natural look of the robot impressed several participants despite it being a projection. One participant stated how it was “human enough, but not creepy”. Additionally, they indicated that giving the robot the name “Noah” further facilitated humanlike interactions since it would not be referred to as just “the robot”. Based on participants’ feedback and comments, we can classify their preferences for humanlikeness of the robot into three categories: visual humanlikeness, behavioral humanlikeness, and patient similarity.

Visual Humanlikeness

Participants explicitly expressed positive feedback about the visual humanlikeness of ROSE in terms of its face and gender. For instance, one participant stated that its "face, for the most part, looks more humanlike than most other robots". They indicated that the natural-looking contours, color, and shading of the skin are contributing factors to the humanlikeness of Noah.

As S-MD stated, "the contours of the face combined with the eyes moving and blinking gave it a much more realistic appearance". In order to improve the realism of the system, some participants suggested adding hair, glasses, and eyebrows can help, while others mentioned how adding hair might make the robot creepy.

One participant expressed that the shape and 3D volume of the robot's face looked humanlike. However, another participant found it unnatural that the face surface is primarily a physical 3D surface, while the facial expressions and movements were a 2D virtual projection on the 3D surface. This may cause users to not be able to precisely see volume when Noah performs some of the NATs, such as when it enlarges its cheeks. They also expressed that a lack of texture makes the face less humanlike, and suggested employing more facial textures, including wrinkles, in future work. On that note, N-MD stated that usually stroke patients are older than what Noah looks like, so the face should depict more wrinkles and accentuated nasolabial flattening.

It is critical for CLs to practice performing a stroke diagnosis on ROSE with diverse backgrounds, to broaden their understanding of stroke presentations across different patient demographics. As one participant suggested, "it's important to see people of different ethnicities, and different races, and different facial features". On the other hand, another participant stated how they were not able to identify Noah's gender, and suggested that the gender-neutrality of the simulator might potentially help remove gender bias during future patient-physician interactions.

Because ROSE does not have a body, some participants deemed it as strange and "a little bit creepy". From a clinical perspective, adding a body to ROSE can make it more realistic as it enables the robot to incorporate a depiction of weakness in the affected side of the body in addition to the face.

Behavioral Humanlikeness

Robot behavior encompasses a wide range of actions and responses exhibited by robots, which contributes significantly to its perceived humanlikeness. With respect to behavioral humanlikeness, a participant mentioned how the robot's voice was humanlike. Several participants also expressed how ROSE had realistic non-verbal backchannel cues enabling the robot to be more humanlike. One participant talked about Noah's gaze, and how Noah's eyes were "moving around a little bit and. . . blinking, using its eyebrows, and those kinds of behaviors feel very natural. It feels like someone that's just sitting and waiting to be spoken to". Additionally, Noah's ability to move its head, such as nodding, impressed a participant, which they believed adds a level of realism. Noah's eye-tracking, rapid and subtle eye movements, blinking actions, and head movements are all baseline behaviors that helped establish a more humanlike system.

However, one participant expressed how the jump from Noah's rest baseline state to an NAT state can be sudden, and can seem animated and unnatural. This suggests that making the transition from Noah's relaxed state to one of the NAT states smoother could increase the realism of the system.

In addition, enabling the robot to both understand and synthesize filler words that occur in natural language—such as "um", "uh", "like", "wait", or "hang on"—can make the system more realistic and naturalistic. Some participants expressed their frustrations about Noah's inability to understand the filler words as linguistic pauses or hesitation markers, and instead of acknowledging these cues, Noah responded by indicating a lack of understanding of the user's input. The robot either does not need to respond to filler words, or, as a participant suggested, to create a category of filler words the robot could take into account to make the interaction more natural.

Excluding the robot's direct responses to filler words, and treating these words as inconsequential data to the conversation can enhance the robot's authenticity. Alternatively, as one participant suggested, ROSE could establish a specific category of filler words to recognize and consider, enabling a more natural interaction and accommodating users' linguistic preferences.

Patient Similarity

It is critical for ROSE to accurately imitate characteristics of PwS to support CLs with their stroke diagnosis skills. This means that the robot should be able to depict various clinical scenarios and characteristics similar to real patients (See Section 6.5.2), as well as realistically perform NATs required to assess signs of stroke (See Section 6.5.2). For example, S-MD expressed how the robot's facial asymmetry was clear even just by looking at its face at rest, evidently suggesting a neurological problem. N-MD stated that the NATs performed by ROSE were very similar to actual human expressions of PwS: "the fact that he's lifting up both sides of his eyebrows is great because, in an acute stroke, you wouldn't expect that as opposed to Bell's Palsy, where you would expect the weak side [of the eyebrow] to not lift up. That's fantastic!".

N-MD also stated that Noah's asymmetric blinking alongside the facial droop made the robot's face look more realistic and similar to a patient with a stroke. How facial droop caused by a stroke looks can vary significantly, ranging from subtle changes to extremely pronounced effects such as a downward angle of the mouth and a partially open eyelid.

Other participants expressed additional similarities between Noah's facial movements and PwS. For example, a participant thought Noah's resting face, puffing cheeks, and raising eyebrows were realistic. Another participant thought the "close" NAT was realistic. (See Figure 6.16 for an overview of NATs performed by Noah).

Some participants found the tasks performed by the robot to be easily identifiable, such as ROSE blinking versus closing their eyes. However, others found the robot's asymmetric facial characteristics associated with stroke less distinct or prominent.

Although ROSE is able to exhibit recognizable facial characteristics associated with a stroke, on average, N-MD stated that PwSs would have more facial droop than what Noah displays. This suggests that the facial droopiness for Noah, in its partial severity setup mode, is more subtle than the usual facial droop seen in PwS on average.

In addition to FP, some PwS may experience speech difficulties such as slurred speech as a result of muscle weakness (dysarthria), or difficulty sequencing and coordination of the muscle and structures (apraxia) [224]. While the control word dictionary enabled participants to effectively test for facial strength and characteristics associated with stroke, the list does not include terms to test for speech deficits, such as dysarthria and apraxia. Both N-MD and S-MD suggested adding speech slurring for the robot system to display as a symptom to help improve its similarity to PwS. N-MD also stated that if the robot had more "facial weakness, it might seem less realistic if the speech was still clear". Additionally, the physician participants stated that they would appreciate if the robot were able to perform a wider set of NATs, such as having the robot visually follow a finger making an "H" shape with its eyes.

Still, it is important to note that achieving high fidelity humanlikeness in robots is still ongoing research, and the level of human likeness that is ultimately possible for a robot may vary due to technological advancements and design choices.

6.7.3 Realistic Communication Capabilities

To ensure the fidelity of stroke diagnosis simulations resembling real-world patient-physician interactions, it is important to establish two-way conversational interactions between users and the robot. Thus, based on participant feedback, how the robot engaged in speech recognition and speech synthesis significantly impacted the realism of conversation with the robot. As per participant recommendations, addressing and refining these will foster more authentic and effective dialogue between users and the robotic system.

Speech Recognition

The robot's speech recognition encompasses its ability to accurately comprehend and interpret user speech inputs. Although some participants expressed a few frustrations with the speech recognition component of the system, the robot consistently demonstrated its ability

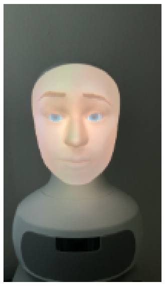
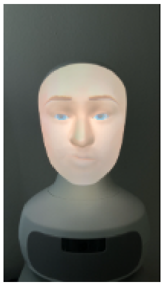
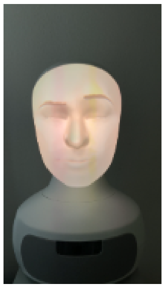
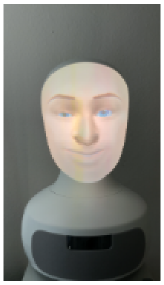
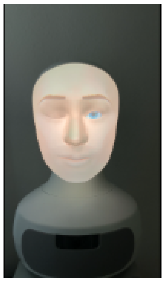
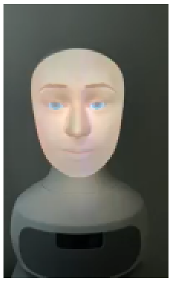
 **RAISE** Robot raises its eyebrows while trying to induce the wrinkling of the forehead muscles.	 **PUFF** Robot makes its cheeks larger and rounder by filling them with air while having lips tightly closed.
 **CLOSE** The robot closes its eyes tightly while wrinkling its inner eyebrows.	 **SMILE** Robot makes a big smile with its lips wide open.
 **BLINK** Robot closes its eyes tightly and then opens them	 **REST** Robot relaxes its face with its teeth in a normal bite and with a neutral facial expression.

Figure 6.16. An overview of NATs performed by Noah.

to appropriately understand and respond to user requests in a timely manner. For instance, N-MD stated how they "didn't have to alter [their] language very much so that part of it was very realistic. . . So [they] used very similar vocabulary to what [they] would do in a clinical setting. [They] didn't have to cater [their] language to this simulation event". Another participant expressed that Noah's speech recognition is fairly good, in that it was able to carry out user requests based on control words.

However, other participants noted occasional challenges with the robot's auditory perception, where Noah encountered difficulties in accurately capturing user speech. For example, because the robot responds based on the provided keywords, if it mishears a word on the list, it results in the robot saying that it doesn't understand the user. As a result, there were instances where Noah did not fully execute some commands due to slight misinterpretations of the user prompts.

Additionally, some participants stated that although the control word dictionary allowed them to connect with the robot, incorporating a broader range of user inputs could enhance the robot's ability to understand and fulfill the user's intentions. For instance, a participant stated that it is difficult to find another way to express a command using only the keywords if the robot didn't understand them initially, so they needed to repeat their commands multiple times. Several participants suggested increasing the possible number of control words and constructing a more extensive vocabulary for the robot, such as having more keyword synonyms.

N-MD provided a few examples of what a clinician would say to a patient: "put air in your cheeks", "close your eyes tight tight tight", "I'm going to examine you now". These words and phrases are potential add-ons to the current list of keywords.

Another participant recommendation would be for the robot to "not exactly recognize the exact word, but infer the meaning from the user" instead of expanding the range of keywords the robot has to understand.

Adding more keywords to the dictionary and enabling the robot to understand the meaning of the user input rather than keyword-matching are two potential methods to increase users' chances of getting a robot response without having to repeat themselves, and they can also make the interaction feel more natural for the user.

Speech Synthesis

The robot's speech synthesis pertains to the robot's capability to generate humanlike and contextually appropriate responses, playing an important role in the system's conversational realism A few participants expressed how it is less of a two-way conversation, and more of the user telling the robot to perform tasks. Although the robot's dialogue was based on real clinical scenarios, some GEs perceived the conversations to be more unidirectional, with the user primarily instructing Noah to perform tasks rather than engaging in a genuine two-way conversation. N-MD noted that while examining a human PwS, clinicians usually introduce themselves to patients and ask them to introduce themselves, so it would be more realistic for Noah to initiate conversation as well. Participants suggested the responses incorporate more niceties, such as “How are you today?”, “I’m good, thanks for asking”, “I’m fine”, “Nice to meet you”, fostering a more balanced and interactive exchange between the user and the robot.

Moreover, participants provided constructive feedback regarding certain aspects of the speech synthesis, such as Noah’s taking too long to respond. One participant expressed frustration about this, especially when they used words outside the predetermined list, leading to the robot not comprehending the input.

To address this, participants suggested implementing shorter prompt responses for the robot when it encounters a lack of understanding or difficulty hearing the user. Other participants recommended streamlining the robot’s dialogue by removing certain words from the robot’s prompt responses to enhance the overall flow and efficiency of the conversation. For instance, one RE’s suggestion was for the robot to eliminate mentioning the phrase “OK” before it performs a task. One participant also noted that shorter dialogue from the robot could help prevent interruptions and enable participants to maintain their train of thought without the risk of forgetting their intended message while waiting for the robot to finish speaking.

Furthermore, participants provided recommendations regarding Noah’s response time. Although several participants appreciated that the robot responded in a timely manner to deliver a natural interaction, there were suggestions from other participants to further enhance the robot’s responsiveness. They recommended that the robot could benefit from even quicker response times when engaging with the user, as the delay between the user command and robot action may “reduce the effectiveness of the interaction”.

Another participant suggested that Noah should ask “Could you repeat that?” more quickly after a user says a command, instead of the robot pausing for 3-4 seconds before asking. Another participant recommended having the user’s facial expressions serve as a cue for the robot to perform turn-taking in order to increase response time. These adjustment suggestions would contribute to a smoother and more efficient conversation, enhancing the overall user experience.

6.7.4 Ways to Improve User Experience

The user experience of the system plays a key role in fostering motivation among CLs to consistently use the system for skill improvement and knowledge enhancement in the area of stroke diagnosis. If ROSE provides a positive and engaging user experience, it can encourage CL usage. Participants identified several potential uses for the system, along with three key aspects of user experience that warrant improvement, discussed below.

Perceived Usefulness

All of our participants with clinical backgrounds found our system very useful for training CLs on a variety of skills. N-MD stated that having CLs train on the robot “would help them identify a facial droop better”, suggesting that training with the robot can improve their overall skills in recognizing the facial characteristics associated with stroke.

S-MD indicated that they “can imagine practicing the sequence of a neuro exam or taking a history [with the robot]”.

S-MD also highlighted the potential power of “an electronic system” with knowledge of completed and incomplete tasks. They expressed the idea that real time assessment and tracking of CLs’ questions and robots’ neuro exams could be valuable. They suggested the system could provide feedback, allowing users to repeat any missed steps. For instance, if a CL performed steps 1-6 using the robot but skipped steps 7-8, the system would identify the omission, notify the user of the missed steps, and allow them to retry.

Additionally, S-MD emphasized the potential for setting automated performance metrics within the system, such as tracking the time taken to complete each assessment. By leveraging this feature, CLs can enhance their skills and proficiency through targeted improvements and iterative learning facilitated by the robot’s automated performance evaluation.

Engagement

User engagement within the system holds significant benefits for CLs, and has the potential to positively contribute to both sustained system usage and a positive perception of the robot's usefulness. One of the GEs expressed how they enjoyed the robot being able to recognize and respond to them fairly accurately, and how it was able to still communicate with them if it did not hear them the first time.

In line with this, S-MD recommended having the robot display expressions in response to receiving a greeting or statement of encouragement. For instance, the robot could respond to phrases such as "well done" or "good job" with a smile, promoting a more interactive experience. S-MD also suggested that the robot could exhibit expressions and responses consistent with frustration. For example, if the robot encounters difficulty with a task and needs to redirect the provider, it could express its limitations by saying,"that's the most I can do."

Ease of Use

An intuitive and accessible system can support CLs to effectively use the system, providing opportunities for learning and skill improvement. A participant explicitly expressed how Noah was "overall easy to interact with and use". In particular, participants found the control word dictionary beneficial for enabling them to know how to command Noah to perform desired actions.

However, some participants found it frustrating to have to repeat commands to Noah. To address this, one GE participant suggested incorporating the ability for users to define the duration or repetition of the robot's actions in order to simplify use of the system. For example, instead of asking Noah to "smile" three separate times, the user could directly ask the robot to "smile three times" in a single command.

Additionally, another participant mentioned the robot's lack of feedback or guidance when the system did not work as expected, forcing the user to repeat the same commands repeatedly. The participant suggested that the robot should provide more informative feedback to explain an issue. For example, when Noah states it cannot see the user, it can add “you’re too close” or “you’re too far away’.

Enjoyment and Sociability

Participants expressed how the robot was fun to interact with, and appreciated its smoothness and politeness, which contributed to an enjoyable experience.

However, a few participants found certain aspects of the system, such as the sociability of the robot, to be less enjoyable. In particular, a few participants experienced frustration with voice recognition, such as having to repeat commands multiple times. Additionally, an RE participant expressed how their natural response to what the robot says does not always align with the vocabulary the robot understands. For example, they stated that if “the robot says something like “Can you say that louder?”, I feel it’s natural to say something like “Yeah, for sure. I can say that louder” or something like that, which the robot can’t understand, so it asks you again, so that can be a little frustrating”. Currently, the robot would not be able to handle this sort of deviation in the dialogue, and could lead to additional user frustration.

These features might demotivate CLs to use the system as their association with the system becomes negative, frustrating, and unenjoyable. Addressing them could significantly enhance the user experience, maintain motivation, and improve the overall satisfaction of CLs.

6.8 Discussion

6.8.1 Contributions to HRI

Introducing a new application space for HRI. Our research brings an approach to support HET by providing an diverse, inclusive, and customizable platform that improves in realism and accessibility. This can provide the foundations for bridging the gap for CLs between theoretical knowledge and real-world clinical practice, promoting a richer understanding of clinical conditions and fostering critical thinking, ultimately improving stroke patient care. To our knowledge, ROSE is the first patient-data-driven, interactive clinical training tool accessible

Table 6.3. Design and development recommendations for supporting the robot's interaction and operation in clinical education.

Increased Humanlikeness	Visual Humanlikeness	The robot's visual appearance should include natural-looking contours, shading, and color of the skin for a more humanlike appearance. To support the appearance of older patients with stroke, the robot should be able to display more wrinkles and accentuated nasolabial flattening. Adding a body can also increase its realism since it can display weakness in the affected side of the body as a stroke symptom.
	Behavioral Humanlikeness	The robot's behaviors should include verbal cues (e.g., language understanding and speech synthesis), as well as non-verbal backchannel cues (e.g., gaze tracking, blinking, head movements). The transition from the robot's relaxed state to an NAT should be smoother, and a category of filler words should also be integrated into the robot's vocabulary (e.g., "um", "uh").
	Patient Similarity	The robot exhibits NATs and facial movements similar to actual stroke patients, such as recognizable facial asymmetry (e.g., asymmetric blinking and facial drooping). The robot should be able to display facial drooping in various levels of severity, perform a wider set of NATs (e.g., eye tracking), and display speech slurring as a symptom to increase the patient similarity of the system.
Realistic Conversation	Speech Recognition	The robot can consistently understand and respond to user requests in a timely manner and allow clinicians to utilize language they would use in a clinical setting. Adding more keywords to the robot's vocabulary and enabling the robot to infer meaning from the user can increase the chance of getting a robot response without having to repeat themselves if the robot mishears or misinterprets the user.
	Speech Synthesis	Some participants suggested that the robot should engage in a more balanced and interactive interaction with the user through the initiation of dialogue, such as the incorporation of an introduction and more niceties in conversation (e.g., "How are you today?", "I'm good, thanks for asking"). The robot should also have a shorter prompt duration and response time to facilitate a more streamlined conversation.
Ways to Improve User Experience	Perceived Usefulness	The robot is useful for training CLs to improve their ability to diagnose facial characteristics associated with a stroke (e.g., facial droop). The robot can also provide support for real-time assessments, tracking of CL's questions, and patient exams. It has the potential to provide automated performance evaluation/feedback for CLs based on clinician-set performance metrics to enhance CL's skills and proficiency.

(Continued on next page)

Table 6.3. Design and development recommendations for supporting the robot's interaction and operation in clinical education (Continued).

	Engagement	The robot is able to recognize and respond to user prompts accurately, which can help users stay motivated to continue utilizing the robot. It was suggested that the robot display expressions in response to greetings, encouragements, and frustrations. For example, the robot could respond to phrases like "well done" with a smile or say "that's the most I can do" to express its limitations and struggles with a task.
	Ease of Use	The control keywords dictionary enables users to command Noah to perform desired actions, but some participants found it frustrating having to repeat commands to Noah. The system could incorporate the ability for users to define the duration or repetition of the robot's actions to simplify the system's usage. The robot should also provide more informative feedback and guidance so the user does not have to repeat commands over and over again.
	Enjoyability	The robot is fun to interact with, and users enjoyed its smoothness and politeness. Users expressed frustration with voice recognition (e.g., repeating commands) and the limitations of the robot's speech recognition, as users' natural responses did not align with the robot's vocabulary.

to clinicians to practice diagnosis and treatment of neurological disorders such as stroke.

Our system demonstrates the feasibility of using expressive robots capable of automatically interacting with humans in the application space of HET, which opens up the door for other researchers to explore robots in this domain. Our work provides a framework for researchers to explore HRI in new experiential learning settings (e.g., build RPS systems to enable CLs to avoid forming biased impressions) and broader domains (e.g., explore methods for designing social robots to enhance people's perception of individuals with FP).

Data-generated models of real patients. Our research advances the state of the art of medical simulation. We leverage our frameworks to create data-generated models of real patients, and overlay them on the robot to enable it to depict realistic verbal and non-verbal cues. Thus, this study demonstrates the visibility of putting more complex, realistic behaviors on robotic platforms in this HET context, affording robots with richer communication modalities. This enables researchers to explore how people perceive and interact with more humanlike robots, and investigate user perceptions and experiences during these interactions. This work will enable the robotics community to leverage our approach to customize a robot's autonomous behaviors, adapt to end-user needs, and promote effective HRI within their own application domains.

Supporting reproducibility for the HRI and HET communities. This work aims to inspire researchers to develop expressive interactive robots that are more accessible, engaging, and capable of supporting individuals within the HRI community. As an artifact to support reproducibility for the HRI and HET communities, the software discussed in this chapter will be made available as open-source.

6.8.2 Collaborative User-centered Design

Given the importance of user-centered design, our work highlights our collaboration with multiple stakeholders, including neurologists, clinical educators, and engineers. Our main objective was to demonstrate the significance of customizing the robot during deployment to address its specific needs and backgrounds. Particularly, we aimed to support varying needs within a system like ROSE for CLs who may possess varying clinical or social knowledge and skill levels due to differences in their level of medical education/experience. By emphasizing user-centered design throughout our development process, we sought to ensure that the system effectively accommodates the unique requirements of CLs and facilitates their learning experiences. This is critical to creating effective and successful robotic systems.

6.8.3 Making RPS systems more humanlike

Traditional simulation-based training typically lacks the features that support humanlike communication, limiting engagement and immersion in learning. For example, using tactile-based methods, such as partial-task trainers or clay models, provides valuable hands-on practice by simulating anatomical structures or procedures, but does not permit interactive conversation. Similarly, simulation-based tools, such as existing RPS systems, provide clinically-relevant learning scenarios, but often rely on the operator's voice rather than enabling autonomous, expressive communication between the user and robot.

These existing tools fail to provide the end-users the opportunity for multimodal humanlike conversation, hindering their ability to fully engage and participate in meaningful interactions with the robot. This deficiency can adversely affect the immersion and effectiveness of the learning experience. Preserving these robot features can help users stay engaged and motivated to continue using the robot.

Our work addresses this gap by introducing a humanized approach to RPS systems. ROSE enables multimodal communication between CLs and robots, which allows real time, two-way communication autonomously. Our robot also incorporates thoughtful and respectful interactions with users, maintaining a polite demeanor throughout the conversation, and allows users to conclude the interaction at their discretion. These features foster a more realistic and engaging learning environment, facilitating deeper involvement and offering the potential to improve the overall effectiveness of HET. By providing a platform that supports natural and interactive communication, our system offers a promising avenue for advancing the capabilities of clinical learning.

ROSE exhibits autonomous behavior in two key aspects: First, our learning tool simulates assessment tasks automatically, eliminating the need for operators to describe clinical conditions verbally or manually adjust visual parameters. Second, our robot autonomously interacts with the end user (e.g., CLs) and automatically performs clinical assessment tasks within a predefined clinical scenario. This approach stands in contrast to other existing RPS systems that usually require an operator to manually control the robot's actions based on the end-user's responses.

6.8.4 Continuous adaption to CL's evolving needs and preferences

As CLs progress in their course of clinical education, their clinical and social skills may evolve. Furthermore, each CL may have different learning styles (visual, auditory, kinesthetic, reading/writing, social, solitary, analytical, or logical learner) or feedback preferences (e.g., positive, negative, binary, or explicit feedback). CEs may have different educational goals for RPS systems based on each individual CL and their capabilities. These attributes necessitate the robot's system to dynamically adapt to the situation over time dynamically. Advancing the robot to adapt to the CLs' needs provides personalized training experiences that promote inclusivity and accommodate different learning styles.

6.8.5 Intersectionality Considerations

Utilizing humanlike robots in HET presents a host of benefits, particularly with respect to the intersectionality considerations within healthcare, which could encompass cultural, age, complexion, ethnicity, gender identity, socioeconomic, and academic background considerations, to name a few.

First, from the perspective of representing diverse patients, these robots are designed to offer a high degree of realism, enabling them to convey a range of diverse backgrounds. This enables CLs to experience and navigate the intricacies of human differences and similarities in communication and healthcare practices. By engaging with robots that accurately reflect the norms of different patient groups, learners can develop the essential skills of competence, empathy, and adaptability. These skills are crucial for healthcare professionals who aspire to provide patient-centered care in multicultural contexts while promoting positive patient outcomes.

Second, from the perspective of diverse end-users (e.g., CLs), humanlike RPSwS systems should be able to inclusively accommodate CLs with various cultural backgrounds and different types of professionals. These systems' interactive and responsive communication features give CLs an excellent opportunity to comprehend the intricacies of cross-cultural interactions in different stages of their training. This improves their communication and clinical skills in a controlled environment, fostering cultural sensitivity and awareness.

While humanlikeness in RPwS helps address some cultural challenges, this approach may not fully address inequities in medicine. For instance, cultural differences can also manifest in non-verbal cues and body language, which robots may not fully (or faithfully) replicate. Additionally, cultural beliefs, values, and expectations regarding healthcare can greatly vary, and a robot's humanlike appearance may not fully convey or accommodate these complex cultural factors. While our advancements in creating diverse and inclusive HET systems are commendable, it is important to recognize that inherent biases and discrimination can still persist within these technologies [202]. Thus, it is important for HET to concurrently improve efforts in anti-racist education, fostering equality in healthcare.

6.8.6 Ethical considerations

There are several benefits to designing humanlike social robots as a tool for HET. First, using humanlike robots allows for a more immersive and realistic training experience resembling human patients. It allows healthcare professionals to refine their clinical skills in a controlled and safe environment before interacting with real patients.

Second, using humanlike robots ensures a standardized and consistent training platform. Unlike human patients, robots can reliably reproduce specific symptoms, responses, and behaviors, providing CLs with consistent training opportunities. This standardization offers all learners an equal opportunity for repeated practice of the same scenarios. Additionally, using this tool eliminates the potential risks and ethical concerns associated with practicing on real patients, such as accidental harm, misdiagnosis, or privacy invasion.

Third, using humanlike robots in research allows for a more accurate assessment and evaluation of CLs' performance. By replicating human gestures, expressions, gaze, and responses, these robots provide learners with realistic feedback, allowing them to gauge their abilities and identify areas of improvement. This feedback is crucial for creating a safe and effective learning environment, promoting continuous professional development, and, ultimately, improving patient care.

Humanlike robots, however, may pose ethical challenges, which should be acknowledged and addressed.

Collaborative policy development. While humanlikeness in robot design offers numerous benefits in healthcare education and research, it is imperative to remain vigilant and address ethical challenges to ensure users' well-being, privacy, and autonomy. Particularly in HET, it's important to engage in collaborative policy development with key stakeholders to protect patients' and learners' rights, privacy, and ethics. Comprehensive legal policies and frameworks should be developed in consultation with relevant parties, including clinicians, educators, and patients, to ensure equitable use of these robots. Additionally, policies must address concerns around data privacy and security, informed consent, and ethical considerations.

Unintended perpetuation of bias. While utilizing humanlike social robots in HET offers numerous benefits, there is a potential for them to perpetuate biases if not carefully managed. Existing common clinical learning modalities tend to lack access to diverse representations of patients (both standardized and simulated), which many have argued represents a key limitation for CLs [85, 247]. Biased programming or robot behavior modeling can negatively impact learners' perceptions and interactions with diverse patient populations, potentially reinforcing existing stereotypes or disparities. Additionally, biased data used to develop facial masks or train the robots can result in unequal learning experiences.

Ongoing monitoring, evaluation, and feedback from diverse stakeholders are essential to identify and address potential biases or shortcomings in using humanlike robots in healthcare contexts. Institutions must ensure that the robots do not reinforce existing biases or stereotypes and that learners receive equal and unbiased training experiences. By doing so, institutions can promote cultural sensitivity, inclusivity, and equality in healthcare education and training, leading to better patient care and outcomes. Furthermore, HET programs must recognize that humanlike robots may not entirely capture cultural differences, as non-verbal cues and attitudes can significantly affect cross-cultural interactions.

Informed Consent in HET-focused HRI. To guarantee the responsible use of robots, healthcare organizations must provide proper instructions and guidelines that clearly explain the goals and objectives of each task, and comprehensively demonstrate proper ways of operating the system. Incorporating informed consent would further accessibility, especially for people without prior knowledge of robotics. This can maximize their utilization of the system. Informed consent can also help foster transparency and ensure participants comprehend the nature of their involvement.

Informed consent is a crucial element of ensuring the responsible use of robots in HET, and can be a requirement in different stages of the robot system design and deployment. Initially, informed consent is required from human patients to gather real-world data, aiding in the design of the humanlike RPS. This is to protect patient privacy while ensuring their understanding of the implications of how their data might be used to affect the development of RPS systems.

It is also important to obtain informed consent from the CL, who will use and interact with the RPS. This guarantees that they are comprehensively informed of a robot's capabilities, behavior, and information-gathering procedures before beginning any training, thereby enhancing an informed engagement and positive interaction based on mutual understanding.

Despite the significance of informed consent in HET, it still retains potential risks. Without full clarity about the robot's components, behavior, or how it collects data, there could be unforeseen outcomes. For example, collecting sensitive information without informing the subjects is an ethical and privacy issue. Moreover, in case of discrepancies between disclosed robot functions and its actual capabilities, CLs might perform based on misleading assumptions. The lack of clarity could adversely affect CLs, leading them to experience confusion, mistrust, and frustration, which in turn, could hamper their concentration and involvement in the learning process.

Autonomy in HET-focused HRI. Supporting personal autonomy is an essential aspect of HET, particularly in providing CLs with the freedom to choose when they want to engage or disengage from robot-enabled activities throughout their course of training. This feature endorses a fundamental concept of autonomy in medical ethics, by allowing people to make their own choices without being forced into anything. This can also provide a learning environment that facilitates personal decision-making and comfort.

When it comes to HRI, autonomy has pitfalls too. If learners feel pressured by an academic atmosphere into using the robotic systems when they do not want to or feel uncomfortable with them, it could lead to tension and anxiety. Moreover, regularity requirements may force a sense of obligation, violating CLs' autonomy. This could potentially cause CLs to become irritated and disengaged. This could also prevent CLs from voicing any concerns they may have, thus, preventing potential upgrades to the system. Ultimately, such experiences could compromise pursuing the primary goals of the learning process.

Power Dynamics in HET-focused HRI. The humanlike characteristics of robots in HET can have a profound impact on human users due to their potential to unintentionally create power imbalances between CLs and robots or between CLs and CEs. Thus, it is essential to consider these dynamics during the design and programming phases of humanlike robots to ensure that the training environment promotes a favorable learning experience that conforms to prevailing ethical standards.

Regarding power imbalances between CLs and robots, designing a robot in HET to manifest authority could inadvertently control CLs, thereby sabotaging their capacity for independent critical thinking. This could encourage an environment in which CLs become excessively dependent on the robot's guidance, adversely affecting their independent decision-making and clinical judgment skills. This dark side could impede the goal of developing competent and confident professionals.

In terms of power imbalances between CLs and CEs, learners may display an overly-submissive position when using a robotic simulator due to its perceived realism and authority. On the other hand, educators may unintentionally reinforce this dynamic by taking an authoritative role, causing learners to lose their autonomy and critical thinking skills.

To address these concerns, adopting a user-centered design approach that encourages CLs in active participation and reinforces independent thinking while encouraging dialogue between learners and instructors is essential.

Uncanny valley effects. There is the potential for uncanny valley effects, where close but not quite humanlike robots can elicit feelings of unease or discomfort in users. This may raise concerns about emotional well-being and the potential for psychological distress toward CLs.

Privacy and data security. Other ethical considerations include privacy and data security, since humanlike robots may collect and process sensitive personal information of CLs. Ensuring this data's protection and proper handling is crucial to maintain privacy and upholding ethical standards.

6.8.7 Limitations and Future Work

Despite our diligent efforts to operationalize the complex constructs of humanlikeness, autonomous behavior, healthcare education delivery, and extensive piloting, we were compelled to make decisions based on the practicality dictated by methodological and platform limitations. In the future, it would be interesting to explore different clinical situations, use various robotic platforms, and test different interaction methods to broaden and deepen our research.

Due to resource and time constraints, we have not been able to fully explore how our robot interacts with clinical learners. However, we acknowledge the importance of this interaction in evaluating the effectiveness and applicability of our system in a real-world HET setting. Thus, our future research plan prioritizes a comprehensive study of the robot's performance with CLs. We believe this study will provide valuable insights about the robot and improve its usefulness in HET.

To further improve the effectiveness of humanlike RPSwS systems in HES, we will address some key limitations in future work. First, we will enhance the system's conversational framework by expanding the repertoire of control words. This will improve keyword matching, minimize the need for users to repeat themselves, and thereby increase the system's utility.

Second, we intend to enhance the visual appearance and behavior of the robot to address participants' comments about increasing humanlikeness and patient similarity. For this purpose, we will modify the design of the robot to include increased wrinkles and accentuated nasolabial flattening, particularly on the unaffected side of the face, to more accurately replicate patient symptoms. We will also work closely with neurologists to adjust the degree of droopiness and asymmetric blinking, two key symptoms for stroke assessment. Furthermore, we will provide customization options for the robot's gender, skin tone, age, voice, and facial characteristics and adjust the severity of symptoms like droopiness and asymmetrical blinking. These enhancements will enable healthcare learners to better understand stroke presentations in diverse patient populations and improve the realism of the simulation scenarios.

Third, we will enhance the naturalistic and engaging perception of human-robot interaction by implementing several measures. For one, we will assign individual names to robots to establish distinct identities and character-based resemblances, fostering relatability. Moreover, we will incorporate self-introduction functionality to allow robots to introduce themselves and establish personal connections. Furthermore, we will shorten prompt responses and implement polite niceties to create a friendly and efficient conversational atmosphere. These changes promote two-way interactive dialogue, and prioritize creating a more personalized and inclusive learning environment.

Fourth, we have plans to include a feature that allows for customization of the robot's speech proficiency, specifically focusing on modifying the level of Dysarthria or Slurred Speech. This is crucial because slurred speech can manifest as a symptom of stroke, and CLs need to practice evaluating the severity of dysarthria. To achieve this, learners will use a rating scale that ranges from none to mild-moderate, severe, to mute.

Fifth, we will focus on enhancing the system's capability to provide comprehensive and personalized feedback to learners regarding their performance. This entails highlighting areas requiring improvement and offering tailored feedback to enhance learners' skills. By incorporating these advanced feedback mechanisms, we aim to optimize the learning experience and empower learners to achieve higher proficiency levels in their clinical skills.

6.9 Chapter Summary

This chapter presented ROSE, an interactive social robot for clinical training tool, that enables CLs to automatically interact and practice their stroke diagnosis skills on the robot.

6.10 Acknowledgments

I thank Dr. Jamie Labuzetta for helping me design and evaluate the clinical scenarios for ROSE. I thank Drs. Sean Evans, Leena Kansal, and Dominic Ferrey for their assistance with prototype evaluation and study design. I thank Suhas Pai for helping me implement and report the MMC framework development during the prototype design phase. I also thank Rainee Pei for her assistant with the data analysis and results reporting. This chapter contains material from "ROSE: An Interactive Social Robot for Medical Education" by M. Pourebadi, S. Pai, R. Pei, and L. D. Riek, which will be submitted to the ACM/IEEE International Conference on Human-Robot Interaction (HRI)[191]. I was the primary investigator and author of this paper.

Chapter 7

Conclusion

This chapter discusses the main contributions of my research to the fields of HRI, robotics, FG, and health technology. I then briefly deliberate on prospects for future research avenues which follow from my work, and broader open questions that will need to be addressed to make interactive, expressive humanlike robots accessible in real-world settings. At the end, I will conclude this work with closing remarks.

7.1 Contributions

7.1.1 Presented the potential of humanlike robotic patient simulators in the context of HET.

Simulation-based training methods such as RPS consistently demonstrate benefits in comprehension, confidence, efficiency, and enthusiasm for learning. These improvements directly contribute to clinicians' ability to support patient safety, reduce preventable harm, minimize risks, enhance the quality of care, and lower healthcare costs [207, 241]. However, it is crucial to identify gaps and opportunities in existing learning modalities in order to recognize the potential of humanlike RPS in the context of HET.

First, in order to identify the gaps and opportunities for using robots in the HET domain, I outlined the root causes of preventable patient harm in healthcare departments, and the application domain of HET as one of the best defenses to reduce the incidence of patient harm (See Chapter 2).

Second, I examined the benefits and challenges that accompany common learning modalities of HET, including virtual and robotic patient simulators. Finally, I presented major gaps in introducing the use of humanlike learning modalities as a possible solution. This work establishes the foundations of designing and deploying expressive RPS systems capable of portraying clinical scenarios, and provides a framework for researchers to support this process.

7.1.2 Created new virtual and physical faces for robots in HET.

In dynamic, real-world environments such as simulation-based HET, social robots need to have realistic humanlike behaviors to accurately exhibit verbal and non-verbal cues. However, many existing RPS systems have limited to no capabilities for humanlike expressiveness in their faces, which may impede emotional engagement, empathy, and social presence, leading CLs to experience reduced motivation, interest, and retention of training content [185]. Thus, in my work, I investigated the effect of expressive mechanical and rendered faces in RPS design and presented my work on building new expressive faces.

First, I discussed the role of humanlike behaviors in social interactions and outlined the benefits and key challenges of enabling virtual and robotic embodiments to depict verbal and non-verbal behaviors (See Chapter 3).

Second, I explored techniques for detecting, modeling, and synthesizing humanlike FE in virtual and physical robotic faces. I explored common methods to build FEA systems for detecting and tracking humanlike expressions to contextualize facial expression analysis in HRI. I described different facial action modeling techniques in order to build FAM systems, while considering various information processing and knowledge modeling methods. I examined technical approaches for building FSA systems to synthesize dynamic FEs on virtual agents and physical robots for a variety of applications.

Finally, I discussed my research on virtual and robot patient simulator faces, enhanced with the capacity to exhibit nuanced verbal and non-verbal behaviors and cues, while displaying diverse appearances and backgrounds. This included my work on creating new embodiments for RPSs, focused on designing physical and virtual faces, while enhancing their expressivity, diversification, and control modalities [196, 195].

This work stands as a potential transformative instrument in HET, opening new frontiers in developing expressive RPS systems. Moreover, this work provides valuable insights to researchers by examining methods for detecting, modeling, and synthesizing FEs, with potential applications in enhancing social interactions, and clinical education.

7.1.3 Built an end-to-end AMS control framework, and developed a general FPM framework to generate accurate representations of patient-like FEs on RPS faces.

Every year, millions of individuals experience conditions such as stroke, Parkinson's disease, Moebius syndrome, and Bell's palsy, leading to facial paralysis and A-FEs. People's misperceptions and biased impressions can make it challenging for them to interact socially with and understand the emotions of people with A-FEs These misperceptions in clinical settings can adversely impact the quality of care provided to FP patients. This highlights the need for new training tools to enhance clinicians' interaction skills and improve care for individuals with facial paralysis. The lack of exploration in using FP patient simulators highlights the need for researchers to develop training tools that consider individuals with FP, aiming to enhance clinician skills in avoiding biased impressions, improve clinical communication, and deliver better care for this population.

To address this problem, we developed two frameworks to make RPS systems able to depict an accurate representation of A-FE on their faces based on real patient's facial characteristics, and demonstrated the complementary relation between these two frameworks (See Chapter 4). This work had two main goals. First, it aimed to enable people to easily synthesize human facial movements on any robotic and virtual faces in real time. Second, it aimed to understand how robots can accurately and realistically depict asymmetric facial expressions.

In this work, we first designed and developed the AMS control framework that integrates three systems presented in Chapter 3 to robustly, easily, and automatically transfer humanlike expressions from a subject's face onto a range of physical or virtual robotic faces. We modularized the AMS framework into three components (FEA, FAM, and FSA), allowing for a more organized and encapsulated structure of the framework. The FEA component enables the AMS framework to more robustly detect and track FE movements in real time. The FAM component overlays a computational representation of a clinical condition onto the tracked FE movements. Finally, the FSA component automatically synthesizes facial movements onto the face of robotic and virtual simulators with different ages, races, and genders, and animates their facial components.

Second, we developed the FPM framework to provide a platform to automatically generate accurate computational models derived from facial characteristics of people with FP, and is constructible in real time. We then integrated these two frameworks by overlaying pre-built FPMs on the facial model of the AMS framework to recreate A-FEs on RPS faces. The AMS framework uses the results of the FPM framework and enables the system to robustly recognize the facial movements of a human operator, mask the generated model on tracked movements, and automatically synthesize the generated models of FEs across a range of RPS embodiments, thereby animating their facial components. Furthermore, to put these frameworks into practice, we presented an FPM framework personalized to model characteristics of a particular type of FP, BP, and synthesized it on virtual RPS. Finally, we reported the results from an expert-based user study, highlighting that experts perceived our expressive virtual patient simulator to be realistic and comparable to humans with BP.

The presented frameworks create a solution that accurately portrays patient-like FEs on RPS faces, situated within a HET context. This research opens new avenues of exploration in Healthcare Robotics, and may trigger a new round of relevant technological innovations by creating the next generation of patient simulator robot technology. Furthermore, the results of this work will enable roboticists to discover platform-independent methods to control the FEs of both robots and virtual agents, and yield new modalities for interaction.

7.1.4 Introduced RPSwS for modeling and synthesizing acute stroke.

As FEs and their intensities exhibit significant inter-individual variability and dynamicity [275], the development of a universal RPS system capable of accurately modeling and presenting neurological impairments across diverse cultural and demographic spectrums poses a daunting challenge [195]. Doing this would require access to a large corpora of data from PwS representing a diverse set of characteristics associated with acute neurological disorders and facial impairments, which is both time and labor-intensive. Moreover, it can be challenging to analyze the data collected from a restricted cohort of PwS and extrapolate it to construct stroke models that depict a more extensive population of PwS. Developing such universal models to design versatile RPS systems with synthesized faces encompassing a diverse patient group is essential. This diverse assortment includes but is not restricted to individuals of varied ages, genders, and ethnicities suffering from various health afflictions [275].

To address these challenges, in Chapter 5, I introduced RPSwS: a new expressive training tool capable of realistically depicting non-verbal, asymmetric FP cues representing acute stroke. The core objective of this work is to architect a comprehensive, holistic system derived from our general FPM and AMS frameworks in order to create data-generated models of people, and overlay them on the robot to enable it to depict realistic verbal and non-verbal cues. This enables RPSs to accurately and effectively depict stroke symptoms, thereby advancing the landscape of HET for stroke diagnosis and treatment.

First, I introduced Stroke FPM: a new framework for generating statistical modeling approaches representing the facial characteristics of stroke. This consists of two parts: 1) a machine learning method to accurately identify and automatically track facial landmarks of PwS that are crucial for analyzing A-FE movements, and 2) a statistical modeling approach to use tracked facial point values to automatically represent the most visually significant features representing stroke in each facial region. I then ran the Stroke FPM on a newly collected dataset of PwS admitted to an urban medical center who have experienced acute ischemic stroke resulting in neurological findings, leading to the generation of computational models representing stroke's characteristics.

Second, I created the RPSwS capable of automatically displaying FP by developing an end-to-end Stroke AMS framework, which applies the generated models using the Stroke FPM onto the face of an RPS system [197, 195].

Third, I reported the results from a perceptual study with seven clinicians to investigate the efficacy of my system for modeling and synthesizing stroke. This study explored the visual differences in realism and similarity between the synthesized stroke robot faces and those of stroke patients. The results of these measurements facilitated the identification of salient attributes for each facial region that can make the stroke robot look more realistic and similar to PwS. The received feedback endorsed the robot's utility, concurrently providing valuable recommendations for potential enhancements.

This work has cross-disciplinary impacts in clinical education, health informatics, FG, robotics, and HRI, as it pioneers a new method for comprehensive stroke-associated FP representation, facilitates realistic FP simulations on various RPS systems, and provides insights for asymmetric FE analysis, social robot design, and understanding the effects of facial asymmetry on social interactions. To my knowledge, the Stroke FPM framework pioneered the implementation of a statistical modeling methodology that can capture stroke-associated facial anomalies, extending across the upper and lower facial regions. This framework provides a systematic and objective way to analyze and interpret facial movement patterns associated with stroke, contributing to extracting various presentations of the medical condition.

Moreover, the RPSwS represents the first utilization of the comprehensive AMS framework to render stroke characteristics on diverse RPS systems, thereby facilitating the production of highly authentic FP simulations. Our RPSwS system can depict realistic facial expressions similar to PwS. By producing a set of 75 facial models representing stroke in various facial regions, my research yielded insights into the best representations of stroke in each facial region based on professional expertise, enhancing the precision and reliability of stroke representations for different facial regions. By employing this mechanism, robotics researchers can create many empirically derived FP facial representations for robots, specifically in HET contexts, and and perform studies on people's perception of and response to facial asymmetry [197]. Furthemore, the results from an expert-based user study with physicians, highlights their strong interest in robots replicating facial characteristics associated with stroke, supporting the indicative of the system's potential as a healthcare education tool.

Our work can also help researchers in the FG community to explore new methods for asymmetric facial expression analysis, modeling, and synthesis. Moreover, our study enables HRI researchers to explore methods for designing social robots to enhance people's perception of individuals with FP and understand the effects of facial asymmetry on social interactions.

7.1.5 Designed and developed ROSE: an interactive social robot for medical education.

Current RPS systems often fail to provide CLs with interactive platforms for humanlike engagement, limiting the development of social interaction skills and confidence in diagnosis, while lacking communication modalities that allow effective interactions and training effectiveness. Additionally, these systems may not replicate automatic manipulation of clinical scenarios, limiting skill acquisition and knowledge transfer, potentially leading to missed opportunities for timely interventions and prevention of harm. Nonetheless, designing an interactive robotic tool to address gaps in engagement and usability may introduce challenges related to advanced technology reliance, complex interfaces, and potential frustration among clinicians, impacting their acceptance and learning outcomes.

To address these challenges, we spearheaded the design and deployment of ROSE: an immersive clinical training tool employing an interactive, socially adept RPSwS to enhance the learning experience for CLs.

The core objective of this work was to understand how to enable a robot with social intelligence to autonomously exhibit realistic behaviors and effectively engage in interactions within real clinical education settings.

We first presented the collaborative user-centered design requirements for building ROSE in the context of HET (See Chapter 6).

Second, we introduced the MMC framework to automatically simulate clinical scenarios, and enable autonomous interaction and engagement on ROSE. We designed this in close collaboration with neurologists and CEs to co-design an interactive robot that could depict neurological impairments. For this purpose, We developed an interactive, expressive RPSwS as the platform of the robot, with customizable expressions, appearance, and characters, to enable the robot to be more humanlike and realistically portray PwS. We enabled the robot to simultaneously show stroke characteristics in all facial parts, leading to a more realistic robot appearance and behavior. We then incorporated a real-world dynamic scenario into our system to suit the users' needs better, and developed a keyword-marching control mechanism to enable the robot to interact with CLs automatically.

Third, we presented ROSE to create a realistic and immersive environment for practicing diagnosis and treatment skills. We report the results from pilot studies and interviews with clinical educators and robotics engineers to investigate the efficacy of our tool for depicting dynamic clinical scenarios, and report the results revealing how they envision using ROSE for stroke diagnosis. To our knowledge, ROSE is the first of its kind, representing an exciting new area of research.

This work has implications for HET, as well as the broader healthcare and HRI communities. To our knowledge, ROSE is the first patient-data-driven, interactive clinical training tool accessible to clinicians to practice the diagnosis and treatment of stroke. The presented collaborative user-centered design requirements offer insights for HRI researchers and developers of interactive, expressive robots, encouraging them to adapt these design requirements to their own applications. Furthemore, the MMC framework enabling the robotics community to leverage our approach to customize a robot's autonomous behavior, adapt to user needs, and promote effective HRI within their own application domains. Our work lays the foundation for extending the accessibility of educational interventions to the healthcare domain, enabling humanlike social robots to support HET through an automated interactive, expressive robot. ROSE enables researchers to create a realistic and immersive environment for practicing stroke diagnosis skills, offering opportunities for repeated clinical practice, and promising avenues for advancing the capabilities of clinical learning. Moreover, our work provides a framework for researchers to explore HRI in new experiential learning settings (e.g., build RPS systems to enable CLs to avoid forming biased impressions) and broader domains (e.g., explore methods for designing social robots to enhance people's perception of individuals with FP).

7.2 Future Work

In the future, it would be interesting to explore different clinical situations, use various robotic platforms, enable our robot to automatically collect and understand physiological signals from the monitoring systems, and test different interaction methods to broaden and deepen our research. To further improve the effectiveness of humanlike RPSwS systems in HET, we will address some key limitations in future work.

Due to resource and time constraints, we have not been able to fully explore how our RPSwS interacts with CLs. However, we acknowledge the importance of this interaction in evaluating the effectiveness and applicability of our system in a real-world HET setting. Thus, our future research plan prioritizes a comprehensive study of the robot's performance with CLs. We believe this study will provide valuable insights about the robot and improve its usefulness in HET.

7.2.1 Situating Social Robots within HET

The findings of this study provide a foundation for future research aiming to enhance the visual appearance and behavior of humanlike robots in HET. Future research could explore broader directions, such as increasing humanlikeness and patient similarity, customizing the robot's characteristics to represent diverse patient populations, and collaborating with domain experts to enhance the reproduction of specific symptoms. These enhancements will contribute to improving realism in simulation scenarios, and further enhance the understanding of various medical presentations in diverse patient populations.

The efforts of this work will open doors to future research in HRI in order to explore various avenues to enhance naturalistic and engaging perceptions of RPS systems. This could include incorporating self-introduction functionality to allow robots to verbally introduce themselves, provide a patient history, and establish personal connections. Furthermore, future research could include designing more effective prompt responses for the robot and implementing polite niceties, in order to enhance friendliness, and create an efficient two-way conversational atmosphere.

7.2.2 Creating automatic, adaptive feedback systems in HET for individualized learning experiences

Moving forward, researchers can explore advancing the system's capacity to provide CLs with personalized feedback regarding their performance, with the goal of enhancing their clinical skills. This research direction could involve leveraging data analytics and machine learning algorithms to develop adaptive feedback models that interpret a CLs' interactions, determine patterns, and suggest recommendations tailored to their needs, thereby creating a more personalized learning experience. These systems adjust feedback format, content, and intensity based on a CL's learning preferences and styles.

Moreover, researchers can investigate mechanisms for improving the process of feedback delivery in order to make it more effective. This could involve using methods such as visualization, augmented reality, or interactive interfaces to increase engagement with CLs. This research direction has the potential to revolutionize HET by effectively delivering personalized feedback which enable CLs to optimize their learning experience and achieve higher levels of proficiency.

7.2.3 Advancing personalized anti-bias education for promoting equitable clinical practices

Developing personalized anti-bias educational interventions using RPS systems with diverse identities (such as complexion and gender) opens up broad avenues for future research. Potential directions could involve exploring how receiving training sessions using such educational systems can impact racial and gender biases among CLs and promote cultural sensitivity throughout their training. This research direction could pave the road to mitigate biases among healthcare professionals while treating real patients, enabling them to deliver more compassionate patient care. Based on the research I have conducted so far, the presence of racial bias, racial disparities, and gender bias in stroke care highlights the need for further exploration and action.

Research has identified racial disparities in diagnosing and treating acute stroke. For example, non-Hispanic Black and Hispanic patients experience longer Emergency Department waiting times compared to non-Hispanic white patients, potentially leading to delays in treatment and sub-optimal stroke care [108]. Another study illustrated that African American patients are less likely to undergo brain imaging within the recommended timeframe compared to white patients [141]. Further investigation into the underlying causes of racial disparities is necessary to identify strategies for mitigating bias, and help inform interventions and policies, ensuring timely and fair clinical care for all patients.

Future research on addressing racial disparities could include developing an intervention for CLs to enhance cultural competence among healthcare providers, increase diversity among healthcare professionals, and implement strategies to address preconceptions and implicit biases that lead to unequal access to medical services for individuals with darker complexions or diverse racial identities. Receiving training on the proposed intervention may positively influence the quality of care, health outcomes, and treatment provided by clinicians to individuals from different racial or ethnic backgrounds.

Research suggests that gender presentation casn impact disparities in stroke presentation, diagnoses, and outcomes, with women being more likely to present nonspecific symptoms, including hiccups, nausea, chest pain, fatigue, shortness of breath, and a racing heartbeat [75]. Research on cerebrovascular diagnosis has identified that women have a 25% higher chance of misdiagnosis than men [75].

Future research can focus on the investigation of understanding and addressing gender bias in stroke care. This may involve determining the ways gender presentation may impact clinically-related symptom recognition, diagnosis, and treatment. Creating training programs and guidelines can prepare clinicians to accurately diagnose conditions, thus, enhancing the ability to address gender disparities in healthcare.

Finally, future research directions could explore ethical and intersectionality considerations. In healthcare education and research, humanlikeness in robot design offers numerous ethical benefits, such as immersive training experiences, accurate assessment, and effective feedback. Moreover, it also supports various benefits of intersectionality, such as enabling the accurate replication of diverse patient backgrounds and inclusive accommodation of CLs with diverse backgrounds.. However, It is crucial to remain alert to the challenges associated with ethics and the acknowledgment of intersectionality in this domain. This includes identifying factors contributing to challenges related to the unintended perpetuation of bias, privacy, data security, informed consent, autonomy, and power dynamics through clear guidelines, protocols, collaborative policy development, and comprehensive legal frameworks.

It is also important to study existing training methods designed to reduce ethical and cultural biases in healthcare, in order to understand their effectiveness and potential areas for improvement. Further comprehensive training modules and cultural sensitivity workshops are necessary to address complex cultural factors in healthcare settings [143, 127]. By supporting empowerment and cultural sensitivity, researchers can create personalized and inclusive learning environments in the realm of HRI, paving the way for exploring broader directions in this field.

7.2.4 Customizing speech proficiency in RPS systems for enhanced patient similarity

Based on the research progress made thus far, future research can focus on customizing speech proficiency in RPS systems by developing a flexible and adjustable robot that allows researchers to alter dysarthria or slurred speech. Creating an adjustable rating scale that covers a diverse range of dysarthria severity would enable CEs to set the degree of slurred speech according to their particular needs and preferences. Such tools could lead to the development of comprehensive educational tools, ultimately revolutionizing the assessment and treatment of dysarthric impairments in HET settings.

7.2.5 Advancing user input understanding methods for improved utility and satisfaction

Based on this work, researchers can explore further advancements in the MMC framework to improve the system's utility and user satisfaction. This exploration can involve expanding the repertoire of control keywords. Incorporating this approach would reduce the need for users to repeat themselves, enhance the robot's ability to understand and fulfill user intentions, and create a more natural and efficient interaction experience.

Additionally, many roboticists are working on improving multimodal communication, by enabling the robot to understand and infer the user's input beyond speech recognition [229]. Future research direction can focus on interring meaning and intent through facial cues, body language analysis, natural language processing, contextual understanding, and sensor integration. This can enhance the robot's ability to more deeply interpret users' emotional states, behavioral intentions, and engagement levels, ultimately leading to more effective and nuanced communication capabilities.

7.3 Open Questions

7.3.1 What are the implications of automating learners' performance assessment?

Performance assessment automation is essential for assessment of all CLs, thereby removing subjective biases that CEs may unconsciously introduce and enabling a fair and unbiased evaluation of their performance. However, it is also necessary to acknowledge and address potential limitations in designing automated assessment systems, considering that humans develop the algorithms used within these systems and, thus, convey their own biases in its design [203, 105]. Moreover, the algorithms are usually trained on datasets that may inherently reflect various biases or disparities; thereby, careful consideration should be given to the training data used for developing such systems.

Moreover, automation has the potential to reduce the burden on CEs, enabling them to focus on controlling the simulator and facilitating the training process. Manually assessing CLs' performance can be time-consuming and resource-intensive, especially when evaluating multiple CLs simultaneously. Thus, RPS systems which provide some autonomous support to the assessment have the potential to free up CE time. Automating the assessment process may help CEs with their workload thereby, enabling them to dedicate more time and energy to provide better guidance and support to CLs.

7.3.2 How does the use of interactive, expressive RPS systems affect the CEs' workload and task disruption in HET?

CEs who act as both operators of the RPS system and instructors in HET scenarios already have many tasks to manage, such as operating the RPS system, guiding CLs, assessing their performance, providing feedback, and ensuring a smooth training session. This raises questions about the effects of implementing shared control modalities for the robot on CEs and their workload.

Some open questions in this area may include: How does the dual role of CEs as robot operators and instructors impact their workload in training clinical scenarios and assessing CLs' performance? To what extent can increased autonomy of the robot effectively reduce the workload of CEs in healthcare education and training? What are the implications of implementing shared control modalities with respect to time efficiency, workload management, and task disruption for CEs? This opens up new opportunities for research to explore the impact of using these robots on CEs' workload management, and the overall effectiveness of HET. It also opens up an avenue for exploring the benefits and challenges of developing shared control modalities for such robots in HET.

7.3.3 How can RPS training and feedback functionalities continually adapt to best support learners' evolving needs and preferences?

Continually adapting educational interventions is essential to keep CLs interested, and consistently engaged in the interaction with the robot, ultimately leading to successful educational outcomes. In the context of HET, the importance of continuously being immersed in the learning necessitates adaptive RPS behaviors that evolve with the changing needs and preferences of CLs over long-term interactions.

As CLs progress in their course of clinical education, their clinical and social skills may evolve. Furthermore, CLs will have different learning styles (e.g., visual, auditory, kinesthetic, etc.), as preferences for different types of training modalities. CEs may have different educational goals for RPS systems based on each individual CL and their capabilities. As a result, an RPS system must be able to adapt to the situation over time dynamically.

Addressing these challenges to enable educational interventions to adapt appropriately raises several questions. For example, what are the key considerations in understanding CLs' specific support requirements, and establishing appropriate behaviors for diverse situations? How can we effectively tailor and diversify educational materials over time to meet the personalized CLs' needs and their learning styles, as well as CEs' educational goals? How can we create diverse educational materials and feedback to adapt to the evolving clinical skills of CLs throughout their training? How can we adjust the assessment feedback over time to align with CLs' preferred feedback methods, optimizing its effectiveness?

7.3.4 How can we effectively collect and use comprehensive datasets to develop realistic representations of clinical conditions for the robot, considering the diverse range of patients?

Developing RPS systems to effectively simulate the diverse range of PwS requires a comprehensive dataset, representing diverse, accurate, and realistic representations of stroke-related symptoms. To address this challenge, we collected a dataset from a subset of the PwS population (See Chapter 5). However, our dataset has potential gaps in terms of diversity of background, age, and other factors. Moreover, our dataset primarily represents NATs related to assessing visual facial cues and speech-related symptoms, lacking representation of other critical stroke-related symptoms such as body weakness.

These gaps raise some open questions in this context. For example, what size and variety of datasets of PwS is required to create a comprehensive dataset of stroke patients, enabling researchers to generate accurate simulations of stroke for the robot? What specific tasks should PwS perform while recording their videos to ensure the data set comprehensively covers the symptoms and their variations? Finally, what are the similarities and differences between facial symptoms in non-stroke facial palsy patients and PwS, and to what degree can the data from non-stroke facial palsy be used in order to create comprehensive robotic stroke models?

7.3.5 What are the impacts and dynamics of trust in RPS HRI?

The dynamics of trust between the robot, CLs, CEs, and developers present new avenues for exploring the impact of trust on various aspects in the context of using interactive, expressive RPS systems for HET. However, there is a gap in understanding of how different levels of responsiveness and speed can affect how reliable and competent the RPS systems appear, and what expressive capabilities could include in the robot design to build trust between humans and robots. Moreover, there are limited comprehensive studies to compare the effect on patient trust when CLs are trained with virtual simulations versus real patients.

These challenges may raise some open questions in this area: How do the responsiveness and speed of the expressive RPS system influence its acceptance and trustworthiness among CLs and CEs? How does the training experience on robotics simulators compared to real patient interactions influence patients' trust in CLs?

7.4 Closing Remarks

My work addresses problems in enabling robots to realistically depict behaviors and appearances of a diverse group of humans and automatically interact with people in the real-world, which will enable robots to effectively and reliably deliver clinical educational interventions, and ultimately, facilitate improved health care outcomes. More specifically, my research addresses fundamental challenges in humanlike robot design and development within the context of HET. My work aims to transform how humanlike robots automatically interact with people, with the ultimate goal of enabling more realistic and effective human-robot interaction.

As humanlike robots integrate into environments centered around human interaction, it becomes crucial for them to effectively and reliably adjust their appearance, communication modalities, and behavior in a suitable, useful, empowering, and diverse way for each individual involved. Throughout my Ph.D., I designed and developed algorithms and systems that enable robots to realistically depict the behaviors and actions of a diverse group of individuals, and automatically interact with people in the real world.

My research opens new avenues of exploration for the advancement of humanlike robot technologies in the fields of robotics, HRI, FG, HET, and health informatics. My work will support healthcare education by triggering a new round of relevant technological innovations by creating the next generation of clinical educational technology. It will enable roboticists to develop robust methods for asymmetric facial expression analysis, modeling, and synthesis, and discover platform-independent methods to control the facial expressions of both robots and virtual agents, yielding new modalities for interaction. This work serves as a bridge between robotics and healthcare research and practice, and offers promising opportunities to reduce misdiagnoses and bias in healthcare, and, ultimately, explore reducing preventable patient harm. It is my aspiration that this research serves as an inspiration for researchers to conscientiously deliberate on the humanlikeness, efficiency, and ethical use of these systems in facilitating substantial support for individuals in their daily lives.

Appendix A

A.1 List of Acronyms

- **HRI**: Human robot interaction
- **FG**: Automatic facial and gesture recognition
- **HET**: Healthcare education and training
- **CEs**: clinical educators
- **CLs**: clinical learners
- **SHP**: Standardized human patients
- **APS**: Augmented reality patient simulators
- **AR**: Augmented reality
- **VPS**: Virtual patient simulators
- **RPS**: Robotic patient simulators
- **RPSwS**: Robotic patient simulators with stroke
- **PwS**: Human patients with stroke

- **PwoS**: Participants without stroke
- **BP**: Bell's Palsy
- **FP**: Facial palsy
- **A-FE**: Asymmetric facial expressions
- **FE**: Facial expressions
- **FL**: Facial landmarks
- **AU**: Action units
- **FACS**: Facial Action Coding System
- **FFPD**: Facial feature point detection
- **FEA**: Facial expression analysis systems
- **FAM**: Facial action modeling systems
- **FSA**: Facial expression synthesis and animation systems
- **AMS**: Analysis modeling and synthesis framework
- **FPM**: Facial paralysis masks framework
- **Stroke FPM**: Stroke-related facial paralysis masks framework
- **MMC**: Multi-modal communication framework
- **CN**: Cranial nerves
- **NAT**: Neurological assessment tests
- **SSM**: Stroke statistical measurement features

- **IMU**: Inertial measurement unit
- **ROSE**: An interactive social robot for medical education
- **FSM**: Finite State Machine
- **LTR**: Learner-to-robot interaction input
- **RTL**: Robot-to-learner interaction output
- **NLU**: natural language understanding
- **FAN**: Face alignment network
- **DNN**: Deep Neural Network
- **CNN**: Convolutional Neural Network
- **R-CNN**: Region-based Convolutional Neural Network
- **MLP**: Multilayer perceptron
- **AAM**: Active appearance models
- **SDM**: Supervised descent method
- **CLM**: Constrained local model
- **DL**: Deep Learning
- **DAE**: Deep autoencoder network
- **DSAE**: Deep sparse autoencoder network
- **CAE**: Contractive autoencoder network
- **GAN**: Generative adversarial networks

- **SVM**: Support vector machine
- **PCC**: Pearson correlation coefficient
- **DOF**: Degrees of freedom
- **ROI**: Region of interest
- **POI**: Points of interest
- **SDK**: Software developer kit
- **LOE**: Left outer eye corner
- **ROE**: Right outer eye corner
- **LIE**: Left inner eye corner
- **RIE**: Right inner eye corner
- **LB**: Left brow
- **RB**: Right brow
- **LLE**: central point of the left lower eyelid
- **RLE**: central point of the right lower eyelid
- **LUE**: central point of the left upper eyelid
- **RUE**: central point of the right upper eyelid
- **NT**: Nose tip
- **LL**: Central point of the lower lip
- **UL**: Central point of the upper lip

- **LC**: left lip corners

- **RC**: Right lip corner

- **LLF**: Live link face

- **IMU**: Inertial measurement unit

- **4-point DVAS**: 4-point discrete visual analogue scale

- **SUS**: System usability scale

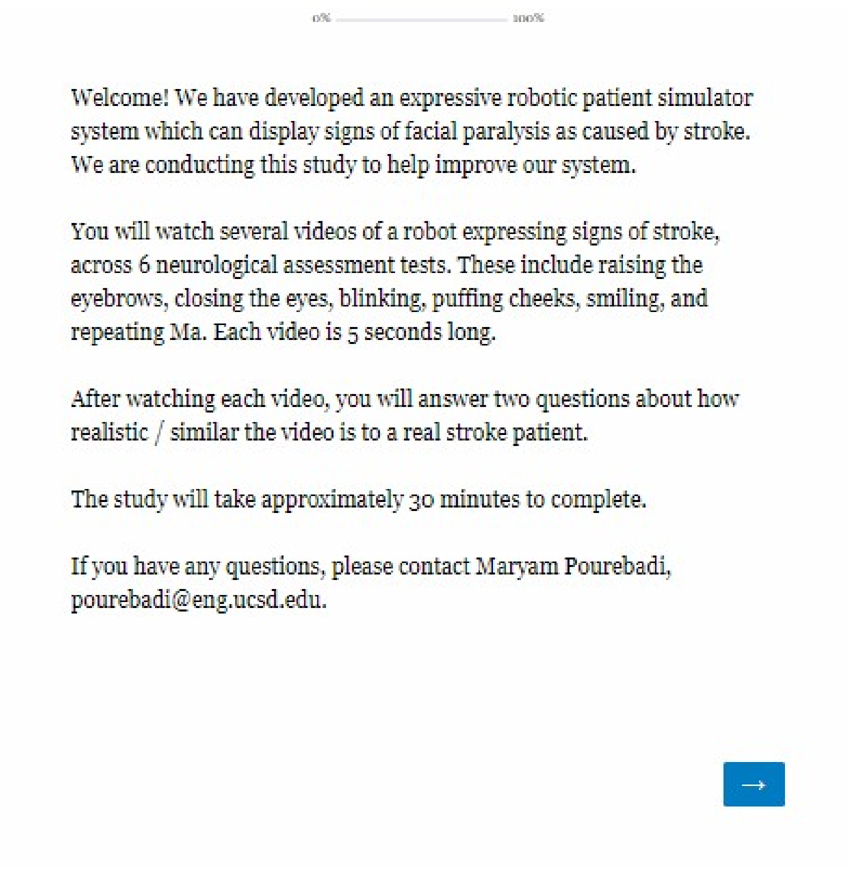
- **N-MD**: Physician with a specialty in neurocritical care

- **S-MD**: Clinical professor of medicine who works in medical education simulation

- **GE**: graduate engineering students

Appendix B

B.1 Interview questionnaire with clinicians to evaluate RPSwS



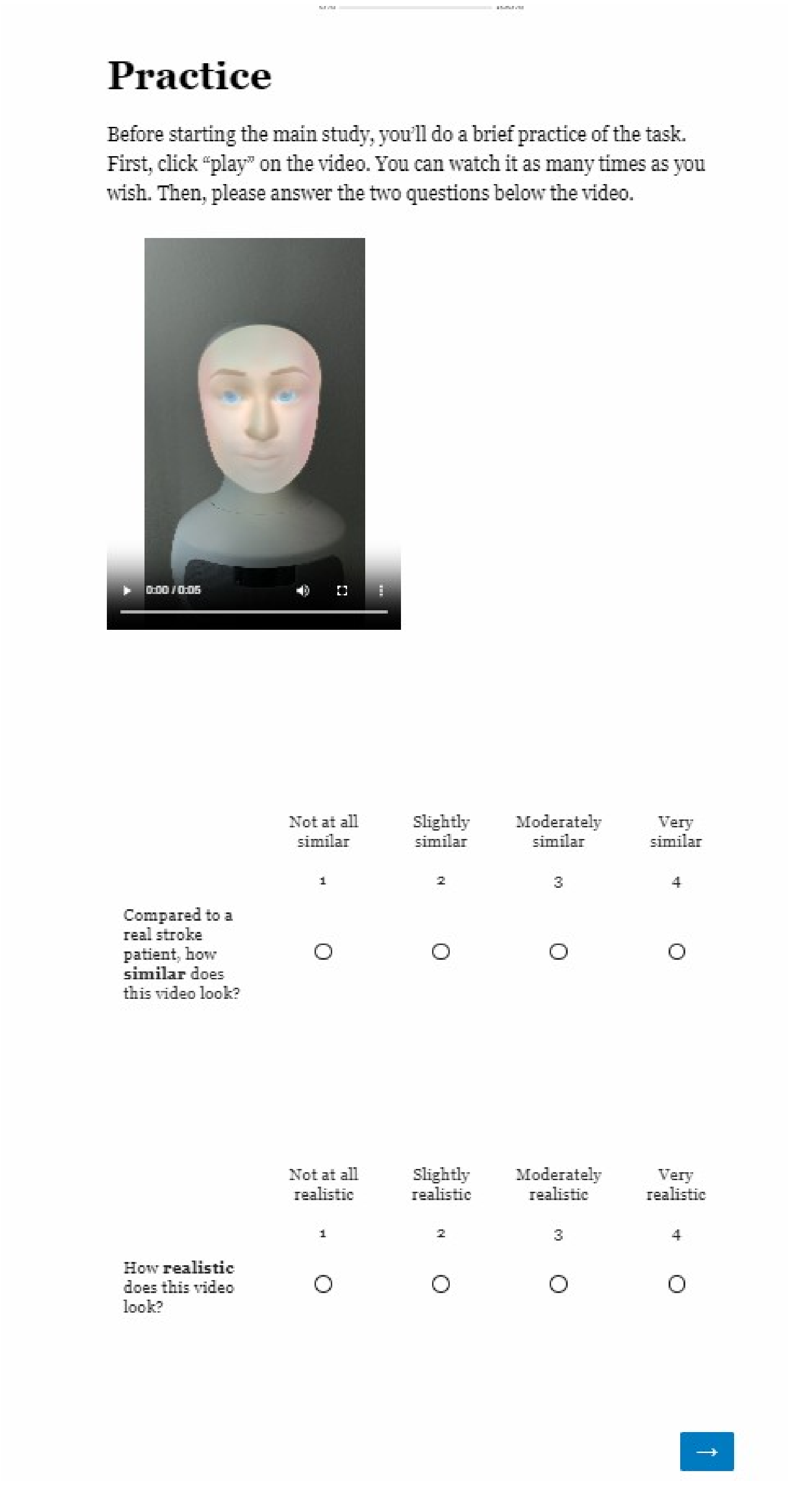
Welcome! We have developed an expressive robotic patient simulator system which can display signs of facial paralysis as caused by stroke. We are conducting this study to help improve our system.

You will watch several videos of a robot expressing signs of stroke, across 6 neurological assessment tests. These include raising the eyebrows, closing the eyes, blinking, puffing cheeks, smiling, and repeating Ma. Each video is 5 seconds long.

After watching each video, you will answer two questions about how realistic / similar the video is to a real stroke patient.

The study will take approximately 30 minutes to complete.

If you have any questions, please contact Maryam Pourebadi, pourebadi@eng.ucsd.edu.

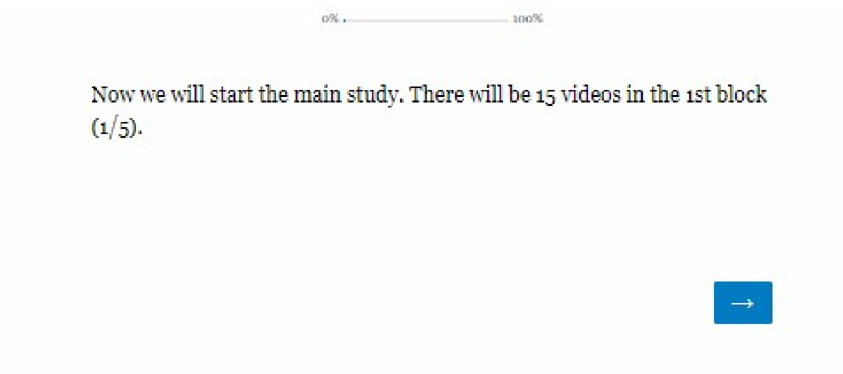


Practice

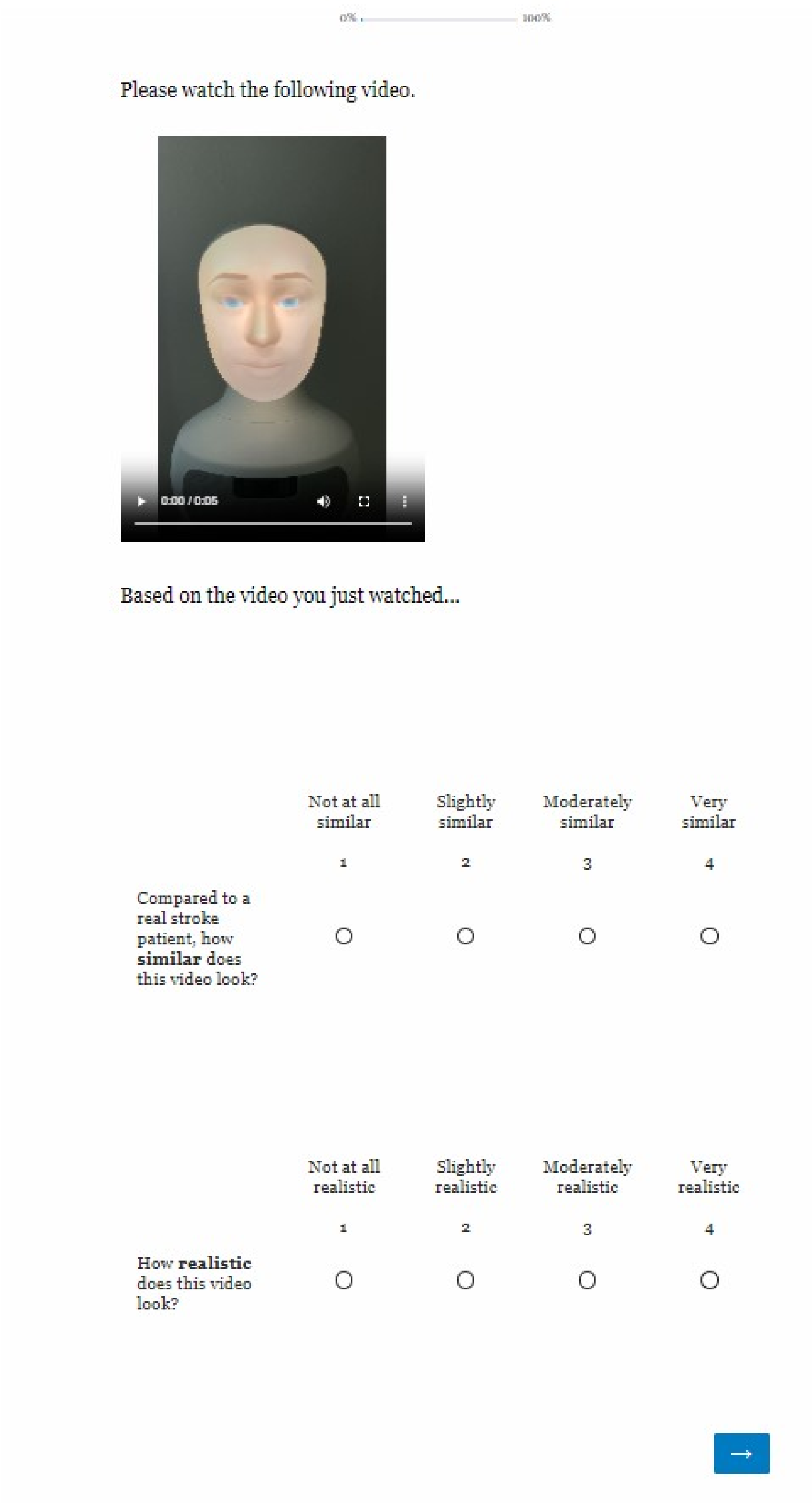
Before starting the main study, you'll do a brief practice of the task. First, click "play" on the video. You can watch it as many times as you wish. Then, please answer the two questions below the video.

	Not at all similar	Slightly similar	Moderately similar	Very similar
	1	2	3	4
Compared to a real stroke patient, how **similar** does this video look?	○	○	○	○

	Not at all realistic	Slightly realistic	Moderately realistic	Very realistic
	1	2	3	4
How **realistic** does this video look?	○	○	○	○



Now we will start the main study. There will be 15 videos in the 1st block (1/5).

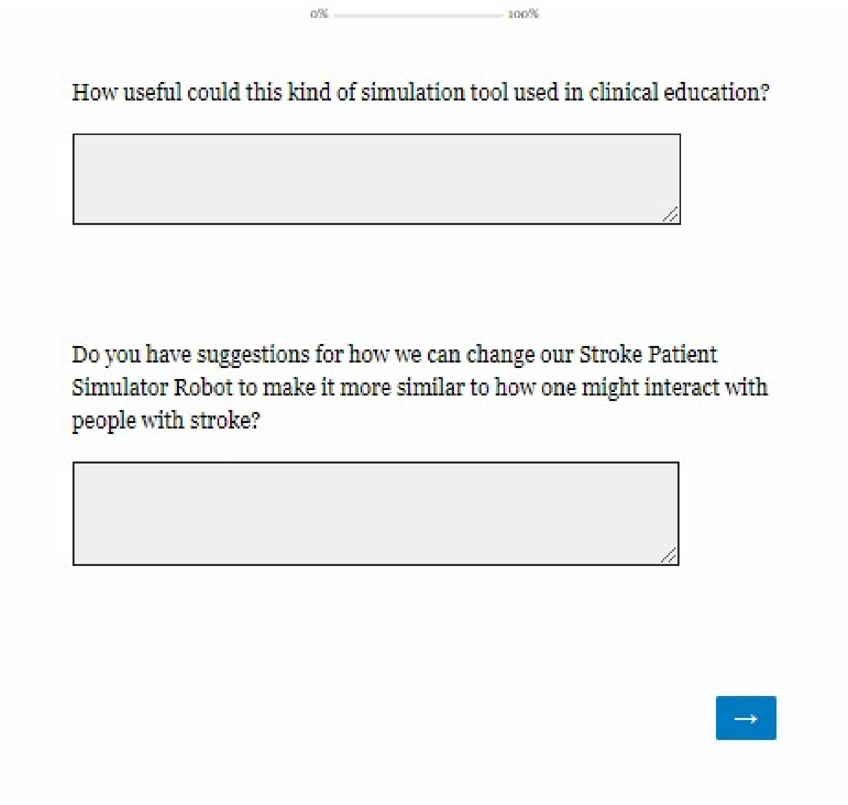


Please watch the following video.

Based on the video you just watched...

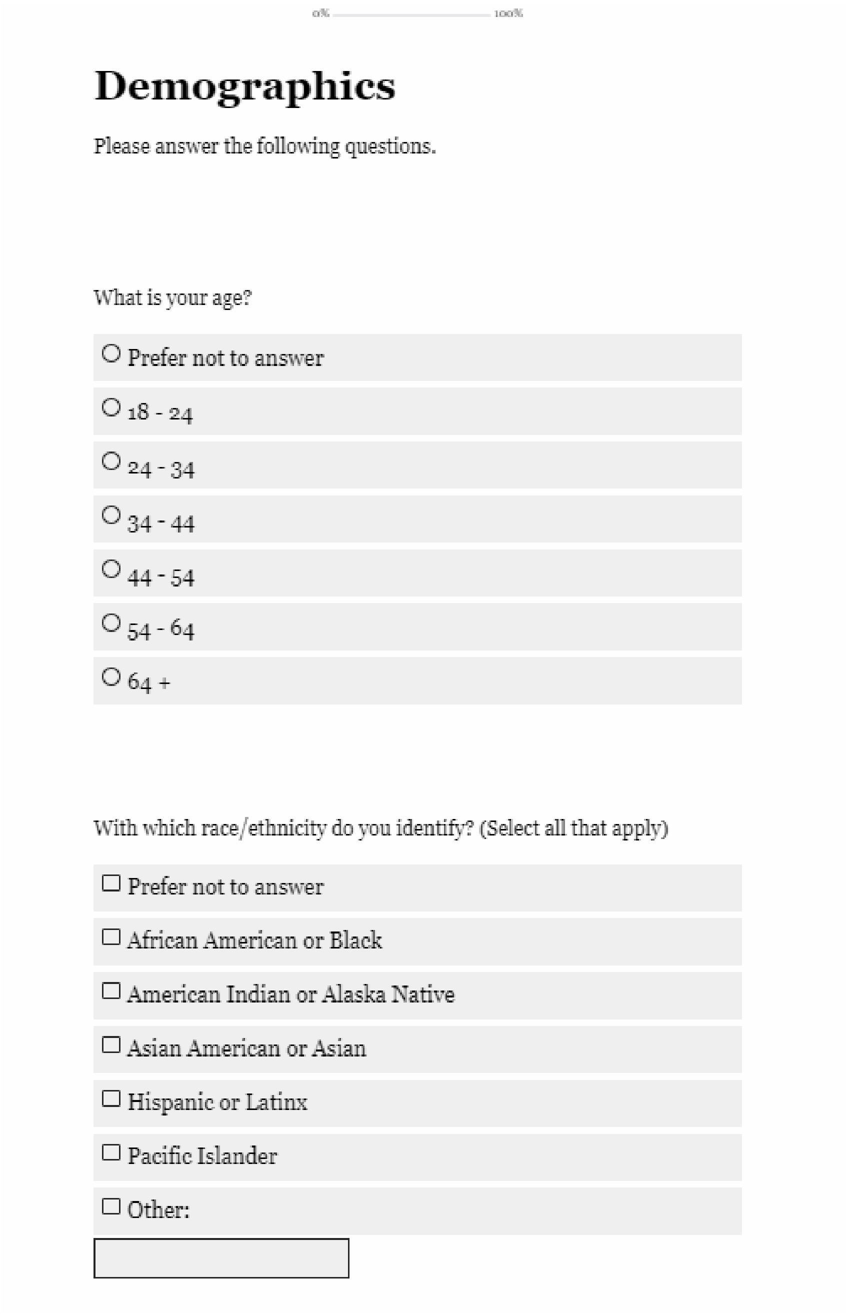
	Not at all similar	Slightly similar	Moderately similar	Very similar
	1	2	3	4
Compared to a real stroke patient, how **similar** does this video look?	○	○	○	○

	Not at all realistic	Slightly realistic	Moderately realistic	Very realistic
	1	2	3	4
How **realistic** does this video look?	○	○	○	○



How useful could this kind of simulation tool used in clinical education?

Do you have suggestions for how we can change our Stroke Patient Simulator Robot to make it more similar to how one might interact with people with stroke?



Demographics

Please answer the following questions.

What is your age?

○ Prefer not to answer

○ 18 - 24

○ 24 - 34

○ 34 - 44

○ 44 - 54

○ 54 - 64

○ 64 +

With which race/ethnicity do you identify? (Select all that apply)

□ Prefer not to answer

□ African American or Black

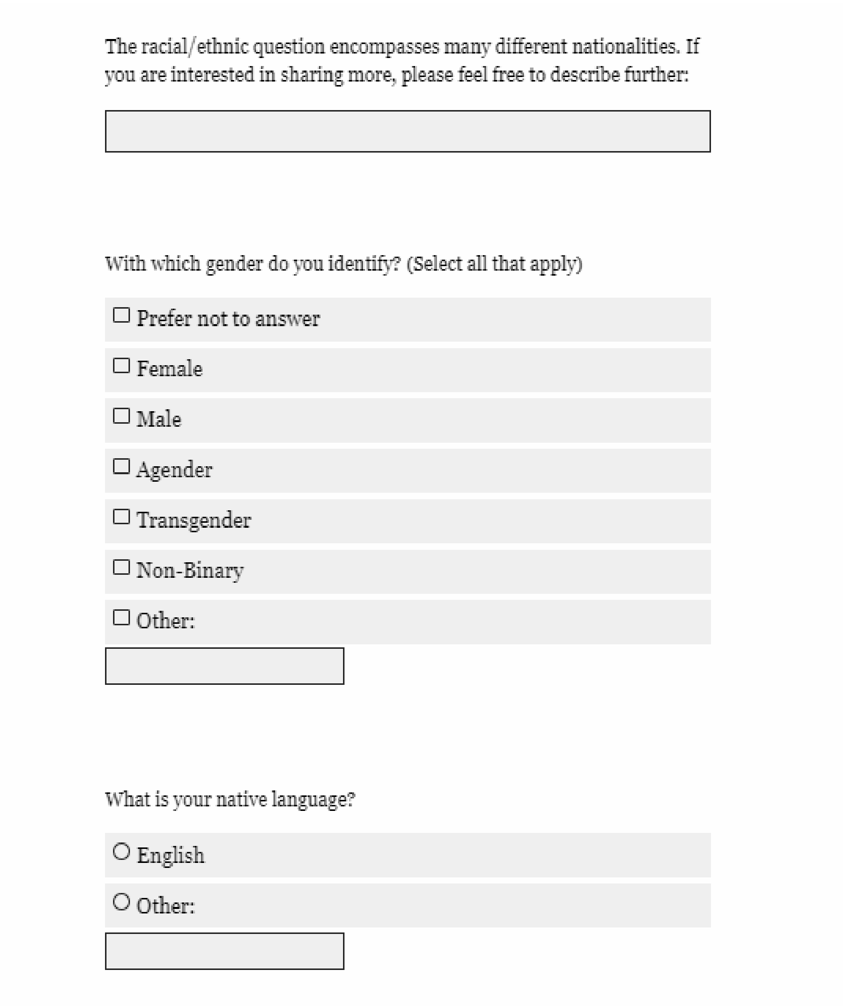
□ American Indian or Alaska Native

□ Asian American or Asian

□ Hispanic or Latinx

□ Pacific Islander

□ Other:



The racial/ethnic question encompasses many different nationalities. If you are interested in sharing more, please feel free to describe further:

With which gender do you identify? (Select all that apply)

□ Prefer not to answer

□ Female

□ Male

□ Agender

□ Transgender

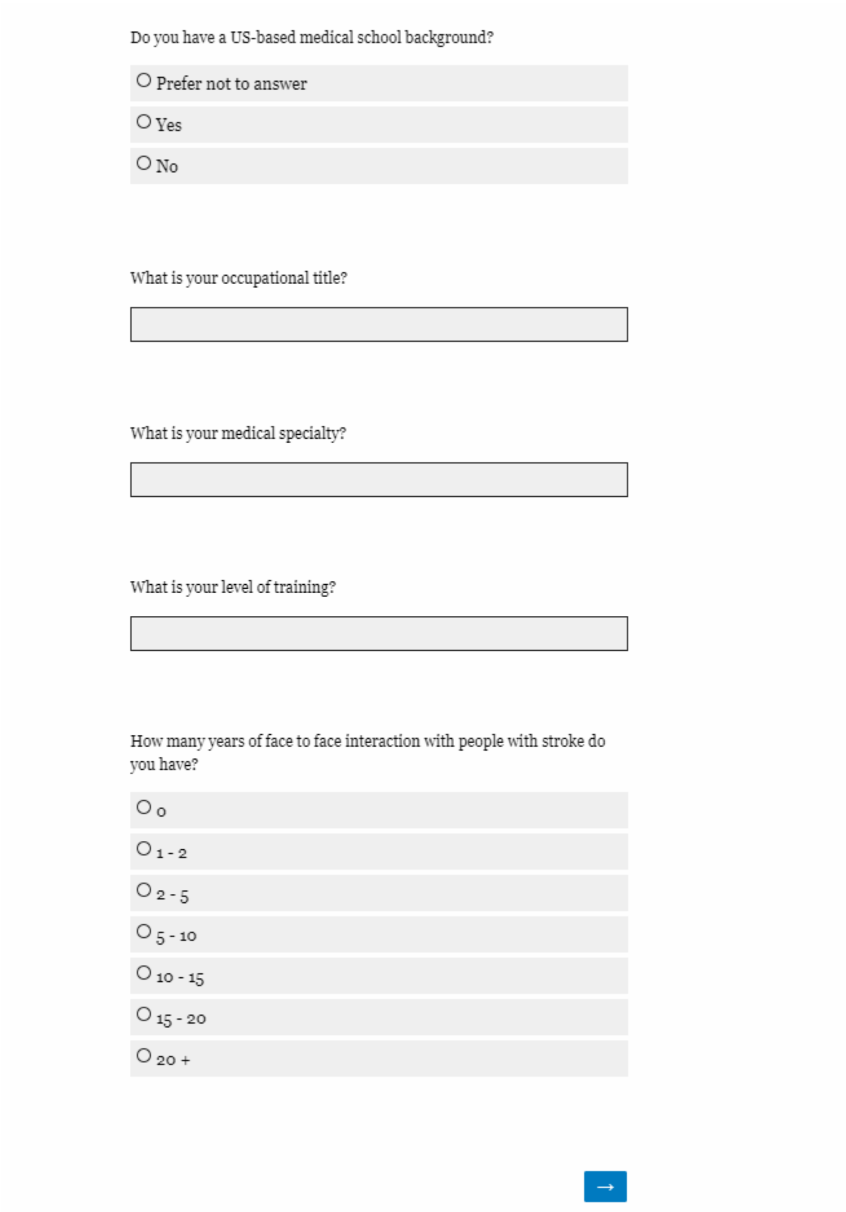
□ Non-Binary

□ Other:

What is your native language?

○ English

○ Other:



Do you have a US-based medical school background?

○ Prefer not to answer

○ Yes

○ No

What is your occupational title?

What is your medical specialty?

What is your level of training?

How many years of face to face interaction with people with stroke do you have?

○ 0

○ 1 - 2

○ 2 - 5

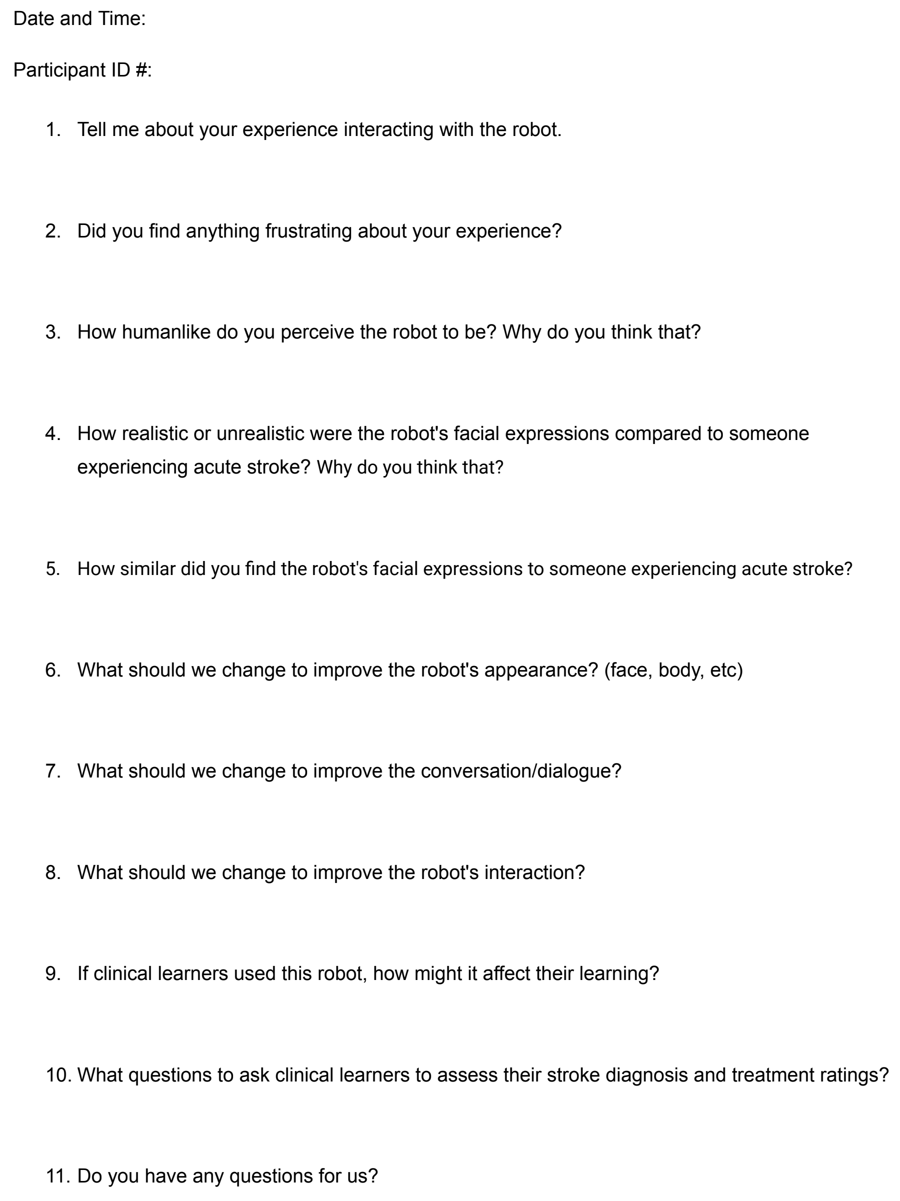
○ 5 - 10

○ 10 - 15

○ 15 - 20

○ 20 +

B.2 Interview questionnaire with clinicians to evaluate ROSE

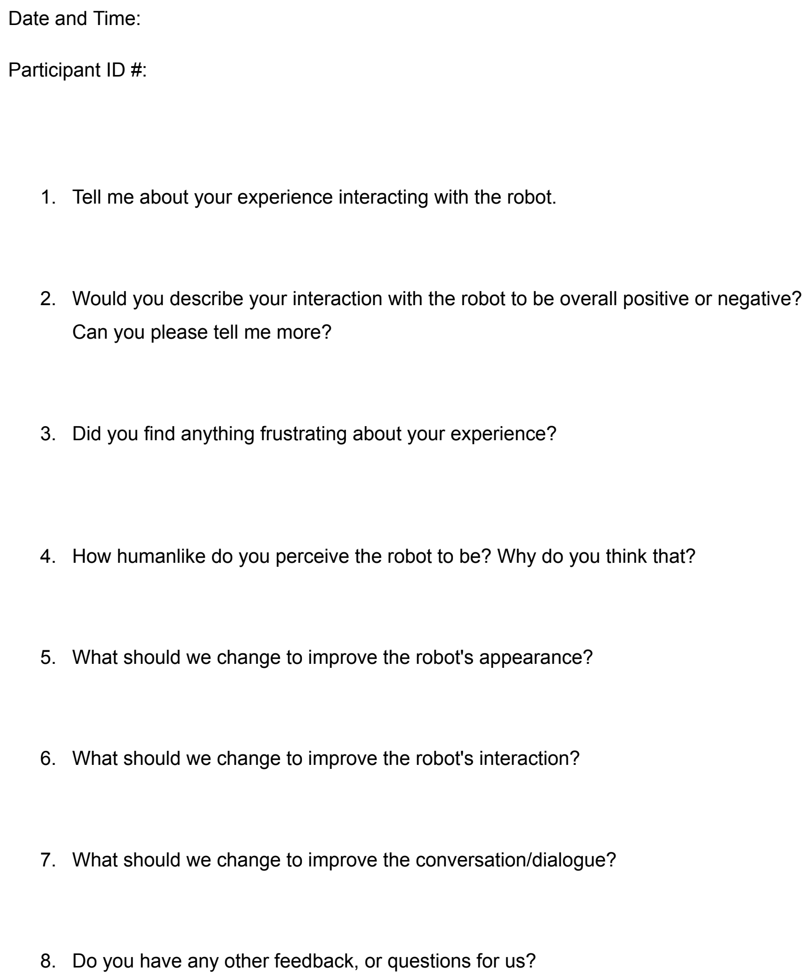


Date and Time:

Participant ID #:

1. Tell me about your experience interacting with the robot.
2. Did you find anything frustrating about your experience?
3. How humanlike do you perceive the robot to be? Why do you think that?
4. How realistic or unrealistic were the robot's facial expressions compared to someone experiencing acute stroke? Why do you think that?
5. How similar did you find the robot's facial expressions to someone experiencing acute stroke?
6. What should we change to improve the robot's appearance? (face, body, etc)
7. What should we change to improve the conversation/dialogue?
8. What should we change to improve the robot's interaction?
9. If clinical learners used this robot, how might it affect their learning?
10. What questions to ask clinical learners to assess their stroke diagnosis and treatment ratings?
11. Do you have any questions for us?

B.3 Interview questionnaire with graduate engineering students to evaluate ROSE

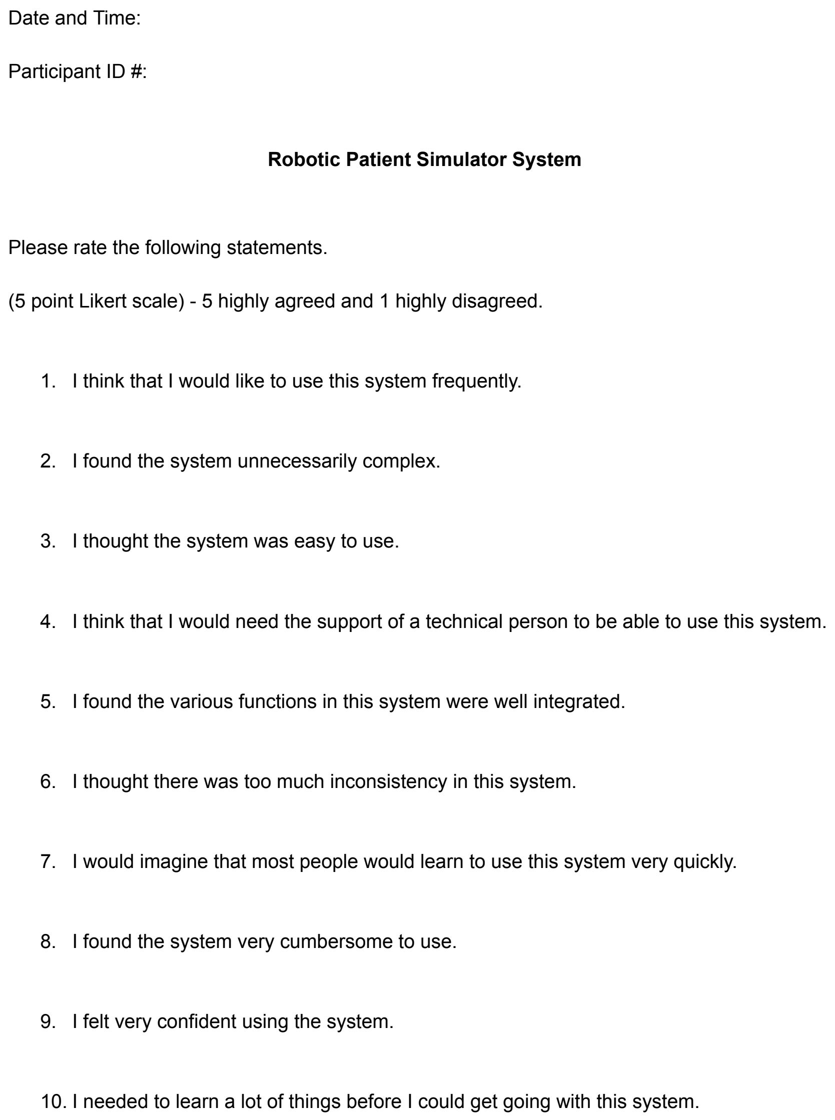


Date and Time:

Participant ID #:

1. Tell me about your experience interacting with the robot.
2. Would you describe your interaction with the robot to be overall positive or negative?
   Can you please tell me more?
3. Did you find anything frustrating about your experience?
4. How humanlike do you perceive the robot to be? Why do you think that?
5. What should we change to improve the robot's appearance?
6. What should we change to improve the robot's interaction?
7. What should we change to improve the conversation/dialogue?
8. Do you have any other feedback, or questions for us?

B.4 System Usability Scale Questionnaire to evaluate ROSE



Date and Time:

Participant ID #:

Robotic Patient Simulator System

Please rate the following statements.

(5 point Likert scale) - 5 highly agreed and 1 highly disagreed.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

Bibliography

[1] Adverse events in rehabilitation hospitals: National incidence among medicare beneficiaries.

[2] Metamotions. www.mbientlab.com/metamotions/. Accessed: 05-05-2023.

[3] Technical product overview. www.furhatrobotics.com/docs/Furhat-Robotics-Technical-Product-Overview.pdf. Accessed: 2023-04-28.

[4] Text to speech voices. www.acapela-group.com/voices/. Accessed: 2023-05-01.

[5] Kismet. robots.ieee.org/robots/kismet/, 1998. Accessed: 06-20-2020.

[6] The world health report 2002: reducing risks, promoting healthy life. www.who.int/whr/2002/en/whr02_en.pdf?ua=1/, 2002. Accessed: 09-09-2020.

[7] Diego-san research robot. www.hansonrobotics.com/diego-san/, 2013. Accessed: 07-31-2020.

[8] Clinispace offers healthcare training applications and engine platform. www.healthysimulation.com/5499/clinispace-offers-healthcare-training-applications-engine-platform/, 2014. Accessed: 06-20-2020.

[9] Advancing care excellence for seniors (ace.s). www.nln.org/professional-development-programs/teaching-resources/ace-s, 2017. Accessed: 06-20-2020.

[10] Openface. multicomp.cs.cmu.edu/resources/openface/, 2018. Accessed: 07-31-2020.

[11] Tug, one platform, multi-purpose. aethon.com/products/, 2018. Accessed: 06-20-2020.

[12] Facial action coding system (facs) – a visual guidebook. imotions.com/blog/facial-action-coding-system/, 2019. Accessed: 06-20-2020.

[13] Future robot. www.futurerobot.com, 2019. Accessed: 06-20-2020.

[14] Global health estimates: Life expectancy and leading causes of death and disability. www.who.int/data/gho/data/themes/mortality-and-global-health-estimates, 2019. Accessed: 08-15-2022.

[15] How far has cpr feedback come? www.laerdal.com/us/information/resusci-anne-then-and-now/, 2019. Accessed: 06-20-2020.

[16] Brain anatomy. www.my-ms.org/anatomy_brain_part4.html, 2020. Accessed: 08-15-2022.

[17] Buddy the first emotional companion robot. buddytherobot.com/en/buddy-the-emotional-robot/, 2020. Accessed: 07-31-2020.

[18] Da vinci by intuitive. www.intuitive.com/en-us/products-and-services/da-vinci, 2020. Accessed: 06-20-2020.

[19] Explore kuri. www.heykuri.com/explore-kuri/, 2020. Accessed: 07-31-2020.

[20] Faceposer. developer.valvesoftware.com/wiki/Faceposer, 2020. Accessed: 06-20-2020.

[21] Facereader. www.noldus.com/facereader, 2020. Accessed: 06-20-2020.

[22] Facial expression analysis. imotions.com/biosensor/fea-facial-expression-analysis/, 2020. Accessed: 06-20-2020.

[23] Furhat robot. furhatrobotics.com/furhat-robot/, 2020. Accessed: 06-20-2020.

[24] Gaumard simulators. www.gaumard.com/aboutsims/, 2020. Accessed: 06-20-2020.

[25] Greta. github.com/isir/greta, 2020. Accessed: 06-20-2020.

[26] Hey, i'm jibo. jibo.com/, 2020. Accessed: 07-31-2020.

[27] How to create 3d face animation the smart and slick way. www.//tinyurl.com/ye3ah2ss, 2020. Accessed: 09-09-2020.

[28] i-human. www.i-human.com, 2020. Accessed: 06-20-2020.

[29] Mamanatalie - birthing simulator. www.laerdal.com/us/mamaNatalie, 2020. Accessed: 06-20-2020.

[30] Mededportal – physician resident scenarios. www.mededportal.org, 2020. Accessed: 06-20-2020.

[31] Meet pediatric hal s2225. www.gaumard.com/s2225, 2020. Accessed: 07-31-2020.

[32] Minnesota simulation alliance. www.mnsimlib.org/, 2020. Accessed: 06-20-2020.

[33] Patient safety. www.who.int/patientsafety/en/, 2020. Accessed: 07-31-2020.

[34] Shadow health. www.shadowhealth.com/, 2020. Accessed: 06-20-2020.

[35] Simnewb. www.laerdal.com/us/doc/88/SimNewB, 2020. Accessed: 06-20-2020.

[36] Simroid patient simulation system for dental education. www.morita.com/group/en/products/educational-and-training-systems/training-simulation-system/simroid/, 2020. Accessed: 07-31-2020.

[37] Socibot. robotsoflondon.co.uk/socibot, 2020. Accessed: 07-31-2020.

[38] Sophia. www.hansonrobotics.com/sophia/, 2020. Accessed: 07-31-2020.

[39] Source sdk 2013. developer.valvesoftware.com/wiki/Source_SDK_2013, 2020. Accessed: 06-20-2020.

[40] Advocacy. www.facialpalsy.org.uk/support/useful-info/advocacy/, 2021. Accessed: 08-15-2022.

[41] Amazon polly developer guide. docs.aws.amazon.com/pdfs/polly/latest/dg/polly-dg.pdf, 2023. Accessed: 2023-05-01.

[42] What is the speech service? learn.microsoft.com/en-us/azure/cognitive-services/speech-service/overview, 4 2023. Accessed: 2023-05-01.

[43] D. Acharya, Z. Huang, D. Pani Paudel, and L. Van Gool. Covariance pooling for facial expression recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 367–374, 2018.

[44] G. Adamo. Simulated and standardized patients in osces: achievements and challenges 1992-2003. volume 25.

[45] A.H. Al-Elq. Simulation-based medical teaching and learning. In *Journal of Family and Community Medicine*, volume 17.

[46] M. Andrejevic and N. Selwyn. Facial recognition technology in schools: critical questions and concerns. volume 45, pages 115–128. Taylor & Francis, 2020.

[47] A.E. Arch, D.C. Weisman, S. Coca, K.V. Nystrom, C.R. Wira, and J.K. Schindler. Missed ischemic stroke diagnosis in the emergency department by emergency medicine and neurology services. In *Stroke*, volume 47, pages 668–673. booktitle of the American Heart Association, 2016.

[48] R. C. Arkin and M. J. Pettinati. Moral emotions, robots, and their role in managing stigma in early stage parkinson's disease caregiving. In *Workshop on New Frontiers of Service Robotics for the Eldery, RO-MAN*, 2014.

[49] R. Armstrong, T. Wright, S. de Ribaupierre, and R. Eagleson. Augmented reality for neurosurgical guidance: an objective comparison of planning interface modalities. In *Medical Imaging and Augmented Reality: 7th International Conference, MIAR 2016, Bern, Switzerland, August 24-26, 2016, Proceedings 7*, pages 233–243. Springer, 2016.

[50] B. Austin and C. Pidcock. Cereproc cerevoice cloud user guide. www.cereproc.com/files/CereVoiceCloudGuide.pdf, 2016. Accessed: 2023-05-01.

[51] C. Bacorn, N. S. Fong, and L. K. Lin. Misdiagnosis of bell's palsy: Case series and literature review. In *Clinical Case Reports*, volume 8, pages 1185–1191, 2020.

[52] T. Baltrušaitis, Amir Zadeh, Y.C. Lim, and L.P. Morency. Openface 2.0: Facial behavior analysis toolkit. 2018.

[53] A. Bandini, J. Green, B. Richburg, and Y. Yunusova. Automatic detection of orofacial impairment in stroke. In *Interspeech*, pages 1711–1715, 2019.

[54] A. Bandini, J. R. Green, B. Taati, S. Orlandi, L. Zinman, and Y. Yunusova. Automatic detection of amyotrophic lateral sclerosis (als) from video-based analysis of facial movements: Speech and non-speech tasks. In *IEEE International Conference on Automatic Face Gesture Recognition (FG)*, pages 150–157, 2018.

[55] A. Bandini, S. Rezaei, D. L. Guarin, M. Kulkarni, D. Lim, M. I. Boulos, L. Zinman, Y. Yunusova, and B. Taati. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE journal of biomedical and health informatics*, 25(4):1111–1119, 2020.

[56] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, 4(3):114–123, 2009.

[57] H.S. Barrows. An overview of the uses of standardized patients for teaching and evaluating clinical skills. volume 68, pages 443–443. ASSOCIATION OF AMERICAN MEDICAL COLLEGES, 1993.

[58] R. F. Baugh, G. J. Basura, L. E. Ishii, S. R. Schwartz, C.M. Drumheller, and R. Burkholder. Clinical practice guideline bell's palsy. *Otolaryngology–Head and Neck Surgery*, 2013.

[59] C. Becker-Asano and H. Ishiguro. Evaluating facial displays of emotion for the android robot Geminoid F. In *IEEE Workshop on Affective Computational Intelligence (WACI)*, 2011.

[60] M. Benedek, D. Daxberger, S. Annerer-Walcher, and J. Smallwood. Are you with me? probing the human capacity to recognize external/internal attention in others' faces. volume 26, pages 511–517. Taylor & Francis, 2018.

[61] R. Benjamin. Race after technology: Abolitionist tools for the new jim code. volume 98, pages 1–3, 2019.

[62] T. Bickmore, A. Rubin, and S. Simon. Substance use screening using virtual agents: Towards automated screening, brief intervention, and referral to treatment (sbirt). In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20)*, 2020.

[63] Y. Blumberg. Here's the real reason health care costs so much more in the us. www.cnbc.com/2018/03/22/the-real-reason-medical-care-costs-so-much-more-in-the-us.html, 2018. Accessed: 07-31-2020.

[64] K. R. Bogart, L. Tickle-Degnen, and N. Ambady. Communicating without the face: Holistic perception of emotions of people with facial paralysis. *Basic and Applied Social Psychology*, 2014.

[65] H. Boughrara, M. Chtourou, B.C. Amar, and L. Chen. Facial expression recognition based on a mlp neural network using constructive training algorithm. volume 75, pages 709–731. Springer, 2016.

[66] C. Breazeal. *Designing sociable robots*. MIT press, 2004.

[67] R. Breuer and R. Kimmel. A deep learning perspective on the origin of facial expressions. 2017.

[68] J. Brooke. Sus: A 'quick and dirty'usability scale, usability evaluation in industry, 1996.

[69] A. Bulat and G. Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1021–1030, 2017.

[70] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *In Proceedings of Machine Learning Research at Conference on Fairness, Accountability, and Transparency*, 2018.

[71] J.K. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli. Social signal processing. 2017.

[72] E. Burkov, I. Pasechnik, A. Grigorev, and V. Lempitsky. Neural head reenactment with latent pose descriptors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13786–13795, 2020.

[73] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. volume 43, pages 172–186. IEEE, 2019.

[74] M. Carbone, R. Piazza, and S. Condino. Commercially available head-mounted displays are unsuitable for augmented reality surgical guidance: A call for focused research for surgical applications. In *Surgical innovation*, volume 27, pages 254–255. SAGE Publications Sage CA: Los Angeles, CA, 2020.

[75] C. Carcel, M. Woodward, X. Wang, C. Bushnell, and E. C. Sandset. Sex matters in stroke: a review of recent evidence on the differences between women and men. *Frontiers in Neuroendocrinology*, 59:100870, 2020.

[76] D.F. Carter. Man-made man: Anesthesiological medical human simulator. volume 3, pages 80–86, 1969.

[77] H. R. Champion and A.G. Gallagher. Surgical simulation—a 'good idea whose time has come'. volume 90, pages 767–768. Oxford University Press, 2003.

[78] K Charmaz. *Constructing grounded theory*. Sage, 2014.

[79] C. Chen, O.G.B. Garrod, J. Zhan, J. Beskow, P.G. Schyns, and R.E. Jack. Reverse engineering psychologically valid facial expressions of emotion into social robots. 2018.

[80] C. Chen, K.B. Hensel, Y. Duan, R.A. Ince, O.G.B. Garrod, J. Beskow, R.E. Jack, and P.G. Schyns. Equipping social robots with culturally-sensitive facial expressions of emotion using data-driven methods. 2019.

[81] H. Christensen, N. Amato, H. Yanco, M. Mataric, W. Choset, A. Drobnis, K. Goldberg, J. Grizzle, G. Hager, J. Hollerbach, et al. A roadmap for us robotics–from internet to robotics 2020 edition. *Foundations and Trends® in Robotics*, 8(4):307–424, 2021.

[82] Riek L. D. et al. Christensen, H. I. A roadmap for us robotics: From internet to robotics 2016 edition. *Computing Community Consortium*, 2016.

[83] W. Chu, F. De la Torre, and J. F. Cohn. Learning facial action units with spatiotemporal cues and multi-label sampling. volume 81, pages 1–14. Elsevier, 2019.

[84] D.C. Classen, R. Resar, F. Griffin, F. Federico, T. Frankel, N. Kimmel, J.C. Whittington, A. Frankel, A. Seger, and B.C. James. Global trigger tool" shows that adverse events in hospitals may be ten times greater than previously measured. volume 30, pages 581–589, 2011.

[85] R. L. Conigliaro, K. D. Peterson, and T. D. Stratton. Lack of diversity in simulation technology: an educational limitation? *Simulation in Healthcare*, 15(2):112–114, 2020.

[86] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. In *IEEE TranS.Pattern AnaK.MacH.Intell.*, 1998.

[87] D. Cristinacce and T. Cootes. Automatic feature localisation with constrained local models. volume 41, pages 3054–3067. Elsevier, 2008.

[88] S. Daher, J. Hochreiter, N. Norouzi, L. Gonzalez, G. Bruder, and G. Welch. Physical-virtual agents for healthcare simulation. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 99–106, 2018.

[89] S. Daher, J. Hochreiter, R. Schubert, L. Gonzalez, J. Cendan, M. Anderson, D.A Diaz, and G.F. Welch. The physical-virtual patient simulator a physical human form with virtual appearance and behavior. In *Simulation in Healthcare*, volume 15, pages 115–121. LWW, 2020.

[90] F. De la Torre, W. Chu, X. Xiong, F. Vicente, X. Ding, and J.F. Cohn. Intraface. 2015.

[91] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5203–5212, 2020.

[92] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou. Retinaface: Single-stage dense face localisation in the wild. 2019.

[93] C. Diana. How our robots will charm us (and why we want them to). sonarplusd.com/en/programs/barcelona-2017/areas/talks/how-our-robots-will-charm-us-and-why-we-want-them-to, 2017. Accessed: 06-20-2020.

[94] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of the 31st International Conference on Machine Learning*, pages 647–655. PMLR, 2014.

[95] E.S. Donkor. Stroke in the 21 st century: A snapshot of the burden, epidemiology, and quality of life. In *Stroke research and treatment*, volume 2018. Hindawi, 2018.

[96] P. Ekman and W. Friesen. *Facial Action Coding System: Investigator's Guide*. Consulting Psychologists Press, 1978.

[97] P. Ekman, E.R. Sorenson, and W.V. Friesen. Pan-cultural elements in facial displays of emotion. volume 164, pages 86–88. American Association for the Advancement of Science, 1969.

[98] R. El Kaliouby and P. Robinson. Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction*. New York: Springer, 2005.

[99] H. A. Elfenbein, M. Beaupré, M. Lévesque, and U. Hess. Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. volume 7, page 131. American Psychological Association, 2007.

[100] J. B. Engelmann and M. Pogosyan. Emotion perception across cultures: the role of cognitive mechanisms. volume 4, page 118. Frontiers, 2013.

[101] N. Ersotelos and F. Dong. Building highly realistic facial modeling and animation: a survey. volume 24, pages 13–30. Springer, 2008.

[102] I. Ertugrul, L. A. Jeni, and J. F. Cohn. Pattnet: Patch-attentive deep network for action unit detection. In *BMVC*, page 114, 2019.

[103] I. O. Ertugrul, J. F. Cohn, L. A. Jeni, Z. Zhang, L. Yin, and Q. Ji. Crossing domains for au coding: Perspectives, approaches, and measures. volume 2, pages 158–171. IEEE, 2020.

[104] S. Ethier, W.J. Wilson, and C. Hulls. Telerobotic part assembly with shared visual servo control. 2002.

[105] S. Fabi, X. Xu, and V.R. de Sa. Exploring the racial bias in pain detection with a computer vision model. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44, 2022.

[106] B. Fasel. Robust face analysis using convolutional neural networks. In *Object recognition supported by user interaction for service robots*, volume 2, pages 40–43. IEEE, 2002.

[107] A.E. Frank, A. Kubota, and L.D. Riek. Wearable activity recognition for robust human-robot teaming in safety-critical environments via hybrid neural networks. 2019.

[108] H. Gardener, R. L. Sacco, T. Rundek, V Battistella, Y. Cheung, and M. Elkind. Race and ethnic disparities in stroke incidence in the northern manhattan study. *Stroke*, 51(4):1064–1069, 2020.

[109] N. F. Garmann-Johnsen, T. Mettler, and M. Sprenger. Service robotics in healthcare: A perspective for information systems researchers? 2014.

[110] B. Gecer, A. Lattas, S. Ploumpis, J. Deng, A. Papaioannou, S. Moschoglou, and S. Zafeiriou. Synthesizing coupled 3d face modalities by trunk-branch generative adversarial networks. In *European Conference on Computer Vision*, pages 415–433. Springer, 2020.

[111] M. Ghayoumi. A quick review of deep learning in facial expression. volume 14, pages 34–8, 2017.

[112] M. Ghayoumi. A quick review of deep learning in facial expression. *J. Commun. Comput*, 14(1):34–38, 2017.

[113] M. Ghayoumi and M. Pourebadi. Fuzzy knowledge-based architecture for learning and interaction in social robots. *arXiv preprint arXiv:1909.11004*, 2019.

[114] J. Giger, N. Piçarra, P. Alves-Oliveira, R. Oliveira, and P. Arriaga. Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies*, 1(2):111–123, 2019.

[115] K.H. Glantz. Conducting research with children: Legal and ethical issues. volume 35, page 1283–1291, 1996.

[116] C. Goldberg. Ucsd's practical guide to clinical medicine. meded.ucsd.edu/clinicalmed/neuro2.html, 2018. Accessed: 12-30-2019.

[117] M.A. Goodrich and A.C. Schultz. *Human-Robot Interaction: A Survey*. Now Publishers Inc., 2007.

[118] T. Gorman, J. Dropkin, J. Kamen, S. Nimbalkar, N. Zuckerman, T. Lowe, J. Szeinuk, D. Milek, G. Piligian, and A. Freund. Controlling health hazards to hospital workers. volume 23, pages 1–169. SAGE Publications Sage CA: Los Angeles, CA, 2014.

[119] S.J. Goyal, A.K. Upadhyay, R.S. Jadon, and R. Goyal. Real-life facial expression recognition systems: A review. volume 77, pages 311–331, 2018.

[120] J.D. Greer, T.K. Morimoto, A.M. Okamura, and E.W. Hawkes. A soft, steerable continuum robot that grows via tip extension. volume 6, pages 95–108, 2019.

[121] C. A. Grimm. Adverse events in hospitals: a quarter of medicare patients experienced harm in october 2018. *Office of Inspector General, I. General*, page 117, 2022.

[122] D. Guarin, A. Dempster, A. Bandini, Y. Yunusova, and B. Taati. Estimation of orofacial kinematics in parkinson's disease: Comparison of 2d and 3d markerless systems for motion tracking. In *IEEE International Conference on Automatic Face Gesture Recognition (FG)*, 2020.

[123] E. Guizzo. Hiroshi ishiguro: The man who made a copy of himself. spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself, 2010. Accessed: 07-31-2020.

[124] J. Hamm, C.G. Kohler, R.C. Gur, and R. Vermaa. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. volume 200, pages 237–256. Elsevier, 2011.

[125] A. Hansson and M. Servin. Semi-autonomous shared control of large-scale manipulator arms. 2010.

[126] T. Hashimoto, S. Hitramatsu, T. Tsuji, and H. Kobayashi. Development of the face robot saya for rich facial expressions. 2006.

[127] N. Hassen, A. Lofters, S. Michael, A. Mall, A. D. Pinto, and J. Rackal. Implementing anti-racism interventions in healthcare settings: a scoping review. *International Journal of Environmental Research and Public Health*, 18(6):2993, 2021.

[128] P. Hellyer. Preventable patient harm is expensive. volume 227, pages 275–275. Nature Publishing Group, 2019.

[129] A. G Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017.

[130] P. Hu and D. Ramanan. Finding tiny faces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 951–959, 2017.

[131] W. Huang. When hci meets hri: the intersection and distinction. 2015.

[132] Y. Huang, F. Chen, S. Lv, and X. Wang. Facial expression recognition: A survey. volume 11, page 1189. Multidisciplinary Digital Publishing Institute, 2019.

[133] R. E Jack. Culture and facial expressions of emotion. volume 21, pages 1248–1286. Taylor & Francis, 2013.

[134] R. E Jack, W. Sun, I. Delis, O. GB Garrod, and P. G Schyns. Four not six: Revealing culturally common facial expressions of emotion. volume 145, page 708. American Psychological Association, 2016.

[135] R.E. Jack, O.G. Garrod, and P.G. Schyns. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. volume 24, pages 187–192. Elsevier, 2014.

[136] J. T. James. A new evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety*, 2013.

[137] P. R. Jeffries. Simulation in nursing education: from conceptualization to education. In *In National League for Nursing*, 2007.

[138] C. O. Johnson, M. Nguyen, G. A Roth, E. Nichols, T. Alam, D. Abate, F. Abd-Allah, A. Abdelalim, H. N. Abraha, N. M. Abu-Rmeileh, et al. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, 18(5):439–458, 2019.

[139] A. Kalegina, G. Schroeder, A. Allchin, K. Berlin, and M. Cakmak. Characterizing the design space of rendered robot faces. 2018.

[140] S. Kaltwang, O. Rudovic, and M. Pantic. Continuous pain intensity estimation from facial expressions. In *International Symposium on Visual Computing*, pages 368–377. Springer, 2012.

[141] S. J. Karve, R. Balkrishnan, Y. M. Mohammad, and D. A. Levine. Racial/ethnic disparities in emergency department waiting time for stroke patients in the united states. *Journal of Stroke and Cerebrovascular Diseases*, 20(1):30–40, 2011.

[142] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.

[143] V. Kerrigan, N. Lewis, A. Cass, M. Hefler, and A. P. Ralph. "how can i do more?" cultural awareness training for hospital-based healthcare providers working with high aboriginal caseload. *BMC Medical Education*, 20:1–11, 2020.

[144] D. Kollias, V. Sharmanska, and S. Zafeiriou. Face behavior a la carte: Expressions, affect and action units in a single network. 2019.

[145] D. Kollias and S. Zafeiriou. Aff-wild2: Extending the aff-wild database for affect recognition. 2018.

[146] A. A. Kononowicz, K.A. Woodham, S. Edelbring, N. Stathakarou, D. Davies, N. Saxena, K.T. Car, J. Carlstedt-Duke, J. Car, and N. Zary. Virtual patient simulations in health professions education: Systematic review and meta-analysis by the digital health education collaboration. In *Journal of medical Internet research*, volume 21, page e14676. JMIR Publications Inc., Toronto, Canada, 2019.

[147] J. Kossaifi, R. Walecki, Y. Panagakis, J. Shen, M. Schmitt, F. Ringeval, J. Han, V. Pandit, A. Toisoul, B. W. Schuller, K. Star, E. Hajiyev, and M. Pantic. Sewa db: A rich database for audio-visual emotion and sentiment research in the wild. volume 43, pages 1022–1040. IEEE, 2019.

[148] A. Kubota, D. Cruz-Sandoval, S. Kim, E. W. Twamley, and L. D. Riek. Cognitively assistive robots at home: Hri design patterns for translational science. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 53–62. IEEE, 2022.

[149] J. Laybourn. Meet the robot that can mimic human emotion. www.cambridge-news.co.uk/news/cambridge-news/cambridge-university-robot-human-emotion-14431300, 2018.

[150] Q. V. Le. Building high-level features using large scale unsupervised learning. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 8595–8598. IEEE, 2013.

[151] M. J. Leo and D. Manimegalai. 3d modeling of human faces- a survey. 2011.

[152] J. Li, D. Zhang, J. Zhang, J. Zhang, T. Li, Y. Xia, Q. Yan, and L. Xun. Facial expression recognition with faster r-cnn. volume 107, pages 135–140. Elsevier, 2017.

[153] S. Li and W. Deng. Deep facial expression recognition: A survey. IEEE, 2020.

[154] W. Li, F. Abtahi, and Z. Zhu. Action unit detection with region adaptation, multi-labeling learning and optimal temporal fusing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1841–1850, 2017.

[155] Y. Li, J. Zeng, S. Shan, and X. Chen. Occlusion aware facial expression recognition using cnn with attention mechanism. volume 28, pages 2439–2450. IEEE, 2018.

[156] Y. Li, J. Zeng, S. Shan, and X. Chen. Self-supervised representation learning from videos for facial action unit detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10924–10933, 2019.

[157] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett. The computer expression recognition toolbox (cert). 2011.

[158] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 57–64. IEEE, 2011.

[159] Q. Mao, Q. Rao, Y. Yu, and M. Dong. Hierarchical bayesian theme models for multipose facial expression recognition. volume 19, pages 861–873. IEEE, 2016.

[160] A. G. Marson and R. Salinas. Clinical evidence: Bell's palsy. *Western Journal of Medicine*, 2000.

[161] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic. Automatic analysis of facial actions: A survey. volume 10, pages 325–347. IEEE, 2017.

[162] S. Matsumoto, P. Ghosh, R. Jamshad, and L. D. Riek. Robot, uninterrupted: Telemedical robots to mitigate care disruption. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 495–505, 2023.

[163] S. Matsumoto, S. Moharana, N. Devanagondi, L. C. Oyama, and L. D. Riek. Iris: A low-cost telemedicine robot to support healthcare safety and equity during a pandemic. In *Pervasive Computing Technologies for Healthcare: 15th EAI International Conference, Pervasive Health 2021, Virtual Event, December 6-8, 2021, Proceedings*, pages 113–133. Springer, 2022.

[164] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn. Disfa: A spontaneous facial action intensity database. volume 4, pages 151–160. IEEE, 2013.

[165] D. Mazzei, N. Lazzeri, D. Hanson, and D. De Rossi. Hefes: An hybrid engine for facial expressions synthesis to control human-like androids and avatars. In *IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012.

[166] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R.E. Kaliouby. Affdex sdk: A cross-platform realtime multi-face expression recognition toolkit. 2016.

[167] D. Moores. Facial paralysis.

[168] M. Moosaei, S. K. Das, D. O. Popa, and L. D. Riek. Using facially expressive robots to calibrate clinical pain perception. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 2017.

[169] M. Moosaei, M. J. Gonzales, and L. D. Riek. Naturalistic pain synthesis for virtual patients. In *International Conference on Intelligent Virtual Agents*, 2014.

[170] M. Moosaei, C. J. Hayes, and L. D. Riek. Performing facial expression synthesis on robot faces: A real-time software system. *In Proceedings of the 4th Int. AISB Symposium on New Frontiers in Human-Robot Interaction, AISB*, 2015.

[171] M. Moosaei, C.J. Hayes, and L.D. Riek. Facial expression synthesis on robots: An ros module. 2015.

[172] M. Moosaei, M. Pourebadi, and L.D. Riek. Modeling and synthesizing idiopathic facial paralysis. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2019.

[173] M. Mori. The uncanny valley: The original essay by masahiro mori. 2012.

[174] P. Morton. Creating a laboratory that simulates the critical care environment. *Critical Care Nursing, 16(6), 76–81. National Association for Associate Degree Nursing*, 1995.

[175] N.L. Nelson and J.A. Russell. Universality revisited. volume 5, pages 8–15, 2013.

[176] D. E. Newman-Toker, E. Moy, E. Valente, R Coffey, and A. L. Hines. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. In *Diagnosis*, volume 1, pages 155–166. De Gruyter, 2014.

[177] S. Nishio, H. Ishiguro, and N. Hagita. Geminoid: Teleoperated android of an existing person. volume 14, pages 343–352. Vienna, Austria: I-Tech Education and Publishing, 2007.

[178] S. U. Noble. *Algorithms of oppression*. NYU Press, 2018.

[179] S. U. Noble. Tech won't save us: Reimagining digital technologies for the public. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, pages 1–1, 2020.

[180] I. Ntinou, E. Sanchez, A. Bulat, M. Valstar, and Y. Tzimiropoulos. A transfer learning approach to heatmap regression for action unit intensity estimation. pages 1–1. IEEE, 2021.

[181] N. Otberdout, A. Kacem, M. Daoudi, L. Ballihi, and S. Berretti. Deep covariance descriptors for facial expression recognition. 2018.

[182] H. Owen. Early use of simulation in medical education. volume 7, pages 102–116. LWW, 2012.

[183] S. Ozturkcan and E. Merdin-Uygur. Humanoid service robots: The future of healthcare? *Journal of Information Technology Teaching Cases*, 12(2):163–169, 2022.

[184] M. Pantic and M.S. Bartlett. Machine analysis of facial expressions. In *Face Recognition*. IntechOpen, 2007.

[185] S. Park and M. Whang. Empathy in human-robot interaction: Designing for social robots. *International Journal of Environmental Research and Public Health*, 19(3):1889, 2022.

[186] C. Pelachaud. Greta, an interactive expressive embodied conversational agent. 2015.

[187] C. Pelachaud. Greta: A conversing socio-emotional agent. 2017.

[188] B. Pierce, T. Kuratate, C. Vogl, and G. Cheng. Dmask-bot 2i": An active customisable robotic head with interchangeable face. In *IEEE-RAS International Conference on Humanoid Robots*, 2012.

[189] H. Y. Ping, K.N. Abdullah, P.S. Sulaiman, and A.A. Halin. Computer facial animation: A review. volume 5, page 658. IACSIT Press, 2013.

[190] M. Pourebadi. *A Deep Learning Approach for Blind Image Quality Assessment*. PhD thesis, Kent State University, 2017.

[191] M. Pourebadi, J. Labuzetta, and L.D. Riek. Modeling and synthesizing stroke on expressive patient simulator robots. In *The ACM Transactions on Human-Robot Interaction (THRI)*, 2023. Pending Submission.

[192] M. Pourebadi, J. N. LaBuzetta, C. Gonzalez, P. Suresh, and L. D. Riek. Mimicking acute stroke findings with a digital avatar. In *STROKE*, volume 51, 2020.

[193] M. Pourebadi, S. Pai, R. Pei, and L.D. Riek. Rose: An interactive social robot for medical education. In *The ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Pending Submission.

[194] M. Pourebadi and M. Pourebadi. Multilayer perceptron neural network based approach for facial expression analysis. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, page 263, 2016.

[195] M. Pourebadi and L. D. Riek. Facial expression modeling and synthesis for patient simulator systems: Past, present, and future. In *ACM Transactions on Computing for Healthcare*, volume 3. Association for Computing Machinery, 2022.

[196] M. Pourebadi and L.D. Riek. Expressive robotic patient simulators for clinical education. In *R4L Workshop on Robots for Learning - Inclusive Learning, Workshop at the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018.

[197] M. Pourebadi and L.D. Riek. Stroke modeling and synthesis for robotic and virtual patient simulators. In *AAAI Artificial Intelligence for Human-Robot Interaction (AAAI AI-HRI): Trust & Explainability in Artificial Intelligence for Human-Robot Interaction*, 2020.

[198] V. Powar and A. Jahagirdar. Reliable face detection in varying illumination and complex background. In *International Conference on Communication, Information & Computing Technology*, pages 1–4. IEEE, 2012.

[199] T. J. Prescott and J. M. Robillard. Are friends electric? the benefits and risks of human-robot relationships. *Iscience*, 24(1):101993, 2021.

[200] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer. Ganimation: Anatomically-aware facial animation from a single image. In *Proceedings of the European conference on computer vision (ECCV)*, pages 818–833, 2018.

[201] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer. Ganimation: One-shot anatomically consistent facial animation. volume 128, pages 698–713. Springer, 2020.

[202] M. Racic, M. I. Roche-Miranda, and G. Fatahi. Twelve tips for implementing and teaching anti-racism curriculum in medical education. *Medical Teacher*, pages 1–6, 2023.

[203] R. Raina, M. Monares, M. Xu, S. Fabi, X. Xu, L. Li, W. Sumerfield, J. Gan, and V. R. de Sa. Exploring biases in facial expression analysis using synthetic faces. In *NeurIPS 2022 Workshop on Synthetic Data for Empowering ML Research*.

[204] M. Ramacciotti, M. Milazzo, F. Leoni, S. Roccella, and C. Stefanini. A novel shared control algorithm for industrial robots. volume 13, page 1729881416682701. SAGE Publications Sage UK: London, England, 2016.

[205] J. Redfern, C. McKevitt, and C. Wolfe. Development of complex interventions in stroke care. In *ASA Stroke*, 2006.

[206] I.M. Revina and W.R.S. Emmanuel. A survey on human face expression recognition techniques. volume 33, pages 619–628. Elsevier, 2021.

[207] L. D. Riek. Healthcare robotics. *Communictions of the ACM, Vol. 60, No. 11. pp. 68-78*, 2017.

[208] L.D. Riek. *Expression Synthesis on Robots*. PhD thesis, University of Cambridge, 2011.

[209] L.D. Riek. System and method for robotic patient synthesis, 2016. US Patent 9,280,147.

[210] L.D. Riek and P. Robinson. Using robots to help people habituate to visible disabilities. In *2011 IEEE International Conference on Rehabilitation Robotics*, pages 1–8. IEEE, 2011.

[211] S. Rifai, V. Pascal, M. Xavier, G. Xavier, and B. Yoshua. Contractive auto-encoders: Explicit invariance during feature extraction. In *International Conference on Machine Learning*, 2011.

[212] D. Rivera-Gutierrez, G. Welch, P. Lincoln, M. Whitton, JJ. Cendan, D. Chesnutt, H. Fuchs, and B. Lok. Shader lamps virtual patients: The physical manifestation of virtual patients. In *Medicine Meets Virtual Reality 19*, pages 372–378. IOS Press, 2012.

[213] K. L. Robey, P. M. Minihan, L. M. Long-Bellil, J. E. Hahn, J. G. Reiss, G. E. Eddey, Alliance for Disability in Health Care Education, et al. Teaching health care students about disability within a cultural competency context. *Disability and Health Journal*, 2013.

[214] S. Robla-Gómez, V.M. Becerra, J.R. Llata, E. González-Sarabia, C. Torre-Ferrero, and J. Pérez-Oria. Working together: A review on safe human-robot collaboration in industrial environments. volume 5, pages 26754–26773. IEEE, 2017.

[215] T.L. Rodziewicz and J.E. Hipskind. Medical error prevention. 2020.

[216] H. Salam and R. Séguier. A survey on face modeling: building a bridge between face analysis and synthesis. volume 34, pages 289–319. Springer, 2018.

[217] E. Sanchez, A. Bulat, A. Zaganidis, and G. Tzimiropoulos. Semi-supervised facial action unit intensity estimation with contrastive learning. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

[218] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. volume 7, pages 413–422. Oxford University Press, 2012.

[219] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813, 2014.

[220] S.R. Shrivastava, P.S. Shrivastava, and J.D. Ramasamy. Reduction in global burden of stroke in underserved areas. In *Journal of neurosciences in rural practice*, volume 4, pages 475–476. Thieme Medical and Scientific Publishers Private Ltd., 2013.

[221] B. Singer. The human simulation lab—dissecting sex in the simulator lab: The clinical lacuna of transsexed embodiment. volume 34, pages 249–254. Springer, 2013.

[222] E. Sjödin. Enhancing the classroom experience with social robots. www.robotminds.se/post/enhancing-the-classroom-experience-with-social-robots, 2023. Accessed: 05-05-2023.

[223] D. P. Smith. Research of strokes and misdiagnosis. www.mccormick.northwestern.edu/news/articles/2022/05/of-strokes-and-misdiagnosis/, 2022. Accessed: 05-05-2023.

[224] Spectrum Speech. www.spectrumspeech.ie/stroke. Accessed: 05-05-2023.

[225] C. R. Stephenson, S. L. Bonnes, A. P. Sawatsky, L. W. Richards, C. D. Schleck, J. N. Mandrekar, T. J. Beckman, and C. M. Wittich. The relationship between learner engagement and teaching effectiveness: a novel assessment of student engagement in continuing medical education. *BMC medical education*, 20(1):1–8, 2020.

[226] S. Stockli, Schulte-Mecklenbeck, S. M, Borer, and A.C. Samson. Facial expression analysis with affdex and facet: A validation study. volume 50, pages 1446–1460. Springer, 2018.

[227] M. Strait, A. Ramos, V. Contreras, and N. Garcia. Robots racialized in the likeness of marginalized social identities are subject to greater dehumanization than those racialized as white. In *IEEE-RAS International Conference on Humanoid Robots*, 2018.

[228] S. Strilciuc, D. A. Grad, C. Radu, D. Chira, A. Stan, M. Ungureanu, A. Gheorghe, and F. D. Muresanu. The economic burden of stroke: a systematic review of cost of illness studies. In *booktitle of medicine and life*, 2021.

[229] H. Su, W. Qi, J. Chen, C. Yang, J. Sandoval, and M Laribi. Recent advancements in multimodal human–robot interaction. *Frontiers in Neurorobotics*, 17, 2023.

[230] C. Suarez, MD. Menendez, J. Alonso, N. Castaño, M. Alonso, and F. Vazquez. Detection of adverse events in an acute geriatric hospital over a 6-year period using the global trigger tool. volume 62, pages 896–900. Wiley Online Library, 2014.

[231] B. Sun, l. Li, g. Zhou, x. Wu, J. He, L. Yu, D. Li, and Q. Wei. Combining multimodal features within a fusion network for emotion recognition in the wild. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 497–502, 2015.

[232] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. 2013.

[233] H. Sunvisson, B. Habermann, S. Weiss, and P. Benner. Augmenting the cartesian medical discourse with an understanding of the person's lifeworld, lived body, life story and social identity. *Nursing Philosophy*, 2009.

[234] L. Surace, M. Patacchiola, E. Battini Sönmez, W. Spataro, and A. Cangelosi. Emotion recognition in the wild using deep neural networks and bayesian classifiers. 2017.

[235] A.A. Tarnutzer, S. Lee, K.A. Robinson, Z. Wang, J.A. Edlow, and D.E. Newman-Toker. Ed misdiagnosis of cerebrovascular events in the era of modern neuroimaging. In *American Academy of Neurology*, volume 88, pages 1468–1477. AAN Enterprises, 2017.

[236] A. Taylor, H. Lee, A. Kubota, and L.D. Riek. Simulation-based medical teaching and learning. 2019.

[237] A. Taylor, S. Matsumoto, and L.D. Riek. Situating robots in the emergency department. 2020.

[238] A. Taylor, S. Matsumoto, W. Xiao, and L. D Riek. Social navigation for mobile robots in the emergency department. 2021.

[239] Y.I. Tian, T. Kanade, and J.F. Cohn. Recognizing action units for facial expression analysis. 2001.

[240] L. Tickle-Degnen, L. A. Zebrowitz, and H. Ma. Culture, gender and health care stigma: Practitioners' response to facial masking experienced by people with parkinson's disease. *Social Science & Medicine*, 2011.

[241] J. Toole, S. Chmieleski, M. Dekker, A. Elstein, D. Jones, R. Kelly, S. Sanghvi, S. Shapiro, C. Taylor, N Walczak, S. Teppema, S. Siegel, and b. Scott. The economic measurement of medical errors. 2010.

[242] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, M. Fiore, H. Hung, O.A.I. Ramírez, M. Joosse, H. Khambhaita, T. Kucner, B. Leibe, A.J. Lilienthal, T. Linder, M. Lohse, M. Magnusson, B. Okal, L. Palmieri, U. Rafi, M. van Rooij, and L. Zhang. Spencer: A socially aware service robot for passenger guidance and help in busy airports. In *Field and service robotics*, pages 607–622. Springer, 2016.

[243] T. Tsai. Using children as standardised patients for assessing clinical competence in paediatrics. volume 89, pages 1117–1120. BMJ Publishing Group Ltd, 2004.

[244] O. Tysnes and A. Storstein. Epidemiology of parkinson's disease. *Journal of Neural Transmission*, 2017.

[245] M. Unbeck, K. Schildmeijer, P. Henriksson, U. Jürgensen, O. Muren, L. Nilsson, and K. Pukk Härenstam. Is detection of adverse events affected by record review methodology? an evaluation of the "harvard medical practice study" method and the "global trigger tool". volume 7, pages 1–12. BioMed Central, 2013.

[246] B. A. Urgen, M. Kutas, and A. P. Saygin. Uncanny valley as a window into predictive processing in the social brain. volume 114, pages 181–185. Elsevier, 2018.

[247] D. Uzelli Yilmaz, A. Azim, and M. Sibbald. The role of standardized patient programs in promoting equity, diversity, and inclusion: A narrative review. *Academic Medicine*, 97(3):459–468, 2022.

[248] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. 2001.

[249] A.W. Walter, C. Julce, N. Sidduri, L. Yinusa-Nyahkoon, J. Howard, M. Reichert, T. Bickmore, and B.W. Jack. Study protocol for the implementation of the gabby preconception care system - an evidence-based, health information technology intervention for black and african american women. volume 20, pages 1–14. Springer, 2020.

[250] N. Wang, X. Gao, D. Tao, and X. Li. Facial feature point detection: A comprehensive survey. volume 275, pages 50–65. Elsevier, 2018.

[251] S. Wang, B. Pan, S. Wu, and Q. Ji. Deep facial action unit recognition and intensity estimation from partially labelled data. volume 12, pages 1018–1030. IEEE, 2019.

[252] S. Wang and G. Peng. Weakly supervised dual learning for facial action unit recognition. volume 21, pages 3218–3230. IEEE, 2019.

[253] W. Wang, Q. Sun, T. Chen, C. Cao, Z. Zheng, G. Xu, H. Qiu, and Y. Fu. A fine-grained facial expression database for end-to-end multi-pose facial expression recognition. 2019.

[254] C. Watson and T.K. Morimoto. Permanent magnet-based localization for growing robots in medical applications. volume 5, pages 2666–2673. IEEE, 2020.

[255] K. Watson. Asymmetrical face: What is it, and should you be concerned? 2023. Accessed: 05-13-2023.

[256] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. Picard. Automatic recognition methods supporting pain assessment: A survey. pages 1–1, 2019.

[257] A. Williams. Facial expression of pain: An evolutionary account. volume 25, pages 439–455. Cambridge University Press, 2002.

[258] A.D. Wilson and S.N. Bathiche. Compact interactive tabletop with projection-vision. In *US10026177B2 Patent*, 2013.

[259] T. Wright, S. de Ribaupierre, and R. Eagleson. Design and evaluation of an augmented reality simulator using leap motion. In *Healthcare technology letters*, volume 4, 2017.

[260] Q. Wu, L. Zhao, and X. Ye. Shortage of healthcare professionals in china. volume 354. British Medical Journal Publishing Group, 2016.

[261] Y. Wu and Q. Ji. Constrained joint cascade regression framework for simultaneous facial action unit recognition and facial landmark detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3400–3408, 2016.

[262] Y. Wu and Q. Ji. Facial landmark detection: a literature survey. volume 127, pages 115–142. Springer, 2019.

[263] X. Xiong and F. De La Torre. Supervised descent method and its applications to face alignment. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[264] X. Xu, K. D. Craig, D. Diaz, M. S. Goodwin, M. Akcakaya, B. Susam, J. S. Huang, and V. R. de Sa. Automated pain detection in facial videos of children using human-assisted transfer learning. In *International Workshop on Artificial Intelligence in Health*, pages 162–180. Springer, 2018.

[265] X. Xu and V. R. de Sa. Exploring multidimensional measurements for pain evaluation using facial action units. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 786–792. IEEE, 2020.

[266] X. Xu and V. R. de Sa. Personalized pain detection in facial video with uncertainty estimation. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 4163–4168. IEEE, 2021.

[267] X. Xu, J. S. Huang, and V. R. De Sa. Pain evaluation in video using extended multitask learning from multidimensional measurements. In *ML4H@ NeurIPS*, pages 141–154, 2019.

[268] A. Zadeh, T. Baltrušaitis, and L.P. Morency. Constrained local model for facial landmark detection. 2017.

[269] E. Zakharov, A. Shysheya, E. Burkov, and V. Lempitsky. Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9459–9468, 2019.

[270] A. Zewe. Giving robots social skills a new machine-learning system helps robots understand and perform certain social interactions. www.news.mit.edu/2021/robots-social-skills-1105, 2021. Accessed: 05-05-2023.

[271] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. volume 23, pages 1499–1503. IEEE, 2016.

[272] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. volume 32, pages 692–706. Elsevier, 2014.

[273] Y. Zhang, H. Jiang, B. Wu, Y. Fan, and Q. Ji. Context-aware feature and label fusion for facial action unit intensity estimation with partially labeled data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 733–742, 2019.

[274] X. Zhao, X. Shi, and S. Zhang. Facial expression recognition via deep learning. 2012.

[275] I. Zubrycki, I. Szafarczyk, and G. Granosik. Project fantom: Co-designing a robot for demonstrating an epileptic seizure. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2018.