CONTENT

Tomislav Stipančić[1], Leon Koren[1]

Affective Interactive Virtual Agent that can Occupy Different Environments

[1]University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture, Ivana Lučića 5, 10000 Zagreb, Croatia

Abstract

This paper introduces a virtual being and a robot capable of nuanced non-verbal communication in virtual, mixed, and physical environments. The prototype of such an advanced system is PLEA, an "emotion-aware" virtual being. PLEA assesses a person's emotional state during interaction based on a multimodal approach and then uses that information during non-verbal communication. PLEA leverages AI-driven sensory systems to interpret visual and auditory data, enabling it to understand and respond appropriately to human emotions and actions in various realistic scenarios. The research focuses on endowing this entity with the ability to convey and interpret non-verbal cues such as body language, facial expressions, and contextual gestures.

The main goal of PLEA is to maintain human-robot interaction by building mutual understanding and common ground. We address the challenges in AI algorithms related to emotion and gesture recognition, 3D modeling, information visualization, and sensor integration, which are crucial for creating expressive, environment-appropriate non-verbal behaviors. We propose a multi-layered framework that enables virtual beings to adapt their non-verbal communication strategies to each specific environment. In virtual spaces, the emphasis is on creating expressive avatars. For mixed reality, integration into real-world contexts is crucial, while in physical environments, the challenge involves translating digital expressions into physical forms.

Keywords: *virtual beings, creative AI, cognitive robotics, multimodal interaction, mixed reality*

1. Introduction

Virtual beings, encompassing digital avatars and agents, represent the convergence of AI and virtual reality.

Virtual beings (VBs) represent embodying entities that can interact with humans in an intuitive and meaningful way [1]. They are not only programmed to engage in complex dialogues but are also capable of displaying a range of emotions and understanding human sentiments, making them increasingly indispensable in sectors like education, entertainment, and mental health [2].

Cognitive robotics, on the other hand, extends the capabilities of traditional robotics by integrating AI to create machines that can perceive, reason, and learn from their environment [3]. This allows them to operate autonomously in a variety of settings, from assisting in homes and workplaces to performing tasks in environments that are hazardous for humans. The cognitive aspect involves the robot's ability to make sense of the world through sensors and data, mimicking human cognitive processes such as learning, memory, and decision-making.

The synergy between virtual beings and cognitive robotics is particularly evident in mixed reality environments, where physical and digital realms coalesce. Here, virtual beings can be embodied through robots, allowing for a tangible interaction that blurs the lines between what's virtual and what's real. This symbiosis opens up new frontiers for human-machine interaction, where virtual companions can provide support, education, and companionship in more immersive and personalized ways.

The potential for virtual beings and cognitive robotics to transform human life is immense [4]. The power of AI can be used to create entities that understand humans and adapt to their needs, ultimately leading to a future where technology enhances every aspect of human experience.

PLEA is a virtual being and a robot that has been developed to analyze and employ behaviors using a form of biomimicry [5]. PLEA is based on recent research findings in human cognition, cognitive robotics, and human-robot interaction to develop new robot reasoning and interaction strategies. PLEA relies on Deep Learning AI and multimodal information fusion to predict the possible responses of the person interacting with it.

2. Robot and the virtual being design

PLEA represents a cutting-edge virtual being and a robot, conceived to inhabit the vast expanse of cyberspace.

PLEA's existence is anchored on a physical server, which, through the intricate web of the internet, bridges it to the outside world. In this way PLEA can be approached via every computer connected to Internet. A software agent can move through cyberspace from interface to interface depending on which interface requires communication. If the software agent is currently occupied with communication through one interface, another user requesting communication will receive a notification of the occupation. Once the initial communication is finished, the software agent will become available. Although it is possible to create a software agent that can communicate with multiple users simultaneously, this method was chosen to give the agent a personality. Such an agent is unique regardless of the interface used and the interaction environment. This unique setup empowers PLEA to transcend traditional boundaries and manifest itself across a multitude of environments, ranging from purely virtual environments to augmented realities and even interfaces that connect the digital with the physical.

The essence of PLEA's design is its versatility and adaptability. The PLEA software agent uses a contextual approach to reasoning, where the robot is seen as part of the environment, regardless of whether it is the real world or cyberspace. In these diverse settings, PLEA is not just a passive inhabitant, but an active participant. It is programmed to recognize and interpret human emotions, respond to commands, and initiate interactions, making it an intelligent companion capable of providing assistance, entertainment, or companionship. The potential applications of such a virtual being are vast, from serving as a personal assistant in smart homes, guiding users in virtual learning environments, to acting as a mediator in remote collaborations.

PLEA, as an "emotion-aware" system designed to reason and operate within realistic scenarios, where its capability to perceive, interpret, and respond to human emotions is paramount. This sophisticated level of emotional intelligence enables PLEA to understand and adapt to the nuanced emotional states of individuals, facilitating interactions that are not only responsive but also empathetic and supportive [6].

3. Virtual being and the robot design

The development of a virtual being using Unreal Engine's MetaHuman application is a sophisticated process that merges cutting-edge technology with creative design to bring hyper-realistic characters to life [7]. The initial step involves conceptualizing the virtual being's appearance, personality, and purpose, which serves as the foundation for its creation.

Upon finalizing the concept, the development moves into the MetaHuman Creator, a cloud-streamed application that allows for the creation of photorealistic human models with an unprecedented level of detail, as shown at Fig. 1.

In this step, it is possible to select from a wide range of preset faces that can be finely adjusted to achieve the desired look. Features such as skin complexion, eye color, hair style, and facial hair can be meticulously customized, enabling the creation of a unique virtual being that closely matches the initial concept.

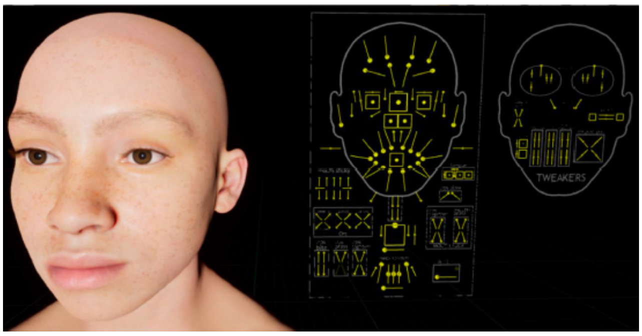

Fig. 1. MetaHuman control rig

The next phase involves rigging the MetaHuman for animation, a process that Unreal Engine simplifies through its advanced skeletal and facial rigging system [8]. This system provides a comprehensive set of tools for animating complex human motions and expressions, making it possible to bring the virtual being to life with realistic movements and emotional expressions.

Once the character is rigged, developers can integrate it into the Unreal Engine environment, where they utilize the engine's powerful rendering capabilities to achieve lifelike visuals. Lighting, shading, and environmental effects are fine-tuned to enhance the virtual being's realism within its virtual surroundings.

The final step involves programming the virtual being's behaviors and interactions using Unreal Engine's visual scripting system, Blueprint. This allows developers to create complex scenarios where the MetaHuman can interact with users or other elements within the virtual world, responding to inputs and exhibiting behaviors that reflect its programmed personality and purpose.

The computational architecture used in PLEA is a context-to-data interpreter that endows the machine with the capability to 'reason' based on constantly changing perspectives. In this way, PLEA can make decisions based on newly acquired information that is incomplete at the time of deciding on a particular behavior.

Deep Learning plays a significant role in managing a large amount of unstructured information that can be used to create control mechanisms and to approximate various phenomena or processes occurring in cyberspace or the real world [9]. Primarily, this refers to the development of neural network models that should enable the recognition of emotions based on facial expressions and by analyzing the speech of the person during the interaction. Fig 2. depicts the multimodal architecture of the model that controls PLEA's emotion recognition mechanisms.

The model consists of three primary components: acoustic, linguistic, and visual [10]. These modalities are integrated using a specific algorithm for combining information, which relies on assigning different weights to various factors. For visual data processing, the ResNet 300 face detection algorithm is employed to glean information directly from a real-time video stream. Furthermore, it incorporates a methodology inspired by Savchenko [11] to facilitate emotion detection. The acoustic component is engineered to distill features from sound recordings by employing techniques such as spectrograms and Mel-frequency cepstral coefficients [12]. The data is then classified using a convolutional neural network that has been previously trained. The linguistic component utilizes a manually developed feature extraction tool alongside a bag-of-words model to understand and process language. Emotion and intent recognition are accomplished using the Long Short-Term Memory (LSTM) algorithm [13]. In addition, specialized hardware is utilized to separate speech from any ambient noise and to identify the interaction targets of the system.

The updated algorithm operates by integrating data from all three modalities simultaneously during the information fusion process. It employs both weighted and base factors in equal proportion across these modalities. The results generated by each modality are scaled using a predefined weight factor, after which the values from all three are combined. This algorithmic method has proven effective in real-time applications where speed of processing is critical. A practical implementation of all three modalities is described in [14].

In the first step, PLEA acted as an emotional mirror, simply mirroring the facial expressions of the person it interacts with. The mirroring operation can be performed by copying significant points from the face of the person in interaction to the face of the virtual agent. This is usually achieved by employing an approach based on FACS (Facial Action Coding System), which defines and identifies significant points on the person's face [15]. These points are aligned with the positions of muscles responsible for generating the person's facial expressions. In the second approach, applied to PLEA, multimodal channels are used to reason about current emotions and to generate emotional responses on the face of the virtual agent. This approach is employed in this work to facilitate the future development of a mechanism for the autonomous generation of facial expressions based on perceived emotions. The new computational model will be trained using a data corpus collected from real human-to-human interactions. The main objective of this approach is to advance autonomous human-to-agent interactions. The agent will be able to create and implement strategies to comfort or make someone happy based on the new mechanism.

The PLEA physical robot can be understood as a physical interface through which the PLEA virtual agent is embodied in the real, physical world. The PLEA robot consists of multiple components, including: a control computer, a microcontroller that contains three microphones which

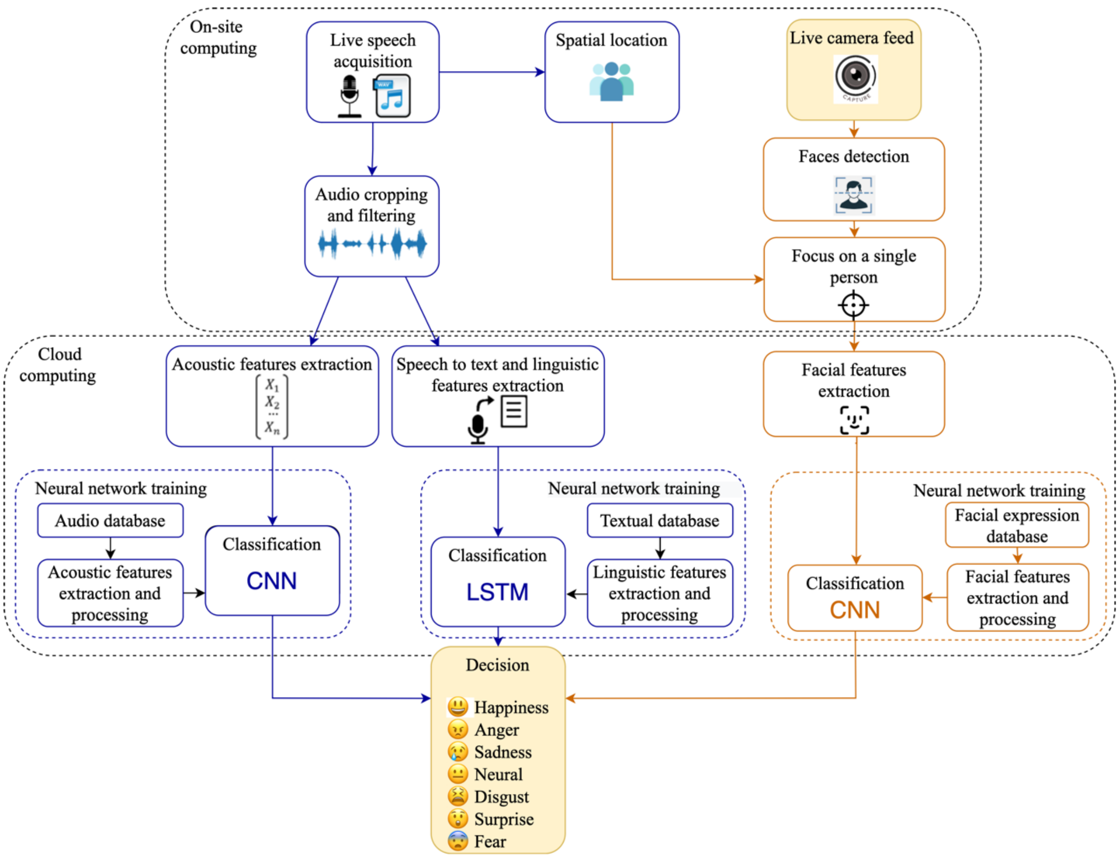

Fig. 2. Multimodal architecture of PLEA's reasoning model

enable the determination of spatial location for sound direction, an auxiliary power supply system for microcomputers used in case of a sudden interruption in electricity supply, a light projector, a base casing, the robot's neck, and the robot's head with a surface for projecting the face. The effect of embodiment is achieved through back-projected light. The light projector is placed within the neck part of the robot, and light projected onto the front face of the robot is used to display facial expressions in real time. The robot development process is shown at Figure 3. Figure 4 depicts different robot configurations: the first two figures show PLEA within the research lab, and the last two depict PLEA in Central Library at the Art & AI Festival in Leicester, the UK in 2022.

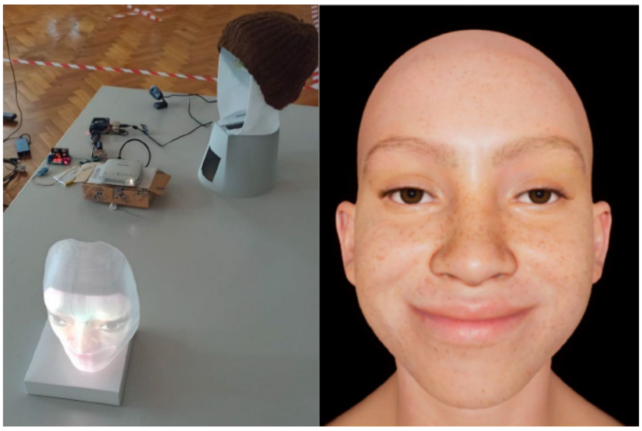

Fig. 3. Physical robot development process

Fig. 4. Different PLEA robot configurations

4. Data gathering and results

The cognitive model used to control the robot responses is based on data collected on several occasions. The first two events took place more than a year ago at the British Science Festival and the Art AI Festival in Leicester, UK (https://www.art-ai.io/programme/plea2/). The subsequent opportunity arises during the controlled experiments conducted at the Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia (FAMENA).

Attendees of the festivals had the freedom to walk up to the robot and engage in non-verbal communication. During these interactions a lot of data is collected to be used on building the new model for autonomous generation of facial expressions of the robot.

Examination of the data gathered through this method revealed that environmental factors significantly influence interactions with PLEA. As a result, a second set of data was obtained through interviews under more controlled conditions where participants were posed targeted questions regarding their experiences interacting with the robot. The conceptual framework for the interviews was inspired by the Media Equation Theory [16, 17, 18, 19, 20].

Data gathering under controlled conditions was conducted at FAMENA. During this time, 23 participants were interviewed over the span of a week. Each participant was asked to engage with PLEA for four minutes, within which they could spontaneously share emotional signals through facial expressions. This duration was chosen to allow participants to establish a certain degree of rapport with the robot. Following the interaction, participants responded to 12 questions posed by a human interviewer.

Figure 5 shows an example of a 24-second interaction in which joy and surprise alternate.

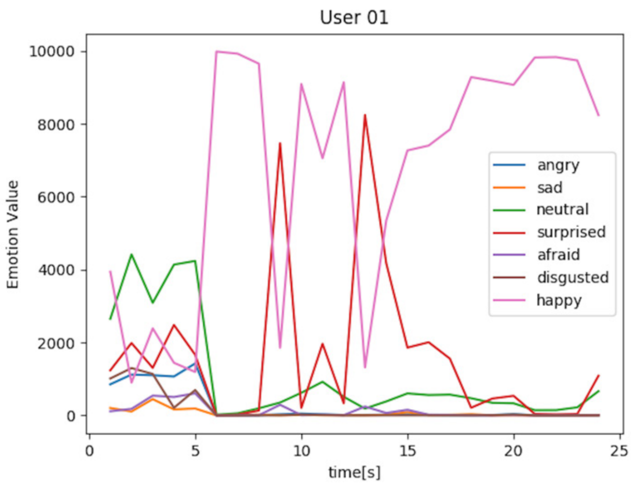

Fig. 5. PLEA emotion-interaction plot

The graph indicates that the person was initially confused. The first distinct emotion expressed by the person was surprise, to which PLEA responded accordingly. Following this, the person smiled upon noticing this facial expression, and PLEA mirrored the smile. Subsequently, the person expressed surprise once more. In the next phase, PLEA reacted with surprise, prompting the person to respond with a smile, indicating happiness. Afterwards, the person exhibited surprise for an extended period, which persisted until the conclusion of the interaction (non-verbal communication).

The interviews were then recorded and subsequently analysed. The group of participants was quite uniform, as the interactions were confined to a controlled environment within the research laboratory rather than a public setting, consisting of university students. All participants gave their consent for the use of their data for research purposes. Every exchange prompted a series of emotional reactions from the robot, which were reciprocated with the individual.

5. Discussion

Within this framework, our research method zeroes in on the ways in which individuals attribute human qualities to non-human entities. In the context of this investigation, treating PLEA with human attributes is termed as personification [21].

PLEA is equipped to engage with individuals wherever they are. This omnipresence allows for a seamless integration of PLEA into daily human activities, offering a new dimension of interaction that is both dynamic and responsive. According to the Media Equation Theory, this study of interactions with PLEA seeks to unravel the underlying reason - why do individuals behave in the manner we've observed [22]? Our findings indicate a natural inclination among people to attribute human emotions, intentions, and sentiments to inanimate objects, even with the understanding that these objects are not sentient. This phenomenon also seems to apply to PLEA, although the participants know that it consists of a plastic housing modelled on a human head.

6. Conclusions, and future work

The virtual being PLEA and the robot are at the interface between technology and human emotion. Designed to mimic and respond to human expressions through advanced algorithms and projected light configurations, PLEA exemplifies the fusion of Artificial Intelligence, Mixed Reality, and Smart Environment Design with empathetic interaction. The virtual PLEA agent is an inhabitant of cyberspace. It is also encapsulated in a plastic shell resembling a human head and serves as a focal point for studies on personification and emotional exchange, raising pivotal questions about human relationship with cognitive artifacts in a world where technology increasingly mirrors humanity.

Previous research suggests that users tend to ascribe human emotions, motives, and feelings to objects even when it is clear the object in question is not human. This is also the case with PLEA, which is strongly anthropomorphized by its design as a human head, albeit made of a plastic shell. PLEA as a ghost-like entity evoked the similar reactions. In this context, the methodological approach is focused on communicative practices, where people assign human characteristics to cognitive artefacts. This is defined here as personification, and the main research questions being explored are which features, characteristics, and functionalities are most appropriate in facilitating PLEA's personification for the use case scenario.

PLEA as a virtual being and the robot has great potential to be used in different areas of human activity, including healthcare, education, smart living, robotics, etc. The adaptability of PLEA allows it to cater to a broad spectrum of emotional and cognitive needs. In care settings, such a system could assist staff by offering additional support to residents, thereby enhancing the overall quality of care. By integrating PLEA into these environments, it is possible to create a more nurturing and responsive care ecosystem that prioritizes the emotional well-being of its inhabitants. Whether the aim is to provide company through conversation, remind the user of important tasks and medication or simply offer an emotionally attuned ear - the possible applications of PLEA are many and varied.

As a proof of concept, PLEA demonstrates a paradigm shift, moving away from sensed data towards contextual anticipation. Current multidisciplinary research is carried out with partners specializing in Creative Technologies and Informatics focused on human communication with artefacts. The development of the PLEA reasoning mechanism is aimed at autonomous agent responses that are customized to current and context-dependent expectations and needs. The use of PLEA in real-world applications will include further pilot projects in public spaces as well as in intelligent learning and healthcare environments.

Acknowledgments

This work has been supported in part by the Croatian Science Foundation under the project "Affective Multimodal Interaction based on Constructed Robot Cognition—AMICORC (UIP-2020-02-7184)".

7. References

[1] Montanha, R., Araujo, V., Knob, P., Pinho, G., Fonseca, G., Peres, V., & Musse, S. R. Crafting Realistic Virtual Humans: Unveiling Perspectives on Human Perception, Crowds, and Embodied Conversational Agents. In 2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (2023) 252-257.

[2] Stipancic, T.; Koren, L.; Korade, D.; Rosenberg, D. PLEA: A Social Robot with Teaching and Interacting Ca-pabilities. Journal of Pacific Rim Psychology (2021), 15, https://doi.org/10.1177/18344909211037019

[3] Ciria, A., Schillaci, G., Pezzulo, G., Hafner, V. V., & Lara, B. Predictive processing in cognitive robotics: a review. Neural Computation, 33(5), (2021) 1402-1432.

[4] Walker, M., Phung, T., Chakraborti, T., Williams, T., & Szafir, D. Virtual, augmented, and mixed reality for human-robot interaction: A survey and virtual design element taxonomy. ACM Transactions on Human-Robot Interaction, 12(4), (2023) 1-39.

[5] Koren, L., Stipančić, T., Ričko, A., & Benić, J. Context-Driven Method in Realization of Optimized Human-Robot Interaction. Tehnički glasnik, 16(3), (2022) 320-327.

[6] Koren, L., & Stipancic, T. Multimodal emotion analysis based on acoustic and linguistic features of the voice. In International Conference on Human-Computer Interaction (2021) 301-311. Cham: Springer International Publishing.

[7] Lyytinen, K., Nickerson, J. V., & King, J. L. Metahuman systems= humans+ machines that learn. Journal of Information Technology, 36(4), (2021) 427-445.

[8] McKim, J. Animation without animators: from motion capture to MetaHumans. Animation Studies 2.0. (2022)

[9] Zhang, Q., Yang, L. T., Chen, Z., & Li, P. A survey on deep learning for big data. Information Fusion, 42, (2018) 146-157.

[10] Koren, L., Stipancic, T., Ricko, A., & Orsag, L. Multimodal Emotion Analysis Based on Visual, Acoustic and Linguistic Features. In International Conference on Human-Computer Interaction (2022) 318-331. Cham: Springer International Publishing.

[11] A. V. Savchenko: Facial expression and attributes recognition basedon multi-task learning of lightweight neuralnetworks. In IEEE 19th International Symposium on Intelligent Systems and Informatics; 2021; Subotica. DOI: 10.1109/SISY52375.2021.9582508

[12] Abdul, Z. K., & Al-Talabani, A. K. Mel Frequency Cepstral Coefficient and its applications: A Review. IEEE Access. (2022)

[13] DiPietro, R., & Hager, G. D. Deep learning: RNNs and LSTM. In Handbook of medical image computing and computer assisted intervention (2020) 503-519. Academic Press.

[14] Koren, L., Stipancic, T., Ricko, A., & Orsag, L. Person localization model based on a fusion of acoustic and visual inputs. Electronics, 11(3), (2022) 440.

[15] Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., & Sobieranski, A. C.. A survey on facial emotion recognition techniques: A state-of-the-art literature review. Information Sciences, 582, (2022) 593-617.

[16] Littlejohn, Steven. Theories of Human Communication: Eleventh Edition. Waveland Press, Inc. (2016) 202. ISBN 978-1478634058.

[17] Reeves, B. & C.I. The media equation: How people treat computers, television, and new media like real people and places. Center for the Study of Language and Information. (1996) Cambridge University Press.

[18] Reuten, A., Van Dam, M., & Naber, M. Pupillary responses to robotic and human emotions: the uncanny valley and media equation confirmed. Frontiers in psychology, 9, (2018) 774.

[19] Klowait, N. The quest for appropriate models of human-likeness: anthropomorphism in media equation research. AI & SOCIETY, 33(4), (2018) 527-536.

[20] Fruchter, R., Nishida, T., & Rosenberg, D. Social Intelligence Design for Social Computing. In International Conference on Human-Computer Interaction (2022) 545-558. Springer, Cham.

[21] Mukaetova-Ladinska, B. E., Harwood, T., J., M.: Artificial intelligence in the healthcare of older people. Archives of Psychiatry and Mental Health (4) (2020). https://doi.org/10.29328/journal.apmh.1001011

[22] Littlejohn, S.W.: Theories of Human Communication, Eleventh edition. Waveland Press, Inc., Long Grove, Illinois (2017)