Embodied chatbots should have what non-verbal expressions, and how and when will they combine/alternate the use of verbal and non-verbal expressions?

Yao Yuanhang*
Peking University
2000017813@stu.pku.edu.cn

Abstract

Communication between people and between people and machines often does not limit to verbal expressions. Language alone cannot fully provide the information needed for communication, designing an embodied chatbot needs to give a series of non-verbal expressions. From facial expressions, gaze, gestures and postures to graphic animations, sounds and music, they are all effective ways to enhance communication. Moreover, our embodied robots often need to choose different communication methods according to different communication scenarios, communication objects, and communication purposes, such as only using verbal expressions, only using non-verbal expressions, and combining or alternating verbal and non-verbal expressions.

1 Introduction

According to the literature review by Van Pinxteren et al. [4], dialogue agents (chatbots, avatars, and robots) can use various communication behaviors to enhance relationship outcomes. These behaviors can be classified along two dimensions: modality (verbal, non-verbal, appearance) and footing (similarity, responsiveness). This essay mainly discusses verbal and non-verbal expressions and their corresponding application scenarios

2 Which non-verbal expressions should an embodied chatbot have?

"In this part, we will discuss one by one which non-verbal expressions an embodied chatbot needs to have and what are the characteristics and functions of these language expressions.

2.1 facial expressions

Firstly, an embodied chatbot should have facial expressions. Emoticons are a simple and effective non-verbal expression that can be used to convey emotions and intentions. For example, a smiling face can indicate happiness or satisfaction, a crying face can indicate sadness or disappointment, and a surprised face can indicate surprise or astonishment. In chatbots, emoticons can be used to enhance the user experience and make the conversation more lively and interesting. At the same time, considering that this is an embodied chatbot, I believe that a robot with appropriate facial expressions is more effective in overcoming the uncanny valley effect than a robot without any expressions, making the robot appear more intelligent and natural.

*Use footnote for providing further information about author (webpage, co-first authors, *etc*.).

2.2 gaze

Another important non-verbal expression is gaze. Gaze can convey a lot of information and emotions. For example, eye contact can express interest, curiosity, respect, challenge, threat, distrust, and other emotions. In social interactions, eye contact can indicate the other party's intentions and emotions, thereby enhancing the effect of communication. In addition, gaze can also be used to control the rhythm and direction of the conversation, for example, by using gaze to guide the other party's attention, or by using gaze to imply the other party's inappropriate behavior. Relevant research shows [1], the eyes represent different levels of signal value, which depends on the state, personality, and emotional state of the sender and receiver. Some areas in the brains of monkeys and humans contain neurons that are selective for facial, body, and gaze responses, and allowing our robots to have gaze functions can make them closer to the real level of human communication. The gaze function can also allow robots to infer shared attention in third-person social scene videos [2]. Shared attention refers to two or more people simultaneously looking at a common target in a social scene. In social activities and understanding of social scenes, perceiving and recognizing shared attention plays a crucial role.

2.3 gestures

Next, we need to mention non-verbal expressions, which are gestures. Gestures are an important non-verbal expression that can be used to convey emotions, intentions, and information. In chatbots, gestures can be used to enhance the user experience and make the conversation more lively and interesting. Gestures can be used to demonstrate operations and interactions, or to explain processes and steps. In chatbots, gestures can be used to explain and demonstrate complex concepts, making the conversation clearer and easier to understand. Gestures have multiple roles in interaction, including indication, signaling, scenario deduction, marking, and manipulation, etc. Gestures can be used to indicate objects or directions, convey meanings, demonstrate specific tasks, convey specific object or entity information, and guide movements, etc. There is even related research that believes that language originally appeared in the form of gestures, so allowing robots to have gesture functions has a huge effect on improving their communication level.

2.4 others

Lastly, let's mention some minor non-verbal expressions. Posture, sound and music, graphics and animations are all alternative non-verbal expressions. Posture is the use of the body as its conveyor, involving examples of attitudes, such as confidence or compliance. Graphics and animations are another form of non-verbal expression that can be used to convey complex information and concepts. For example, charts can be used to display data and trends, flowcharts can be used to explain processes and steps, and animations can be used to demonstrate operations and interactions. In chatbots, graphics and animations can be used to explain and demonstrate complex concepts, making the conversation clearer and easier to understand. Sound and music are a very useful non-verbal expression that can be used to convey emotions and atmosphere. For example, cheerful music can make people feel happy and excited, and soft sounds can make people feel relaxed and at ease. In chatbots, sound and music can be used to enhance the user experience and make the conversation more lively and interesting.

2.5 problem

I believe that embodied chatbots should generally have the above-mentioned non-verbal expressions. But there is still a point worth noting, that is, the content expressed by non-verbal expressions may be different in different cultures. When it comes to non-verbal expressions, differences between different cultures may lead to misunderstandings or confusion. Here are some examples of non-verbal expressions that may have different meanings in different cultures: Smile: In most cultures, a smile usually indicates happiness or pleasure. But in Russian culture, smiling is considered impolite, and in Asian culture, smiling can also be used to express pain and embarrassment. Eye contact: In Western culture, eye contact usually indicates confidence and honesty. But in Japanese culture, staring at someone's eyes is considered impolite, and in Indian culture, eye contact is usually seen as a provocation. Nodding: In most cultures, nodding usually indicates agreement or approval. But in Bulgarian culture, nodding usually indicates disagreement or opposition, and in Indonesian culture, nodding usually indicates understanding or acceptance. These are just a few examples, differences

between different cultures may lead to more misunderstandings or confusion. Therefore, chatbots need to pay attention to cultural differences in cross-cultural communication and avoid using non-verbal expressions that may be misunderstood as much as possible.

3 How and when will an embodied chatbot combine or alternate the use of verbal and non-verbal expressions?

Non-verbal expressions are an important way to enhance communication between robots and users, making communication more lively and emotional. However, how and when to combine the use of verbal and non-verbal expressions needs to be determined according to specific communication scenarios and user needs. First consider the user, if the chat object of the robot has a hearing impairment, then verbal expression may not be applicable, at this time gestures, images and other non-verbal expressions are needed to better communicate. A common application scenario for embodied chatbots is early childhood education. Human children of a younger age do not have fully developed language functions and cannot understand the complex concepts expressed by language. In this case, focusing on the action demonstration of real examples is more effective than focusing on the language teaching of abstract concepts. Taking chess as an example, the action demonstration can directly move the chess pieces, allowing the educated toddlers to have a direct impression of the rules and methods, while the verbal simulation demonstration is a non-symbolic communicative behavior, relying on a common understanding of the problem domain. Research shows [3], language teachers are good at conveying abstract information. However, if there are no object features, they find it challenging to convey the same information with specific objects. In contrast, demonstration teachers find it difficult to convey complex concepts.

However, the accuracy of verbal expression is still irreplaceable by non-verbal expression. As mentioned in the previous section, the same non-verbal expression may have different meanings in different cultures, and non-verbal expression may cause ambiguity to some extent, which requires the accuracy of verbal expression to eliminate this ambiguity. When it comes to the combination of verbal and non-verbal expressions, we need to give full play to the intuitive characteristics of non-verbal expressions and the accuracy of verbal expressions. The timing and frequency of non-verbal expressions depend on the context of the conversation and the behavior of the user. For example, if the user speaks quickly or uses a lot of technical terms, the robot may nod more frequently to indicate that it is following up. On the contrary, if the user speaks slowly or hesitates, the robot may use more facial expressions to express sympathy and encouragement. In combining verbal and non-verbal expressions, robots can use various techniques to create more natural and engaging conversations. For example, it can use non-verbal expressions to emphasize key points or provide additional context for its verbal responses. Or, it can use verbal expressions to clarify or expand its non-verbal cues. Moreover, while expressing the main information and content in language, non-verbal expressions can be used to emphasize emotions and convey emotions. The goal of incorporating non-verbal expressions into embodied chatbots is to create more intuitive and human-like interactions, which are both informative and attractive. However, the specific non-verbal expressions and verbal expressions that chatbots should use depend on the context and goals of the interaction. Regarding the combination of verbal and non-verbal expressions, it should be noted that the timing and frequency of these expressions will affect the effect of the interaction. In general, the choice of non-verbal expressions and their combination with verbal expressions should be based on the specific goals and context of the interaction, as well as the preferences and expectations of the user.

References

[1] Nathan J Emery. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6):581–604, 2000. URL ^9^. 2

[2] Lifeng Fan, Yixin Chen, Ping Wei, Wenguan Wang, and Song-Chun Zhu. Inferring shared attention in social scene videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6460–6468, 2018. URL ^1^. 2

[3] Theodore R Sumers, Mark K Ho, Robert D Hawkins, and Thomas L Griffiths. Show or tell? exploring when (and why) teaching with language outperforms demonstration. *Cognition*, 232: 105326, 2023. URL ^13^. 3

[4] Michelle ME Van Pinxteren, Mark Pluymaekers, and Jos GAM Lemmink. Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management*, 31(2):203–225, 2020. doi: 10.1108/JOSM-06-2019-0175. URL ^6^. 1