https://doi.org/10.31705/BPRM.v2(2).2022.4

Adapting the Concept of Human-Human Multimodal Interaction in Human-Robot Applications

Introduction

Human communication is multimodal in nature. In a normal environment, people use to interact with other humans and with the environment using more than one modality or medium of communication. They speak, use gestures and look at things to interact with nature and other humans. By listening to the different voice tones, looking at face gazes, and arm movements people understand communication cues. A discussion with two people will be in vocal communication, hand gestures, head gestures, and facial cues, etc. [1]. If textbook definition is considered synergistic use of these interaction methods is known as multimodal interaction [2]. For example, , a wheelchair user might instruct the smart wheelchair or the assistant to go forward, as shown in Fig. 1(a). However, with a hand gesture shown in the figure, he or she might want to go slowly. In the same way as of Fig. 1(b), a person might give someone a direction with a vocal command 'that way' and gesture the direction with his or her hand.

In most Human-Robot Interaction (HRI) developments, there is an assumption that human interactions are unimodal. This forces the researchers to ignore the information other modalities carry with them. Therefore, it would provide an additional dimension for interpretation of human robot interactions. This article provides a concise description of how to adapt the concept of multimodal interaction in human-robot applications.

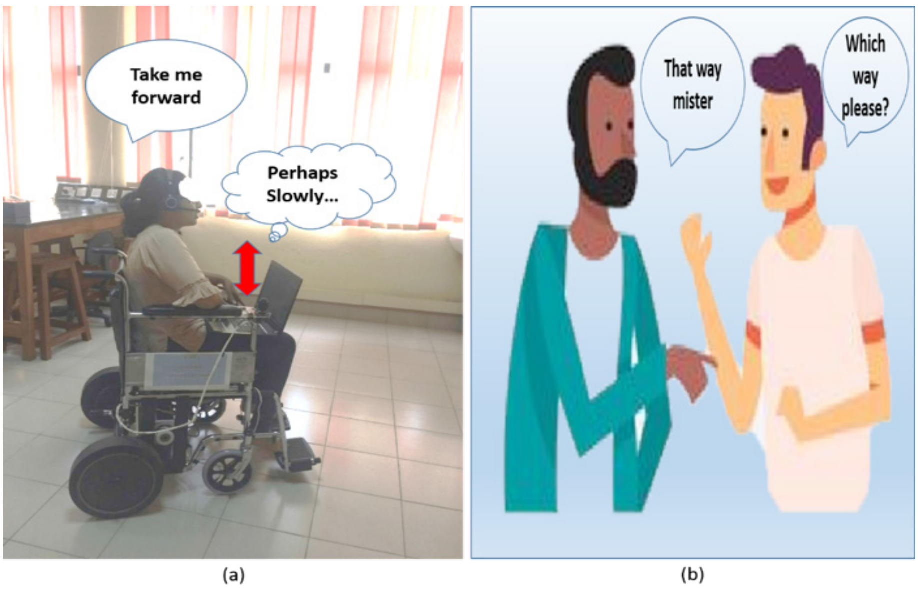

Figure 1: Example scenarios where multimodal interaction affect a human-robot encounter (a) represents an occasion where intelligent wheel-chair user use a command with speech and hand gestures (b) represents a general human-human communication where both uses their hand gestures as a way of communication

Why Multimodal Interaction?

Interaction among humans and animals tend to show multimodal behavior due to their biological build. Studies regarding superior colliculus have revealed that different senses are isolated initially at the neural level as shown in fig. 1.[3] When those senses reach the brain, sensory signals will be converged to the same target area in the superior colliculus. It also receives signals from the cerebral cortex and in turn modulates resultant behavior. Also studies shows that about 50% of neurons leaving the superior colliculus are multisensory as shown in Fig. 2.[4] These particulars encourage the multimodal interaction in HRI especially in assistive robots.

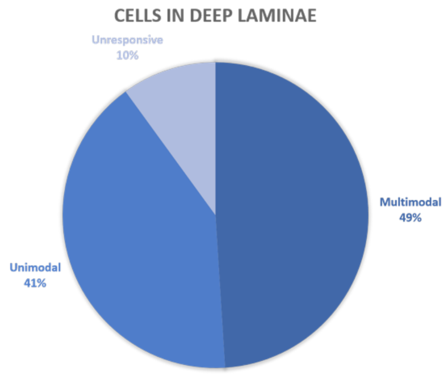

Figure 2: Percentages of unimodal, multimodal, and unresponsive cells present in the deep laminae of the superior colliculus

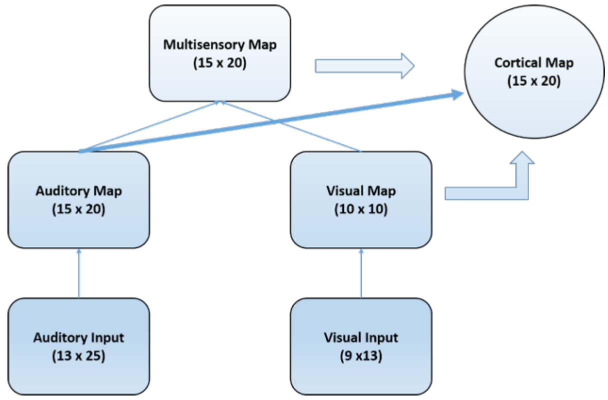

Fig. 3 Multisensory integration in Super colliculus model

The most convenient practical reason for multimodal interaction is the ineffectiveness of existing interaction devices and methods. When the systems and concepts in HRI getting advanced, using interaction devices like a joystick, keyboard, mouse, and touchpads will become impractical. Even if you develop an advanced and fast system, user interaction using a single modality can derail the system altogether. Another issue of the current advanced single-modality HRI is that it lacks robustness and accuracy. Statistical data analysis goes hand in hand with data fusion. One serious disadvantage of a single modality system is that it will not reduce the uncertainty associated with decision-making. Uncertainties occur when features are missing or when the sensors cannot measure all relevant attributes or when observations are ambiguous. Also, it is a proven fact that by using multiple observations from the same source, the statistical accuracy can be improved. [5]

Hand gesture information to interpret vocal uncertainties

Vocal commands or Speech is considered the primary modality among most humans. However, whenever a wheelchair user or any other person who uses an assistive robotic device gives out a navigational vocal command, these commands could contain uncertainties. These uncertainties are diverse and could affect the understanding of the command. There are some systems that have attempted to enhance the understanding of these uncertainties without considering another modality. However these systems are limited to few uncertainties and since general human-human communication contain multimodal commands, another modality could enhance the understanding of these uncertainties. Therefore

“Another issue of the current advanced single-modality HRI is that it lacks robustness and accuracy. Statistical data analysis goes hand in hand with data fusion.”

researchers have focused on using hand gestures for this purpose.

There are four major types of gestures associated with vocal commands

1. Deictic or pointing gestures
2. Iconic gestures
3. Metaphoric gestures
4. Beat gestures

1. Deictic Gestures

Deictic gestures are defined as the gestures which are performed to guide a recipient's attention towards a specific target or object or referent in the immediate or extended environment. The pointing gestures are one form of deictic gestures and often direct the recipient towards a direction or a specific object.

2. Iconic Gestures

Iconic gestures are the movements that represent meaning closer to the semantic content of the vocal phrases that they accompany. Therefore these types of gestures are sometimes termed redundant gestures.

3. Metaphoric Gestures

Metaphoric gestures are performed when a human being wanted to create a physical representation of an abstract idea or concept. These gestures provide additional semantic meaning that complements the ongoing verbal content.

4. Beat Gestures

Beat gestures are the gestures that do not add any informative meaning to the speech and these types of gestures aren't considered as a part of the verbal message. These gestures convey nonverbal meanings and are more in tune with the intonations of the speech.

Fusion of partial hand gesture information to interpret spatial uncertainties

Hand gestures carry partial information related to the vocal instruction. For example distance and direction related uncertainties present in navigation commands can be interpreted using these partial information. Deictic gestures which include pointing gestures carry information of direction and distance. Further gestures like iconic and metaphoric gestures carry additional semantic meanings of the same command which reduces the uncertainty. In addition the presence of beat gestures are synonymous with the vocal prosody which interprets the emotional state of the person and the effect of it on the meaning of the command. Intelligent Service Robotics Group (ISRG), Department of Electrical Engineering have conducted research on this and managed to enhance the understanding of spatial uncertainties in static and dynamic environments using partial hand gesture information. They are continuing on the research to explore the emotional effect of the hand gestures. Research have been implemented on the Intelligent Wheelchair Robot that they have developed.

References:

1. K. Priyanayana, A. B. P. Jayasekara, and R. Gopura, "Multimodal behavior analysis of human-robot navigational commands," in 2020 3rd International Conference on Control and Robots (ICCR). IEEE
2. Raisamo R (1999) Multimodal Human-Computer Interaction: a constructive and empirical study. Tampere University Press
3. 2020, pp. 79–84.Guerguiev J, Lillicrap TP, Richards BA (2017) Towards deep learning with segregated dendrites. ELife 6:e22901
4. Xu J, Yu L, Rowland BA, Stein BE (2017) The normal environment delays the development of multisensory integration. Scienti_c reports 7(1):4772
5. Thiran JP, Marques F, Bourlard H (2009) Multimodal Signal Processing: Theory and applications for human-computer interaction. Academic Press

Article by

Sahan Priyanayana[1], Buddhika Jayasekara[1], Ruwan Gopura[2]

Intelligent Service Robotics Group(ISRG)
[1]Department of Electrical Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka
[2]Department of Mechanical Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka