

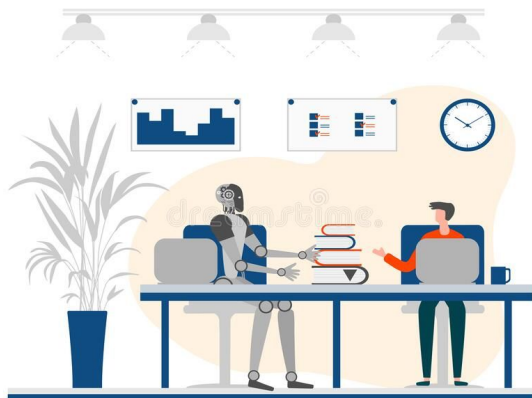


Detecting Interlocutor Confusion in Situating Human-Avatar Dialogue: A Pilot Study

Na Li, John D.Kelleher, Robert Ross
School of Computer Science
Technological University Dublin

Research Motivation

Human and Robot Interaction



**Confusion:
a unique mental
state**



**Confusion detection is
in online learning
system**

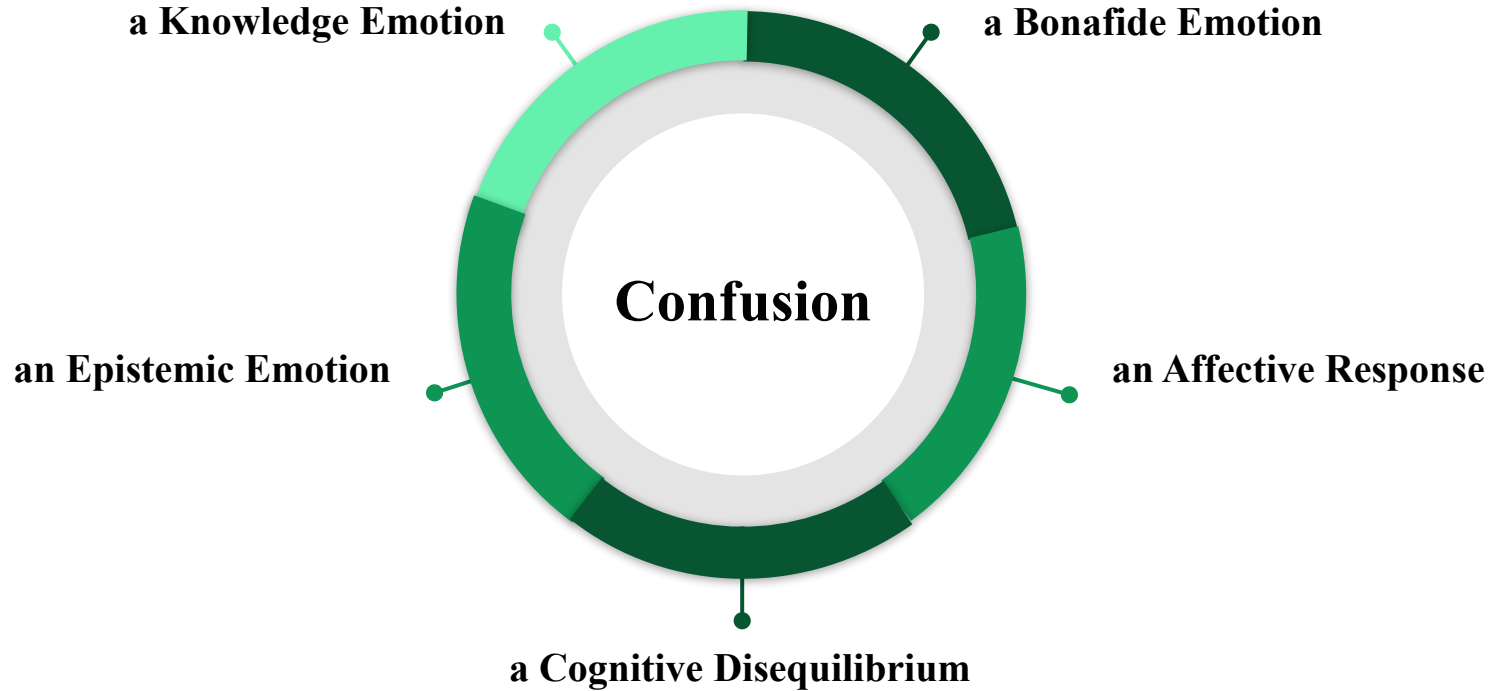


Research Question

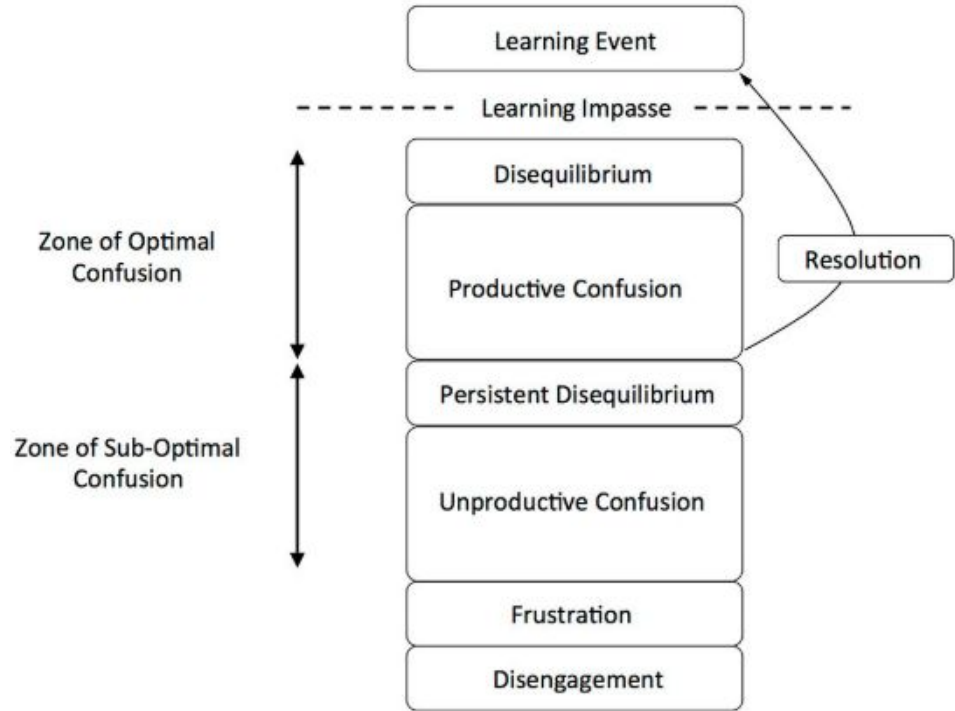
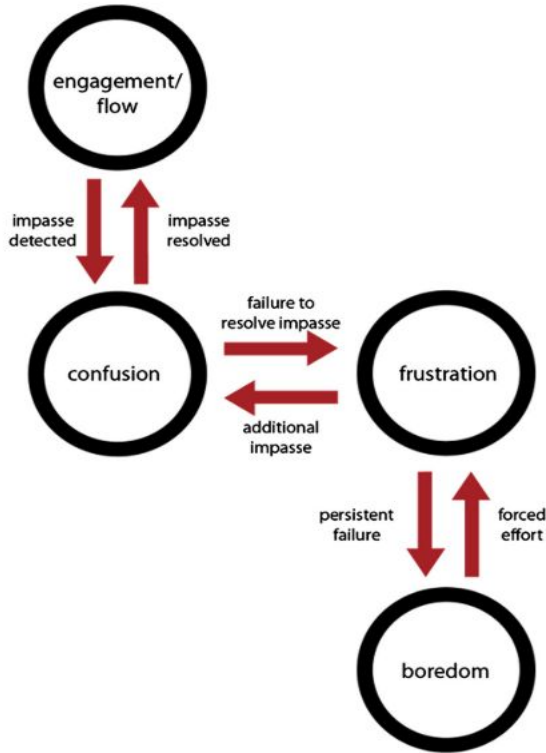
- Are participants **aware** they are confused if we give them a specific confusing situation?

- Do participants express **different** physical or verbal/non-verbal **behaviours** when they are confused that we can detect?

Confusion definitions in Fields



Conceptual Framework of Confusion



Observed Emotion Transition, S. D'Mello et. al (2014)

ZOSOC, Conceptual framework of ZOC and sub-optimal confusion, J. M. Lodge et. al (2014)

Definition of Confusion

Confusion is a **mental state** where under certain circumstances, a human experiences obstacles in the flow of interaction. A series of **behaviour responses** (which may be nonverbal, verbal, and, or non-linguistic vocal expression) may be **triggered**, and the human who is confused will typically want to **solve** the state of cognitive disequilibrium in a reasonable duration. However, if the confusion state is maintained **over a longer duration**, the interlocutor may become **frustrated**, or even **drop out** of the ongoing interaction.

Methods to Trigger Confusion in HRI



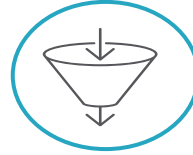
Complex Information



Insufficient Information



Contradictory Information

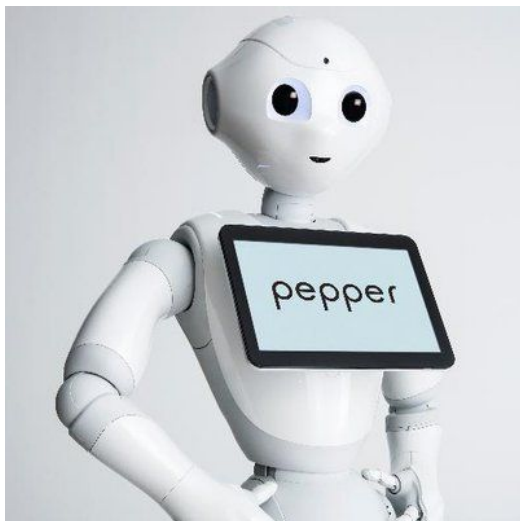


Inconsistent Information



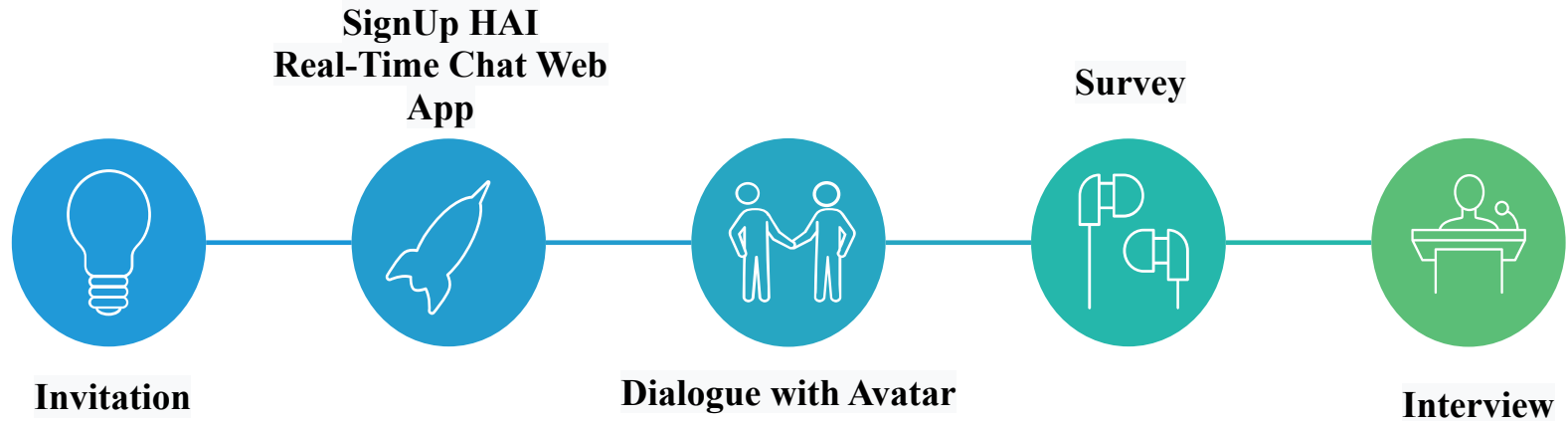
Feedback

Study Design



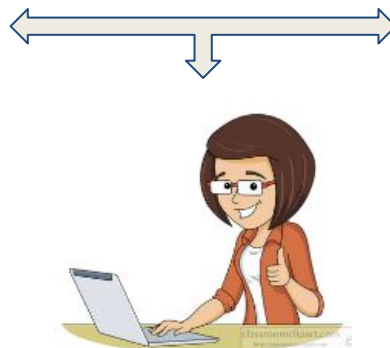
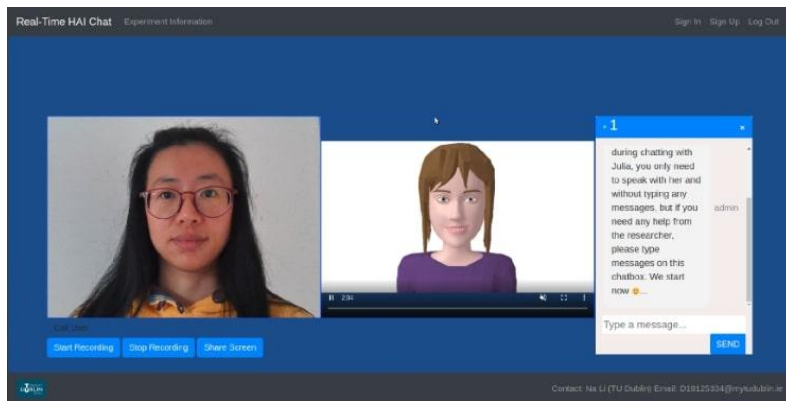
Sloan et. al (2020)

Process of Online Experiment

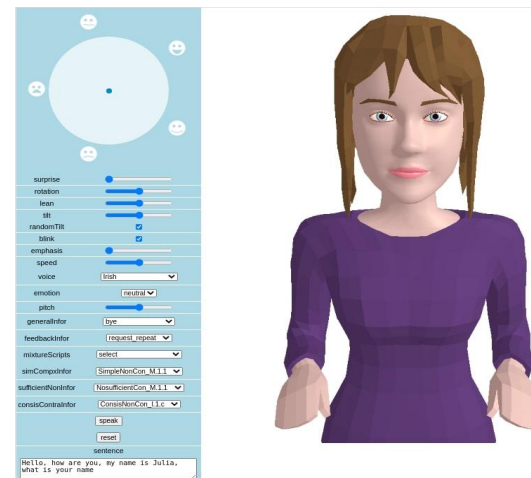


Study Approach

Client Side: HAI Real-Time Chat Web App



Researcher side: Avatar Web App



- Wizard-of-Oz experiment
- Semi-spontaneous one-to-one conversation
- < 15mins (5mins for task centric part)

Situated Dialogue Design

Definition of Tasks	
Task 1	a Simple logical problem
Task 2	a Word problem
Task 3	a Math question
Definition of Conditions	
Condition A	A task in a complex way to invoke confusion in the participant.
Condition B	A task in a straightforward way and should avoid confused states

Participant 1		
Stimulus	Task	Condition
1st	Task 1	A
2nd	Task 2	B
3rd	Task 3	A
Participant 2		
Stimulus	Task	Condition
1st	Task 1	B
2nd	Task 2	A
3rd	Task 3	B

An example of the experiment sequence for two separate study participants

Situated Dialogue Design

Patterns of Confusion with two Conditions	
Condition A	Condition B
Complex information	Simple information
Insufficient information	Sufficient information
Correct-negative feedback	Correct-positive feedback



Conversational Responses	Conversational Behaviours
1. Correct-positive feedback 2. Positive response	
1. Correct-negative feedback 2. Negative response	

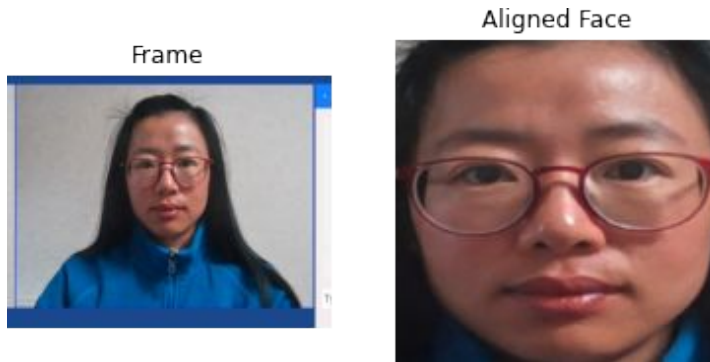
- Insufficient information in condition A: “*There are 66 people in the playground including 28 girls, boys and teachers. How many teachers were there in total?*”;
- Sufficient Information in condition B: “*There are 5 groups of 4 students, how many students are there in the class?*”.



Data Preparation

- Data collection: 23 participants in 6 countries, over 18 years of age from different colleges.
- Frame data: 19 participants' videos (8 males, 11 females) and labelling (ABA or BAB)

Condition	Frames
A	4084
B	3273



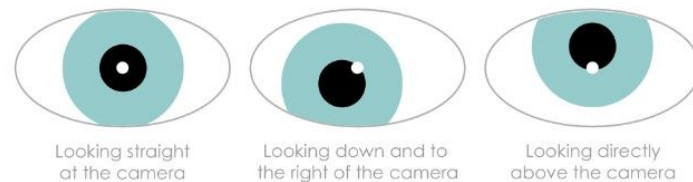
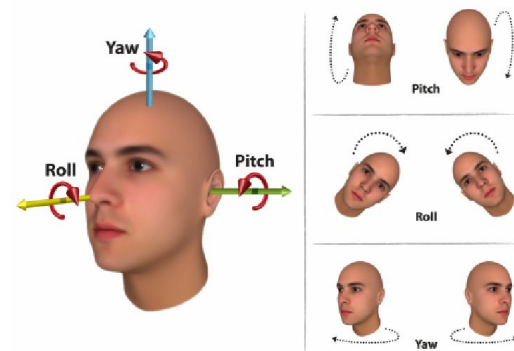
- Survey: 10 questions using a 5-level Likert scale
 - Each use survey has two conditions with three tasks (ABA or BAB)
 - Separate two sub-datasets by condition A and condition B
 - Calculate the average of confusion levels scores for three tasks

Data Analysis



Frame Data Measurement

- Emotion Detection (Savchenko, 2021)
- Head-pose estimation (Patacchiola and Cangelosi, 2017)
- Eye-gaze estimation (Zhang et al., 2020)



Data Analysis

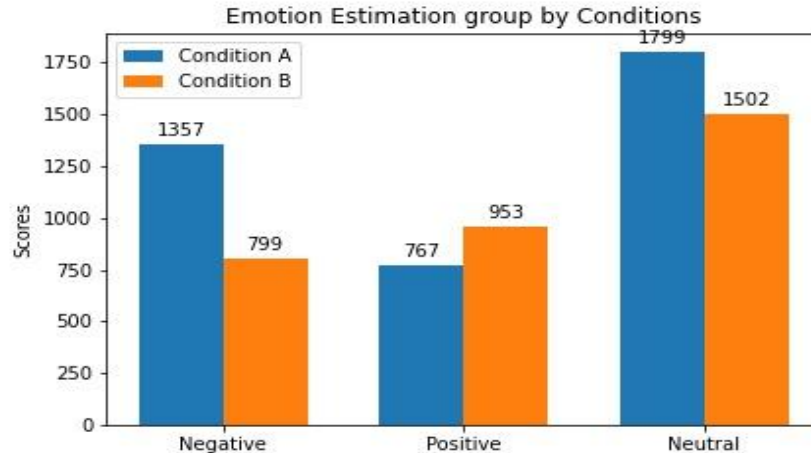


Frame Data Measurement



- Emotion Detection (Savchenko, 2021)

condition <chr>	neutral <int>	anger <int>	disgust <int>	fear <int>	sadness <int>	happiness <int>	surprise <int>	overall <int>
A	1799	262	282	136	677	702	65	3923
B	1502	77	165	57	480	858	95	3234



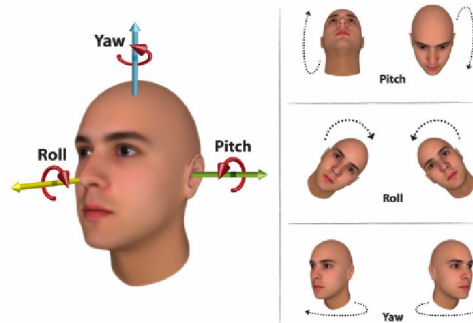
An independent-sample t-test is that there is a significant difference in the **three emotional categories** (negative, positive and neutral) and two conditions ($M = 0.77$, $SD = 0.94$ for condition A, $M = 0.48$, $SD = 0.60$ for condition B), $t(715) = 5.05$, $p - \text{value} < 0.05$.

Data Analysis



Frame Data Measure

- Head-pose estimation (Patacchiola and Cangelosi, 2017)
 - ➔ Calculated the sum of abs (pitch) and abs (yaw) and abs (roll) angles as an independent feature with conditions.
 - ➔ An independent-sample t-test: A significant difference in the sum of absolute values of these three angles and two conditions ($M = 21.96$, $SD = 9.46$ for condition A, $M = 27.40$, $SD = 12.21$ for condition B), $t(703) = -6.61$, $p\text{-value} < 0.05$.

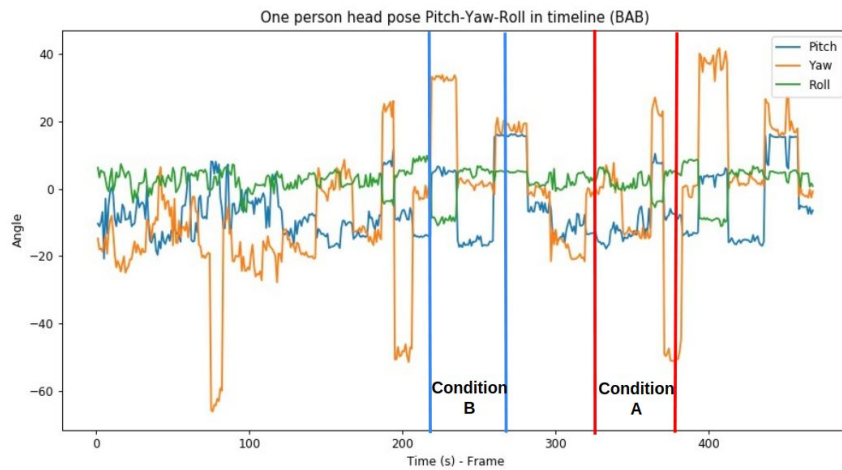
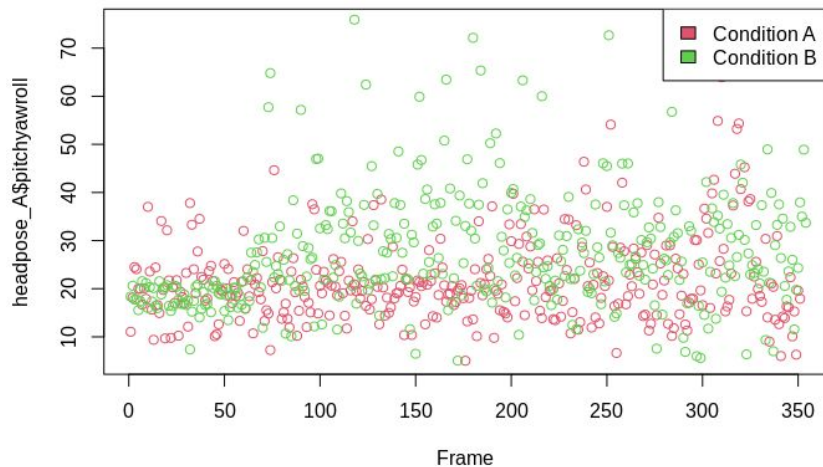
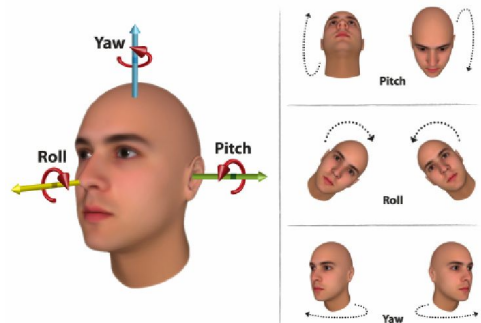


Data Analysis



Frame Data Measurement

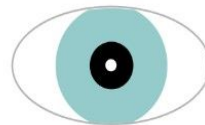
- Head-pose estimation (Patacchiola and Cangelosi, 2017)



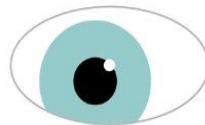
Data Analysis



Frame Data Measurement



Looking straight
at the camera



Looking down and to
the right of the camera



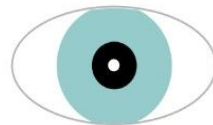
Looking directly
above the camera

- Eye-gaze estimation (Zhang et al., 2020)
 - Calculated the sum of $\text{abs}(\text{pitch})$ and $\text{abs}(\text{yaw})$ angles as an independent feature.
 - An independent-sample t-test: a significant difference in the sum of absolute values of pitch and yaw and two conditions was found ($M = 0.44$, $SD = 0.26$ for condition A, $M = 0.49$, $SD = 0.22$ for condition B), $t(728) = -2.58$, $p\text{-value} < 0.05$.

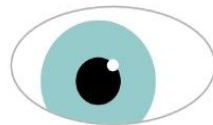
Data Analysis



Frame Data Measurement



Looking straight
at the camera

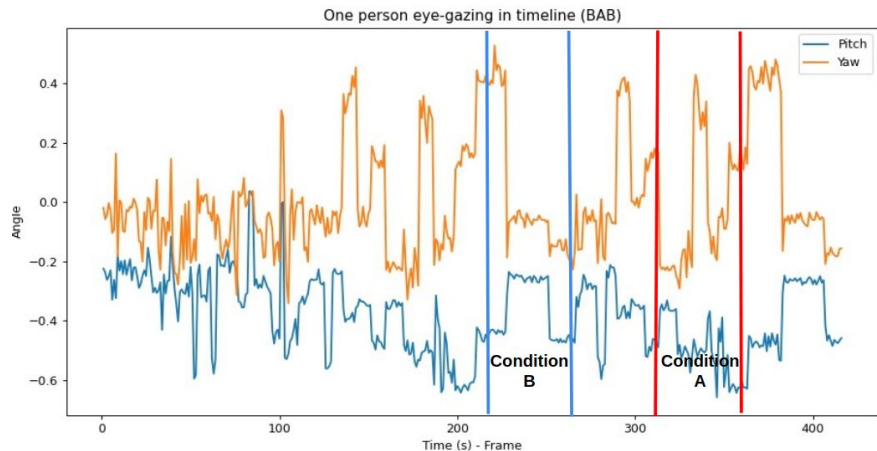
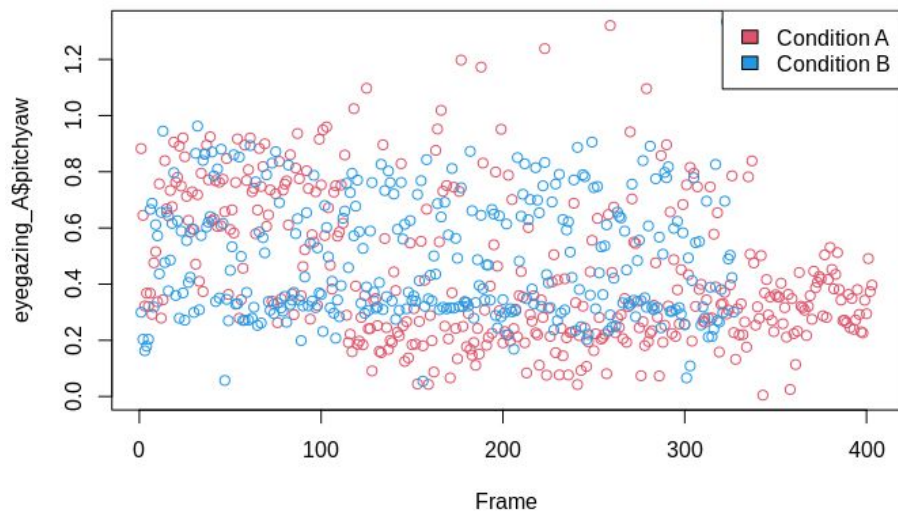


Looking down and to
the right of the camera



Looking directly
above the camera

- Eye-gaze estimation (Zhang et al., 2020)



Data Analysis



Subjective Measurement

Tasks	Results
Task 1: Logical problems and two conditions	There is no significant difference in the confusion scores for task 1 with two conditions was found ($M = 3.00$, $SD = 1.07$ for condition A, $M = 2.44$, $SD = 1.33$ for condition B), $t(15) = 0.94$, $p - \text{value} = 0.36$.
Task 2: Word problems and two conditions	There is no significant difference in the confusion scores for task 2 with two conditions was found ($M = 3.09$, $SD = 1.22$ for condition A, $M = 3.10$, $SD = 1.29$ for condition B), $t(19) = -0.02$, $p - \text{value} = 0.99$.
Task 3: Math problems and two conditions	There is a significant difference in the confusion scores for task 3 was found ($M = 4.38$, $SD = 0.74$ for condition A, $M = 3.00$, $SD = 1.12$ for condition B), $t(15) = 2.94$, $p - \text{value} < 0.05$.
Average scores of three tasks and two conditions	There is no significant difference between the average of confusion scores of the three tasks and two conditions ($M = 3.50$, $SD = 1.40$ for condition A, $M = 2.97$, $SD = 1.12$ for condition B), $t(36) = 1.28$, $p - \text{value} = 0.21$

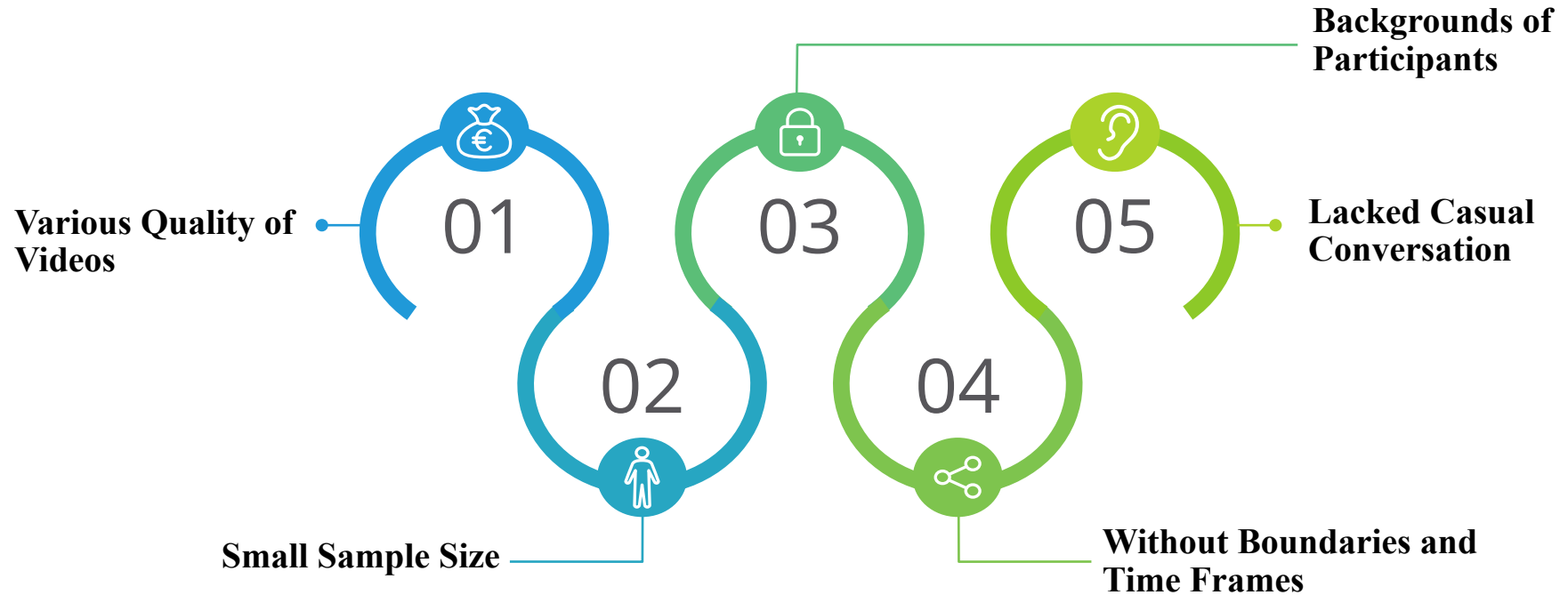
Discussion

--- Research Results

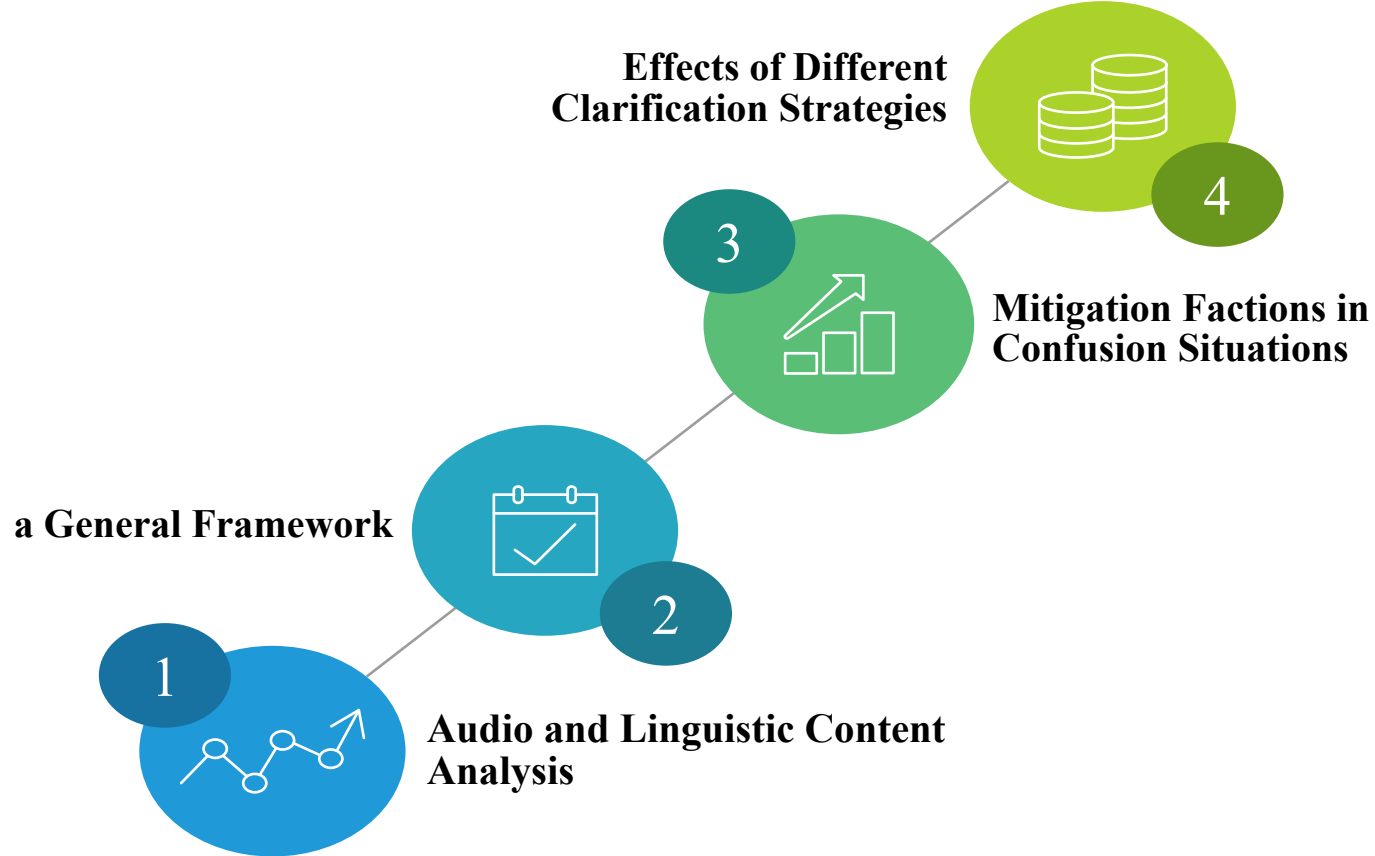
- Participants **are not always aware** they are confused if we gave them a specific confusing situation.
- When they are confused, their emotion is **more negative** than when they are not confused.
- When they are confused, the range of **angles of eye gazing is more than** when they are not confused.
- When they are confused, the range of the **angles of head shaking is less than** when they are not confused.

Discussion

--- Study Limitation



Further Study



Conclusion

- Confusion as an **important factor** in improving dialogue
- A new working **definition of confusion**
- Study to **induce confusion** in HAI
- No significant relationship between confusion scores and induced confusion states from **self-report**
- Significant relationship between **physical states** and induced confusion states
- A **crucial initial step** to build a computational model of confusion



References

A. Arguel and R. Lane, “Fostering deep understanding in geography by inducing and managing confusion: An online learning approach,” ASCILITE 2015 - Australasian Society for Computers in Learning and Tertiary Education, Conference Proceedings, no. November, pp. 374–378, 2019.

Andrey V. Savchenko. 2021. Facial expression and attributes recognition based on multi-task learning of lightweight neural networks. CoRR, abs/2103.17107.

B. Lehman, S. D’Mello, and A. Graesser, “Confusion and complex learning during interactions with computer learning environments,” Internet and Higher Education, vol. 15, no. 3, pp. 184–194, jun 2012.

D. Yang, R. E. Kraut, C. P. Rośe, and R. Rośe, “Exploring the Effect of Student Confusion in Massive Open Online Courses,” Tech. Rep. [Online]. Available: <http://www.katyjordan.com/MOOCproject.html>

J. M. Lodge, G. Kennedy, L. Lockyer, A. Arguel, and M. Pachman, “Understanding Difficulties and Resulting Confusion in Learning: An Integrative Review,” Frontiers in Education, vol. 3, 2018.

John Sloan, Daniel Maguire, and Julie Carson Berndsen. 2020. Emotional response language education for mobile devices. In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI ’20, New York, NY, USA. Association for Computing Machinery.


Lehman, Blair & D’Mello, Sidney & Graesser, Art. (2012). Confusion and Complex Learning during Interactions with Computer Learning Environments. The Internet and Higher Education. 15. 184-194. 10.1016/j.iheduc.2012.01.002.

Massimiliano Patacchiola and Angelo Cangelosi. 2017. Head pose estimation in the wild using convolutional neural networks and adaptive gradient methods. Pattern Recognition, 71:132–143

O. Celiktutan, S. Profile, H. Gunes, and E. Skordos, “Multimodal Human-Human-Robot Interactions (MHHRI) Dataset for Studying Personality and Engagement Intelligent,” 2017. [Online]. Available: <https://www.researchgate.net/publication/320179510>

S. D’Mello, B. Lehman, R. Pekrun, and A. Graesser, “Confusion can be beneficial for Learning,” Learning and Instruction, vol. 29, pp. 153–170, feb 2014.

Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. 2020. Ethxgaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In European Conference on Computer Vision (ECCV).

A background image showing a white robotic hand shaking a human hand. The text 'Thank you Any Questions' is overlaid in the center.

Thank you Any Questions

na.li@tudublin.ie

ACKNOWLEDGEMENT

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

