Supplementary Materials

Chi-Hsuan Wu[1], Shih-yang Liu[1], Xijie Huang[1], Xingbo Wang[1], Rong Zhang[1], Luca Minciullo[2]
Wong Kai Yiu[2], Kenny Kwan[2], Kwang-Ting Cheng[1]
[1]Hong Kong University of Science and Technology, [2]LifeHikes
{cwuau, sliuau, xhuangbs, xingbo.wang, rzhangab} @connect.ust.hk
{luca.minciullo, tim.wong, kenny.kwan} @lifehikes.com, timcheng@ust.hk

1. Behaviors and their Engagement Scores

We invited two psychology experts to suggest and evaluate the importance of certain engagement-related behaviors. A lower score from the expert means that the behavior may indicate disengaging. In contrast, a higher score means the behavior may be engaging. The detail is listed in Table 1.

Feature	Expert1	Expert2	Type
Arms crossed	4	7	Body
Consistent pose	5	6	Body
Changing seating position	6	6	Body
Slouching	3	4	Body
Sudden behavior change	8	5	Body
Yawning	3	5	Body
Back from breaking room	8	10	Facial
Speaking	6	10	Facial
Smile	7	7	Facial
Active hand movements	8	7	Hand
Hand at the back of head	3	4	Hand
Drinking or eating	5	6	Hand
Gesture+Speaking	7	7	Hand
Playing hands	3	5	Hand
Hand on mouth (thinking)	6	8	Hand
Hand stretching	5	6	Hand
Modify glasses	None	5	Hand
Moving closer to screen	8	9	Head
Nodding	9	10	Head
Head tilting towards screen	4	6	Head
Looking down	4	4	Gaze
Blank stare	3	4	Gaze
Eye rolling	4	5	Gaze
Focus on other objects	3	5	Gaze
Focus on a point on screen	5	7	Gaze
Consistent gaze direction	5	6	Gaze

Table 1. Different behaviors and their received scores from psychology experts.

2. Input Features and their corresponding shape

In this section, we list out the shape of all the input features of the model. The detail is listed in Table 2.

Feature	Shape
Gaze	$(B \times 8)$
Head Pose	$(B \times 6)$
Facial Action Units	$(B \times 35)$
I3D vector	$(B \times 1024)$
Speech	String
Acoustics	Volume: $(B \times n)$ Pitch: $(B \times m)$ n,m depends on speech length

Table 2. Format of different extracted features