eHRI Database

The Engagement in Human-Robot Interaction (eHRI) database contains natural interactions between two human participants and a robot under a story-shaping game scenario. The audio-visual recordings provided with the database are fully annotated at 5-intensity scale for head nods and smiles as well as with speech transcription and continuous engagement values. The database includes 24 video clips from recordings of 12 distinct groups of participants. Total duration is 142 minutes.

The eHRI database will be available soon for academic purposes. You may contact eerzin@ku.edu.tr to access the released database.

Elements of the eHRI Database (to be released)

Folders and Files:

- **Annotations:** Folder containing annotation files of visual smiles, head nods, adjacency, mutual gaze and engagement.
    - a_headnod: Annotated head nods.
    - a_visualsmile: Annotated visual smiles.
    - c_engagement : Computed engagement.
    - r_adj : Recognized adjacent pair.
    - r_gaze: Recognized mutual gaze.
- **Transcription:** Folder containing transcription files of videos.
- **Videos:** Folder containing videos.
- **sessions.csv :** File containing information on sessions.

Descriptions and Format:

- **Annotations:** Files are in csv format.
    - a_headnod and a_visualsmile: 2 people annotated head nods and visual smiles's event start and end time. Then different 3 people ranked these events with 1-5 intensity scale. At the beginning of the videos there are warm up dialogs, game start times are also included in the files. Game start is not ranked by annotator. It has 1 as default value.

Format:

event_name	start_time	end_time	annotator_1	annotator_2	annotator_3

- r_adj and r_gaze: These events are not annotated, recognized by the system. Gaze is recognized by head location, adjacency pair is from participants answering time. Their definition and recognition method is explained in detail in the paper.

Format:

event_name	event_time

- c_engagement**:** Engagement is calculated with 4 connection events; annotated head nods, visual smiles, and recognized adjacency pair, gaze looking past 15 second at a 4Hz rate. In the files, their changed time and values are shown.

Format:

changed time	engagement_values

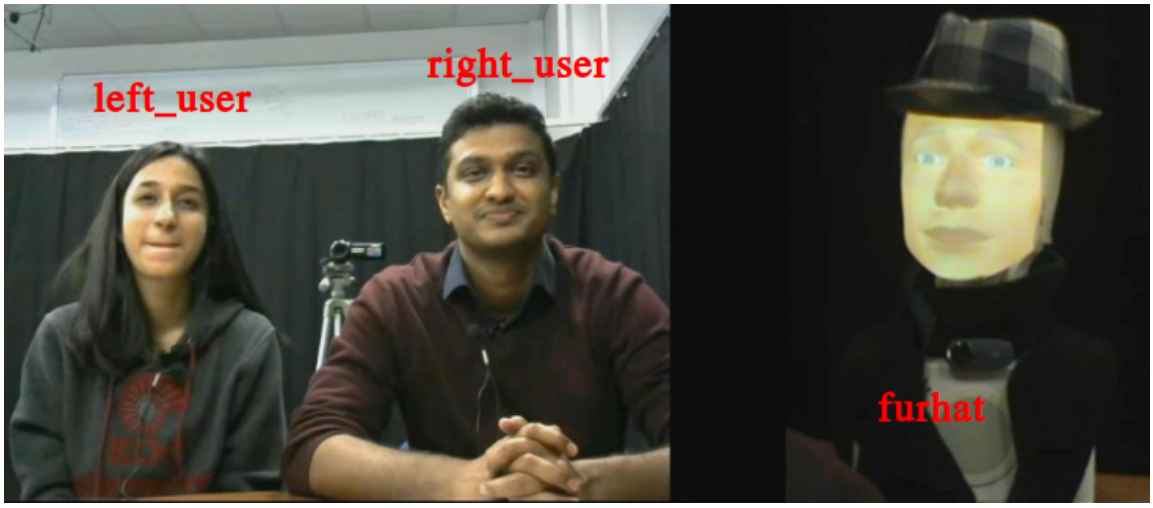
- **Transcription:** Files are in srt (subtitle format). Can be imported with media players and annotation software.

Format example:

1 (number of transcription)
00:00:00,030 --> 00:00:01,080 (start time --> end time)
furhat: Hi there! Who am I seeing today? (name of the person*: transcription)
2
00:00:01,680 --> 00:00:03,360
My name is Furhat. What is your name?
* Robot's name is given as furhat. Person sitting on the left in the videos is given as left_user person sitting on the right is given as right_user. Name of the person is **only given** when another person takes the turn. If a person says multiple sentences back to back the name will not be given in the following transcription.

- **Videos:** Files are given in .mp4 format. Video names are given as video_(video number)_(policy name).mp4 . There are two policies which are learning based policy (LBP) and rule based policy (RBP). These policies determine the backchannel generation of the robot. In each session only one policy is used.

Details can be found in the paper. Each clip contains recordings of two users and the robot. An example frame from the videos is shown below.



Important note: In video_04_rbp after 04:02 robot's recording is not given due to a camera problem. However it is available up to 04:02. In video_12_rbp after 02:49 synchronization of audio and video is lost due to frame drops. The time difference between audio and video increases until the end of the video.

- **sessions.csv :** There are three columns in this file. First column consists of the names of each video. Second column gives information whether a session includes warm up or not. If a session includes a warm up phase it is indicated as 1 if not it is indicated as 0. Last column indicates the type of policy used in that session. Learning based policy is given as lbp, rule based policy is given as rbp in the file.

Related Publications:

Ege Kesim, Tugce Numanoglu, Oyku Bayramoglu, Bekir Berker Turker, Nusrah Hussain, Metin Sezgin, Yucel Yemez and Engin Erzin, "The eHRI database: {A} multimodal database of engagement in human-robot interactions," submitted for publication.

N. Hussain, E. Erzin, T. M. Sezgin and Y. Yemez, "Training Socially Engaging Robots: Modeling Backchannel Behaviors with Batch Reinforcement Learning," in IEEE Transactions on Affective Computing, 2022, doi: 10.1109/TAFFC.2022.3190233.