Overview

This folder contains the data obtained from 10 sessions of public speaking tasks (PRE, TEST01,...,TEST08, POST), arranged according to sessions. There is another folder named 'misc' that contains topics of public speaking tasks and VR scenarios used in the study. The contents of the session folders are-

1. Actiwave: This folder contains participants' ECG and heart rate (HR) data obtained from actiwave sensor in .csv format. It includes data for relaxation (_RELAX), preparation (_PREP), and presentation (_PPT) periods. Data for relaxation and preparation contains only the signal value, while data for the presentation period contains the corresponding timestamps, which are synchronized to the audio signals.

2. Annotation: This folder contains the time-continuous ratings of stress labeled by 4 annotators (R1, R2, R4, R5) and their aggregated gold standard (Fused) rating. Files are in .csv format where the first column refers to timestamp at 1 second step size, and the remaining columns refer to the ratings from the annotators and the fused ratings. Naming convention of the file is <Session_Name>_<Participant_id>_annotation.csv. Timestamps are synchronized with audio.

3. Audio: This folder contains the audio files of presentation speech in .wav format. Naming convention of the file is <Session_Name>_<Participant_id>_AUDIO_PPT.wav

4. E4: This folder contains participants' physiological signals obtained from E4 sensor in .csv format. It includes data for relaxation (_RELAX), preparation (_PREP), and presentation (_PPT) periods. Data for relaxation and preparation contains only the signal value, while data for the presentation period contains the corresponding timestamps, which are synchronized to the audio signals. For each period the following data are included: 3-axis acceleration (ACC), blood volume pulse (BVP), electrodermal activity (EDA), heart rate (HR), interbeat interval (IBI), and wrist temperature (TEMP).

5. Features: This folder contains the summative features (i.e., over the entire session) extracted from different modalities. Features from E4 sensors, audio and actiwave are extracted using Ledalab, OpenSMILE and BioSPPY libraries, respectively. Features are computed separately for the relaxation, preparation, and presentation periods.

6. Self_Reports : This folder contains the self-reported public speaking anxiety of participants, obtained by conducting various surveys before and after the session. The aggregated scores of the surveys are placed here in .csv format.

7. participant_id.csv : This file contains the participant id (PID) of all the participants that completed the PRE session.

8. README.txt : This file contains information about participants ids with missing data.

The 'train_pid_annotation.csv' and 'test_pid_annotation.csv' files are for participant independent train-test split for continuous stress detection when using the ratings from the annotation folder.

Additional details about individual folders are provided in the following sections.

Actiwave

This folder contains the participants' actiwave sensor data, organized in folders which are named according to PIDs. Data for relaxation (_RELAX), preparation (_PREP), and presentation (_PPT) are available in PRE, TEST01, TEST05, and POST. Only data for preparation (_PREP) and presentation (_PPT) are available in the remaining folders. Each folder may contain-

1. ECG_RELAX.csv: This contains the ECG value (in Volts) during relaxation period, sampled at 512 Hz. There is one column in the file, whose header is ECG.
2. ECG_PREP.csv: This contains the ECG value (in Volts) during preparation period, sampled at 512 Hz. There is one column in the file, whose header is ECG.
3. ECG_PPT.csv: This contains the ECG value (in Volts) during the presentation, sampled at 512 Hz, and corresponding timestamps. There are two columns in the file, whose headers are Time (s) and ECG. Timestamps are synchronized with audio.

4. HR_RELAX.csv: This contains the heart rate (in bpm) calculated from ECG during relaxation period, sampled at 1 Hz. There is one column in the file, whose header is HR.
5. HR_PREP.csv: This contains the heart rate (in bpm) calculated from ECG during preparation period, sampled at 1 Hz. There is one column in the file, whose header is HR.
6. HR_PPT.csv: This contains the heart rate (in bpm) calculated from ECG during presentation, sampled at 1 Hz, and corresponding timestamps. There are two columns in the file, whose headers are Time (s) and HR. Timestamps are synchronized with audio.

Annotation

This folder contains the time-continuous ratings of stress labeled by 4 annotators (R1, R2, R4, R5) and their aggregated gold standard (Fused) rating for each public speaking session. Files are in .csv format where the first column refers to timestamp at 1 second step size, and the remaining columns refer to the ratings from the annotators and the fused ratings. Naming convention of the file is <Session_Name>_<Participant_id>_annotation.csv. Ratings were collected on a likert scale from 1 (No Stress) to 5 (Extremely High Stress) and then they were converted to [0, 0.25, 0.5, 0.75, 1] scale. These ratings were aggregated into the Fused ratings for each session that can be used as labels for continuous stress detection. Timestamps are synchronized with audio.

Audio

This folder contains the audio files of presentation speech in .wav format. Speech is sampled at 16 kHz. Naming convention of the file is <Session_Name>_<Participant_id>_AUDIO_PPT.wav

E4

This folder contains the participants' Empatica E4 sensor data, organized in folders that are named according to PIDs. Data for relaxation (_RELAX), preparation (_PREP), and presentation (_PPT) are available in PRE, TEST01, TEST05, and POST. Only data for preparation (_PREP) and presentation (_PPT) are available in the remaining folders. Each folder may contain-

1. ACC_RELAX.csv: This contains the data from a 3-axis accelerometer (in (1/64)g) sensor during relaxation period, sampled at 32 Hz. There are 3 columns in the file, whose headers are x, y, and z.
2. ACC_PREP.csv: This contains the data from a 3-axis accelerometer (in (1/64)g) sensor during the preparation period, sampled at 32 Hz. There are 3 columns in the file, whose headers are x, y, and z.
3. ACC_PPT.csv: This contains the data from 3-axis accelerometer (in (1/64)g) sensor during preparation period, sampled at 32 Hz, and corresponding timestamps. There are 4 columns in the file, whose headers are Time (s), x, y, and z. Timestamps are synchronized with audio.

4. BVP_RELAX.csv: This contains the blood volume pulse data from a PPG sensor during the relaxation period, sampled at 64 Hz. There is 1 column in the file, whose header is BVP.
5. BVP_PREP.csv: This contains the blood volume pulse data from PPG sensor during preparation period, sampled at 64 Hz. There is 1 column in the file, whose header is BVP.
6. BVP_PPT.csv: This contains the blood volume pulse data from PPG sensor during presentation, sampled at 64 Hz, and corresponding timestamps. There are 2 columns in the file, whose headers are Time (s), and BVP. Timestamps are synchronized with audio.

7. EDA_RELAX.csv: This contains the electrodermal activity data (in µS) from GSR sensor during relaxation period, sampled at 4 Hz. There is 1 column in the file, whose header is EDA.
8. EDA_PREP.csv: This contains the electrodermal activity data (in µS) from GSR sensor during preparation period, sampled at 4 Hz. There is 1 column in the file, whose header is EDA.
9. EDA_PPT.csv: This contains the electrodermal activity data (in µS) from GSR sensor during presentation, sampled at 4 Hz, and corresponding timestamps. There are 2 columns in the file, whose headers are Time (s), and EDA. Timestamps are synchronized with audio.

10. HR_RELAX.csv: This contains the heart rate (in bpm) calculated from PPG sensor during relaxation period, sampled at 1 Hz. There is one column in the file, whose header is HR.
11. HR_PREP.csv: This contains the heart rate (in bpm) calculated from PPG sensor during preparation period, sampled at 1 Hz. There is one column in the file, whose header is HR.
12. HR_PPT.csv: This contains the heart rate (in bpm) calculated from PPG sensor during presentation, sampled at 1 Hz, and corresponding timestamps. There are two columns in the file, whose headers are Time (s) and HR. Timestamps are synchronized with audio.

13. IBI_RELAX.csv: This contains the Inter-Beat Interval (in seconds) data from PPG sensor during relaxation period, and the corresponding timestamp. There are 2 columns in the file, whose headers are Time (s) and IBI. Timestamps are synchronized with audio.

14. IBI_PREP.csv: This contains the Inter-Beat Interval (in seconds) data from PPG sensor during preparation period, and the corresponding timestamp. There are 2 columns in the file, whose headers are Time (s) and IBI. Timestamps are synchronized with audio.
15. IBI_PPT.csv: This contains the Inter-Beat Interval (in seconds) data from PPG sensor during presentation, and the corresponding timestamp. There are 2 columns in the file, whose headers are Time (s) and IBI. Timestamps are synchronized with audio.

16. TEMP_RELAX.csv: This contains the body temperature (in Celsius) during relaxation period, sampled at 4 Hz. There is 1 column in the file, whose header is TEMP.
17. TEMP_PREP.csv: This contains the body temperature (in Celsius) during preparation period, sampled at 4 Hz. There is 1 column in the file, whose header is TEMP.
18. TEMP_PPT.csv: This contains the body temperature (in Celsius) during presentation, sampled at 4 Hz, and corresponding timestamps. There are 2 columns in the file, whose headers are Time (s), and TEMP. Timestamps are synchronized with audio.

Features

This folder contains the summative features and related information for different modalities. The contents of this folder are-

1. vad_timing: This folder contains the start and end time of voice activity segments for each audio. Files are in .txt format, where each line contains two numbers; the start and end time of VAD segment. These time points are extracted using OpenSMILE library.
2. vad_features: This folder contains two folders-IS2009 and IS2010 that contains the Interspeech 2009 and 2010 feature set for individual VAD segments for each audio. OpenSMILE is used for feature extraction.
3. IS09.csv: This file contains the aggregate Interspeech 2009 features for all audios.
4. IS10.csv: This file contains the aggregate Interspeech 2010 features for all audios.
5. Pause_features.csv: This file contains the pause related features for all audios. Features include number of pause, mean pause frequency, mean pause interval, mean, and percent.
6. E4_measures.csv: This file contains the mean signal values obtained from E4 for all available participants. Along with that, Skin Conductance Response (SCR) amplitude and frequency are provided. Ledalab is used for this purpose.
7. td_feature.csv: This file contains the aggregated time-domain features obtained from ECG signals for each participant. BioSPPY, pyHRV, and hrv-analysis libraries are used for this purpose.
8. fd_feature.csv: This file contains the aggregated frequency-domain features obtained from ECG signals for each participant. Previously mentioned libraries are used for this purpose.
9. Mean_HR.csv: This file contains the mean heart rate obtained from actiwave sensor for all participants.

Self_Reports

This folder may contain the following files

1. <Session_Name>_beforePPT.csv: This contains the aggregated score of several self-reported trait anxiety ratings for each participant. These surveys were conducted at the beginning of the session. This is available in PRE and POST only.

2. <Session_Name>_afterPPT.csv: This contains the aggregated score of several self-reported state anxiety ratings for each participant. These surveys were conducted at the end of the session. This is available for all sessions.

3. <Session_Name>_dailyexp.csv: This contains participants' response to a survey about their experience on that day. This is available in PRE, TEST01, and POST only.

4. <Session_Name>_demographics.csv: This contains participants' response to a survey about their demographics. This is available in PRE only. A readme_demographics.txt is associated with it.

5. <Session_Name>_BFI.csv: This contains the aggregated scores of self-reported personality scores collected using Big Five inventory. This is available only in TEST01.

6. <Session_Name>_DayStart.csv: This contains the aggregated score of several self-reported trait anxiety ratings for each participant. These surveys were conducted at the beginning of the day of TEST sessions. This is available in TEST01 and TEST05 only.

7. <Session_Name>_DayEnd.csv: This contains the aggregated score of several self-reported trait anxiety ratings for each participant. These surveys were conducted at the end of the day of TEST sessions. This is available in TEST04 and TEST08 only.

What's new in the updated version?

1. Time continuous ratings of stress are added in Annotation folder that can be used for continuous stress detection. Train-test split information for this use case is also provided.

2. All features and signals are in .csv format instead of .xlsx format.

3. Physiological signals for the relaxation and preparation time are provided for TEST sessions.

4. Actiwave features in the Features folders are updated.