

mOral Dataset: A dataset for inferring Oral Hygiene Behaviors in-the-wild using wrist-worn inertial sensors

Updated version of this file can be found in the following link:

https://docs.google.com/document/d/1QhoBCyCofl-Tl0NzUzWt_AKWOf-Rmr79sJcS69Py_-Q/e/dit?usp=sharing

Publication: [mORAL: An mHealth Model for Inferring Oral Hygiene Behaviors in-the-wild Using Wrist-worn Inertial Sensors](https://dl.acm.org/citation.cfm?id=3314388)

(<https://dl.acm.org/citation.cfm?id=3314388>)

Summary

- Large scale: 3,886 hours of sensor data were collected over 290 days from 25 users.
- Everyday devices: sensors from smartphone and smartwatch.
- In-the-Wild: data was collected from users that were engaged in their regular natural behavior continuously for more than 7 days for each user.
- Publicly available: everyone is invited to download the dataset for free and use it (conditioned on citing our original paper -- linked below).

This dataset was collected in 2017-2019 at the MD2K Center, University of Memphis:

- Sayma Akther
- Dr. Nazir Saleheen
- Shahin Samiei
- Dr. Vivek Shetty
- Dr. Emre Ertin
- Dr. Santosh Kumar

Dataset description

The original dataset that used for the mORAL paper contains data from 25 participants (12 males, 13 females; mean age 28.5 ± 7.6 years), each identified with a random identifier, who completed one-week-long brushing and flossing study. Data from three out of 25 participants are excluded from this released dataset since they did not consent to release their data.

Devices and Sensors

- MotionSense ("wristband") that included 3-axis accelerometers sampled at 16 Hz and 3-axis gyroscopes sampled at 32 Hz. Data were collected via a wireless Bluetooth connection to a study smartphone platform.
- Samsung Galaxy S5 smartphone platform
- Bluetooth-enabled Oral-B toothbrush ("SmartBrush")

Data collection protocol

The study team conducted a field study to collect labeled sensor data in participants' natural environments. The study was approved by the Institutional Review Board (IRB) and all participants provided written informed consent. At the time of recruitment, potential study participants were screened for their brushing habits (at least twice daily) and flossing habits (at least once daily). Once enrolled, participants wore the wristbands during their waking hours for seven (7) days. Participants were asked to use the study smartphone daily to record videos of themselves when performing their typical oral health routines (e.g., toothbrushing or flossing) in a comfortable setting (e.g., at home), while wearing the wristband. Non identifiable sensor data was periodically uploaded from the smartphone platform via a dedicated HTTPS connection to a secure server.

Video data collection

The videos provided the ground truth labels for toothbrushing and flossing events. The phone stored these data on an encrypted microSD card. Video data were recovered from the phone at the end of the study period.

Annotation of Oral Hygiene Behaviors (OHB) from Video Data

To create labeled data for model development and evaluation, we annotated the video data to record the timing of each OHB. In total, 362 videos were collected, each with an average duration of 3.12 minutes. We annotated the brushing type (normal or SmartBrush), flossing type (string or pick), and their corresponding start/end times from the videos.

For brushing, special attention was paid to the interval between the lowering and raising of the brush-holding hand from the mouth. This interval was marked as a pause. We also marked the orientation configuration of both wrists, the brushing wrist (left or right), use of manual or SmartBrush, flossing with string or picks, flossing wrist (left or right), and video pause times.

Sensor data and annotation files

Directory Structure:

Directory Structure

data/	annotation/	Configurations.csv		
		9a6b_20170420_000900.csv		
		9a6b_20170420_092852.csv		
		...		
		be71_20170514_231429.csv		
	sensor_data/	9a6b (User id)	d2017-04-19 (day)	Sensor data: left-wrist-accelx.csv left-wrist-accely.csv left-wrist-accelz.csv left-wrist-gyrox.csv left-wrist-gyroy.csv left-wrist-gyroz.csv right-wrist-accelx.csv right-wrist-accely.csv right-wrist-accelz.csv right-wrist-gyrox.csv right-wrist-gyroy.csv right-wrist-gyroz.csv
			d2017-04-20	
			...	
			...	
			Be71	
	features_and_Mlresults/		...	
	code/			

Accelerometer Data:

Data filenames:

- right-wrist-accelx.csv
- right-wrist-accely.csv
- right-wrist-accelz.csv
- left-wrist-accelx.csv
- left-wrist-accely.csv
- left-wrist-accelz.csv

Example:

timestamp	value
1505585435721	0.001220703125

Column Details:

1. **timestamp (UTC):** Coordinated Universal Time indicating the number of milliseconds since January 1, 1970 (Unix time)

- Unit: milliseconds
- Type: long
- 2. **value**: Acceleration MotionSense accelerometer in the particular axis
 - Unit: g (G-forces)
 - Type: double
 - Sampling frequency: 16 Hz
 - Range: -2g to +2g

Gyroscope Data

Data filenames:

- right-wrist-gyrox.csv
- right-wrist-gyroy.csv
- right-wrist-gyroz.csv
- left-wrist-gyrox.csv
- left-wrist-gyroy.csv
- left-wrist-gyroz.csv

Example:

timestamp	value
1505585435721	47.30224609375

Column Details:

1. **timestamp (UTC)**: Coordinated Universal Time indicating the number of milliseconds since January 1, 1970 (Unix time)
 - Unit: milliseconds
 - Type: long
2. **value**: the rate at which the MotionSense device rotates around the x-axis
 - Unit: degree/second
 - Type: double
 - Sampling frequency: 32 Hz
 - Range: -250 to +250

Configuration file (Configuration.csv)

One single file contains metadata information of each recorded video

Example:

video_file_name	pid	wrist	is_manual_brushing	is_smart_brushing	is_flossing	flossing_with	is_rinsing	video_duration	is_video_paused	left_orientation	right_orientation	is_new_device
820c_20170318_111735	820c	R	1	0	0	0	0	1:51	0	0	0	0

Column Details:

- **Video_file_name:** For any video file one row is generated in the form of 'uuuu_YYMMDD_HHmmSS', here 'uuuu' is the four-character user id and YYMMDD_HHmmSS is the video start time.
 - Format: uuuu_YYMMDD_HHmmSS
 - Type: String
- **Pid: Randomly generated four character participant id**
 - Format: four character
 - Type: String
- **wrist:** brushing or flossing wrist
 - Type: String
 - Values: L (Left), R (Right), or B (Both)
- **is_manual_brushing:** True/False
 - Type: Boolean
- **is_smart_brushing:** True/False
 - Type: Boolean
- **is_flossing:** True/False
 - Type: Boolean
- **flossing_with:**
 - Type: String
 - Values:
 - String
 - Picks
 - 0 (for not flossing)
- **is_rinsing:** True/False
 - Type: Boolean
- **Video_duration:** Values are available only when manual brushing or string flossing
 - Format: m:ss
- **left_orientation:** Orientation of the left wrist.
 - Type: int

- values: 0, 1, 2, and 3 as defined in the paper.
- **right_orientation**: Orientation of the right wrist.
 - Type: int
 - values: 0, 1, 2, and 3 as defined in the paper.
- **is_new_device**:
 - Type: int
 - values:
 - 0 (Bottom device as shown in Figure 3(a) of the mORAL paper)
 - 1 (Top device as shown in Figure 3(a) of the mORAL paper)

Video Annotation

For each video one csv file is created for annotation. Filename is in the form of 'uuuu_YYMMDD_HHmmSS' which can be used to link with the configuration file.

Example:

start_offset(sec)	end_offset(sec)	start_timestamp	end_timestamp	label
0	46	1531581838000	1531581884000	manual_brushing

Column Details:

- **start_offset(sec)**: Time offset of the start of the event since the video start in seconds
 - Type: Integer
 - Unit: seconds
- **end_offset(sec)**: Time offset of the end of the event since the video start in seconds
 - Type: Integer
 - Unit: seconds
- **start_timestamp (UTC)**: Timing of the first video frame when the participant began a defined activity (name of the activity is found in 'label' column)
 - Unit: milliseconds
 - Type: long
- **end_timestamp (UTC)**: Timing of the last video frame when the participant ends the activity
 - Unit: milliseconds
 - Type: long
- **label**: name of the activity. Possible values (as defined in the paper):

- Values
 - manual_brushing
 - smart_brushing
 - string_flossing
 - pick_flossing
 - pause
- Type: String

Code:

Latest code can be found in the following github link:

<https://github.com/sakther/mORAL>

main file:

MAIN_train_and_test_mORAL.ipynb