**Comparisons of datasets**

Head gesture duration MEAN / MEDIAN [seconds]

(L = listener, S = speaker)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **nod** | **shake** | **tilt** |
| NVB | 1.03 / 0.88 | 1.91 / 1.60 | 0.97 / 0.86 |
| SEWA | 2.24 / 1.80 | 1.71 / 1.27 |  |
| CCDb (both) | 1.41 / 1.16 | 1.25 / 1.12 | 0.94 / 0.82 |
| Hatice2010 | 1.27 / 1.18 | 1.06 / 1.03 |  |
| FIPCO |  |  |  |
| VRA1 06/07  (L nods, consolidated)  (all nods & S tilts, w/o consolidation) | **1.76 / 1.41**  1.59 / 1.18 |  | 0.91 / 0.77 |

Overall mean and median: Mean: 1.5469605263157895 Median: 1.18

# occurences of head gestures

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **nod** | **shake** | **tilt** | **other** |
| NVB | 493 | 280 | 373 | ? |
| SEWA | 352 | 142 | 0 | ? |
| CCDb (both) | 306 | 191 | 83 | ? |
| Hatice2010 | 154 | 104 | 0 | 25 |
| FIPCO | 1563 +879tick | 60 + 605turn | 477 | ? |
| VRA1 06/07  (nods of L)  (nods, tilts of S) | 681 | **443**  259 | 0  0 | 0  73 |  |

Total duration of head gestures in datasets (second)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **nod** | **shake** | **tilt** | **other** |
| NVB | 508 | 533 | 360 | ? |
| SEWA | 788 | 242 | 0 | ? |
| CCDb (both) | 432 | 238 | 78 | ? |
| Hatice2010 | 196 | 110 | 0 | 141 |
| FIPCO | ~ 5500 (7000) | ~ (2000) | ~ 1800 | ? |
| VRA1 06/07  (nods of L)  (nods, tilts of S) | 1119 | **780**  371 | 0  0 | 0  67 |  |

Comparison

[all spontaneous gestures / conversations]

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Type** | **Modalities** |  |
| NVB | Answer questions | Video, audio |  |
| SEWA | Conversation | Video, audio? |  |
| CCDb (both) | Conversation | Video, audio |  |
| Hatice2010 | Conversation? | Video, audio |  |
| FIPCO | Conversation | Video, audio, eye-gaze |  |
| VRA1 06/07  (L)  (S) | listener  speaker | Video, audio  Video, audio |  |

IEMOCAP

* human-2-human
* <https://sail.usc.edu/iemocap/iemocap_info.htm>
* English
* 10 Actors = 5 + 5
* **Modalities:** Motion Capture Face Information, Speech, Videos, Head Movement and Head Angle Information, Dialog Transcriptions, Word level, Syllable level and Phoneme level alignment
* **Annotations:** segmented into utterances - annotated by at least 3 human annotators with Categorical Attributes: anger, happiness, excitement, sadness, frustration, fear, surprise, other and neutral state and Dimensional Attributes: valence, activation, dominance
* TOTAL: 12 hours
* But probably these are not useful annotations?

SEMAINE

* VA-2-human / VA-operator-2-human
* 240 conversations, 80 fully annotated & transcribed
* 5 min
* **Annotations:** 
  + Turn taking & transcripts
  + Head nods, shake
  + <https://semaine-db.eu/pages/about/>

Cardiff Conversation Database (CCDb) [ASKED FOR ACCESS, GOT]

* human-2-human
* <http://www.cs.cf.ac.uk/CCDb/>
* 30 conversations
* 5 min
* Audio-visual
* 16 speakers = 12+4
* TOTAL: 2x150 min = 5 hours
* Manual **annotations** (performed using ELAN) for 8 conversations (40min) consisting of:
  + Backchannel
  + Frontchannel
  + Agree
  + Disagree
  + Utterance (verbal and non-verbal)
  + Happy (smile or laugh)
  + Surprise
  + Thinking
  + Confusion
  + Head Nodding
  + Head Shake
  + Head Tilt
  + Transcripts as well

FIPCO [ASKED FOR ACCESS]

* 15 hours,
* 20 participants
* 1st person or far camera
* Per-frame annotated
  + 11 gesture categories
  + None (background class), Nod (head up and down), Jerk (head down and up), Up (pitch up for a while), Down (pitch down for a while), Tick (repeated nods), Tilt, Shake, Turn, Forward (lean forward), Backward (lean backward).
* Bigger than CCDb

NOMCO corpus [ASKED FOR ACCESS]

* <https://link.springer.com/content/pdf/10.1007/s10579-016-9371-6.pdf>
* [http://sskkii.gu.se/nomco/examples.php#](http://sskkii.gu.se/nomco/examples.php)
* Danish
* acquaintance dialogues
* 12 conversations, 1h total … small
* Whole body
* Counts of gestures
  + Nod746
  + Tilt496
  + SideTurn437
  + HeadForward357
  + Shake337
  + HeadBackward264
  + HeadOther212
  + Jerk180
  + Waggle88
* TOO FAR, AND VARIABLE POSES, WHOLE UPPER BODY
* But got the OpenFace data from Patrizia Paggio, Senior Researcher, University of Copenhagen, Centre for Language Technology, [paggio@hum.ku.dk](mailto:paggio@hum.ku.dk)

BUHMAP DB

* <https://www.cmpe.boun.edu.tr/pilab/pilabfiles/databases/buhmap/>
* 3 classes useful => ⅜ \* 440 \* 2 sec = 5 min in total!
* Not spontaneous

Mimicry DB [ASKED FOR ACCESS, GOT]

* <https://mahnob-db.eu/mimicry/>
* <https://ibug.doc.ic.ac.uk/media/uploads/documents/sun2011multimodal.pdf>
* human-2-human
* 40 participants = 28+12
* 54 recordings (discussion on a political topic, and a role-playing game)
  + 1 recording has videos of one subject only and without FaceFar2
* 10-15 min per recording => 9 hours
* **annotations** for many different phenomena, including dialogue acts, turn-taking, affect, head gestures, hand gestures, body movements and facial expressions.
* Audio signals - separate sources
* Videos “Front”
* Use **FaceFar2**
* Experiment A:
  + Presentation
  + Discussion
* Experiment B
  + Conversation - find a roommate
* **Annotations**
  + Of speaker / listener => **can use for turn-taking**
  + Of behav. cues
  + Of mimicry
  + behavioural expression labels:
    - smile, head nod, headshake, body leaning away, and body leaning forward
    - Conscious: with DESIRED GOAL
      * - e.g. to flatter others, to emphasize understanding, to express agreement, to share rapport/empathy, to increase acceptance.
    - Unconscious: with SOCIAL SIGNAL EXPRESSION
      * - e.g. understanding, agreement, liking, confused, or uncertain

**MultiLis Corpus [**suggested by LP Morency**]**

* “The MultiLis Corpus – Dealing with Individual Differences in Nonverbal Listening Behavior” <https://ris.utwente.nl/ws/portalfiles/portal/5370313/dekok_2010_multilis.pdf>
* 1 speaker - 3 listeners
* One-way! Not a conversation! [2 tasks: retell a video, give instructions for cooking recipe]
  + Not spontaneous conversation
* In Dutch
* 8 x 4 = 32 sessions
* 32 subjects
* 393 min of listener data; 131 min of speaker data
* Annotations
  + Personality - big 5
  + Speaker
    - Eye gaze
      * binary
    - Mouth movements
  + Listener
    - Head
      * 12 cat
    - Eyebrows
    - Mouth movements
    - Speech
      * “Huh”, “okay”, “ja”

Lixing data [looking at their files]

* What is composed video?
* What are fruits videos? (all in root directory)
* /videos/
  + 46 videos
  + 2-3 min
  + AV of 1 human giving backchannels
  + Questions asked by virtual human (female/male)
  + Not dialogue
  + disclosure
* /tmp/
  + 7 videos
  + 5 min
  + Noisy conversations
  + Very silent
* RapportSpeakerVideos
  + 47 videos WMV
  + 2-3 min
  + Story telling
  + 1 AV of human speaker
  + ANNOTATIONS where to make backchannels by listener
* /Sound/
  + VA sounds and associated BML
* Sinhwa’s data?
* /AAMAS/ conference?
* /rapport\_evaluate/
  + Side-by-side videos - evaluation, backchannel moments highlighted by color border
  + Human speaker + virtual human rapporting
* Rapport Data Collection
  + 45 videos MP4
  + Like in RapportSpeakerVideos
* Rapport
  + Dyad
    - 16 Story tellings
    - 2-3 min
    - Separate AV of speaker and listener
  + Mp4
    - 45 FLV videos
    - Story
    - AV of speaker
  + Other
    - 1 AV story teller
    - 24 videos
    - 2-3 min
* Per\_question\_turn\_taking\_time
  + Legend?
* Multiculture sounds? And videos of speakers, story
* Multiculture audio single
  + Videos of speakers, story
* Multiculture aligned video
  + Videos of listener and speaker
  + Stories
  + L @ left audio channel, S @ right audio channel
  + 174/2=87 conversations
* Marshall
  + Fd
* Keyboard head nod label
  + Fd

[Rapport.ict.usc.edu](http://rapport.ict.usc.edu/)