The KTH talking head in space - a vehicle for situated multi-party interaction



Björn Granström-CTT, KTH



Speech synthesis developments at KTH





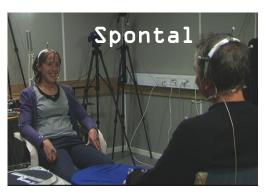
Talking Heads Models and Applications for Multimodal Speech Synthesis



Jonas Beskow Doctoral Dissertation Stockholm 2003

KTH interactional data collections

Human-human



Simulated human-computer





Human-computer





Collection of audio-visual databases: interactive spontaneous dialogues

Eliciting technique: information seeking scenario

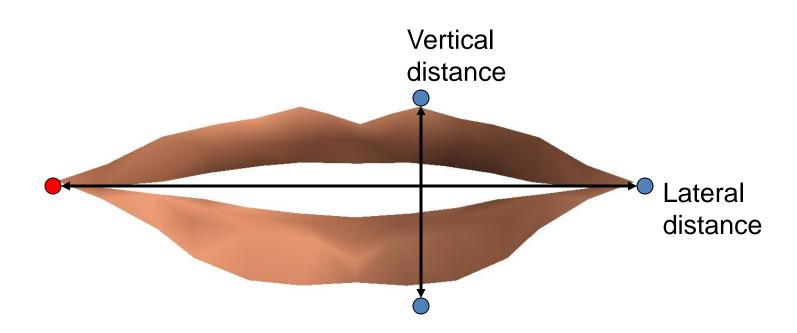
Focus on the speaker who has the role of information giver

The speaker seats facing 4 infrared cameras, a digital video-camera, a microphone The other person is only video recorded.



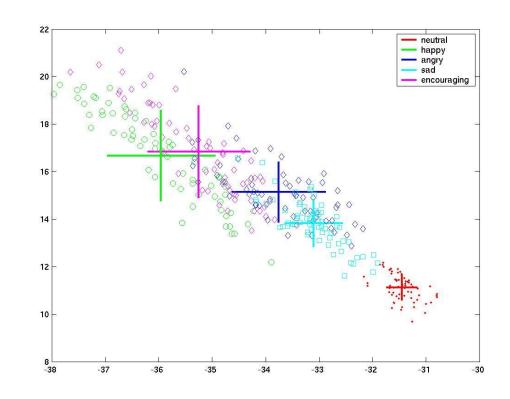


Measurement points for lip coarticulation analysis



The expressive mouth

- All vowels (sentences)
 - Encouraging
 - Happy
 - Angry
 - Sad
 - Neutral

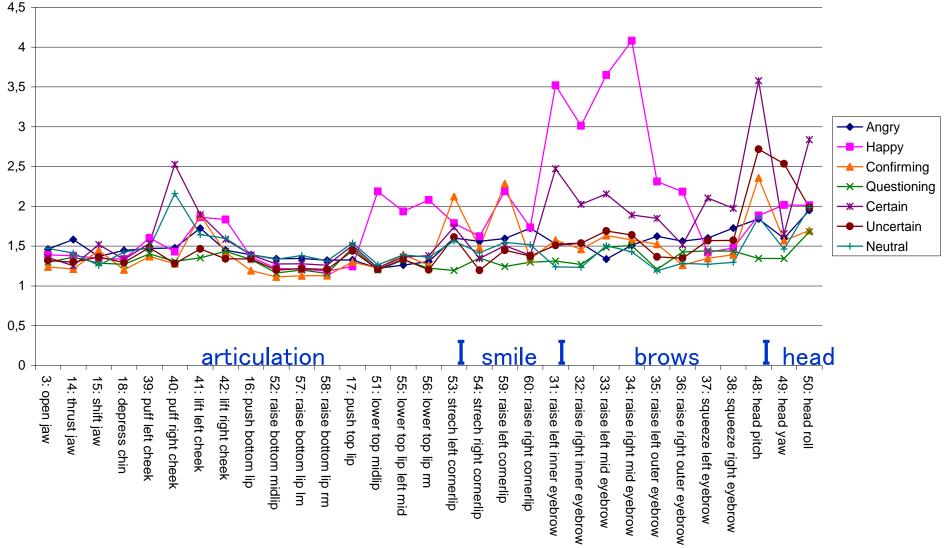


"left mouth corner"

Visual prosody Word focus vs. articulation

- Simple 3-word sentences with systematically varied focus
- Pronounced in seven different expressive modes
- Lab/acted speech
- Motion capture of points on face mapped on facial animation parameters (FAP)
- Focal motion quotient (FMQ) calculated, as variation in FAP over the focal word divided by the (mean) variation in FAP of the same word in non-focal position)
- If FMQ=1, focus does not affect that point on the face

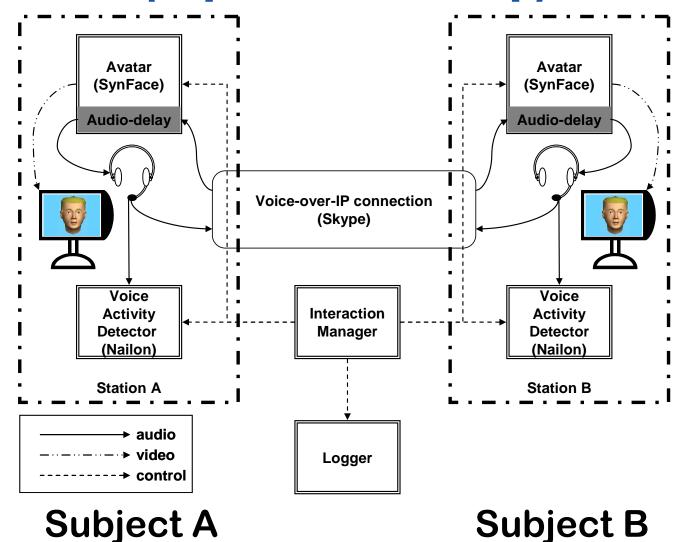
The focal motion quotient, FMQ, averaged across all sentences, for several expressive modes



Turntaking in mediated audio-visual human-human communication

- Based on "SynFace" technology originally used as lip reading support for hard-of-hearing making telephone calls
- Natural audio combined with synthetic (avatar) faces
- Can manipulated visual feedback affect the turntaking behaviour of subjects?

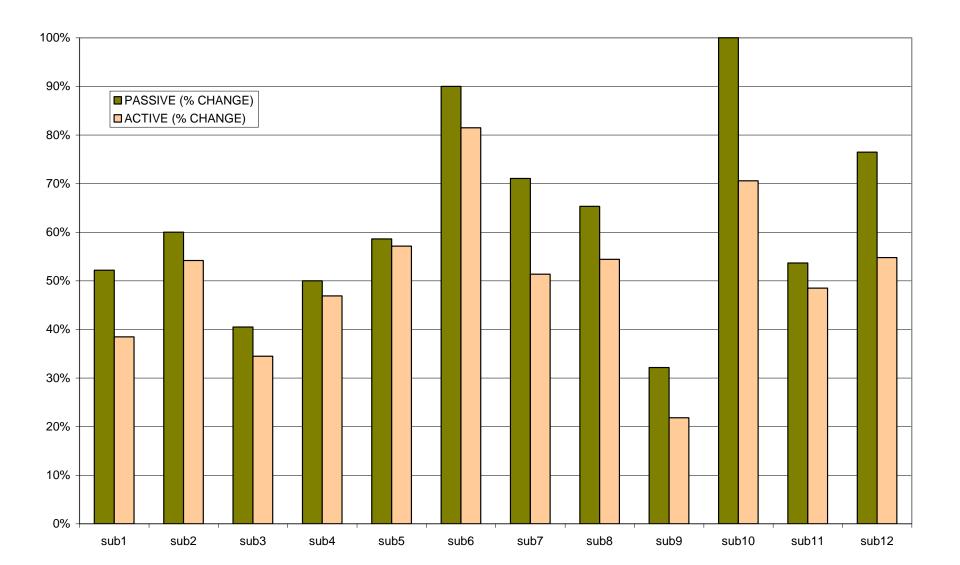
Manipulating turntaking in mediated human-human communication (experimental set-up)



The interaction manager

- Controls the visual turntaking behaviour of the avatar by controlling facial gestures:
 - Turntaking/keeping gesture (head turn and looking away (active)
 - Turn yielding/listening gesture (looks at the subject with slightly raised eyebrows (passive)
- Switching after ten detected silences between avatar A/B being neutral/neutral, active/passive, passive/active

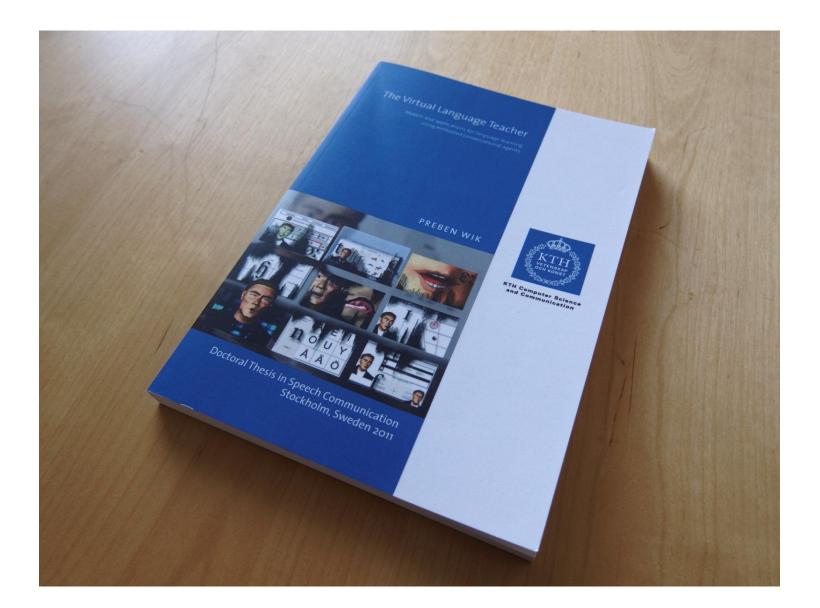
Percentage of contributions followed by a change of turn for twelve subjects represented by passive vs. active avatars



Summary of the three early studies

- Emotions more important for (lip) articulation than vowel identity
- The whole face is affected by focal accents, but differently for different kinds of expressive speech
- Visual cues often override audio speech cues
- Interaction behaviour can be manipulated by avatars, useful in e.g. multimodal dialogue systems

Applications in dialogue systems, virtual tutors and rehabilitation



Visual impact on intelligibility. Early application in the TELEFACE project

Concept video

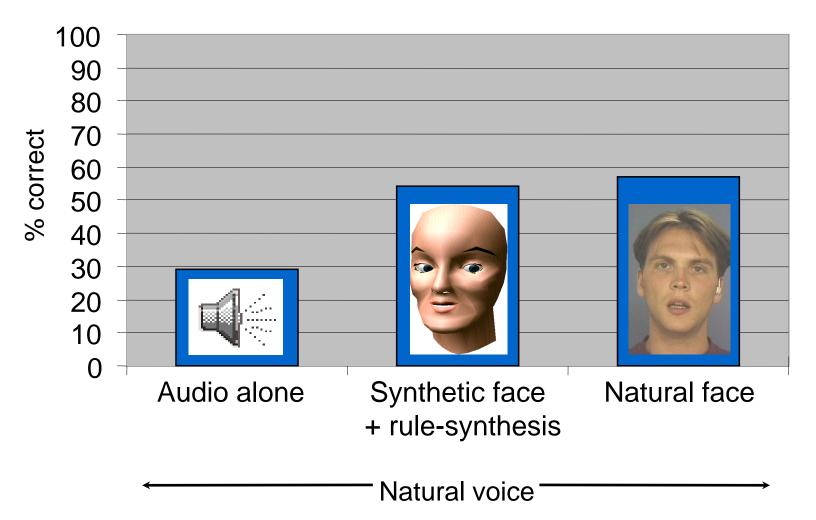


Multi-modal speech communication for hearing impaired persons



Continued in EU projects SYNFACE and Hearing at Home,

Results for VCV-words (hearing impaired subjects)



Better than humans?

| | | bil | labd | den | pal | vel |
|-------------|-------------|-----|------|------|------|------|
| | bilabia | 100 | | | | |
| ~ C~ | labiodental | | 96,3 | 3,7 | | |
| aCa | dental | | 3,0 | 78,0 | 5,5 | 13,4 |
| | palatal | | | 9,9 | 70,4 | 19,8 |
| | velar | | | 4,9 | 16,0 | 79,0 |

| | bil | labd | den | pal | vel |
|---|------|------|------------------|------|------|
| (| 96,3 | | 2,5 | 1,3 | |
| | | 92,6 | [/] 5,6 | 1,9 | |
| | | | 85,8 | 7,4 | 6,8 |
| | | 1,2 | 17,3 | 71,6 | 9,9 |
| | | | 2,5 | 25,0 | 72,5 |

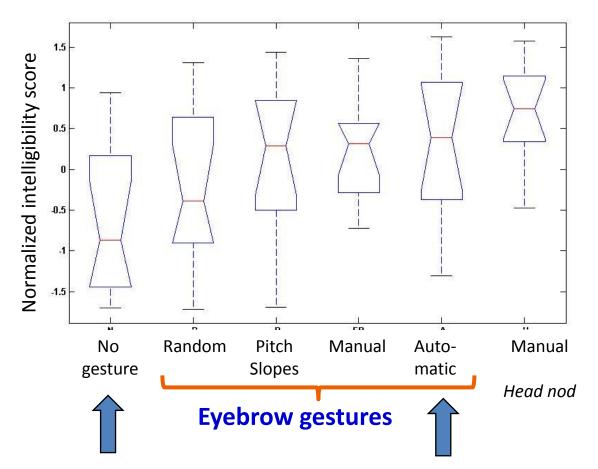
Synthetic face

Natural face

Expanded use: Reading talking books with SynFace and TOBII Eye-tracker (EU/HaH project)

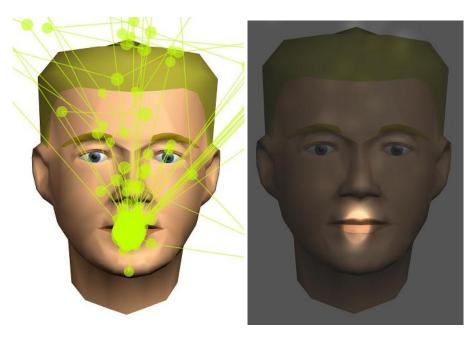


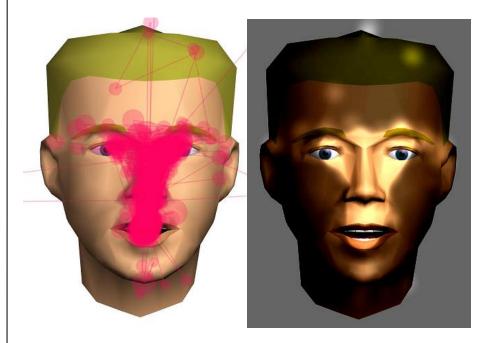
Effect of non-verbal information Addition of prominence gestures



Significant intelligibility increase for gestures (eyebrows & head nod) based on pitch slopes, automatic prominence detection and manual annotation

Eye-tracking results





Without prominence gestures

With gestures



LARGE IP - 4 years

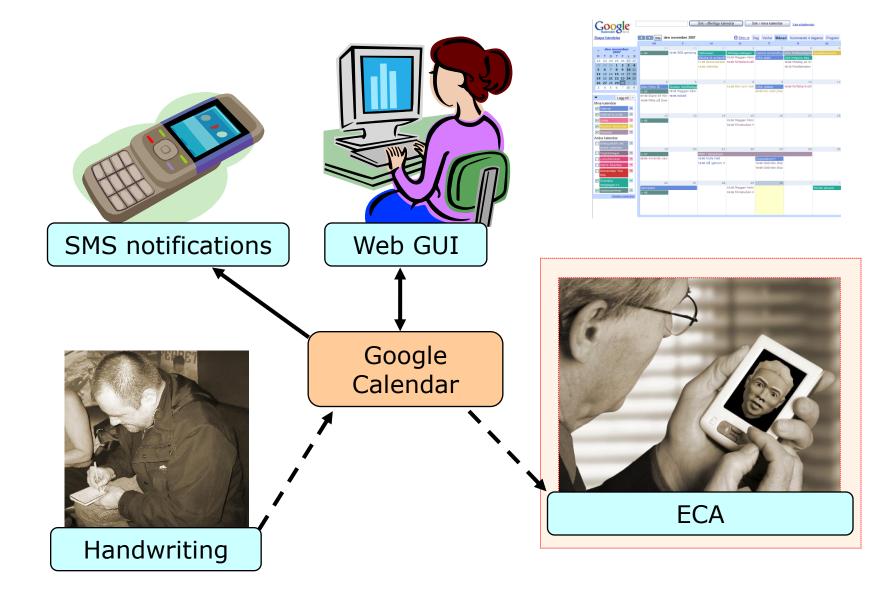
KTH Work Package (WP8)

Innovative Interfaces

Björn Granström Samer Al Moubayed, Jonas Beskow, Mats Blomberg, Jens Edlund, Joakim Gustafsson, Daniel Neiberg, Alec Seward, Gabriel Skantze



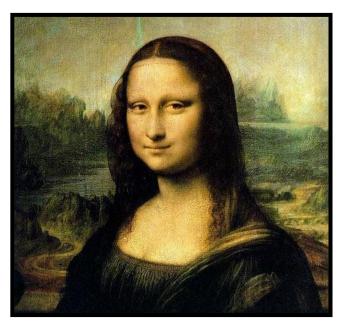
Google Calendar interfaces



Tutoring session in the MonAMI domain

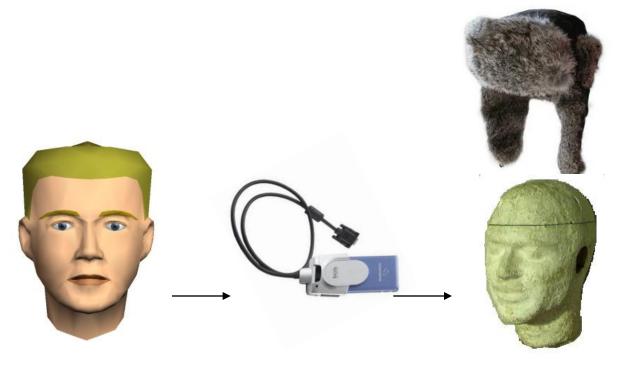


The Mona Lisa Effect



- The Mona Lisa gaze effect is when all viewers perceive the same gaze direction, independently of where they are seated.
- Good when all should see the same scene
- Also applies to avatars presented on flat screens
- Confusing in situated multi-party interaction

Proposal: 3D Displays for Avatars



Animated face

Laser projector

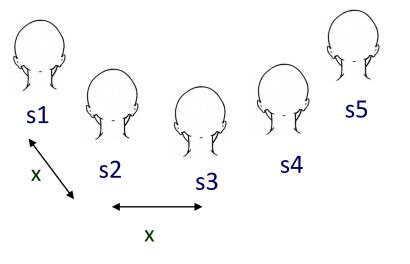
Static head model

Investigating Gaze Accuracy with 3D Projected Avatars

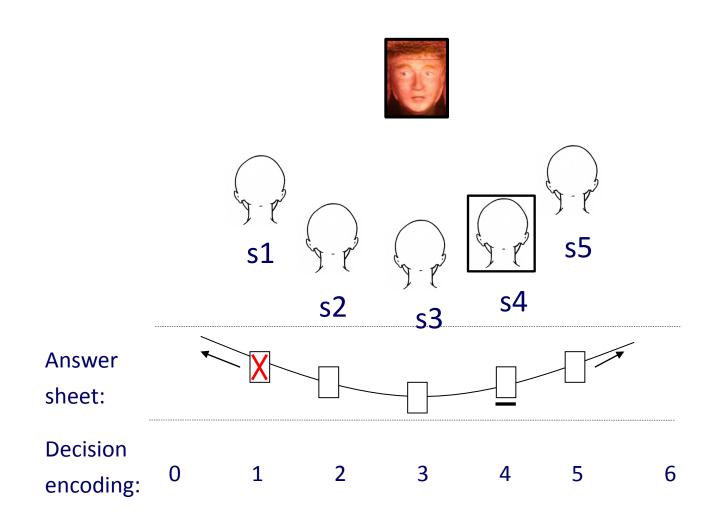
Experiment setup

- Two display methods:
 - Projection on 2D surface
 - Projection on 3D head model
- 5 Subjects, seated at equal distance from each other. with equal distance to the display.
- Gaze varied in 20 angles.
- Five runs with each subject in all positions.
- Task: Decide who is being gazed at.



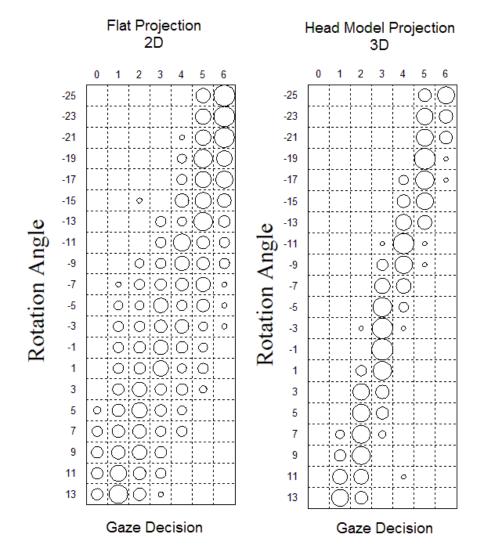


Question: On whom is he looking?



28

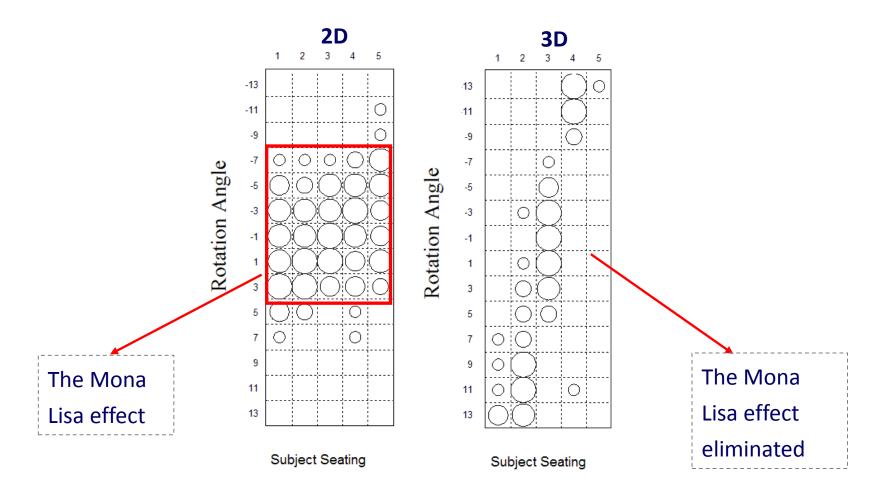
Results: Gaze Accuracy



Raw Data Plot

The Mona Lisa gaze effect in the data

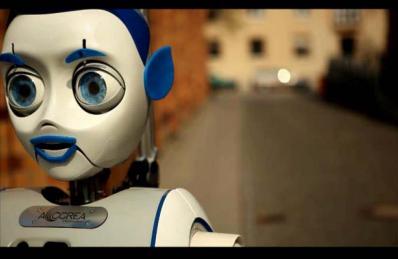
Only looking at data when eye-contact is reported



The EU/IURO project 2010-2013

The goal of the Interactive Urban Robot (IURO) project is to develop and implement methods and technologies enabling robots to navigate and interact in densely populated, unknown human-centred environments and retrieve information from human partners in order to achieve a given navigation or interaction goal.













IURO Research Issues at KTH

- How can a robot extract information from people?
 - How can we encourage people to speak?
 - How can we interpret spoken language into route graphs?
 - How can we deal with partial understanding?
- How can the robot handle multi-party dialogue?
 - Multiparty turn-taking, competition for the floor
 - Gaze, head pose and turn-taking gestures



The IURO project investigates two head solutions

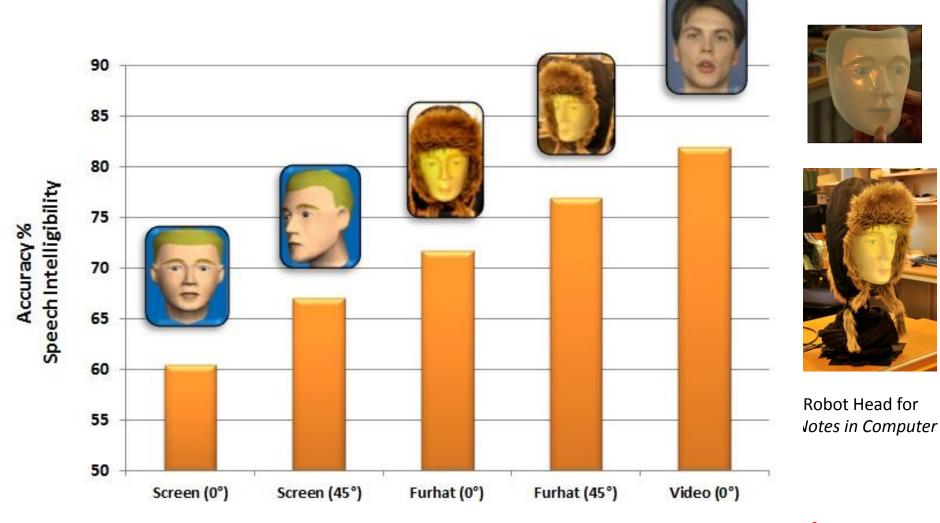
A back-projected face A mechatronic face







Developing the back-projected robot head FurHat





Condition

FurHat's usage of head and gaze direction

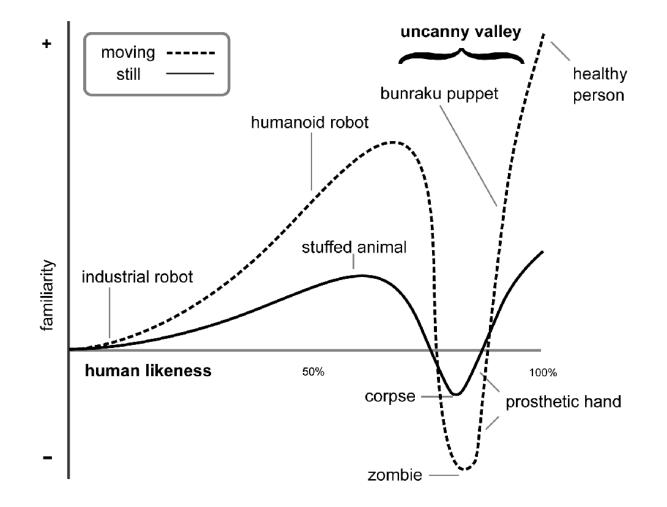
- *Indicate idle/no contact* by looking down
- *Establish contact* by turning head and gaze to newly arrived
- Put newly arrived on hold while in conversation by giving the new visitor a brief glance while saying "please wait a while!"
- Deflecting understanding problems by turning head and gaze towards the other user and ask "What do you think about that?"
- *Pose open question* to both visitors by directing the head straight ahead, then switching the gaze between them
- "Pointing" with gaze



Multi-party dialogue test at the London Science Museum:



Did we experience the "uncanny valley"?



Remember what happened to the answering machine.



Gaze and head pose for pointing



Experimental setting: Human-agent collaborative (task solving) scenario

Task: indicate where the agent is "pointing"

- 18 subjects
- 9 gaze targets per condition
- Conditions * targets randomized for each subject

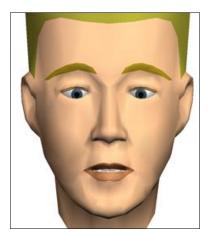
| | DC 13D | 135 134 | 140 3 | 41 142 | 145 144 | 142 | 146 14 | 40 | 142 | 144 1 | 4B 14C | 14D | 142 142 | 150 | 134 1 | 12 139 | 14 12 | 156 | 127 | 138 1 | 59 D.A | 108-3 | se lup | 102 |
|--|----------|------------|-------|----------|---------|--------|---------|-------|------|--------|----------|-----|---------|------|---------|---------|---------|--------|------|--------|--------------|--------|---------|---------------------|
| | 132 180 | 161 162 | 183 1 | 64 165 | 168 15 | 168 | 169 16 | 50 | 15C | 160 1 | dE 1dP | 170 | 171 173 | 175 | 114 1 | 12 176 | 77 17 | 179 | 178 | 178 1 | 110 | 178 1 | 19 | 181 |
| | 582 383 | 184 183 | 196 1 | \$7 188 | 189 18/ | 1110 | 18C 18 | 1 | 107 | 190 1 | 91 192 | 193 | 194 193 | 198 | 197 1 | 98 199 | A 19 | 1 190 | 19D | 19E 1 | 91° 1.40 | 1.41 1 | A2 LAD | 1.84 |
| | 145 146 | 142 141 | 139 1 | AA LAB | TAC LA | 1.AE | 1AF 19 | 81 | 182 | 180 1 | B4 185 | 196 | 187 18 | 138 | 1BA 18 | 68 1BC | D 18 | E SBF | 100 | 101 | C2 1C3 | 104 3 | C5 108 | 9C7 |
| ed for each subject | 1C8 1C9 | ICA 1CB | 100 1 | CD ICE | DCF 1D | 1Dt | 1D2 1D |) [D4 | 105 | 106 1 | D? 1D8 | 104 | IDA 1D | 1DC | 100 11 | DE 1DF | 80 I.E | 162 | 18 | 164 1 | E5 136 | 1E7 1 | BR 189 | 1EA |
| ea for each subject | 1EB 1EC | IED IEE | 12F 1 | F0 1F1 | 1F2 1F | 194 | 195 19 | 6 19 | 1198 | 159 11 | FA IFB | IFC | IFD 1FB | 199 | 200 2 | 01 202 | 0 204 | 295 | 309 | 207 2 | 08 309 | 20A 2 | 0B 20C | 20D |
| | 39E 30F | 210 213 | 213 2 | 13 254 | 215 210 | 217 | 218 219 | P. IA | 218 | 210 2 | SD 21E | 218 | 220 221 | 222 | 233 Z | 14 225 | 56 22 | 228 | 229 | 22A 2 | 2B 23C | 210 2 | 2E 22F | 230 |
| | 231 232 | 233 234 | 235 2 | 38 237 | 238 211 | 23A | 23B 23 | 30 | 238 | 21F 2 | 48 241 | 242 | 343 244 | 245 | 246 2 | 47 248 | 8 24 | 348 | 340 | 340 2 | 4E 34E | 290 Z | 33 232 | 258 |
| | 234 233 | 236 237 | 236 2 | 254 | 258 250 | 29D | 23E 23 | 10 | 261 | 262 2 | 63 264 | 267 | 100 100 | 298 | 269 3 | 6A 288 | C 28 | 242 | 247 | 270 2 | | 215 2 | 74 273 | 276 |
| | 277 276 | 279 27A | 278 | - 0 | 276 273 | 290 | 281 28 | 10 | 384 | 285 2 | 86 257 | 285 | | 219 | 3BC 31 | ED 285 | E 2H | 291 | 292 | 299 2 | h | 286 2 | 97 298 | 299 |
| | 29A 29B | 39C 29D | 296 2 | 9F 2A0 | 3A1 3A | 2 2.43 | 244 24 | 5 46 | 2A? | 2A8 2 | A9 2.4.4 | :AB | Ψ | 34E | 2.4F [1 | 80 281 | 82 28 | 284 | 286 | 286 2 | | 239 2 | RA 189 | 29C |
| | 190 19E | 19F X4 | 201 2 | C2 3C3 | 3C4 2C | 5 206 | 303 X | 1 (S | 10.8 | X8 3 | CC XD | 1CE | ICF XD | 301 | 3D1 3 | 08 304 | × 20 | e apri | 208 | 209 2 | DA 308 | BC 3 | 00 108 | 10F |
| | 2B0 2E1 | 2E) 2E) | 264 3 | ES 166 | 207 25 | 3B9 | 2EA 2B | 8 8C | 2BD | ZEE 3 | EF 250 | 291 | 1F2 1F3 | 354 | 395 2 | 96 (AP) | F8 7.P | 2FA | 2FB | ZPC Z | PD 2FE | 297 3 | 00 301 | 302 |
| THE REAL PROPERTY AND A RE | 303 304 | 307 308 | 307 3 | 08 389 | 308 318 | 5 SIC | 30D 30 | t of | 310 | 311 3 | 12 313 | 314 | 315 319 | 307 | 318 3 | 19 31A | 19 710 | 31D | 51E | 50F 3 | 39 921 | 322 3 | 23 324 | 325 |
| | 320 227 | 3,18 3,29 | SIA 3 | 18 320 | 320 32 | 5.2F | 330 33 | 32 | 333 | 334 3 | 35 338 | 337 | 158 1.8 | 23,4 | 138 3. | 9C 33D | R 33 | 340 | 341 | 342 3 | 43 344 | 345 5 | 46 341 | 348 |
| | 349 344 | 340 340 | 34D 3 | 42 347 | 100 35 | 252 | 353 ,35 | 55 | 155 | 287 2 | 18 219 | 25A | 350 350 | 350 | 258 3 | 57 350 | 51 36 | 363 | 364 | 365 3 | 66 767 | 348 3 | 69 36A | 368 |
| | ISC HD | 368 369 | 370 3 | 31 335 | 373 371 | 375 | 376 37 | 72 | 379 | 27.A 2 | 18 310 | 91D | 178 179 | 380 | 181 3 | 82 283 | 14 38 | 386 | 387 | 383 3 | 89 78A | 165 1 | C HD | 16 |
| | 387 396 | 164 162 | 393 3 | HA 195 | 386 397 | 398 | 200 30 | 99 | HC. | HD 3 | 96 39F | 348 | AL DAL | 143 | 1A1 3. | A3 3A6 | 17 2.20 | 3,40 | 344 | 348 3 | AC 14D | AAE A | AF IBO | 394 |
| | 382 383 | jB4 jB5 | 394 3 | 187 XB8 | 100 IB/ | 3388 | 18C 38 | 86 | 18F | 300 3 | ci 3c2 | 363 | K4 K | 306 | HC7 H | C8 9C9 | A 30 | 8 800 | RCD | KE 3 | (F 3D0 | 304 3 | 102 300 | 3@4 |
| | 306 3D6 | 3D7 8D4 | 309 3 | DA 3DB | 3DC 3D |) IDE | HDF HB | Đ | 厄 | 疤 3 | E4 JE5 | 386 | YE7 HE | 389 | NEA 3 | EB JEC | D HE | 3EF | 3F8 | 3F1 3 | F2 (F3 | 3F4 3 | ES 1196 | \hat{H}^{\dagger} |
| | 3F8 3F9 | JFA 3FB | 3FC 3 | FD | 3FF 40 | 401 | 402 40 | 104 | 465 | 406 4 | 07 408 | 4 | 400 | 400 | 400 4 | 0E 40E | 10 411 | 412 | 413 | 234 4 | 0 | 417 4 | 18 419 | 41A |
| | 41B 41C | 41D 41D | 417 4 | 421 | 432 423 | 424 | 425 42 | 127 | 425 | 429 4 | 2A 42B | 4 | 421 | 435 | 433 4 | 11 432 | 53 434 | 435 | 436 | 437 4 | 4 | 41A 4 | 18 400 | 40D |
| | 43E 63F | 440 443 | 40 4 | 43 444 | 445 444 | 467 | 442 44 | 64 | 46 | 44C 6 | 4D 44E | 46 | 450 451 | 482 | 457 6 | 54 455 | 86 4.5 | 431 | 4:19 | 45.4 4 | - ac | 45D 4 | SE 457 | 450 |
| | 461 452 | 453 454 | 465 4 | 65 467 | 443 453 | 46A | 45B 48 | sp. | 462 | 167 S | 78 473 | 472 | (2) 43 | 675 | 678 6 | 17 478 | 19 ATZ | 479 | 470 | 470 4 | E (7) | 480 4 | 81 482 | 415 |
| | 484 485 | 496 487 | 438 4 | 199 48.4 | 458 450 | 35D | 455 45 | F 90 | 491 | 682 -8 | 69 494 | 485 | 896 897 | 398 | 499 4 | A AR | C ait | 455 | 487 | 4A0 4 | A1 8A2 | 443 6 | A4 4A3 | 446 |
| | 48.7 448 | 4.49 4.4.4 | AB 4 | AC 4AD | 44E 14 | 490 | 481 48 | 89 | 434 | ABC 4 | 86 497 | 498 | e a | 458 | 49C 4 | 8D 486 | ₩ #C | 401 | 403 | ACS 4 | C4 4C5 | ace a | C7 4C8 | 409 |
| | 22.22.23 | 2223 | 22.2 | 589 | (28) | 122 | 22.7 | 200 | | | | | | - | | | | | | | | | | - |
| | | | | | | | | 2 | | | | | | | | | | | | | | | | |
| | | | | | | | | 1 | | | | | | | | | | | | | | | | |
| | | | | | | | | 13 | | | | | | | | | | | | | | | | |
| | | | | | | | | 137 | | | | | | | | | | | | | | | | |
| | | | | | - | - | | - | | | | | | | | | | | | | | | | |
| | | | | | | | | _ | - | | | | | | | | | | | | | | | |

Gaze vector was estimated using a highly detailed geometric model

Human vs. agent conditions



Human-human condition

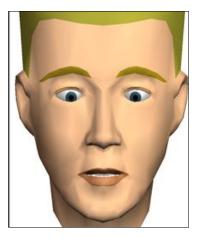


The normal condition (baseline)

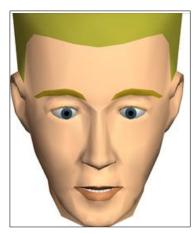
- Static head, moving eyes
- Eyelids follow the eyeballs vertically
- Subject sitting in front of the head

Three more, "anomal", conditions

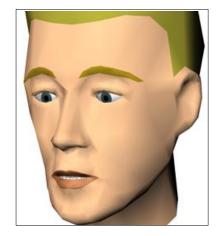




Eyelids condition Static Eyelids Static head Moving eyes



Neck condition Neck moving Eyes frontal

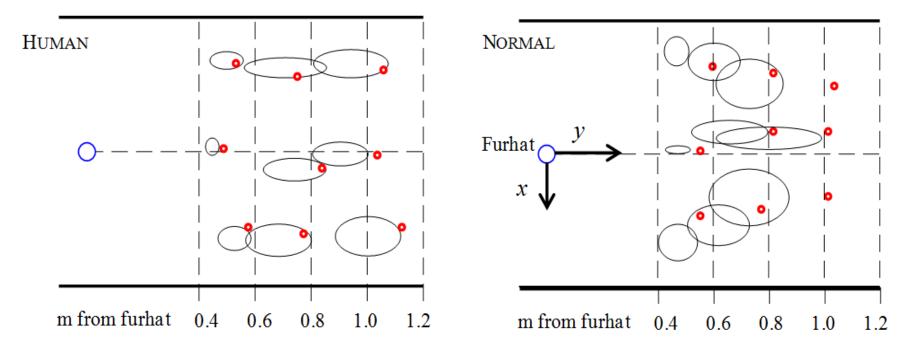


Side condition Subject seated at 45 degrees from the head.



Results

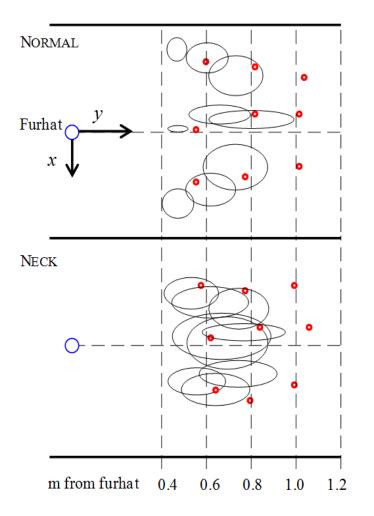




dots: intended target points ellipses: distributions of answers on x and y

Results



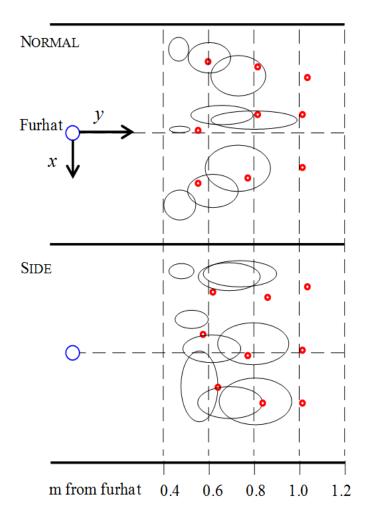


Eyelids: Perceived target is 1.47 degrees lower.

Neck: Resolution very bad on the y axis (5,86). Displacement on x (7.1) and y (2.3).

Results



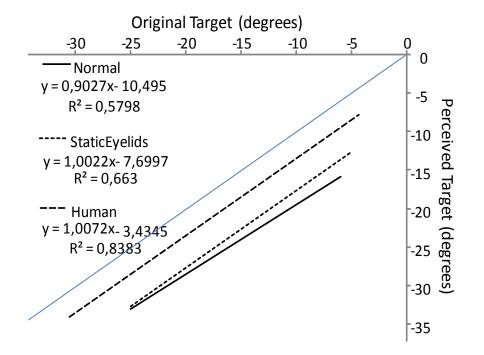


Eyelids: Perceived target is 1.47 degrees lower.

Neck: Resolution very bad on the y axis (5,86). Displacement on x (7.1) and y (2.3).

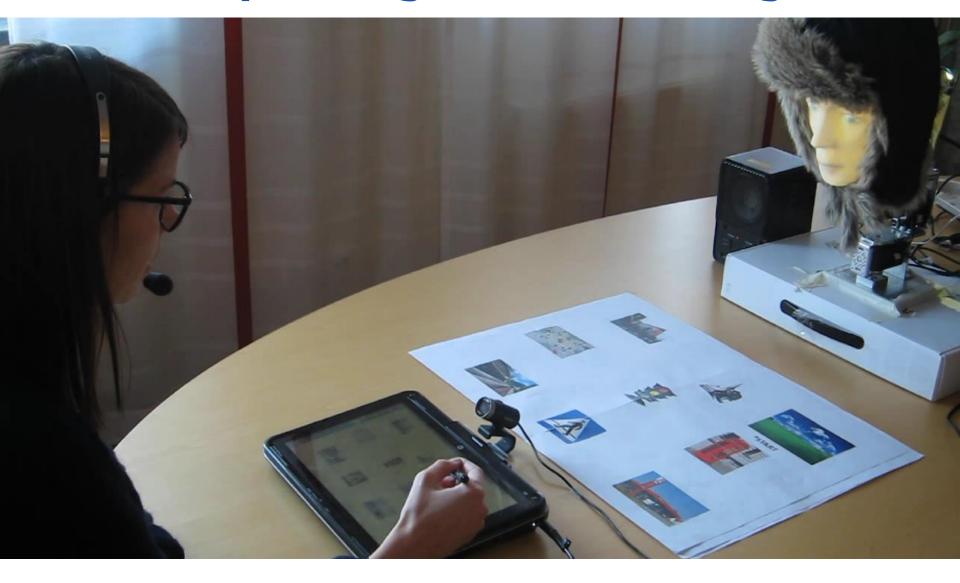
Side: resolution lower on *x* Displacement on *x*

The setup serves as a gaze calibration method



Psychometric functions for three of the conditions on the *y*-axis.

Experiment on "pointing" and interaction Map task – gaze and turntaking



Mapping FurHat to the IURO mechatronic head Tällberg forum domain



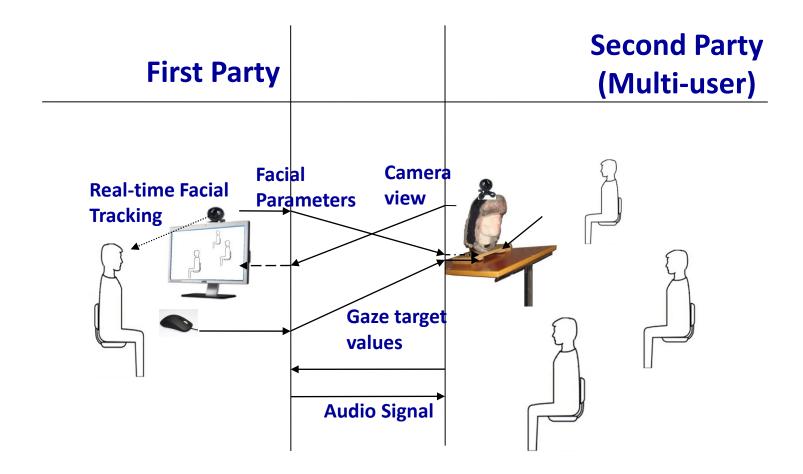
Possible non-robotic applications in mediated communication



No shared access to environment No ability to have exclusive eyecontact (the Mona Lisa effect) User always looks at the video instead of the camera (no shared attention)

Standard video conferencing setup

Proposal: use a 3D head for the remote subject



Tele-presence on one side

In summary:

- Something Olde, Something New, Something Borrowed, Something Blue.....
- Due thanks go to all my co-workers at KTH, in particular: Samer Al Moubayed Jonas Beskow Jens Edlund Joakim Gustafson Anna Hjalmarsson Gabriel Skantze



Furhat at our Christmas party

The End

Questions?





www.facebook.com/furhat.svensson