Recognizing Whole People in Natural Environments

Carina A. Hahn[1], Alice J. O'Toole[1], P. Jonathon Phillips[2]

[1]The University of Texas at Dallas, Richardson, TX, [2]The National Institute of Standards and Technology, Gaithersburg, MD

Background

Person recognition work has concentrated mostly on the face
- In natural viewing environments, we see *whole people in motion*
- As person approaches, information from the face and body changes over time

Goal: Examine person recognition in natural environments when viewing a person approaching

1) *When do we make decisions of recognition* over the time course of viewing a person approaching?
2) *Do judgments gradually build* based on changing information as a person approaches, *or are decisions based on the quality of information* immediately at hand?
3) *How much do the face and body contribute* to recognition?

Information that supports person recognition: face, body, gait
- How is this information combined over time when making recognition judgments?
  - Face & body contributions to recognition[1,3,4,5]
  - Gait[2]

Present approach: Build robust representation of identities -> Recognition *tests to examine time course of recognition and contribution of faces and bodies*
- Possible outcomes:
  - *accumulated* information determines recognition accuracy
  - *quality* of information from face and body at time of recognition decision determines accuracy
  - *contribution of faces and bodies varies as a function of person's distance* from viewer
  - *contribution of faces and bodies is static* across distances

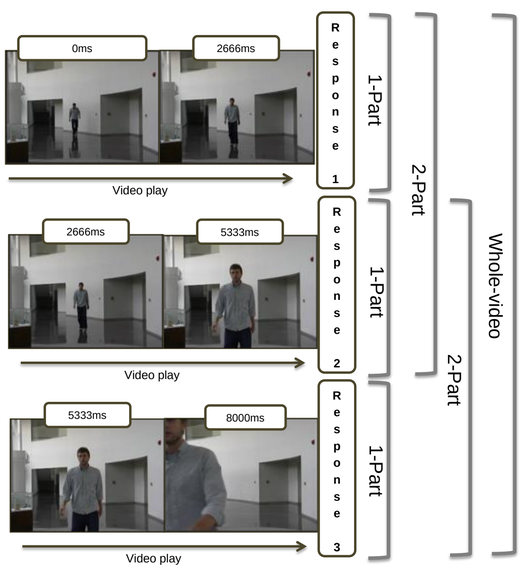
Method

Training (all experiments)

- Familiarized with 30 identities
  - 4 motion-based actions
  - 120 videos total

Experiment 1 – Segmented Videos

Three between subjects *video-quantity conditions*:
1-part: see only 1/3 of video per identity (*n* = 16)
2-part: see 2/3 of video per identity (*n* = 18)
Whole-video: see entire video per identity (*n* = 16)



At each response, participants indicated if identity was unfamiliar to familiar on scale of 1 to 5 (sure unfamiliar to sure familiar)

Experiment 2 – Blurred Face or Bodies

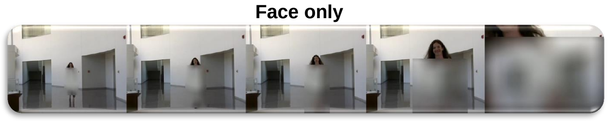
Three between subjects person-visibility conditions:
Whole Person (*n* = 33); Face Only (*n* = 18), or Body Only (*n* = 15)

All participants took two types of tests:
Prompted Response and Free Response

Stimuli

Whole person

Face only



Body only

Prompted responses test

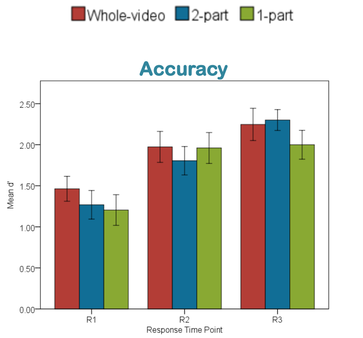
Same as Experiment 1's Whole-video condition

Free responses test

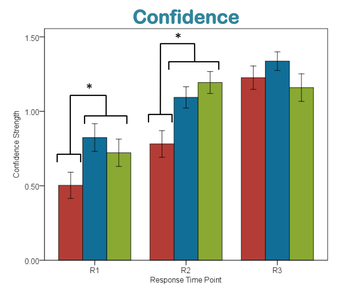
Respond "as soon [they] as feel confident"
Binary response: "familiar" or "unfamiliar"

Results

Experiment 1 Segmented Videos



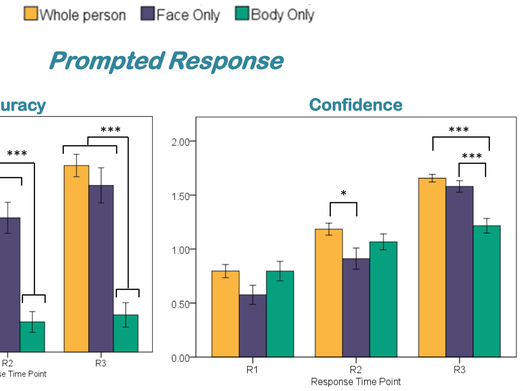
Accuracy improves with proximity
Last response time: no benefit of seeing more video beforehand



Confidence increases with proximity
More confident when more video is not expected

		Accuracy			Confidence		
		R1	R2	R3	R1	R2	R3
		Mean	Mean	Mean	Mean	Mean	Mean
Condition	Whole-video	1.46	1.97	2.25	.50	.78	1.23
	Two-part	1.27	1.81	2.30	.82	1.09	1.34
	One-part	1.20	1.96	2.00	.72	1.19	1.16

Experiment 2 Blurred Face or Bodies



Accuracy improves with proximity in all cases except with bodies only
Whole person and Face Only > Body Only

Greater increase in confidence with whole people and faces only
At R3: Whole person and Face Only > Body Only

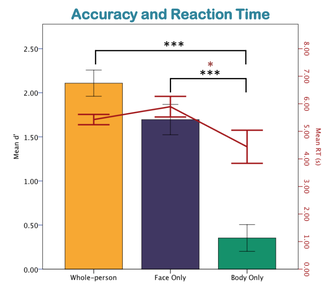
Accuracy (d')

		R1	R2	R3
		Mean	Mean	Mean
Condition	Face Only	.85	1.58	1.97
	Body Only	.42	.35	.44
	Whole person	1.46	1.87	2.20

Confidence

		R1	R2	R3
		Mean	Mean	Mean
Condition	Whole person	.80	1.18	1.66
	Face Only	.58	.91	1.58
	Body Only	.80	1.07	1.22

Free Response



Most accurate with whole people and faces. Fastest with bodies only.

		d'		RT (s)	
		Mean	Standard Deviation	Mean	Standard Deviation
Condition	Whole-person	2.11	.85	5.43	1.08
	Face Only	1.70	.73	5.90	1.59
	Body Only	.35	.59	4.44	2.33

Conclusion

High quality information from the face = accurate recognition of people

Information accumulation ✗

- Information available immediately prior to recognition decisions determines accuracy
  - Accuracy and confidence are higher as the person in the video approaches, however…
  - …same at the latest response time in the video regardless of how much video of that person was seen prior.

Contribution of faces and bodies

- Accuracy is determined by view of face – improves as person gets closer
  - Quality of information from the face improves with proximity
  - Fastest responses with bodies only – but poor accuracy overall
- With whole people, faces only, and bodies only: confidence increases with proximity
  - Greater increases in confidence with whole people and faces only compared to bodies only
  - When person is close to camera, higher confidence when able to view face

References

[1]Burton, A. M., Wilson, S., Cowan, M. & Bruce, V. (1999). Face recognition in poor-quality video. *Psychological Sciences, 10*, 243-248.
[2]Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movements. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 210–220.
[3]O'Toole, A. J., & Roark, D. (2010). Memory for moving faces: The interplay of two recognition systems. In *Dynamic Faces: Insights from Experiments and Computation* (pp. 30–45).
[4]Rice, A., Phillips, P. J., & O'Toole, A. J. (accepted). The role of the face and body in person identification. *Psychological Science.*
[5]Robbins, R. A., & Coltheart, M. (2012). Left – right holistic integration of human bodies. *The Quarterly Journal of Experimental Psychology, 65*(10), 1962–1974.