Designing an Expressive Avatar of a Real Person

9/20/2010

Sangyoon Lee, Gordon Carlson, Steve Jones,
Andrew Johnson, Jason Leigh, and Luc Renambot

University of Illinois at Chicago

Outline

- Background / Introduction
- Related Works
- Lifelike Responsive Avatar Framework
- Design of an Avatar
- Pilot Study
- Conclusion

Background / Introduction

- An avatar, a human-like computer interface, has been actively studied in various area and is becoming more and more prevalent
- With widely spread advanced technology, a lifelike avatar becomes capable of increasingly natural interaction
- However, developing such realistic avatar is non-trivial process in many aspects (i.e. requires a fair amount of manual intervention and/or high fidelity hardware)

Background / Introduction

- A proposed framework and design method reduces such initial barrier and is applicable to various domain
  - We especially focus on expressiveness of an avatar and its realistic visualization
  - Developed prototype applications and avatar for a specific person
  - Pilot study confirmed that our method is partially successful in conveying expressions

Related Work

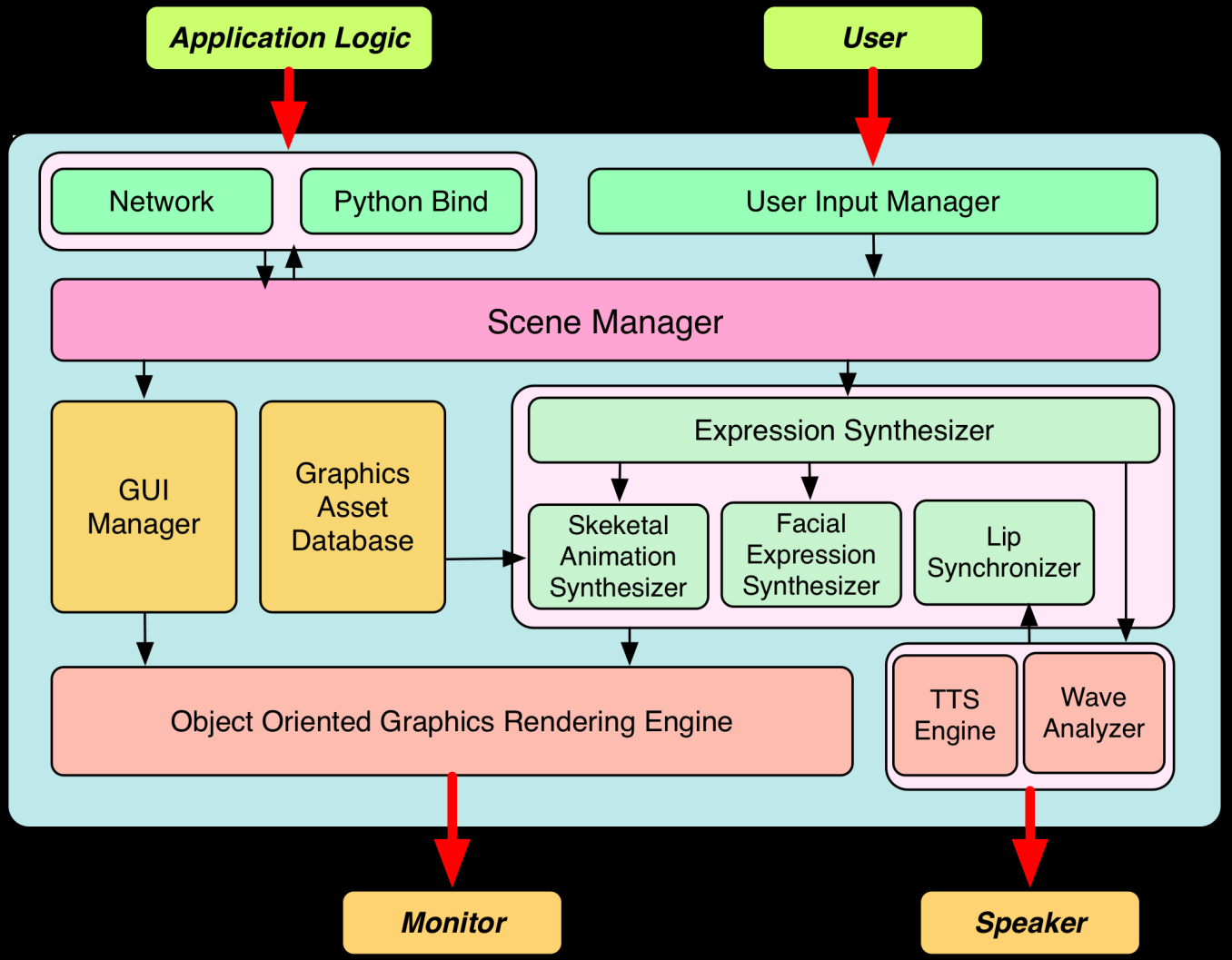
- Effectiveness of avatar: better in subjective responses but not much in performance (Yee 07)
- In contrast, more natural agent model illustrated better results in both measures (Bickmore 05,09)
- With respect to expressive avatar, various graphics techniques contribute to subjective certainty and quality of its conveyance (Wallraven 05, 06, Courgeon 09, de Melo 09)

Lifelike Responsive Avatar Framework

Lifelike Responsive Avatar Framework

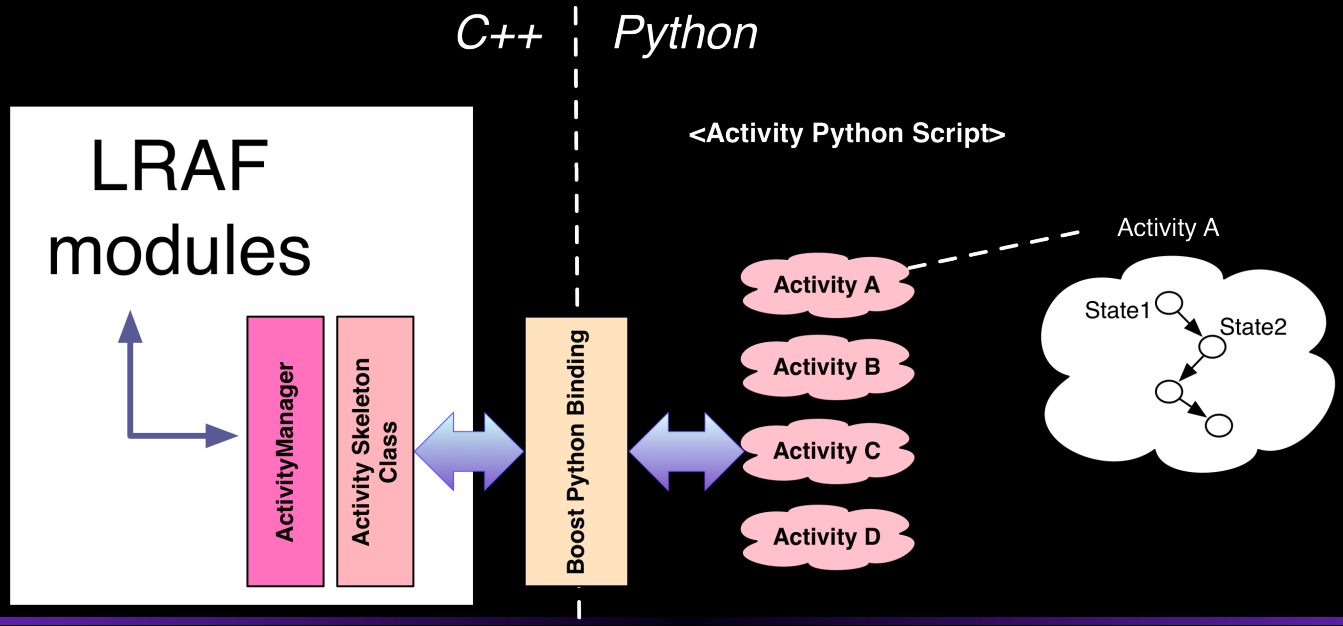
- Speech-enabled lifelike computer interface
- Support various aspects that are necessary to realize natural interaction between a user and computer - an avatar
- Framework is written in C++ / Python
- Underlying modules of framework relies on several open-source libraries (i.e. graphics and speech)

System Architecture

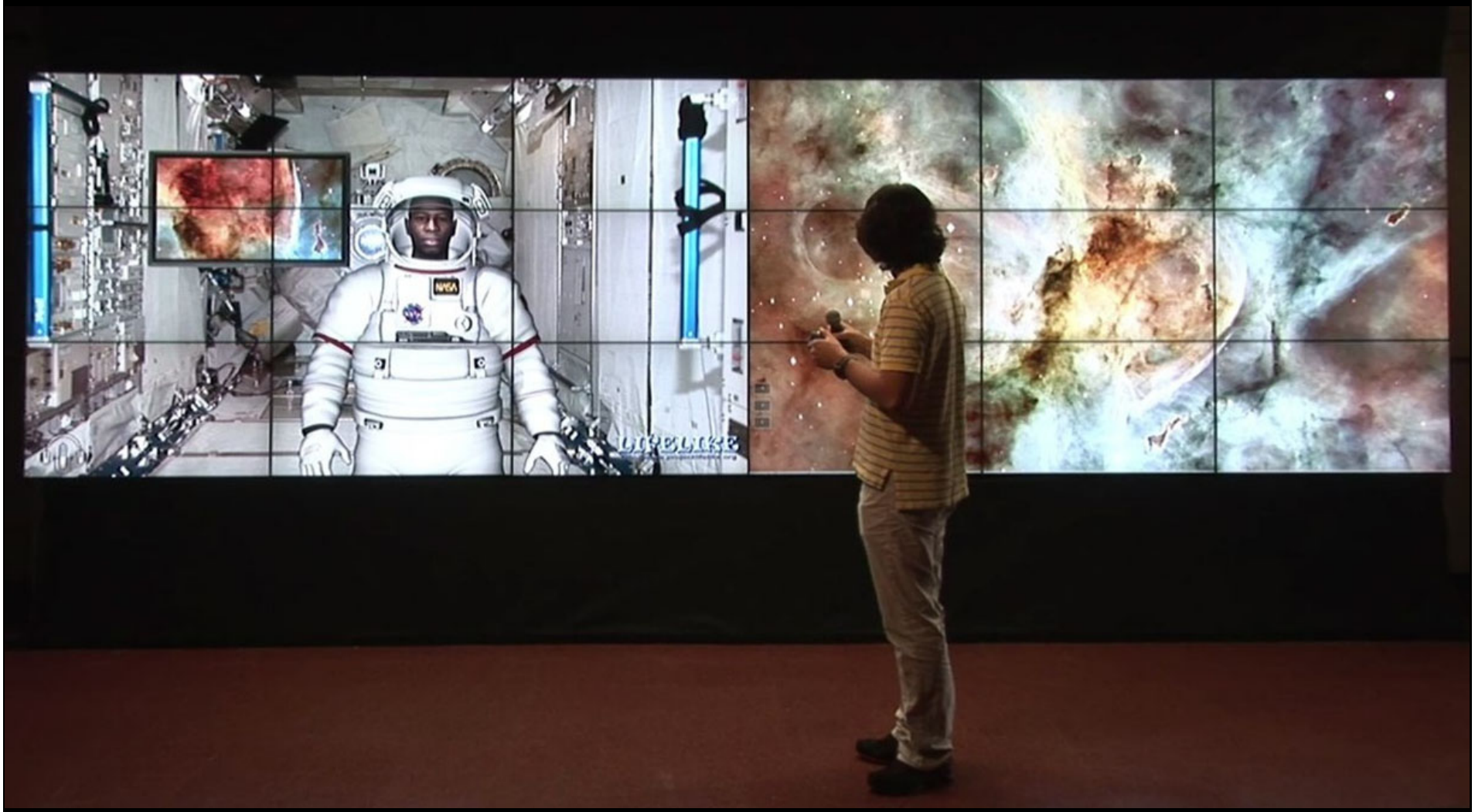


Application Logic Control

- Two separate methods can be used to control avatar behavior by external logic
  - Network communication via TCP
  - Python bindings to LRAF activity manager



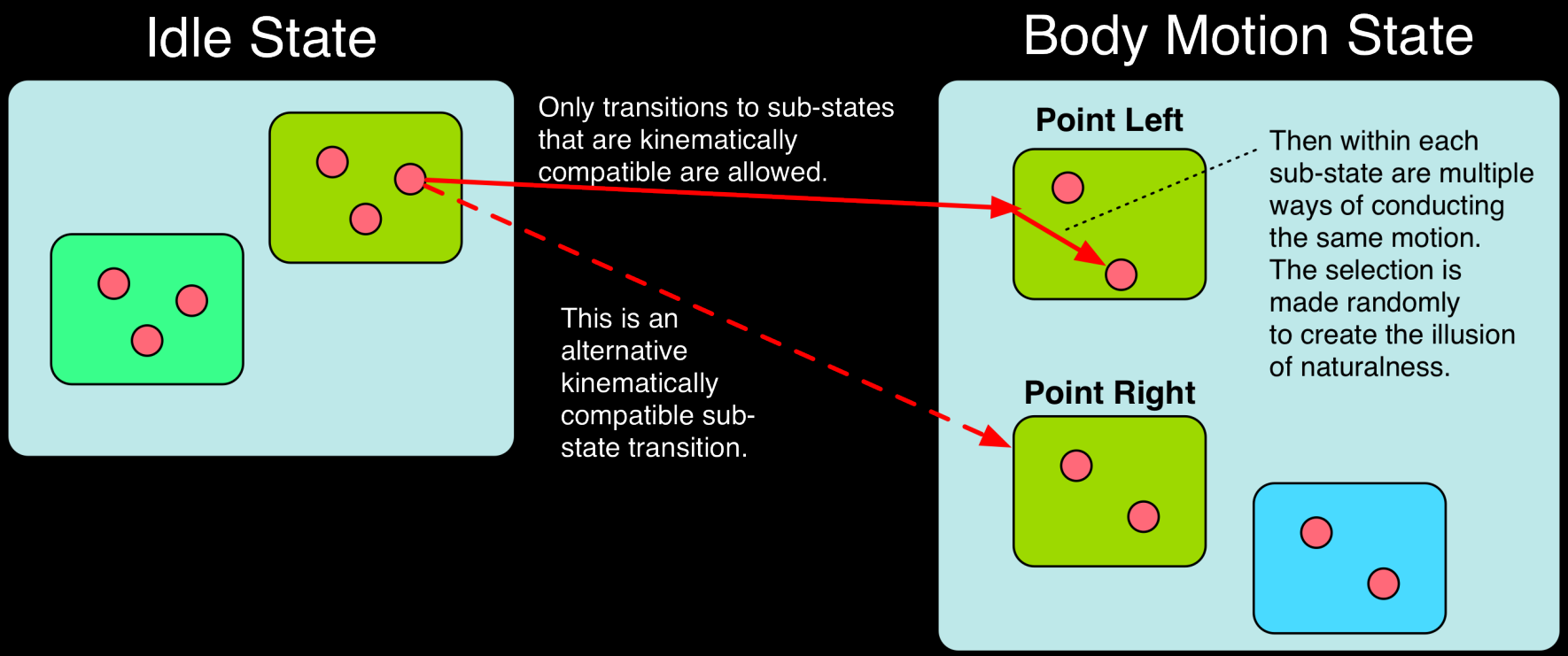
Python Binding Example



Fullbody Motion Synthesis

- Motion Acquisition via Optical Motion Capture System
- Constructed Action-based motion DB
- Semi-Deterministic Hierarchical Finite State Machine (SDHFSM)
  - Enables random selection of motion within a pool of motion DB (categorized motion clips)
  - Minimize monotonous repetition of motion

SDHFSM Model



Facial Expression Synthesis

- Blendshape-based weight parametric model
- Total 40 shape models are used
  - 7 base emotions, 16 modifiers, 16 Viseme
- Three phases blending
  - Emotional state change
  - Pseudo-random expressions (i.e. blinking)
  - Lip-Synchronization to speech (TTS, Wave)

Facial Expression Samples

Design of Avatar

A target person: Dr. Alex Schwarzkopf

An avatar of Dr. Alex Schwarzkopf

In Brief

- Head model is generated from photos of a target person using FaceGen software
  - 40 blendshape morph target is available for facial expression
- Skeleton-based full body skin mesh model
  - Modeled in commercial modeling tool (Maya)
  - Physically matches to a target person
  - 70 bones rigged to body meshes

Skin Texture Refinement

- High-res photo-based texture map enhances the quality of renderings in great degree

Default Texture 512 x 512 resolution

Photo-based Texture 4k x 4k resolution

Skin Texture Refinement

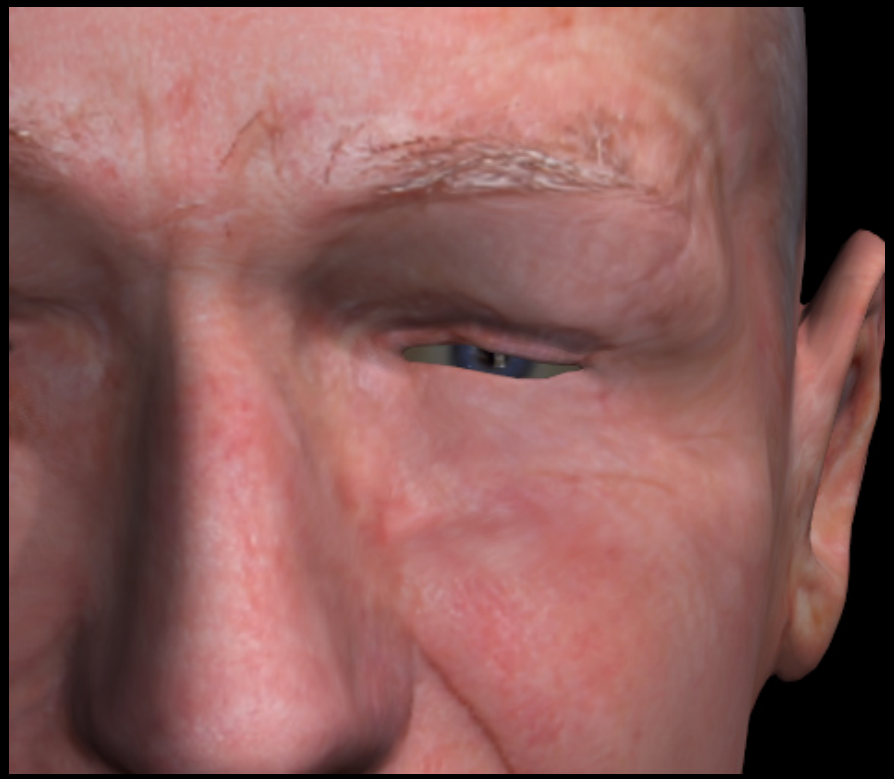
- High-res photo-based texture map enhances the quality of renderings in great degree

Rendering result /w default texture

Rendering result /w photo-based texture

Face Normal Map

- Most effective way to preserve skin details (i.e. skin pores & shallow wrinkles)



Rendering result /w diffuse texture

Rendering result /w diffuse & normal texture

Reflective Material Shader

Pilot Study

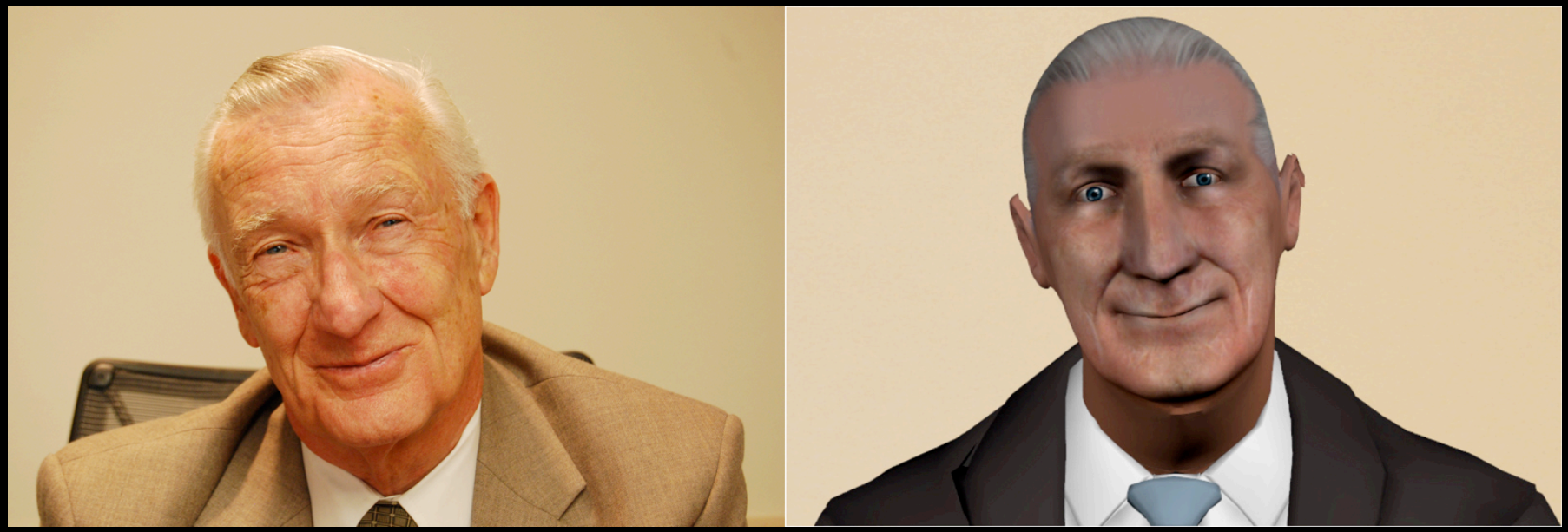
Overview

Can we identify emotions correctly in both human and avatar?

- Ekman's 6 classic emotions used
- Three pairs of images for each emotions
- n = 1,744 (online survey by Univ. students)
  - Even split of gender (864 vs. 867)
  - Ages from 18 to 64 (mean 23.5)

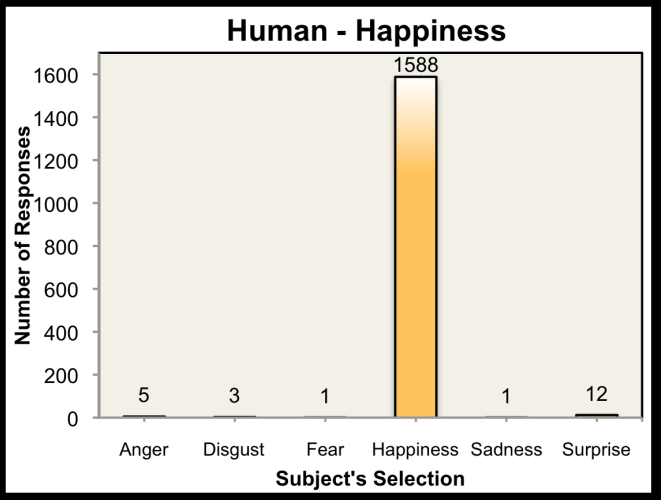
Sample Preparation

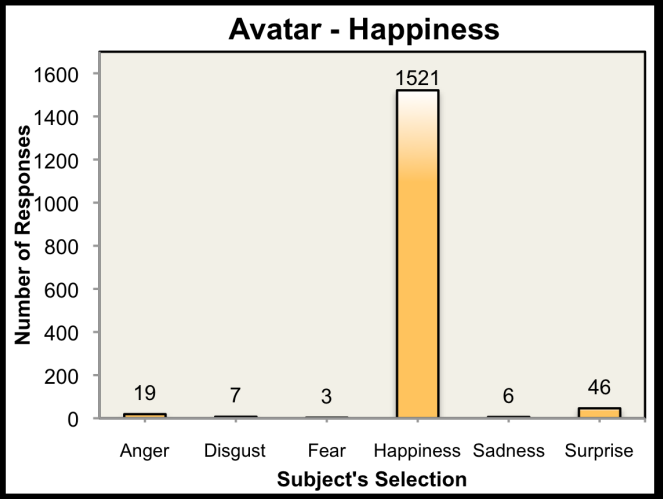
- Photos of the target person's expression
- Adjusted avatar expression parameters



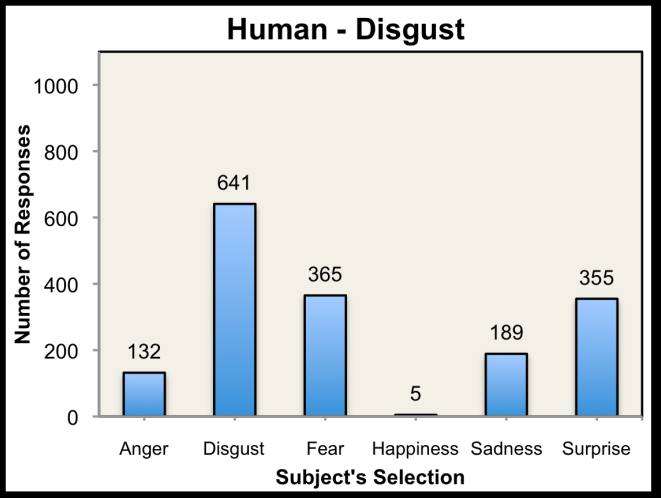
Parameters: Smile(1.0), BlinkLeft(0.2), BlinkRight(0.1), BrowUpLeft(0.2), BrowUpRight(0.1), and Phoneme B(0.65)

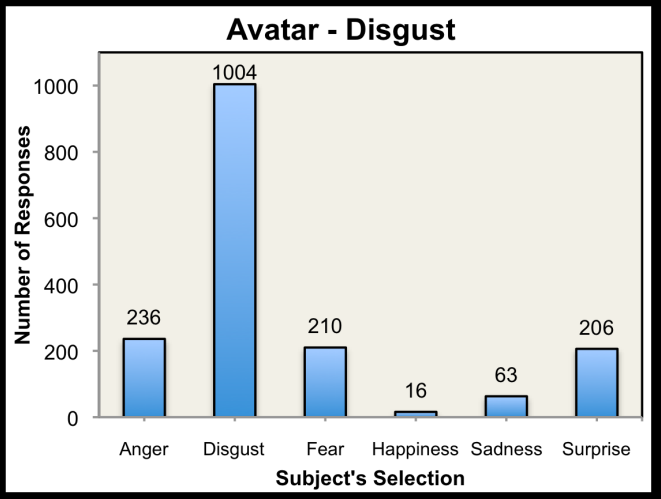
Sample Emotions





Sample Emotions





Results

Emotions	Anger	Disgust	Fear	Happiness	Sadness	Surprise	*n*
Anger	**41.9** / 17.6	27.4 / **34.1**	2.4 / 12.3	17.6 / 8.5	3.3 / 22.5	7.5 / 5.1	1597
	30.5 / **31.8**	18.3 / 19.3	10.2 / 3.1	2.9 / 27.7	13.1 / 10.6	25.1 / 7.4	1624
	49.1 / 17.0	28.0 / 21.4	3.5 / 12.1	0.4 / 10.0	14.0 / **32.5**	4.9 / 6.9	1587
Disgust	6.4 / 17.1	35.3 / **27.3**	0.6 / 3.4	**49.5** / 26.5	2.8 / 19.4	5.4 / 6.3	1587
	7.8 / 13.6	**38.0** / **57.9**	21.6 / 12.1	0.3 / 0.9	11.2 / 3.6	21.0 / 11.9	1684
	13.6 / **44.1**	**42.3** / 30.5	7.4 / 3.9	0.2 / 5.3	34.7 / 15.0	1.9 / 1.3	1593
Fear	3.2 / 3.0	12.5 / 8.3	39.6 / 16.4	0.3 / 1.1	**40.2** / **59.9**	4.2 / 11.3	1592
	3.9 / **78.8**	22.6 / 9.5	25.6 / 4.7	0.4 / 0.7	**42.4** / 5.4	5.0 / 0.9	1612
	30.8 / **69.1**	**32.7** / 11.0	14.8 / 2.1	0.4 / 12.0	13.6 / 1.6	7.6 / 4.3	1601
Happiness	0.0 / 0.7	0.1 / 0.8	0.1 / 0.3	**98.7** / **93.9**	0.2 / 0.9	0.9 / 3.3	1599
	0.2 / 1.1	0.2 / 0.5	0.4 / 0.5	**93.5** / **89.1**	0.2 / 0.5	5.5 / 8.2	1685
	0.3 / 1.2	0.2 / 0.4	0.1 / 0.2	**98.6** / **94.9**	0.1 / 0.4	0.7 / 2.9	1600
Sadness	0.7 / 20.9	20.2 / 13.5	1.8 / 5.8	0.8 / 9.5	**74.7** / **46.7**	1.7 / 3.6	1595
	0.9 / 1.9	2.7 / 4.7	2.2 / 6.3	0.2 / 0.6	**93.6** / **85.5**	0.4 / 1.1	1610
	1.1 / 1.6	7.7 / 4.1	3.2 / 3.7	0.2 / 1.8	**85.6** / **87.6**	2.2 / 1.2	1586
Surprise	5.3 / 4.1	13.2 / 10.5	23.3 / 27.7	0.9 / 1.0	8.6 / 27.6	**48.5** / **28.2**	1666
	5.6 / 8.8	7.3 / 6.3	18.4 / 2.7	2.7 / 21.1	**36.4** / 2.6	29.6 / **58.5**	1604
	5.9 / 7.2	18.4 / 6.2	8.8 / 5.2	1.9 / 31.5	29.2 / 10.3	**35.8** / **39.6**	1594

Summary

- Two emotions (Happiness, Sadness) were recognized successfully in both cases
- Other four emotions showed mixed results
  - Chosen person's expression may be not prototypical
  - Reading human emotion is not trivial task
  - Context-less still image might be not enough to convey emotions

Conclusion

- Demonstrated our avatar framework and design method for a specific person
- Our implementation is partially capable of successfully conveying human emotions
- In the future study
  - Better visualization need to be studied
  - Temporal aspect of expression
  - Context-sensitivity in expression
  - Task oriented performance measures

Thank you!

- More information on project website

  http://www.projectlifelike.org

* This project is supported by NSF, awards CNS-0420477, CNS-0703916

Also like to thank to all co-authors!