What make a Virtual Human Alive ?

1. Avatar & Autonomous Virtual Humans
2. The complexity of expressive movements
3. From artificial to real: the uncanny valley
4. Motion capture is part of the solution (film)
5. Perception of real-time animation
6. Core real-time VH believability factors
7. Other R&D efforts & exercise

1. Avatar & Autonomous Virtual Human

- **Avatar** : [W]
  - (from sanskrit): is a term used in Hinduism for a material manisfestation of a deity
  - (computing): the graphical representation of a user. In VR the avatar movement is expected to be partially or completely driven by the user body movement

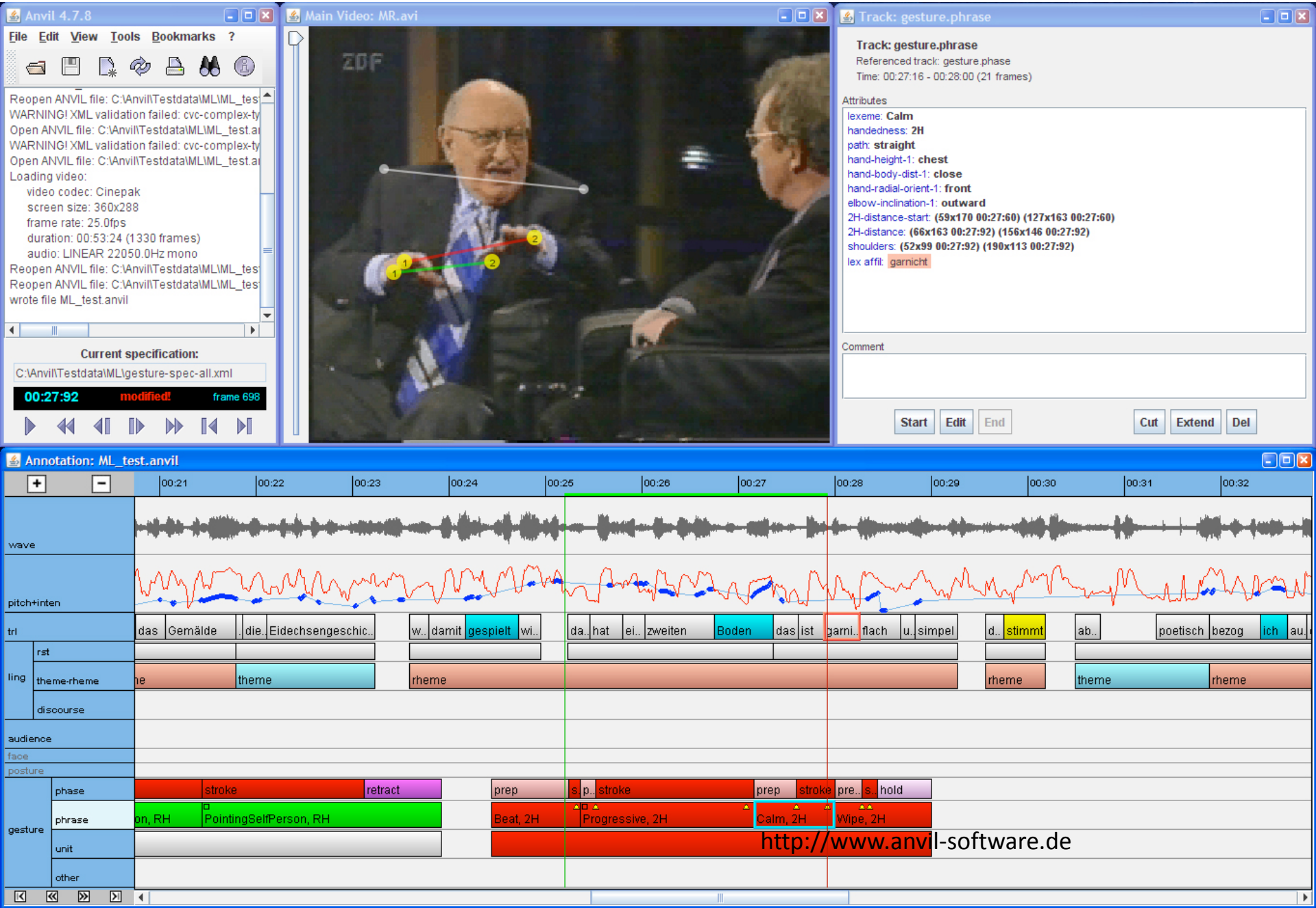
- **Autonomous/Intelligent Virtual Human**
  - for the evaluation of a Virtual environment (e.g. Pedestrian from a crowd in an emergency simulation)
  - For training purpose: the VH takes an active part in a scenario, e.g. Audience in a public speaking to overcome such a phobia

2. The complexity of expressive movements

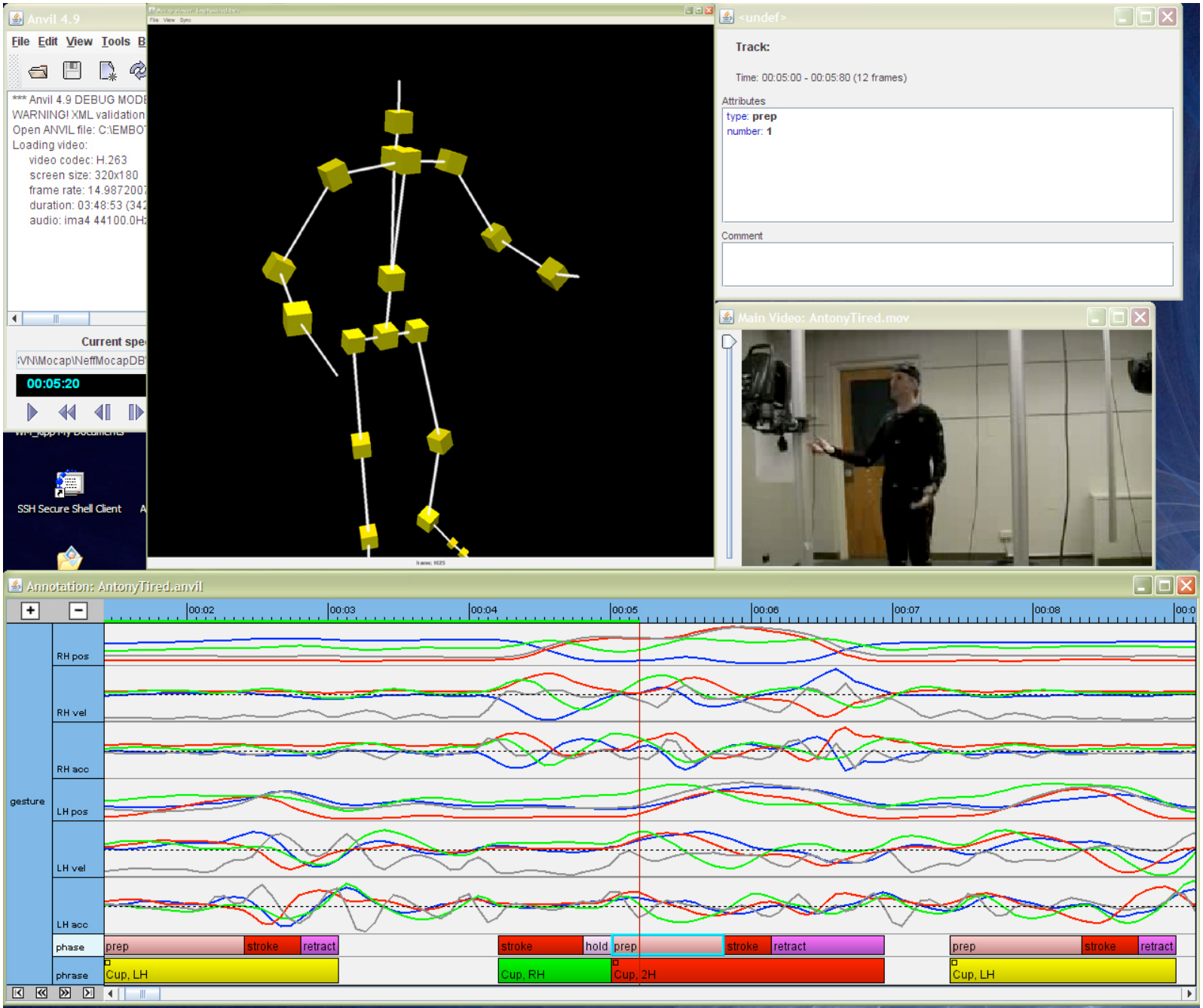
- Human expression is multi-modal:
  - Gestures should be considered to be “full-body” even if they seem to involve only the hands and arms.
  - Gestures production always includes some balance control
  - The body movement is linked to the gaze & facial expression
  - Verbalization & emotions animate the mouth and eyes
  - The vocal prosody reflects intentions and emotions
  - The tongue makes complex movements when speaking
  - Cloth, accessory, hairs, sweat, tears, human tissue dynamics can be important ***secondary movements***

- Analysis tools are necessary to understand part of these subtle interactions [K 2011]:
  - ANVIL (open source project) http://www.anvil-software.de

Annotating multi-modal human expression with ANVIL [K 2011]



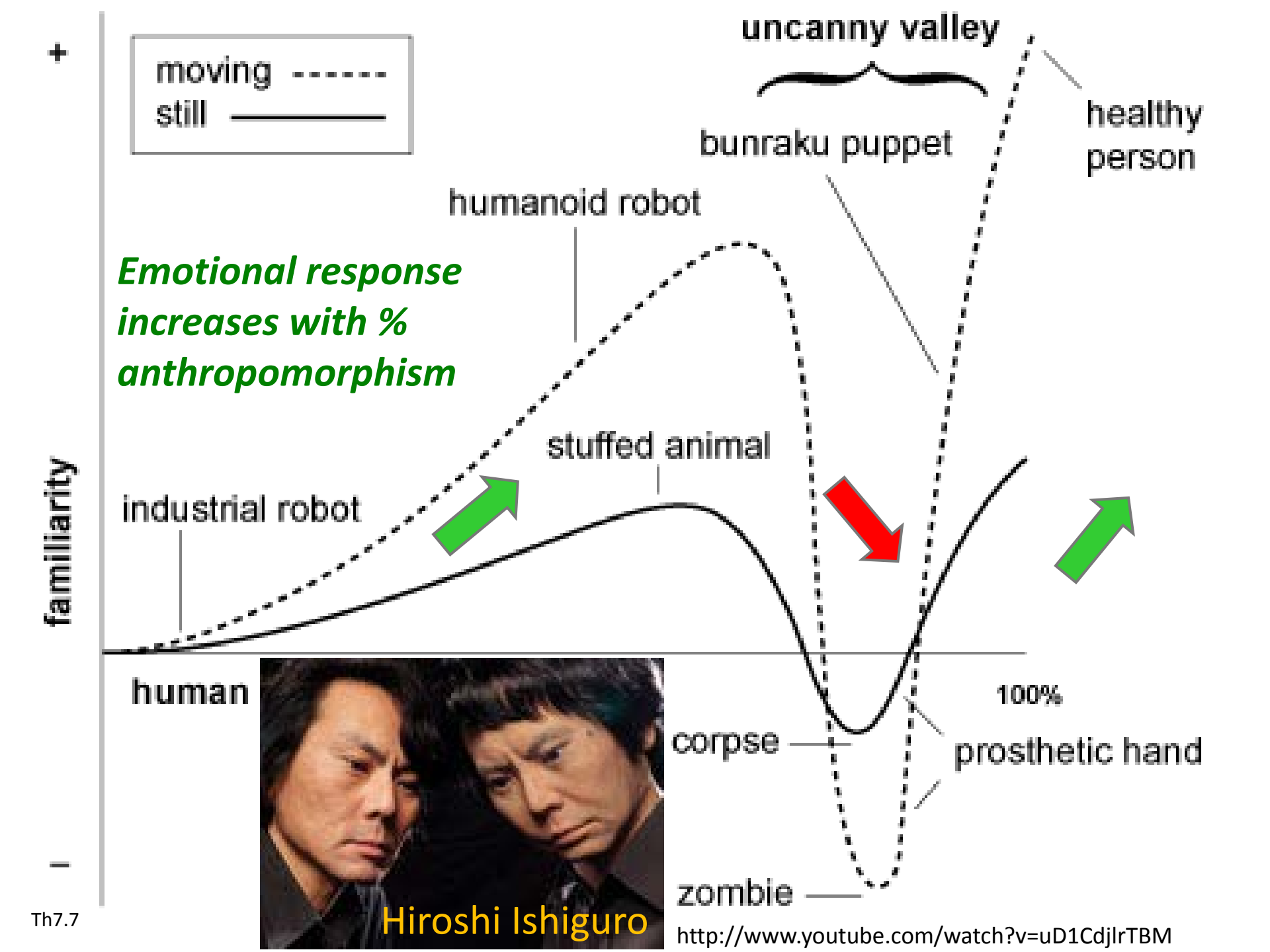
Analyzing body expression with ANVIL [K 2011]



3. From artificial to real : the uncanny valley

- **uncanny** *: (Merriam-Webster)*
  - a : seeming to have a supernatural character or origin : EERIE, MYSTERIOUS
  - b : being beyond what is normal or expected : suggesting superhuman or supernatural powers

- In the 70s Masahiro Mori studied in Robotics the emotional response effect to increasing human-like appearance of still or moving entities.
  - His key article has been translated by McDorman


+
–
familiarity
moving
still
uncanny valley
healthy person
bunraku puppet
humanoid robot
Emotional response increases with % anthropomorphism
stuffed animal
industrial robot
human
100%
corpse
prosthetic hand
zombie
Hiroshi Ishiguro
http://www.youtube.com/watch?v=uD1CdjlrTBM

3. From artificial to real : the uncanny valley (2)

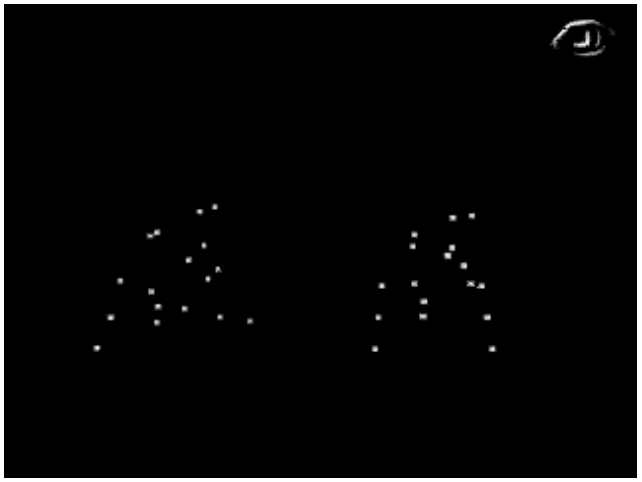
- The paper from M. Mori is questioned regarding its scientific validity (empirical experience rather than rigorous experimental protocol)

- However the concept of uncanny valley has been adopted(and extended) in the field of Computer animation to adjust the human-likeness of a character's design to maximize public acceptance

  - Very realistic human appearances are now feasible in terms of shape, cloth, hairs, skin texture and lighting

  - BUT the quality of the associated animation must match the ***expected*** quality level for that level of verisimilar appearance

High Human sensitivity to human motion perception



Turing test for computer-generated movement (Hodgins et al ~1997)
Question: which one is synthesized from a model vs motion captured ?

Differences between the left and right movements :

- Variety:
  - temporal, style, texture, …
- Coherence of the behavior:
  - Synergy of the whole body involved in the behavior

Unsuccessful tradeoffs (feature films)

2001: Final Fantasy (Square)

Successful tradeoffs (films)

2010: Avatar(J. Cameron)

4. Motion capture is part of the solution for films

– High human-likeness can be recovered through motion capture provided that :

- Professional actors are hired for performance
- The actors learn text and performs as if they were filmed
- The actors are native speakers of the language
- The mocap session is also video recorded - from many viewpoints - to recover subtleties that cannot be measured
- Capturing **eye motions** is essential for the coherence of the synthesized behavior (http://www.mocaplab.com/services/eye-mocap/eye-tracker/)

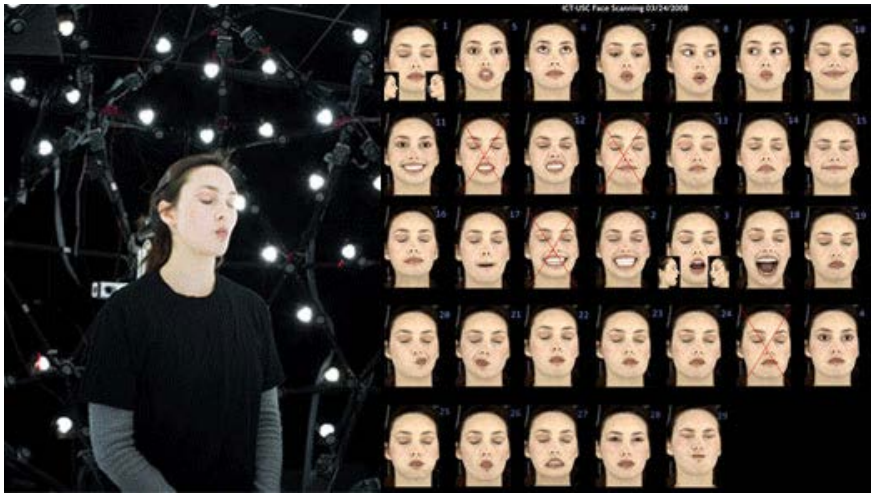
- Capturing **micro-expression**s is a must for the expression of emotions
  - Check the TV series “lie to me” & the youtube ref on micro-expressions

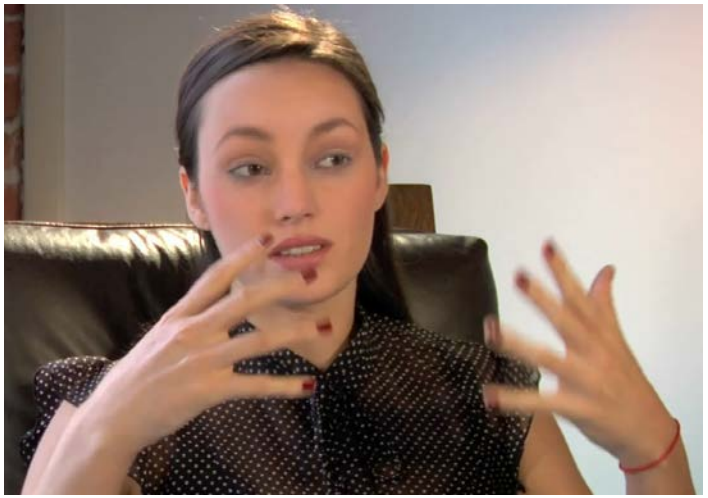
V

2010: Avatar(J. Cameron)

4. Motion capture is part of the solution for films (2)

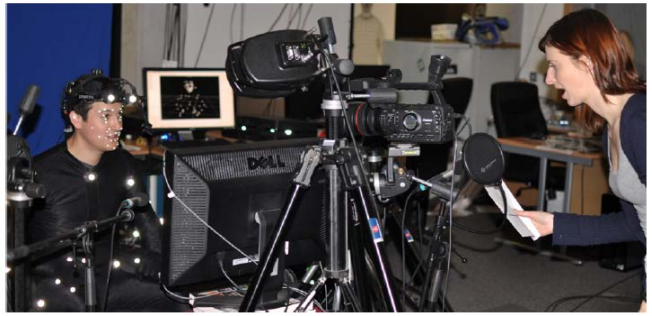
- Alternate motion capture technology based on Computer Vision :
  - Interview presenting Image Metrics technology (2008) [youtube / Emily / Advertizement]





http://www.youtube.com/watch?v=JF_NFmtw89g&feature=fvwrel

- Numerous on-going studies to assess the influence of rendering [McDonnell[2012]:



No simple mapping between the degree of realism and appeal/familiarity/friendliness

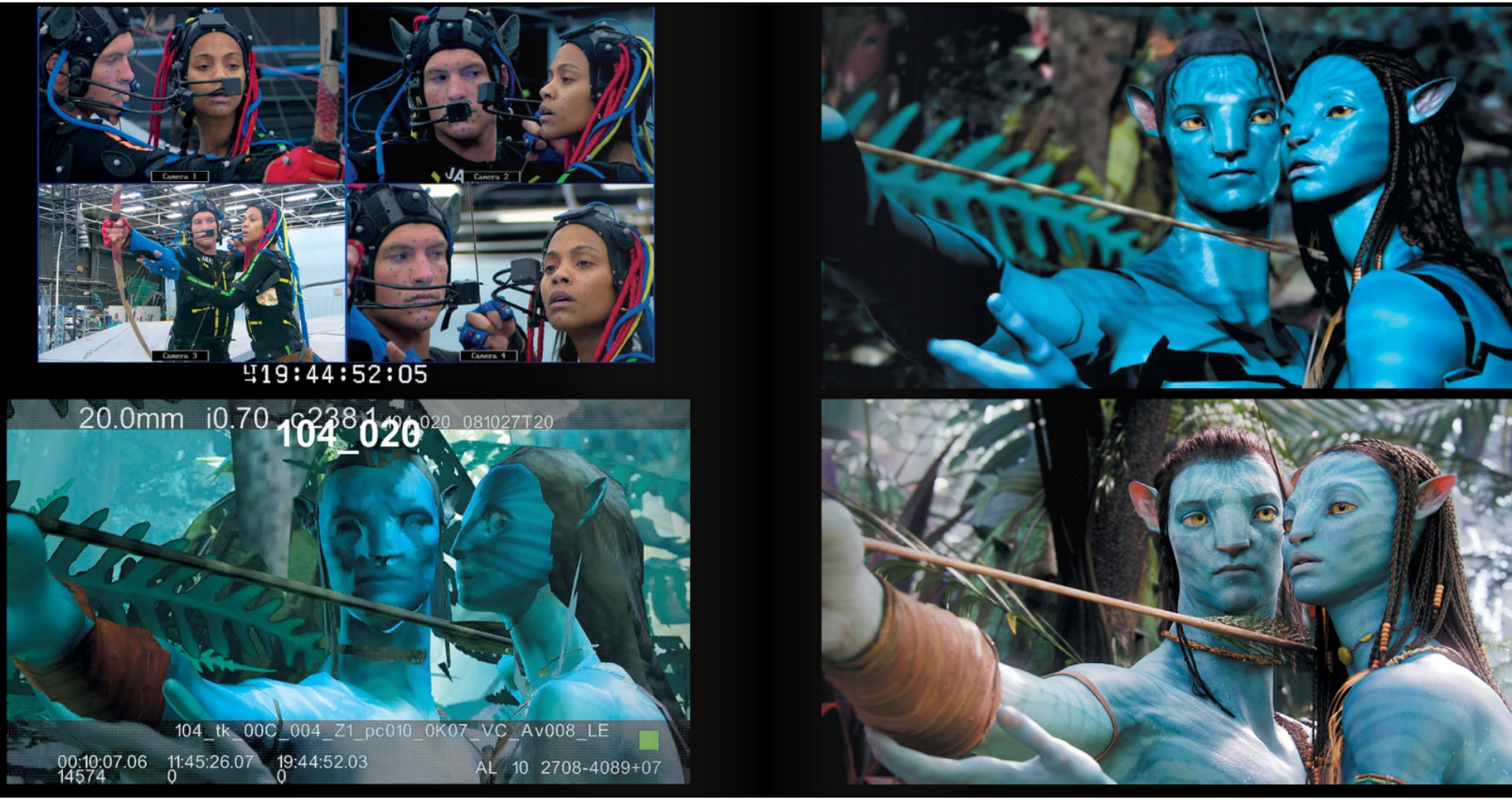
4. Motion capture is part of the solution for films (3)

– However, a very high resolution of facial meshes is not compatible with real-time display in VR, such as the *"swing cam"* concept introduced by James Cameron at the shooting stage to design camera trajectories.

[Cinefex on-line edition 2010]

4. Motion capture is part of the solution for films (3)

– However, a very high resolution of facial meshes is not compatible with real-time display in VR, such as the *“swing cam”* concept introduced by James Cameron at the shooting stage to design camera trajectories.



[Cinefex on-line edition 2010]

5. Perception of real-time animation

The purpose of perception studies is to determine two tradeoffs regarding CPU/GPU use.

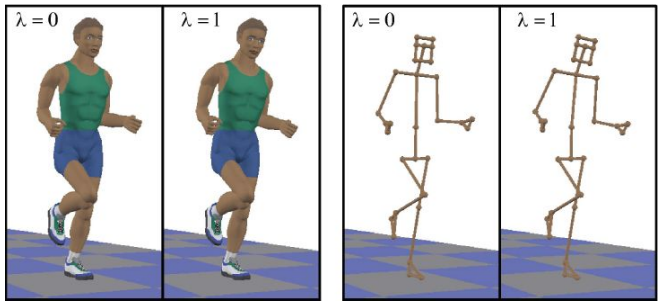
Context: a few **ms** to update the state of Virtual Humans

- **Uncanny valley:** matching animation quality with mesh resolution
  - Rationale: use only a VH degree of realism that can be supported by the available animation resources.
    - Don't add mobile accessories if they cannot be animated, such as long hairs, ear rings, floating pieces of cloth, etc…

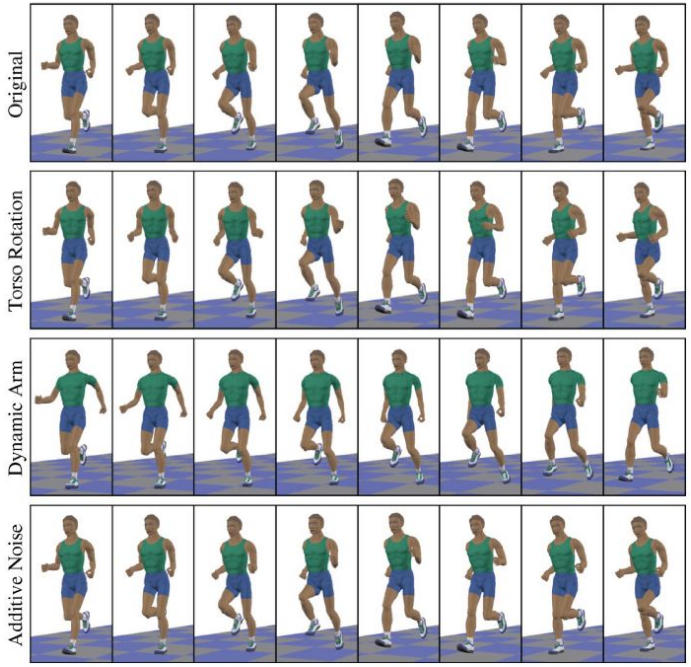
- **Compute what you see:**
  - Rationale: do NOT compute what is NOT perceived.
    - Levels of Details: decrease the resolution of human graphical models as distance increases to reduce display cost and simplify the movement to reduce animation cost.

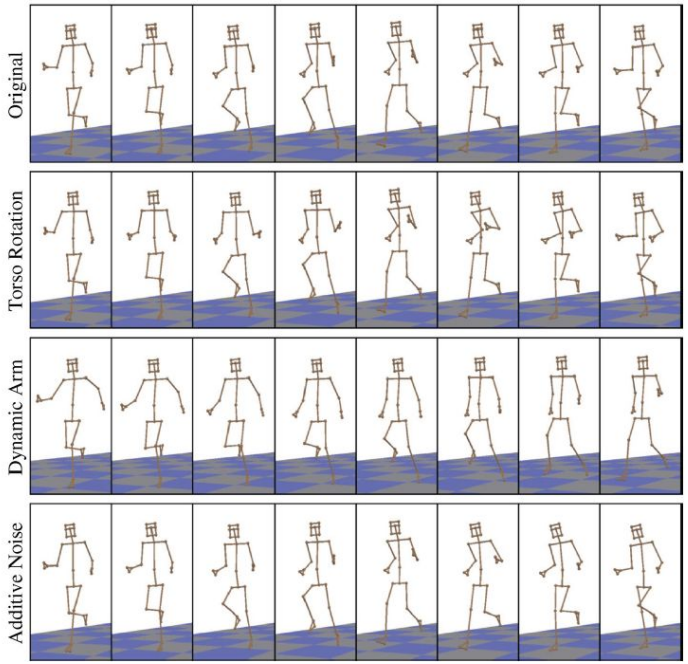
5. Perception of real-time animation (2)

In 1998; Hodgins et al showed that the geometric model type used to represent the human affected people's ability to **perceive the difference between two human motions.**



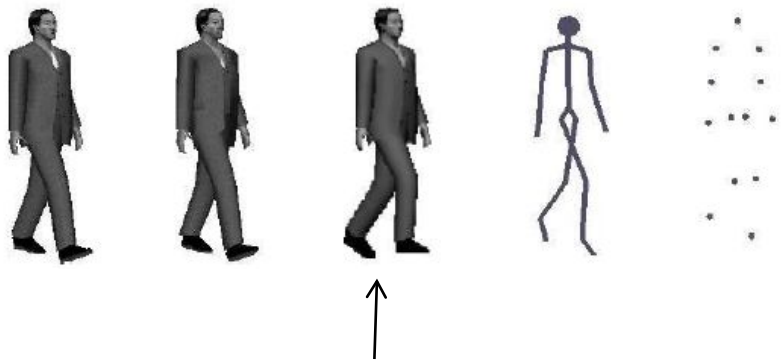
Subjects were more able to tell the difference between 2 motions when they were displayed on the polygonal character.



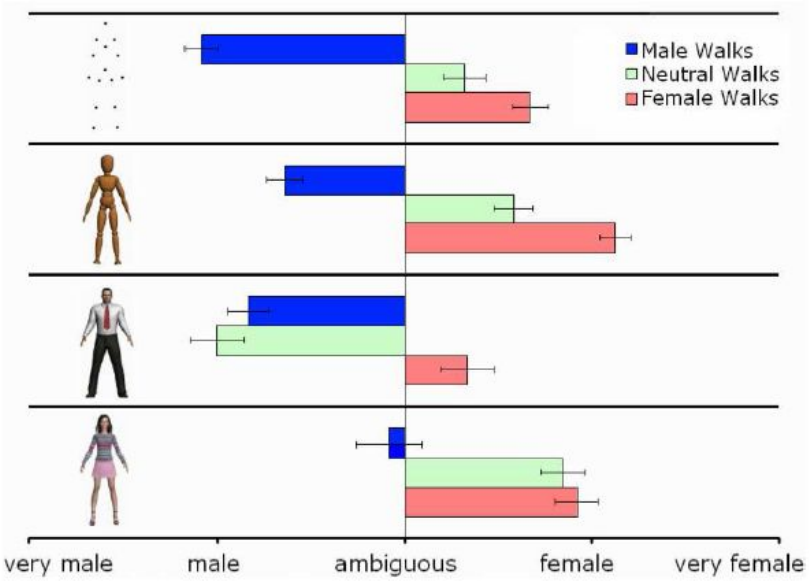


5. Perception of real-time animation (3)

• People are most sensitive to differences in human motions for high-resolution geometry (2022 pol) and *impostor* (i.e., image based rendering) representations, less sensitive for low resolution geometry (800 pol) and stick figures, and least sensitive for point-light representations [M 2005].



Impostor = 17x8 precomputed texture from high resolution geometry



Hodgins, O'Sullivan, Newell, McDowell [M 2007] found that:

- The graphical model may alter the perception of walking style (e.g. neutral).
- Gender-specific style should not be used for the other gender.

5. Perception of real-time animation (4)

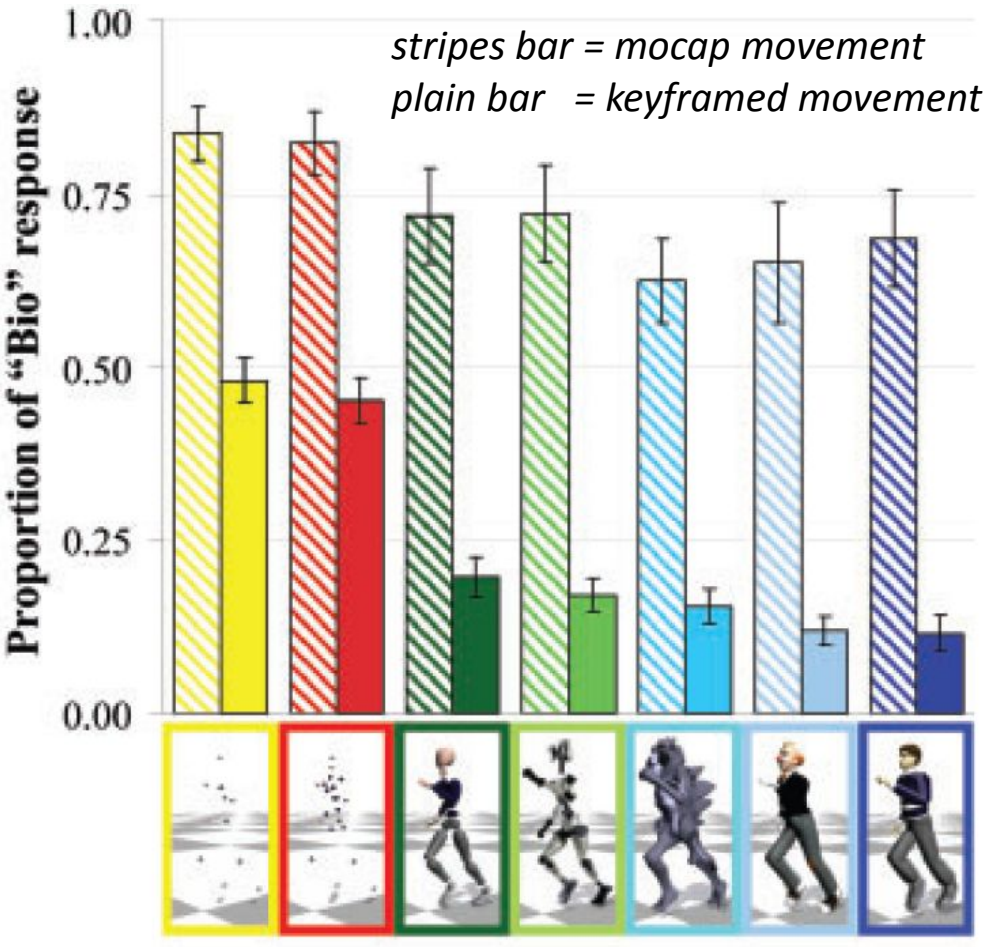
In 2007, Chaminade et al. investigated how the appearance of computer animated characters influenced the perception of a running movement.

Task: indicate whether a running motion is ***biological*** or ***artificial***

Setup: 4 sessions (7 minutes) x 7 characters x 6 motions (1 s)

Results:

- Bias: subjects are more inclided to perceive a ***biological*** motion for simplified characters.
- Motion rendered with anthropomorphic characters are perceived as less natural.
- Emotion is not involved (fMRI)



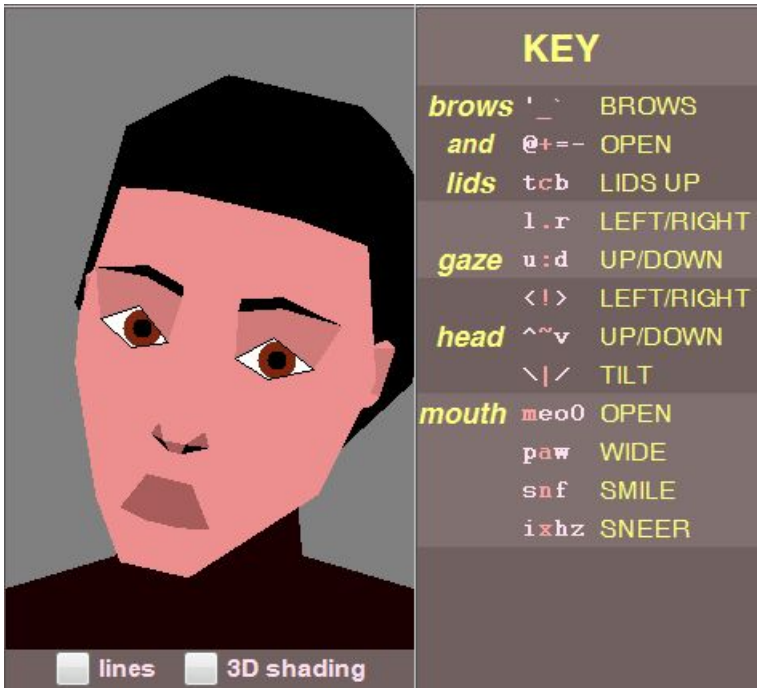
6. Core real-time VH believability factors (1)

- The first key factor is "**animation**" :
  - from latin word "**anima**" : animal life, breath, soul,mind
  - Hence the Virtual Human **MUST NOT BE STILL** otherwise it appears at best as a statue or worse as a dead body.
  - Movement can be procedurally generated or re-synthetized from captured movement through motion graphs [vW 2010]
  - Many commercial chatterbots, e.g. from Virtuoz (FR): http://www.ameli.fr/assures/index.php
    (USA) http://sitepal.com/howitworks/

6. Core real-time VH believability factors (2)

- Minimal animation while "waiting":
  - Breathe gently : sine wave in the spine at the thorax level
  - Eye blinking (5 to 20 /min)
  - Gentle random head movements, possibly coordinated with gaze
  - Gentle balance swaying if standing, possibly with idle movements

- Face demo from K. Perlin:



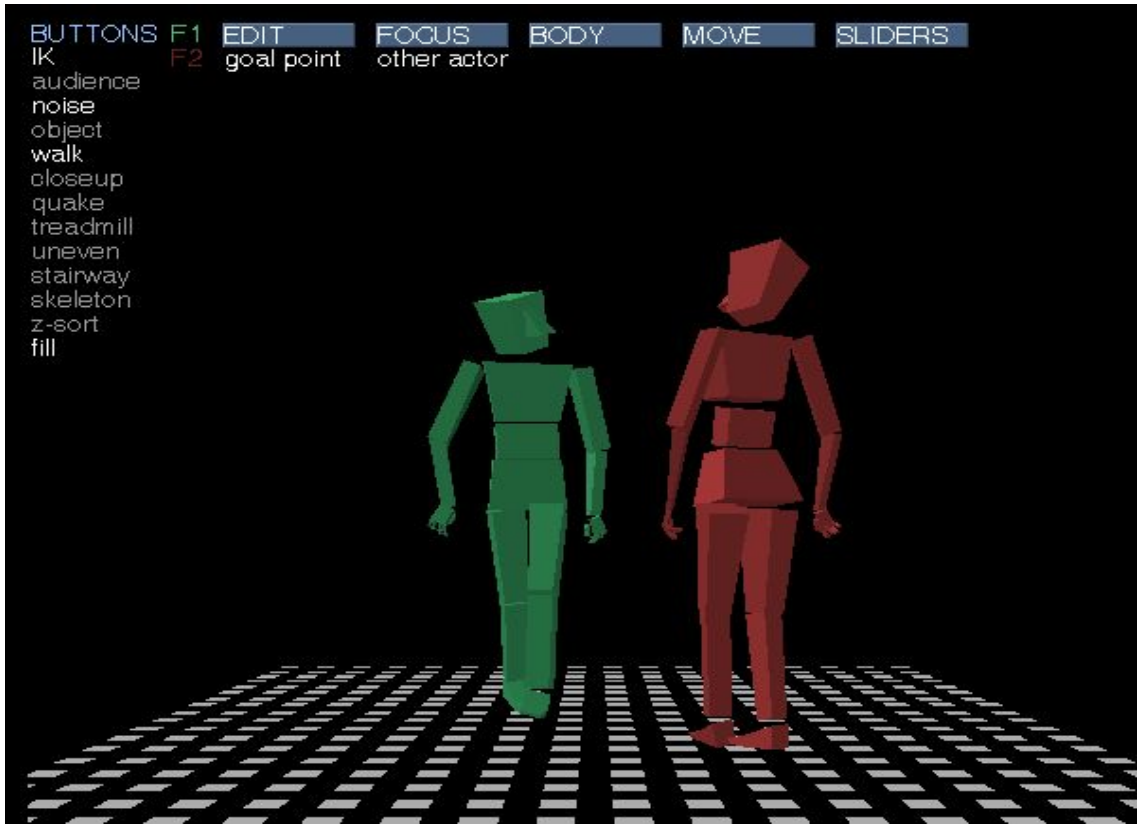
http://www.mrl.nyu.edu/~perlin/

6. Core real-time VH believability factors (3)

- Animation has to be coherent with the second key factor : **interaction**, i.e. being responsive to user input [TVR], including :
  - Plausible ***speech understanding & generation*** *: minimize delays*
  - **Must be coordinated**: facial expressions, head movement and eye gaze
  - Gestures: handle or precompute ***transitions*** between prerecorded gestures instead of sequences of gestures that always start and end with the same neutral posture
  - continuous flow of **idle movement** when not actively interacting
  - Handle eye contact with care: *gaze to express the wish to speak [K2014]*
  - Emotion display is application-dependant: happiness, surprise, interest, smile is generally a safe default.
  - If possible, subtle **mimicry** of the user head movement by the virtual human (e.g. with 4s delay) produces social influence but it backfires if detected because considered as a form of deception [Bailenson 2008]

6. Core real-time VH believability factors (4)

- Key contributor in expressive procedural RT characters: Ken Perlin (NYU)
  - Known for the “Perlin noise” for generating low cost textures
  - Applied the perlin noise to produce a continuously smooth movement
    - Emotive Actors demo:



- Principle of Perlin noise:

- Add noise functions with decreasing amplitude as frequency increases:

  F= 1 Hz, amplitude: 128

  + F= 2 Hz, amplitude: 64

  + F= 4 Hz, amplitude: 32

  + etc..

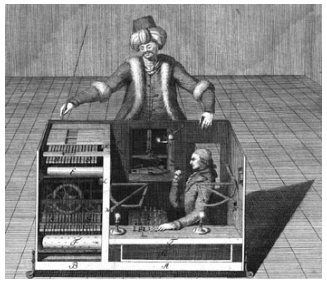
- Smooth/interpolate the result to produce in-between frames at display rate (20 to 60 Hz)

- More at [PerlinNoise]

http://mrl.nyu.edu/~perlin/

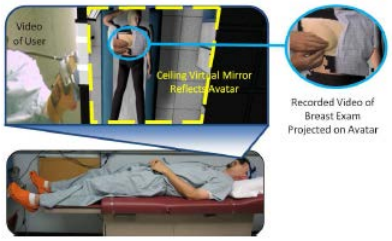
6. Core real-time VH believability factors (5)

- integrate a ***hidden operator*** when real-time constraint prevent the synthesis of sufficient quality movement or social experience:



- Performance animation for animating a synthetic character in TV shows or theme parks to interact with the public.
- Mechanical Turk (inspired by a false chess automaton from the XVIII century), e.g. teleoperated realistic puppet of Hiroshi Ishiguro (see uncanny valley slide) for fairs, theme park, etc...

- Wizard of Oz (inspired by the novel from F. Baum), e.g. for scientific experiments or training of complex social skills: the operator select predefined actions, sentences, behaviors etc based on the instantaneous user input (cf Presence course).



- in case a touch or haptic feedback is also needed, the VH should be collocated with a tangible interface, e.g. in [R 2009] a physical mannequin is manipulated by the trained medical doctor (e.g. for a breast exam) while seing a VH patient in a HMD.

7. Other R&D efforts

- Other academic groups involved in RT Autonomous VH:
  - INRIA-BUNRAKU/ Golaem (FR) : normalized postural control, Behavior
  - Paris-Tech (FR) : speaking agent GRETA, Catherine Pelachaud
  - Grenoble GIPSA-lab: Prosody & emotions, Gérard Bailly, Rémy Ronfard
  - DFKI (DE): Thomas Rist, Michael Kipp
  - UK teams: Ruth Aylett, Marc Cavazza
  - Other US teams: Justine Cassell, Andrew Cowell, Ari Shapiro

- Industrial solutions:
  - Numerous full body 3D assets available with UNITY3D (e.g. MORPH3D MCS: Morphable Character System, Mixamo)
  - Web site characters focus on spoken interactions with “chatterbots”: often limited to a 2D/3D speaking head/torso
  - coupled with text understanding and Text-To-Speech tools
  - Heavy trend of integrating an emotional dimension
  - Highfidelity.io is an on-going VR-upgrade of Second Life

7. Exercise (1): spot key factors in this RT demo

Real-time spoken interaction demo from the EU project SEMAINE “the sensitive agent project” involving Paris-Tech, DFKI, Imperial College, QUB, TUM, Univ. of Twente (2010):

7. Exercise (2): spot key factors in this RT demo

Example of 3D avatar mediating text-based communication

[prototype software from the **CyberEmotions** EU project]

https://www.youtube.com/watch?v=UGbW8nDNO24&feature=youtu.be

Purpose : express the emotions that is conveyed by the **text messages** through facial expressions and body language (but no sound).

Question: what are the key factors of believability ?

7. Exercise (3): spot key believability factors

Gallery of chatterbot demos from Sitepal.com
http://www.sitepal.com/howitworks/
http://content.oddcast.com/vhss/vhss_v5.swf?doc=http://vhss-d.oddcast.com/php/playScene/acc=1194891/ss=1902652/sl=0&acc=1194891&bgcolor=0x&embedid=41c5a82f0286836d9bef315621d4e366

Commercial Library of full-body 3D characters from Rocketbox studio

https://www.youtube.com/watch?v=zIqtWivC4Hg

Morph3D

https://www.youtube.com/watch?v=csQoCBZ4gWA

Mixamo

https://www.youtube.com/watch?v=kPb6cF8rnB8

- Consider playing with the UNITY CyberEmotions demo from EPFL-IIG providing real-time facial expression with (symmetric or asymmetric) emotions :

http://iig.epfl.ch/page-40268-en.html

[References]

[Bailenson 2008]J. N. Bailenson, N. Yee, K. Patel, and A. C. Beall. 2008. Detecting digital chameleons. Comput. Hum. Behav. 24, 1 (January 2008), 66-87.

[H 1998] Hodgins et al.: Perception of Human Motion With Different Geometric Models, IEEE Transactions on Visualization and Computer Graphics, 4(4), 307-316

[K 2010] Kipp, M. , Multimedia Annotation, Querying and Analysis in ANVIL. In: Multimedia Information Extraction, M. Maybury (ed.), IEEE Computer Society Press, in press

[M 2005] R. Mc Donnell, S. Dobbyn, C O'Sullivan Optimising and Evaluating the Realism of Virtual Crowds: Perceptual Experiments and Metrics, in EG07 tutorial on crowd animation.

[P 1995] K. Perlin, “Real Time Responsive Animation with Personality,” IEEE Trans. Visualization and Computer Graphics, vol. 1, no. 1, pp. 5-15,Mar. 1995

[R 2007] A. B. Raij, K. Johnsen, R. F. Dickerson, B. C. Lok, M. S. Cohen, M. Duerson, R. Rainer Pauly, A. O. Stevens, P. Wagner, and D. Scott Lind, Comparing Interpersonal Interactions with a Virtual Human to Those with a Real Human,IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 13, NO. 3, MAY/JUNE 2007

[R2009] A. Raij et al, Virtual Experiences for Social Perspective-Taking, IEEE VR 2009

[TRV 2006] Traité de Réalité Virtuelle, Ed. P. Fuch, vol 2, chap 17, Eds A. Berthoz & J.L. Vercher

[W 2009] van Welbergen, H., van Basten, B.J.H., Egges, A., Ruttkay, Z., Overmars, M.H.: Real Time Animation of Virtual Humans: A Trade-off Between Naturalness and Control. In: Eurographics - State of the Art Reports, Eurographics Association, pp. 45–72 (2009)

[K2014] Kerstin Ruhland, Sean Andrist, Jeremy Badler, Christopher Peters, Norman Badler, et al.. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. Eurographics 2014 - State of the Art Reports, Apr 2014,Strasbourg, France. pp.69-91, 2014, <10.2312/egst.20141036>

[Web References]

http://spectrum.ieee.org/robotics/humanoids/hiroshi-ishiguro-the-man-who-made-a-copy-of-himself

http://en.wikipedia.org/wiki/Lie_to_Me : with Prof. Paul Ekman as consultant.

Doc on microexpressions : http://www.youtube.com/watch?v=k2rb7pAP7hk

Image Metrics: http://www.youtube.com/watch?v=JF_NFmtw89g&feature=fvwrel

Demo of the interacting agent: http://www.semaine-project.eu/

Web site of Prof. Ken Perlin: http://www.mrl.nyu.edu/~perlin/
[PerlinNoise] : http://freespace.virgin.net/hugo.elias/models/m_perlin.htm

[W] [http://en.wikipedia.org/wiki/Uncanny_Valley]