

What makes a Virtual Human Alive ?

- video1*
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 (offline/online)
 5. Perception of real-time animation
 6. Core real-time VH believability factors
 7. Other R&D efforts & exercises

1. Terminology: Avatar vs Agent

- **Avatar** : [W]
 - (from sanskrit): is a term used in Hinduism for a material manifestation of a deity
 - (computing): the graphical representation of a **human user**. In VR the avatar movement is expected to be partially or completely driven by the user body movement (similarly to [VTubers](#))
- **(Autonomous) Agent or (Intelligent) Virtual Human**
 - fully computer controlled
 - 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. coach, instructor, assistant, or audience in a public speaking to overcome such a phobia, etc...
 - either pre-scripted or reactive to the user action or actived through a Wizard-of-Oz approach (e.g. NPC)

*note: the recent AI buzz tends to misuse the word **avatar** for purely offline productions (e.g. meteo/news/tutorial video presenters or [virtual influencers](#))*

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>



ANVIL [K2011-17]

Video image analysis

The screenshot shows the ANVIL 4.7.8 interface. The main window displays a video of two men in suits. A gesture track is overlaid on the video, showing a hand moving from a starting point to an ending point. The 'Track: gesture.phase' window on the right lists attributes for the gesture phase:

- Track: gesture.phase
- Referenced track: gesture phase
- Time: 00:27:16 - 00:28:00 (21 frames)
- Attributes:
 - lexeme: Calm
 - handedness: 2H
 - path: straight
 - hand-height-1: chest
 - hand-body-dist-1: close
 - hand-radial-orient-1: front
 - elbow-inclination-1: outward
 - 2H-distance-start: (59x170 00:27:60) (127x163 00:27:60)
 - 2H-distance: (66x163 00:27:92) (156x146 00:27:92)
 - shoulders: (52x99 00:27:92) (190x113 00:27:92)
 - lex affil: garnicht

Audio track analysis

Words timing



Gestures timing

The screenshot shows the ANVIL 4.7.8 interface with the audio track analysis window open. The window displays a timeline from 00:21 to 00:32. The audio track is shown as a waveform. The pitch contour is shown as a red line with blue dots. The text track is shown as a table of words and their timing. The gesture track is shown as a table of gesture phases and their timing.

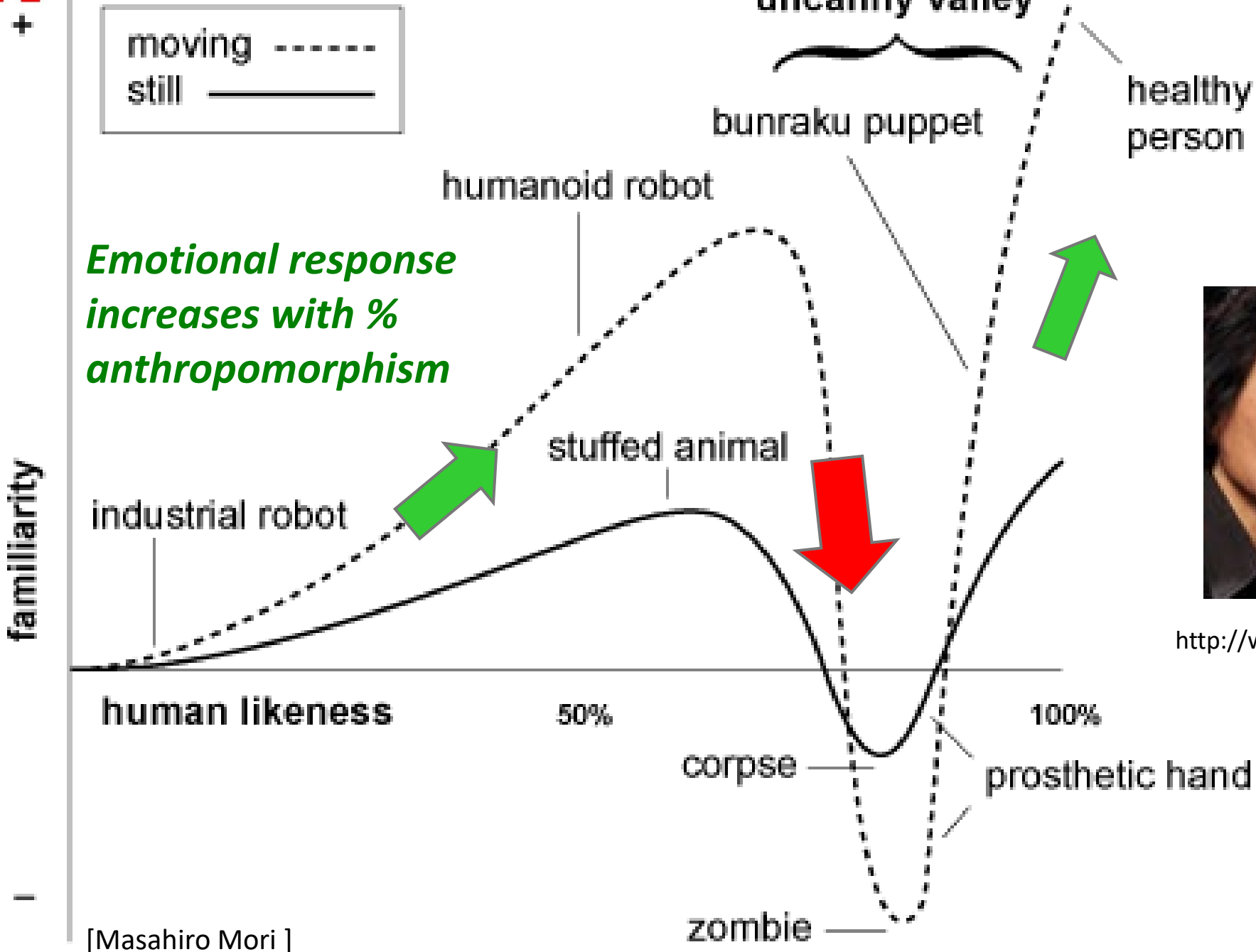
Time	Waveform	Pitch Contour	Text	Gesture
00:21	[Waveform]	[Pitch]	das Gemälde die Eidechsen...	stroke retract
00:22	[Waveform]	[Pitch]	w. damit gespielt wi...	prep s.p. stroke
00:23	[Waveform]	[Pitch]	da hat ei.. zweiten Boden das ist garni. flach u. simpel d. stimmt ab.. poetisch bezog ich au.	prep stroke pre. s. hold
00:24	[Waveform]	[Pitch]		Wipe, 2H

The screenshot displays the ANVIL 4.9 software interface. The central window shows a 3D skeletal model of a human figure with yellow markers at various joints. To the right, a video window titled 'Main Video: AntonyTired.mov' shows a person in a black motion capture suit performing a gesture. Below the video, a timeline window titled 'Annotation: AntonyTired.anvil' shows a series of colored bars representing different phases and phrases of the movement. The timeline includes a time axis from 00:02 to 00:08. The phases are labeled as 'prep', 'stroke', and 'retract'. The phrases are labeled as 'Cup, LH', 'Cup, RH', 'Cup, 2H', and 'Cup, LH'. The interface also includes a menu bar (File, Edit, View, Tools), a status bar, and a console window on the left showing debug messages.

Full-body Body
motion capture
& analysis

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 (in Japanese) has been translated by McDorman

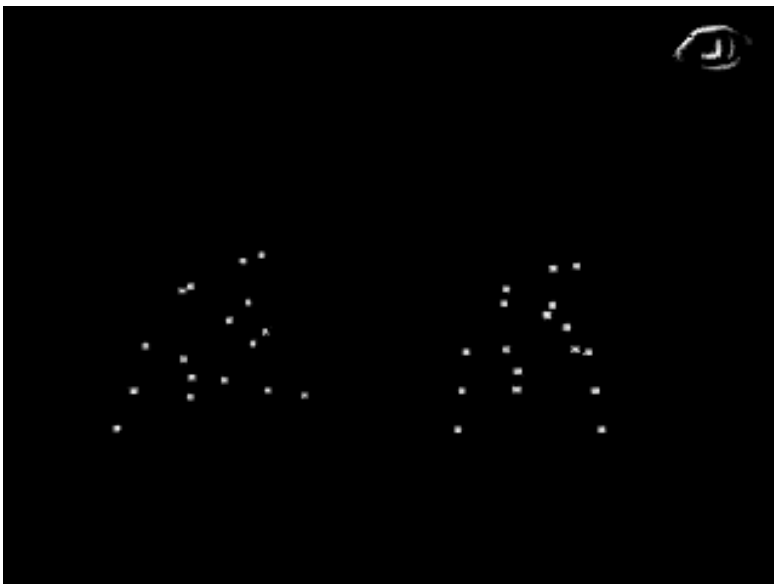


Hiroshi Ishiguro

<http://www.youtube.com/watch?v=uD1CdjlrTBM>

[Masahiro Mori]

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



High Human sensitivity to human motion perception

Turing test for computer-generated movement (Hodgins et al ~1997-98)

Question: which one is synthesized from a model vs motion captured ?

Differences between the left and right movements :

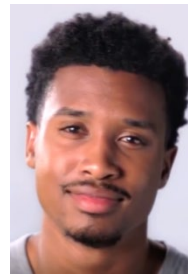
- Variety:
 - temporal, style, texture, ...

- Coherence of the behavior:
 - Synergy of the whole body involved in the behavior

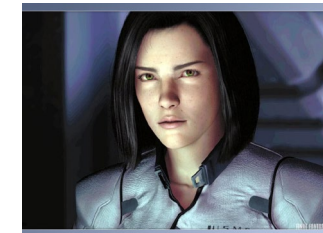
3. From artificial to real : the uncanny valley (3)

- The paper from Masahiro 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*



(Samsung NEON real-time Virtual-Human, CES2020), no update since 2020



Unsuccessful tradeoffs (films)
2001: *Final Fantasy* (Square)

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



Miquela Sousa : virtual influencer since 2016 [K2023] (among 35 others).

Successful tradeoffs (films)



2010: *Avatar* (J. Cameron)

[References]

[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

[K 2023] Kluger L. Virtual influencers in the real world, CACM march 2023,
<https://cacm.acm.org/news/virtual-influencers-in-the-real-world/>

[Web References]

<http://www.anvil-software.org>

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

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