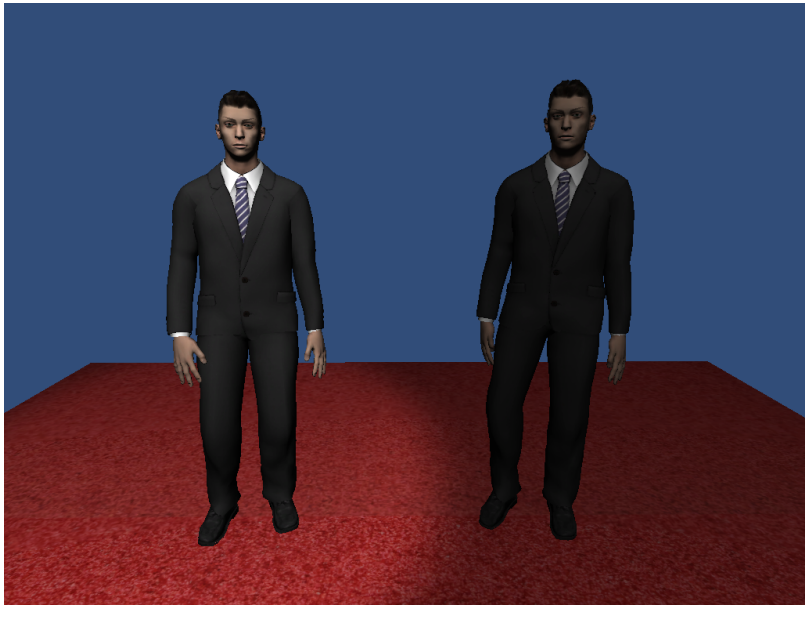
The web links in this slide deck were updated in 2024.

Impacting the social presence of virtual agents by scaling the fidelity of their speech and movement



Julian Fietkau

University of Hamburg

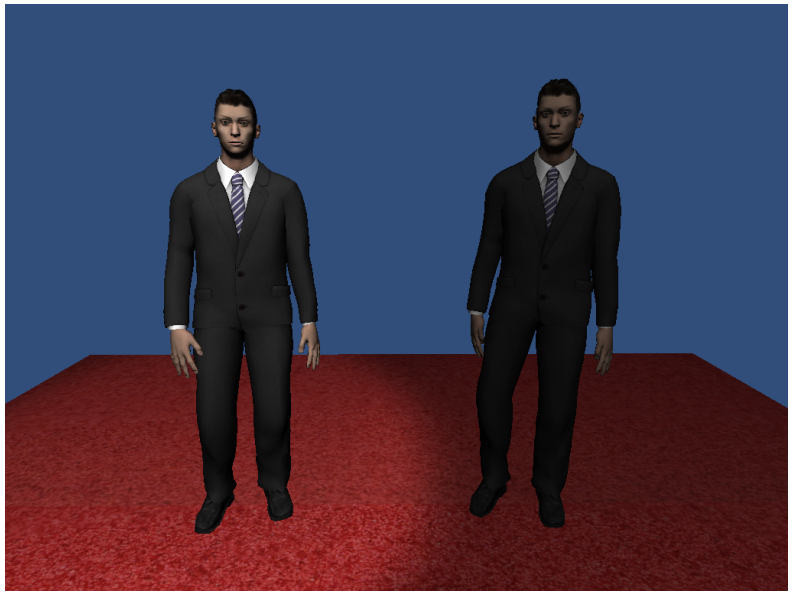
January 23, 2015

Overview

Overview

Virtual agents

- 3D model + sound (voice) + animation
- used in training/simulation, art, games, etc.



Fidelity of speech and movement

- speech may be
  - recorded in advance
  - produced by a text-to-speech system
  - somewhere in between
- movement may be
  - fully based on high-quality motion capturing data
  - partially based on algorithmic animation
  - nonexistant

The experiment does not deal with model or rendering quality.

Social presence

Based on an existing definition (Biocca, Harms, Gregg 2001):

1. **co-presence:** the belief that one is not alone, peripherally or focal awareness of the other
2. **psychological involvement:** the allocation of focal attention to the other person, response to emotional states, perceived insight into motivation and thought
3. **behavioral engagement:** interdependence and connectedness to the actions of the other person

Central question

How does the fidelity (of speech and movement) influence the perceived social presence?

Additional consideration: VR HMD

Why use a head-mounted display for this experiment?

- increased importance of HMDs and other immersive display technology in VR
- perception of social presence could reasonably be impacted by display technology – the HMD context has been explored less

Overview

Idea

- produce combinations of various realizations of the above-mentioned variables (speech and movement)
- display them in pairs
- let test subjects compare the social presence and say which one of a pair has more of it
  - do this a whole lot
- aggregate the results to look at the influence of each factor

Model and engine

- use Unity 4.x for rendering, scripting etc.
- 3D model and textures created with Makehuman 1.0

Speech

Consideration: The speech acts of the virtual agents should not carry any actual message content, because that would be distracting and different speech acts could have an influence on the perceived social presence (e.g. information vs. question vs. order). $\rightarrow$ use vocal gibberish

- take the 1000 most common words in German (source: "Wortschatz" project, Universität Leipzig)
- feed them into Wuggy, a pseudo-word generator (Ghent University, Belgium) to create random sentences that look/sound "believably German"

Gibberish used in the experiment

Putaun ehte pflon veßten düfflich La Fing hürte Kopp geripten Südchen Daude.
Kie Verpreils Hopitie Phraxe metes scheches krumciespiel Dimen wor klück Mozualiin Zaß.
Lychte rafen Fahl toswenden lält luchsgans gorm dadee Spresten ebstbals vesses Newage.
Sis fist Lab Wuderfet kühe Hamte veuten Läuen alny Bopie schäler belögte.
Allerlochs spöbten stekken hanuß bes Beren Rie fal rereis Piedes lanter dabbte.
Tonzerr for Turicht gopen Gander fürr jor nasen hührend rusband zusel Händern.
Vorkau hind nirgst ehka ätmehin umhächst zondern zöln giesen kolst begids Belsallem.
Gesprals Marf hillten fiesen Rottel zockte Jen arrhen peit rafe Wuloner zührend.

Audio assets

1. feed the gibberish into the IVONA text to speech software
2. imitate the IVONA output for a full recording of my own
3. manipulate the recording to reduce the fidelity for the in-between step

Movement

- use the Unity Mecanim example idle animation (high-quality motion capturing data) for the highest fidelity
- replace parts of the animation with stillness and inverse kinematics for the middle fidelity
- use the same initial pose but with no animation for the lowest fidelity

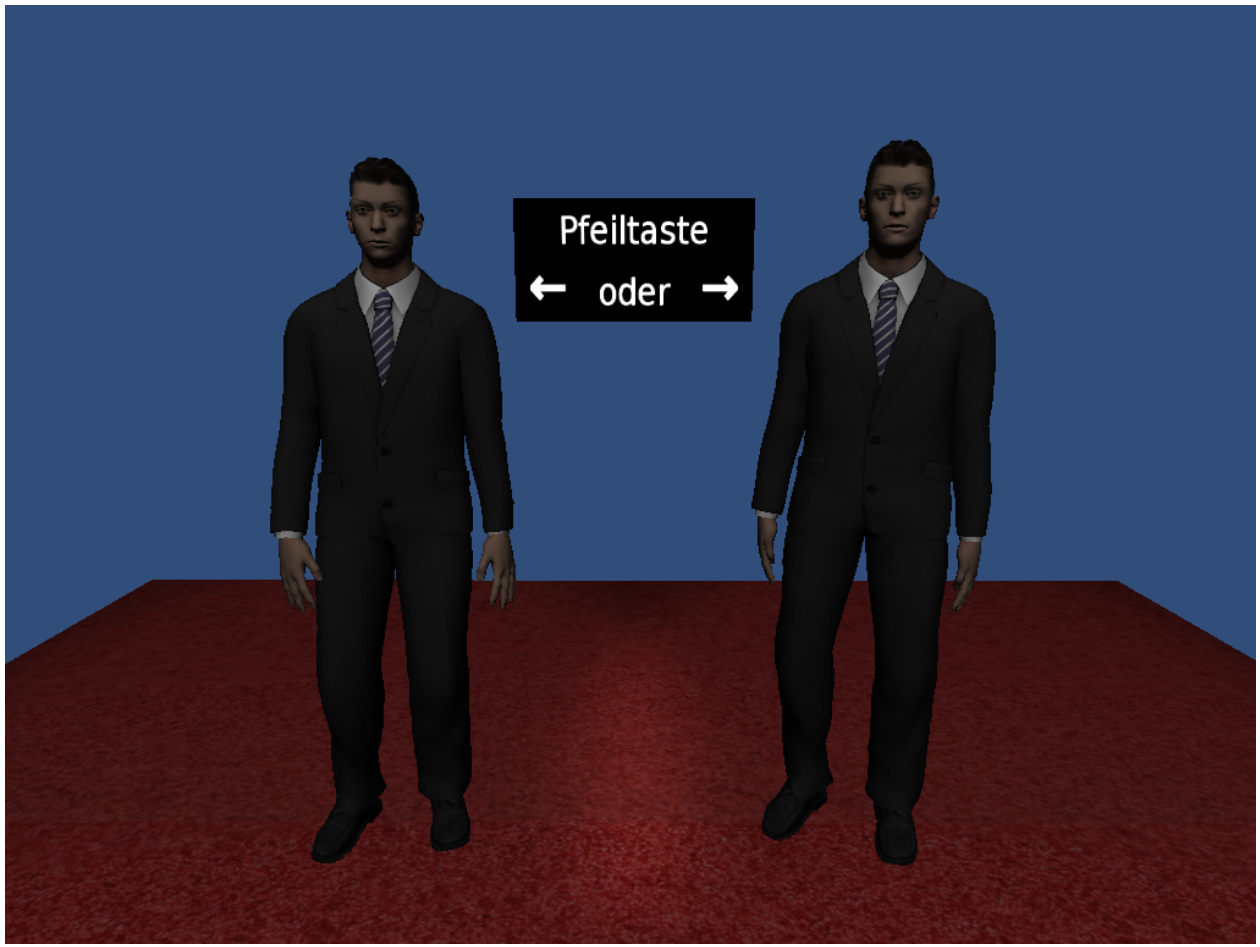
Animation: middle fidelity planning changes

The original intention was to use a manually created animation for the middle fidelity.

- could have been custom-created by a skilled animator (which I am not)
- would need some sort of "measure of similarity" with the mocap animation
- existing available idle animations tend to be tailored for games, not realistic movement

→ decision to repurpose and simplify the mocap animation instead

Realization of idea



Additional considerations

Other components of the experimental procedure:

1. surrounding questionnaire
   - general questions: age, height, gender etc., VR experience
   - potential health-related issues (sight and hearing)
   - The Lateral Preference Inventory (Coren 1993)
   - Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, Lilienthal 1993)
   - Slater-Usoh-Steed Presence Measurement Questionnaire (Slater, Usoh, Steed 1994)
2. short hearing assessment

Experiment

15 test subjects, 144 trials per subject,
9 invalid trials in total $\rightarrow$ 2151 comparisons

Overview

Results I: Choice duration (in milliseconds)

- min: 189, max: 14850
- median: 1031, mean: 1384, 94th percentile: 2957 (mostly short decision times)
- standard deviation: 1303

Results II: Fidelity preferences

- movement – mocap: 853, reduced: 785, still: 513
- speech – recording: 921, altered: 727, tts: 503

Combined pairs:

	tts	altered	recording
still	111	162	240
reduced	175	283	327
full	218	282	354

χ^2-Test with $p < 0.1$: null hypothesis (factors are independent) can not be disproved

Results III: Other influences on preferences

ANOVA of subject, gibberish sentence, display order influence on movement and speech preference:

- very significant influence ($p < 0.001$) of subject and sentence, slightly significant influence ($p < 0.1$) of display order on movement preference, no significant interactions between the factors
- very significant influence ($p < 0.001$) of subject, somewhat significant influence ($p < 0.01$) of sentence on speech preference, no significant interactions between the factors

Results IV: More to come

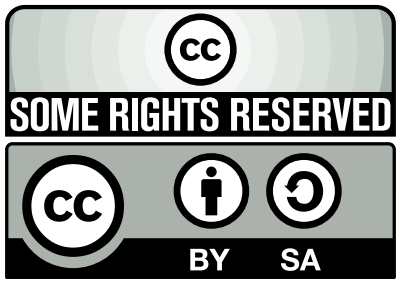
The statistical analysis is not completely finished yet.

Overview

Summary

- Both of the examined fidelity modes (speech and movement/animation) have a significant influence on social presence.
- Mocap beats reduced movement by a little, beats no movement by a lot.
- High-quality speech recording helps social presence significantly.
- There might be interaction effects between the two modes, but if so they are weak.
- High-quality speech recording helps social presence significantly.
- Technical fidelity (in terms of speech and movement) **is highly important for perceived social presence**.

Release and Download



Download and comment:

https://fietkau.science/social_presence_scaling_fidelity