If it Moves Like a Person, and it Talks Like a Person, it is Probably an Avatar

[1,2]Philine Link, [1]Duangkamol Srismith, [3]Maria Alejandra Quiros Ramirez, [1]Judit Vari, [1]Susumu Shikano, [3]Stephan Streuber, [1]Tamara Rathcke, [1,4]Diego Frassinelli

[1]*University of Konstanz,* [2]*Utrecht University,* [3]*Coburg University,* [4]*LMU Munich*

Background and Aims. Software for the creation of customizable, hyper-realistic virtual agents has recently become widely accessible. In scientific contexts, these avatars can offer researchers significant control and ecological validity while also ensuring the anonymity of human subjects (Tromp et al., 2018). Despite their high potential, their use in (psycho-)linguistics remains largely unexplored (Peeters, 2019). So far, most studies employing avatars have used either stylized, low-resolution human representations (Parmar et al., 2022; Thomas et al., 2022) or require complex 3-D scanners that are not widely accessible. Additionally, they often do not disclose the software used or the procedures adopted. The goals of the present study are two-fold: (1) develop and describe a pipeline that uses existing software and a smartphone camera to produce high-fidelity human representations within hours; and (2) test whether these avatars can successfully substitute human stimuli in behavioral studies.

A Pipeline for Stimuli Creation. Using a smartphone camera, we recorded one-minute videos of nine native English speakers (five females) talking about climate change. To create the avatars of the speakers, we used the following software: 1) *RealityCapture*, to create a 3D mesh model of participants' faces; 2) *Unreal Engine 5.0 with MetaHuman Creator*, to automatically generate an avatar from the mesh model; 3) *Faceware Studio*, to add facial animations; and 4) *DaVinci Resolve 18*, to edit the audio-visual signal. The resulting avatars closely resemble their human donors (see Figure 1).

Evaluation Methods. In a rating pre-test, 6 participants confirmed the high similarity (on average 91%) of avatars to their donors (Figure 1). During the main online study, 72 English participants assessed personality characteristics and behavioural traits of the original human videos, avatar videos, and the audio signals extracted from the recordings. The stimuli were presented in a Latin-square design, with each participant rating three individual speakers per condition (human, avatar, audio) on nine 8-point Likert scales (Figure 2). All participants were recruited online via Prolific and reimbursed for their efforts.

Results and Conclusions. As shown in Figure 2, audio-only stimuli received the highest ratings on all scales, followed by human videos. Avatar stimuli were rated significantly lower than the human videos, though there was a high positive correlation between the ratings of all stimuli (Table 1), indicating that the overall relationship between the ratings of human and avatar videos was stable. We conclude that avatar-based methodologies are very promising tools in multimodal behavioral research as they allow a precise manipulation of the different extra-linguistic components supporting language processing.

Figure 1: Example of two humans (donors) and the corresponding generated avatars.

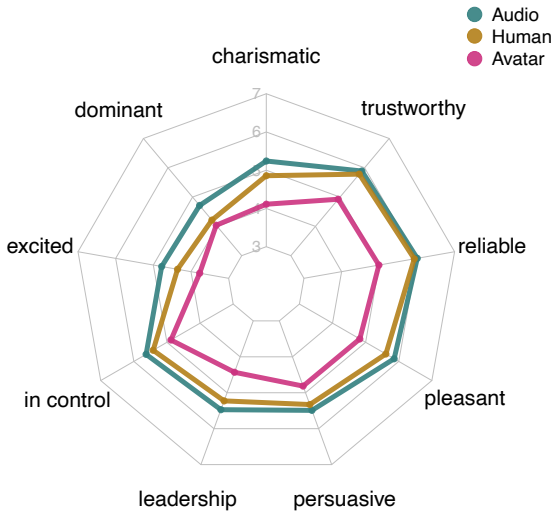

Figure 2: Zoom in on the average ratings (3–7) of each behavioral dimension across stimulus types (audio vs. human vs. avatar). A Linear Mixed-Effects Model analysis with `donor` and `participant` as random intercepts shows that the differences in scores between the three conditions are all highly significant ($p < 0.001$).

Condition	Coefficient
Human vs. Audio	+0.96 ***
Avatar vs. Audio	+0.84 ***
Human vs. Avatar	+0.91 ***

Table 1: Spearman correlation coefficients (ρ) between the average scores obtained for the three experimental conditions across all scales (cf. Figure 2); significance level is $p < 0.001$.

References

Parmar, D., Olafsson, S., Utami, D., Murali, P., & Bickmore, T. (2022). Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Auton. Agents Multi-Agent Syst., 36*(1), 17.

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychon. Bull. Rev., 26*, 894–900.

Thomas, S., Ferstl, Y., McDonnell, R., & Ennis, C. (2022). Investigating how speech and animation realism influence the perceived personality of virtual characters and agents. In *2022 IEEE VR* (pp. 11–20).

Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behav. Res. Methods., 50*, 862–869.