

Do humans distinguish deepfake from real vocal identity? Insights from the perceptual and neurocognitive system

Claudia Roswadowitz

Department of Psychology & Department of Computational Linguistics, University of Zurich, Zurich, Switzerland

claudia.roswadowitz@psychologie.uzh.ch

Deepfakes artificially re-create and manipulate original human data, with the main purpose of spreading social and political misinformation. High-quality deepfakes are viral ingredients of digital environments, and they can trick human cognition into misperceiving the fake as real. However, experimental research on how the human neurocognitive system processes deepfake information has been greatly neglected so far. In this talk, I will present perceptual and neuroimaging data on the sensitivity of the human brain to detect or be deceived by instances of deepfake voice identities. By using advanced deepfake technologies, we created voice identity clones that are acoustically like the natural human voices. During an identity recognition task, humans were mainly deceived by deepfake voice identities, but showed some remaining resources for deepfake detection. On the brain level, we identified a potential „deepfake sensor” including the subcortical ventral striatum, which assigns social reward to natural but not to deepfake identities, and sensory auditory cortex evaluating the acoustic degree of artificiality in human utterances. With our study, we present neurocognitive findings on the potential but also limitations of emerging deepfakes as artificial social signals for humans. Our findings highlight the relevance of the reward level of social cues for successful and effective human-computer interactions.

