The uncanny valley in virtual reality – the impact of categorical ambiguity and mind attribution on emotion elicitation

We live in a rapidly changing world in which new technologies affect our everyday functioning to a large extent. We are surrounded by more and more characters imitating people, with whom we often have to make contact, consider for example Google assistant listening to us, robot from Boston Dynamics carrying loads, Pepper robot keeping the elderly company, or a character in a computer game asked about the direction. The – so called – Uncanny Valley Hypothesis concerns the emotions experienced during a contact with such human-like characters. It suggests that contact with characters very similar to a human, causes a feeling of discomfort. Since there are more and more humanlike characters around us, it is important to know what exactly causes the discomfort and how to diminish such effects if needed.

The goal of the project is to test two explanations of the Uncanny Valley – based on the categorization of objects and on the attribution of Theory of Mind. Two experiments using virtual reality will be carried out. In both of them, the subjects will find themselves in a virtual café together with robotic, almost-human or human characters.

The first experiment focuses on categorical ambiguity, an explanation that refers to the difficulty in assessing whether the character belongs to the robot or to the human category. The experiment is designed in such a way that it will imitate a real meeting, giving the opportunity to check whether almost-human characters evoke negative emotions. The second experiment focuses on an explanation based on the Theory of Mind. Some of the studies indicate that the effect of eeriness (fear or disgust) occurs when the subjects attribute the ability to *experience* human emotions to the artificial characters. However, it has not been explained how attribution of the *experience* affects subjects' emotions by association with visual humanlikeness. The aim of the experiment is to check whether there is a boundary on the spectrum of humanlikeness after which the attribution of human emotions evokes the Uncanny Valley or whether these two variables (humanlikeness and mind attribution) are unrelated. To assess the emotions experienced by the subjects, we will compare the answers to questionnaires about their emotions with psychophysiological measures (more precisely: the electrodermal activity and electrocardiographic data).

We expect that the closeness of humanlike characters will translate into a stronger effect of the Uncanny Valley and will allow us to broaden the explanations for what causes negative emotions. The results of the research will bring a better understanding of the phenomenon and emotions related to characters imitating humans, and also will improve the accuracy of future research on human-robot interaction topic.