Секция «Когнитивная психология»

Neurocognitive processing of attitude-consistent and attitude-inconsistent deepfakes: N400 study

Монахова Элиана
Student (master)
Национальный исследовательский университет «Высшая школа экономики», Факультет социальных наук, Москва, Россия
E-mail: eliana98@mail.ru

Nowadays the mass distribution of fake content has acquired a significant scale and spread to various topics. One of the modern varieties of fake representation is the technology for synthesizing video and audio formats called deepfake. With the development of neural networks, such new method of transforming faces and voices has become very popular [6]. Certainly, a novel area of interest attracts scientists of different directions, involving neuroscientists and cognitive psychologists, but the amount of research on deepfake materials in this field cannot be called exhaustive. Hence, a few papers are devoted to specific neurophysiological correlations in deepfake format processing.

The current study, in turn, involves completely new attitude-consistent and attitude-inconsistent audial deepfakes dedicated to the topic of vaccination against the COVID-19 virus in Russia. The author analyzes electrophysiological brain response towards such deepfakes manifestation and observes whether congruence or incongruence of internal attitudes and the degree of analytical thinking influence the level of trust to deepfakes. The selected stimuli represent two opinion leaders covering the issue of vaccination against COVID-19 virus. The main distinction between real and deepfake materials is that the neural network copies the actors' voices and generates artificial speech recordings, broadcasting opposite positions to the topic under discussion. Thus, participants will be misled by the realistic materials and will probably believe in the presented information due to low level of analytical thinking and congruent attitudes towards COVID-19 vaccination topic.

Concerning methodology, the study involves several questionnaries for the behavioral part after the main experiment (such as Cognitive Reflection Test (CRT) [2], Need for Cognition Scale (NFCS) [1], Likert scale [4] and Conformity Scale [5]) and EEG technology for the N400 component observation [3]. The neural hypothesis of the study suggests that N400 amplitude will reflect larger negativity of audio deepfakes that are incongruent to the participants' opinion and contradict celebrity's public opinion. The behavioral one, in turn, reflects that people tend to believe more in those deepfakes whose attitudes correspond to their own, regardless of whether they coincide with the public opinion of the deepfake celebrity. The study comprises two participant groups with opposite attitudes towards the COVID-19 vaccination issue. As for the materials, the research includes 2 audio deepfake materials representing Russian celebrities, switching roles, and broadcasting the opposite opinion to public one about the vaccination. Each deepfake contains 40 experimental phrases with unexpected endings and 40 control phrases for the N400 component elicitation, expected at the sentence-final target word, if it contradicts participant's attitudes. The evoked response between conditions will be analyzed by a three-way ANOVA method, involving groups, congruence, and electrodes factors, the N400 amplitude as a dependent variable. In addition, participants will likewise pass classical N400 component test with written sentences containing unexpected endings as a control condition.

Concerning the expected results, the author intends to observe a higher level of trust to the deepfakes, whose views coincide with the participants' internal attitudes, regardless of whether they correspond to the public view of the deepfake's character, having a significant correlation with their ability for analytical thinking. What is more, it is expected to observe the stronger N400 in the region of the chosen electrodes in the deepfake sentences that do not correspond to the public opinion of the celebrities, particularly when the statement mismatches the attitudes of the participant. Overall, such research will form a deeper understanding of whether the congruence of internal attitudes and the level of analytical thinking affect the degree of trust in deepfake materials, as well as identify its neurophysiological correlate with the N400 component to the unexpected congruent or incongruent material.

References

1) Cohen A. R., Stotland E., Wolfe D. M. An experimental investigation of need for cognition //The Journal of Abnormal and Social Psychology. – 1955. – Т. 51. – №. 2. – С. 291.
2) Frederick S. Cognitive reflection and decision making //Journal of Economic perspectives. – 2005. – Т. 19. – №. 4. – С. 25-42.
3) Kutas M., Federmeier K. D. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP) //Annual review of psychology. – 2011. – Т. 62. – С. 621-647.
4) Likert R. A technique for the measurement of attitudes //Archives of psychology. – 1932.
5) Mehrabian A., Stefl C. A. Basic temperament components of loneliness, shyness, and conformity //Social Behavior and Personality: an international journal. – 1995. – Т. 23. – №. 3. – С. 253-263.
6) Thies J. et al. Face2face: Real-time face capture and reenactment of rgb videos //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 2387-2395.