INTERNATIONAL JOURNAL OF EDUCATION AND INFORMATION TECHNOLOGIES
DOI: 10.46300/9109.2025.19.12 Volume 19, 2025

Mirror Neurons cannot be Fooled by Artificial Voices – a study with Implications for Education using Magnetic Resonance Imaging (MRI) and Convolutional Neural Network (CNN)

Costin Andrei Bratan[1,2], Andreea Marinescu[3], Elena Terecoasa[3], Ana Voichita Tebeanu[1], Bogdan Morosanu[1], Eduard Franti[2,4], Monica Dascalu[1,2], Alexandra Andrei[1], Claudia Tocila-Matasel[5], Bogdan Ionescu[1], Gheorghe Iana[5], Ana Maria Oproiu[3], Gabriela Iorgulescu[3],
[1]Politehnica University of Bucharest, Romania
[2]Research Institute for Artificial Intelligence "Mihai Drăgănescu", Bucharest, Romania
[3]University of Medicine and Pharmacy „Carol Davila", Bucharest, Romania
[4]National Institute for Research and Development in Microtechnology, Bucharest, Romania
[5]Medima Health, Otopeni, Romania

Received: June 9, 2024. Revised: March 17, 2025. Accepted: May 21, 2025. Published: July 30, 2025.

Abstract— Mirror neurons have a crucial role in detecting and reproducing the actions of others as if the observer himself were performing the specific action. In this paper, four different audio voice files are used to determine, in two methods, the idea that the mirror neurons can be activated only by a real human original voice with a strong emotional load. The first method is the magnetic resonance imaging (MRI), which gives information about the brain activity regarding the specific areas where the mirror neurons are located when the four different audio files are listened to by a group of ten volunteers. The second method implies a deep learning approach, using two convolutional neural network (CNNs) architectures, one used to recognize the timbre of the audio speaker and the second one to determine the level of remnant (residual) emotion in the audio files listened to by them. The four audio files used are an audio text recorded by a Romanian actress with a specific emotion, two different actresses' voice recordings with the same text and emotion, and with very similar voice features to the main actress, and the last one is an artificially generated voice using AI algorithm. The results show a promising response from both perspectives- the hypothesis that mirror neurons can't be fooled by an artificial voice is confirmed, and that the intensity of emotion is higher in the original voice than the two imitating voices.

Keywords— artificial voice for education, convolutional neural network, emotion intensity detection, learning outcomes, magnetic resonance imaging, mirror neurons, timbre recognition.

I. INTRODUCTION

The advances in the field of AI, in recent years, have allowed the development of highly performing artificial voices similar to human voices. There are already software packages on the market that generate artificial voices almost identical to the voices of famous politicians or actors, such as Donald Trump, Joe Biden, Morgan Freeman, or, why not, Gabriela Bobes, our colleague and co-author. Although these voices have almost identical acoustic characteristics to the voices they copy, they fail to convey as intense emotions as the original voices. This limitation has been analyzed by many researchers in the hope of finding solutions for artificial voices to equal the power of emotional influence of the human voice.

Researchers who have analyzed this limitation have compared the mechanisms by which the human voice conveys intense emotion to humans with those by which artificial voices are supposed to do so, [1], [2], [3], [4], [5], [6]. It was thus found that the human voice succeeds in transmitting with great fidelity intense emotions to a human listener due to both the speaker's emotional experiences and the mirror neurons in the listener's brain. Mirror neurons duplicate, by means of complex neural processes, in the brain of the person listening, the emotional state of the speaker. If the speaker does not have any kind of emotional experience, then his voice will not convey any kind of emotions, because the mirror neurons in the brain of the person listening will not trigger any kind of emotions in them. Returning to artificial voices, they are produced by computer systems that do not have emotional experiences but only imitate them by copying some tonalities and inflections of the human voice. The researchers, therefore, believe that the mirror neurons in the brain of a human listener cannot trigger emotions that the computer system only mimics and does not experience, [1], [2], [3], [4], [5], [6]. This hypothesis was also subjected to testing in this research, where the brains of ten human judges were imaged through magnetic resonance equipment (MRI) while they listened alternately to audio recordings with famous politicians' and actors' voices and recordings with artificial voices mimicking them.

II. State of the Art

Mirror neurons are a group of neurons that were initially identified in the brains of macaque monkeys during the 1990s by researchers led by, [7]. They are an unusual type of brain cell that "mirrors" other people's behaviors as if the observer himself were performing the act. The discovery of mirror neurons turned everything upside down and indoors when it comes to the function of the brain in social knowledge, empathy, and imitation of behavior. Mirror neurons have a very significant function in understanding emotional processing and are responsible for the neural phenomenon of empathic and emotional resonance with others. The neurons enable a person to not only know and comprehend other people's actions but also feel their emotions to some extent. This process is merely called 'emotional empathy' or 'affective empathy', [8]. When we see someone else experiencing an emotion like happiness, sadness, or pain, the mirror neurons in our brain get activated, repeating the same neural pattern of that emotion as if we are actually going through it ourselves. This mirroring in the brain allows us to understand other individuals' emotional state, which is essential for social interactions, empathy, and being capable of interpreting other people's intentions and emotions, [9]. The mirror neuron firing in emotional processing goes beyond the degree of details that go into the observation of emotions in other people; they are the ones that allow the mechanism through which emotional contagion takes place, where one person's emotional state can "infect" or influence another person's emotional state.

The importance of mirror neurons to emotional processing is further underscored by their alleged function in some psychological disorders that have their origins in empathy and emotional regulation. As an example, impairments of the mirror neuron system have been shown to be involved in such conditions as autism spectrum disorders, wherein patients should have difficulty identifying and comprehending other people's emotions. Understanding the neurobiological underpinnings of emotional empathy by means of mirror neurons can tell us about the mechanisms behind these disorders and guide treatment, [8].

The activity of mirror neurons was followed using nuclear magnetic resonance technology. Several neuroimaging studies have been able to put forward solid proof of the linkage between emotional resonance in the brain and the activation of mirror neuron activity. Studying the activation of brain function when witnessing emotions or activities has enabled scientists to map out brain areas displaying patterned activity consonant with that of mirror neuron function. One important study [10] employed the technique of functional magnetic resonance imaging (fMRI) to study the neural correlations of pain and touch empathy in the human somatosensory cortex. Here, the researchers discovered that, as the participants watched others hurt or be touched, there was considerable activation of the mirror neuron system in their somatosensory cortex. What this suggests is that the mirror neuron system plays an important role in empathy and caring about others. In their examination of emotional contagion, [11] employed fMRI to contrast the brain systems for affective and cognitive empathy. They discovered that emotional contagion, catching and transmission of emotion to others was connected with activation in the mirror neuron areas, proposing mirroring to relate to understanding emotion.

These experiments together offer strong evidence for the activation of mirror neurons with emotional content and empathy. The activation of the mirror neuron system by observing others' emotional expressions or actions implies its central role in relating and understanding others' emotions, a critical component of social interactions, and human empathy. The processing of emotional information derived from human voice is mediated through intricate neural circuitry that enables us to notice and interpret the various emotional signals, including tone, pitch, and intonation. Functional neuroimaging studies have yielded significant findings regarding neural mechanisms of processing emotional voice. One important area of the brain involved in the perception of emotional voices is the superior temporal sulcus (STS). The STS is also linked to the detection of socially pertinent stimuli, such as facial emotional expressions and affective expressions of voice. fMRI research has repeatedly found heightened activity in the STS when subjects listened to emotional voices, [12], [13]. The STS is considered to have a pivotal role in the processing and extraction of acoustic cues for emotional information in voice. But another significant area engaged in the emotional voice processing is the insula. The insula supports interoceptive awareness and is responsible for subjective feeling regarding emotions. It has been found to be activated when exposed to emotional voices, which indicates its role in connecting auditory emotional content with subjective emotional experience, [14], [15], [16]. Furthermore, the amygdala, a key structure of the limbic system, is implicated in emotional processing, such as the processing of voices. Amygdala activation has been found during the perception of emotional prosody, showing that it plays a part in the rapid identification and analysis of emotional cues in voices, [17], [18]. Finally, the prefrontal cortex (PFC) plays a role in higher-order emotional processing of voices. The PFC, especially the ventromedial prefrontal cortex (vmPFC) and orbitofrontal cortex (OFC), is involved in evaluating the emotional significance of voices and integrating emotional information with cognitive and memory processes. PFC activation has been linked with the identification of emotional valence and regulation of emotional responses evoked by voices, [19], [20]. Finally, the mirror neuron system (MNS) has also been found to be involved in emotional voice processing. The MNS, originally discovered as part of the motor behavior domain, is also found to contribute to perceiving and interpreting the emotions that people express. It has been recently shown through neuroimaging techniques that the MNS is activated when people listen to emotional voice sounds, pointing towards its function in empathetic responses to the emotions expressed by voice, [19].

Mirror neurons and emotional processing studies have broader implications than human-to-human communication. As artificial intelligence has emerged and AI voices have been added to different applications, data regarding emotional reactions to AI voices are pertinent now. Whether AI voices can evoke emotions in listeners is an issue that receives increasing attention since it affects user experience on virtual assistants, customer support portals, and entertainment sites, [20].

The arrival of artificial voices, especially those from AI-based speech synthesis, has raised some interesting questions regarding how our brains process and react to these artificial voices versus natural human voices. Studies examining brain activation patterns in response to natural and artificial voices have found both similarities and differences in neural processing of the two kinds of auditory stimuli. Functional neuroimaging experiments, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), have revealed that the primary auditory cortex (PAC) is generally engaged when individuals perceive both natural and synthetic voices, [21], [22]. The PAC receives auditory input first, whether from a natural source or an artificial one. This finding suggests that, at a basic level, the brain treats both voices as acoustic stimuli that trigger sensory processing in the auditory cortex. There are differences, however, in brain activation when examining higher-order cognitive processes and emotional reactions to natural versus artificial voices. A number of studies have indicated that natural human speech is more likely to produce greater emotional reactions and more robust activations of brain areas engaged in emotion processing, for example, the amygdala and anterior cingulate cortex, [23], [24]. Human voices also convey subtle emotional information, including prosody, intonation, and vocal timbre, difficult to precisely simulate in artificial voices. Therefore, natural voices are better at evoking emotions and engaging emotional processing networks in the brain of the listener. Artificial voices produced by AI algorithms, however, can engage brain areas involved in cognitive processing, especially when comprehending and recalling from memory. They are meant to be readable, transparent, and concise, and this may prove useful in particular uses, e.g., learning a language or spreading information, [25], [26], [27]. Experiments on voice synthesis perception have presented higher activations within brain areas involved in processing languages, e.g., Broca's area and the left inferior frontal gyrus, [28]. Also, brain activation differences could be caused by the uncanny valley effect, whereby nearly human-sounding but not human-sounding voices can elicit a sense of unease or eeriness, [29]. Such a dissonance between the anticipated human-like voice and the actual features of the synthetic voice can cause changes in brain activity, potentially via social cognition and human-likeness perception regions. In general, though there are commonalities in early auditory processing of human and artificial voices, the unique emotional and cognitive characteristics of human voices and the current shortcomings of AI-generated voices lead to differences in brain activation patterns. Additional investigation is necessary to better understand the neural basis of the listener's response to artificial voices and how to better utilize them in many applications.

Mirror neurons which are known to be active while observing other people's emotional understanding and empathy, may also be involved when dealing with emotions induced by AI voices. Although AI voices do not include the entire palette of emotional signals included in natural human voices, they are yet capable of provoking emotional responses from listeners. Listeners have been discovered to attribute emotional qualities to AI voices by applying speech prosody, intonation, and rhythm, [23], [30].

In addition, the brain's emotional processing ability can be the same across auditory modalities. It has been shown in studies that cross-modal activation exists between mirror neurons in proving that face emotion perception triggers the mirror system across motor and premotor regions, [31], [32]. Likewise, when people hear emotional voices, mirror neurons can be activated not just with the sound stimulus but also with the corresponding visual and somatosensory representations of the emotions, which can help toward a more holistic comprehension of the emotional content of AI-generated voices. In addition, the process of emotional contagion, in which people involuntarily imitate and "catch" other people's emotional states, can also be applied to the emotional processing of AI-generated voices. As mirror neurons have been reported to be responsible for emotional contagion [33], they might be accountable for making listeners vulnerable to being emotionally affected by AI-created voices. This could have significant implications in scenarios where emotional response is needed, e.g., customer service or virtual friends. Still, the emotional processing elicited by AI voices will probably be more intricate and complicated than that based solely on mirror neurons. Other regions of the brain, including the amygdala for emotional processing and the prefrontal cortex for cognitive evaluation, also have essential roles in emotional processing, [15], [34]. The interaction of mirror neurons with these areas could influence the overall emotional experience and its related behavioral and cognitive response to AI voices.

The comprehension of the role played by mirror neurons in emotional resonance also has deeply significant implications for education, stressing the importance of properly integrating emotional information to better motivate and engage students, [35], [36]. Such realization is crucial in crafting personalized learning experiences that change in real-time according to the emotions of the students, potentially transforming educational technology and practice, for example, [37]. In addition, utilization of the results can improve SEL curricula, more specifically empathic development, by utilizing multimedia tools for the simulation of emotional experiences, [38], [39].

In addition, providing teachers with awareness of mirror neurons through professional development can prepare them to develop more empathetically stimulating learning environments, making informed choices regarding the use of technology and human-to-human interaction in classrooms, [40], [41].

III. Methodology

In this study, human voice and artificial voice imitation for the voice of a famous Romanian actress were analyzed with AI algorithms and human evaluators. The methodology contains the following stages:

1. An audio recording of the famous Romanian actress Gabriela Bobes was chosen. In that recording (AUD-GAB) Gabriela's voice transmits positive emotions (appropriate for love) with a high intensity.
2. The transcription of the AUD-GAB was then used by two Romanian actresses to imitate Gabriela's voice in all aspects: timbre, rhythm, inflections, tonality, emotions transmitted, etc. Two audio recordings were thus obtained (AUD-IMIT-HUMAN1 and AUD-IMIT-HUMAN2) that imitate Gabriela's voice.
3. The transcription of the AUD-GAB was also used to imitate Gabriela's voice with an artificial voice in all aspects: timbre, rhythm, inflections, tonality, emotions transmitted, etc. Another audio recording (AUD-IMIT-AI) was thus obtained.
4. Ten human evaluators listened to the four audio recordings in the laboratory and rated them in terms of vocal timbre and emotion transmitted (valence and intensity).
5. The ten human evaluators listened to the four audio recordings while their brains were scanned with MRI equipment and evaluated them again in terms of vocal timbre and emotion transmitted (valence and intensity). The MRI images were then analyzed in order to identify the degree of oxygenation of their brain's areas related to the positive emotion transmitted by the audio recordings.
6. The four audio recordings were analyzed with two AI algorithms: one AI algorithm for voice timbre identification in each of the four recordings and one AI algorithm for detection of the emotions transmitted intensities in each of the four recordings.

IV. Results

The analysis carried out with human evaluators showed that the artificial voices used in this study had a lower emotional load than the human voices they imitated. Volunteers who participated in that study assessed the emotional impact of artificial voices with an average grade of 3 (on a scale of 1 to 10) and grade 8 for human voices. Of the ten human evaluators, seven considered that the recordings with artificial voices had an emotional charge with an intensity two times lower than that of the recordings with human voices, and three evaluators considered that the recordings with artificial voices had the same emotional charge as the recordings with human voices.

In Fig.1 and Fig.2 the differences are shown between the MRI images when the participant did not hear any audio and the images when the participant listened to a human voice during the MRI at the hippocampus level (Fig. 1) and for the frontal lobe of the brain (Fig.2).

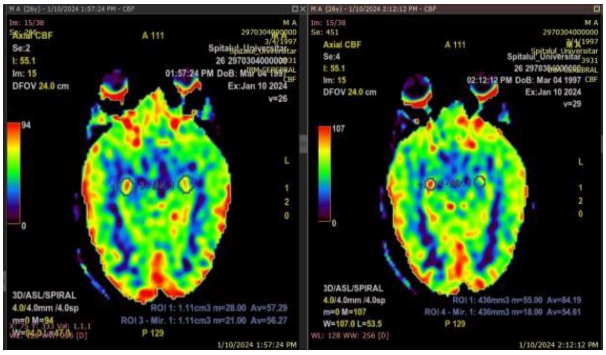

Fig.1 MRI images with no audio vs. human voice audio for the hippocampus area of the brain

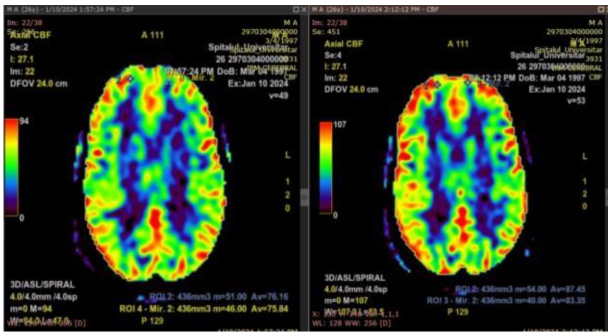

Fig. 2 MRI images with no audio vs. human voice audio for the frontal lobe of the brain

In Fig.3 and Fig.4 the differences are shown between the MRI images of the artificial voice and the human voice at the hippocampus level (Fig. 3) and for the frontal lobe of the brain (Fig. 4).

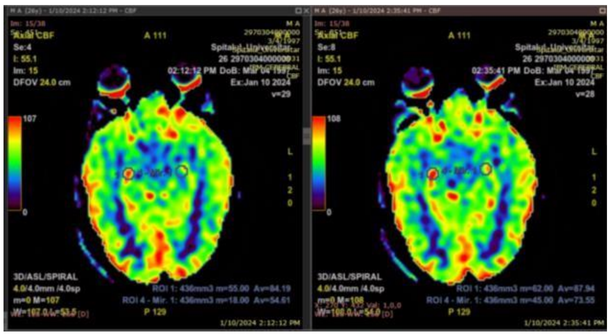

Fig. 3 MRI images with artificial voice audio vs. human voice audio for hippocampus area of the brain

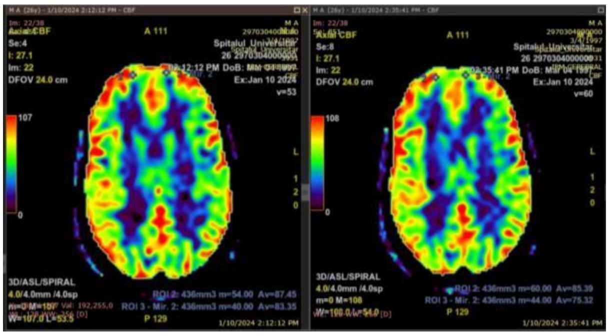

Fig. 4 MRI images with artificial voice audio vs. human voice audio for the frontal lobe of the brain

The MRI images from all human raters while listening to recordings of human voices showed greater oxygenation of mirror neuron areas and those responsible for inducing emotions, compared to MRI images from artificial voices. These results showed that the brains of all evaluators perceived the differences in emotional load between human voices and artificial voices, even if three subjects did not realize these differences. In these three human evaluators, the perception of differences in emotional load worked at the brain level, but it seems that they did not realize it.

In Fig.5 is shown the differences between the MRI images of the two imitating voices (left and right) and Gabriela's original voice (middle).

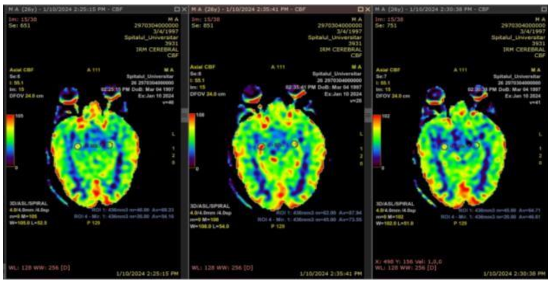

Fig. 5 MRI images with imitating voices (left and right) vs. the real voice (middle).

MRI measurements and their analysis indicated that the mechanisms by which artificial voices induce emotions in human listeners are more related to brain memory processes than those involving mirror neurons. The MRI images showed that recordings of artificial voices triggered the activation of brain areas responsible for memory and those responsible for emotions. These results suggest the idea that artificial voices can reactivate in the brains of human listeners only those emotions that are already in their memory, and that they will experience them with a low intensity. The results of this study also show that artificial voices may not be able to convey new emotions to human users that they have never experienced before and therefore would not be in their memory. The low intensity of emotions is due to brain memory mechanisms that reactivate only part of the stimuli that previously triggered the emotion, because in the memorization process, many stimuli were not included in the memory or were removed along the way. The results showed that the limitation in intensity of the emotions transmitted by the atypical voices is due to the cerebral memory mechanisms that reactivate only a part of the emotion-triggering stimuli.

The evaluation of audio recordings with the proposed CNN architecture is shown in Table 1 and suggests the idea that there is a big difference between the emotional load of human voices compared to the artificial voices used. The AI algorithms used in this stage were trained with the specific emotional load of each human voice for the people chosen in this study.

Table 1. The AI algorithm results of the emotion level for each type of audio file

Type of audio	**Emotion Intensity**
artificial voice	1
imitation human voice #1	1.5
imitation human voice #2	1
original human voice	2.5

V. CONCLUSIONS

The results obtained in this study showed that human voices can be imitated very well both by professional actors and by artificial voices. Both types of imitations are so good that human evaluators do not notice any difference in tonality, timbre, and inflections between the original and the imitations with human voice or those with artificial voice. However, these differences will be noticed by the AI algorithms specialized in detecting the voice fingerprint of each person.

The results showed that human evaluators noticed the differences between the emotions transmitted by the original voice and those transmitted by imitations with a human voice and an artificial voice. These differences were also noticed by the AI algorithms specialized in detecting the emotions embedded in the voice and their intensity.

MRI investigations showed that human evaluators were able to detect the differences in the emotions conveyed by the original voice and the two types of imitations due to the mirror neurons in the listeners' brains. MRI measurements have also shown that artificial voices can convey emotions but not as intensely as the emotions conveyed by human voices. This is because human voices manage to convey emotions to the listeners due to the brain mechanisms in the brains of the listeners that are based on mirror neurons. MRI measurements showed that artificial voices do not trigger brain mechanisms based on mirror neurons, but brain mechanisms based on memory.

All the evaluations carried out in this study showed that although the audio recordings with human voice/artificial voice sounded the same, the emotional charges of artificial voices were lower than those of human voices. The obtained results disproved the hypothesis according to which artificial voices cannot convey emotions to human listeners. Measurements have shown that artificial voices can induce emotions in human listeners due to neural processes other than those related to mirror neurons, namely through memory processes. The artificial voices can trigger emotions in human listeners, but with a lower intensity than human voices. The study thus shows that, due to the way they are constructed, artificial voices today cannot convey intense emotions to human interlocutors, and that to overcome this limitation they would have to be reconfigured differently, but it is not yet known how.

As research in this area is still in its early stages, further investigations are needed to fully understand the role of mirror neurons in processing emotions elicited by AI-generated voices. Unraveling the neural mechanisms involved will not only advance our understanding of how AI-generated voices influence human emotions but also inform the development and design of emotionally engaging and responsive AI systems.

The implications of these findings extend beyond the field of neuroscience into practical applications within educational settings. The differential impact of human and artificial voices on emotional resonance and mirror neuron activation underscores, for example, the potential for optimizing AI-driven educational tools. By better understanding how mirror neurons respond to various auditory stimuli, educators and technologists can design AI systems that more effectively mimic the emotional nuances of human speech. This could enhance the learning experience and support the development of empathy and social skills through more emotionally responsive educational content. Consequently, the insights from this study advocate for a possible more nuanced integration of AI voices in educational technology, aiming to harness the full potential of emotional engagement in learning environments.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The authors wrote, reviewed and edited the content as needed and they have not utilised artificial intelligence (AI) tools. The authors take full responsibility for the content of the publication.

References

[1] Bachorowski, J.-A. (1999). Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science,* 8(2), 53-57. https://doi.org/10.1111/1467-8721.00013 (Original work published 1999).

[2] Abelin, Å., Allwood, J. (2000) Cross linguistic interpretation of emotional prosody. *Proceedings ITRW on Speech and Emotion*, 110-113.

[3] Chen, X., Yang, J., Gan, S., & Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. *PloS one*, 7(1), e30278. https://doi.org/10.1371/journal.pone.0030278.

[4] Bänziger, T., Hosoya, G., & Scherer, K. R. (2015). Path Models of Vocal Emotion Communication. *PloS one*, *10*(9), e0136675. https://doi.org/10.1371/journal.pone.0136675.

[5] Mitchell, R. L., & Xu, Y. (2015). What is the Value of Embedding Artificial Emotional Prosody in Human-Computer Interactions? Implications for Theory and Design in Psychological Science. *Frontiers in psychology*, 6, 1750. https://doi.org/10.3389/fpsyg.2015.01750.

[6] Huang, K. L., Duan, S. F., & Lyu, X. (2021). Affective Voice Interaction and Artificial Intelligence: A Research Study on the Acoustic Features of Gender and the Emotional States of the PAD Model. *Frontiers in psychology*, *12*, 664925. https://doi.org/10.3389/fpsyg.2021.664925.

[7] Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews. Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060.

[8] Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral cortex (New York, N.Y.: 1991)*, *19*(6), 1239–1255. https://doi.org/10.1093/cercor/bhn181.

[9] Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends in cognitive sciences*, *15*(11), 512–519. https://doi.org/10.1016/j.tics.2011.09.003.

[10] Bufalari, I., Aprile, T., Avenanti, A., Di Russo, F., & Aglioti, S. M. (2007). Empathy for pain and touch in the human somatosensory cortex. *Cerebral cortex (New York, N.Y.: 1991)*, *17*(11), 2553–2561. https://doi.org/10.1093/cercor/bhl161.

[11] Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *NeuroImage*, *43*(3), 571–580. https://doi.org/10.1016/j.neuroimage.2008.08.014.

[12] Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in cognitive sciences*, *8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008.

[13] Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., … Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage,* 30(2), 580-587. https://doi.org/10.1016/j.neuroimage.2005.09.059.

[14] Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, *11*(4), 277–281. https://doi.org/10.1080/17549500902943043.

[15] Frühholz, S., & Grandjean, D. (2013). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex: a journal devoted to the study of the nervous system and behavior*, *49*(5), 1394–1403. https://doi.org/10.1016/j.cortex.2012.08.003.

[16] Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in neurobiology*, *123*, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003.

[17] Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature neuroscience*, *8*(2), 145–146. https://doi.org/10.1038/nn1392.

[18] Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal

emotional processing. *Trends in cognitive sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009.

[19] Liu, Y., Tang, Q., Zhao, X., Lu, H., Yuan, G., Wang, G., & Liu, G. (2020). Neural activation of different music styles during emotion-evoking. *Psychology of Music*, 49(6), 1546-1560. https://doi.org/10.1177/0305735620968260 (Original work published 2021).

[20] de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature reviews. Neuroscience*, 18(8), 498–509. https://doi.org/10.1038/nrn.2017.72.

[21] Kotz, S.A., Paulmann, S. (2011), Emotion, Language, and the Brain. *Language and Linguistics Compass*, 5: 108-125. https://doi.org/10.1111/j.1749-818X.2010.00267.x.

[22] Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37 (4), 417-435. https://doi.org/10.1016/j.wocn.2009.07.005.

[23] Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion (Washington, D.C.)*, *9*(5), 691–704. https://doi.org/10.1037/a0017088.

[24] Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: a neurobiological model of multimodal affective information processing. *Physics of life reviews*, *8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002.

[25] Frühholz, S., Trost, W., & Grandjean, D. (2014). The role of the medial temporal limbic system in processing emotions in voice and music. *Progress in neurobiology*, *123*, 1–17. https://doi.org/10.1016/j.pneurobio.2014.09.003.

[26] Zhang, W., Li, J., Ji, L. *et al.* fNIRS experimental study on the impact of AI-synthesized familiar voices on brain neural responses. *Scientific Reports* **15**, 16872 (2025). https://doi.org/10.1038/s41598-025-92702-5.

[27] Harris L. T. (2024). The Neuroscience of Human and Artificial Intelligence Presence. *Annual review of psychology*, *75*, 433–466. https://doi.org/10.1146/annurev-psych-013123-123421.

[28] Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *32*(11), 3942–3948. https://doi.org/10.1523/JNEUROSCI.3814-11.2012.

[29] Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology*, *6*, 390. https://doi.org/10.3389/fpsyg.2015.00390.

[30] Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The neural correlates of emotional prosody comprehension: disentangling simple from complex emotion. *PloS one*, *6*(12), e28701. https://doi.org/10.1371/journal.pone.0028701.

[31] Wallenwein, L.A., Schmidt, S.N., Hass, J., & Mier, D. (2024). Cross-modal decoding of emotional expressions in fMRI—Cross-session and cross-sample replication. *Imaging Neuroscience, 2*, 1-15. https://doi.org/10.1162/imag_a_00289.

[32] Likowski K.U., Mühlberger A, Gerdes A.B.M., Wieser M.J., Pauli P., Weyers P. (2012) Facial mimicry and the mirror neuron system: simultaneous acquisition of facial electromyography and functional magnetic resonance imaging. *Frontiers in Human Neuroscience* **6**:214. https://doi.org/10.3389/fnhum.2012.00214.

[33] Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). Face to face: blocking facial mimicry can selectively impair recognition of emotional expressions. *Social neuroscience*, *2*(3-4), 167–178. https://doi.org/10.1080/17470910701391943.

[34] Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of cognitive neuroscience*, *21*(7), 1255–1268. https://doi.org/10.1162/jocn.2009.21099.

[35] Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. PLoS biology, 3(3), e79. https://doi.org/10.1371/journal.pbio.0030079.

[36] Mara, D. (2017). The function of mirror neurons in the learning process. 8th International Conference on Manufacturing Science and Education – MSE 2017 "Trends in New Industrial Revolution", 12012, https://doi.org/10.1051/matecconf/201712112012.

[37] Choksi, K., Chen, H., Joshi, K., Jade, S., Nirjon, S., & Lin, S. (2024). SensEmo: Enabling affective learning through real-time emotion recognition with smartwatches. *arXiv*. https://doi.org/10.48550/arXiv.2407.09911.

[38] Sun, F., Li, L., Meng, S., Teng, X., Payne, T., & Craig, P. (2025). Integrating emotional intelligence, memory architecture, and gestures to achieve empathetic humanoid robot interaction in an educational setting. *arXiv*. https://doi.org/10.48550/arXiv.2505.19803.

[39] Bachen, C. M., Hernández-Ramos, P. F., & Raphael, C. (2012). Simulating REAL LIVES: Promoting global empathy and interest in learning through simulation games. *Simulation & Gaming,* 43(4), 437–460. https://doi.org/10.1177/1046878111432108.

[40] Trieu, M., Foster, A.E., Yaseen, Z.S., Beaubian, C., Calati, R. (2019). Neurobiology of Empathy. In: Foster, A.E., Yaseen, Z.S. (eds) Teaching Empathy in Healthcare. Springer, Cham. https://doi.org/10.1007/978-3-030-29876-0_2.

[41] Acharya, S., & Shukla, S. (2012). Mirror neurons: Enigma of the metaphysical modular brain. *Journal of natural science, biology, and medicine*, *3*(2), 118–124. https://doi.org/10.4103/0976-9668.101878.

Contribution of individual authors to the creation of a scientific article (ghostwriting policy)

We confirm that all Authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of funding for research presented in a scientific article or scientific article itself

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.