The Observation of Actors' Vocal Emotion Exercises with Deep Learning and Spectral Analysis

COSTIN ANDREI BRATAN[1,2], CLAUDIA TOCILA-MATASEL[3,4], ALEXANDRA-GEORGIANA ANDREI[1], ANA VOICHITA TEBEANU[5], EDUARD FRANTI[2,6], MONICA DASCALU[1,2], BOGDAN IONESCU[1], GHEORGHE IANA[3], GABRIELA BOBEȘ[7], BOGDAN MOROSANU[1], ANA-MARIA OPROIU[8], GABRIELA IORGULESCU[8,9]

[1]Faculty of Electronics, Telecommunications and Informational Technology,
National University of Science and Technology Politehnica,
Splaiul Independenței 313, Bucharest 060042,
ROMANIA

[2]Romanian Academy Research Institute for Artificial Intelligence,
Calea Victoriei 125, Bucharest 010071,
ROMANIA

[3]MEDIMA Health,
Odăii 42, Otopeni 075100,
ROMANIA

[4]University of Medicine and Pharmacy "Iuliu Hațieganu",
Victor Babes 8, Cluj-Napoca 400347,
ROMANIA

[5]Departamentul Pentru Pregătirea Personalului Didactic,
National University of Science and Technology Politehnica,
Splaiul Independenței 313, Bucharest 060042,
ROMANIA

[6]National Institute for Microtechnologies,
Erou Iancu Nicolae 126A, Voluntari 077190,
ROMANIA

[7]OKaua Theater Company and Pink Stil SRL,
Bucharest,
ROMANIA

[8]Faculty of Medicine,
University of Medicine and Pharmacy "Carol Davila",
Bulevardul Eroii Sanitari 8, Bucharest 050474,
ROMANIA

[9]Academic Society of Behavioral Sciences,
Bucharest,
ROMANIA

Abstract: - This paper presents two distinct methods that demonstrate the increased intensity of a specific emotion when the induced emotion is trained daily for 30 days. For this study, four actors participated in a 30-day exercise trial and were recorded each day using high-level audio equipment. The first method supporting our hypothesis is a deep learning approach. A convolutional neural network pre-trained on Mel-frequency cepstral coefficients analyzed the actors' recordings and delivered the intensity of the detected emotion. The

CNN tested 3,561 segments of 0.2-second length, and the results showed a higher level of intensity on the final day of training for each participant. The second method is spectral analysis. The spectrograms generated on the first and final days of the experiment showed that the spectral composition on the final day had a wider range of frequencies than on the first day, further supporting our hypothesis.

Key-Words: - artifficial intelligence algorithm, convolutional neural network, emotion detection, Mel-frequency cepstral coefficients , spectral analysis, spectrogram.

Received: May 26, 2023. Revised: December 29, 2023. Accepted: January 22, 2023. Published: March 1, 2024.

1 Introduction

Actors' voices have a strong emotional impact on the audience. Actors do not have this ability from birth, but acquire it after a long and persevering practice of many diction, breathing, and stage manifestation exercises. The efforts that the actors make in the preparation of each show are unimaginably great for the rest of us, and they involve repeating all the lines thousands of times until they are "charged" with those intense emotions that the actors want to convey to the public. An actor's voice sounds different after repeating a text a thousand times compared to the first reading. The more times the actor repeats the role he is playing, the greater will be the emotional impact of his voice on the audience. This is the subject of this article: measuring the intensity of the emotions imbued in the voices of five actors during 30 days of training with acting vocal techniques. The five actors were audio recorded daily during training and the audio files were analyzed with AI algorithms to determine the intensity of emotions imbued in their voices. Additionally, on the first and last day of training, the actors were monitored with AI algorithms to identify brain processes that favored the amplification of their power of emotional influence through the voice.

Some psychologists claim that, through the techniques they practice during rehearsals, the actors enter a special mental state that makes for them to be in full unison between what they say, think, intend, and feel, and the stage movements they perform. This state is called in psychology the state of flow and it helps the actors to have a strong emotional impact on the audience, [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

Some authors argue that the intense emotions that the actors transmit in the state of flow are due to the functioning mechanisms of the mirror neurons in the brains of the spectators. Mirror neurons are connected to those areas in the brain that are responsible for emotions and perceptions through the senses. Thanks to these neurons, everything we hear, see, or perceive in another person is duplicated in our brain along with their emotional state. The duplication of another person's emotional state in our brain is all the stronger the more we have perceptions of them (auditory, visual, olfactory, etc.). This perhaps explains why the emotions generated by the actors in a 3D film are much more intense than those conveyed by the actors in a silent film, [11].

2 The State of Flow

The state of flow, also known as being "in the zone," is a psychological concept that describes a highly focused and fully immersive mental state in which a person is fully engaged and completely absorbed in an activity. Coined by psychologist Mihaly Csikszentmihalyi, flow is characterized by a sense of energized focus, intense concentration, and enjoyment in the present moment. When in a state of flow, individuals often lose track of time and experience a deep sense of satisfaction and fulfillment. Flow is commonly linked to activities that push a person's abilities and skills all the while offering objectives and instant feedback. It often happens when the difficulty of a task aligns, with or slightly surpasses an individual skill level resulting in an equilibrium that enables them to perform. The experience of flow is characterized by a sense of engagement effortless focus, a blending of action and awareness, and a lack of self-awareness.

Key characteristics that define the state of flow include:

- Concentration: the person's attention is completely immersed in the activity, fully engaged and absorbed in what they're doing.
- Clear objectives and feedback: the activity has well-defined goals and the individual receives immediate feedback on their performance enabling them to make adjustments and stay focused.
- Fusion of action and awareness: the person becomes one, with the activity seamlessly carrying out actions without conscious effort or overthinking.

- Altered perception of time: time seems to fly or even stand still as the individual becomes fully engrossed in the task losing track of the world.
- Sense of control: the person experiences a sense of mastery and control over the activity as their skills align, with the task's demands.
- Enjoyment and fulfillment: flow brings about a sense of enjoyment, satisfaction, and intrinsic motivation that stems directly from engaging in the activity itself.

Flow can manifest in a variety of endeavors, such as sports, creative pursuits, work, hobbies, and everyday tasks. It is often associated with performance and creativity because individuals in a state of flow display heightened focus, productivity, and innovation.

Attaining a state of flow offers benefits like improved performance enhanced learning capabilities increased well-being and a sense of fulfillment. By understanding the characteristics and circumstances that flow people can strive to create conducive environments and engage in activities that promote this ideal psychological state. Flow refers to an experiential state that arises when individuals are fully engaged in an activity that aligns with its demands. In [12], the authors discovered that demanding tasks elicited higher levels of flow among those with greater fluid ability but lower levels among those with lower fluid ability. To truly experience a state of "flow" one must find the balance, between their skill level and the challenge they face. It involves maintaining focus setting goals and immersing oneself in the task at hand. During this state, individuals may lose track of time. Temporarily set aside their sense of self. Finding this sweet spot between skill and challenge is essential for achieving flow, [13]. Additionally, a 2018 study suggests that understanding the concept of flow can be valuable in designing information systems (IS) that promote optimal user interactions. With advancements, in NeuroIS and psychophysiology, it is now possible to assess flow levels during IS usage, [14]. Additionally, experiencing flow states can greatly enhance performance and it is suggested that transcranial direct current stimulation (tDCS) could be a method to induce these states, [15].

Flow occurs when individuals meet a challenge with the skills have clear goals focus their concentration feel in control of their actions lose track of time temporarily and experience a temporary loss of self-awareness. An interesting finding in [16], suggests that reading fiction that individuals choose themselves and matches their skill level can induce a state of flow. A research in 2013 proposes that flow arises when implicit motivations are triggered by incentives within the task itself without any conflicting explicit motivations being activated. Moreover, it is crucial for individuals to perceive themselves as capable of achieving the task at hand, [17]. Another perspective on measuring flow was presented in [18], where the authors consider it as a combination of engagement in the task enjoyment derived from it and having control, over it.

The studies indicate that achieving a state of flow requires finding the balance between challenge and skill having clear goals maintaining focus feeling in control and experiencing a sense of enjoyment.

Flow is a state that allows individuals to reach their full potential and enhances their overall experience. According to [19], everyday flow is characterized by levels of motivation, cognitive efficiency, activation, and satisfaction. Also, [20], suggests a connection between flow and mindfulness. In [21], the authors examined the characteristics of flow in the process, while in [22], the authors explored its role in sports, exercise, and performance. Both studies concluded that flow leads to peak performance and is associated with experiences. In summary, these papers indicate that experiencing flow offers benefits such as increased motivation, cognitive efficiency, activation, satisfaction, and peak performance.

Lately, there has been a growing interest, in exploring the link between the state of flow and the emotions conveyed through voice. Voice is recognized as a tool for expressing emotions and researchers have studied its ability to evoke responses from listeners, [23].

When investigating how emotions conveyed through voice impact the state of flow researchers propose that emotional vocal signals play a role in initiating and sustaining the flow experience. Research findings indicate that adjusting expressions can influence the induction of flow. Studies revealed that individuals who were emotionally primed demonstrated levels of flow during tasks compared to those who received neutral priming, [24]. This suggests that conveying cues through voice has the potential to enhance engagement and absorption in activities.

In contexts, Johnson, study discovered a relationship between flow and emotional expression in speech. Participants who reported experiencing a state of flow during a writing task exhibited variability in their expressions of emotions. This indicates an interaction, between experiences, creative engagement, and modulation of one's voice, [25]. The study also found that people who

experience flow show changes in how they express their emotions through their voices. This suggests that there is a relationship, between experiences and the immersive engagement of being in a state of flow.

Looking at the connection between flow and the transmission of emotions through voice has implications across fields. In a 2013 study, the authors explored how music, as a form of expression affects stress response, [26]. Their findings demonstrated that music can actively regulate stress levels highlighting the potential for using stimuli like voice to enhance and enrich the experience of being, in a state of flow.

3 Methodology and Equipment

3.1 Methodology of the Study

In this study, four actors were monitored and audio recorded for 30 days while practicing exercises to amplify the emotional impact of the voice. All actors agreed to voluntarily participate in this study and were informed that they were to be monitored by video and audio equipment during training. The actors were audio-recorded for 30 days while they interpreted the same text and aimed to charge it with the same positive emotion and with as much intensity as possible. The audio recordings from the actors were then analyzed with AI algorithms and spectral analysis equipment to determine and monitor the intensity of the emotions imbued in the actors' voices, and the spectral component of the voices (what frequencies the actors' voices contained).

Additionally, after the end of the training, other measurements were made. After the last day, the actors were made to read (at first glance) a new text and to load it with emotions opposite to those from the previous days' training. The new audio recordings were then analyzed with AI algorithms and spectral analysis equipment, and the results were compared with those during the training.

3.2 Utilised Equipment

During the thirty-day experimental trial, all four participants recorded their speech in a well-equipped studio with special audio conditions to obtain better and clearer voice recordings.

The recordings were made using a microphone called CMC5 which had an MK4 capsule attached to it. The microphone Signal, to Noise Ratio (SNR), was measured at 80 dB indicating its ability to capture the desired sound while minimizing background noise interference. With a Maximum Sound Pressure Level (SPL) of 131 dB, the microphone could handle pressures without distortion.

For interface, we used Apogee Symphony i/o in this setup. It has a Total Harmonic Distortion plus Noise (THD+N) rating of 115 dB at 22 dBu meaning it accurately reproduces the input signal without introducing distortion or noise. The interface also boasts a range of 122 dB A weighted allowing it to effectively capture both loud sounds. This wide range facilitated the recording of details.
To facilitate the recording process we employed the audio workstation Pro Tools System. Widely recognized in the audio industry this software offers an array of features and tools, for capturing, editing, and processing audio. We strategically positioned the Schoeps CMC5 with the MK4 capsule during recording to ensure the capturing of our intended source.

The Apogee Symphony i/o ensures that audio is converted to resolution, with distortion and noise. The microphone has a signal-to-noise ratio. Can handle high sound pressure levels ensuring accurate and distortion-free recordings. We used the Pro Tools System to direct the signal through the Symphony i/o for recording, editing, and processing. To minimize noise and unwanted reflections we designed the recording environment with a treated booth that has a reverberation period of 0.4 seconds. This creates a controlled audio capture for specific types of recordings, like vocals or instruments that require minimal room ambiance.

4 Results

The analysis of the audio recordings with AI algorithms showed a continuous increase in the intensity of emotions impregnated in the voices of the actors during the 30 days of training.

The current application utilizes a Convolutional Neural Network (CNN) structure, with connected layers. These layers have been trained using the Mel-frequency cepstral coefficients (MFCCs) extracted from the input audio files. Each audio file is divided into segments of two seconds and a fixed window size is applied to extract 40 features using the Discrete Fourier Transform (DFT). The CNN consists of a total of 87,944 parameters. Employs a ReLu activation function. Additionally, it includes a dropout rate of 20% and a connected layer that reduces output from 640 components to 8 components, [27].

These final eight elements correspond to the emotions the algorithm can identify; Neutral, Calm,

Happy, Sad, Irritable, Fearful, Disgust, and Surprised. Emotion levels range between 0 and 3 with increments of 0.5.

To train the CNN model mentioned above we used 1,440 speech files from the Ryerson Audio Visual Database of Emotional Speech and Song (RAVDESS) dataset along with 2,800 files from the Toronto speech set (TESS) dataset. Both datasets contain audio .wav files sampled at a rate of 16 bits, per sample. Have a sample rate of 48kHz. The dataset consists of recordings, from 24 actors with several females and males. These actors delivered two sets of statements that were lexically similar all spoken in a North American accent, [27].

The results of the study involving the CNN input and the 3561 segments generated and tested are presented in Table 1 for the first and 30th day of training for each actor. It is evident from the table that the intensity of the induced emotion increased significantly from the first day of training to the last.

Table 1. The CNN results for each participant

Actor No.	**Intensity value on the 1st day**	**Intensity value on the 30th day**
#1	0.5	1.5
#2	1	2.5
#3	1	2
#4	1	2.5

Figure 1 and Figure 2 plot the spectrograms on the first and the last day of training. The spectral analysis of the audio recordings showed that the actors' voices contained a wider range of frequencies on the last day of training compared to the first day. This indicates that the breathing techniques practiced by the actors helped them to use all their resonant cavities during the interpretation of the text and to charge it with more and more intense emotions.

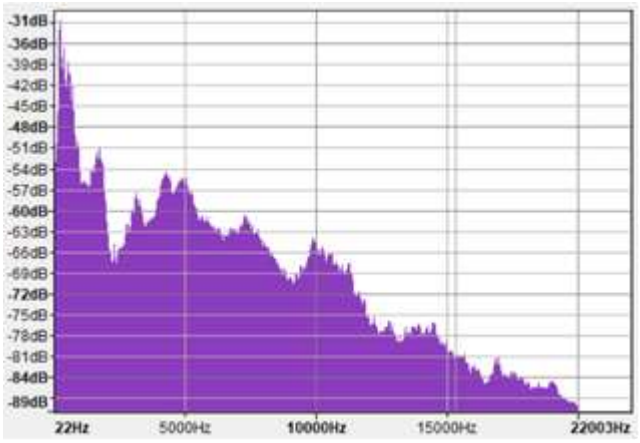

Fig. 1: Spectrogram on the first day of training

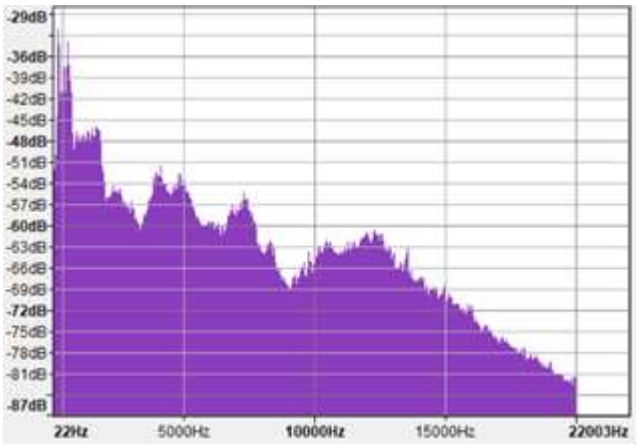

Fig. 2: Spectrogram on the last day of training

Following the training, the actors were instructed to read a new text while simultaneously grappling with emotions that were contrary to those experienced during the initial training. As depicted in Figure 3, the spectrogram of the new text was examined. Employing the same spectral analysis equipment, the new audio recordings were scrutinized, and the findings revealed that both the spectral composition of the voices and the emotions conveyed therein were significantly lower than those observed after the training with the initial text. These measurements demonstrate that acting techniques may effectively augment the emotional content of the voice, provided they are systematically practiced and sustained over an extended period.

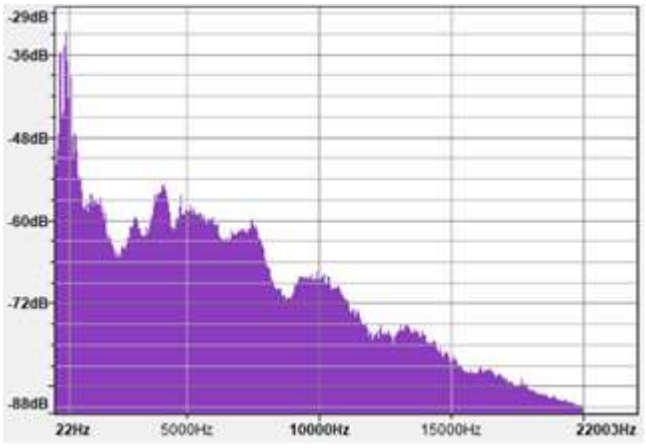

Fig. 3: Spectrogram of the new text interpretation

A comparison of the outcomes from the assessment of the intensity of emotions and the spectral composition of the voices after the completion of training with the new text indicates that acting techniques aid in enhancing the emotional impact of the voice, provided they are consistently applied and endured for a considerable duration. It is worth noting that after being trained to

exhibit a specific type of emotion, actors often require an equally prolonged period of training to express other types of emotions. This can be observed in some actors who, after portraying the same role in a series for numerous years, subsequently struggle to effectively portray other roles.

5 Conclusion

The results of this study show that the practice of acting techniques of voice, breathing, and stage interpretation ensure the amplification of the power of emotional influence of the voice. Practicing these techniques helps actors to efficiently use all the resonant cavities along the vocal tract, and their voices thus become capable of conveying intense emotions to the audience. Acting techniques also help actors develop their power of concentration and enter the state of flow, which proved to have an important role in amplifying the emotional charge of their voices.

The intersection of flow state and voice-carried emotions highlights the intricate relationship between psychological experiences and communicative cues. Understanding how emotions conveyed through voice contribute to the emergence and maintenance of a flow state opens new avenues for interventions in education, performance, and therapeutic settings. Future research could delve deeper into the neurological mechanisms that underlie this connection, shedding light on the intricate interplay between emotional processing, cognitive engagement, and optimal performance.

References:

[1] Csikszentmihályi M., *FLOW: The Psychology of Optimal Experience*. Harper and Row, 1990. ISBN: 978-0-06-016253-5.

[2] Nakamura J., Csikszentmihályi M., *"Flow Theory and Research"*. In Snyder CR, Lopez SJ (eds.). Handbook of Positive Psychology. Oxford University Press. pp. 195–206, 20 December 2001. ISBN: 978-0-19-803094-2.

[3] Csikszentmihályi M., *"The flow experience and its significance for human psychology"*. *Optimal experience: psychological studies of flow in consciousness*. Cambridge, UK: Cambridge University Press. pp. 15–35, 1988. ISBN: 978-0-521-43809-4.E.

[4] Wrigley W. J., Emmerson S. B., "The experience of the flow state in live music performance". *Psychology of Music*., 41 (3): 292–305, May 2013. doi:10.1177/0305735611425903. S2CID 144877389.

[5] O'Neill S., "Flow Theory and the Development of Musical Performance Skills". *Bulletin of the Council for Research in Music Education*. 141 (141): 129–134. JSTOR 40318998 – via University of Illinois Press, 1999.

[6] Csikszentmihályi M., Flow: *"The Psychology of Happiness"*. Rider. 1992. ISBN: 978-0-7126-5477-7.

[7] Landhäußer A., Keller J., "Flow and its affective, cognitive, and performance-related consequences.". In Engeser S (ed.). *Advances in flow research*. New York, NY.: Springer. pp. 65–85, 2012, doi:10.1007/978-1-4614-2359-1_4. ISBN: 978-1-4614-2359-1.

[8] Csíkszentmihályi M. (2004), *Good Business: Leadership, Flow, and the Making of Meaning*, Penguin Books.

[9] Bruya B., *"Effortless Attention: A New Perspective in the Cognitive Science of Attention and action"*. Bradford Book. Cambridge, Mass.: MIT Press, 2010. ISBN: 978-0-262-26943-8. OCLC 646069518.

[10] Ashinoff Brandon K., Abu-Akel A., (2021-02-01). "Hyperfocus: the forgotten frontier of attention". *Psychological Research*. 85 (1): 1–19. doi:10.1007/s00426-019-01245-8. ISSN: 1430-2772. PMC 7851038. PMID 31541305.

[11] Chen D., Haviland-Jones J., Human olfactory communication of emotions, *Perceptual and Motor Skills*, 2000, 91 (3), pp.771-781.

[12] Payne B. R., Jackson J. J., Noh S. R., Stine-Morrow E. A.. In the zone: flow state and cognition in older adults. *Psychology and aging*, 26 3, 738-43, 2011.

[13] Towey, C.A. Flow. *The Acquisitions Librarian*, 13, 131 – 140, 2000.

[14] Knierim, M.T., Rissler, R., Dorner, V., Maedche, A., Weinhardt, C. (2018). The Psychophysiology of Flow: A Systematic Review of Peripheral Nervous System Features. In: Davis, F., Riedl, R., vom Brocke, J., Léger, PM., Randolph, A. (eds) *Information Systems and Neuroscience. Lecture Notes in Information Systems and Organisation*, vol 25. Springer, Cham. https://doi.org/10.1007/978-3-319-67431-5_13.

[15] Gold J. R., Ciorciari J., A Review on the Role of the Neuroscience of Flow States in the Modern World. *Behavioral Sciences*, 10(9): 137, 2022, doi: 10.3390/bs10090137.

[16] Towey C. A., Katz B., Flow: The Benefits of Pleasure Reading and Tapping Readers' Interests. In Readers, *Reading, and Librarians* (1st ed.). Routledge, 2001. doi: 10.4324/9781315862309.

[17] Schiepe-Tiska, A., *In the Power of Flow: The Impact of Implicit and Explicit Motives on Flow Experience with a Special Focus on the Power Domain*, 2013, [Online]. https://d-nb.info/1035502828/34 (Accessed Date: February 21, 2024).

[18] Sodhi K., Luthra M., Mehta D., Yerkes-Dodson Law for Flow: A Study on the Role of Competition and Difficulty in the Achievement of Flow. *International Journal of Education and Management Studies*, 6, 95, 2016.

[19] LeFevre J., *Flow and the quality of experience during work and leisure*. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), Optimal experience: Psychological studies of flow in consciousness (pp. 307–318). Cambridge University Press, 1988.

[20] Jackson S., *Flow, and mindfulness in performance*. In A. L. Baltzell (Ed.), Mindfulness and performance (pp. 78–100). Cambridge University Press, 2016. https://doi.org/10.1017/CBO9781139871310.005.

[21] Biasutti M., *Flow and Optimal Experience*. In M. A. Runco, & S. R. Pritzker (Eds.), Encyclopedia of Creativity (2nd ed., Vol. 1, pp. 522-528). London: Academic Press, 2011.

[22] Carter L., River B., Sachs M., Flow in Sport, Exercise, and Performance: A Review with Implications for Future Research. *Journal of Multidisciplinary Research*, 5, 17, 2013.

[23] Juslin P. N., Laukka P., Communication of emotions in vocal expression and music performance: different channels, same code?. *Psychological Bulletin*, 129(5), 770–814, 2003. doi: 10.1037/0033-2909.129.5.770.

[24] Chen M., Bargh J. A., Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, 25(2), 215–224, 1999. doi: 10.1177/0146167299025002007.

[25] Johnson A. M., Eerola T., Huovinen E., Flow as a musical emotion: A replication and extension of Zentner and Eerola (2010). *Music Perception: An Interdisciplinary Journal*, 34(2), 219-234, 2017.

[26] Thoma M. V., La Marca R., Brönnimann R., Finkel L., Ehlert U., Nater U. M., The effect of music on the human stress response, 2013, PLoS ONE 8(8): e70156. https://doi.org/10.1371/journal.pone.0070156.

[27] De Pinto M. G., Polignano M., Lops P., Semerano G., Emotions Understanding Model from Spoken Language using Deep Neural Networks and Mel-Frequency Cepstral Coefficients, *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 2020, Bari, Italy, 27-29 May 2020.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The 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.

Conflict of Interest

The authors have no conflicts of interest to declare.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US