Jonas Schewior*, Roman Grefen, Rodolfo Verde, Alina Ergardt, Ying Zhao and Walter H. Kullmann

Speech-Controlled Robot Enabling Cognitive Training and Stimulation in Dementia Prevention for Severely Disabled People

https://doi.org/10.1515/cdbme-2024-2133

Abstract: The current German medical S3-guideline on dementia recommends the use of cognitive training and stimulation, e.g. through shared games, for mild cognitive impairment and mild dementia. The use of cobots, which allow direct human-robot collaboration, enables people with paralyzed upper limbs to actively participate in social activities. This research presents a novel approach through the development of a voice-controlled board game specifically tailored to the inclusion needs of people with severe disabilities. The human voice commands recorded via USB microphones are digitally filtered. Speech recognition of the control commands is performed using a Convolutional Neural Network (CNN) based on the VGG-16 architecture. The robot´s activity is controlled utilizing the ROS 2 robot operating system. A portable table-based complete system with a 3D-printed Tic-Tac-Toe playing field and robot assistant for severely disabled paralyzed people has been developed. The robot control system employs a pick-and-place mechanism, seamlessly integrating with speech recognition to enhance gameplay interaction. The CNN model achieves an impressive accuracy rate of 97%, ensuring reliable speech recognition performance throughout gameplay. The targeted integration of robotic technologies and artificial intelligence opens new avenues in the prevention of mental illness, care support and inclusion of older and disabled people.

Keywords: Healthcare-Robotics, Vulnerable disabled people, Speech-Control, Convolutional Neural Networks (CNN), Dementia prevention

***Corresponding author: Jonas Schewior:** Center Robotics (CERI), Technical University of Applied Sciences Wuerzburg-Schweinfurt, Konrad-Geiger-Strasse 2, Schweinfurt, Germany, e-mail: jonas.schewior@study.thws.de

Roman Grefen, Rodolfo Verde, Alina Ergardt, Ying Zhao, Walter H. Kullmann: Center Robotics (CERI), Technical University of Applied Sciences Wuerzburg-Schweinfurt, Schweinfurt, Germany

1 Introduction

The usage of speech control in medical robot-assisted applications varies depending on the specific task at hand, with different systematic approaches tailored to each application. Currently, speech-controlled systems are employed in various medica, such as minimal invasive endoscopic procedures [1], instrument transfer during orthopedic operations [2], and hysterectomies [3]. In therapeutic and rehabilitation settings, robotic systems have shown potential in improving social and cognitive abilities in patients, including those with severe disabilities [4]. Implementing gaming elements for this vulnerable group could enhance their cognitive capabilities [5] and potentially decelerate symptoms of conditions like mild dementia. The German medical S3-guideline on dementia recommends the use of cognitive training and stimulation, e.g. through shared games, for mild cognitive impairment and mild dementia [6]. Introducing such interventions within the framework of robotic systems not only augments their therapeutic potential but also offers avenues for personalized and engaging rehabilitation experiences. Consequently, speech control can be utilized to operate the robot and facilitate gameplay for the user. This paper introduces a novel approach: a Table-top and speech-controlled robotic system integrated with the classic game of Tic-Tac-Toe. This system is meticulously designed to cater to the needs of individuals with severe disabilities, e.g. paralyzed upper limbs, offering both single player matches against the robot and multiplayer functionality for interactive sessions between two users. Through the utilization of speech control, users can seamlessly operate the robot and engage in gameplay, thereby fostering cognitive training and stimulation within an immersive and accessible framework.

By exploring the fusion of speech control, robotics, and gamification, this research aims to contribute to the burgeoning field of robot-based assistive technologies, particularly in the realm of inclusion, cognitive training, and dementia prevention.

2 System Overview

To realize such a system, the project was compartmentalized into three distinct groups: Tic-Tac-Toe board creation utilizing 3D-printing technology, robot control within the software environment ROS2, and speech control employing convolutional neural networks (CNN) based on the VGG-16 architecture [7]. Each of these focus groups necessitates distinct hardware and software applications to fulfill its respective objectives.

2.1 Hardware

The hardware utilized for the implementation of the project, which can be seen in Figure 1, is as follows:

1. Robot-Arm (ReactorX 150, 5 DoF, arm span: 630 mm, working payload: 100 g, Trossen Robotics) to manipulate the figures.
2. One Raspberry Pi-4b microcomputer for speech classification and the robot control.
3. Two LCD displays (16x2) with an I2C interface for visualizing game information and speech recognition.
4. Two Lavelier Microphones (SJ-B02, Sujeetec) for speech detection.
5. A wooden board (400x800x18 mm) serving as the system's foundation.
6. A 3D-printed Tic-Tac-Toe playing field (240x240x5 mm) with 12 circular figures (6 blue, 6 red) with a radius of 20 mm and a height of 6 mm as game figures.
7. Additional cases for housing the Raspberry Pi and managing cabling and power supply.

Figure 1: Hardware setup of Tic-Tac-Toe speech-controlled robotic assistant system.

Each component is meticulously integrated to create a portable table-based game board. The system requires a two-slot power socket for the robot and Raspberry Pi to function.

2.2 Software

To maintain a unified software environment Python is used in every implementation of the system. Therefore, the TensorFlow library is used to implement the Deep Learning CNN model for the speech recognition. The training data is saved as .wav format with the audio editor and recorder Audacity.

The Raspberry Pi is suited with Ubuntu 22.04 and ROS2 Humble for controlling the Robot.

3 Speech Control

For the game implementation of the system via German speech commands, the commands must be categorized into the game mechanics. Therefore, nine different classes/commands are constructed (six for positioning and three for control).

- "A", "B", "C", "Eins", "Zwei", "Drei" for the coordinates of the Tic-Tac-Toe game field
- "Rex", as the starting command for the robot to process the coordinate command afterwards.
- "Stopp", as the safety command to stop the robot immediately in case of an accident and for the user to change the coordinate command in case necessary.
- "Other", which is used as the class belonging to any other word other than the commands of the game.

These commands build the necessary classes to be classified for the system.

3.1 Data gathering and preprocessing

To effectively operate the CNN, a substantial amount of training data must be generated. This process begins with the creation of audio files using Audacity, followed by manual labeling.

For the command-based approach, specific preprocessing steps are essential to ensure the smooth operation of the system. To this end, a dataprocessor is developed to preprocess the input data and ensure that a word serves as the output of the class and input to the CNN. The preprocessing steps are as follows:

1. **Data filtering:** Utilizing Band-pass filters (30-3400 Hz) within the dominant frequency range of human speech to enhance data quality.
2. **Automatic Gain Control (AGC):** Normalizing the input data to a specific sound pressure level (-30 dB) range through AGC to ensure uniformity.
3. **Voice Activity Detection (VAD):** Employing VAD to detect human voice (-40 dB) within the input data, facilitating accurate recognition.
4. **Word Extraction:** Extracting each command from the preprocessed data, ensuring precise input for classification by the CNN.

Following the preprocessing steps, the extracted words undergo transformation into Mel Frequency Cepstral Coefficients (MFCC) to serve as input for the CNN model. In addition to the manually created data (five male voices, five female voices, ages 20-70 years), augmentation techniques [8] are employed on the dataset to expand the training data size of 5,900 instances, resulting in a total of 61,072 instances. Figure 2 showcases one MFCC image of the command word “Rex”.

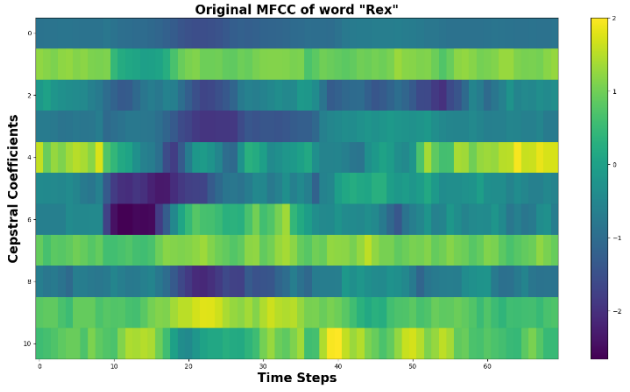

Figure 2: The original MFCC image without augmentation of a command word “Rex” with all 11 cepstral coefficients after removing the first coefficient due to it containing the pitch.

3.2 Neural network

The classification method utilized in this study employs a command-based CNN classification approach. Therefore, a model architecture had to be developed which could quickly classify a complex given image. Our approach entails the utilization of a network, derived from the VGG-16 architecture [7]. The network receives as input a (two-dimensional) MFCC image representing the extracted (one-dimensional) letter sequence from the dataprocessor and generates as output the probability of the word belonging to each of the nine predefined classes (A, B, C, 1, 2, 3, Rex, Stop, Other). The maximum probability translates to the understood word for the system.

This classified word is passed over to a game logic module which ensures the rules and commands eligible for the game.

3.2.1 Neural network evaluation and Raspberry Pi integration

Due to the significant computational demands, the model is trained utilizing the V100-GPU provided by Google Colab Pro, owing to its ample 14.8 GB of GPU-RAM. Through this resource, the training process achieves an accuracy of 97% after 60 epochs with the dataset which is shown in Figure 3.

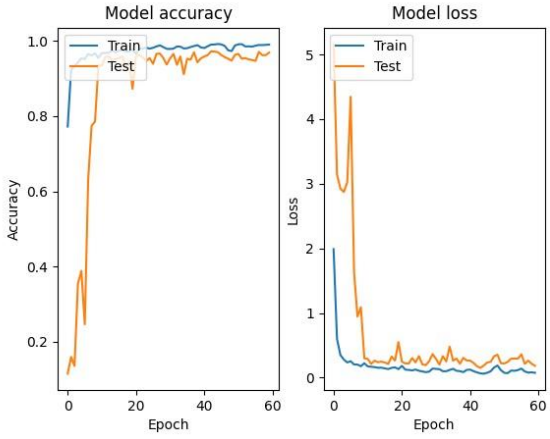

Figure 3: Training accuracy and loss graph of model based on the VGG-16 architecture with 61,072 instances of data.

Following initial testing of the model on the Raspberry Pi, it is observed that the accuracy closely aligns with the training accuracy. However, a notable concern arose regarding the considerable time and computational resources required for the Raspberry Pi to perform predictions using the original 746 MB-sized model, sometimes necessitating up to 240 ms for each classification.

In response to this issue, a smaller model is developed, achieving a size of 11 MB while retaining the 97% accuracy. This optimized model is derived from the VGG-16 architecture, however, compresses the model by reducing the number of convolutions in combination with MaxPooling-layers. This significantly expedited the prediction process, requiring only around 50 ms per classification.

To bolster the speech classification process, queuing is implemented, a threading technique, ensuring swift predictions crucial for seamless Tic-Tac-Toe gameplay. This optimization enhances system responsiveness, facilitating a smooth user experience.

4 Application and Discussion

This system development focuses on the care sector for older, severely disabled people with paralyzed upper extremities. A robotic assistance system is intended to support the inclusion of this vulnerable group and at the same time enable dementia prevention in a playful way, or, in the case of mild dementia, slow down the progression of the disease by means of cognitive training and cognitive stimulation. The players control the robot arm using speech, detected via microphones. Player safety is guaranteed as there is no physical contact with the robot arm. A Raspberry Pi single-board computer transforms the voice commands into control commands for the robot. The robot arm places the 3D printed game pieces in the corresponding game positions. The LCD displays on the microphone stands indicate the currently active player and the coordinates of the gaming field to be occupied. The robot automatically recognizes the end of the game when one party wins or when the game is tied and then cleans up the game pieces independently. The outcome is shown on the displays.
Overall, the system allows both joint games with two game participants and a game version in which a single participant plays against the robot.
In addition to the game, the robot assistant enables the caregiver to be relieved. To start the game for the first time, the caretaker only must switch on the entire system, which is mounted on a wooden board, using an electrical switch. Once the system is on there is no further external inputs needed for the game to be played repeatedly.

5 Conclusion

In conclusion, this research has presented a novel approach to cognitive training and stimulation in the context of dementia prevention for individuals with severe disabilities, such as paralyzed upper limbs. By integrating speech control, robotics, and gamification, the study demonstrates the potential of assistive technologies to enhance the quality of life for vulnerable populations.
In a playful way, the robotic assistance system allows cognitive training and cognitive stimulation to promote attention, concentration, and problem solving for people with restricted freedom of movements of their upper limbs.
In addition, the robot system relieves the nursing staff because after the game is initially switched on, the game participants can act autonomously with the robot via voice control.
Looking ahead, the findings of this study lay the groundwork for further advancements in the field of assistive technologies. The Table-top and voice controlled robot system enables persons with impaired or paralyzed upper limbs to play board games for cognitive training. Upcoming testing in nursing homes will evaluate the system in practice.

Author Statement
Research funding: The authors state no funding involved. Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study.

References

[1] K. Zinchenko, C. -Y. Wu and K. -T. Song, "A Study on Speech Recognition Control for a Surgical Robot," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 607-615, April 2017, doi: 10.1109/TII.2016.2625818.W. Khaewratana, E. S. Veinott and S. M. Ramkumar, "Development of a Generalized Voice-Controlled Human-Robot Interface: One Automatic Speech Recognition System for All Robots," *2020 3rd International Conference on Control and Robots (ICCR)*, Tokyo, Japan, 2020, pp. 38-42, doi: 10.1109/ICCR51572.2020.9344123.

[2] A. S. Nakhusheva, R. S. Nakhushev, A. M. Sabanchiev, V. N. Konstantyan and T. I. Kuliev, "Surgical Instrument Supply Automated System," *2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, Sochi, Russia, 2019, pp. 363-366, doi: 10.1109/ITQMIS.2019.8928369.

[3] M. Fang, P. Li, L. Wei, X. Hou and X. Duan, "Voice Control of a Robotic Arm for Hysterectomy and Its Optimal Pivot Selection," *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, Irkutsk, Russia, 2019, pp. 644-649, doi: 10.1109/RCAR47638.2019.9043990.

[4] N. Guo *et al*., "SSVEP-Based Brain Computer Interface Controlled Soft Robotic Glove for Post-Stroke Hand Function Rehabilitation," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1737-1744, 2022, doi: 10.1109/TNSRE.2022.3185262.

[5] E. Dulau, C. R. Botha-Ravyse, M. Luimula, P. Markopoulos, E. Markopoulos and K. Tarkkanen, "A virtual reality game for cognitive impairment screening in the elderly: a user perspective," *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Naples, Italy, 2019, pp. 403-410, doi: 10.1109/CogInfoCom47531.2019.9089973.

[6] DGN e.V. & DGPPN e.V. (Hrsg.), S3-Leitlinie Demenzen, Version XX, 8.11.2023, https://register.awmf.org/de/leitlinien/detail/038-013 (last access: March 28, 2024)

[7] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale ImageRecognition.” arXiv, Apr. 10, 2015. Accessed: Feb. 28, 2024. [Online]. Available: http://arxiv.org/abs/1409.1556

[8] S. Wei, S. Zou, F. Liao, and W. Lang, “A Comparison on Data Augmentation Methods Based on Deep Learning for Audio Classification,” *J. Phys. Conf. Ser.*, vol. 1453, no. 1, p. 012085, Jan. 2020, doi: 10.1088/1742-6596/1453/1/01208