Radar-Based Gesture Recognition Towards Supporting Communication in Aphasia: The Bedroom Scenario

Luís Santana, Ana Patrícia Rocha(✉), Afonso Guimarães, Ilídio C. Oliveira,
José Maria Fernandes, Samuel Silva, and António Teixeira(✉)

Department of Electronics, Telecommunications and Informatics,
Institute of Electronics and Informatics Engineering of Aveiro,
University of Aveiro, Aveiro, Portugal
{luis.santana,aprocha,afonso.guima,ico,jfernan,sss,ajst}@ua.pt

Abstract. Aphasia and other communication disorders affect a person's daily life, leading to isolation and lack of self-confidence, affecting independence, and hindering the ability to express themselves easily, including asking for help. Even though assistive technology for these disorders already exists, solutions rely mostly on a graphical output and touch, gaze, or brain-activated input modalities, which do not provide all the necessary features to cover all periods of the day (e.g., night-time). In the scope of the AAL APH-ALARM project, we aim at providing communication support to users with speech difficulties (mainly aphasics), while lying in bed. Towards this end, we propose a system based on gesture recognition using a radar deployed, for example, in a wall of the bedroom. A first prototype was implemented and used to evaluate gesture recognition, relying on radar data and transfer learning. The initial results are encouraging, indicating that using a radar can be a viable option to enhance the communication of people with speech difficulties, in the in-bed scenario.

Keywords: Smart environments · Communication · Gestures · FMCW radar · In-bed scenarios · Aphasia

1 Introduction

People suffering from communication impairments have much more difficulty expressing their needs in ways that other people can understand. These difficulties can lead to problems socialising and limit the person's independence, namely in asking for help when needed.

Existing assistive technology for augmenting or replacing speech includes devices providing a graphical interface and relying on non-verbal interaction modalities, such as touch, gaze, or brain-activated, together with speech-generation [2]. These solutions require interacting with a given device (e.g., tablet),

---

This work was supported by EU and national funds through the Portuguese Foundation for Science and Technology (FCT), in the context of the AAL APH-ALARM project (AAL/0006/2019), and funding to the research unit IEETA (UIDB/00127/2020).

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 500–506, 2022.
https://doi.org/10.1007/978-3-030-94822-1_30

which may not be easily reached in some situations (e.g., lying in bed), rely on the use of cameras, which raises privacy concerns, and/or are too intrusive.

An alternative approach for assisting communication at a distance is the use of gestures. However, most contributions focus specifically on sign language [8]. Gesture recognition has also been explored in the context of human-computer interaction, relying on wearable sensors [5], vision-based sensors [7], or radars [1,3,4,9]. The latter have advantage of being the less intrusive and also preserving the user's privacy.

The ongoing project APH-ALARM – Comprehensive safety solution for people with Aphasia[1] – aims at allowing people suffering from aphasia (e.g., after a stroke) to communicate more easily with other people anywhere and anytime. In the scope of this project, our main objective is to enhance communication for people with speech difficulties, in the in-bed scenario (i.e., user lying in bed).

Towards this goal, we propose a system based on a Frequency Modulated Continuous Wave (FMCW) radar for supporting communication through gestures, in the considered scenario. A first prototype, where gesture recognition is performed by a model obtained through transfer learning and radar data, was developed to explore the viability of the technology. To the best of our knowledge, gesture interaction for the in-bed scenario, where some patients may spend a large part of their time, has not yet deserved much attention.

2 Radar-Based Gesture Recognition System

We propose the architecture of a system that aids communication when the user is alone in a bedroom, lying in bed, and may need to communicate with other people (e.g., caregiver, family member) to ask for help, for instance. As a first step towards a novel communication support system for patients with aphasia, we present a first prototype for radar-based gesture recognition.

2.1 General Architecture

An overview of the system is depicted in Fig. 1. A radar captures data from the detected targets, in this case the human body. These data are sent to a processing unit, where they are pre-processed by removing outliers. Features are then extracted and used to recognise the gesture being carried out. A final decision is made and sent to a smartphone.

2.2 First Prototype

As a proof-of-concept, we implemented a first prototype that relies on the setup shown on the left side of Fig. 1, which includes a bed and a radar. The radar is elevated 0.55 m from the ground, and placed at 1 m from the bed, on the left side of the subject, parallel to the longest side of the bed. The radar's 2D coordinate system is shown in Fig. 1.

---

[1] https://www.aph-alarm-project.com.

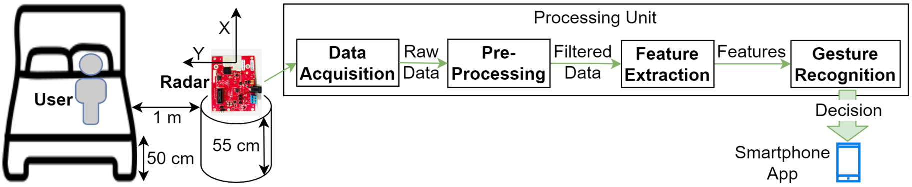

Fig. 1. Overview of the proposed system, including a possible setup for the bed and radar, as well as the pipeline for gesture recognition.

The radar is a Frequency Modulated Continuous-Wave (FMCW) radar from Texas Instruments, the AWR1642, with notable configurations entailing: frame rate (20 fps); resolution for range (4 cm) and radial velocity (0.22 m/s); maximum range (10.28 m) and radial velocity (3.47 m/s); for object detection, peak grouping in the Doppler (on) and range (off) directions, without clutter removal.

Data Acquisition: The data provided by the radar includes the X and Y coordinates, as well the Doppler index, for each detected moving target. In this first prototype, the data captured by the radar are saved to a computer.

Pre-processing: The acquired data are processed using a sliding window of 5 s without overlap. For each window, pre-processing consists of removing outliers corresponding to unwanted reflections or noise. A detected target is considered as an outlier if its Euclidean distance to the radar is outside the interval [0.5, 3] m or its absolute Doppler index is outside the interval [1e−5, 10]. All data samples with X and Y coordinates outside the intervals [−1.5, 1.5] m and [0, 2.25] m, respectively, were also discarded.

Feature Extraction: From the filtered data, three different maps are created, one for each data type versus the elapsed time (X-Time, Y-Time, and Doppler-Time). The beginning and ending of the window where no movement is detected are discarded. An example of the X-Time and Doppler-Time maps for a repetition of the third gesture described below ("Back and Forth") is presented in Fig. 2, where the colour represents the number of detected targets (bright yellow corresponds to the maximum value for each map, while dark blue corresponds to no detected target). The matrix associated to each map is used to obtain a normalised greyscale image. The three images are then combined into a single image (X-Time above Y-Time above Doppler-Time).

Gesture Recognition: The images resulting from feature extraction are fed into a model that performs gesture recognition. This model is previously trained using the transfer learning method, relying on a pre-trained deep neural network model for image classification, and a given dataset. For this prototype, the focus was on the recognition of the following three arm gestures, all starting with the arm parallel to the body and resting on the bed: (a) **Wave** – Move the arm and hand from left to right and back; (b) **Raise Arm** – Raise the arm until a 90° angle is formed with the body and then lower it back to initial position; (c)

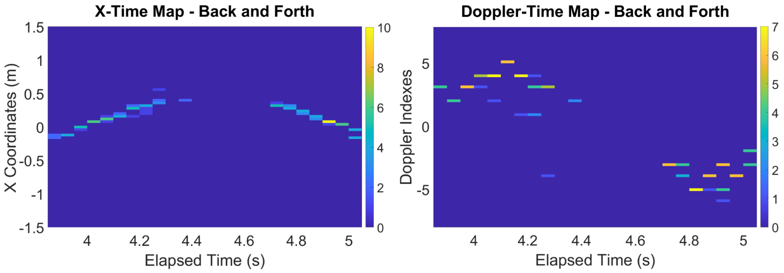

Fig. 2. Example of X-Time (left) and Doppler-Time (right) maps, for the "Back and Forth" gesture. (Color figure online)

Back and Forth – "Come to me" motion, where the forearm is moved towards the arm making an angle below 90°, and then returning to full extension.

These gestures were selected aiming at simplicity and based on initial feedback from therapists and carers on the gestures' suitability for aphasic patients lying in bed. Moreover, they can be used for generating simple messages (e.g., "I need help") and "Yes/No" answer.

3 Evaluation

An initial evaluation of the prototype was performed to explore the possibility of recognising the defined gesture set using radar data together with transfer learning, in the context of the in-bed scenario.

Subject and Experimental Protocol: Radar data were captured from a 23-year-old, right-handed male subject. The used setup is the one included in Fig. 1, where the subject was lying in bed on their back. Each considered gesture was executed 50 times. Even though the subject is right-handed, all gestures were performed with the left arm, due to the radar being on the left side of the bed. For each repetition, data recording was initiated before the gesture execution and stopped automatically after 5 s.

Datasets: The obtained dataset includes 150 images (50 per gesture). Since deep learning requires a large dataset to obtain reasonable results, we expanded the dataset relying on offline data augmentation, to obtain a better performance and avoid overfitting. For each image in the original dataset, 5 or 10 new images were created by adding noise to that image (resulting in two augmented datasets). The type of noise added to the image was randomly chosen and can be a combination of the following types: Gaussian, salt and pepper, and Poisson. For all except Poisson, the amount of noise was limited to a proportion of image pixels to replace of 0.002 (chosen empirically).

Gesture Recognition Models: To obtain a model that recognises the considered gestures, we used the transfer learning method. Since our aim is to run gesture recognition in a processing unit with limited memory and computing capability, from the pre-trained models directly available in Keras [6], we explored three that achieved a top-5 accuracy equal or greater than 90% (ImageNet validation dataset) and have less than 10 million parameters: MobileNetV2, NASNet-Mobile, DenseNet121. For each pre-trained model, the top layers were replaced by a single fully connected layer with 256 neurons (ReLU as the activation function) and an output layer with 3 neurons (softmax activation function). The used optimizer was ADAM (default parameters). Crossentropy was used as the loss function, and accuracy as the evaluation metric during training and validation.

Evaluation Method: Each model was evaluated using the 10-fold cross-validation approach, where 80% of the dataset is used for training, 10% for validation, and 10% for testing, in each iteration. Training is stopped when the validation loss has not decreased more than 0.1 for 5 epochs. The resulting model is evaluated on the test data of the corresponding iteration.

4 First Results

Results were obtained for the three pre-trained models listed above and for three different datasets: original (150 images); augmented 1 (750 images); augmented 2 (1500 images). The boxplots for the accuracy, F1 score, train time, and prediction time per image, considering all 10 folds, are shown in Fig. 3.

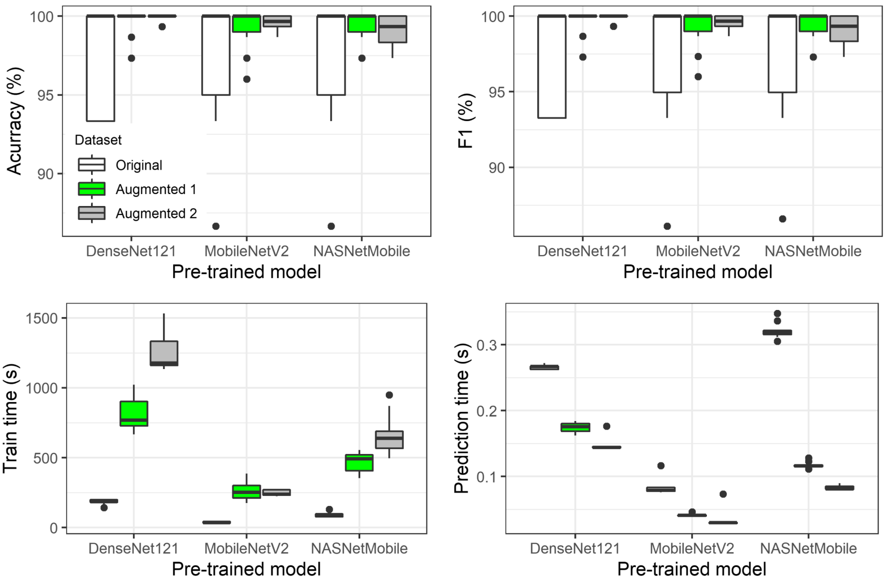

Fig. 3. Boxplots for the accuracy (left-top), F1 score (right-top), train time (left-bottom), and prediction time (right-bottom), for each model and dataset.

We can see that augmenting the data has an overall positive effect when it comes to the variability of accuracy and F1 score, for all three models. On the other hand, the train time increases when the size of the dataset increases, as expected, but the prediction time per image decreases. For both train and prediction times, the difference among datasets is lower for the MobileNetV2 model, which also has the lowest median train and prediction times: 35 to 252 s and 0.03 to 0.08 s, respectively, versus 84 to 639 s and 0.08 to 0.32 s for NASNetMobile, and 185 to 1177 s and 0.14 to 0.27 s for DenseNet121. This was also expected, since MobileNetV2 is the smallest of the three pre-trained models (≈3.5 M parameters), followed by NASNetMobile (≈5.3 M parameters; DenseNet21 has ≈8.1 M). Despite its smaller size, MobileNetV2 still leads to a model with a median accuracy and F1 score similar to the other models (≥99% for all datasets). Although these results are quite good, it can be because only three gestures were considered and all used data came from the same subject.

5 Conclusion and Future Work

Our long-term research goal is the implementation of gesture-based communication support system for people with speech difficulties, such as aphasics. This system would provide its users with a more assisted and independent life, including at night-time. Our initial results on gesture recognition are in line with those reported in other similar contributions using radars (in scenarios different from the in-bed setting) [3,4,9]. They show the feasibility of recognising a simple set of gestures, in the specific in-bed scenario, based on a radar, which is not invasive or intrusive and can be placed in the environment.

Our study has some limitations, such as a small dataset limited to one subject and three gestures. However, we intend to obtain a larger dataset including more gestures and data from a greater number of subjects. This dataset will allow us to investigate if a model trained with data from a given subject(s) can be used to recognise gestures performed by never seen subjects.

References

1. Ahmed, S., Kallu, K.D., Ahmed, S., Cho, S.H.: Hand gestures recognition using radar sensors for human-computer-interaction: a review. Remote Sens. **13**(3), 527 (2021)
2. Elsahar, Y., Hu, S., Bouazza-Marouf, K., Kerr, D., Mansor, A.: Augmentative and Alternative Communication (AAC) advances: a review of configurations for individuals with a speech disability. Sensors **19**(8), 1911 (2019)
3. Hazra, S., Santra, A.: Robust gesture recognition using millimetric-wave radar system. IEEE Sens. Lett. **2**(4), 1–4 (2018)
4. Ishak, K., Appenrodt, N., Dickmann, J., Waldschmidt, C.: Human gesture classification for autonomous driving applications using radars. In: IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), pp. 1–4, November 2020
5. Jiang, S., Kang, P., Song, X., Lo, B., Shull, P.B.: Emerging wearable interfaces and algorithms for hand gesture recognition: a survey. In: IEEE Reviews in Biomedical Engineering, p. 1 (2021)

6. Keras: Keras applications. https://keras.io/api/applications/
7. Wang, T., et al.: A survey on vision-based hand gesture recognition. In: Basu, A., Berretti, S. (eds.) Smart Multimedia, pp. 219–231. Springer, Cham (2018)
8. Yasen, M., Jusoh, S.: A systematic review on hand gesture recognition techniques, challenges and applications. Peer J. Comput. Sci. **5**, e218 (2019)
9. Yu, M., Kim, N., Jung, Y., Lee, S.: A frame detection method for real-time hand gesture recognition systems using CW-radar. Sensors **20**(8), 2321 (2020)