Wi-Nod: Head Nodding Recognition by Wi-Fi CSI Toward Communicative Support for Quadriplegics

Marwa R. M. Bastwesy[1,2], Kiichiro Kai[3], Hyuckjin Choi[1], Shigemi Ishida[4], Yutaka Arakawa[1]
[1]*ISEE, Kyushu University, Fukuoka 819-0395, Fukuoka, Japan*
[2]*CCE, Tanta University, Tanta, Egypt*
[3]*EECS, Kyushu University, Fukuoka 819-0395, Fukuoka, Japan*
[4]*Dept. Media Architecture, Future University Hakodate, Hokkaido, Japan*
{bastwesy.marwa.377, kai.kiichiro.418}@s.kyushu-u.ac.jp,{choi,arakawa}@ait.kyushu-u.ac.jp, ish@fun.ac.jp

2023 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-9122-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCNC55385.2023.10118666

***Abstract*—Recently, the studies of wireless device-free human sensing technology have dramatically advanced with enabling a variety of applications, from activity recognition to vital sign monitoring. In this paper, we propose Wi-Nod which leverages the Wi-Fi Channel State Information (CSI) to detect head nodding gestures for each Morse code symbol based on time-frequency features for accurate recognition accuracy in multi-human context environment. The system consists of three basic modules: data collection, data preprocessing, and learning part based on the inception model. The model was trained to perform the head movement detection based on the CSI spectrogram collected by the ESP32 nodes. We evaluated the performance of the system on four different data sets collected in two different sessions. Our system achieves over 95% recognition accuracy that reveals the feasibility of Wi-Nod system for real-life deployment.**

***Index Terms*—Wi-Fi CSI, head gesture recognition, signal processing, quadriplegic, deep learning**

I. Introduction

In 2016, the American Spinal Injury Association conducted a study revealing that between 1.3 and 2.6 million disabilities cases have spinal cord injuries of different degrees every year, which affects their mobility and leads to quadriplegia [1]. Human activity recognition (HAR) systems could help quadriplegia patients whose limbs are impaired communicate with others easier. Recently, HAR sensing techniques have gained a lot of attention because they aim to serve a variety of applications such as fall detection [2], human-computer interaction [3], indoor localization [4], etc. There is rapid development in sensing techniques since they can be classified into three categories, namely, vision-based [5], wearable sensor-based [6], and wireless-based mechanisms [7]. Vision-based systems can achieve an acceptable recognition accuracy, however they require sufficient lighting conditions and high computing performance, can only detect the objects in the line-of-sight scenario (LOS), and have some privacy issues. Although wearable sensor-based systems are very light and inexpensive, they can be fatal if the user forgets to wear them, especially in healthcare applications. Due to the restricted coverage of the camera-based system and user inconvenience of the wearable sensors, the wireless sensing mechanism gains considerable attention as it leverages the radio frequency (RF) signals to detect, identify, and recognize the objects in LOS and Non-LOS without the need of wearing any special device.

In this paper, we introduce and validate a Wi-Nod, a contactless sensing system, with ESP32 nodes as a CSI toolkit. To the best of our knowledge, this is the first work that collected the Wi-Fi signals in multi-human context environment. It means that there is not only a target patient, but also a caregiver who is always along with a quadriplegic in the real-world scenario and acts as a scatter which provides more multi-path propagation in the sensing area, unlike other existing studies that collected the data only in a single user environment. By using Wi-Fi CSI, we represent Morse symbols, dot and dash, by moving the head down and right, we also add a third symbol, space, to separate a word from the previous word, by moving the head to the left. Specifically, the proposed system extracts and analyzes variations in the amplitude of each symbol to represent a signature for each head motion. Then, a learning model based on the inception module is used to perform the head motion classification.

By this work, our proposed system has addressed various challenges. First, extracting the informative context is a difficult task because of the existence of multiple objects around the patient. Second, wireless signals are influenced by the temperature and humidity of the sensing area. Finally, different patients can perform the same motion at different speeds which affected the variations of the signals. To tackle these challenges and verify the robustness, we remove the outliers, smooth the amplitude, and convert it to the time-frequency domain to capture more meaningful information which is fed as input to the efficient learning algorithm for the classification. The main contributions of this paper are as follows.

- We examined the possibility of Wi-Fi CSI to become a base of head motion-based Morse code system that supports Quadriplegics' communication in multi-human context environment, for the first time.
- We evaluated the performance of the proposed systems using practical data collected from the two participants, one male and the other female wearing Hijab which can be represented as a scatter.
- The performance metric shows that the Wi-Nod can achieve accuracy up to 95% for different persons. These results emphasize that the system can be deployed in real-life scenario to make communication between persons with quadriplegia and others easier.

The rest of the paper is structured as follows: Section II provides related works of Wi-Fi CSI sensing. Section III represents our system design including signal preprocessing and learning algorithm. Section IV describes our experimental

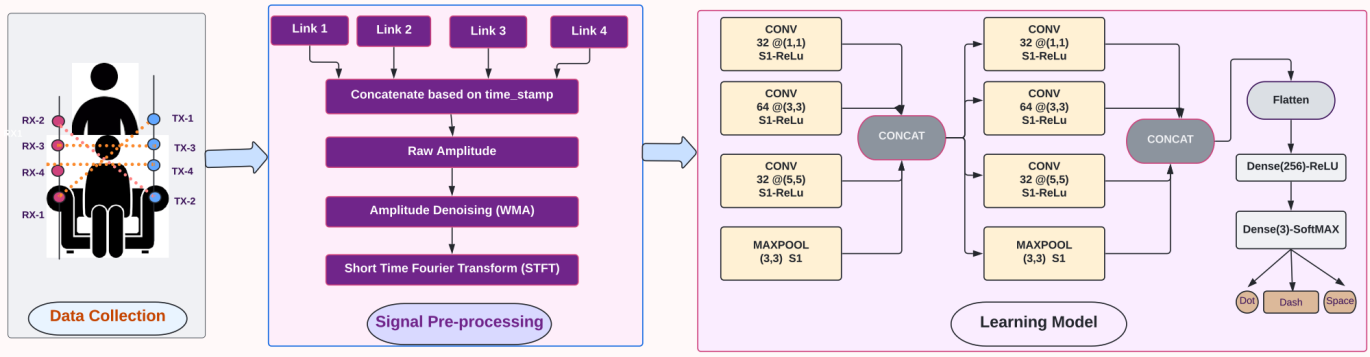


Fig. 1: Wi-Nod System Framework

setup and results. Section V discusses the impact of the different links configuration, time and user diversity. In Section VI, we conclude our work and discuss our future plans.

II. Related Work

The Wi-Fi CSI sensing mechanism has been involved in a wide variety of applications, e.g. gesture recognition [8], localization [9], and health care [10]. Several works based on CSI measurements are listed below.

ViHOT [11] is a passive wireless driver head tracking system based on the CSI phase. The system consisted of two phases, namely, profiling and run-time. The profiling stage wad responsible for collecting CSI data of the driver's head positions and orientations. In the run-time tier, the system mapped the CSI phase readings to unique patterns to perform the tracking task.

The WiHead system was introduced by [12] is the CSI-based system for tracking human head orientation in various directions yaw, roll, pitch, and different combinations of them to get a user feedback of online courses. WiHead used the Atheros CSI extractor tool [13] with 56 subcarriers at 2.4 GHz. It extracted both phase and amplitude from the CSI signals and applied the denoising techniques, low-pass filter for amplitude noise removal, and phase calibration algorithm for removing the randomness of the phase. The combination of filtered amplitude and phase is then fed to the principal component analysis (PCA) algorithm for dimensional reduction since the adjacent subcarriers are very correlated. However, important information can be lost in this process. furthermore, WiHead built a CNN model and achieved a 90% k-fold cross-validation for 3 head motion angles: pitch, roll, and yaw.

WiSense [14] proposed a human activity recognition system that includes four activities, namely picking up objects from the floor, falling on a mattress, sitting on a chair, and walking indoors, based on Wi-Fi CSI. Firstly, the authors collected the data from nine volunteers using 2 laptops each equipped with Intel 5300 NIC to use one as a transmitter with one external antenna and the other as a receiver with two external antennas. The system contained three modules, the first for data preprocessing using the CSI ratio for phase correlation followed by PCA to reduce the dimensionality, and finally low pass filter to remove high-frequency noise caused by the environment. The second stage is to use STFT to compute the spectrogram of each activity and save it as a PNG image with $224\times224\times3$ dimension to be used as input to the final module. the final module is a CNN classifier which includes 14 layers. The system achieves a 97.78% recognition accuracy.

EfficientFi [15] investigated the performance of HAR and human identification based on CSI compression based on VQ-VAE. the authors evaluated the accuracy of the system using different compression rates and achieved better performance when increased the number of embedding space which means the compression rate is low because when the compression rate increases this means loss of more information. It achieved 98% and 83% accuracy for HAR and human identification, respectively.

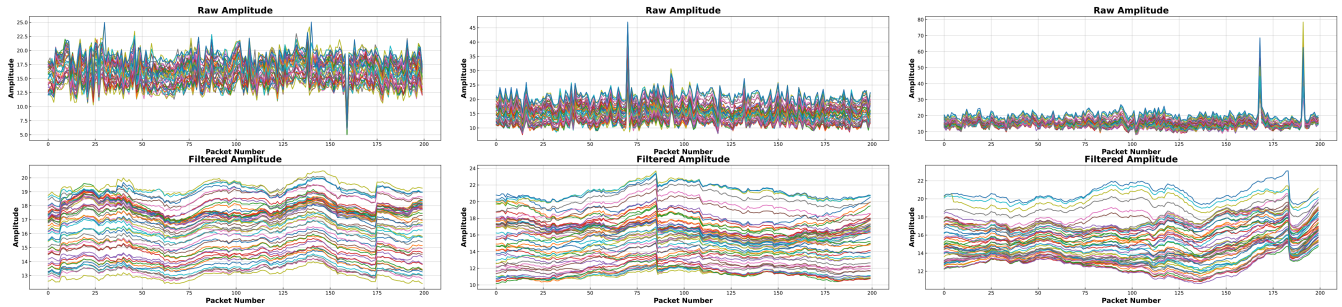
Lastly, our idea is inspired by the WiMorse [16] that employed Intel 5300 NIC to collect the CSI waveforms produced by a finger. The authors created their own code that encoded the two Morse symbols base on the subtle finger movements. The authors built a mathematical model to detect the characters and numbers via WiFi CSI measurements generated from the finger movements. WiMorse is a position-independent system that can be deployed in different environments, and the system achieved an average accuracy of 95%.

III. System Design

The challenge of this paper is to use Wi-Fi CSI waveforms to recognize head motion in multi-human context environment since persons act as scatters. Also, different users can perform the same head nodding at different speeds which provides different CSI patterns as it depends on the face shape and head size. Additionally, the person moving behind, assuming a caregiver is carrying the wheelchair, affects the received signal. We apply simple amplitude denoising and extract the Doppler spectrogram since the Doppler of the static objects is zero and it includes more informative characteristics of the signals to emphasize the head nodding and distinguish between each symbol.

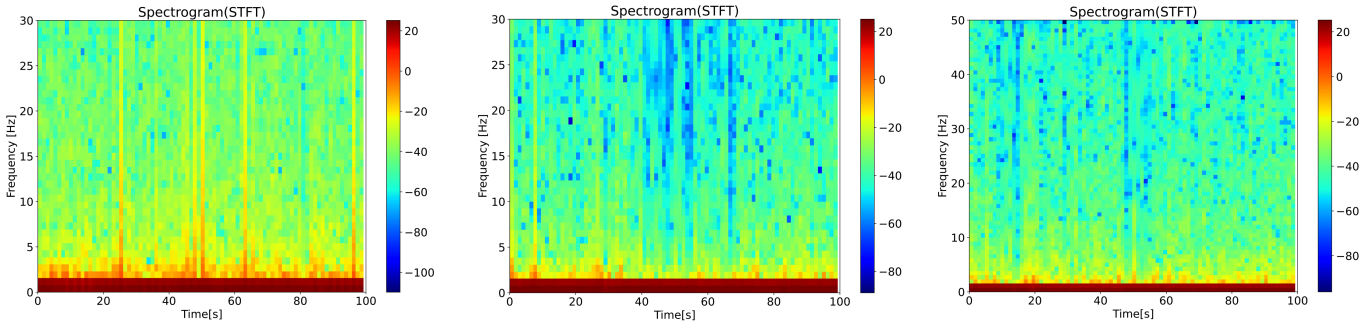
A. ESP32 CSI Toolkit

In this paper, we collected data using the ESP32 CSI toolkit [17] which is considered a promising CSI sensing solution due to its cost and power efficiency. ESP32 Wi-Fi system has a single antenna and works in 2.4-GHz frequency band with 20-MHz bandwidth and the packet sampling rate is set as 50 Hz in our system. The ESP32 node has only about 64 subcarriers, including null and data subcarriers. We remove the null subcarriers and extract only the amplitude from the 52 data subcarriers.



(a) Dot (b) Dash (c) Space

Fig. 2: Raw and Filtered Amplitudes of Three Symbols across All Subcarriers in 1st Link



(a) Dot (b) Dash (c) Space

Fig. 3: Spectrograms of 13th Subcarrier in 3rd Link for Three Symbols

B. System overview

For our proposed system, we conduct three key modules to reveal the patterns of channel variation correlated with head nodding. Fig. 1 illustrates the general architecture of the Wi-Nod system, that includes three modules, namely data collection, signal preprocessing, and the learning model. We collected the CSI streams by eight ESP32 nodes that are split into a half working as transmitters and the other half are the receivers. We parse the CSI measurements and extract the amplitude. Then, the raw CSI amplitudes are fed to the preprocessing module. Finally, the spectrogram is used as input to our learning model, which is based on the inception module. These three modules are described in detail in the following subsections.

C. Data Preprocessing

As it is known that the CSI waveform is noisy and may not significantly reduce the performance of the learning model if its raw form is used, this module aims to prepare the CSI measurements for the classification stage. This tier consists of several stages which are described as follows:

1) **Data Segmentation**: The objective of signal segmentation is to split the CSI measurements of each link based on their time stamp to be able to fuse the signals of each link with each other to build a unique pattern for each user's head motion and to be able to map it to the corresponding Morse code.
2) **Amplitude Extraction**: In our work, we extract the amplitude as the base signal by analyzing its variations due to head motion because it is reliable and has less randomness than the CSI phase. We parse the CSI files and utilize the amplitude as a base signal that feeds to the next stage.
3) **Amplitude Noise Removal**: The purpose of the noise removal stage is to smooth and remove outliers of the raw amplitude caused by environmental changes. To achieve this objective, we apply the weighted moving average filter (WMA). In general, the filtered amplitude can be calculated as:

$$A'_t = \frac{1}{m+(m-1)...+1}.[m.A_t + (m-1).A_t - 1 + (m-2).A_t - 2 + ... + A_t - m + 1]$$

A'_t is the weighted average amplitude within a window size m for time t. Fig. 2 illustrates the results of the weighted moving average for each symbol sample, and the color curves represent the amplitude of each subcarrier within the first link, as it is observed that the amplitude is smoother and outliers are eliminated.
4) **Spectrogram Extraction**: The head motions and human movements cause complex variations in the CSI amplitude since each user can perform the same motion at different speeds which can be revealed by the spectrogram with different frequencies. By applying a sliding window on the filtered amplitude to get equal-sized segments of the signal and then performing FFT on the samples in each segment, which transfers the signal from the time domain to the frequency domain, spectrograms are produced through STFT. Fig. 3 depicts the spectrogram of the subcarrier with index 13th in each link for the

(a) Dot (down) (b) Dash (right) (c) Space (left)

Fig. 4: Head Motions Used in Experiment: Dot, Dash, and Space

three different symbols. As observed in Fig. 3 there are variations in the frequency information of the head nodding for each symbol.

D. Learning Algorithm

Currently, the deep learning (DL) methods achieve higher recognition accuracy than the traditional machine learning (ML) models since the performance of any ML model depends on the quality of hand-crafted features as the low quality of the features degrades the accuracy rate. Therefore, the motivation for using the DL, especially the inception model in our work, is its ability to improve recognition accuracy and performance based on automatic feature extraction with low computational complexity. Compared to the other convolution neural networks (CNN), the inception model is usually wider rather than deeper, to speed up the learning process as the classifier input is manipulated in parallel, as we can see the architecture of the inception model in Fig. 1. The spectrogram is fed to the first inception module to extract meaningful features by using different CNN layers with different kernel sizes working in parallel. The learning process consists of two stages: The first stage is based on two inception layers for feature extraction, and the second stage is for the classification task.

- Feature extraction: The two inception block consists of three parallel convolution layers with different number and size of kernels followed by ReLU activation function and one maximum pooling (MaxPool) with the same stride value equal to one followed by the concatenation layer. The first convolution layer with 32 kernels with size (1×1), the parameters of the second convolution layer are 64 kernels with size (3×3), the third one includes 32 kernels with (5×5) size, and finally the MaxPool with (3×3) kernel size.
- Classifier: The output of the second inception layer is flattened and then fed to the fully connected layer to combine all the extracted features in the classification layer, which is represented in the Softmax layer with three classes that refer to our symbols, dot, dash, and space.

IV. Performance Evaluation

To evaluate our proposed system, we test the performance of Wi-Nod in multi-human context environment to verify the robustness of the system.

A. Experiment Setup and Data Collection

In this stage, eight ESP32 nodes are deployed to build our system. Four of them are worked as transmitters connected to mini-PCs and the others serve as receivers. During the data collection, the Lenovo laptop works as a server and monitors the received CSI waveforms. Two participants, including one female and one male, for necessary data collection to perform head nodding evaluation in a laboratory environment. Each participant performs the head motion for four minutes per symbol in two different sessions, in the morning and in the evening. The collected datasets are for three symbols, dot, dash, and space (moving the head down, right, and left, respectively), as illustrated in Fig. 4. We ask the participants to perform each gesture for four minutes. In total, we have four different datasets about 720 samples for each one. Additionally, the data collection tier was done in a multi-human environment with one person holding the frame and moving behind the participant to simulate the real-world scenario of the wheelchair for quadriplegia patients and there are several people around the performers. For robustness evaluation, we collected the data in two different time sessions, in the morning and evening, in which there are some variations in the perceived environment around the frame since in the first session, there were few people in the lab, around three persons, however, in the evening, there were numbers of students, around 10 people. The data is processed by python and Keras platform.

B. Evaluation

We evaluated the performance of our proposed system based on two subjects who collected data at different times in our laboratory with several students. For each subject, a person holding the frame behind him/her and walking, and the CSI waveforms are captured over four minutes for each symbol. This means that there is four different datasets, two for each subject collected in the morning and evening. Let F1 and F2 refer to the data set collected by the female participant in the morning and evening with a person moving behind her and M1 and M2 refer to the male data set collected in the morning and evening with the female moving behind him.

1) Accuracy: The performance of our proposed model is evaluated using the four different datasets described above in which each dataset is randomly split into 70% training and 30% for testing. We evaluate the performance based on the test accuracy which can be represented by the percentage of the number of head motions correctly detected over the number of test samples and can be calculated as the following

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (1)$$

TABLE I: Overall System Accuracy

Dataset	L1_2				L3_4				L_all			
	RAW_ AMP	WMA_ PCA_STFT	RAW_ STFT	WMA_ STFT	RAW_ AMP	WMA_ PCA_STFT	RAW_ STFT	WMA_ STFT	RAW_ AMP	WMA_ PCA_STFT	RAW_ STFT	WMA_ STFT
F1	97.7	83.4	98.16	99.5	94.93	85.71	96.3	98.6	98.62	84.79	99.1	**99.5**
M1	92.1	67.59	92.1	95.4	92.6	77.42	91.24	96.31	99.1	69.91	95.37	**99.0**
F2	98.15	80.09	98.6	99.07	92.13	78.2	92.13	**99.0**	98.7	84.26	97.7	98.1
M2	93.1	44.0	92.1	92.6	91.5	71.23	89.15	**94.8**	96.7	68.87	86.1	93.4
F1_M1	43.3	36.57	34.6	35.4	63.66	38.6	**68.2**	64.8	27.2	28.1	34.3	44.7
F1_F2	58.6	53.8	78.8	**79.1**	32.2	33.3	35.6	32.9	41.3	43.1	60.1	61.5

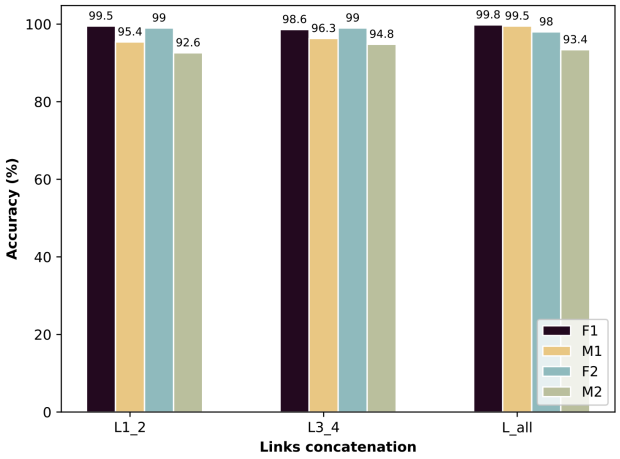


Fig. 5: The Recognition Accuracy of Different Dataset

where TP, TN, FP, FN are true positive, true negative, false positive, and false negative, respectively.

Fig. 5 shows the classification accuracy per each link concatenation which L1_2, L3_4, and L_all refer to the concatenation between link 1 and 2 (the cross configuration), link 3 and 4 (the horizontal one) and all links together, respectively. We perform the fusion between different links based on the subcarrier indices. Table I summarizes the accuracy scores of our system. As it is obviously showing, integrating all links gives better accuracy in the first session and slightly decreases in the other session but still the accuracy is high.

2) Confusion Matrix: Confusion matrix summarizes the number of instances correctly and mistakenly classified by the learning model. Fig. 6 illustrates the confusion matrix of four links concatenation for each dataset. In the first session, the model predicted 1.4% of the dot samples as dashed compared to the second session, where the misclassification rate is 11% between dashed and space in M2 dataset and about 5% in F2.

V. Discussion

A. User Diversity

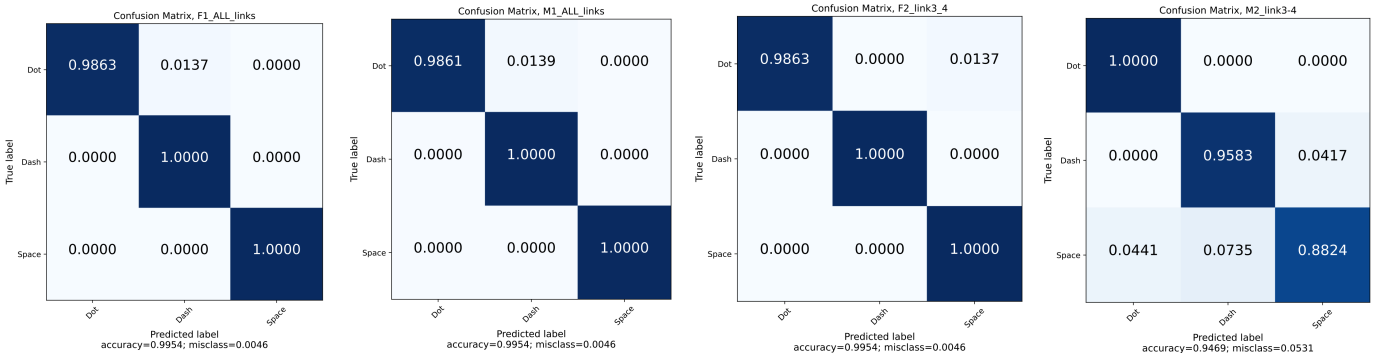
To evaluate the robustness of the user diversity, we use all the F1 dataset as a training dataset and test the model by all M1 dataset. F1_M1 raw in Table I refers to the performance of different base signals including raw amplitude (RAW_AMP), applying weighted moving average followed by Principal Component Analysis (PCA) and STFT (WMA_PCA_STFT) to reduce the dimensionality of the CSI waveforms, raw amplitude followed by the STFT (RAW_STFT), and finally weighted moving average followed by STFT (WMA_STFT), respectively, for different links fusion. As shown in the table, the combination of the 3rd and 4th links achieves the highest accuracy using the raw spectrogram as a base signal which is slightly higher than the filtered one because both the amplitudes of link 1 and 2 are affected by the movement of the subject holding the frame and moving. Also, we investigated the impact of the dimensionality reduction based on the PCA which is the lowest accuracy in Link3_4 since it eliminates the correlated variables, however within this process, most informative data can be lost. Furthermore, the results highlight the impact of using the spectrogram as a base signal in the robustness of the user diversity because the speed of different movements generates different frequencies in the frequency domain. Fig. 7a shows the accuracy based on the leave one subject out validation. The confusion matrix of the raw spectrogram in the integration of links 3 and 4 is shown in Fig. 7b and reveals that the highest rate of miss classification is between the dot and space samples, since more than half of the dot samples are predicted as space, 30% as a dash symbol, and only 12% are correctly classified.

B. Time Diversity

We investigated the time diversity robustness by training the model using the dataset collected in the morning and testing by the evening dataset of the same subject. The last row in Table I shows that the first and second link integration has the highest accuracy using the weighted moving average amplitude followed by STFT with a 79.1% rate since the link 1 and 2 covers the user head motion and the movement of the person behind him and the inception model extracts meaningful features for each symbol from these movements based on the spectrogram, which reflects different speeds to different frequencies and maps them to unique patterns. The summary of the accuracy is shown in Fig. 7c. The confusion matrix in Fig. 7d shows the model inaccurately predicted about 23% of the dash samples as space and about 20% of space samples as dot.

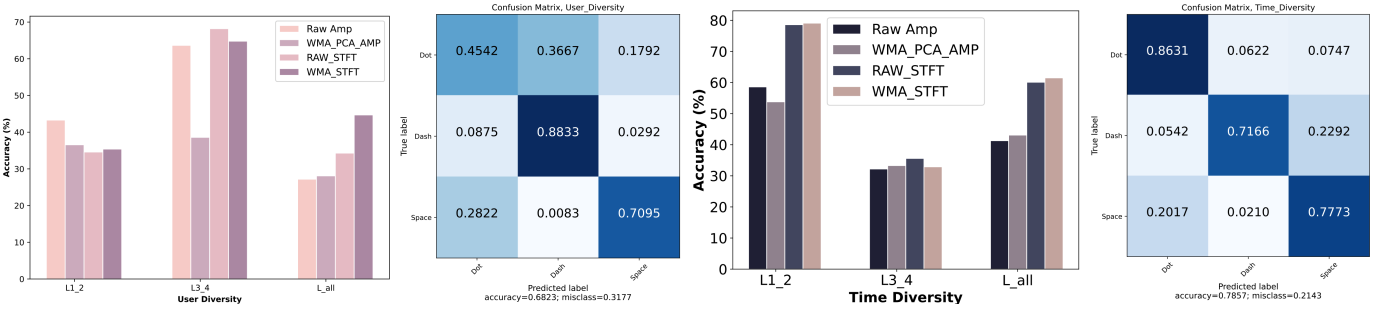
VI. Conclusion

In this paper, we proposed Wi-Nod, a Wi-Fi CSI-based head nodding recognition system that will possibly become a base of Morse code system operated by head motions, to assist the communication between quadriplegia patients and others. The details of Wi-Nod system are introduced from the data collection using the compact and inexpensive ESP32 nodes followed by the data preprocessing module for data segmentation, concatenation, amplitude extraction, outliers removal



(a) F1 confusion matrix (b) M1 confusion matrix (c) F2 confusion matrix (d) M2 confusion matrix

Fig. 6: The Confusion Matrix of Different Subjects



(a) Acc. (User Diversity) (b) CM (User Diversity) (c) Acc. (Time Diversity) (d) CM (Time Diversity)

Fig. 7: The Accuracy (Acc.) and Confusion Matrix (CM) of User/Time Diversity

filter, and frequency domain transformation based on the STFT fed to the inception model to perform the classification task. Our system evaluation with four datasets collected with two different users on two different time sessions in a multi-human context environment revealed that the system could achieve a head motion recognition accuracy of over 95%. We are planning to improve our system in terms of user and time diversity robustness and collect more data from a large number of users in different environments, additionally, we will assess the environmental independent robustness in our future work.

ACKNOWLEDGMENT

This work was partially supported by JSPS KAKENHI (JP19H05665).

REFERENCES

[1] T. T. Roberts, G. R. Leonard, and D. J. Cepela, "Classifications in brief: American spinal injury association (asia) impairment scale," 2017.

[2] M. J. Al Nahian, T. Ghosh, M. H. Al Banna, M. A. Aseeri, M. N. Uddin, M. R. Ahmed, M. Mahmud, and M. S. Kaiser, "Towards an accelerometer-based elderly fall detection system using cross-disciplinary time series features," *IEEE Access*, vol. 9, pp. 39413–39431, 2021.

[3] S. Ahmed, K. D. Kallu, S. Ahmed, and S. H. Cho, "Hand gestures recognition using radar sensors for human-computer-interaction: A review," *Remote Sensing*, vol. 13, no. 3, p. 527, 2021.

[4] H. Obeidat, W. Shuaieb, O. Obeidat, and R. Abd-Alhameed, "A review of indoor localization techniques and wireless technologies," *Wireless Personal Communications*, vol. 119, no. 1, pp. 289–327, 2021.

[5] Y. Gao, X. Li, X. V. Wang, L. Wang, and L. Gao, "A review on recent advances in vision-based defect recognition towards industrial intelligence," *Journal of Manufacturing Systems*, 2021.

[6] J. Henderson, J. Condell, J. Connolly, D. Kelly, and K. Curran, "Review of wearable sensor-based health monitoring glove devices for rheumatoid arthritis," *Sensors*, vol. 21, no. 5, p. 1576, 2021.

[7] C. Li, Z. Cao, and Y. Liu, "Deep ai enabled ubiquitous wireless sensing: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–35, 2021.

[8] M. R. Bastwesy, N. M. ElShennawy, and M. T. F. Saidahmed, "Deep learning sign language recognition system based on wi-fi csi," *International Journal of Intelligent Systems & Applications*, vol. 12, no. 6, 2020.

[9] H. Choi, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-cal: Wifi sensing and machine learning based device-free crowd counting and localization," *IEEE Access*, vol. 10, pp. 24395–24410, 2022.

[10] D. Zhang, Y. Zeng, F. Zhang, and J. Xiong, "Wifi csi-based vital signs monitoring," in *Contactless Vital Signs Monitoring*, pp. 231–255, Elsevier, 2022.

[11] X. Xie, K. G. Shin, H. Yousefi, and S. He, "Wireless csi-based head tracking in the driver seat," in *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*, pp. 112–125, 2018.

[12] Y. Liu and S. Konomi, "Wihead: Wifi-based head-pose estimation," in *International Conference on Human-Computer Interaction*, pp. 69–86, Springer, 2022.

[13] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in *Proceedings of the 21st Annual international conference on Mobile Computing and Networking*, pp. 53–64, 2015.

[14] M. Muaaz, A. Chelli, M. W. Gerdes, and M. Pätzold, "Wi-sense: A passive human activity recognition system using wi-fi and convolutional neural network and its integration in health information systems," *Annals of Telecommunications*, vol. 77, no. 3, pp. 163–175, 2022.

[15] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "Efficientfi: Towards large-scale lightweight wifi sensing via csi compression," *IEEE Internet of Things Journal*, 2022.

[16] K. Niu, F. Zhang, Y. Jiang, J. Xiong, Q. Lv, Y. Zeng, and D. Zhang, "Wimorse: A contactless morse code text input system using ambient wifi signals," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9993–10008, 2019.

[17] S. M. Hernandez and E. Bulut, "Lightweight and standalone iot based wifi sensing for active repositioning and mobility," in *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)*, pp. 277–286, IEEE, 2020.