Real-Time American Sign Language Recognition System by Using Surface EMG Signal

Celal Savur
Dept. of Electrical and Microelectronic Engineering
Rochester Institute of Technology
Rochester, NY
cs1323@rit.edu

Ferat Sahin
Dept. of Electrical and Microelectronic Engineering
Rochester Institute of Technology
Rochester, NY
feseee@rit.edu

***Abstract*—Sign Language Recognition (SLR) system is a novel method that allows hard of hearing people to communicate with general society. In this study, American Sign Language (ASL) recognition system was proposed by using the surface Electromyography (sEMG). The objective of this study is to recognize the ASL alphabet letters and allow users to spell words and sentences. For this purpose, sEMG data are acquired from subject's right forearm for twenty-six American Sign Language gestures of twenty-six English alphabets. Time domain information were used in the feature extraction process. As a classification method, Support Vector Machine was used and its results are compared with tabulated results. The experiment result of offline system is reaching a recognition rate of 91.1% accuracy and real-time system has a recognition rate of 82.3% accuracy. In conclusion, the results of this study show that sEMG signal can be used for SLR systems.**

***Index Terms*—Real-time EMG classification, Support Vector Machine, American Sign Language (ASL).**

I. Introduction

According to World Health Organization, there are approximately 360 million people having disability of hearing in 2015 [1]. The number was 278 million people in 2005 [2]. In 10 years, the number of the people having hearing problems increased about 14%. Sign Language is the only method hard of hearing people use in their daily life. Sign Language is a hand gesture language that hearing impaired people use to express their thoughts, feeling, and knowledge instead of verbal communication. There are different sign languages available in the world which have their own alphabet and hand gestures. American Sign Language (ASL) is one of the most popular sign language in the world. According to Tamar, ASL is the fourth most popular language in the USA and it is trending in college students [3]. Hearing impaired people need a communication method which allow them to communicate with society in order to accomplish their daily task. Therefore, it is really important to have an accurate SLR system.

In this study a novel SLR method system implemented by using sEMG signal. The sEMG signal is the differential voltage in the muscles. The advantage of the sEMG signal is that it is not affected from weak lightning area and easy to use when other methods have limitation such as lightning. The proposed method uses sEMG signals to predict American Sign Language alphabets. With this novel method hard of hearing people will be able to type any word and sentences to communicate with society.

The contributions of this study can be summarized as the following: (1) Introduction of existed SLR systems and their imperfections, (2) Proposed a new approach to real-time sign language recognition by using sEMG signals.

The rest of the paper is organized as follows: Section II presents the existed SLR systems. Section III describes noise types, filters, feature extraction methods, and classification methods. Section IV shows how to setup the experiments. Section V presents experimental results of both offline and real-time. Lastly, Section VI presents the conclusion and future work.

II. Background and Literature Survey

This section presents a literature review of the American Sign Language and machine learning algorithms used in this study. Sign Language Recognition (SLR) is a topic which has gained a lot of attention by researchers recently. There have been many different approaches proposed to solve this problem. Existing approaches may be divided into three main groups: Cyber Gloves, Camera and 3D Sensors, and EMG signal.

A. Cyber Gloves

In 1993, Fels et al. were one of the first researchers who worked on Sign Language Recognition system [4]. The authors came up with the idea of a Glove-Talk that acquired data from Cyber Glove, which was later used to recognize a text. Neural Network (NN) was used as the classification technique. In comparison to recent advances, and the lack of technology twenty years ago, the research had impressive results. The SLR system produced wrong words less than 1% of time in 203 gesture vocabularies. Two years later, Liang (1995) et al. used the Cyber glove to recognize ASL alphabet rather than words [5]. The researchers did not use any classification algorithm, instead, they used rule-based voting algorithm. The experiment had satisfactory results, which recognized 3 to 4 characters per second, although the researchers did not use any classifier. Meanwhile, Jiangqin et al. published the paper about Chinese Sign Language Recognition system [6]. In 2002, Mehdi et

978-1-5090-0287-0/15 $31.00 © 2015 IEEE
DOI 10.1109/ICMLA.2015.212

al. tried to developed an ASL system without training their system [7]. The authors ignored letter ’J’ and ’Z’ because Cyber Glove captures only the shape of the hand and both letters required wrist motion. Subsequently, a Neural Network based classifier was used to discriminate twenty four letters and two special character (twenty six classes). The researchers did not use any preprocessing (filtering) method, thus acquired signal from gloves were directly fed into the system. As a result, the system had lower accuracy than the trained system. Sole et al. used Extreme Learning Machine (ELC) algorithm to classify the Auslan (Australian Sign Language) alphabet [8]. The reason, the researchers chose ELC because it is a simplified version of the Neural Network (NN).

According to reported results, the system had 95% accuracy. Hayek et al. used an efficient way to show the result of the system displayed on LCD [9]. Similar to [5], the author of [9] did not use any classification algorithm. Instead, they used their own binary algorithm and did not share any performance result. The researchers did not mention how the letters that have a similar shape such as letter ’U’ and ’V’ were handled. The only difference between those letters is the angle between index finger and middle finger as shown in Figure 1 [10]. Using their binary approach will not differentiate those letters from each other. In addition, they did not mention how they handled letter ’J’ and ’Z’. The methods discussed above are using Cyber Glove for data acquisition. Although, those studies have valuable results, it is not comfortable for daily life until there are great advances in wearable technology.

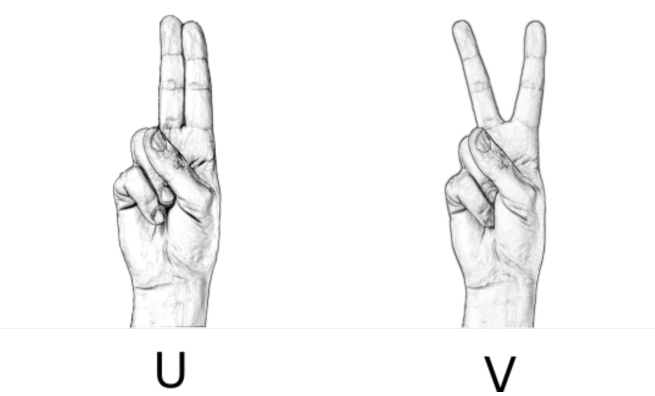

Fig. 1: The represantation of letter ’u’ and ’v’ in ASL such as taken from [10]

B. Camera and 3D Sensors

Research into Computer Vision has a long history. Therefore, there are many different applications in SLR systems. In this section, most related studies are mentioned.

In 1998, Starner et al. mounted a camera to a cap, in order to capture hand tracking [11]. The researcher used Hidden Markov Models (HMM) to recognize sentence-level ASL rather than explicitly modeling the fingers. The researchers found that having a camera mounted on a cap had better accuracy than having desk mounted camera. Otiniano-Rodriguez et al. used the Kinect sensor for ASL recognition system [12]. They mainly focused on comparing RGB, depth and RGB-Depth performances on ASL system. They found that combining RGB and depth information gave better results than single information system. Keskin et al. also used the Kinect sensors in their study [13]. The authors used a Random Decision Forest (RDF) and had satisfactory result of 99.9%. Results of RDF and ANN were compared and tabulated. Chuan et al. developed an SLR system using Leap Motion sensor [14]. The researcher used Leap Motion sensor because it is easier to transport and economical in comparison to the Kinect sensor and Cyber glove. They used the k-nearest neighborhood and Support Vector Machine (SVM) to classify 26 letters of American Sign Language. Meanwhile, Raut et al. used the dataset of 312 ASL hand gestures images [15]. The LGB Vector Quantization method was used to solve SLR system. However, in those experiments, they were using RGB information in SLR system. Therefore, in poor light conditions, these methods failed.

C. EMG Signal

The third approach is surface Electromyography (sEMG) signals. It is usually used by researchers to control robot or prosthetic arm [16] [17]. In addition, EMG signal was used to increase the usability of Human Computer Interaction (HCI) [18] [19]. Erkilinc and Sahin [20] designed a surveillance camera control system for people who cannot use joysticks. In order to extract features, Fast Fourier Transform (FFT) was applied to the raw sEMG data. The researcher also used the Principal Component Analysis (PCA) to eliminate uncorrelated features. Sahin et al. implemented a system, which control the computer cursor, using sEMG signal [21]. They used wavelet transformation with PCA to extract features and fed them into SVM classifiers. Kosmidou et al. classified most common ten English words without explicitly classifying the letters [22]. Their system was an offline system.

Unlike Cyber Glove, camera, and 3D sensor that are restricted to well lightning areas, EMG signals are not affected by light. There is one difficulty for acquiring sEMG signal, i.e. having too many wires around the arm. Thanks to the advance wearable technologies, such as Myo Armband, acquiring SEMG signals is a lot simpler and less clumsy [23].

III. Methodology

A. Time Domain Features

In machine learning, having distinguishable features are the most important keys as they improve accuracy of the system significantly. In order to have a good SLR system, following feature extraction techniques were used:

1) Mean Absolute Value (MAV):

- X represents a channel information.
- X_i is a data point in the channel.
- N is a size of the channel which is range between 1 and 1920.

The MAV is a way to represent the signal in the time domain that makes easy to see the pattern of the muscle activity, given in Equation 1. Since each letter in American Sign Language has different gesture, the MAV might increase the accuracy of SLR system.

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|Xi| \tag{1}$$

2) Modified Mean Absolute Value (MMAV): The MMAV is the extension of the MAV. In addition to the MAV, it uses weighted window. With this feature, we can set different weights for a certain range. The Equation 2 was used.

$$MMAV = \sum_{i=1}^{N} Wi|Xi|; Wi = \begin{cases} 1, & \text{if } 0.25N \le i \le 0.75N \\ 0.5, & \text{otherwise} \end{cases} \tag{2}$$

3) Simple Square Integral (SSI): The SSI is a representation of energy in the sEMG signal [24], given in Equation 3.

$$SSI = \sum_{i=1}^{N}|Xi|^2 \tag{3}$$

4) Root Mean Square (RMS): Similar to the MAV, the RMS reflects the activity of muscles. Although, the MAV and the RMS are similar, the features in the RMS are more advanced than the MAV [24]. Therefore, both methods were used in feature extraction step. This feature calculated by using Equation 4.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} Xi^2} \tag{4}$$

5) Log Detector: Log detector provides an estimate of exerted muscle force. It is computed by the following formula 5.

$$LOG = e^{\frac{1}{N}\sum_{i=1}^{N}\log(|Xi|)} \tag{5}$$

6) Average Amplitude Change (AAC): The AAC measures an average of the amplitude change in a signal. It can be calculated with Equation 6.

$$AAC = \frac{1}{N}\sum_{i=1}^{N-1}|X_{i+1} - X_i| \tag{6}$$

7) Maximum Fractal Length (MFL): The MFL is a way of measuring small changes in muscle activity. [25]. This feature calculated by using equation 7.

$$MFL = log10(\sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(X_i - X_{i+1})} \tag{7}$$

8) Minimum: Minimum of each channel contributes as a feature. This feature calculated by using Equation 8.

$$MIN = min(X) \tag{8}$$

9) Maximum: Maximum of each channel is another feature for the system. The Equation 9 was used to calculate the maximum.

$$MAX = max(X) \tag{9}$$

10) Standard Deviation: Similar to Min and Max, standard deviation is taken into account as a feature,given in Equation 10.

$$STD = std(X) \tag{10}$$

B. Classification

There are two different classification methods in machine learning, supervised learning and unsupervised learning. Supervised learning is a method of predicting a function from labelled dataset. A dataset contains training samples. Each sample has features and labelled classes. In other words, supervised learning algorithm analyzes the given dataset and produces a predicted function which is called hypothesis. In order to have a good hypothesis, the dataset should contain distinguishable features. A good algorithm is the algorithm that predicts unseen samples correctly. Unsupervised learning is a method that finds the invisible structure in unlabeled dataset such as clustering, K Nearest Neighbor.

In this research Support Vector Machine (SVM) algorithm was used.

1) Support Vector Machine: The SVM is the model that represents samples as points in the space, then divides them into separate groups with as large gap as possible as shown in Figure 2. There are kernel functions can be used to separate classes in higher dimensions. In this study, Radial Basis Function (RBF) kernels were used with SVM. The RBF is non-linear kernel which is a good choice for ASL dataset with multiple classes. The SVM is a binary classification. To achieve multi class classification, one verses all approach was taken.

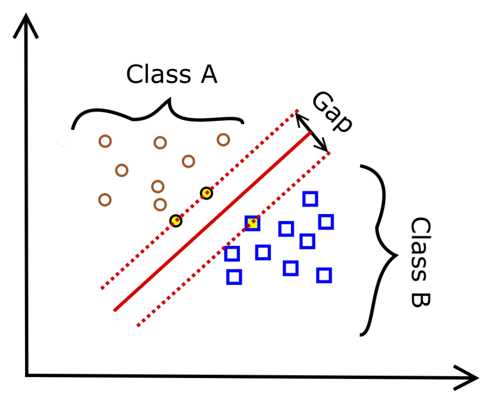

Fig. 2: The SVM classifier

IV. Experiment

A. Data Acquisition

A subject in this experiment was a 27 years old male. The subject was selected because he knows American Sign

Language alphabet. The sEMG signal was collected using eight channel Bio Radio 150 CleveMed device. All eight channels were used and attached to the right forearm of the subject via electrode. The electrode placement can be seen in the Figure 3. The placement of the electrodes were selected to mimic the Myo Armband [23]. The time window method was used and each time window size was set at two seconds length. Bio Radio 150 sampling frequency was set at 960 Hz. Thus, each time window contains 1920 data points.

Fig. 3: Right hand forearm electrode placement

A custom data collection system software implemented by using Psychtoolbox [26]. The Psychtoolbox is an interface between Matlab and computer hardware. The data collection system application read the raw data from the device .

The sEMG signal was continuously obtained from the subject and buffered in the system. The visual stimuli shows each American Sign Language letter for two-second interval and subject was asked to perform the corresponding gesture during this period. Each letter repeated twenty times in each sections. EMG data was recorded for all alphabets (Twenty Six letter of English Alphabet) and labeled and saved in the files. Figure 4 is a screen capture of the simulation application.

Fig. 4: Sample screen capture of data acquisition

The Data collection repeated 4 times. Therefore, each letter was repeated 80 times and each class sample consist of eight channels and each channel contribute ten features. Thus, the whole dataset was consist of 81 features (last feature indicates response) and 2080 samples. Afterward, the dataset was normalized using zero mean method and the dataset were randomly shuffled. Afterwards, the dataset was divided into three categories. First category was utilized as a training set which was 50% of the dataset. Second category was selected as a cross validation set which was 25% of the dataset. Remain 25% of the dataset was used for testing the system. The test set was never used in the training process.

B. Filtering

The first step after obtaining the raw data was to filter the raw sEMG data. Samples of each letter were transform into time window which is common technique for sEMG processing. Then two filters were applied to the raw dataset. The first filter was a band-pass filter consisting of a high-pass filter with 20 Hz cutoff frequency and a low-pass filter with 480 Hz cutoff frequency. The second filter was a notch filter with 60 Hz was applied for removing the power line noise. Remain data is valuable EMG information which used in feature extraction step.

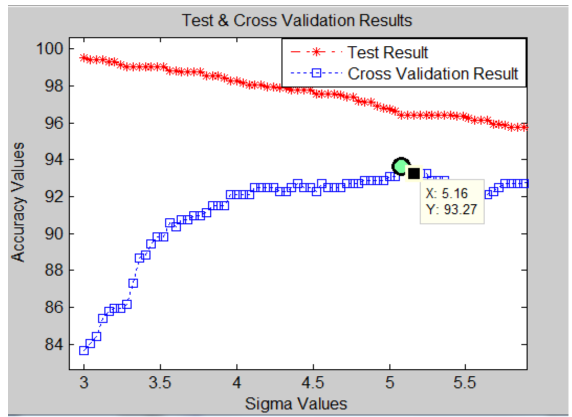

Fig. 5: Test result vs. Cross Validation result and best sigma value

C. Optimization

In this study, we chose the SVM as classifier and Radial Basis Function (RBF) kernel. The reason we chose the SVM and this kernel was that they were effective for multiple classes (26 classes). The key to get good performance with SVM is choosing right kernel function and find the optimal sigma value. Therefore, we have tested the dataset against different sigma values in range of 3.00 to 6.00 increment by 0.04 (75 steps). After those iterations the best value, which have highest accuracy and smallest distance between test result and cross validation result, was selected. Figure 5 shows the cross validation result vs. training result and the green dot indicates the best sigma value for the system.

V. Result

A. Offline Experiment

Fig. 6: Offline experiment main steps

There are four steps in off-line experiment as it can be seen in Figure 6. In first step, we collect data from the subject right forearm. Then in the second step, collected raw data was filtered. In this step, two filters were used, band-pass and notch filter. The third step was the feature extraction. In this step, the methods which mentioned in Section III, were used to extract features. In the fourth step, the SVM with RBF kernel were used to classify 26 American Sign Language Alphabets. The classification step has two sub steps, optimization and testing. In optimization, off-line system trained and tested with different sigma values. Afterwards, the off-line system trained with the best sigma value. In the second sub step, the trained offline system tested with the test set. In table I, performance of the offline experiment can be seen.

Tests	Percentage of dataset	Accuracy
Training set	50%	95.00%
Cross Validation set	25%	92.12%
Test set	25%	91.73%

TABLE I: Offline experiment results

B. Confusion Matrix

Confusion matrix of the offline system can be seen in Figure 7. Letter U and V in Figure 1 are presented. It is clear that both letters have similar hand shape. Therefore, If we look at the 'U' line, it can be seen that letter 'U' confused 15% with letter 'V'.

Fig. 7: Partial confusion matrix of the system [27]

C. Real-time Experiment

As it can be seen in Figure 8, the real-time system was similar to the offline system Fig. 6. However, there were two differences in this real-time system. First, the real-time system did not have any training process. The real-time system uses the classifiers was trained in offline experiment. Second, the real-time system was in the loop which keep acquiring sEMG data from subject in two second interval. Then the acquired data was passed through preprocessing, feature extraction and fed into the trained system.

In order to test the real time performance of the system, we asked the subject to repeat each letter ten times. This process repeated three times. The result of experiments can be seen in the Figure 9. Our method provided 82.3% accuracy in 26 classes. The real-time system's accuracy was lower than offline system accuracy. This difference was caused by action time. Time windows of offline data set has similar action time. On the other hand, in the real-time system each time windows is independent than other therefore real-time system has lower accuracy. As a future study we will focus on this issue overcome this problem.

VI. CONCLUSION AND FUTURE WORK

This study has shown that the feasibility of developing real-time sign language recognition system by using sEMG signal. In the study Bio Radio 150 was used to acquire sEMG signal from subject right forearm. The raw sEMG signals were preprocessed (filtering), feature extracted, and classified in order to predict sign gesture. Ten features were extracted in each channel and twenty six letters were used for classification. Although letter J and Z required wrist motion, both letters were included.

Although the current study has limited participants, the result of the offline system had 90% accuracy. Therefore, our system results are promising, which shows sEMG signal can be used in SLR system.

It is recommend that further research be undertaken, extracting new features in frequency domain may increase the accuracy of the SLR system. In addition to that, current data acquisition device has many wires connection and prevents subject to perform a gesture comfortably. Replacing this device with better one may increase the quality of the sEMG signal.

REFERENCES

[1] W. H. O. (WHO), "Deafness and hearing loss," 2015.

[2] L. Kin, T. Tian, R. Anuar, Z. Yahya, and A. Yahya, "Sign Language Recognition System using SEMG and Hidden Markov Model," *Conference on Recent Advances in Mathematical Methods, Intelligent Systems and Materials*, pp. 50–53, 2013.

[3] T. LEWIN, "Colleges See 16% Increase in Study of Sign Language," 2010.

[4] S. S. Fels and G. E. Hinton, "Glove-talk: a neural network interface between a data-glove and a speech synthesizer," *IEEE Transactions on Neural Networks*, vol. 4, no. 1, pp. 2–8, 1993.

[5] R.-H. Liang and M. Ouhyoung, "A Real-time Continuous Alphabetic Sign Language to Speech Conversion VR System," *Computer Graphics Forum*, vol. 14, no. 3, pp. 67–76, 1995.

[6] W. J. W. Jiangqin, G. W. G. Wen, S. Y. S. Yibo, L. W. L. Wei, and P. B. P. Bo, "A simple sign language recognition system based on data glove," *ICSP '98. 1998 Fourth International Conference on Signal Processing (Cat. No.98TH8344)*, vol. 2, pp. 1257–1260, 1998.

[7] S. A. K. Mehdi Y. N., "Sign language recognition using sensor gloves," *Neural Information Processing, 2002. ICONIP '02. Proceedings of the 9th International Conference on*, vol. 5, pp. 2204–2206 vol.5, 2002.

[8] M. M. Sole and M. S. Tsoeu, "Sign Language Recognition using the Extreme Learning Machine," *IEEE Africon 2011 - The Falls Resort and Conference Centre, Livingstone, Zambia*, no. September, pp. 13–15, 2011.

[9] H. E. Hayek, J. Nacouzi, and P. O. B. Z. Mosbeh, "Sign to Letter Translator System using a Hand Glove," *The Third International Conference on e-Technologies and Networks for Development (ICeND2014)*, pp. 146–150, 2014.

[10] G. W. Vicars, "American Sign Language Alphabet," 2015.

[11] T. Starner, J. Weaver, and a. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998.

[12] K. O. Rodriguez and G. C. Chavez, "Finger spelling recognition from RGB-D information using kernel descriptor," *Brazilian Symposium of Computer Graphic and Image Processing*, pp. 1–7, 2013.

[13] C. Keskin, F. Kraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," *IEEE International Conference on Computer Vision Workshops*, pp. 1228–1234, 2011.

[14] C.-H. Chuan, E. Regina, and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," *2014 13th International Conference on Machine Learning and Applications*, pp. 541–544, 2014.

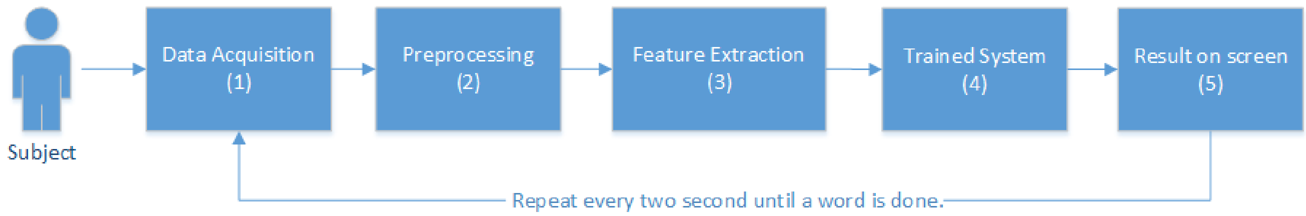

Fig. 8: Flowchart of the real-time system

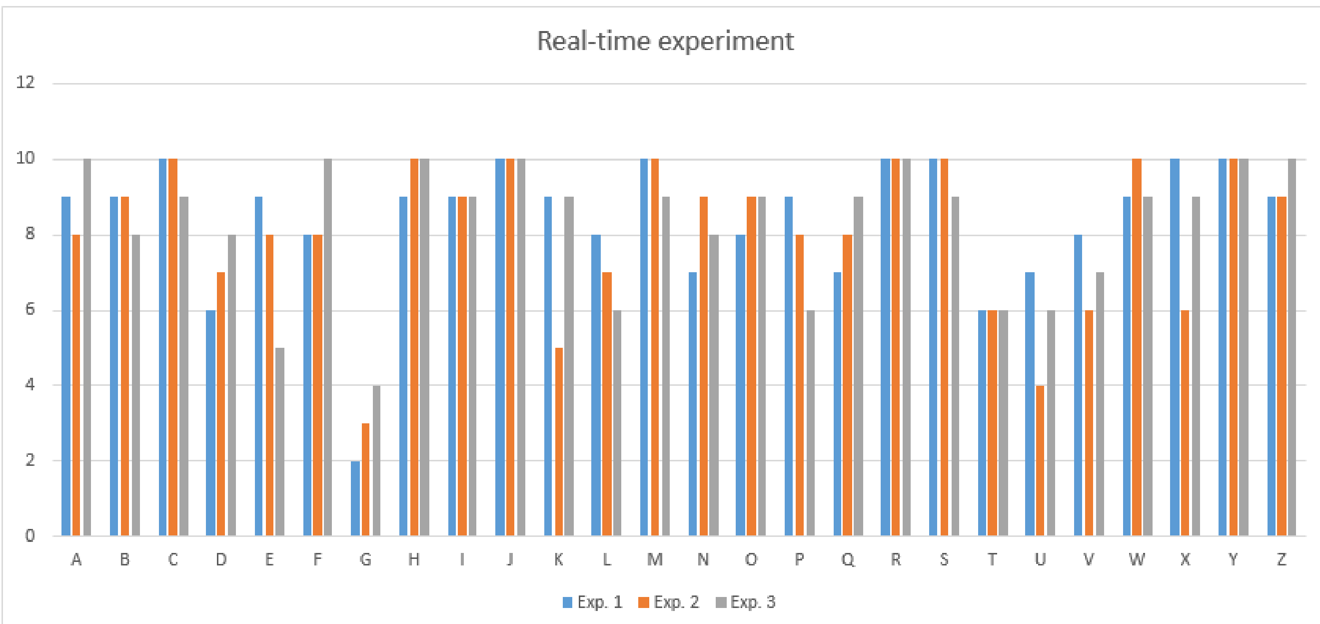

Fig. 9: Real-time experiment results

[15] K. S. Raut, "Recognition of American Sign Language Using LBG Vector Quantization," *2014 International Conference on Computer Communication and Informatics (ICCCI -2014)*, pp. 2–6, 2014.

[16] M. A. Soltanmoradi, V. Azimirad, and M. Hajibabazadeh, "Detecting finger movement through classification of electromyography signals for use in control of robots," *Proceeding of the 2nd RSI/ISM International Conference on Robotics and Mechatronics*, pp. 791–794, 2014.

[17] J.-U. C. J.-U. Chu, I. M. I. Moon, Y.-J. L. Y.-J. Lee, S.-K. K. S.-K. Kim, and M.-S. M. M.-S. Mun, "A Supervised Feature-Projection-Based Real-Time EMG Pattern Recognition for Multifunction Myoelectric Hand Control," *IEEE/ASME Transactions on Mechatronics*, vol. 12, no. 3, pp. 282–290, 2007.

[18] T. S. Saponas, D. S. Tan, D. Morris, J. Turner, and J. a. Landay, "Making muscle-computer interfaces more practical," *Proceedings of the 28th international conference on Human factors in computing systems CHI 10*, no. TR-I4-01-07, p. 851, 2010.

[19] T. S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces," *Proceeding of the twentysixth annual CHI conference on Human factors in computing systems CHI 08*, vol. 00, p. 515, 2008.

[20] M. S. Erkilinc and F. Sahin, "Camera control with EMG signals using Principal Component Analysis and support vector machines," *Systems Conference (SysCon), 2011 IEEE International*, pp. 417–421, 2011.

[21] U. Sahin and F. Sahin, "Pattern recognition with surface EMG signal based wavelet transformation," *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 295–300, 2012.

[22] V. E. Kosmidou, L. J. Hadjileontiadis, and S. M. Panas, "Evaluation of surface EMG features for the recognition of American Sign Language gestures," *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, vol. 2, no. 4, pp. 6197–6200, 2006.

[23] Thalmic, "Myo Armband," 2015.

[24] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A Novel Feature Extraction for Robust EMG Pattern Recognition," *JOURNAL OF COMPUTING*, vol. 1, no. 1, pp. 71–80, 2009.

[25] T. Portfolio, *Fractal features of Surface Electromyogram : A new measure for low level muscle activation*. PhD thesis, 2008.

[26] D. Brainard, "The psychophysics toolbox," *Spatial vision*, pp. 433–436, 1997.

[27] T. Lajnef, "Confusion Matrix Matlab Library," 2015.