Decoding Speech from Single Trial MEG Signals Using Convolutional Neural Networks and Transfer Learning

Debadatta Dash[1], Paul Ferrari[2], Daragh Heitzman[3], and Jun Wang[1,4]

Abstract— **Decoding speech directly from the brain has the potential for the development of the next generation, more efficient brain computer interfaces (BCIs) to assist in the communication of patients with locked-in syndrome (fully paralyzed but aware). In this study, we have explored the spectral and temporal features of the magnetoencephalography (MEG) signals and trained those features with convolutional neural networks (CNN) for the classification of neural signals corresponding to phrases. Experimental results demonstrated the effectiveness of CNNs in decoding speech during perception, imagination, and production tasks. Furthermore, to overcome the long training time issue of CNNs, we leveraged principal component analysis (PCA) for spatial dimension reduction of MEG data and transfer learning for model initialization. Both PCA and transfer learning were found to be highly beneficial for faster model training. The best configuration (50 principal coefficients + transfer learning) led to more than 10 times faster training than the original setting while the speech decoding accuracy remained at a similarly high level.**

I. INTRODUCTION

Patients with locked-in syndrome (fully paralyzed but aware) lose motor functions and eventually their ability to speak as they enter into a state of complete paralysis, although being cognitively intact [1]. Brain damage or neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) may cause this locked-in syndrome. Communication through the brain might be the only way to provide communication assistance to these patients. Current electroencephalography (EEG)-based BCIs use attention/visual signals to drive a letter selection on the screen, which results in a slower word synthesis rate (one to a few words per minute) [2] [3]. Ergo, there is a critical need of better and faster communication interfaces for these patients such that they can resume a meaningful life.

In recent years, research on neural speech decoding has been actively growing. Although EEG is considered as the present standard for BCIs, its low spatial resolution results in intermediate accuracy [4]–[6]. Also, the electric fields recorded with EEG are distorted at the skull and scalp. Furthermore, EEG recordings are done with respect to a reference point and hence, the measurements are more sensitive to error. Besides EEG, functional Magnetic resonance imaging (fMRI) has also been investigated for speech perception and production decoding [7], owing to its high spatial resolution [8] [9]. However, fMRI suffers from low temporal resolution characteristics [10] [11] and hence is not suitable for the development of fast BCIs. Very recently, electrocorticography (ECoG) has been shown to have the potential for direct neural decoding [12]–[15]. Despite the high spatial resolution and data quality (less noise), ECoG, however, is invasive and thus not practically preferable for some patients and it is almost impossible to obtain training data from healthy subjects.

Magnetoencephalography (MEG), a non-invasive neuroimaging technique, has an optimal spatial (higher than EEG) and temporal resolution (higher than fMRI) and hence can be the preferred option for building speech decoding based next generation-faster BCIs. Moreover, it is quiet, less prone to error (as it's reference free), and more importantly, non-invasive, hence, more practically suitable. MEG records the magnetic field and the change in the magnetic field generated by post-synaptic electric current transmission in cortical pyramidal neurons. Its high temporal resolution has been shown to be effective in capturing the fast non-stationary neural dynamics of speech from a multitude of brain regions [16]–[19]. However, with higher resolution, the complexity of MEG data (i.e., noise and data dimension) increases thereby increasing the computation time for algorithm training. Hence, reduction in machine training time is a challenge which we need to address.

CNN is a powerful machine learning algorithm for learning the latent features of images [20] [21]. In this work, we have leveraged this attribute of CNN, by training it on the spectrogram images generated from the MEG recorded neuromagnetic signals. To address the long training time issue, we applied both PCA and transfer learning in our neural speech decoding task. Prior to training, we first performed PCA to reduce the data dimension (MEG channels). Further, we explored subject knowledge transfer via transfer learning where we first performed neural speech decoding by training CNNs on a single subject data and then transferred the learned weights of the trained CNNs for next subjects.

The rest of the paper is organized as follows. Section II provides a detailed description of data collection and initial pre-processing. Methods implemented in this study such as wavelet analysis, PCA, CNN, and transfer learning have been described in Section III. Section IV represents the results and discussions, followed by conclusion in Section V.

[1]Debadatta Dash is a Ph.D. Candidate with the Department of Bioengineering, The University of Texas at Dallas, TX 75080, USA. debadatta.dash@utdallas.edu

[2]Paul Ferrari is a Research Associate Professor with the Department of Psychology, The University of Texas at Austin, and the Research Director of the MEG lab, Dell Children's Medical Center, Austin, TX 78712, USA. pferrari@utexas.edu

[3]Daragh Heitzman is a Neurologist with ALS/MDA Center, Texas Neurology, USA. dheitzman@texasneurology.com

[1,4]Jun Wang is an Assistant Professor in the Department of Bioengineering and Callier Center for Communication Disorders, The University of Texas at Dallas, TX 75080, USA. wangjun@utdallas.edu

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

Fig. 1. A Neuromag Elekta MEG unit with a subject

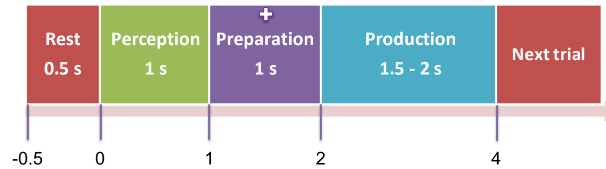


Fig. 2. The Protocol of the Time-locked Experiment

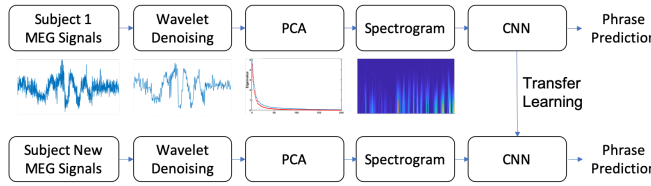


Fig. 3. Activity diagram describing the methodology

II. DATA COLLECTION

Four right-handed healthy speakers (2 females) without any speech, language or cognitive disorders participated in the data collection. Figure 1 shows the Elekta Triux MEG machine used for this study, which has 204 planar gradiometers and 102 magnetometers, housed within a magnetically shielded room (MSR). The experiments were conducted at the MEG Center, Cook Children's Hospital, Fort Worth, Texas. Proper voluntary consent was taken from the subjects prior to the experimentation. This study is approved by the institutional review board of the participating institutions.

The experiment was divided into four stages namely, pre-stimuli (rest), perception, preparation (imagination) and production (articulation) as shown in Figure 2. In the first stage, the subjects were at rest. During the perception stage, one of the five short commonly used augmentative and alternative communication (AAC) phrases (i.e., *1. Do you understand me? 2. That's perfect. 3. How are you? 4. I need Help. 5. Good-Bye.*) was displayed at a time on a back-projected screen for 1 second via a computer interfaced DLP projector in a pseudo-randomized order to avoid response suppression to repeated exposure [22]. In the imagination stage, a fixation cross was displayed on the screen which signaled the subjects to prepare for the articulation of the previously shown phrase. Then, the subjects articulated the phrase with their natural speaking rate after being signaled by the termination of fixation in the final stage of production. This procedure was repeated 100 times for each phrase to collect 100 single trials per phrase from each participant.

The recorded signals were epoched into trials from -0.5 to $+4.0$ s centered on stimulus onset. Erroneous samples containing large motion artifacts or incorrect articulations were visually inspected and discarded. Trials containing artifacts from head motion, cardiac distortion (via ECG) and eye blinks (via EOG) were removed from the signal. After preprocessing, a total of $1,635$ valid samples out of $2,000$ (4 subjects $\times$ 5 phrases $\times$ 100 trials) trials from the four subjects were remained for analysis. The MEG signals were recorded at 4 kHz sampling frequency and then band-pass filtered and resampled to 1 kHz. Four sensors out of the 204 gradiometers showed high channel noise, hence, removed from the analysis.

III. METHODS

This study performs a subject dependent 5 class classification task on each stage of the experiment where each phrase is one class (Figure 3). First, the MEG signals were denoised with wavelets and then subjected to PCA for spatial dimension reduction. Then, we generated the spectrograms of the PCA coefficients and trained a 3 layered 2D-CNN for each of the 200 gradiometer sensors. Once the CNNs were trained, the conv3 features were extracted from the 200 numbers of trained CNNs to further classify the phrases. Each of these methods is briefly explained below.

A. Wavelet Analysis

In our prior works on neural speech decoding [18] [23] wavelets were found to be highly effective for denoising the MEG signals. Here, Daubechies-4 wavelet was applied on the MEG signals of 1 kHz sampling frequency with 2 level decomposition which generated 2 detail (high frequency) signals: d_1 (250-500Hz), d_2 (125-500 Hz), and one approximation (low frequency) signal: a_2(0.1-125 Hz). The wavelet decomposition of a signal s can be simply represented as:

$$s = a_1 + d_1 = a_2 + d_2 + d_1 \tag{1}$$

It has been shown in the literature that the brain waves are limited within high gamma frequency (< 125 Hz) range [24]. Hence, we discarded the detail coefficients as noise and used the approximation (a_3) as our signal of interest.

B. PCA

PCA performs an orthogonal transformation of the data into a set of linearly uncorrelated principal components and their projected coefficients to reveal the internal variance structure of the data. It has been extensively used for artifact

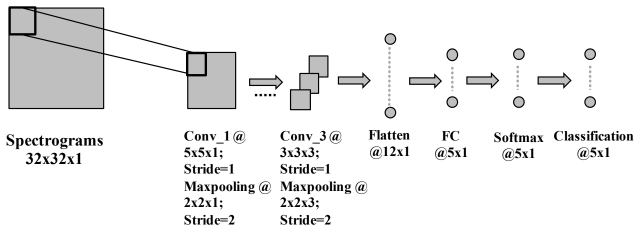


Fig. 4. Proposed 2D-CNN Architecture

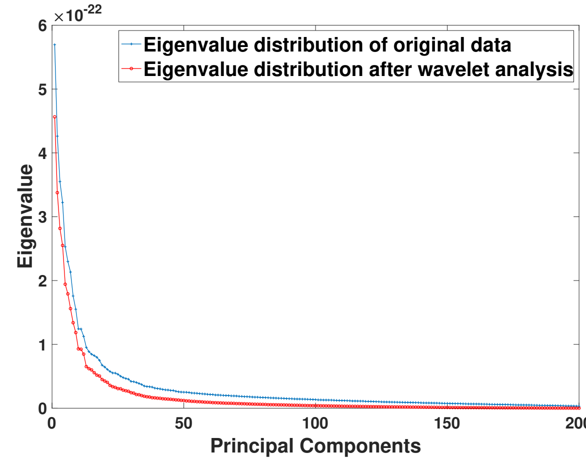


Fig. 5. Eigenvalue distribution of one trial of MEG data

reduction in MEG data [25]. Here, we used PCA for the spatial dimension reduction of MEG data to reduce the data complexity with the aim of reducing the training time of CNN. We performed the eigenvalue analysis on the wavelet denoised 200-dimensional data (per trial) as shown in Figure 5, which clearly suggests that the first 20 principal components represent the maximum variability within the data. Hence we chose the initial sets of principal coefficients (20, 30, 40, 50, 100) starting with 20 and computed their spectrograms to train the CNNs.

C. Convolutional Neural Network (CNN)

CNN, also known as ConvNet, has been heavily used for image feature analysis. It operates as a variational approach of multilayer perceptrons modeled to require minimal processing [26]. CNN is scale and shift invariant based on its shared weight and translation invariance characteristics. Its architecture is inspired from the visual cortex of the brain such that cortical neurons work on a restricted area of the visual field (called receptive field) by partial overlapping with each other to cover the whole visual space. The major advantage of CNN is it learns the relevant features from the input data, which are required to be hand-engineered by traditional algorithms.

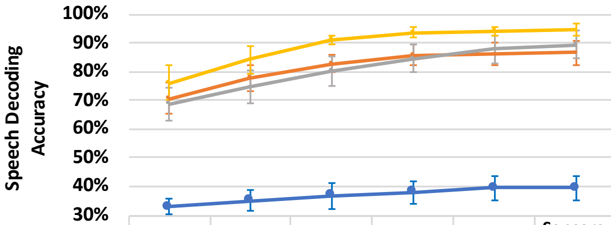
Spectral features of MEG signals are believed to carry important latent attributes of neural responses [27]. Spectrograms of the non-stationary MEG signals can represent the time-frequency variation of the sensor acquisitions. Hence, to benefit from the spectral information of the MEG signals, we computed the spectrograms of the sensors to be used as the input to the CNN. Spectrograms were generated with an analysis window of 200 ms and with a hop size of 50 ms to retain maximum frequency information (100 Hz as per Nyquist criteria) in each frame. Based on the uncertainty principle of time-frequency resolution this setup of window and hop size was optimal. However, since there were 200 spectrograms (from 200 sensor signals) obtained for a single trial of a particular stimulus, we trained the same 2D-CNN model on these grayscale spectrogram images for 200 times separately for each sensor. The objective here was to extract the CNN features from each sensor separately, which can be used for further classification of phrases.

The architecture of CNN was designed to have a conv, batch normalization, ReLU, dropout, and MaxPool layer with appropriate filter size in sequence as shown in Figure 4. For this data, 3 convolution layers were sufficient to obtain the reduced relevant feature set. For this model, the dropout probability thresholds were chosen to be 0.1, 0.2 and 0.3 for the 1^{st}, 2^{nd} and 3^{rd} layer respectively to avoid data over-fitting considering the redundant information in spectrograms. The output of these sequences is densed with a fully connected (FC) layer of 12 nodes, followed by another FC layer, softmax layer, and classification layer each with 5 nodes. Data augmentation with linear shifts was done to increase the data size as well as to compensate for the variability within the data due to the changes in the speech stimulus onset. After training, the learned FC1 features of dimension 12 were extracted from all the CNNs and then concatenated to form the input feature matrix and then trained with a shallow neural network with 200 hidden layer nodes followed by a fully connected layer, softmax, and classification layer each with 5 nodes. The data were divided into 70%, 15%, and 15% for training, testing, and validation respectively. The architecture of the network remained the same, although the input feature size changed with the change in PCA components.

However, training 200 number of CNNs is computationally very expensive, although it resulted in very good classification accuracy. Hence, in further analysis, we took the spectrograms of only the initial sets of PCA representations to train the CNN. For instance, with PCA_{50} (only the first 50 principal coefficients were selected), we computed the spectrograms of 50 PCA projected coefficients for all the trials and then used the grayscale spectrogram images (size: $32 \times 32 \times 1$) to train 50 CNNs. Dense Conv3 features were extracted from the 50 CNNs, and then used as the input for classification with the shallow ANN.

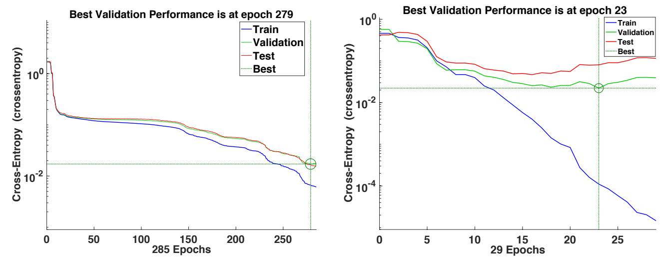
D. Transfer Learning

Transfer learning is used to reuse the pre-trained models for new data. Transfer learning has been heavily used by researchers, where the weights of pre-trained networks (e.g., GoogleNet, AlexNet, VGGNet) are transferred to a new network for faster convergence and better performance. However, considering the uniqueness of our data (spectrograms



	PCA-20	PCA-30	PCA-40	PCA-50	PCA-100	Sensors-200
Rest	32.92%	35.00%	36.67%	37.92%	39.58%	39.58%
Perception	70.83%	77.92%	82.92%	85.83%	86.25%	86.67%
Preparation	68.75%	74.58%	80.42%	84.58%	87.92%	89.58%
Articulation	76.25%	84.17%	91.25%	93.75%	94.17%	94.58%

Fig. 6. Average phrase classification accuracy during rest (pre-stimuli), perception, preparation, and articulation across the 4 subjects with the error bars representing standard deviation)



(a) CNN with 200 sensors (b) CNN+PCA$_{50}$+TL

Fig. 7. Performance curves of CNN showing training epochs and best validation performance with early stopping.

of MEG signals), we did not use these predefined nets, rather, first performed the neural speech decoding by training CNNs on a single subject data and then transferred the learned weights of the trained CNNs of that subject for speech decoding in the next subjects. The weights of the classification layer, softmax layer, and FC layers were not used for training on the data of another subject.

IV. RESULTS AND DISCUSSIONS

Figure 5 illustrates the effectiveness of wavelet for noise reduction in data processing. The variability of the signal has been greatly reduced after wavelet analysis as the red curve representing the Eigenvalue distribution of MEG data is lower to the blue curve which is for the original data. Further, it can be seen that the magnitude of the first Eigenvalue is lower after wavelet denoising representing much reduction in variability within the data. Furthermore, this helped in a stricter selection of the threshold on the number of principal components as the Eigenvalues saturated faster to a lower value with wavelet analysis compared to the Eigenvalues of the original data.

Figure 6 represents the average neural speech decoding (phrase classification) accuracy across 4 subjects during rest/pre-stimuli, perception, preparation, and production stages, obtained from CNN features by taking different numbers of PCA components at a time. It can be observed that during articulation (speech production) stage, the speech decoding performance was the highest, then preparation and then during the perception stage when all the sensor data were used for training. The rest stage showed a significantly lower accuracy and was equivalent to chance level as expected since no speech information was present in this stage. However, a trend of decrease in accuracy in all the 4 stages can be observed with the decrease in the number of PCA components. Up to 50 components (from the higher side), the accuracy for speech perception, preparation and articulation stages remained more or less the same. This conveys that, 50 number of projected coefficients might be necessary and sufficient for this small neural speech decoding task.

TABLE I

ACCURACY, EPOCH AND TRAINING TIME OF DIFFERENT METHODS DURING ARTICULATION AVERAGED ACROSS SUBJECTS (STANDARD DEVIATIONS IN THE PARENTHESIS)

Method	Accuracy (%)	Epoch	Training (hours)
CNN-200 (Baseline)	95.00 (2.89)	282 (14)	24.61 (2.52)
CNN+TL	95.00 (2.89)	53 (23)	3.67 (0.69)
PCA-100+CNN+TL	93.89 (1.92)	27 (9)	2.41 (1.06)
PCA-50+CNN+TL	93.89 (1.67)	23 (5)	2.02 (0.53)
PCA-40+CNN+TL	91.67 (1.67)	23 (4)	1.94 (0.63)
PCA-30+CNN+TL	84.44 (5.85)	21 (2)	1.43 (0.12)
PCA-20+CNN+TL	77.22 (6.74)	16 (3)	1.19 (0.15)

Table I further illustrates the effectiveness of PCA and transfer learning in training time reduction by enlisting the average accuracy, epoch and training time with their standard deviation across the three subjects. The results with transfer learning (represented as TL) are when the 1^{st} subject was first trained with CNNs and then the learned weights were transferred for the training of the other three subjects. The first two rows in Table I clearly indicates that the number of epochs and consequently the training time was significantly reduced with transfer learning, while accuracy remained the same. PCA helped in further reduction of training time by limiting the data dimension. With 50 PCA coefficients and transfer learning, the number of training epochs was reduced from 282 to 23 with more than 10 times faster training time while maintaining a similar level of accuracy. The green circle in Figure 7 represents the epoch with the best validation performance before it started to over-fit. To check for data over-fitting a validation patience of 6 epochs was used such that if the validation error increases continuously for 6 epochs training will halt. The training was done with a single GPU (Nvidia Titan X Pascal Video card 12GB graphic memory) with MATLAB R2018a on Linux (Ubuntu 16.04) platform. The system consists of Dual Intel Broadwell-EP Xeon E5 – 2667 V4 CPUs with 3.2 GHz (16 Cores / 32 Threads total) and 1 TB SDRAM. Please note, the training time shown here is exemplary and highly subjective across subjects and runs, however, the trend of decrease in epochs and training time was highly consistent across subjects.

A limitation of this study is the small data size from only four healthy subjects. Further studies with data from more subjects are needed to verify the findings. Nevertheless, the recurring performance validations across all the four subjects obtained with this approach showed great potential for direct speech decoding from the brain signals. Currently, the study is being escalated with data collection from more number of healthy subjects as well as from ALS patients. In addition, current MEG has limitations for BCI applications such as high cost, non-portability, and large size. Fortunately, continuous advancements of cost-effective MEG acquisition techniques including a recently developed movable MEG [28] strengthen the potential of MEG for BCI applications in the future.

V. CONCLUSION

This study investigated speech decoding from the brain using convolutional neural networks. Speech decoding accuracy was the best during the speech production stage over perception and imagination stages. Even during the imagination stage, the accuracy was satisfactory, which suggests the ability for decoding intended or covert speech. Further, to overcome the high training requirement of CNN, PCA based dimension reduction and transfer learning were used. Based on the experimental results, 50 number of PCA coefficients were found to be necessary and sufficient for this small speech decoding task whereas inter-subject transfer learning was proven to be efficient for significant reduction of the training epochs and time of CNNs.

VI. ACKNOWLEDGMENT

This study was supported by the University of Texas System Brain Research grant and NIH under award number R03DC013990. We thank Dr. Mark McManis, Dr. Ted Mau, Dr. Saleem Malik, Dr. Angel W. Hernandez-Mulero, and the volunteering participants.

REFERENCES

[1] S. Laureys, F. Pellas, P. V. Eeckhout, S. Ghorbel, C. Schnakers, F. Perrin, J. Berr, M. E. Faymonville, K. H. Pantke, F. Damas, M. Lamy, G. Moonen, S. Goldman, The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless?, The BChakrabarty2015Progressoundaries of Consciousness: Neurobiology and Neuropathology, vol. 150, pp. 495-611, 2005.

[2] J. S. Brumberg, A. Nieto-Castanon, P. R. Kennedy, and F. H. Guenthera, Brain-computer interfaces for speech communication, Speech Commun., vol. 52, no. 4, pp. 367-379, 2010.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, Brain-computer interfaces for communication and control, Clinical Neurophysiology, vol. 113, pp. 767-791, 2002.

[4] A. R. Sereshkeh, R. Trott, A. Bricout and T. Chau, EEG Classification of Covert Speech Using Regularized Neural Networks, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(12), pp. 2292-2300, 2017.

[5] K. Brigham and B. V. K. V. Kumar, Imagined speech classification with EEG signals for silent communication: A preliminary investigation into synthetic telepathy, in 2010 4th International Conference on Bioinformatics and Biomedical Engineering, June 2010, pp. 1–4.

[6] S. Martin; P. Brunner; I. Iturrate; J. d. R. Millan; G. Schalk et al., Word pair classification during imagined speech using direct brain recordings, Scientific Reports 6, 25803, 2016.

[7] E. Formisano, F. D. Martino, M. Bonte, R. Goebel, Who Is Saying What? Brain-Based Decoding of Human Voice and Speech. Science 322(5903), pp. 970-973, 2008.

[8] D. Dash, V. Abrol, A. Sao, and B. Biswal, Spatial sparsification and low rank projection for fast analysis of multi-subject resting state fMRI data, IEEE 15th International Symposium on Biomedical Imaging (ISBI), pp. 1280-1283, 2018.

[9] D. Dash , B. Biswal, A. K. Sao, J. Wang, Automatic Recognition of Resting State fMRI Networks with Dictionary Learning. In: Brain Informatics (BI), Lecture Notes in Computer Science, vol. 11309, Springer, pp. 249-259, 2018.

[10] D. Dash, A. K. Sao, J. Wang, B. Biswal, How many fMRI scans are necessary and sufficient for resting brain connectivity analysis?, IEEE 6th International Global Conference on Signal and Information Processing (GlobalSIP), pp. 494-498, 2018.

[11] D. Dash, V. Abrol, A. Sao, and B. Biswal, The model order limit: Deep sparse factorization for resting brain, IEEE 15th International Symposium on Biomedical Imaging (ISBI), pp. 1244-1247, 2018.

[12] C. Herff, D. Heger, A. de Pesters, D. Telaar, P. Brunner, G. Schalk, T. Schultz, Brain-to-text: decoding spoken phrases from phone representations in the brain, Frontiers in Neuroscience, 9(217), 2015.

[13] E. C. Leuthardt, J. Cunningham, and D. Barbour, Towards a speech BCI using ECoG, in Brain-Computer Interface Research: A State-of-the-Art Summary, pp. 93-110, 2013.

[14] Q. Rabbani, G. Milsap, and N.E. Crone, The Potential for a Speech Brain–Computer Interface Using Chronic Electrocorticography, Neurotherapeutics (2019) 16: 144.

[15] Chakrabarti, S., Sandberg, H.M., Brumberg, J.S. et al., Progress in speech decoding from the electrocorticogram, Biomed. Eng. Lett. (2015) 5(1): 10–21.

[16] N. Memarian, P. Ferrari, M. J. Macdonald, D. Cheyne, L.F. De Nil, and E. W. Panga, Cortical activity during speech and non-speech oromotor tasks: A magnetoencephalography (MEG) study, Neurosci. Lett., vol. 527, pp. 34-39, 2012.

[17] J. Wang, M. Kim, A. W. Hernandez-Mulero, D. Heitzman, and P. Ferrari, Towards decoding speech production from single-trial magnetoencephalography (MEG) signals, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3036-3040, 2017.

[18] D. Dash, P. Ferrari, S. Malik, J. Wang, Overt speech retrieval from neuromagnetic signals using wavelets and artificial neural networks., IEEE 6th International Global Conference on Signal and Information Processing (GlobalSIP), pp. 489-493, 2018.

[19] D. Dash, P. Ferrari, S. Malik, J. Wang, Automatic Speech Activity Recognition from MEG Signals Using Seq2Seq Learning, in International IEEE EMBS conference on Neural Engineering (NER), pp 340-343, 2019.

[20] A. S. Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: An astounding baseline for recognition, in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2014.

[21] X. Lu, X. Duan, X. Mao, Y. Li, and X. Zhang, Feature Extraction and Fusion Using Deep Convolutional Neural Networks for Face Detection, in Mathematical Problems in Engineering, 9, 2017.

[22] K. Grill-Spector, R. Henson, and A. Martin, Repetition and the brain: neural models of stimulus-specific effects, Trends in Cognitive Sciences, 10(1), pp. 14-23, 2006.

[23] D. Dash, P. Ferrari, S. Malik, A. Montillo, J. Maldjian, and J. Wang, Determining the Optimal Number of MEG Trials: A Machine Learning and Speech Decoding Perspective, in Brain Informatics (BI), Lecture Notes in Computer Science, vol 11309, Springer, pp. 163-172, 2018.

[24] L. M. Ward, Synchronous neural oscillations and cognitive processes, Trends in Cognitive Sciences, vol. 7, no. 12, pp. 553-559, 2003.

[25] N. T. Haumann, L. Parkkonen, M. Kliuchko, P. Vuust, and E. Brattico, Comparing the performance of popular MEG/EEG artifact correction methods in an evoked-response study, Intell. Neuroscience, vol. 2016, pp. 3-12, 2016.

[26] D. Ciresan, M. Ueli, M. Jonathan, G. M. Luca, S. Jurgen, Flexible, High Performance Convolutional Neural Networks for Image Classification, Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, vol 2, pp.1237–1242, 2011.

[27] H. L. Halme and L. Parkkonen, Comparing Features for Classification of MEG Responses to Motor Imagery, PLoS ONE 11(12): e0168766, 2016.

[28] E. Boto, N. Holmes, J. Leggett, G. Roberts, V. Shah, S. S. Meyer, L. Duque Muoz, K. J. Mullinger, T. M. Tierney, S. Bestmann, G. R. Barnes, R. Bowtell and M. J. Brookes, Moving magnetoencephalography towards real-world applications with a wearable system, Nature, 555(7698), pp. 657-661, 2018.