INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Towards a Speaker Independent Speech-BCI Using Speaker Adaptation

Debadatta Dash[1], *Alan Wisler*[1], *Paul Ferrari*[2,3], *Jun Wang*[1,4]

[1]Department of Bioengineering, University of Texas at Dallas, USA
[2]Department of Psychology, University of Texas at Austin, USA
[3]MEG Lab, Dell Children's Medical Center, Austin, TX, USA
[4]Callier Center for Communication Disorders, University of Texas at Dallas, USA

debadatta.dash@utdallas.edu, alan.wisler@utdallas.edu, pferrari@utexas.edu, jun.wang@austin.utexas.edu

Abstract

Neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) can cause locked-in-syndrome (fully paralyzed but aware). Brain-computer interface (BCI) may be the only option to restore their communication. Current BCIs typically use visual or attention correlates in neural activities to select letters randomly displayed on a screen, which are extremely slow (a few words per minute). Speech-BCIs, which aim to convert the brain activity patterns to speech (neural speech decoding), hold the potential to enable faster communication. Although a few recent studies have shown the potential of neural speech decoding, those are focused on speaker-dependent models. In this study, we investigated speaker-independent neural speech decoding of five continuous phrases from Magnetoencephalography (MEG) signals while 8 subjects produced speech covertly (imagination) or overtly (articulation). We have used both supervised and unsupervised speaker adaptation strategies for implementing a speaker independent model. Experimental results demonstrated that the proposed adaptation-based speaker-independent model has significantly improved decoding performance. To our knowledge, this is the first demonstration of the possibility of speaker-independent neural speech decoding.

Index Terms: silent speech interface, magnetoencephalography, brain-computer interface, artificial neural network

1. Introduction

Biosignal-based speech communication has shown increasing promise towards various clinical applications [1] such as silent speech interface (SSI), which directly converts non-audio articulatory information to speech to help individuals who have lost their ability of speech production but can still articulate silently (e.g., laryngectomees) [2]. Besides novel tongue and lip motion tracking devices, current SSI research is focused on developing algorithms that can directly map the articulatory information to speech (text/acoustics) accurately and efficiently [2–6].

Locked-in-syndrome, which can be caused by brain damage or neurodegenerative diseases (e.g., amyotrophic lateral sclerosis, ALS), refers to a state in which patients are fully paralyzed but their cognition may be normal [7]. For these patients, who lack the motor control for other forms of communication, brain-computer interfaces (BCIs) might be the only option to provide them some level of communication assistance. Although some commercial Electroencephalography (EEG) based BCIs have been introduced in the past years, these technologies are limited by extremely slow rate (a few words per minute) [8]. Moreover, current EEG-BCIs are driven by control character selection prompted with attention/visual correlates, which decode the perception of the subjects rather their speech [9]. BCIs that can achieve a higher speaking rate are needed for improving these patients' communication efficiency.

Speech-BCI or neural silent speech interface (NSSI), a novel technological paradigm that aims to convert neural signals to speech (text, acoustics, or articulatory parameters) and then drive a text- or articulatory-to-speech synthesizer [10], has been recently demonstrated to be possible from either invasive or non-invasive neural signals [11–13]. Although a few studies on speech decoding based EEG-BCIs have been investigated such as "yes" or "no" classification [14], binary phoneme [15] or syllable classification [16, 17], they suffer from intermediate performance possibly due to the low spatial resolution and low signal-to-noise ratio of EEG signals. Invasive Electrocorticography (ECoG) and non-invasive Magnetoencephalography (MEG) have recently shown a higher potential in speech decoding [11–13, 18–20].

Despite the recent progress in demonstrating the possibility of neural speech decoding, current works are speaker-dependent, where the train and test data are from the same speakers. It is extremely challenging to decode speech or any other information from non-specific person's neural signals. Exactly what neural activity patterns are getting mapped to speech is currently poorly understood and the inter-subject cognitive variance in speech processing must be tremendous, which explains the reason behind current studies on only speaker dependent speech-BCIs.

In this study, we investigated the possibility of speaker-independent neural speech decoding. To our knowledge, this is the first study for speaker-independent speech-BCI. MEG was used to collect neural signals while the subjects produced speech covertly (imagination) or overtly (articulation). MEG records the intracellular post-synaptic neuronal current induced magnetic fields and their changes in the brain [21] using highly sensitive magnetometers and gradiometers. It has an optimal temporal resolution (real-time) and has higher spatial resolution and records less distorted signals than EEG. MEG has been shown to be successful in our prior speaker-dependent neural speech decoding studies [13, 20, 22]. Here, two types of speaker adaptation strategies (supervised and unsupervised) were used in a classification task with five phrases. Supervised adaptation utilizes a small set of labeled data from the target speakers for training. Unsupervised strategies exploit data of the target domain but without labels to identify a data representation in which patterns in the out-of-speaker data are more likely to generalize to the target speaker. The performance of the two strategies was evaluated both separately and in combination. Performance of speaker-independent model without any adaptation was used as the baseline (low boundary) with speaker-dependent decoding accuracy as the ceiling (high boundary).

http://dx.doi.org/10.21437/Interspeech.2019-3109

Figure 1: *The MEG system with a subject*

2. Data Collection

2.1. The MEG System

Two identical Neuromag (Elekta ltd.) MEG machines were used for collecting the neuromagnetic signals from the subjects corresponding to different speech stimuli, one is situated at Cook Children's Hospital, Fort Worth, Texas (Figure 1) and the other at Dell Children's Medical Center, Austin, Texas. The MEG machine is 306 channeled with 204 gradiometer sensors and 104 magnetometer sensors. It is housed within a magnetically scheduled room (MSR) which discards the external magnetic field interference. A computer interfaced DLP projector, situated at about 90 cm distance from the subject within the MSR, was used to display the stimuli.

2.2. Participants

This study involved 8 subjects (3 females and 5 males; age 41 ± 14 years). No vision, speech, auditory or cognitive disorder history was reported by the subjects. Voluntary consent has been obtained from each of the subjects prior to the experiment. Subjects were first trained on some sample stimuli for compliance during data collection. The study has been approved by the institutional review board of the participating institutions.

2.3. Stimuli

A set of 5 commonly used phrases used in augmentative and alternative communication (AAC) were chosen as the stimuli for this study. They are: *1. Do you understand me, 2. That's perfect, 3. How are you, 4. Good-bye. and 5. I need help.* The stimuli were visual, or in other words, these 5 stimuli (written in English) were displayed to the subjects via the projector.

2.4. Protocol

Prior to data collection, a common coordinate system based on three fiducial points was created for the subjects. For head motion tracking, 5 head-position-coils were used, placed at the lower lip and the anterior and posterior lateral sides of the head. Two integrated electrooculography (EOG) sensors were used to record the eye blink artifacts. Two bipolar integrated electrocardiograph (ECG) sensors were used to record the cardiac artifacts. All the sensors were calibrated and checked for noise.

The protocol was designed as a time-locked experiment with four discrete stages as pre-stimuli (rest), perception, preparation (imagination) and production (articulation) as shown in Figure 2. The pre-stimuli stage was the stage of rest which was a period of 0.5 s prior to the stimulus onset. In perception stage, one out of the 5 phrases was displayed on the screen. The stimulus remained on the screen for 1 second and then was replaced by a fixation cross (+), where the subject prepared for (or imagined) the articulation of the previously shown stimulus. After 1 s, the subject articulated the phrase overtly (loudly) at his/her natural speaking rate. For each subject, 100 trials of data recording were done per phrase following the standard practice of MEG experiment [13, 23]. A pseudo-random order of the stimuli was maintained to avoid response suppression due to repeated exposure [23–25]. The whole experiment lasted for an average of 45 minutes per subject.

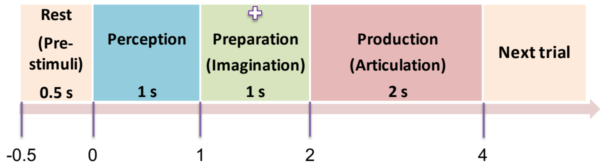


Figure 2: *Protocol of the time locked experiment.*

2.5. Data Preprocessing

Through visual inspection, high amplitude artifacts and untimely articulated trials were discarded. EOG and ECG recorded artifacts were also removed from the data. Although the articulated speech was recorded with a built-in microphone, that information was not used in data analysis for this study. A total of 3046 valid trials were retained after preprocessing from 4000 ($8\, subjects \times 5\, phrases \times 100\, trials$) collected trials with an average of 75 trials per phrase per subject. Only gradiometer sensors were considered in this study considering their effectiveness in noise suppression. However, some gradiometer sensors showed high channel noise, which were discarded. A total of 196 gradiometers out of 204 were considered. The neuromagnetic signals were recorded at 4 kHz sampling frequency which were then band-pass filtered and resampled to 1 kHz. Considering the effectiveness of wavelets in de-noising the MEG signals [20, 22, 26], a 2 level Daubechies (db)-4 discrete wavelet decomposition was performed to denoise and restrict the signal up to high-gamma frequency.

3. Methods

This study performs a speaker independent 5-phrase classification task using an artificial neural network (ANN). The ANN was trained on the root mean square (RMS) features of the neuromagnetic signals and tasked with discriminating the neuromagnetic signals corresponding to 5 different speech stimuli during imagination and articulation. The effectiveness of the RMS features with ANNs has been verified in our previous studies on speaker dependent speech decoding [20, 22] and hence was chosen as the standard to perform the speaker independent analysis. Both speaker-dependent and speaker-independent analysis were performed and compared. Additionally, we tested several supervised and unsupervised adaptation strategies to try and improve the performance of the speaker-independent system on new speakers.

3.1. Speaker Adaptation

The commonly used domain adaptation techniques were adopted here for performing speaker adaptation. Domain adptation aims at solving a learning problem in one domain (target domain) by utilizing the training data in a slightly different domain (source domain) [27]. This process is useful when the data in the target domain is insufficient for training an independent model, but can still be useful in adapting a similar model to the challenges of this specific domain.

Let X_s and Y_s be the feature data and corresponding labels (phrase 1, phrase 2..., phrase 5) in the source domain. A machine learning algorithm tries to learn a mathematical model (a hypothesis) h: $X_s \rightarrow Y_s$ such that it commits as little error as possible in labeling the target domain data: $(x_t, y_t) \,\epsilon\, (X_t \times Y_t)$. However, machine learning algorithms fail to generalize the hypothesis h learned from the source domain to target domain when the distributions of these domains are different as in the case of neural signals across subjects. Hence, for the building of a speaker independent speech-BCI, use of a domain adaptation strategy is prevalent.

We divided the 8 speaker data into source and target domains each with data of 4 speakers such that the source consisted the data acquired at Cook's Children hospital and the target consisted the data acquired at Austin. Such a data division was done to overcome the instrumental bias. Training was done on the complete source data (4 speakers), whereas, the testing was performed by taking one speaker at a time from the target domain. Although the average number of valid trials per class per speaker was 75, the minimum number of valid trials for one speech stimulus was 63 for a speaker in the target domain. Hence, for an unbiased comparison across the speakers, a total of 60 trials per class per speaker was considered for this study.

3.1.1. Supervised Adaptations: Knowledge Transfer and Instance Weighting

Supervised adaptation refers to the use of labeled target data in training. The simplest supervised adaptation strategy may be to include some samples from the target domain $(x_i, y_i) \,\epsilon (X_t, Y_t)$ for training. We refer to this process as *knowledge transfer*. The objective is to improve the generalization of the mapping h that has been learned in the source domain by adding partial knowledge from the target domain into the source domain. For this experiment, we started by adding 10% of the target-speaker data and continued to add in 10% increments up to 50% to observe the effect of increase in knowledge transfer in the speech decoding performance on the target domain. Of course, the training and test samples still remained unique.

A limitation of the basic knowledge transfer approach is that it assigns equal weight to both the source and target domain data. This may be problematic as it allows the small amount of target-domain data to be overshadowed by the large amount of source-domain data which is less relevant to the target application. A simple solution to this is the instance weighting strategy, which applies a greater weight to the training data from target domain. A number of supervised and unsupervised weighting strategies have been proposed in the literature [28–30]. The general motive of these strategies is to assign instance-dependent weights to the loss function, that selectively values instance based on their importance to the given learning problem. In this study, we utilized a simple weighting algorithm, similar to the approach outlined in [28], which weighted the target-speaker data eight times more heavily than the source-speaker data. The factor of eight was selected so that when 50% of the data has been transferred, the amount of weight assigned to source-speaker and target-speaker data equal in training samples (1200 = $1\ subject\ \times\ 5\, classes\ \times\ 30\, samples\ \times\ 8$).

3.1.2. Unsupervised Adaptations: Subspace Alignment and Geodesic Flow Kernel

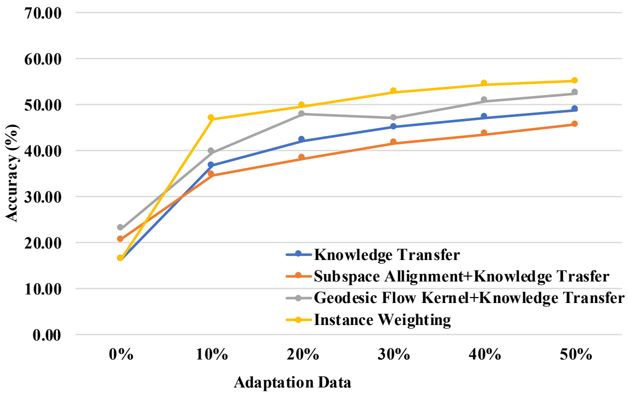
Here, unsupervised adaption refers to the use of target data but without labels [30]. A common strategy is to transform the input data into a representation that increases alignment between the two domains [31–34]. The first unsupervised adaptation method we investigated was *subspace alignment* [32]. This method learns an optimized mapping function (a transformation matrix) which aligns the source and target sub-space by exploiting the global covariance statistical structure of the two domains. The second unsupervised adaptation strategy we considered was the *geodesic flow kernel*, which models the domain shift by integrating an infinite number of sub-spaces to characterize changes in the geometric and statistical properties from the source to the target [33]. Learning from all these changes it extracts the most dominant subspace directions that are truly domain-invariant. In our experiment, the unsupervised strategies were conducted separately and also combined with the previously described knowledge transfer approach.

3.2. Classifier - Artificial Neural Network

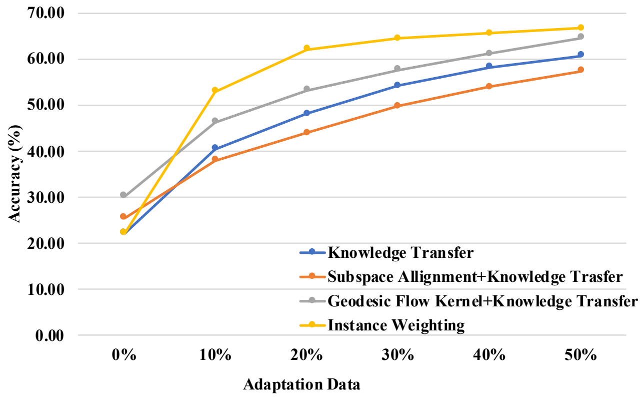
A shallow ANN classifier with 256 nodes in the hidden layer was used as the classifier. ANN is an expanded, non-linear version of linear perceptron classifier and is believed to be robust in computational modeling for pattern classification. Normalized RMS values from each gradiometer signal were used as the input (196 dimensional) to this shallow ANN. The training was performed via back-propagation using scaled conjugate gradient optimization. The model consisted of a sigmoid activation function, a fully connected softmax layer and a classification layer each with 5 nodes for this 5 class classification task. The learning rate was fixed to 0.01 and the maximum number of epochs to 1000. Based on the type of speaker adaptation strategy used the number of train and test samples were varied.

4. Results and Discussions

Experimental results are shown as a function of the amount of adaptation data in Figure 3, for imagination and articulation respectively. The data points at the 0% knowledge transfer label (the first points from the left) indicate the speaker independent performance without any target knowledge. The speaker independent decoding performance without any adaption was about the chance level performance (20%). However, a noticeable increase was observed with unsupervised domain adaptation methods even without any knowledge transfer from the target domain. A continuous increase in the speaker independent decoding accuracy was obtained during both imagination and articulation when the degree of knowledge transfer was increased. The same result was observed when knowledge transfer was combined with unsupervised domain adaptation strategy. The best performances were obtained when all adaptation data (50% of the target data) were used. Although the performance of the speaker independent decoding with adaptation remains lower than speaker-dependent decoding, the improvement suggests that some degree of speaker distinctiveness can be suppressed even by unsupervised adaption. Further, the accuracy was better for articulation compared to imagination as can be seen from these figures, possibly because the additional information of auditory and motor cortex during articulation might have contributed to this improvement. We also observed that instance weighting based supervised domain adaptation outperformed the other methods. Although supervised



(a) Speaker adaptation performance for imagination



(b) Speaker adaptation performance for articulation

Figure 3: *Performance evaluation with increase in adaptation data during (a)Imagination and (b)Articulation*

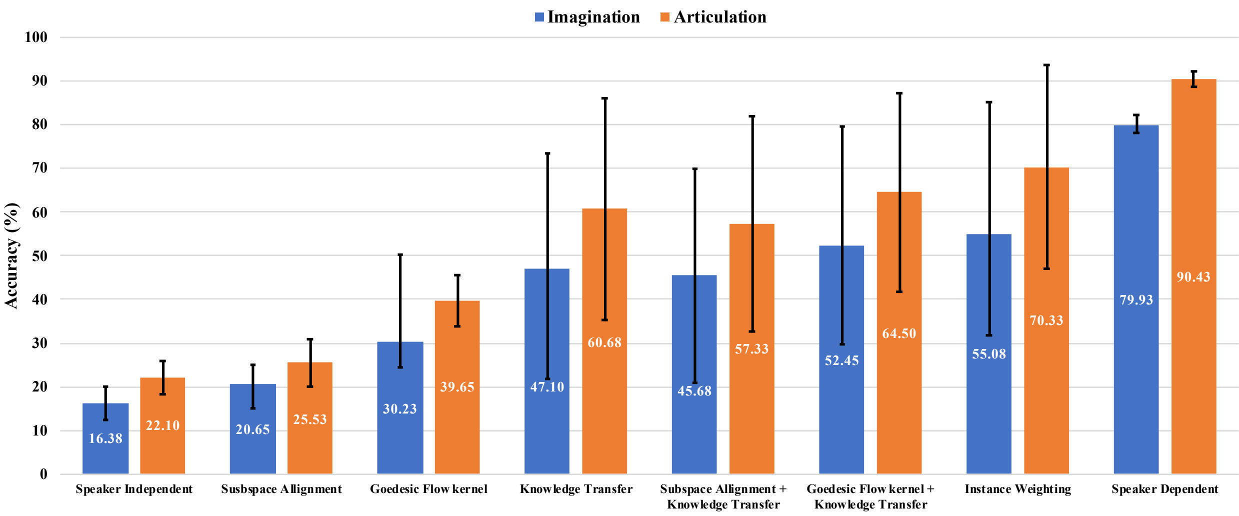


Figure 4: *Performance evaluation of implemented adaptation strategies during imagination and articulation*

adaptation may not be always feasible, the results indicated that even a small amount of adaptation data can significantly improve the decoding performance. This finding also suggests that the supervised adaption may generally outperform unsupervised adaption, which is not surprising because supervised adaption uses labeled data from the target speakers.

Figure 4 shows the best performance of these methods averaged across the 4 test speakers, individually or combined, compared with speaker dependent decoding. These results convey that (1) all adaption methods improved the decoding performance, and (2) supervised adaption outperformed unsupervised methods. The best performance was obtained with instance weighting. However, a large variation in instance weighting was observed compared to speaker-dependent decoding. Although the unsupervised adaptation strategies offered significant improvements when no labeled adaptation data was available, this improvement reduced with increase in labeled adaptation data.

5. Conclusions

In this study, we investigated the possibility of a speaker-independent speech-BCI using non-invasive nerual (MEG) signals, with several supervised and unsupervised adaption strategies. Experimental results suggested that both unsupervised and supervised adaptations can improve the performance of speaker-independent neural speech decoding, however, supervised methods significantly outperformed the unsupervised ones. Current MEG machine is limited with non-portability and high cost, but a recent study [35] shows that these limitations can be overcome making MEG suitable for speech-BCIs. Future work will focus on reducing the variability in performance of the speaker-adapted BCIs across different test speakers.

6. Acknowledgments

This study was funded by the UT System Brain Research grant and NIH (R03DC013990). We thank Dr. Ted Mau, Dr. Saleem Malik, Dr. Daragh Heitzman, Dr. Angel W. Hernandez-Mulero, and the volunteering participants.

7. References

[1] T. Schultz, M. Wand, T. Hueber, D. J. Krusienski, C. Herff, and J. S. Brumberg, "Biosignal-based spoken communication: A sur-

vey," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2257–2271, Dec 2017.

[2] B. Denby, T. Schultz, K. Honda, T. Hueber, J. Gilbert, and J. Brumberg, "Silent speech interfaces," *Speech Communication*, vol. 52, no. 4, pp. 270 – 287, 2010.

[3] T. G. Csapo, T. Grosz, G. Gosztolya, L. Toth, and A. Marko, "DNN-based ultrasound-to-speech conversion for a silent speech interface," in *Proc. Interspeech 2017*, 2017, pp. 3672–3676.

[4] M. Janke and L. Diener, "EMG-to-speech: Direct generation of speech from facial electromyographic signals," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 25, no. 12, pp. 2375–2385, Dec. 2017.

[5] M. Kim, B. Cao, T. Mau, and J. Wang, "Speaker-independent silent speech recognition from flesh-point articulatory movements using an LSTM neural network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 12, pp. 2323–2336, 2017.

[6] B. Cao, M. Kim, J. R. Wang, J. van Santen, T. Mau, and J. Wang, "Articulation-to-speech synthesis using articulatory flesh point sensors orientation information," in *Proceedings of Interspeech*, 2018, pp. 3152–3156.

[7] S. Laureys, F. Pellas, P. V. Eeckhout, and et al., "The locked-in syndrome : what is it like to be conscious but paralyzed and voiceless?" in *The Boundaries of Consciousness: Neurobiology and Neuropathology*. Elsevier, 2005, vol. 150, pp. 495 – 611.

[8] N. Birbaumer, "Brain-computer-interface research: Coming of age," *Clinical Neurophysiology*, vol. 117, no. 3, pp. 479 – 483, 2006.

[9] J. d. R. Millan, R. Rupp, G. Mueller-Putz, R. Murray-Smith, C. Giugliemma, M. Tangermann, C. Vidaurre, F. Cincotti, A. Kubler, R. Leeb, C. Neuper, K. Mueller, and D. Mattia, "Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges," *Frontiers in Neuroscience*, vol. 4, p. 161, 2010.

[10] F. Bocquelet, T. Hueber, L. Girin, C. Savariaux, and B. Yvert, "Real-time control of an articulatory-based speech synthesizer for brain computer interfaces," *PLOS Computational Biology*, vol. 12, no. 11, pp. 1–28, 11 2016.

[11] G. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, pp. 493–498, 04 2019.

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

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

[14] A. Rezazadeh 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*, vol. 25, no. 12, pp. 2292–2300, Dec 2017.

[15] S. Iqbal, P. M. Shanir, Y. U. Khan, and O. Farooq, "Time domain analysis of EEG to classify imagined speech," in *Proceedings of the Second International Conference on Computer and Communication Technologies*. Springer India, 2016, pp. 793–800.

[16] 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.

[17] C. Cooney, R. Folli, and D. Coyle, "Neurolinguistics research advancing development of a direct-speech brain-computer interface," *iScience*, vol. 8, pp. 103 – 125, 2018.

[18] M. Angrick, C. Herff, E. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, "Speech synthesis from ECoG using densely connected 3D convolutional neural networks," *Journal of Neural Engineering*, vol. 16, no. 3, p. 036019, apr 2019.

[19] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *Journal of Neural Engineering*, vol. 7, no. 5, p. 056007, sep 2010.

[20] D. Dash, P. Ferrari, S. Malik, and J. Wang, "Overt speech retrieval from neuromagnetic signals using wavelets and artificial neural networks," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2018, pp. 489–493.

[21] D. Cohen and B. Cuffin, "Demonstration of useful differences between magnetoencephalogram and electroencephalogram," *Electroencephalography and Clinical Neurophysiology*, vol. 56, no. 1, pp. 38 – 51, 1983.

[22] D. Dash, P. Ferrari, S. Malik, A. Montillo, J. A. Maldjian, and J. Wang, "Determining the optimal number of MEG trials: A machine learning and speech decoding perspective," in *Brain Informatics*. Springer International Publishing, 2018, pp. 163–172.

[23] J. Gross, S. Baillet, G. R. Barnes, R. N. Henson, A. Hillebrand, O. Jensen, K. Jerbi, V. Litvak, B. Maess, R. Oostenveld, L. Parkkonen, J. R. Taylor, V. van Wassenhove, M. Wibral, and J.-M. Schoffelen, "Good practice for conducting and reporting MEG research," *NeuroImage*, vol. 65, pp. 349 – 363, 2013.

[24] D. Cheyne and P. Ferrari, "MEG studies of motor cortex gamma oscillations: evidence for a gamma "fingerprint" in the brain?" *Frontiers in Human Neuroscience*, vol. 7, p. 575, 2013.

[25] K. Grill-Spector, R. Henson, and A. Martin, "Repetition and the brain: neural models of stimulus-specific effects," *Trends in Cognitive Sciences*, vol. 10, no. 1, pp. 14 – 23, 2006.

[26] D. Dash, P. Ferrari, S. Malik, and J. Wang, "Automatic speech activity recognition from MEG signals using seq2seq learning," in *IEEE EMBS International conference on Neural Engineering Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2018, pp. 340–343.

[27] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, Feb 2011.

[28] J. Jiang and C. Zhai, "Instance weighting for domain adaptation in nlp," in *Proceedings of the 45th annual meeting of the association of computational linguistics*, 2007, pp. 264–271.

[29] G. Foster, C. Goutte, and R. Kuhn, "Discriminative instance weighting for domain adaptation in statistical machine translation," in *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2010, pp. 451–459.

[30] A. Margolis, "A literature review of domain adaptation with unlabeled data," *Tec. Report*, pp. 1–42, 2011.

[31] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proceedings of the 2006 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2006, pp. 120–128.

[32] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 2960–2967.

[33] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 2066–2073.

[34] A. Wisler, V. Berisha, J. Liss, and A. Spanias, "Domain invariant speech features using a new divergence measure," in *IEEE Spoken Language Technology Workshop (SLT)*, Dec 2014, pp. 77–82.

[35] E. Boto, N. Holmes, and J. L. et al., "Moving magnetoencephalography towards real-world applications with a wearable system," *Nature*, vol. 555, no. 7698, pp. 657–661, Mar 2018.