Neural Speech Decoding with Magnetoencephalography

Debadatta Dash

Department of Electrical and Computer Engineering
Department of Neurology, Dell Medical School
The University of Texas at Austin

debadatta.dash@utexas.edu

Abstract

Neural speech decoding retrieves speech information directly from the brain, providing promise towards better communication assistance to patients with locked-in syndrome (e.g. due to amyotrophic lateral sclerosis, ALS) using their brain signals. Magnetoencephalography (MEG) is a non-invasive neuroimaging modality that has an excellent spatio-temporal resolution, suitable to study neural mechanism of speech. However, MEG had not been explored for speech decoding before. In my doctoral research I have thoroughly investigated neural speech decoding using MEG in the following aspects: Intended, imagined, and overt speech decoding, subject generalization in decoding, continuous articulation (jaw motion) and acoustic (VAD and speech envelope) synthesis, and MEG applications for ALS population. Here, promising results have been shown with MEG for speech decoding, providing foundation towards future wearable MEG based speech-BCI applications.

Index Terms: speech decoding, brainwaves, MEG, BCI, ALS

1. Introduction

Motivation: Neurodegenerative disorders such as amyotrophic lateral sclerosis (ALS) may lead the patients towards a state of complete paralysis otherwise being cognitively intact, i.e., locked-in syndrome. The brain might be the only source of communication for these patients. Current commercially available brain-computer interface (BCI) spellers can help these patients communicate to a level but at a very slow communication rate (less than 10 words/min). Neural speech decoding paradigm attempts to decode speech information directly from the brain providing promise towards faster communication assistance, thereby, improving the quality of life for these patients.

Literature: Electrocorticography (ECoG) has been extensively used to validate the efficacy of neural speech paradigm recently, starting from discrete recognition of speech units [1,2] to continuous speech synthesis [3–5]. Non-invasively Electroencephalography (EEG) has also been proven effective for discrete classification of a few speech units [6–8]. Magnetoencephalography (MEG) is non-invasive, has an excellent spatio-temporal resolution, and has been proven to be effective for studying neural speech processing [9, 10]. However, MEG had not been explored for speech decoding before.

2. Data Collection

We collected MEG data from 8 healthy participants and 3 ALS patients speaking 5 phrases both covertly and overtly as per a time locked protocol (Figure 1). We also collected data from 8 healthy participants speaking either ‘yes’ or ‘no’ randomly, without any cue, for natural speech decoding. Moreover, we also collected Q & A data from 12 participants answering ‘yes’/‘no’/‘I don’t know’/‘maybe’ to abstract questions (e.g. can birds smell?). We have simultaneously acquired ECG, EOG, and jaw motion data for preprocessing and developed several artifact removal pipelines including visual inspection, wavelets, and ICA.

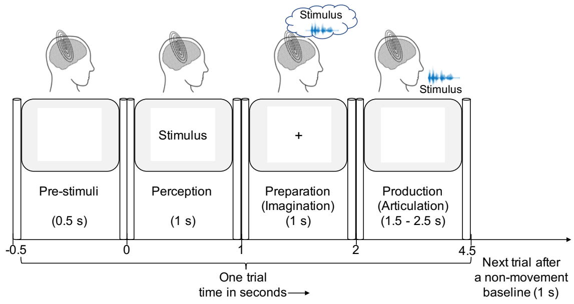

Figure 1: *The time-locked experimental protocol*

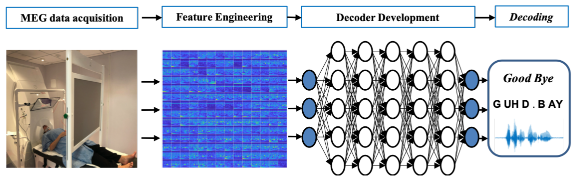

Figure 2: *Methodology for Neural speech decoding with MEG*

3. Results and Discussions

Overt and Covert Speech Decoding: Investigation on various feature engineering, and decoding pipelines (Figure 2) to properly identify speech from MEG signals resulted in the finding of root mean square and bandpower of neural oscillations to be most discriminative for decoding. We observed that adding jaw motion data to MEG signals significantly improves decoding performance indicating complementary information in kinematics and neural data [11]. Investigations on the role of neural oscillations for speech decoding showed high performance for gamma and delta band but combining all the brainwaves resulted in best performance, indicating the dynamic speech-motor coordination [12]. Spatial-spectral-temporal features were found to perform best when learned with pre-trained Deep-CNNs (e.g. Inception) which resulted in state of the art accuracy of 93% for classifying 5 imagined phrases (Figure 4(a)) [13]. Furthermore, for the first time, we were able to succesfully demonstrate that covert and overt speech can be decoded from ALS patients [14], although the performance was lower than healthy participants yet significantly higher than the chance level accuracy.

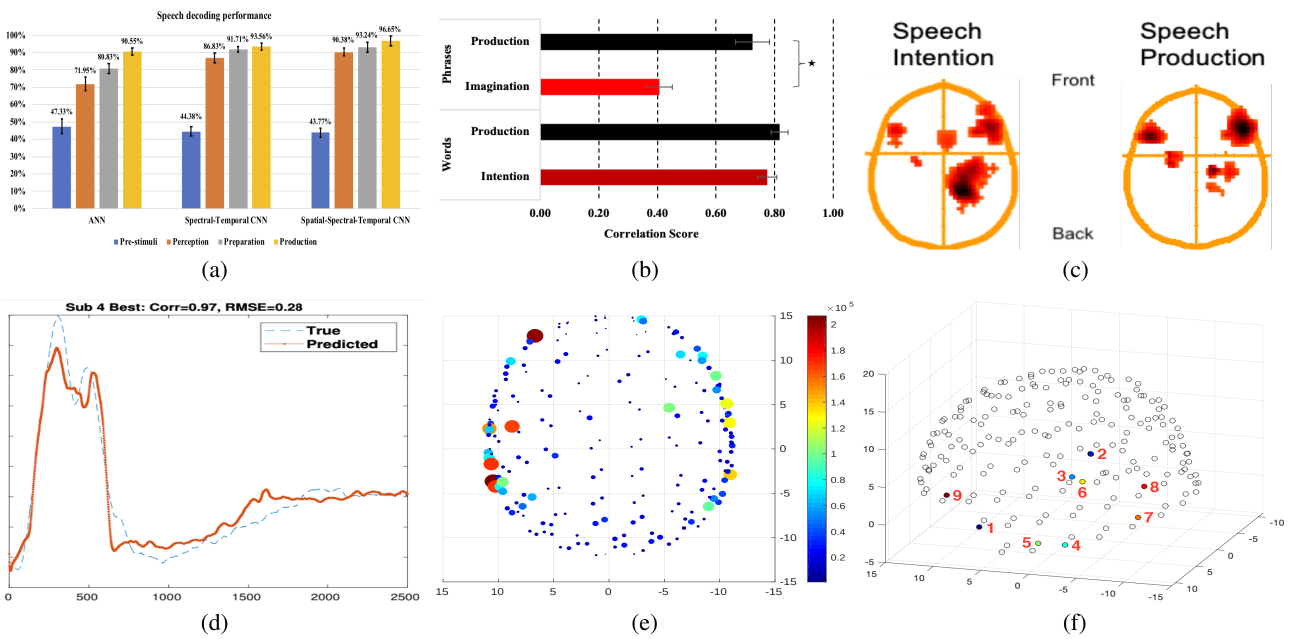

Figure 3: *(a) Decoding performance for pre-stimuli, perception, imagination, and production using ANN and CNN, (b) Correlation score between synthesized and true speech envelope for speech intention, imagination, and production (c) beta band similarity between speech intention and production, (d) an example for jaw motion regression (e) Sensors with high discriminability for VAD from MEG, (f) Saggital plot showing automatic identified optimal MEG sensor set for speech decoding*

Subject Generalization in Decoding: A major challenge in decoding research is cognitive variance across subjects making it difficult to develop a generalized decoder. Using domain adaptation, we were able to successfully develop a speaker independent decoder which was although less efficient than speaker dependent model but was significantly above than the generic speaker independent model, indicating domain adaptation might be useful in decoding research [15]. Moreover, more investigation resulted in the temporal lobe sensors to contain the subject specific characteristics for speech processing [16]. Furthermore, we were able to show that subject knowledge can be transferred to decoders for faster training of deep models [17].
Articulation and Acoustics Synthesis: Going beyond classification, we successfully reconstructed intended and articulated speech envelope with similar correlation scores about ~ 0.77 showing the possibility of continuous speech intention decoding (Figure 3(b)). This was further illustrated by the similarity in the beta band source space (Figure 3(c)). Efficacy of MEG signals in kinematics decoding was illustrated by synthesizing jaw motion with 0.8 correlation score [18]. We obtained about $\sim 90\%$ accuracy for real-time voice activity detection ((Figure 3(d)) which was resulted from the contribution of temporal sensors (Figure 3(e)) [19]. The effectiveness of temporal sensors for speech processing was further highlighted by a data driven selection of optimal sensor set (Figure 3(f)) [20] which also resulted in sensors near Broca's area to be the most effective cortical region for discriminating speech stimuli.

4. Future Work

We are increasing the speech vocabulary to 400+ phrases for open set decoding using both MEG and OPMs which will validate the use of speech evoked magnetic fields for speech-BCI applications. Further works on ALS biomarker detection from speech-MEG, Q&A speech decoding, and efficacy of magnetometers in speech decoding [21] are in progress.

5. Contributions

The major contribution is to use MEG and to show its efficacy for speech decoding which has not been attempted before. Another contribution to the speech community would be the MEG-speech data of about 30 healthy subjects and 3 ALS patients doing various protocols. In regards to the advancement of speech decoding research, major contributions are: successful use of deep CNNs with transfer learning for application on low resource (MEG) data [13], demonstrations of- subject generalization for decoding [15, 17], speech fingerprints in the brain [16], identifying the location of optimal sensors group for improvement in decoding [20], real-time VAD from MEG signals [19], intended, imagined, and spoken speech envelope synthesis from neural signals [22], simultaneously acquired articulation (jaw motion) synthesis from non-invasive signals [18], and most importantly, successful decoding of imagined, intended, and articulated speech for ALS patients. [14]. Furthermore, recent developments of wearable, movable, and low-cost MEGs, known as, optically pumped magnetometers (OPM) [23, 24], have huge potential in future wearable speech-BCI applications and these research studies are the stepping stone towards the future OPM-MEG-Speech-BCI.

6. Acknowledgements

This work was supported by the UT System Brain Initiative and NIH. I would like to thank my Ph.D. adviser Dr. Jun Wang, MEG adviser Dr. Paul Ferrari, coauthors, lab members, and the volunteering participants.

7. References

[1] 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.

[2] E. M. Mugler, J. L. Patton, R. D. Flint, Z. A. Wright, S. U. Schuele, J. Rosenow, J. J. Shih, D. J. Krusienski, and M. W. Slutzky, "Direct classification of all american english phonemes using signals from functional speech motor cortex," *Journal of neural engineering*, vol. 11, no. 3, p. 035015, 2014.

[3] 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, 2019.

[4] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, 2019.

[5] C. Herff, L. Diener, M. Angrick, E. Mugler, M. C. Tate, M. A. Goldrick, D. J. Krusienski, M. W. Slutzky, and T. Schultz, "Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices," *Frontiers in Neuroscience*, vol. 13, p. 1267, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2019.01267

[6] S. Saminu, G. Xu, Z. Shuai, A. H. Jabire, I. A. Karaye, I. S. Ahmad, A. Abdulkarim *et al.*, "Electroencephalogram (EEG) based imagined speech decoding and recognition," *Journal of Applied Materials and Technology*, vol. 2, no. 2, pp. 74–84, 2021.

[7] G. Krishna, C. Tran, Y. Han, M. Carnahan, and A. H. Tewfik, "Speech synthesis using EEG," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1235–1238.

[8] R. A. Sharon, S. S. Narayanan, M. Sur, and A. H. Murthy, "Neural speech decoding during audition, imagination and production," *IEEE Access*, vol. 8, pp. 149 714–149 729, 2020.

[9] J. Gehrig, M. Wibral, C. Arnold, and C. Kell, "Setting up the speech production network: How oscillations contribute to lateralized information routing," *Frontiers in Psychology*, vol. 3, p. 169, 2012.

[10] N. Memarian, P. Ferrari, M. J. Macdonald, D. Cheyne, F. Luc, and E. W. Pang, "Cortical activity during speech and non-speech oromotor tasks: A magnetoencephalography (MEG) study," *Neuroscience letters*, vol. 527, no. 1, pp. 34–39, 2012.

[11] 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)*. IEEE, 2018, pp. 489–493.

[12] D. Dash, P. Ferrari, and J. Wang, "Role of brainwaves in neural speech decoding," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1357–1361.

[13] D. Dash, P. Ferrari, and J. Wang, "Decoding imagined and spoken phrases from non-invasive neural (meg) signals," *Frontiers in Neuroscience*, vol. 14, p. 290, 2020. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2020.00290

[14] D. Dash, P. Ferrari, A. Hernandez, D. Heitzman, S. G. Austin, and J. Wang, "Neural speech decoding for amyotrophic lateral sclerosis," *Proc. Interspeech 2020*, pp. 2782–2786, 2020.

[15] D. Dash, A. Wisler, P. Ferrari, and J. Wang, "Towards a speaker independent speech-bci using speaker adaptation." 2019, pp. 864–868.

[16] D. Dash, P. Ferrari, and J. Wang, "Spatial and spectral fingerprint in the brain: Speaker identification from single trial MEG signals." in *INTERSPEECH*, 2019, pp. 1203–1207.

[17] D. Dash, P. Ferrari, D. Heitzman, and J. Wang, "Decoding speech from single trial meg signals using convolutional neural networks and transfer learning," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 5531–5535.

[18] D. Dash, P. Ferrari, and J. Wang, "Decoding speech evoked jaw motion from non-invasive neuromagnetic oscillations," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[19] D. Dash, P. Ferrari, S. Dutta, and J. Wang, "NeuroVAD: Real-time voice activity detection from non-invasive neuromagnetic signals," *Sensors*, vol. 20, no. 8, p. 2248, 2020.

[20] D. Dash, A. Wisler, P. Ferrari, E. M. Davenport, J. Maldjian, and J. Wang, "MEG sensor selection for neural speech decoding," *IEEE Access*, vol. 8, pp. 182 320–182 337, 2020.

[21] D. Dash, P. Ferrari, A. Babajani, A. Borna, P. D. D. Schwindt, and J. Wang, "Magnetometers vs gradiometers in neural speeh decoding (submitted)," in *IEEE EMBC*, 2021.

[22] D. Dash, P. Ferrari, K. Berstis, and J. Wang, "Imagined, intended, and spoken speech envelope synthesis from neuromagnetic signals (submitted)," in *SPECOM*, 2021.

[23] E. Boto, N. Holmes, J. Leggett, G. Roberts, V. Shah, S. S. Meyer, L. D. Muñoz, K. J. Mullinger, T. M. Tierney, S. Bestmann *et al.*, "Moving magnetoencephalography towards real-world applications with a wearable system," *Nature*, vol. 555, no. 7698, pp. 657–661, 2018.

[24] E. J. Pratt, M. Ledbetter, R. Jiménez-Martínez, B. Shapiro, A. Solon, G. Z. Iwata, S. Garber, J. Gormley, D. Decker, D. Delgadillo *et al.*, "Kernel flux: a whole-head 432-magnetometer optically-pumped magnetoencephalography (OP-MEG) system for brain activity imaging during natural human experiences," in *Optical and Quantum Sensing and Precision Metrology*, vol. 11700. International Society for Optics and Photonics, 2021, p. 1170032.