



Determining the Optimal Number of MEG Trials: A Machine Learning and Speech Decoding Perspective

Debadatta Dash¹(✉) , Paul Ferrari^{2,3}, Saleem Malik⁴ , Albert Montillo^{5,6}, Joseph A. Maldjian⁵ , and Jun Wang^{1,7} 

¹ Department of Bioengineering, University of Texas at Dallas, Richardson, USA
`debadatta.dash@utdallas.edu`

² Department of Psychology, University of Texas at Austin, Austin, USA
`pferrari@utexas.edu`

³ MEG Laboratory, Dell Children's Medical Center, Austin, USA

⁴ MEG Lab, Cook Children's Hospital, Fort Worth, TX, USA
`saleem.malik@cookchildrens.org`

⁵ Department of Radiology, UT Southwestern Medical Center, Dallas, USA
`Joseph.Maldjian@UTSouthwestern.edu`

⁶ Department of Bioinformatics, UT Southwestern Medical Center, Dallas, USA
`Albert.Montillo@UTSouthwestern.edu`

⁷ Callier Center for Communication Disorders, University of Texas at Dallas, Richardson, USA
`wangjun@utdallas.edu`

Abstract. Advancing the knowledge about neural speech mechanisms is critical for developing next-generation, faster brain computer interface to assist in speech communication for the patients with severe neurological conditions (e.g., locked-in syndrome). Among current neuroimaging techniques, Magnetoencephalography (MEG) provides direct representation for the large-scale neural dynamics of underlying cognitive processes based on its optimal spatiotemporal resolution. However, the MEG measured neural signals are smaller in magnitude compared to the background noise and hence, MEG usually suffers from a low signal-to-noise ratio (SNR) at the single-trial level. To overcome this limitation, it is common to record many trials of the same event-task and use the time-locked average signal for analysis, which can be very time consuming. In this study, we investigated the effect of the number of MEG recording trials required for speech decoding using a machine learning algorithm. We used a wavelet filter for generating the denoised neural features to train an Artificial Neural Network (ANN) for speech decoding. We found that wavelet based denoising increased the SNR of the neural signal prior to analysis and facilitated accurate speech decoding performance using as few as 40 single-trials. This study may open up the possibility of limiting MEG trials for other task evoked studies as well.

Keywords: MEG · Speech · Wavelets · Artificial Neural Network

1 Introduction

Speech is an important inherent attribute of the humans for effective communication. Speech centers of the brain along with speech articulators function synergistically to produce speech. Epochs of air, originating from the lungs, with the help of pulmonary pressure are excited using the vocal cords at specifically designed frequencies, passed through the vocal tract to the oral cavity, modulated through various articulators such as the tongue, lips, and jaw and then are radiated as ‘speech’ from the mouth. Similarly, the speech centers of the brain include Wernicke’s area: responsible for language recognition; Broca’s area: responsible for constructing the sentence structure of speech; motor cortex: directs the motion of articulators; and auditory cortex: provides auditory feedback. With regard to the motor cortex, the left cerebral hemisphere of the brain consisting of bilateral supplementary motor area, the left posterior inferior frontal gyrus, the left insula, the left primary motor cortex, and temporal cortex is the major contributor towards the motor control of speech production [1]. Also, several sub-cortical areas of the brain such as basal ganglia help in voluntary motor control for language processing [2]; and cerebellum aids in the rhythmical organized sequencing of speech syllables [3]. While it is known that a myriad of brain regions participate in speech production, the true neural basis for speech production is still poorly understood.

A better understanding of the speech mechanism is critical to help the patients with severe neurological conditions (e.g., locked-in syndrome). Locked-in syndrome (patients are fully paralyzed but aware) usually occurs due to quadriplegia, severe brain damage, or neurodegenerative disease (e.g., amyotrophic lateral sclerosis, ALS) [4] and results in the inability to speak in otherwise cognitively intact individuals [5]. Neural signal based communication might be the only way to help these patients to resume a meaningful life with some level of verbal communication. Current Brain Computer Interfaces (BCIs) send directional commands to a computer based on the signal acquired from the brain without needing any acoustic sound production [6]. EEG is the present de-facto standard for BCIs owing to its characteristics of non-invasiveness, easy setup requirement and high-quality signal acquisition with high temporal resolution [7,8]. EEG-BCIs are believed to remain as the optimal choice for communication in paralyzed and completely locked-in patients with debilitating neurological diseases [9]. However, typical EEG-BCI experiments require the subjects to select letters from a screen with a visual/attention cue. This is time consuming with an average synthesis rate of 1 Word/Minute [10] and hence not suitable for spontaneous conversations in daily life. Recently, ECoG, which measures electric potentials directly from the brain surface [11], has also been used for decoding continuous spoken speech from the cortical surface [12]. However, it is invasive in nature and hence can not be used for data collection from healthy subjects. fMRI, on the other hand, estimates the neural activity from the voxels of the brain by measuring changes in blood oxygenation [13]. Although fMRI has a very high spatial precision [14] and has been used in related neuroscientific studies [15], its slow nature hinders in continuous speech recognition.

MEG measures the weak magnetic field induced synchronized neuronal ionic currents during synaptic transmission using very sensitive magnetometers positioned around the head [16]. The higher temporal resolution and quiet nature of MEG are advantageous over fMRI. It is also non-invasive and hence more practically suitable than ECoG. In contrast to EEG, the magnetic fields recorded by MEG are less distorted than the EEG recorded electric fields at skull and scalp. Also, it is reference-free and provides higher spatial resolution. Compared to other modalities MEG might be a better choice to analyze the neural dynamics during oromotor tasks of speech production. Moreover, prior studies on MEG for decoding speech [17–19] support its advantages over other approaches.

Despite the advantages, MEG signals are sensitive to background noise and motion artifacts. The motion of the facial muscles during speech utterance and eye blinks yield large artifacts in the MEG signals. The recorded magnetic field gradients are typically temporally averaged across trials to obtain an effective signal above background activity [20]. However, recording a large number of trials for the study of a particular stimulus evoked response can be very time consuming. The number of required trials for the study of a stimulus evoked response is a trade-off between the brain physiology (e.g., cell density, cell types, and location) and the maximum number of possible trials with similar performance level, maintaining a stable head position and avoiding eye blinks, etc. [21]. It has also been shown that ensembles containing too many trials can also be problematic [22]. With time, there is always a higher chance of motion artifact induction (due to head movement or eye blinks) which subsequently reduces the signal quality. Hence, the optimal number of trials of a MEG experiment is highly variable and is still debated. Very few studies have suggested a reasonable number of MEG trials as beyond either as many as possible [20] or typically between 100–300 [22]. However, no experimental evidence is given in these studies and the conclusions are made based on assumptions of good practice.

In this study, we aimed to determine the optimal number of MEG trials that are necessary and sufficient for effective speech decoding from the brain using a machine learning algorithm, which has rarely been studied, to our knowledge. We performed a speech decoding analysis based on a set of total trials from 5 to 70, respectively, where a high decoding accuracy indicates that more information is encoded in the signals. An optimal number of trials N will be determined if the decoding accuracy is decreased when the total number of trials is less than N and also if the accuracy is not significantly increased (remains at the similar level or even decreased) when the total number of trials is greater than N .

2 Data Collection

2.1 The MEG Unit

A 306 channel Elekta Triux MEG machine as shown in Fig. 1(a) was used in the experiment. The machine is equipped with 204 number of planar gradiometers and 102 number of magnetometers. The experiments were conducted at the MEG Center, Cook Children’s Hospital, Forth Worth, Texas. The machine

was housed within a two-layered magnetically shielded room, which reduced the interference of unwanted background magnetic fields. A 3 fiducial point based coordinate system was created for the subjects. Polhemus Fastrack digitalized 5 head-position-coils were used for the head positioning of the subjects inside the MEG scanner. Comfortable seating of the subjects inside the MEG unit was ensured with their arms resting on a table. A computer interfaced DLP projector was used to display the visual cue of the stimulus on a back-projection screen. The projector was situated at a distance of 90 cm from the MEG unit.

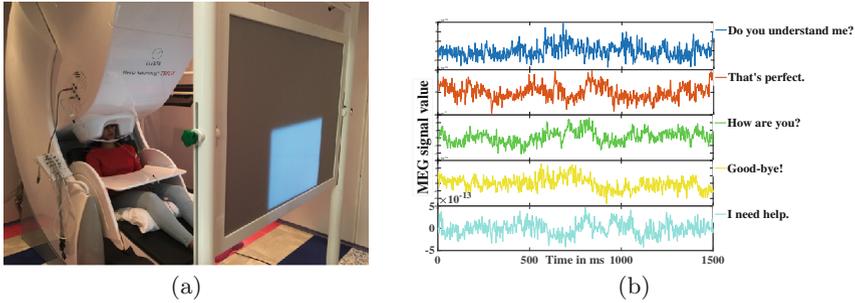


Fig. 1. (a) The MEG unit, (b) MEG signal for 5 spoken phrases with a sensor

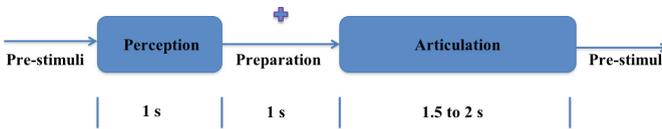


Fig. 2. Design of the experiment.

2.2 Participants and Protocol

Four young healthy, right-handed, English-speaking adults (2 males and 2 females) participated with their consent in the data collection. All the subjects had normal vision and had no speech, language, and/or cognitive disorders.

The experiment was designed as a delayed overt reading task. Subjects were asked to speak five short commonly used phrases: *Do you understand me*, *That's perfect*, *How are you*, *Good-bye*, and *I need help*. These sentences were chosen from Augmentative and Alternative Communication (AAC) data. The experiment was designed as 4 consecutive stages/segments: Pre-stimuli, Perception, Imagination (or preparation), and Articulation (Fig. 2). The pre-stimuli stage was a state of rest designed for 0.5 s. In the perception stage, a phrase stimulus

was displayed on the screen for 1 s. During preparation, the subjects were asked to prepare the phrase displayed in the previous stage with a 1-s fixation cross displayed on the screen. In the final stage, the fixation was terminated by displaying a blank screen that signaled the subjects to speak the phrase (stimulus) at their natural speaking rate. The average time period of this articulation stage was 2 (up to 2.5) s. A total of 100 trials for each of the 5 stimuli were recorded after providing for roughly 1 s of non-movement baseline. To avoid the response suppression to repeated exposure the stimuli were presented in a pseudo-randomized order [23]. A prior training of subjects on sample stimuli was conducted to ensure subject compliance. The entire experiment lasted for an average of 45 min per subject. Sample MEG signals recorded through one of the sensors for the five phrases are displayed in Fig. 1(b).

2.3 Data Preprocessing

MEG signals were acquired with a 4 kHz sampling frequency. A band-pass filter of frequency range 0.1 Hz–1.3 kHz was applied to the MEG signals. Eye-blinking and cardiac signals were recorded using integrated bipolar EOG and EEG channels. Head motion of the subjects was tracked using a scanner based continuous head localization technique [24]. The articulated speech was recorded with a standard microphone attached to a transducer placed outside the magnetically shielded room. A custom air-pressure sensor was connected to an air-filled bladder attached below the jaw of the subjects. This recorded the jaw movements of the subjects during articulation by measuring the depression in the bladder. Both speech and jaw movement analog signals were fed into the MEG ADC channels and were digitized in real-time as separate channels. The data recorded through the MEG sensors were then epoched into trials from -0.5 to $+4.0$ s centered on stimulus onset. The trials that contained high motion artifacts were inspected visually and discarded [24]. Further, signals with erroneous movements due to incorrect articulation (started either before the cue to speak or existed within the baseline period for the next trial) were also removed. The remaining trials were then down sampled to 1 kHz. After preprocessing, a total of 1635 valid samples collected from the four subjects remained for analysis.

3 Methods

For unbiased data length among subjects, a maximum of 70 trials per phase per subject was considered for this study, as the minimum number of valid trials remained after processing for one subject was 73 for one phrase. In order to find the optimal number of trials, the data was partitioned into batches consisting $10n$ trials (for $n = 0.5, 1, 2, \dots, 7$). Each batch was considered separately for analysis, denoised and then spatiotemporal wavelet features were extracted. An ANN model was trained with these features separately for each batch and the classification accuracies were compared. In our prior work [25], we had experimented with Gaussian Mixture Model (GMM) and Support Vector Machines (SVM),

but these classifiers were not able to produce good decoding accuracy. We also tested Deep Neural Networks (DNN) for this purpose, but, as the data size was limited, it resulted in data generalization even with two hidden layers.

3.1 Wavelet Analysis

Before analysis, first the MEG acquired signals were denoised using a complex Morlet wavelet. These wavelets have a sinusoidal shape and are weighted by a Gaussian kernel. Morlet wavelets effectively capture the local harmonic components in the MEG time series and hence are popularly used in MEG data processing [26]. Complex Morlet Wavelet w in the time domain t at different frequencies f is given as:

$$w(t, f) = A \exp(-t^2/2\sigma_t^2) \exp(2i\pi ft) \quad (1)$$

where, $A = (\sigma_t \sqrt{\pi i})^{-1/2}$, t = time, σ_t = wavelet time period, and $i = \sqrt{-1}$. In our analysis $f_0/\sigma_f = 5$ was used, where σ_f is the shape of the Gaussian in the frequency domain and f_0 is the central frequency. The denoising was performed in the frequency range of 0.1–120 Hz to accommodate for neuromagnetic signals up to the high gamma frequency range with a central frequency at 1 Hz intervals.

Further, we decomposed the denoised signals into respective neural frequency bands by using a Discrete Wavelet Transform (DWT) approach to select the spatiotemporal features for further analysis. The underlying principle of wavelet analysis is to express a signal \mathbf{X} as a linear combination of a particular set of functions, obtained by shifting and scaling a single function $\Psi(t)$ as:

$$\mathbf{X}(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \overline{\Psi\left(\frac{t-b}{a}\right)} \mathbf{x}(t) dt, \quad (2)$$

where, a and b are the scaling and shifting factors respectively. We have used Daubechies (db)-4 wavelet transform and performed a 5 level decomposition of the denoised MEG signal to find 5 different signals in the range of 0.1–4 Hz (delta), 4–8 Hz (theta), 8–15 Hz (alpha), 15–30 Hz (beta), 30–60 Hz (gamma), and 60–120 Hz (high gamma). Mathematically, the wavelet decomposition is represented as

$$s = a_5 + d_5 + d_4 + d_3 + d_2 + d_1 \quad (3)$$

where a_5 is the 5th level approximation of the signal representing the delta band frequency component whereas d_{1-5} are the respective layer detail components such that d_1 to d_5 represent high gamma, gamma, beta, alpha, and theta band frequency range respectively. Root Mean Square (RMS) values of the decomposed signals of all the trials were taken as the feature set to train the model.

3.2 Artificial Neural Network

A shallow ANN classifier was used in this study for its robust and effective non-linear computational modeling attributes of data classification. ANNs have been

widely used in pattern classification problems to model the particular inputs leading to specific target outputs. The ANN model was designed to take the RMS values of the wavelet decomposed sensor signals in the input layer. A single hidden layer consisting of 256 nodes were used with random weights during initialization. The maximum number of epochs was set to 100 to properly train the model. A sigmoid activation function was used after the hidden layer which transformed the learned weights into a non-linear hyper-dimensional space. A 5 dimensional fully connected softmax layer was used as the final layer to represent the minimized cross-entropy of the 5 phrases. The weights (ω_{ij}) for nodes in the hidden layer of the ANN at iteration ($t + 1$) are updated based on iteration (t) via back-propagation using a stochastic gradient descent as:

$$\omega_{ij}(t + 1) = \omega_{ij}(t) + \eta \frac{\partial \mathbf{C}}{\partial \omega_{ij}}, \quad (4)$$

where \mathbf{C} is the cost function, η is the learning rate set to 0.01, i and j are the input and hidden layer neuron labels respectively. The data was divided into three parts as training, testing and validation data. Training data consisted 70% of the whole data whereas the testing and validation data consisted 15% each of the whole data. The validation data was used to check for data over-fitting during training. Early stopping of the training resulted when the data starts to generalize with this approach. Further, we have experimented with various number of nodes to train the model to find that with a lower number of nodes the accuracy decreased whereas with higher node selection the performance of the model remained constant. Data over fitting resulted with more than 512 nodes in the hidden layer even after the 3rd epoch.

4 Results and Discussions

Figure 3 shows a comparison of the average speech decoding accuracies of the 4 subjects, obtained for each of the 4 stages of the speech production task at different trials. A peak at the 40th can be observed for the articulation, imagination and perception stages indicating the best performance with 40 number of trials. Although, as hypothesized, for the articulation and perception stage the speech decoding performance more or less saturated after 40 trials, the accuracy of imagination stage resulted in a continuous decrease. Although at present, there is no clear explanation for this behavior of the imagination stage, the absence of external stimuli or overt movement (For perception there were visual stimuli and during articulation, the produced overt speech sound was the feedback stimuli) might allow for more endogenous variability in the neural signal. After reaching optimal SNR, this variability may override any increased value of more trials.

To more accurately verify, if 40 is the optimal number, we performed the analysis with 32, 35, 37, 39, 41, 43, 45, 47 and 49 trials. The pattern remained consistent with the best accuracy obtained with either 39 or with 40 number of trials. Further, for statistical validation, we implemented a 4-way ANOVA across trials with the four subjects as the 4 factors for each stage (pre-stimuli, perception,

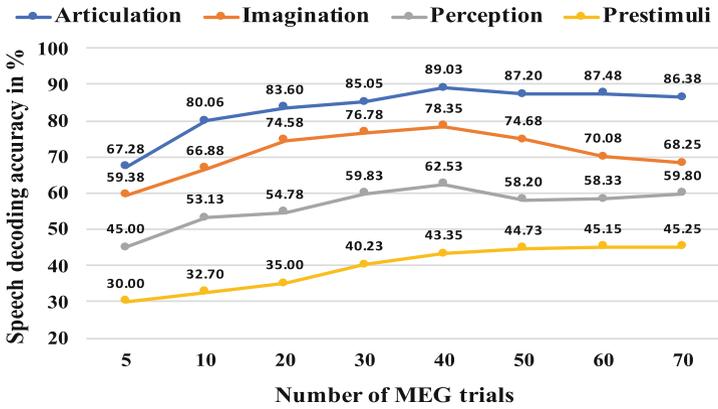


Fig. 3. Speech decoding accuracy at various total number of MEG trials

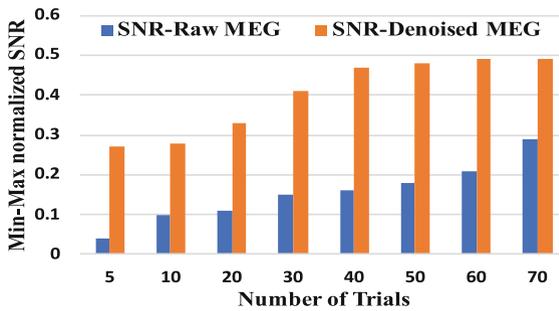


Fig. 4. Comparison of SNR for raw and denoised MEG signal at different trials

imagination, and articulation) respectively. Results indicated no statistically significant differences across subjects ($p > .05$) for each stage.

Figure 4 manifests the variation of the min-max normalized SNR of the raw MEG signal and wavelet decomposed MEG signal at different numbers of trials. It can be seen that, in case of the raw MEG signal, the SNR value was increased continuously with an increase in the number of trials. However, after denoising with wavelets, the SNR values were saturated after 40 trials. Hence, with proper denoising of the signal prior to analysis, 40 number of MEG trials could be sufficient for best speech decoding performance.

Among the 4 stages, during articulation, the highest model performance was obtained with an average accuracy of 89.03%. The average accuracy of 78.35% was maximum for classification of covert speech, i.e., during the stage of imagination. The additional involvement of the motor and auditory cortex regions of the brain to facilitate articulation and auditory feedback respectively could be the reason for the resulting higher accuracy in articulation. Perceived speech phrase classification was poorer. The pattern of increasing accuracy during the pre-stimuli stage can be clearly observed in the Fig. 3. Although, the subjects

were at rest and the accuracy should be at the chance level for this stage, with an increase in the number of trials the speech decoding accuracy was increased. The most probable reason for this is that the subjects may have started to memorize the stimulus with time, as in total 5 number of stimuli were used.

Limitation. Although the results are promising, the dataset considered in this study is relatively small. A further analysis with a larger data set from more number of subjects is needed to validate these findings.

5 Conclusion

In this study, we investigated to determine the optimal number of MEG trials from a machine learning and speech decoding perspective. A total of 40 trials were found to be sufficient and necessary for the maximum speech decoding performance. Wavelet denoising resulted in a saturated SNR after 40 trials. Limiting the MEG trials will facilitate for huge improvement in experiment and computation time as well as cost reduction. We also demonstrated that speech decoding directly from the brain is possible with an accuracy of 89.03% during articulation whereas the imagined speech information can be extracted with an accuracy of 78.35%. Although 40 trials were obtained to be optimal for speech decoding task, it may not be optimal for other task-evoked MEG experiments.

Acknowledgment. This project was supported by the University of Texas System Brain Research grant and by the National Institutes of Health (NIH) under award number R03 DC013990. We thank Dr. Mark McManis, Dr. Ted Mau, Dr. Angel W. Hernandez-Mulero, Kanishk Goel and the volunteering participants.

References

1. Indefrey, P., Levelt, W.J.M.: The spatial and temporal signatures of word production components. *Cognition* **92**(1), 101–144 (2004)
2. Booth, J.R., Wood, L., Lu, D., Houk, J.C., Bitan, T.: The role of the basal ganglia and cerebellum in language processing. *Brain Res.* **1133**, 136–144 (2007)
3. Ackermann, H.: Cerebellar contributions to speech production and speech perception psycholinguistic and neurobiological perspectives. *Trends Neurosci.* **31**(6), 256–272 (2008)
4. Laureys, S.: The locked-in syndrome: what is it like to be conscious but paralyzed and voiceless? *Progress Brain Res.* **150**, 495–611 (2005). *The Boundaries of Consciousness: Neurobiology and Neuropathology*
5. Duffy, J.: *Motor Speech Disorders Substrates, Differential Diagnosis, and Management*, 3rd edn, p. 295. Elsevier, St. Louis (2012)
6. Herff, C., Schultz, T.: Automatic speech recognition from neural signals: a focused review. *Front. Neurosci.* **10**, 429 (2016)
7. Wolpaw, J.R., Mcfarland, D.: Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *PNAS* **51**, 49–54 (2004)
8. Yoshimura, N., et al.: Decoding of covert vowel articulation using electroencephalography cortical currents. *Front. Neurosci.* **10**, 175 (2016)

9. Birbaumer, N.: Brain computer-interface research: coming of age. *Clin. Neurophysiol.* **117**(3), 479–483 (2006)
10. Brumberg, J.S., et al.: Brain computer interfaces for speech communication. *Speech Commun.* **52**(4), 367–379 (2010)
11. Leuthardt, E.C., Cunningham, J., Barbour, D.: Brain-computer interface research. In: Guger, C., Allison, B., Edlinger, G. (eds.) *Towards a Speech BCI Using ECoG*. SpringerBriefs in Electrical and Computer Engineering, pp. 93–110. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36083-1_10
12. Herff, C., et al.: Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci.* **9**(217), 1–11 (2005)
13. Dash, D., Abrol, B., Sao, A., Biswal, B.: The model order limit: deep sparse factorization for resting brain. In: *IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, pp. 1244–1247 (2018)
14. Dash, D., Abrol, B., Sao, A., Biswal, B.: Spatial sparsification and low rank projection for fast analysis of multi-subject resting state fMRI data. In: *IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, pp. 1280–1283 (2018)
15. Formisano, E., De Martino, F., Bonte, M., Goebel, R.: Who is saying what? Brain-based decoding of human voice and speech. *Science* **322**, 970–973 (2008)
16. Cohen, D., Cuffin, B.N.: Demonstration of useful differences between magnetoencephalogram and electroencephalogram. *Electroencephalogr. Clin. Neurophysiol.* **56**(1), 38–51 (1983)
17. Chan, A.M., et al.: Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage* **54**(4), 3028–3039 (2011)
18. Wang, J., Kim, M., Hernandez-Mulero, A.H., Heitzman, D., Ferrari, P.: Towards decoding speech production from single-trial Magnetoencephalography (MEG) signals. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3036–3040 (2017)
19. Dash, D., Ferrari, P., Malik, S., Wang, J.: Overt speech retrieval from neuromagnetic signals using wavelets and artificial neural networks. In: *IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (2018)
20. Gross, J., Baillet, S., Barnes, G.R., Henson, R.N., Hillebrand, A., Jensen, O., Schoffelen, J.-M.: Good practice for conducting and reporting MEG research. *Neuroimage* **65**(100), 349–363 (2013)
21. Attal, Y., et al.: Modelling and detecting deep brain activity with MEG and EEG. *IRBM - Biomed. Eng. Res.* **30**, 133 (2009)
22. Burgess, R.C., Funke, M.E., Bowyer, S.M., Lewine, J.F., Kirsch, H.E., Bagi, A.I.: American clinical magnetoencephalography society clinical practice guideline 2: presurgical functional brain mapping using magnetic evoked fields. *J. Clin. Neurophysiol.* **28**, 355–361 (2011)
23. Grill-Spector, K., Henson, R., Martin, A.: Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.* **10**(1), 14–23 (2006)
24. Cheyne, D., Ferrari, P.: MEG studies of motor cortex gamma oscillations: evidence for a gamma fingerprint in the brain? *Front. Hum. Neurosci.* **7**, 575 (2013)
25. Dash, D., Kim, M., Ferrari, P., Wang, J.: Brain activation pattern analysis for speech production decoding from MEG signals. In: *25th Annual meeting of Biomedical Engineering Society (BMES)* (2018)
26. Tadel, F., Baillet, S., Mosher, J.C., Pantazis, D., Leahy, R.M.: Brainstorm: a user-friendly Application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **8**, 8:1–8:13 (2011)