2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Classification of Phonological Categories in Imagined Speech using Phase Synchronization Measure

Jerrin Thomas Panachakel[1] and Ramakrishnan A G[4]

Abstract— **Phonological categories in articulated speech are defined based on the place and manner of articulation. In this work, we investigate whether the phonological categories of the prompts imagined during speech imagery lead to differences in phase synchronization in various cortical regions that can be discriminated from the EEG captured during the imagination. Nasal and bilabial consonant are the two phonological categories considered due to their differences in both place and manner of articulation. Mean phase coherence (MPC) is used for measuring the phase synchronization and shallow neural network (NN) is used as the classifier. As a benchmark, we have also designed another NN based on statistical parameters extracted from imagined speech EEG. The NN trained on MPC values in the beta band gives classification results superior to NN trained on alpha band MPC values, gamma band MPC values and statistical parameters extracted from the EEG.**
Clinical relevance: Brain-computer interface (BCI) is a promising tool for aiding differently-abled people and for neurorehabilitation. One of the challenges in designing speech imagery based BCI is the identification of speech prompts that can lead to distinct neural activations. We have shown that nasal and blilabial consonants lead to dissimilar activations. Hence prompts orthogonal in these phonological categories are good choices as speech imagery prompts.

I. Introduction

In imagined, covert, or inner speech, the subjects imagine speaking without any intentional movement of their articulators [1]. During speech imagery, articulatory planning occurs in premotor cortex. Since motor movements are not intended during speech imagery, the information flow is terminated at M1 [2], [3]. Nevertheless, a motor efference copy is sent to inferior parietal cortex for somatosensory estimation [4]. The perceptual efference copy generated at the inferior parietal cortex is sent to posterior superior temporal gyrus (pSTG) and superior temporal sulcus (STS), leading to an activation in the auditory cortex [5].

This study investigates whether neural correlates exist in EEG corresponding to phonological categories even while imagining speech. Specifically, we look at whether we can classify the imagined prompt as a bilabial or a nasal from the EEG captured during the imagination. This is different from the task addressed in [6]–[8]. Nasal and bilabial consonants are chosen since the articulation of prompts from these categories require different motor actions and hence these can lead to difference in the functional connectivity in the premotor cortex, where motor planning takes place. Also, due to the perceptual efference copy sent to pSTG and STS, which are part of the auditory cortex, a significant difference may show up in the auditory cortex also. Mean phase coherence (MPC) [9] is used for extracting the features for classifying nasal from bilabial consonants. Nasal consonants are produced by an occlusion produced with a lowered velum, allowing air to escape freely through the nose whereas a bilabial consonant is articulated with both the lips. These two phonological categories are chosen due to their difference in the place and manner of articulation. This difference may lead to a difference in neural activation in different human cortices when prompts from these categories are imagined.

[1]Jerrin Thomas is with Department of Electrical Engineering, Indian Institute of Science, Bangalore, India jerrinp@iisc.ac.in
[2]Ramakrishnan A G is with Department of Electrical Engineering, Indian Institute of Science, Bangalore, India agr@iisc.ac.in

II. Cortical regions considered

Speech production involves a large neocortical area along with subcortical regions. The information is processed at somatosensory, motor, auditory and prefrontal cortices [10] and hence they are considered in this study. In addition, preomotor cortex is considered based its role in motor planning and sensorimotor cortex based on its role in motor representation [11].

III. EEG Bands Considered in the Study

The three EEG rhythms considered in this study are:

- **Alpha & Mu Bands:** Alpha band (8 to 13 Hz) is usually associated with inhibitory control and attention including auditory attention [12]. An increase in alpha is observed in the visual cortex when the sensory input to the visual cortex is disrupted [13] and also when the brain processes potentially distracting information [14], [15]. A reduction in alpha is related to an increased ability to discriminate visual targets [16]. An alpha suppression is also associated with speech preparation [17]. Mu band lies in the same frequency band but is differentiated from alpha based on topography and responsivity. Mu band is observed in the sensorimotor cortex and its suppression is associated with motor movements or their observation [18]. Mu band suppression was also reported by Tamura et al. [19] during speech imagery.
- **Beta band:** Beta band (13 Hz to 30 Hz) is associated with motor tasks including motor imagery. An event-related synchronization (ERS) during motor planning and an event-related desynchronization (ERD) after task execution are reported in the beta band. This is true for actual [20] and imagined motor tasks [11]. This ERD/S is observed in sensorimotor cortex [20]. Beta activity is observed in the prefrontal cortex during motor inhibition [21]. Since the articulators move to produce speech,

changes in the sensorimotor and prefrontlal cortices may be observed during speech imagery also.

- **Gamma band:** Gamma activity arises from negative feedback between excitatory and inhibitory neurons. Beta-gamma frequency components have an evolutionary basis which lies in the six-layered neocortex to maintain its functional stability under constant exposure of sensory stimuli [22]. Gamma band in cortical and subsortical regions is found to be modulated in cognitive processing of different memory types and high level of vigilance [23]. Frontal and temporal areas communicate during speech production and increased coherence was found in high gamma range [24]. In this work, lower gamma band from 30 Hz to 45 Hz is considered since the publicly available dataset is bandpass filtered from 1 Hz to 50 Hz.

IV. Dataset Used for the Study

This study uses the publicly available dataset Kara One [1] [6]. It consists of EEG acquired from 14 subjects when they imagined articulating 7 phonemic/syllabic prompts (/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/) and 4 words derived from Kent's list of phonetically-similar pairs (i.e., pat, pot, knew, and gnaw). The EEG data was acquired using 64-channel Neuroscan Quick-cap and SynAmps RT amplifier at a sampling rate of 1 KHz. The ocular artifacts are removed using an automated algorithm based on blind source separation [25]. /piy/, "pat" and "'pot" are considered as bilabial and "/n/", "knew" and "gnaw" as nasal category. /m/ is excluded from the analysis since it is both nasal and bilabial.

V. Mean Phase Coherence (MPC)

Mean phase coherence (MPC) is a statistical measure of phase synchronisation between two time-varying signals [9] such as two EEG channels. This measure is closely related to phase locking value (PLV) defined for the condition in which phase differences between the studied signals are attributed to evoked activity [26], [27]. PLV measures the phase synchronisation between two channels across different trials under the assumption that every trial is time-locked to a specific stimulus. This assumption does not hold good for EEG acquired during speech imagery since the imagination is not time-locked across trials albeit the presence of cues for the participant.

MPC between electrodes i and k is defined as,

$$MPC_{i,k} = \frac{1}{N}\left|\sum_{n=0}^{N-1} e^{-j(\phi_i(n)-\phi_k(n))}\right| \tag{1}$$

where $\phi_i(n)$ and $\phi_k(n)$ correspond to the instantaneous phases (obtained using Hilbert transform) of channels i and k at discrete time n. The value of MPC lies between $[0, 1]$ where a value close to zero indicates that the phase difference between the two signals are random whereas a value of one means that they are phase synchronized during most of the time interval considered [28].

[1] http://www.cs.toronto.edu/~complingweb/data/karaOne/karaOne

VI. Classification

A shallow neural network classifier is trained on MPC values to classify nasals and bilabials. As a benchmark, we have also designed a shallow neural network classifier trained on the features followed by Zhao and Rudzicz in the paper where they introduced the Karaone dataset [6].

Zhao and Rudzicz have used several statistical features such as mean, median, variance, skewness and kurtosis along with the first and second differences of these features, resulting in 1197 features per channel. A feature selection is employed based on the Pearson correlation between these EEG features and the same features extracted from the audio signal recorded during each trial. The audio signal contains the actual utterance of the prompt which the subject imagined articulating during the EEG recording. We have tested the effect of the number of features selected using Pearson correlation by varying the number of features from 5 to 24, in steps of 1. More details about the feature extraction and selection method by Zhao and Rudzicz can be found in [6].

For the classifier trained on MPC values, the feature vector contains both the intracortical and intercortical MPC values of all the six cortical regions. For computing the MPC of each cortical region listed in Section II, MPC between all possible pairs of channels in the cortical region are computed and averaged. Similarly, intercortical MPC values between all possible pairs of channels in each pair of cortical region are computed and averaged. This is similar to the method followed by Lee et al. [29]. The classifiers are trained and tested separately on the MPC values computed from the alpha, beta and gamma frequency bands. This helps evaluate the relative discriminability of MPC values of different frequency bands. The length of the feature vector is 21 (six intracortical and 15 intercortical MPC values).

The number of neurons in the single hidden layer of the neural network is empirically set at $N/2$ where N is the length of the feature vector. The network is trained using Levenberg-Marquardt algorithm (LMA) [30], [31]. LMA approximates the Hessian matrix $\mathbf{H}$ from the Jacobian matrix $\mathbf{J}$ containing first derivatives of the network errors with respect to the weights and biases using the following relation:

$$\mathbf{H} = \mathbf{J}^T\mathbf{J} \tag{2}$$

The gradient g is computed using the following relation:

$$\mathbf{g} = \mathbf{J}^T\mathbf{e} \tag{3}$$

where $\mathbf{e}$ is the vector of network error. The LMA update equation is given by,

$$\mathbf{x_{n+1}} = \mathbf{x_n} - \left[\mathbf{J}^T\mathbf{J} + \mu\mathbf{I}\right]^{-1}\mathbf{J}^T\mathbf{e} \tag{4}$$

where μ is the combination coefficient which is always positive and $\mathbf{I}$ is the identity matrix. Equation 4 approximates Gauss–Newton algorithm when μ is very small, and steepest descent method when it is very large. Thus, the update rule switches between Gauss–Newton and steepest descent algorithms. More details about LMA is available in [32].

VII. Results

Figure 1 shows the classification results using MPC values and the statistical features. The classification follows a subject-independent model, where the classifier is trained on the data of all the subjects other than the test subject and is tested on the data of the test subject. This is a challenging test paradigm since the data of the test subject is unseen during training.

VIII. Discussion

Alpha and beta band MPC values have the lowest and the highest discriminatory powers, respectively. The performances of gamma band MPC and statistical features (benchmark classifier) are comparable, although gamma band MPC results in higher variance as shown in Fig. 3. This correlates with Fig. 1 where the classification based on gamma band MPC of subjects S1, S5 and S10 is only as good as chance whereas for S3, S4 and S13 it is comparable to that of beta band. The ages of the fourteen participants in the Karaone dataset have a range of 14 years and recent studies show that gamma oscillations weaken with age [33]. This may be a reason for the observed variance with the caveat that the subjects in the study by Murty et al. [33] are aged 58-80 years but the mean age of the participants in Karaone is only 27.4 years. Further studies are required to explain the variance in gamma band.

Figures 1 and 3 show that the classification accuracy exceeds 75% for beta band MPC for every subject except S1. S1 has the lowest accuracy for both statistical and beta band MPC features, which may be because the subject is BCI-illiterate. It is known that some participants have difficulty in imagining the prompts for a BCI, although the few studies on BCI-illiteracy are on motor imagery based BCI [34]. These results are inline with the results presented in [35] where beta band showed statistically significant difference in MPC for all the cortical regions considered whereas alpha band MPC did not show statistically significant difference in any cortical region.

IX. Conclusion

We have shown that phase synchronization among different cortical regions differs between imagining nasal and bilabial consonants. This difference can be captured with sufficient accuracy using MPC values from the EEG recorded during speech imagery. This result helps in choosing the right prompts when designing a system for decoding speech imagery. Specifically, bilabial and nasal consonant prompts, being orthogonal, can lead to speech imagery BCI systems with better accuracy.

Acknowledgment

The authors express their gratitude to Dr. Anusha A.S. and Dr. Kanishka Sharma, Indian Institute of Science, Bangalore and Dr. Frank Rudzicz, University of Toronto for the support extended to this work.

References

[1] J. T. Panachakel, A. Ramakrishnan, and T. Ananthapadmanabha, "A novel deep learning architecture for decoding imagined speech from EEG," *arXiv preprint arXiv:2003.09374*, 2020.

[2] J. T. Panachakel and R. A. Ganesan, "Decoding covert speech from EEG-a comprehensive review," *Frontiers in Neuroscience*, vol. 15, p. 392, 2021.

[3] X. Tian and D. Poeppel, "Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation," *Frontiers in human neuroscience*, vol. 6, p. 314, 2012.

[4] T. J. Whitford, B. N. Jack, D. Pearson, O. Griffiths, D. Luque, A. W. Harris, K. M. Spencer, and M. E. Le Pelley, "Neurophysiological evidence of efference copies to inner speech," *Elife*, vol. 6, p. e28197, 2017.

[5] X. Tian and D. Poeppel, "Mental imagery of speech and movement implicates the dynamics of internal forward models," *Frontiers in psychology*, vol. 1, p. 166, 2010.

[6] S. Zhao and F. Rudzicz, "Classifying phonological categories in imagined and articulated speech," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 992–996.

[7] P. Mini, T. Thomas, and R. Gopikakumari, "EEG based direct speech BCI system using a fusion of SMRT and MFCC/LPCC features with ann classifier," *Biomedical Signal Processing and Control*, vol. 68, p. 102625, 2021.

[8] J. T. Panachakel, A. Ramakrishnan, and T. Ananthapadmanabha, "Decoding imagined speech using wavelet features and deep neural networks," in *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE, 2019, pp. 1–4.

[9] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, "Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients," *Physica D: Nonlinear Phenomena*, vol. 144, no. 3-4, pp. 358–369, 2000.

[10] F. H. Guenther, "Cortical interactions underlying the production of speech sounds," *Journal of communication disorders*, vol. 39, no. 5, pp. 350–365, 2006.

[11] J. T. Panachakel, N. N. Vinayak, M. Nunna, A. G. Ramakrishnan, and K. Sharma, "An improved EEG acquisition protocol facilitates localized neural activation," in *Advances in Communication Systems and Networks*. Springer, 2020, pp. 267–281.

[12] W. Klimesch, "Alpha-band oscillations, attention, and controlled access to stored information," *Trends in cognitive sciences*, vol. 16, no. 12, pp. 606–617, 2012.

[13] S. Palva and J. M. Palva, "New vistas for a-frequency band oscillations," *Trends in Neurosciences*, vol. 30.

[14] J. J. Foxe and A. C. Snyder, "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention," *Frontiers in psychology*, vol. 2, p. 154, 2011.

[15] M. Wöstmann, S.-J. Lim, and J. Obleser, "The human neural alpha response to speech is a proxy of attentional control," *Cerebral cortex*, vol. 27, no. 6, pp. 3307–3317, 2017.

[16] H. Van Dijk, J.-M. Schoffelen, R. Oostenveld, and O. Jensen, "Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability," *Journal of Neuroscience*, vol. 28, no. 8, pp. 1816–1823, 2008.

[17] D. A. Bridwell, S. Henderson, M. Sorge, S. Plis, and V. D. Calhoun, "Relationships between alpha oscillations during speech preparation and the listener N400 ERP to the produced speech," *Scientific Reports*, vol. 8, no. 1, pp. 1–10, 2018.

[18] H. M. Hobson and D. V. Bishop, "The interpretation of mu suppression as an index of mirror neuron activity: past, present and future," *Royal Society Open Science*, vol. 4, no. 3, p. 160662, 2017.

[19] T. Tamura, A. Gunji, H. Takeichi, H. Shigemasu, M. Inagaki, M. Kaga, and M. Kitazaki, "Audio-vocal monitoring system revealed by mu-rhythm activity," *Frontiers in psychology*, vol. 3, p. 225, 2012.

[20] M. Zaepffel, R. Trachel, B. E. Kilavik, and T. Brochier, "Modulations of EEG beta power during planning and execution of grasping movements," *PloS one*, vol. 8, no. 3, p. e60060, 2013.

[21] R. Schmidt, M. H. Ruiz, B. E. Kilavik, M. Lundqvist, P. A. Starr, and A. R. Aron, "Beta oscillations in working memory, executive control of movement and thought, and sensorimotor function," *Journal of Neuroscience*, vol. 39, no. 42, pp. 8231–8238, 2019.

[22] W. J. Freeman, "Origin, structure, and role of background eeg activity. part 2. analytic phase," *Clinical Neurophysiology*, vol. 115, no. 9, pp. 2089–2107, 2004.

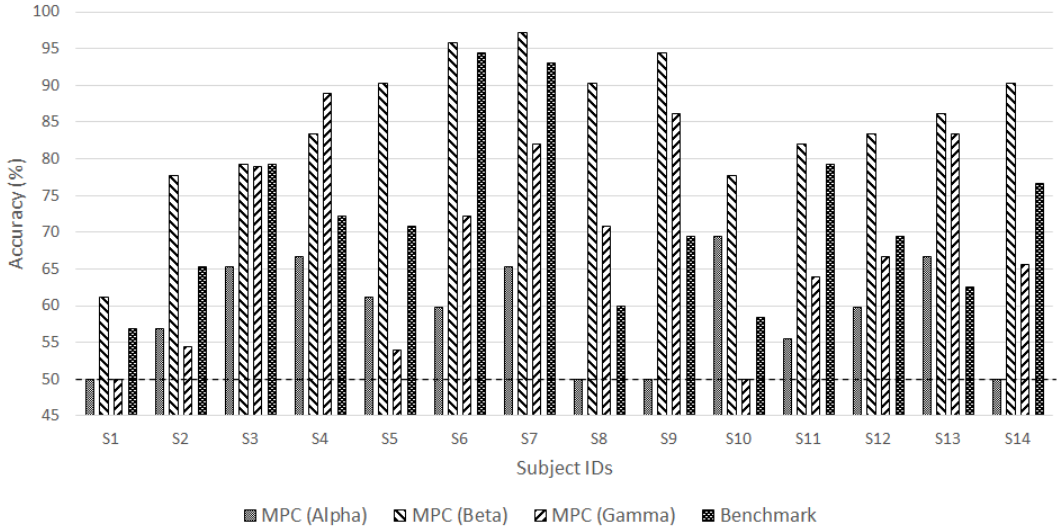

Fig. 1: Mean accuracies across trials for classifying nasal and bilabial consonants using MPC values and statistical features. The subject IDs are given in the horizontal axis and the accuracies are given in the vertical axis. The chance accuracy is given by the dashed line. For the benchmark classifier, the number of input features is $N = 19$.

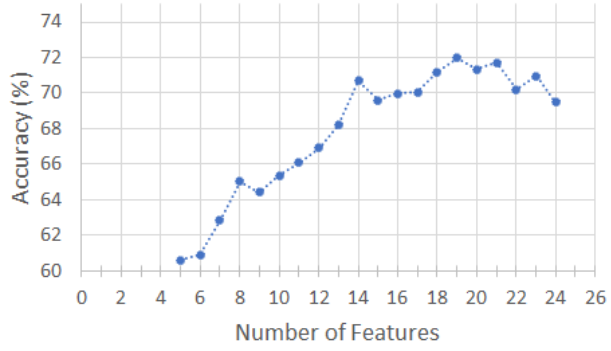

Fig. 2: Mean accuracies over all the 14 subjects as a function of the number of features when statistical features and Pearson correlation based feature selection are used for classifying nasal and bilabial consonants. The highest accuracy is obtained for $N = 19$.

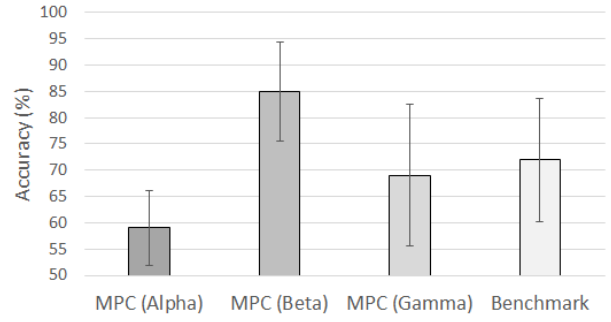

Fig. 3: Accuracies averaged over all the 14 subjects for all the four types of features. For the benchmark classifier, the number of input features is $N = 19$. The error bar shows the standard error.

[23] O. Jensen, J. Kaiser, and J.-P. Lachaux, "Human gamma-frequency oscillations associated with attention and memory," *Trends in neurosciences*, vol. 30, no. 7, pp. 317–324, 2007.

[24] L. Lu, J. Sheng, Z. Liu, and J.-H. Gao, "Neural representations of imagined speech revealed by frequency-tagged magnetoencephalography responses," *NeuroImage*, vol. 229, p. 117724, 2021.

[25] G. Gómez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, "Automatic removal of ocular artifacts in the EEG without an EOG reference channel," in *Proceedings of the 7th Nordic Signal Processing Symposium-NORSIG 2006*. IEEE, 2006, pp. 130–133.

[26] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human brain mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[27] R. Bruña, F. Maestú, and E. Pereda, "Phase locking value revisited: teaching new tricks to an old dog," *Journal of neural engineering*, vol. 15, no. 5, p. 056011, 2018.

[28] R. Q. Quiroga and S. Panzeri, *Principles of neural coding*. CRC Press, 2013.

[29] S.-H. Lee, M. Lee, and S.-W. Lee, "Functional connectivity of imagined speech and visual imagery based on spectral dynamics," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2021, pp. 1–6.

[30] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944.

[31] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[32] H. Yu and B. M. Wilamowski, "Levenberg-marquardt training," *Industrial electronics handbook*, vol. 5, no. 12, p. 1, 2011.

[33] D. V. Murty, K. Manikandan, W. S. Kumar, R. G. Ramesh, S. Purokayastha, M. Javali, N. P. Rao, and S. Ray, "Gamma oscillations weaken with age in healthy elderly in human eeg," *NeuroImage*, vol. 215, p. 116826, 2020.

[34] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High theta and low alpha powers may be indicative of bci-illiteracy in motor imagery," *PloS one*, vol. 8, no. 11, p. e80886, 2013.

[35] J. T. Panachakel, A. Ramakrishnan, A. AS, and K. Sharma, "Can we identify the category of imagined phoneme from EEG?" in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2021.