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

Spatial and Spectral Fingerprint in The Brain: Speaker Identification from Single Trial MEG Signals

Debadatta Dash[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, jun.wang@austin.utexas.edu, pferrari@utexas.edu

Abstract

Brain activity signals are unique subject-specific biological features that can not be forged or stolen. Recognizing this inherent trait, brain waves are recently being acknowledged as a far more secure, sensitive, and confidential biometric approach for user identification. Yet, current electroencephalography (EEG) based biometric systems are still in infancy considering their requirement of a large number of sensors and lower recognition performance compared to present biometric modalities. In this study, we investigated the spatial and spectral fingerprints in the brain with magnetoencephalography (MEG) for speaker identification during rest (pre-stimuli) and speech production. Experimental results suggested that the frontal and the temporal regions of the brain and higher frequency (gamma and high gamma) neural oscillations are more dominating for speaker identification. Moreover, we also found that two optimally located MEG sensors are sufficient to obtain a high speaker classification accuracy during speech tasks whereas at least eight optimally located sensors are needed to accurately identify these subjects during rest-state (pre-stimuli). These results indicated the unique neural traits of speech production across speakers.

Index Terms: brain-computer interface, magnetoencephalography (MEG), speaker identification, support vector machines

1. Introduction

The basic requirements of any physiological features to be used for identity recognition involve universality, uniqueness, stability, and collectibility [1]. Conventional biological features used in biometric systems include but are not limited to fingerprints, iris, facial features, voice, palm print, DNA, and gait. However, all of these features have the limitation of being forged or stolen. Brain activity signals, on the other hand, are unique biological features that can neither be stolen nor be forged, hence hold promising inferences for next-generation biometric identifiers. It also satisfies all the critical requirements mentioned above for identity recognition. It is universal as every human being has this feature; unique across individuals; stable as the neural functional connectivity does not change with time; and is also collectible as it can be quantitatively measured with various neuroimaging modalities such as electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and electrocorticography (ECoG).

Although the existence of genetic information in the brain waves had been proven in the 1930s [2, 3], only very recently, research on cognitive biometrics has captured the pace [4]. EEG has been the major neuroimaging modality investigated for brain based identity recognition considering its non-invasiveness, low cost, and easy setup requirements [5]. Some research among the numerous studies carried out with EEG based user identification can be found in the references [6–25]. However, EEG-Biometrics are still in infancy containing major bottlenecks such as the requirement of a large number of sensors, still relatively low performance compared to conventional biometric systems, inconsistency across different sessions (template aging effect), and the requirement of larger test recordings (lower performance with single trials).

Another major drawback of the EEG-biometric systems is the heavy dependency on the resting state data. The underlying ambiguity of the instruction 'to rest' is that there would be always inconsistency in interpretation and action across trials even by the same subject leading to incommensurable data [5]. Resting state brain activity is the major source of the template aging effect, i.e., the reliability of the recorded brain signals during rest across different sessions portray a greater distinctiveness [5]. Although to overcome this, there are studies on sensory stimuli based EEG signals for user identification, these are mostly based on visual [21, 22] or auditory perception [24] to some external stimulus. The additional need of external devices in such systems result in more complexity and cost.

In this study, we investigated magnetoencephalography (MEG) signals for speaker identification, which has not been reported before, to our knowledge. MEG records the magnetic field induced by the postsynaptic current from the cortical surface neurons with a higher spatial resolution and less distortion compared to EEG. In addition, MEG signals are reference free. Hence, MEG might be more suitable to investigate the inter-subject brain wave distinctiveness. Besides resting state, we also experimented on speaker classification from the brain activity signals during speech production tasks. Speech is an inherent trait of the speakers, and it is void of external stimuli, and hence we believe it is more reliable. We implemented artificial neural networks (ANN) and support vector machines (SVM) with different kernels to classify eight speakers from the single trial recordings of MEG signals. Further, we conducted additional experiments using different combinations of sensors and frequency bandwidths to find an optimal set of MEG sensor locations and frequency bandwidths for subject/speaker identification, respectively.

Besides subject identification for security purpose, speaker classification from the brain could also enable in the development of speaker independent speech decoding based brain computer interfaces (BCIs), which are currently speaker dependent [26, 27]. A better speaker independent speech decoding performance could be achieved by under-weighting the sensors with higher contribution in speaker classification.

Copyright © 2019 ISCA

http://dx.doi.org/10.21437/Interspeech.2019-3105

Figure 1: *The MEG system with a subject*

2. Data Collection

Eight subjects (5 males and 3 females; age=41±14 years) participated in the data collection for this study. The subjects had normal or corrected to normal vision and had no speech, language, hearing, or cognitive disorder history. Written consent was obtained from each subject prior to the experiment. This study was approved by the institutional review boards (IRB) of the participating institutions.

The MEG machine (Figure 1) has 306 sensors in total out of which 204 sensors are gradiometers and rest 102 are magnetometer sensors. The machine is housed within a magnetically shielded room (MSR) to discard the unwanted environmental magnetic field interferences. The experiment was designed as an delayed overt reading (articulation) task as shown in Figure 2. Five commonly used English phrases (*1. Do you understand me, 2. That's perfect, 3. How are you, 4. Good-bye, and 5. I need help*) were used as the stimuli of the experiment. First, the subjects were at rest for half of a second in the pre-stimuli (rest) stage. Then, a stimulus (written in English) among the 5 stimuli was displayed for 1 second on a screen situated at about 90 cm from the subject. It was followed by an imagination/preparation stage of 1 second where a fixation cross replaced the stimulus on the screen. After this stage, the fixation cross was removed, and the blank screen signaled the subjects to articulate the previously shown stimulus at their natural speaking rate. This protocol was repeated for 100 trials per stimulus per subject in a pseudo-randomized order following the standard practice of MEG experiments [28]. The whole experiment lasted for an average of 45 minutes per subject.

The MEG signals were acquired with a 4 kHz sampling frequency, those were then band-pass filtered and resampled to 1 kHz. Head motion correction was performed with the default continuous head localization technique. Through visual inspection, trials containing large eye blink, cardiac or other high amplitude motion artifacts were discarded. Further, trials with untimely articulations were also removed. After data preprocessing a total of 3046 valid trials remained out of 4000 (8 subjects $\times$ 5 phrases $\times$ 100 trials) recorded trials with an average of 75 trials per phrase per subject. Only gradiometers sensors were considered in this study since they are more effective in noise reduction than magnetometers. The sensors containing high channel noise were also discarded from the data set. For an unbiased inter-subject study, out of 204 gradiometers, a total of 187 sensors were considered, which were consistent across subjects and speech stimuli.

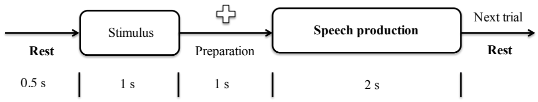

Figure 2: *The design of the time locked experiment*

3. Methods

3.1. Wavelet Analysis

Wavelets have been widely used for extracting meaningful spectral features from non-stationary brain signals for biometric applications [29,30]. Especially, Daubechies(db)-4 based discrete wavelet transform (DWT) has been shown to be effective for denoising as well as multiresolution analysis of both MEG signals [27,31] and EEG signals [30]. Db4 disintegrates the signal of interest using a range of scaled and shifted wavelet functions to capture the time-frequency attributes. Hence, for the objective of spectral feature analysis, we used db4 based 7 level DWT, which decomposed each MEG sensor signal of 1 kHz sampling frequency into 8 sub-signals (1 approximation signal: a_7 and 7 detail signals: d_{1-7}). The first 2 detail components (d_1 : $250 - 500$ Hz and d_2 : $125 - 250$ Hz) were discarded as noise components, since it is believed that the majority of the functional oscillations are restricted up to $<$ 125 Hz (high-gamma) [32]. The rest of the detail components d_{3-7} carried the frequency bandwidth range of the high-gamma, gamma, beta, alpha, and theta oscillations of the neural signals respectively, whereas the approximation signal (a_7) represented the low-frequency delta band oscillation. Root mean square (RMS) features from these individual oscillations were then obtained for further analysis considering their efficacy in our prior works on speech decoding from MEG signals [27,33].

3.2. Classification Algorithms and Experimental Setup

SVMs have been the prevalent choice of classifier in EEG based identity recognition systems [5, 18, 19] for its reliable performance on limited data. Integrating the concept of supporting vectors and kernel tricks, SVM estimates the maximum-margin hyperplane in the feature space by maximizing the 'gap' between the classes [19]. Here, we have used linear, polynomial, and radial basis function (RBF) kernel SVM to classify the 8 speakers with 4-fold cross-validation and 'one-vs-one' multi-class classification strategy. In addition, a shallow ANN classifier with 10 hidden nodes followed by a sigmoid activation function, a fully connected softmax, and a classification layer each with 8 nodes with a learning rate of 0.01 was also trained via back propagation using stochastic gradient descent up to a maximum number of 500 epochs, where 70% of the whole data was used for training and the remaining 30% data was divided equally for validation and testing. To investigate the subject distinctiveness within different regions of the brain and neural oscillations, we experimented on different cases of sensor-frequency (oscillation) combinations which are detailed below. All the classification tasks were done on the brain signals of both pre-stimuli (considered as rest) and speech production stages separately. For pre-stimuli, a single trial duration is of 0.5 second, whereas the average single trial duration for speech production is about 2 seconds. Out of the 4 classifiers considered (Linear-, Polynomial-, RBF-SVM, and ANN), RBF-SVM outperformed the remaining 3 classifiers for all the cases, yet with a very small margin (STD $<$ 3.4%). The results shown are based on the accuracy obtained with RBF-SVM.

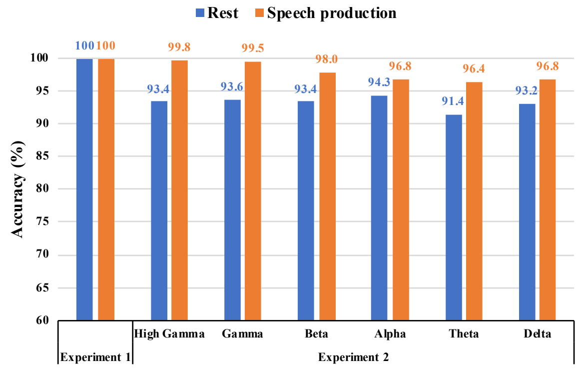

Figure 3: *Results of Experiment 1: All sensors + all oscillations and Experiment 2: all sensors + individual oscillations*

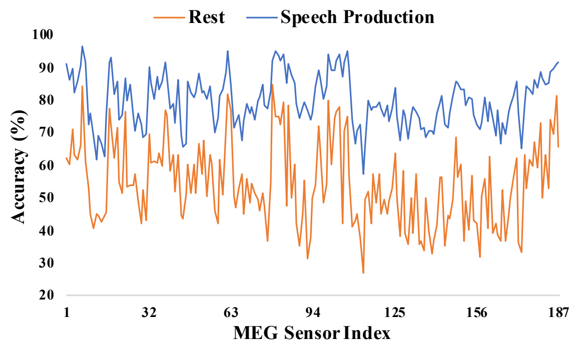

Figure 4: *Results of Experiment 3: Speaker classification accuracy obtained with each of the 187 sensors with all oscillations during rest (pre-stimuli) and speech production*

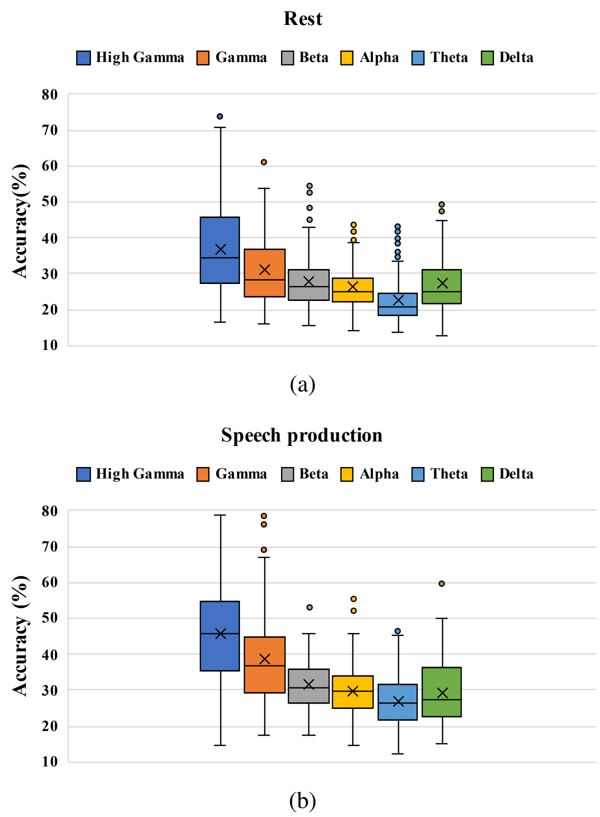

Figure 5: *Results of Experiment 4: Speaker classification accuracy distribution across the 187 sensors with individual oscillations during (a) rest and (b) speech production*

4. Results and Discussions

4.1. Experiment 1: All Sensors + All Oscillations

First, we performed speaker classification when the input feature had information from all the sensors with the brain signals containing all 6 neural oscillations (delta, theta, alpha, beta, gamma, and high-gamma). As expected, a perfect (100%) classification accuracy was obtained for both pre-stimuli and speech production stages (Figure 3). Earlier studies with EEG have shown the performance of $95\% - 100\%$ accuracy with an average of $9 - 56$ sensors. Hence, with 187 sensors and with better quality MEG signals, this performance was not surprising.

4.2. Experiment 2: All Sensors + Individual Oscillations

To identify the dominance of different neural oscillations irrespective of the brain regions, we performed our classification analysis on each oscillation data separately with all the sensors. The results of this analysis can be seen in Figure 3. Although it seems that with the speech production based brain signals, the speaker classification accuracy was better compared to the pre-stimuli state, from statistical analysis (1-tailed t-test: $p < 0.05$), it was found that there was no significant difference in the performances of both stages. However, a relatively better performance by higher frequency signals for the classification of speakers during the speech production task can be observed, unlike the resting state. Another reason for this observation is the use of all sensors (brain regions) in the analysis. Since different brain regions have distinct neural oscillation dominance, the combined effect of all the regions can be reasoned for the contribution of similar performance across the 6 oscillations.

4.3. Experiment 3: Individual Sensors + All Oscillations

To investigate the contributions of different brain regions individually on the distinctiveness of speakers, we performed the classification task with each sensor separately, irrespective of the neural oscillations. In other words, we ran the classification experiment 187 times, once for each sensor, with the information from all the 6 brain waves. Clearly from Figure 4, each sensor performed better during speech production compared to resting state with a significant difference (1-tailed t-test: $p > 0.05$). Another observation with the comparison between rest and speech production states can be made from Figure 4, that the line graph for the speech production stage is similar to a vertical shift of the line curve shown for rest. This infers that the brain regions containing the distinguishable traits across speakers are task independent (rest or speech stimulus).

4.4. Experiment 4: Individual Sensors + Individual Oscillations

Further, we experimented on the speaker classification performance across the whole spatial-spectral domain by considering each of the sensors for each neural oscillation separately (Figure 5). The dominance of high-frequency oscillations both during rest and articulation can be inferred from the two box

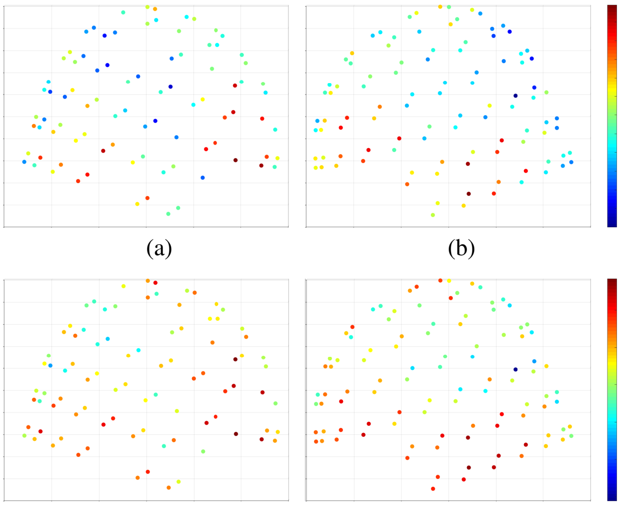

Figure 6: *Accuracy based heat map on MEG sensor locations during rest: (a)right, (b) left and during speech production: (c) right, (d) left.*

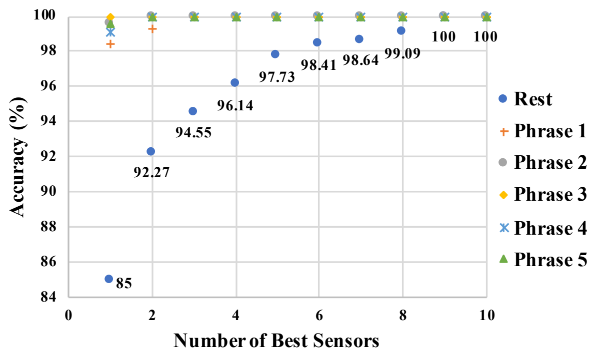

Figure 7: *Optimal number of sensors required vs. speaker classification accuracy for rest and speech production*

plots. Particularly during speech production, the majority of sensors with high-gamma and gamma band oscillations significantly contributed more towards the speaker classification compared to the remaining 4 oscillations which were found to have no statistically significant difference (1-Way ANOVA; post hoc Tukey test, $p > 0.05$). From this analysis as well, the higher performance of speech stimuli based classification compared to the resting brain data is prevalent.

4.5. Spatial Analysis

Figure 6 shows the heat map based on the accuracy obtained via each sensor with all neural oscillations during rest and speech production. In both cases, the left and right frontal and temporal (some parietal) regions of the brain show dominance for speaker classification indicating the possibility of the existence of fingerprint in those regions of the brain. The overlapping of brain regions for both rest and speech production also suggest that the spatial fingerprint in the brain is task independent.

4.6. Optimal Number of Sensors for Subject and Speaker Identification

Lastly, we investigated the optimal number of sensors required for obtaining 100% classification accuracy across the 8 speakers for rest and speech task, respectively. We started with the sensor with the highest performance and then kept on adding the information from the sensor with the next best performance until 100% classification accuracy was achieved. Based on our previous results of spectral dominance, in this analysis, we considered all the 6 neural oscillations for the optimal combination of sensors. Experimental results indicated that at least 8 optimally positioned sensors are required for a perfect classification in resting state whereas only 2 sensors were found to be sufficient to obtain the perfect classification accuracy during speech production (Figure 7). One possible explanation for this observation maybe that the behavioral traits across speakers during speech production might be contributing towards speaker distinctiveness. Nevertheless, the recurring superior performance of speech stimuli for speaker classification suggests that a better cognitive biometric system can be developed with speech tasks.

In addition, we also analyzed the speaker classification performance for each speech stimulus (phrase) separately, where we evaluated the performance by training and testing the features corresponding to each of the 5 phrases separately. The speaker classification performance across the different stimuli was consistently high (Figure 7) without any significant difference among the stimuli.

5. Conclusions

In this study, we investigated the spatial and spectral fingerprints in the brain by classifying 8 subjects from their neuromagnetic signals during rest (pre-stimuli) and speech production tasks. Regarding frequency analysis, we found that the higher frequency components (gamma and high-gamma) were more dominating than lower frequency bands in distinguishing speakers. Moreover, overlapping regions of frontal and temporal areas of the brain were obtained to be dominating in speaker classification, both during rest and speech production. Regards to biometric applications, we found that only 2 MEG sensors were needed for a 100% subject classification during speech task whereas 8 sensors was the minimum requirement for the resting state. These results motivate the possibility of MEG based biometric systems in the future. Although the results are promising, the inferences are needed to be verified with a larger sized speaker data and with intersession scans, towards which the current study is being escalated.

Although the present MEG system has the limitations of non-portability and high cost, a recent study on movable MEG [34] indicated that MEG can be portable in the future. Further, this mobile MEG system has the capability of using different number of sensors starting from one, unlike the present MEG sensor cap where all the sensors are grouped. Hence, this flexibility combined with the lower number of optimal MEG sensor requirements could contribute to a better low cost cognitive biometric system in the future.

6. Acknowledgments

This study was supported by the University of Texas System Brain Research grant and NIH grant R03DC013990. We thank Dr. Ted Mau, Dr. Saleem Malik, Dr. Daragh Heitzman, Dr. Angel W. Hernandez-Mulero, Dr. Alan Wisler, and the volunteering participants.

7. References

[1] W. Liang, L. Cheng, and M. Tang, "Identity recognition using biological electroencephalogram sensors," *Journal of Sensors*, vol. 2016, no. 1831742, p. 9, 2016.

[2] H. Berger, "Über das elektrenkephalogramm des menschen," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 87, no. 1, pp. 527–570, Dec 1929. [Online]. Available: https://doi.org/10.1007/BF01797193

[3] F. Vogel, "The genetic basis of the normal human electroencephalogram (EEG)," *Humangenetik*, vol. 10, no. 2, pp. 91–114, Mar 1970.

[4] P. Campisi, D. La Rocca, and G. Scarano, "EEG for automatic person recognition," *Computer*, vol. 45, no. 7, pp. 87–89, July 2012.

[5] S. Yang and F. Deravi, "On the usability of electroencephalographic signals for biometric recognition: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 958–969, Dec 2017.

[6] M. Poulos, M. Rangoussi, and N. Alexandris, "Neural network based person identification using EEG features," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 2, March 1999, pp. 1117–1120.

[7] S. Yang and F. Deravi, "Novel HHT-based features for biometric identification using EEG signals," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 1922–1927.

[8] P. Campisi and D. L. Rocca, "Brain waves for automatic biometric-based user recognition," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 782–800, May 2014.

[9] D. L. Rocca, P. Campisi, B. Vegso, P. Cserti, G. Kozmann, F. Babiloni, and F. D. V. Fallani, "Human brain distinctiveness based on EEG spectral coherence connectivity," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2406–2412, Sep. 2014.

[10] M. Fraschini, A. Hillebrand, M. Demuru, L. Didaci, and G. L. Marcialis, "An EEG-based biometric system using eigenvector centrality in resting state brain networks," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 666–670, June 2015.

[11] K. V. R. Ravi and R. Palaniappan, "Recognising individuals using their brain patterns," in *Third International Conference on Information Technology and Applications (ICITA'05)*, vol. 2, July 2005, pp. 520–523.

[12] R. Palaniappan, "Vision related brain activity for biometric authentication," in *IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics*, Nov 2006, pp. 3227–3231.

[13] R. Palaniappan and D. P. Mandic, "EEG based biometric framework for automatic identity verification," *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 49, no. 2, pp. 243–250, Nov 2007.

[14] S. Marcel and J. D. R. Millan, "Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 743–752, April 2007.

[15] Q. Gui, Z. Jin, M. V. Ruiz Blondet, S. Laszlo, and W. Xu, "Towards EEG biometrics: pattern matching approaches for user identification," in *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015)*, March 2015, pp. 1–6.

[16] J. Chuang, H. Nguyen, C. Wang, and B. Johnson, "I think, therefore I am: Usability and security of authentication using brainwaves," in *Financial Cryptography and Data Security*, A. A. Adams, M. Brenner, and M. Smith, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–16.

[17] I. Nakanishi, S. Baba, and C. Miyamoto, "EEG based biometric authentication using new spectral features," in *2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Jan 2009, pp. 651–654.

[18] K. Brigham and B. V. K. V. Kumar, "Subject identification from electroencephalogram (EEG) signals during imagined speech," in *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sep. 2010, pp. 1–8.

[19] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein, "Low-cost electroencephalogram (EEG) based authentication," in *2011 5th International IEEE/EMBS Conference on Neural Engineering*, April 2011, pp. 442–445.

[20] S. Yeom, H. Suk, and S. Lee, "EEG-based person authentication using face stimuli," in *2013 International Winter Workshop on Brain-Computer Interface (BCI)*, Feb 2013, pp. 58–61.

[21] R. Palaniappan and K. V. R. Ravi, "Improving visual evoked potential feature classification for person recognition using PCA and normalization," *Pattern Recogn. Lett.*, vol. 27, no. 7, pp. 726–733, May 2006. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2005.10.020

[22] A. Yazdani, A. Roodaki, S. H. Rezatofighi, K. Misaghian, and S. K. Setarehdan, "Fisher linear discriminant based person identification using visual evoked potentials," in *2008 9th International Conference on Signal Processing*, Oct 2008, pp. 1677–1680.

[23] H. J. Lee, H. S. Kim, and K. S. Park, "A study on the reproducibility of biometric authentication based on electroencephalogram (EEG)," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, Nov 2013, pp. 13–16.

[24] F. Chunying, L. Haifeng, M. Lin, and J. Bing, "Induced event-related coherence measures during auditory change detection," in *2014 International Conference on Medical Biometrics*, May 2014, pp. 118–124.

[25] Y. Bai, Z. Zhang, and D. Ming, "Feature selection and channel optimization for biometric identification based on visual evoked potentials," in *2014 19th International Conference on Digital Signal Processing*, Aug 2014, pp. 772–776.

[26] 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.

[27] 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.

[28] 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.

[29] C. N. Gupta, Y. U. Khan, R. Palaniappan, and F. Sepulveda, "Wavelet framework for improved target detection in oddball paradigms using P300 and gamma band analysis (special issue; biosensors: Data acquisition, processing and control)," *International Journal of Biomedical Soft Computing and Human Sciences: the official journal of the Biomedical Fuzzy Systems Association*, vol. 14, no. 2, pp. 63–69, 2009.

[30] S. Yang and F. Deravi, "Wavelet-based EEG preprocessing for biometric applications," in *2013 Fourth International Conference on Emerging Security Technologies*, Sep. 2013, pp. 43–46.

[31] D. Dash, P. Ferrari, S. Malik, and J. Wang, "Automatic speech activity recognition from MEG signals using seq2seq learning," in *2018 IEEE EMBS International Conference on Neural Engineering (NER)*, Mar 2019, pp. 340–343.

[32] A. Ahnaou, H. Huysmans, T. V. de Casteele, and W. H. Drinkenburg, "Cortical high gamma network oscillations and connectivity: a translational index for antipsychotics to normalize aberrant neurophysiological activity," in *Translational Psychiatry*, 2017.

[33] 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.

[34] 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.