Compact and interpretable architecture for speech decoding from stereotactic EEG

Artur Petrosyan
Center for Bioelectric Interfaces
Higher School of Economics Moscow
Russia, 101000
Email: apetrosyan@hse.ru

Alexey Voskoboynikov
Center for Bioelectric Interfaces
Higher School of Economics Moscow
Russia, 101000
Email: avoskoboinikov@hse.ru

Alexei Ossadtchi
Center for Bioelectric Interfaces
Higher School of Economics Moscow
Russia, 101000
Email: aossadtchi@hse.ru

Abstract—**Background: Brain-computer interfaces (BCIs) decode neural activity and extract from it information that can be meaningfully interpreted. One of the most intriguing opportunities is to employ BCIs for decoding speech, a uniquely human trait, which opens up plentiful applications from rehabilitation of patients to a direct and seamless communication between human species. To decipher neuronal code complex deep neural networks furnish only limited success. In such solutions an iffy performance gain is achieved with uniterpretable decision rules characterised by thousands of parameters to be identified from a limited amount of training data. Our recent experience shows that when applied to neural activity data compact neural networks with trainable and physiologically meaningful feature extraction layers [1] deliver comparable performance, ensure robustness of the learned decision rules and offer the exciting opportunity of automatic knowledge discovery.**

Methods: We collected approximately one hour of data (from two sessions) where we recorded stereotactic EEG (sEEG) activity during overt speech (6 different randomly shuffled phrases and rest). We have also recorded synchronized audio speech signal. The sEEG recording was carried out in an epilepsy patient implanted for medical reasons with an sEEG electrode passing through Broca area with 6 contacts spaced at 5 mm. We then used a compact convolutional network-based architecture to recover speech mel-cepstrum coefficients followed by a 2D convolutional network to classify individual words. We then interpreted the former network weights using the theoretically justified approach devised by us earlier [1].

Results: We achieved on average 44% accuracy in classifying 26+1 words (3.7% chance level) using only 6 channels of data recorded with a single minimally invasive sEEG electrode. We compared the performance of our compact convolutional network to that of the DenseNet-like architecture that has recently been featured in neural speech decoding literature and did not find statistically significant performance differences. Moreover, our architecture appeared to be able to learn faster and resulted in a stable, interpretable and physiologically meaningful decision rule successfully operating over a contiguous data segment nooverlapping with the training data interval. Spatial characteristics of neuronal population pivotal to the task corroborate the results of active speech mapping procedure and frequency domain patterns show primary involvement of the high frequency activity.

Conclusions : Most of the speech decoding solutions available to date either use potentially harmful intracortical electrodes or rely on the data recorded with impractically massive multielectrode grids covering large cortical area. Here we for the first time achieved practically usable decoding accuracy for the vocabulary of 26 words + 1 silence class backed by only 6 channels of cortical activity sampled with a single sEEG shaft. The decoding was implemented using a compact and interpretable architecture which ensures robustness of the solution and requires small amount of training data. The proposed approach is the first step towards minimally invasive implantable BCI solution for restoring speech function.

I. Introduction

Brain-computer interfaces (BCIs) directly link the nervous system to external devices [2] or even other brains [3]. While there exist many applications of BCIs [4], clinically relevant BCIs are of primary interest since they hold promise to rehabilitate patients with sensory, motor, and cognitive disabilities [5],[6].

BCIs can deal with a variety of neural signals [7], [8] such as, for example, electroencephalographic (EEG) potentials sampled with electrodes placed on the surface of the head [9], or neural activity recorded invasively with intracortical electrodes penetrating cortex [10] or placed onto the cortical surface [11]. A promising and minimally invasive way to directly access cortical activity is to use stereotactic EEG (sEEG) electrodes inserted stereotactically via a twist drill or a burr hole made in the skull. Recent advances in implantation techniques including the use of brain's 3D angiography, MRI and robot-assisted surgery help to further reduce the risks of such an implantation and make sEEG technology an ideal trade-off for BCI applications [12].

Current study deals with restoration of speech function, one of the most exciting potential applications of the BCI technology. Several attempts have already been made and certain progress is achieved in decoding both individual words [13], [14] and phonetic features [15] with practically usable accuracy. However, these studies relied on heavily multichannel brain activity measurements implemented either with intracortical arrays [16] or with massive ECoG grids [17], [13], [18] covering significant cortical area. Both solutions for reading off brain activity are not intended for a long term use and are associated with significant risks to a patient [19]. sEEG is a promising alternative that has already being tried for the speech decoding task [20] with some success. This study, however, relies on the high count of sEEG channels distributed over a large part of the left frontal and left superior temporal lobes which hinders practical applications.

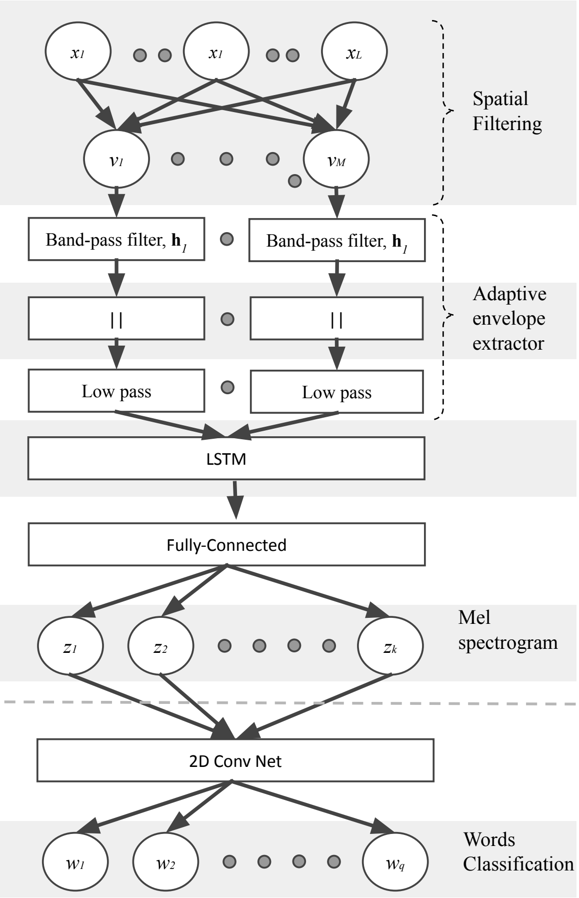

Fig. 1. The architecture based on[1] and adapted for speech classification task. We used the same envelope detector technique to extract robust and meaningful features from the sEEG. We then used the LSTM layer to account for the sequential structure of the mel-spectrogram and finally decoded it with a fully connected layer over the LSTM hidden state. A separate 2D convolutional network was trained and used to classify separate words on top of the decoded mel-spectrogram.

Here we explore the possibility of decoding individual words from the sEEG data sampled with a single 6-channel shaft penetrating the brain and passing though critical speech areas. To ensure reliable and interpretable decoding we extend and employ our compact deep neural network architecture with factorised spatial and temporal processing [1].

II. Methods

We collected approximately one hour of data (split into two sessions) where we recorded sEEG activity during overt speech (6 different randomly shuffled phrases and rest). We have also recorded synchronized audio speech signals. The sEEG recording was carried out in an intellectually intact epilepsy patient implanted for medical reasons with an sEEG electrode passing through Broca area with 6 contacts spaced at 5 mm as shown in Figure 2.

We first employed a compact convolutional network architecture developed by us earlier for motor BCI purposes [1] and augmented it with a single LSTM layer. Since here we aim to decode our intermediate target comprising several mel-spectral coefficients we have also modified the fully connected layer so that it has $M = 80$ output neurons each corresponding to a mel-spectral coefficient whose temporal profile we are aiming to reconstruct from the 6-channel sEEG data. Note that unlike in [20] we do not specify upfront the feature extraction parameters and let our architecture learn them during the training process which aimed to optimise mean Pearson's correlation coefficient between the original and sEEG decoded mel-spectrum timeseries. For training we used the first 70% of data which corresponded to the contiguous block of data of duration approximately 40 minutes. The last 20 minute segment was used for testing. As our intermediate target we have also experimented with speech envelope and linear predictive coding (LPC) coefficients.

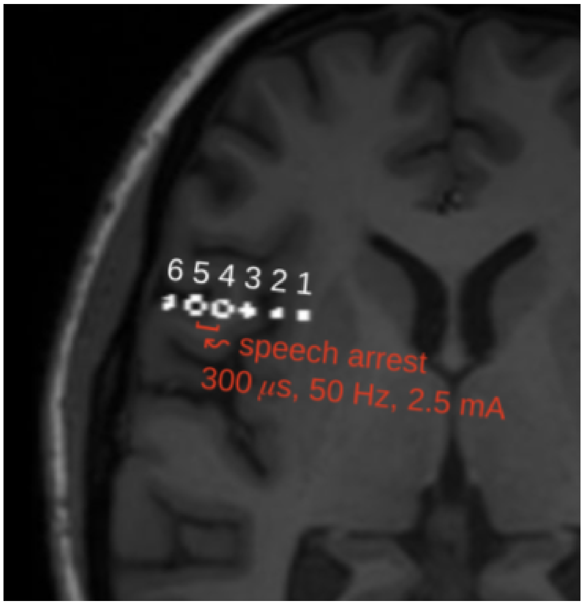

Fig. 2. Stereotactic EEG (sEEG) electrode contacts used in this study extracted from post-surgical CT scan and superimposed onto subject's MRI. Bipolar stimulation between the 4th and the 5th electrodes consistently caused speech arrest in this patient.

After having trained our compact architecture to decode the intermediate target (mel-spectrum, speech envelope, LPC) we used a 2D-convolution network to perform discrete classification of 26 words and the silent class. For this step we prepared the data by cutting the segments around each word using our in house developed simple threshold based procedure. Here again, we used the first 70% percents of word repetitions for training and the last 30% for testing which roughly ensured no overlap of the contiguous time intervals from which the training and testing data samples were collected. We used cross-entropy as a loss function to train the 2-D convolutional network, see Figure 1.

III. Results

Our compact architecture processing only 6 sEEG channels form a single sEEG shaft achieved 62% mean correlation

TABLE I
DECODING PERFORMANCE ACHIEVED WITH COMPACT ARCHITECTURE. THE TABLE SHOWS CORRELATION COEFFICIENT FOR MEL-SPECTORGRAM, LPC-COEFFICIENTS AND SPEECH ENVELOPE. AND ALSO ACCURACY FOR WORDS CLASSIFICATION

	Mels	LPC	Envelope	Words
Compact DNN	0.62	0.51	0.52	0.44

coefficient over $M = 80$ mel-spectral coefficients which is comparable to the accuracy reported in [18] where significantly greater count of data channels collected by ten sEEG shafts was used. Decoding the timeseries of ten LPC coefficients and speech envelope yielded 51 % and 52 % correlation correspondingly, see Table I. An example of the original and sEEG-decoded 80 mel-spectral coefficients is shown in Figure 3. The achieved so far decoding accuracy does not yield intelligible speech when the recovered mel-spectrum is converted back into the sound. Nevertheless, the decoded mel-spectral patterns support the classification of discrete words sufficiently well.

We achieved on average 44% accuracy in classifying 26+1 words (3.7% chance level) using only 6 channels of data recorded with a single minimally invasive sEEG electrode. Figure 4 shows the corresponding confusion matrix. Interestingly, according to this matrix words 6,7 and 8 tend to be confused and at the same time these words are characterised by the presence of prominent fricative sounds ”[sh]” and ”[zh]”. Also words 15 and 16 get confused and both share a very pronounced ”[l]” sound.

Spatial characteristics of neuronal population pivotal to the task as shown in the top panel of Figure 5 that emphasise the importance channel 6 for decoding partly corroborate the results of active speech mapping procedure which found that bipolar electrical stimulation of electrodes 4 and 5 resulted into transient speech arrest as shown in Figure 2. Frequency domain patterns presented in the bottom panel of Figure 5 show primary involvement of the high frequency activity which is in agreement with most invasive BCI studies.

Using the task at hands we have compared the performance of our extended compact convolutional network to that of the DenseNet-like architecture [21] that has recently been featured in neural speech decoding literature [18] and did not find statistically significant performance differences. Moreover, due to its compactness our architecture appears to learn faster and results in a stable, interpretable and physiologically meaningful decision rule successfully operating over a contiguous data segment non-overlapping with the training data time interval.

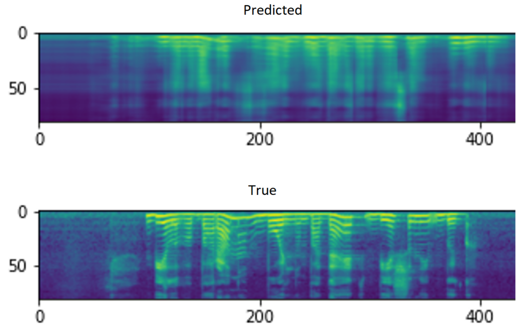

Fig. 3. Example of a true mel-spectrogram and decoded from sEEG mel-spectrogram

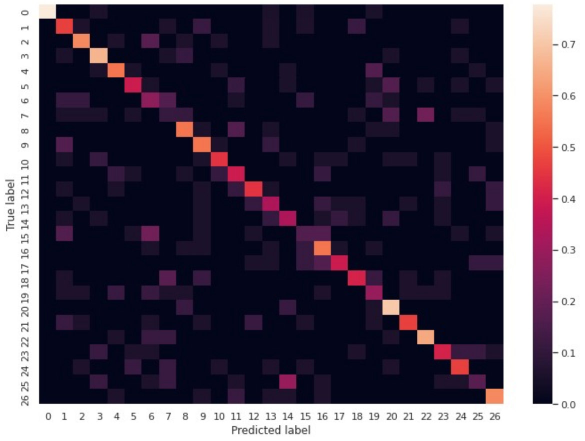

Fig. 4. Confusion matrix of classified words. Words list: 0. silence, 1. zhenia, 2. shiroko, 3. shagaet, 4. zheltykh, 5. shtanakh, 6. shuru, 7. uzhalil, 8. shershen, 9. lara, 10. lovko, 11. krutit, 12. rul, 13. levoi, 14. rukoi, 15. liriku, 16. liubit, 17. lilia, 18. babushka, 19. boitsia, 20. barabanov, 21. belogo, 22. barana, 23. bolno, 24. bodaet, 25. beshenyi, 26. byk

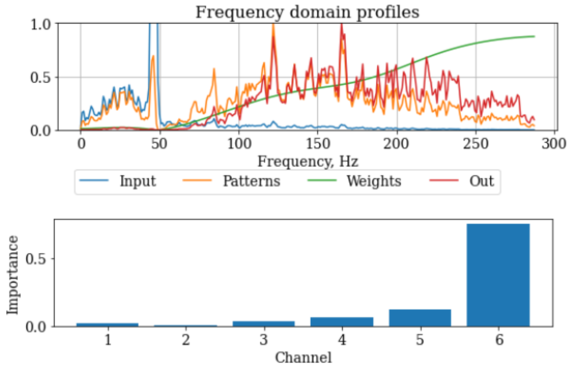

Fig. 5. Theoretically justified weights interpretation applied to the most relevant branch of architecture in Figure 1. Orange trace in the top panel shows power spectral density pattern of the activity of the neuronal population this branch is tuned to. The bottom panel shows the spatial pattern of this population. We can conclude that this source dominantly projects onto the 6th contact located at the lateral part of the sEEG electrode (shaft), see Figure 2.

IV. CONCLUSION

Most of the speech decoding solutions available to date either use potentially harmful intracortical electrodes or rely on the data recorded with impractically massive multi-electrode grids covering large cortical area. Here we for the first time achieved practically usable 44% of decoding accuracy for the vocabulary of 26 words + 1 silence class backed by only 6 channels of cortical activity sampled with an sEEG electrode. The decoding was implemented with a compact and interpretable architecture which ensures robustness of the solution and requires small amount of training data. Our experiments (not described here) show that the compact architecture delivers the accuracy comparable to that obtained with a larger and less interpetable architectures. The proposed approach is the first step towards practical minimally invasive implantable BCI solution for restoring speech function.

Acknowledgment

This work is supported by contract TS 20210517019 /6.12-20/250521-1 ”Brain activity based speech decoding” from LLC Huawei to the Center for Bioelectric Interfaces, NRU HSE.

References

[1] A. Petrosyan, M. Sinkin, M. Lebedev, and A. Ossadtchi, “Decoding and interpreting cortical signals with a compact convolutional neural network,” *Journal of Neural Engineering*, vol. 18, no. 2, p. 026019, 2021.

[2] N. G. Hatsopoulos and J. P. Donoghue, “The science of neural interface systems,” *Annual review of neuroscience*, vol. 32, pp. 249–266, 2009.

[3] M. Pais-Vieira, M. Lebedev, C. Kunicki, J. Wang, and M. Nicolelis, “A brain-to-brain interface for real-time sharing of sensorimotor information,” *Scientific reports*, vol. 3, p. 1319, 02 2013.

[4] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, “Brain computer interfacing: Applications and challenges,” *Egyptian Informatics Journal*, vol. 16, no. 2, pp. 213–230, 2015.

[5] J. N. Mak and J. R. Wolpaw, “Clinical applications of brain-computer interfaces: current state and future prospects,” *IEEE reviews in biomedical engineering*, vol. 2, pp. 187–199, 2009.

[6] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, “Brain–computer interfaces for communication and rehabilitation,” *Nature Reviews Neurology*, vol. 12, no. 9, p. 513, 2016.

[7] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[8] M. A. Lebedev and M. A. Nicolelis, “Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation,” *Physiological reviews*, vol. 97, no. 2, pp. 767–837, 2017.

[9] S. Machado, F. Araújo, F. Paes, B. Velasques, M. Cunha, H. Budde, L. F. Basile, R. Anghinah, O. Arias-Carrión, M. Cagy *et al.*, “Eeg-based brain-computer interfaces: an overview of basic concepts and clinical applications in neurorehabilitation,” *Reviews in the Neurosciences*, vol. 21, no. 6, pp. 451–468, 2010.

[10] M. L. Homer, A. V. Nurmikko, J. P. Donoghue, and L. R. Hochberg, “Sensors and decoding for intracortical brain computer interfaces,” *Annual review of biomedical engineering*, vol. 15, pp. 383–405, 2013.

[11] G. Schalk and E. C. Leuthardt, “Brain-computer interfaces using electrocorticographic signals,” *IEEE reviews in biomedical engineering*, vol. 4, pp. 140–154, 2011.

[12] C. Herff, D. J. Krusienski, and P. Kubben, “The potential of stereotactic-eeg for brain-computer interfaces: current progress and future directions,” *Frontiers in neuroscience*, vol. 14, p. 123, 2020.

[13] J. G. Makin, D. A. Moses, and E. F. Chang, “Machine translation of cortical activity to text with an encoder–decoder framework,” *Nature Neuroscience*, vol. 23, no. 4, pp. 575–582, 2020.

[14] P. Sun, G. K. Anumanchipalli, and E. F. Chang, “Brain2char: a deep architecture for decoding text from brain recordings,” *Journal of Neural Engineering*, vol. 17, no. 6, p. 066015, 2020.

[15] N. F. Ramsey, E. Salari, E. J. Aarnoutse, M. J. Vansteensel, M. G. Bleichner, and Z. Freudenburg, “Decoding spoken phonemes from sensorimotor cortex with high-density ecog grids,” *Neuroimage*, vol. 180, pp. 301–311, 2018.

[16] G. H. Wilson, S. D. Stavisky, F. R. Willett, D. T. Avansino, J. N. Kelemen, L. R. Hochberg, J. M. Henderson, S. Druckmann, and K. V. Shenoy, “Decoding spoken english from intracortical electrode arrays in dorsal precentral gyrus,” *Journal of Neural Engineering*, vol. 17, no. 6, p. 066007, 2020.

[17] H. Akbari, B. Khalighinejad, J. L. Herrero, A. D. Mehta, and N. Mesgarani, “Towards reconstructing intelligible speech from the human auditory cortex,” *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[18] M. Angrick, C. Herff, E. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, “Speech synthesis from ecog using densely connected 3d convolutional neural networks,” *Journal of neural engineering*, vol. 16, no. 3, p. 036019, 2019.

[19] P. Jayakar, J. Gotman, A. S. Harvey, A. Palmini, L. Tassi, D. Schomer, F. Dubeau, F. Bartolomei, A. Yu, P. Kršek, D. Velis, and P. Kahane, “Diagnostic utility of invasive eeg for epilepsy surgery: Indications, modalities, and techniques,” *Epilepsia*, vol. 57, no. 11, pp. 1735–1747, 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.13515

[20] M. Angrick, M. Ottenhoff, L. Diener, D. Ivucic, G. Ivucic, S. Goulis, J. Saal, A. J. Colon, L. Wagner, D. J. Krusienski *et al.*, “Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity,” *bioRxiv*, 2020.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.