Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-78

APPROACHES TO ZERO-SHOT STIMULUS DECODING IN ECOG FOR POTENTIAL BCI APPLICATIONS

C. Ratto[1], C. Caceres[1], M. Roos[1], K. Rupp[2], G. Milsap[2], N. Crone[3], and M. Wolmetz[1]

[1] The Johns Hopkins University Applied Physics Laboratory, Laurel, MD, United States
[2] Dept. of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States
[3] Department of Neurology, Johns Hopkins University, Baltimore, MD, United States

E-mail: christopher.ratto@jhuapl.edu

ABSTRACT: Most Brain-Computer Interface (BCI) work has focused on detecting specific sensory or motor information, but BCIs are beginning to be applied to more abstract domains like covert speech and communication of semantic thought. One potential approach to decoding more abstract information is linear zero-shot classification via semantic attributes, which is computationally efficient and may facilitate real-time processing. In this work, several variations of this model are applied to electrocorticography (ECoG) data recorded during a picture-naming task with nine patients. Performances of encoding and decoding models are compared, and results are discussed in the context of BCI applications.

INTRODUCTION

Sensory and motor information can be understood and encoded in terms of physical functions and attributes, and brain-computer interface (BCI) applications typically utilize models based on these characteristics, e.g. limb motion [1] or speech [2]. These approaches are not applicable in more abstract domains, such as lexical semantics and conceptual thought. However, words or concepts can be decomposed into sets of meaningful attributes [3], so one can study how those attributes are encoded in the brain. For example, the concept of "lettuce" might be encoded with heavy weight on the attributes "green," "edible," and "plant" and low weight on the attributes "black," "manmade," and "hard."

By using machine learning methods to derive these decompositional models from neural data, e.g. functional magnetic resonance imagery (fMRI), a better understanding of how more abstract concepts are represented in the human brain has been achieved [4, 5, 6, 7]. Stimuli can be represented by their constituent semantic attributes, and mappings can be learned between each attribute and the observed neural responses. In the same way, stimuli can be recovered by induction after applying the mapping to novel neural data.

While this approach has been a boon for studying how abstract representations are semantically *encoded* in neural activations, there are clear advantages for neural *decoding* applications as well. Semantically decoding neural signals in this manner allows for the classification of novel classes of stimuli. This process, coined *zero-shot* classification [5], differs from traditional pattern recognition, in which models are tested on new data from the same classes used to train the model. In zero-shot classification, models are tested on data from *new* classes that were *not* used to train the model. Zero-shot classification has been successfully demonstrated in several applications such as computer vision [8] and target detection [9].

Zero-shot classification on neural signals would allow for BCIs to handle novel stimuli more robustly. In previous work [10], it was demonstrated that zero-shot classification of recognized objects was possible from electrocorticography (ECoG) at high levels of performance on par with whole-brain fMRI [4]. Demonstrations of reliable neural decoding performance from electrophysiological responses like this suggest a viable path to BCIs for more abstract domains. For example, zero-shot decoding could potentially be used to classify and/or semantically annotate novel stimuli that produce P300 responses [11], such as anomalous images [12] or frames of video [13]. Furthermore, communication BCIs, such as those used by locked-in individuals [14], could potentially use zero-shot classification to decode conceptual thought as opposed to individual characters.

While promising, applications of zero-shot decoding to ECoG are new and not well-explored. The mapping between semantic attributes and neural features may be learned as an *encoding model*, i.e. a map from attributes to neural features [4, 10], or as a *decoding model*, *i*.e. direct prediction of attributes from neural signals [5, 6], but the efficacy of these approaches have not been compared. The mapping is often assumed to be linear, and typically learned by either least-squares [4] or ridge regression [5, 6, 10] to limit the possibility of overfitting. Support vector machines (SVMs) have been used for classifying neural data in past studies of human-computer interaction [15], as well as in zero-shot classification for computer vision [8]. However SVMs have not yet been investigated for zero-shot ECoG decoding. In this paper, we build on the encoding model described by our group in [10] to compare different approaches to zero-shot decoding in ECoG.

 DOI: 10.3217/978-3-85125-533-1-78

MATERIALS AND METHODS

The experiments carried out for this work utilized ECoG recordings collected during a picture-naming task. Data was recorded from 9 patients with intractable epilepsy (2 female, 31-44 years old) during inpatient monitoring for pre-surgical localization of their ictal onset zone and eloquent cortex. All patients provided informed consent according to a protocol approved by the Johns Hopkins Medicine Institutional Review Boards.

The stimulus set was originally reported in [4], and the data collection paradigm and analyses were originally reported in [10]. White line drawings of objects were briefly presented on a black background, and a white fixation cross was shown between stimuli. Each image was shown for one second, with a rest interval varying randomly between 3.5 and 4.5 seconds. Participants were instructed to name the image as soon as possible, or pass when necessary. Six blocks of data were collected per patient, with all 60 objects being shown in pseudo-random order within each block. ECoG signals were sampled at 1000 Hz, digitized, and recorded using the BlackRock Neuroport system.

The stimuli consisted of line drawing representations of 60 nouns from 12 semantic categories as listed in Tab. 1. Each of the 60 nouns was uniquely mapped to a vector of $P = 218$ semantic attributes originally used in [5]. The attributes were generated by crowdsourcing answers to a series of 218 questions via Amazon Mechanical Turk. All 218 questions were asked of 1,000 different objects, including all 60 of the objects included in this study. Questions probed a variety of semantic properties, including size, usage, composition, and category, with answers on an ordinal scale $[-1, -0.5, 0, 0.5, 1]$. It was empirically determined that regression models tended to perform better when the vector of attributes for a given noun was normalized to unity length.

A high-level illustration of the neural feature extraction process is shown in Fig. 1. After data collection, excessively noisy channels were discarded,

Table 1: List of stimulus nouns and their categories.

Category	Nouns
Animals	bear, cat, cow, dog, horse
Body parts	arm, eye, foot, hand, leg
Buildings	apartment, barn, church, house, igloo
Building parts	arch, chimney, closet, door, window
Clothing	coat, dress, pants, shirt, skirt
Furniture	bed, chair, desk, dresser, table
Insects	ant, bee, beetle, butterfly, fly
Kitchen Utensils	bottle, cup, glass, knife, spoon
Manmade Objects	bell, key, refrigerator, telephone, watch
Tools	chisel, hammer, pliers, saw, screwdriver
Vegetables	carrot, celery, corn, lettuce, tomato
Vehicles	airplane, bicycle, car, train, truck

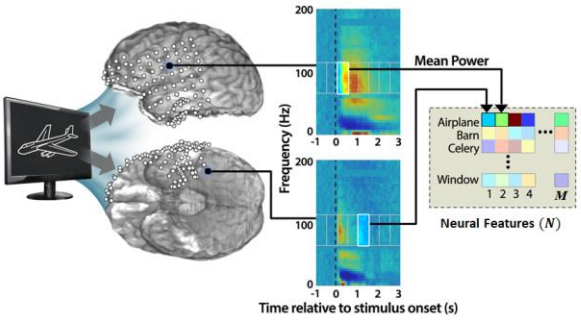

Figure 1. Illustration of neural feature extracton from ECoG recordings, adapted from [10].

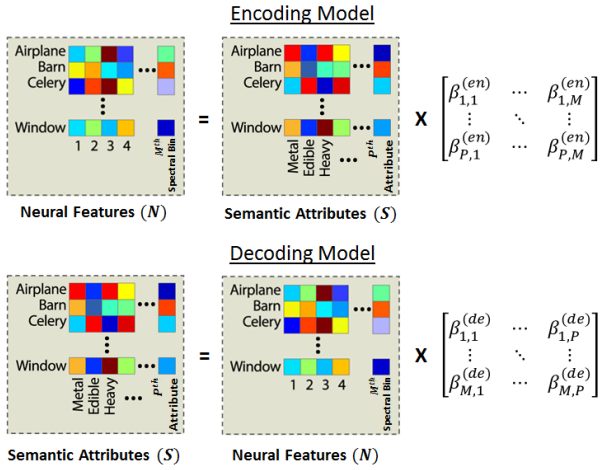

Figure 2: Illustration of encoding (top) and decoding (bottom) models, adapted from [10].

and the signals that were retained were spatially filtered using local common-average referencing. Signals were then low-pass filtered, resampled to 256 Hz, and time-gated from stimulus onset to one second post stimulus onset. Features were extracted from the FFT spectrogram by integrating over 12 octaves with center frequencies spaced by half-octaves beginning at 2 Hz and time offsets of 250 and 500 ms post-stimulus. Because the number of electrodes varied per subject, the number of potential ECoG features varied as well. Features were down-selected by ranking them according to their stability over stimulus presentation, which has precedence in similar studies [4, 5, 6]. The stability of a particular neural feature was calculated by averaging all pairwise Pearson correlations between responses in blocks of trials. Up to 200 of the most stable neural features were considered.

The collected ECoG features and the accompanying semantic attributes for each stimulus can be used to learn an encoding or decoding model. The manner by which these models relate the neural and semantic features to one another are illustrated in Fig. 2.

Let $\boldsymbol{s}$ be a P-dimensional vector of semantic attributes, and $\boldsymbol{n}$ be an M-dimensional vector of neural features. The *encoding model* takes the form of a linear mapping of $\boldsymbol{s}$ onto each n_m, for $m = 1,2,\dots,M$:

$$\hat{n}_m = \boldsymbol{s}^T \boldsymbol{\beta}_m^{(en)} \qquad (1)$$

The parameter vector $\boldsymbol{\beta}_m^{(en)}$ consists of the regression coefficients for encoding the mth feature. In prior work,

 DOI: 10.3217/978-3-85125-533-1-78

$\boldsymbol{\beta}_m^{(en)}$ was learned using ridge regression [10], and the same is done here since the model output $\hat{n}_m$ is continuous-valued. The ridge regression solution for $\boldsymbol{\beta}_m^{(en)}$ is given by

$$\boldsymbol{\beta}_m^{(en)} = \left(\boldsymbol{S}^T\boldsymbol{S} + \lambda^{(en)}\boldsymbol{I}\right)^{-1}\boldsymbol{S}^T\boldsymbol{n}_m, \quad (2)$$

where $\boldsymbol{S}$ is the $T \times P$ matrix of semantic attributes, where T is the total number of trials used to train the model, $\boldsymbol{n}_m$ is the $T \times 1$ vector of values of the mth neural feature (normalized to zero-mean, unit-variance), and $\lambda^{(en)}$ is a regularization parameter determined empirically by grid search amongst five values between 1 and 10.

Conversely, the *decoding model* takes the form of a linear mapping of $\boldsymbol{n}$ onto each s_p, for $p = 1,2,\dots,P$:

$$\hat{s}_p = \boldsymbol{n}^T\boldsymbol{\beta}_p^{(de)}, \quad (3)$$

The parameter vector $\boldsymbol{\beta}_p^{(de)}$ consists of the regression coefficients for decoding the pth attribute. The ridge regression solution for $\boldsymbol{\beta}_p^{(de)}$ is given by

$$\boldsymbol{\beta}_p^{(de)} = \left(\boldsymbol{N}^T\boldsymbol{N} + \lambda^{(de)}\boldsymbol{I}\right)^{-1}\boldsymbol{N}^T\boldsymbol{s}_p, \quad (4)$$

where $\boldsymbol{N}$ is the $T \times M$ matrix of neural features, $\boldsymbol{s}_p$ is the $T \times 1$ vector of values of the pth attribute, and $\lambda^{(de)}$ is a regularization parameter determined empirically by grid search amongst five values between 100 and 1000.

The discrete-valued output $\hat{s}_p$ suggests that a classifier may be more appropriate than regression for learning the decoding model. As suggested by [8], a linear SVM is also used to learn $\boldsymbol{\beta}_m^{(de)}$. For a binary problem where $s_p \in [-1,1]$, the SVM solves the following optimization problem, which maximizes the margin between the classes:

$$\boldsymbol{\beta}_p^{(de)} = \arg\min_{\boldsymbol{\beta}}\left\{c\sum_{t=1}^{T}\max\left[0,1 - s_p(\boldsymbol{\beta}^T\boldsymbol{n}_t)\right]^2 + \|\boldsymbol{\beta}\|^2\right\} \quad (5)$$

To train the SVM, the elements of $\boldsymbol{s}$ were re-quantized to $[-1,0,1]$ by combining the [-0.5, 0, 0.5] responses. Re-quantization casts the original attribute values as simple answers: *no*, *don't know*, and *yes*. The `liblinear` software package was used to train three *one-versus-one* SVM classifiers to discriminate each pair of values [16]. Tuning parameters $c = \{1, 10, 100\}$ were considered, with per-class weighting according to the number of training samples. Voting amongst the three classifiers is used at test to predict $\hat{s}_p$. To assess the effect of modifying the attributes in this manner, ridge regression was also applied to the re-quantized attributes in another version of the decoding model.

After learning the mapping between neural features and attributes, a novel stimulus can be decoded by a distance-based classifier in neural space (if an encoding model was used) or semantic space (if a decoding model was used). Let the cosine distances resulting from the encoder output be denoted as

$$d_\phi^{(en)} = \frac{\hat{\boldsymbol{n}} \cdot \boldsymbol{n}_\phi}{\|\hat{\boldsymbol{n}}\| \cdot \|\boldsymbol{n}_\phi\|}, \quad (6)$$

where $\boldsymbol{n}_\phi$ is the output of the trained encoder applied to $\boldsymbol{s}_\phi$, the true attribute vector for noun ϕ, and $\hat{\boldsymbol{n}} = [\hat{n}_1, \hat{n}_2, \dots, \hat{n}_M]^T$. Therefore, neural decoding by means of an encoding model takes the form of

$$\hat{\phi}^{(en)} = \arg\min_{\phi}\left\{d_\phi^{(en)}\right\}. \quad (7)$$

Similarly, neural decoding by means of a decoding model takes the form of

$$d_\phi^{(de)} = \frac{\hat{\boldsymbol{s}} \cdot \boldsymbol{s}_\phi}{\|\hat{\boldsymbol{s}}\| \cdot \|\boldsymbol{s}_\phi\|}, \quad (8)$$

$$\hat{\phi}^{(de)} = \arg\min_{\phi}\left\{d_\phi^{(de)}\right\}, \quad (9)$$

where $\boldsymbol{s}_\phi$ is the decoder output and $\hat{\boldsymbol{s}} = [\hat{s}_1, \hat{s}_2, \dots, \hat{s}_P]^T$.

RESULTS

Experiments were conducted to assess performance of zero-shot stimulus prediction using four different modeling approaches: *Ridge Encoder*, *Ridge Decoder*, *Ridge Decoder Re-Quantized Attributes*, and *SVM Decoder with Re-Quantized Attributes*. The zero-shot problem was simulated by employing leave-one-noun-out cross-validation; feature selection and training were performed using 59 of the 60 nouns, and one noun was held out for testing. Therefore, the number of trials used to train the models was $T = 6 \times 59 = 354$ per subject. Two options for testing were compared: predicting from the average ECoG feature vector over all 6 trials, and predicting from single trials. Performance was measured via the *mean rank accuracy* (MRA). The MRA represents the average rank accuracy (RA) of the zero-shot test class, taken across the full set of 60 nouns ranked according to the cosine distance,

$$MRA = \frac{1}{60}\sum_{\phi=1}^{60} RA_\phi, \quad (10)$$

where RA_ϕ is the relative (percentage) rank of the test noun ϕ within a ranked list of potential classes,

$$RA_\phi = 100 \times \left(\frac{60 - r_\phi}{59}\right), \quad (11)$$

and r_ϕ is the rank of $d_\phi^{(en)}$ (if an encoding model was used) or $d_\phi^{(de)}$ (if a decoding model was used). The MRA could also be calculated on a per-category basis by averaging the MRA of all nouns within the same category.

The per-noun and per-category MRAs were tested for significance using a Monte Carlo procedure. A total of 1,000 null encoding and decoding models were trained for each subject by permuting the rows of $\boldsymbol{S}$, and the maximum MRA was calculated over all choices of M and $\lambda^{(en)}$ or $\lambda^{(de)}$. The p-values for the MRAs achieved by the alternative models were then computed using the distribution of the MRAs achieved by the null models.

The observed *per-noun* MRAs of the four decoding approaches are summarized in Tab. 2 and Tab. 3. The reported values represent maximum performance over all numbers of neural features and choices of regularization/tuning parameters that were considered. Tab. 2 summarizes the performance for decoding block-averaged neural responses, and Tab. 3 summarizes the performance of decoding single-trial neural responses.

The MRA for block-averaged neural features were

 DOI: 10.3217/978-3-85125-533-1-78

Table 2: *Per-noun* MRA using *block-averaged* neural features at test. Boldface indicates significance at $p < 0.01$, italics indicates significance at $p < 0.05$, and the highest performance per subject is underlined.

	Ridge Encoder	Ridge Decoder	Ridge Decoder w/ Requant	SVM Decoder w/ Requant
S1	***84.11***	***79.19***	***80.08***	***75.25***
S2	***83.92***	***79.75***	***78.33***	***77.33***
S3	*65.64*	*62.86*	*62.47*	60.79
S4	***66.69***	***65.69***	***64.47***	*62.81*
S5	***68.17***	***67.42***	***67.58***	***65.93***
S6	***67.14***	***70.81***	***70.11***	*64.19*
S7	***67.31***	***69.67***	***69.47***	*64.07*
S8	***75.61***	***73.00***	***72.36***	***71.03***
S9	***87.69***	***82.75***	***81.75***	***80.58***

Table 3: *Per-noun* MRA using *single-trial* neural features at test. Boldface indicates significance at $p < 0.01$, italics indicates significance at $p < 0.05$, and the highest performance per subject is underlined.

	Ridge Encoder	Ridge Decoder	Ridge Decoder w/ Requant	SVM Decoder w/ Requant
S1	***74.80***	***71.07***	***71.17***	***64.05***
S2	***75.70***	***72.65***	***71.81***	***70.82***
S3	*57.55*	*56.31*	*55.77*	54.99
S4	***59.15***	***57.86***	***57.13***	*56.80*
S5	***60.13***	***60.62***	***60.58***	***57.75***
S6	***59.13***	***61.74***	***61.59***	57.15
S7	***65.07***	***66.19***	***65.38***	*60.16*
S8	***64.15***	***62.28***	***61.81***	***61.18***
S9	***80.63***	***76.65***	***75.56***	***73.09***

Table 4: *Per-category* MRA using *block-averaged* neural features at test. Boldface indicates significance at $p < 0.01$, italics indicates significance at $p < 0.05$, and the highest performance per subject is underlined.

	Ridge Encoder	Ridge Decoder	Ridge Decoder w/ Requant	SVM Decoder w/ Requant
S1	***80.54***	***79.17***	***79.86***	***73.87***
S2	***82.19***	***79.63***	***79.19***	***78.31***
S3	***62.99***	***62.99***	***62.83***	***61.33***
S4	***65.55***	***65.12***	***64.42***	***63.18***
S5	***64.77***	***67.01***	***66.09***	***67.03***
S6	***66.52***	***70.63***	***70.46***	***64.79***
S7	***68.79***	***70.13***	***69.96***	***65.49***
S8	***70.18***	***73.34***	***73.36***	***70.92***
S9	***82.23***	***81.99***	***81.22***	***80.59***

Table 5: *Per-category* MRA using *single-trial* neural features at test. Boldface indicates significance at $p < 0.01$, italics indicates significance at $p < 0.05$, and the highest performance per subject is underlined.

	Ridge Encoder	Ridge Decoder	Ridge Decoder w/ Requant	SVM Decoder w/ Requant
S1	***72.39***	***70.63***	***70.77***	***63.37***
S2	***74.16***	***72.65***	***72.31***	***70.41***
S3	***56.52***	***56.00***	***55.85***	***55.34***
S4	***58.56***	***58.01***	***57.81***	***57.16***
S5	***58.50***	***59.48***	***59.25***	***57.55***
S6	***58.87***	***61.79***	***61.78***	***57.77***
S7	***66.47***	***66.55***	***66.01***	***60.63***
S8	***62.11***	***62.37***	***62.04***	***60.97***
S9	***76.48***	***75.84***	***75.34***	***72.63***

higher than single-trial MRA because averaging repeated trials mitigates noise. Performance varied within 5% MRA for most subjects. For all but one subject, S2, the Ridge Encoder/Decoder MRA was significant at $p < 0.01$. For most of the subjects (S1-S4, S8, and S9 for block-average and single trial decoding, S5 for block-average decoding only) the Ridge Encoder also yielded the highest MRA and is consistent with performance in similar fMRI studies, e.g. [4]. For the other subjects (S5 for single-trial decoding, S6 and S7 for both types of decoding), the Ridge Decoder was slightly better and re-quantizing the attributes did not significantly affect performance.

The *per-category* MRA, for both block-averaged and single-trial neural features at test, is summarized in Tab. 4 and Tab. 5, respectively. All per-category MRAs were significant at $p < 0.01$. For each subject, the highest per-category MRA tended to be no more than 5% less than the highest per-noun MRA. For three subjects, the Ridge Decoder performed best, and the Ridge and SVM Decoders with Re-Quantized Attributes were each best for one subject.

The RA of each noun was analyzed by comparing the results of the best-performing (S1) and worst-performing (S3) subjects. Those results are illustrated in Fig. 3 and Fig. 4, respectively. The nouns are listed in descending order of RA by the Ridge Encoder. Note that the performance of the three decoder models does not follow the same trend as the performance of the encoder model. In fact, some of the nouns that are decoded poorly by the Ridge Encoder (e.g., *hand* and *foot* for S1, *fly* and *bed* for S3) are actually decoded with much higher RA by the three decoder models.

DISCUSSION

For a majority of subjects, zero-shot stimulus prediction via an encoding model was superior to decoding models for ECoG signals recorded primarily from temporal and basal occipital regions. The difference in MRA between the encoding model and the best decoding model was within 5%. A possible explanation may be that the encoding model is more robust. The semantic attributes used to fit the encoding model are deterministic, while the ECoG features used to fit the decoding model are noisy.

The SVM Decoder was the worst performing model for most subjects and yielded less significant MRAs, suggesting that it makes several incorrect assumptions about the decoding problem. One might be the training

 DOI: 10.3217/978-3-85125-533-1-78

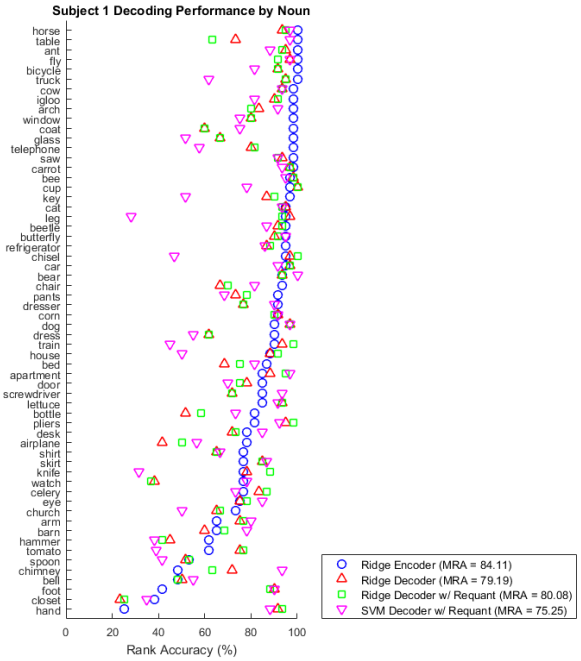

Figure 3: Per-noun RA using block-averaged neural features for S1 (best performer). Nouns are listed in order of descending RA by the Ridge Encoder.

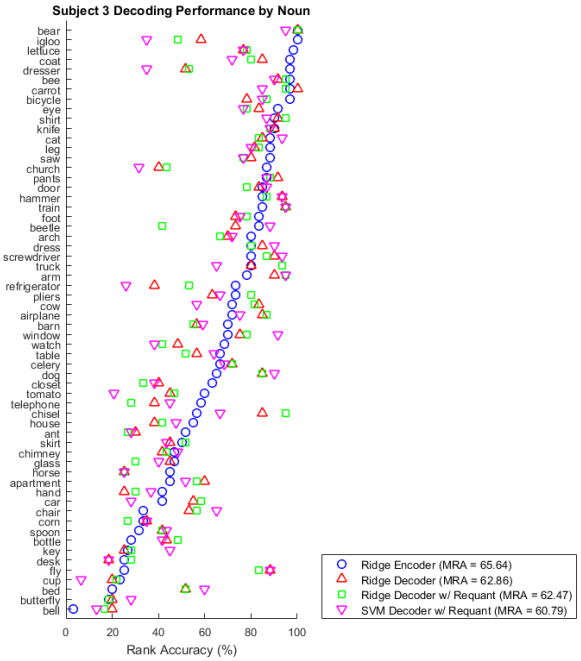

Figure 4: Per-noun RA using block-averaged neural features for S3 (worst performer). Nouns are listed in order of descending RA by the Ridge Encoder.

set size. In [8], the SVM was applied to visual features extracted from a set of 75,489 images belonging to 57 classes – a much larger data set than what was considered in this study. It is possible that the smaller training set may be bolstered by use of a kernel function, but this introduces another tuning parameter which would need to be optimized for each attribute to avoid overfitting.

The SVM also assumes the training set is balanced between classes, so that maximizing the margin minimizes the classification error. However, of the 218 attributes, only 50 had reasonable balance between the re-quantized classes $s = -1, s = 0,$ and $s = 1$. We attempted to soften this assumption by weighting the SVM cost parameter (c) proportionally to the size of each class, but the effect was negligible. One could soften the balanced-class assumption further by optimizing c for each attribute, but that was not explored in this study.

Several aspects of this work suggest that some form of a semantic BCI may be viable. First, we demonstrated that the semantic zero-shot learning approach to semantic decoding can be fruitfully applied to ECoG using encoding or decoding models. As pointed out in [5], "It is intractable to collect neural training images for every possible word in English, so to build a practical neural decoder we must have a way to extrapolate to recognizing words beyond those in the training set." Our study only focused on the ability to classify among 60 zero-shot nouns, and the problem will become more difficult as the scale increases to more classes, especially when the attribute-to-noun mapping is ambiguous or completely unknown. In such cases, simply outputting the most highly-weighted semantic attributes may still be useful for communicating a novel word or concept.

One issue that would need to be addressed is collection of adequate training data, which can be time consuming and costly. But efforts in fMRI demonstrating how voxel-wise models can be built from large datasets of activity elicited by natural stimuli, like movies and stories, suggests these might be more economical strategies for collecting training data [17, 18]. Furthermore, efforts in large-scale pattern classification using semantic hierarchies to trade specificity for accuracy suggest a potential avenue towards robust classification of a wide variety of novel classes [19].

Another obstacle that must be overcome in developing a practical semantic BCI is consistency in real-time decoding performance. While the highest MRAs were achieved when averaging across multiple trials, there was a modest drop (< 10%) in performance when decoding nouns or categories from single trials. In addition, our encoding model did not account for temporal variability in semantic processing, as our features were extracted from fixed time windows post stimulus onset. However, the superior performance observed for the encoding model implies that accurate real-time decoding may be possible by cross-correlating a recorded neural signal against a pre-computed lookup table of signals predicted from various combinations of attributes.

CONCLUSION

Four approaches to zero-shot stimulus prediction were compared for predicting recognized objects from ECoG signals evoked during a picture-naming task. All four approaches attempt to learn a mapping between neural

 DOI: 10.3217/978-3-85125-533-1-78

features and semantic attributes, but differ in what they consider to be the direction of the mapping and in how the mapping is learned from training data. Performance was relatively consistent from subject to subject between the four approaches, though in most cases the Ridge Regression Encoding model yielded the best performance. These results represent an initial step toward realizing semantic BCI, and suggest that the next generation of neuroimaging technologies paired with the algorithms demonstrated here could help new BCI applications to come to fruition.

REFERENCES

[1] Anderson NR, Blakely T, Schalk G, Leuthardt EC, Moran DW. Electrocorticographic (ECoG) correlates of human arm movements. Experimental Brain Research 2012; 223(1): 1-10

[2] Martin D, Brunner P, Holdgraf C, Heinze HH, Crone NE, Rieger J, et al. Decoding spectrotemporal features of overt and covert speech from the human cortex. Frontiers in Neuroengineering 2014; 7(14)

[3] Rosch E. Principles of Categorization. In: Cognition and Categorization, Erlbaum Associates, Hillsdale NJ, 1978, pp. 27-48

[4] Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave V, Manson R, et al. Predicting Human Brain Activity Associated with the Meanings of Nouns. Science 2008; 320(5880): 1191-1195

[5] Palatucci M, Pomerleau D, Hinton GE, Mitchell TM, Zero-shot learning with semantic output codes, in Proc. NIPS, Vancouver, Canada, 2009.

[6] Sudre G, Pomerleau D, Palatucci M, Wehbe L, Fyshe A, Salmelin R, et al. Tracking neural coding of perceptual and semantic features of concrete nouns. NeuroImage. 2012; 62(1): 451-463

[7] Pereira F, Botvinick M, Detre G. Using Wikipedia to learn semantic feature representations of concrete concepts in neuroimaging experiments. Artificial Intelligence 2013; 194: 240-252

[8] Burlina P, Schmidt AC, Wang I-J. Zero shot deep learning from semantic attributes, in Proc. ICMLA, Miami, USA, 2015.

[9] Colwell KA, Collins LM. Attribute-driven transfer learning for detecting novel buried threats with ground-penetrating radar, in Proc. SPIE Defense + Security, Baltimore, USA, 2016

[10] Rupp K, Roos M, Milsap G, Caceres C, Ratto C, Chevillet M, et al. Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. NeuroImage 2017; 148: 318-329

[11] Fazel-Rezai R, Allison BZ, Guger C, Sellers EW, Kleih SC, Kubler A. P300 brain computer interface: current challenges and emerging trends, Frontiers in Neuroengineering 2012; 5(14)

[12] Sajda P, Pohlmeyer E, Wang J, Parra L, Christoforou C, Dmochoski J, et al. In a blink of an eye and a switch of a transistor: cortically coupled computer vision. Proc. IEEE 2010; 98(3): 462-478

[13] Khosla D, Bhattacharyya R, Tasinga P, Huber DJ. Optimal detection of objects in images and videos using electroencephalography (EEG), in Proc. SPIE Defense, Security, and Sensing, Orlando, USA, 2011

[14] Vansteensel M, Pels EGM, Bleichner MG, Branco MP, Denison T, Freudenburg ZV, et al. Fully Implanted Brain-Computer Interface in a Locked-In Patient with ALS, N Engl J Med 2016; 375: 2060-2066

[15] Quitadamo LR, Cavrini F, Sbernini L, Riillo F, Bianchi L, Seri S, et al. Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review. Journal of Neural Engineering 2017; 14(4)

[16] Fan R-E, Chang K-W, Hsieh C-J, Wang X-R, Lin C-J. LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research 2008; 9: 871-1874

[17] Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 2012; 76(6): 1210-1224

[18] Wehbe L, Murphy B, Talukdar P, Fyshe A, Ramdas A, Mitchell T. Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. PLoS ONE 2014: 9(11)

[19] Deng J, Krause J, Berg A, Fei-Fei L. Hedging your bets: optimizing accuracy-specificity trade-offs in large scale visual recognition, in Proc. IEEE CVPR, Providence, USA, 2012