Neurophotonics

Neurophotonics.SPIEDigitalLibrary.org

Decoding semantic representations from functional near-infrared spectroscopy signals

Benjamin D. Zinszer
Laurie Bayet
Lauren L. Emberson
Rajeev D. S. Raizada
Richard N. Aslin

SPIE.

Benjamin D. Zinszer, Laurie Bayet, Lauren L. Emberson, Rajeev D. S. Raizada, Richard N. Aslin, "Decoding semantic representations from functional near-infrared spectroscopy signals," *Neurophoton.* **5**(1), 011003 (2017), doi: 10.1117/1.NPh.5.1.011003.

Neurophotonics 5(1), 011003 (Jan–Mar 2018)

Decoding semantic representations from functional near-infrared spectroscopy signals

Benjamin D. Zinszer,[a,b,*] Laurie Bayet,[b,c,d] Lauren L. Emberson,[e] Rajeev D. S. Raizada,[a,b] and Richard N. Aslin[b]
[a]University of Rochester, Brain and Cognitive Sciences Department, Rochester, New York, United States
[b]University of Rochester, Rochester Center for Brain Imaging, Rochester, New York, United States
[c]Boston Children's Hospital, Boston, Massachusetts, United States
[d]Harvard Medical School, Boston, Massachusetts, United States
[e]Princeton University, Psychology Department, Princeton, New Jersey, United States

Abstract. This study uses representational similarity-based neural decoding to test whether semantic information elicited by words and pictures is encoded in functional near-infrared spectroscopy (fNIRS) data. In experiment 1, subjects passively viewed eight audiovisual word and picture stimuli for 15 min. Blood oxygen levels were measured using the Hitachi ETG-4000 fNIRS system with a posterior array over the occipital lobe and a left lateral array over the temporal lobe. Each participant's response patterns were abstracted to representational similarity space and compared to the group average (excluding that subject, i.e., leave-one-out cross-validation) and to a distributional model of semantic representation. Mean accuracy for both decoding tasks significantly exceeded chance. In experiment 2, we compared three group-level models by averaging the similarity structures from sets of eight participants in each group. In these models, the posterior array was accurately decoded by the semantic model, while the lateral array was accurately decoded in the between-groups comparison. Our findings indicate that semantic representations are encoded in the fNIRS data, preserved across subjects, and decodable by an extrinsic representational model. These results are the first attempt to link the functional response pattern measured by fNIRS to higher-level representations of how words are related to each other. [DOI: 10.1117/1.NPh.5.1.011003]

Keywords: functional near-infrared spectroscopy; neural decoding; multivariate pattern analysis; semantic model.

Paper 17090SSR received Apr. 11, 2017; accepted for publication Jul. 31, 2017; published online Aug. 23, 2017.

1 Introduction

Multivariate statistical analyses for neuroimaging data have provided a new window into the contents of neurocognitive representations. Using multivariate pattern analysis (MVPA), cognitive neuroscientists can decode functional brain responses into signals of theoretical significance, such as representational models of word meaning or visual perception. This decoding is achieved by comparing the distributed neural response (that is, the multivariate pattern) to an analogous theoretical model. Such an approach was first used for decoding the meanings of words by Mitchell et al.,[1] who used a semantic model based on word co-occurrence frequencies to decode the meanings of concrete nouns from adult subjects using functional magnetic resonance imaging (fMRI). MVPA studies have demonstrated that neural activity encoding task-relevant information is both broadly distributed across the cortex and decodable into distinct, localizable components.[2–4] Thus, recent fMRI research has confirmed the utility of separable components of the neural response to classes of stimuli, contrasting representational models derived from computer vision, distributional (corpus-based) semantics, and conceptual features.[1,5–7]

The adaptation of multivariate techniques, such as neural decoding, to other imaging modalities is critical for extending these advances in cognitive neuroscience beyond the MRI scanner and enabling research with participants who are ineligible or not well-suited for fMRI studies (e.g., infants, various clinical populations) as well as tasks that are simply not possible to investigate in the MR environment. In particular, infants usually cannot be scanned while awake, with very few exceptions,[8–10] limiting our ability to investigate the emergence of language in the brain during the most important developmental years. Brain imaging with functional near-infrared spectroscopy (fNIRS) is one especially promising approach for developmental neuroscience research. fNIRS has gained popularity for its advantages in testing infants and children, due to its tolerance for head and body motion and its applicability outside the confines of the MR environment. These unique advantages have also enabled research on social cooperation and face-to-face conversation.[11–13]

However, to date, the application of MVPA to fNIRS remains limited. One recent paper[14] described the first demonstration of neural decoding in infants using fNIRS, based on representational similarity techniques developed for fMRI.[15,16] This work opens the door for using multivariate methods in developmental research, such as measuring semantic representations for words from the earliest stages of language acquisition to the adult-like state. Such fine-grained studies of language are now commonplace in fMRI but completely untested with fNIRS.

Although fNIRS measures a very similar physiological signal to fMRI, these techniques have important differences that affect the amount of information they provide for neural

*Address all correspondence to: Benjamin D. Zinszer, E-mail: bzinszer@gmail.com

2329-423X/2017/$25.00 © 2017 SPIE

decoding. Most notably, the spatial resolution of fNIRS is considerably lower than fMRI, with individual channels covering only the surface of the cortex and sampling from regions a few centimeters in diameter. Further, fNIRS is sensitive to hemodynamic variations in the scalp and other tissues between the detector and the cortex. Some of this noncortical signal can be removed with statistical techniques (such as principal components analysis) to locate and subtract global responses, such as heartbeat or fluctuations in blood pressure. The relatively coarser spatial resolution (i.e., portion of the cortex covered by each measurement channel) and poorer signal-to-noise ratio of fNIRS compared to fMRI likely diminish the power of fNIRS for detecting the small changes in signal that would underlie comparisons among several semantic categories.

However, other aspects of fNIRS are very well-suited to multivariate pattern analyses. In particular, the fNIRS image is spatially specific. Although a single channel may cover a few square centimeters of the scalp, we can be highly confident that the hemodynamic response measured by that channel originates in that spatial area and is not susceptible to the volume conduction effects that apply to most electrophysiological measures, which have also been used in adult neural decoding.[17] Although magnetoencephalography (MEG) has provided improvements in combined spatial and temporal resolution relative to electroencephalography and fMRI, respectively,[18,19] it remains very difficult to apply MEG to infants and young children, especially in visual or audiovisual paradigms. Further, fNIRS devices typically sample at a higher rate (10 Hz or more) than fMRI, providing many more observations of an individual event response curve than available from fMRI (i.e., in a 2-s time window, 20 fNIRS samples are obtained but typically only 1 fMRI sample). Thus, on the one hand, the spatial limitations of fNIRS may limit how much information is encoded in its relatively low-dimensional data (i.e., a few dozen channels versus thousands of voxels). On the other hand, fNIRS provides both a large amount of data about each channel and maintains spatial specificity.

In this study, we ask whether fNIRS data contain information suitable for neural decoding among subjects and for decoding based on extrinsic representational models. We present subjects with eight different matching audiovisual pairs (i.e., a spoken word and its corresponding picture) and attempt to identify specific stimulus representations based either on generalization from other subjects' neural responses (between-subjects decoding) or on a distributional semantic model of the meanings of the stimulus words (semantic decoding). To achieve these goals, we integrate recently developed tools for representational similarity-based decoding in fMRI[16,20] and similarity-based MVPA for fNIRS.[14] We thus reveal the representational contents of the fNIRS signal encoded in multichannel response patterns, and we aim to motivate follow-up studies that can apply these methods to language research in infants and young children.

2 Experiment 1

2.1 *Method*

2.1.1 *Participants*

Eight adults (three males and five females) participated in a 15-min passive viewing and listening task. Participants were students and staff recruited from the Brain and Cognitive Sciences Department at the University of Rochester.

2.1.2 *Procedure*

Informed consent procedures and experimental methods were approved by the institutional review board of the University of Rochester. Participants were presented with eight audiovisual stimuli, featuring a photograph of an object and simultaneous auditory presentation of the object's name. Participants were asked to simply focus on the audiovisual stimuli and to think about the meaning of that stimulus or any memory it evoked. We directed the adult participants toward this more ecological activity (as compared to covert feature generation used in previous studies)[1] in order to best compare with children's responses to the stimuli in a future experiment.

The stimuli were drawn from two broad categories (animals and body parts) and were all objects that would be familiar to infants, as one aim of this study is to validate the design for infants in future work. The objects were: bunny, bear, kitty, dog, mouth, foot, hand, and nose. Visual stimulus presentation lasted for 3 s, with the auditory word presented immediately at onset. A jittered 6- to 9-s interstimulus interval followed each trial.[21] The interstimulus interval was composed of fireworks and a short musical clip, designed for infant paradigms and included for consistency between the adult data and future infant experiments. See Fig. 1 for illustration. This procedure was repeated over 12 blocks, with stimulus order randomized in each block. Participants' blood oxygen levels were measured using the Hitachi ETG-4000 fNIRS system throughout the exposure.

2.1.3 *fNIRS measurement and preprocessing*

The fNIRS probes were arranged in two arrays: 24 channels in a posterior 4×4 array approximately covering the occipital lobe and 18 channels in a lateral 3×5 array over the left temporal, parietal, and prefrontal lobes. (The lateral array typically provides 22 channels, but 4 of the anterior channels were excluded due to a broken laser.) Because the representational similarity-based analysis makes all of its channel-wise comparisons within-subject, only the overall positions of the arrays needed to be controlled across the subjects, allowing for variations in channel positions due to head size and shape. The lateral array was positioned directly above the left ear, approximately centered (anterior-to-posterior) over the ear and with the most anterior channel just beyond the hairline. The posterior array was centered on the back of the head, with the most inferior row of channels just over the inion.

Preprocessing of the fNIRS data was performed in Homer2.[22] We computed a principal components analysis on the optical density data, removing the first component to reduce nonneural physiological signals. Next, the data were bandpass filtered (high pass: 0.01 Hz and low pass: 1.0 Hz). Oxygenated and deoxygenated hemoglobin concentrations were computed by Homer2 according to the modified Beer–Lambert law. The oxygenated hemoglobin concentration data were further processed using custom scripts to remove motion artifacts by masking a 1-s window around any observations that exceeded five standard deviations of the overall mean in that channel.

Finally, we performed a channel stability analysis adapted from fMRI decoding (voxel stability)[1] to determine which channels produced reliable responses across multiple blocks, independent of the discriminability among the stimulus classes. This procedure reduces the dimensionality of the fNIRS data and is intended to increase the signal-to-noise ratio by including channels that respond reliably while excluding channels that

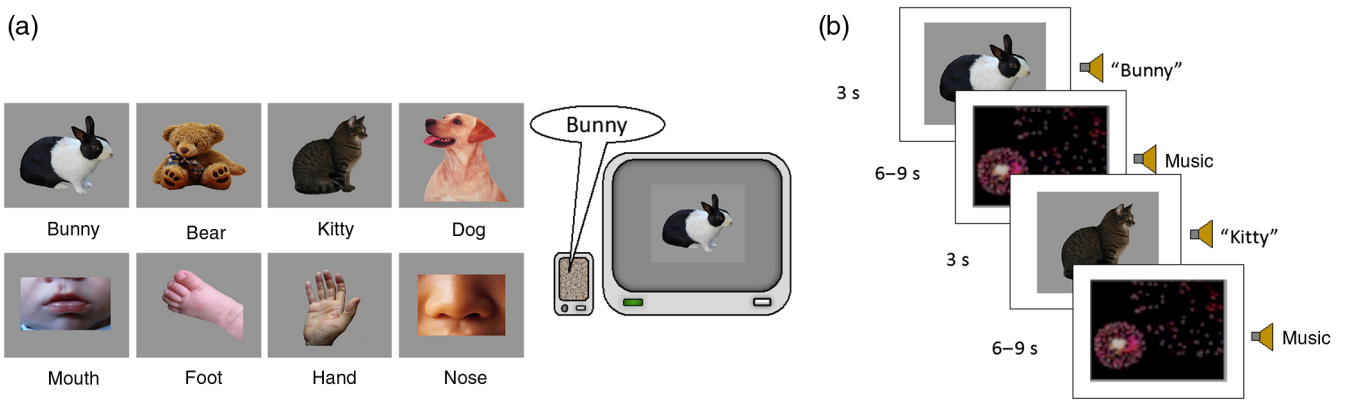

Fig. 1 Schematic of experimental procedures. (a) Eight visual stimuli depicting common objects were presented to participants for 3 s at a time, with immediate-onset auditory presentation of the object's name. (b) Each block consisted of the eight randomly ordered stimuli, followed by a jittered 6- to 9-s interstimulus interval.

contain noise or change their response magnitude over the course of the experiment. For a given channel, the responses for each stimulus type are correlated across blocks. Thus, each block is represented by an n-dimension vector for n stimulus types, and the correlation between block 1 and block 2 is the Pearson r statistic between these two n-dimension vectors. These correlations are repeated for every possible pair of blocks, and the r values are averaged to produce a mean stability value for that channel (further details are provided in the supplementary materials of Ref. 1).

Importantly, the channel stability procedure is independent of the differences between experimental conditions and thus does not need to be performed within each cross-validation loop, i.e., it does not lead to double dipping. For each subject, all channels are assigned stability scores, and the 50% most stable channels are retained (Fig. 2). The 50th percentile stability threshold was applied separately to each subject, so the specific set of channels is not constant across subjects. However, the representational similarity-based decoding procedures described below summarize fNIRS responses at the subject level. This abstraction away from channel-specific data distinguishes similarity-based decoding from other popular methods, and it allows comparisons among the subjects' overall patterns of fNIRS response even when individual channels may differ in quality or location among individual subjects.

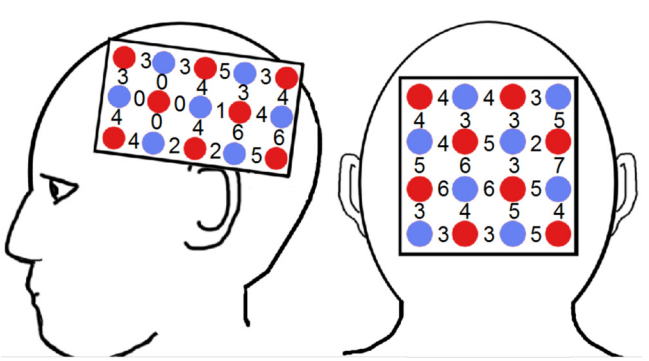

Fig. 2 Approximate positions of fNIRS probes on the left lateral and posterior areas of the head. Red circles indicate lasers and blue circles indicate detectors. Each channel (between a laser and detector) is marked by the number of participants (out of eight, total) for whom it was retained as one of the top 50% most stable channels, independent of decoding accuracy.

2.1.4 *Semantic model*

We selected the compositional operations in semantic space model (COMPOSES), a well-known distributional semantic model,[23] to estimate semantic representations of the eight stimuli. The COMPOSES model describes the meanings of the stimulus words based on an analysis of the words' use in large text corpora, primarily by measuring how often different words co-occur with each other. COMPOSES produces a 400-dimension vector representation of each word. While these representations are unique to the model, we applied representational similarity methods to abstract this model and the fNIRS data from their respective sources into a shared similarity space. That is, each of the eight COMPOSES vectors (one for each experimental stimulus) was correlated with the other seven COMPOSES vectors to yield an 8×8 similarity (cross correlation) matrix of Pearson's correlation coefficients representing how similar the eight stimuli were to each other, according to that model [Fig. 3(c)]. An analogous 8×8 similarity matrix was derived for the fNIRS data (according to a method described below). This procedure allows the model and the fNIRS data to be directly compared in terms of their representational similarity structures [as captured by the 8×8 matrices, see Fig. 3(c)].

2.1.5 *Analyses*

Changes in oxygenated hemoglobin levels following each stimulus were epoched, baseline corrected, and averaged according to Emberson et al.'s[14] MVPA procedures for fNIRS. Epoching was performed in a time window of 6.5 to 9.0 s after stimulus onset, following the optimal onset time identified for integrated audiovisual events,[14] and terminating before the earliest onset of the next stimulus. Epoched and baseline corrected time series data for trials of each stimulus type were averaged across blocks to yield a mean response curve. The mean oxygenated hemoglobin level for this response curve was calculated in each of the 42 channels to produce a 42-dimension response pattern for each stimulus type. Each subject's response patterns were abstracted to representational similarity space by cross correlating the eight 42-dimension vectors into an 8×8 similarity structure.

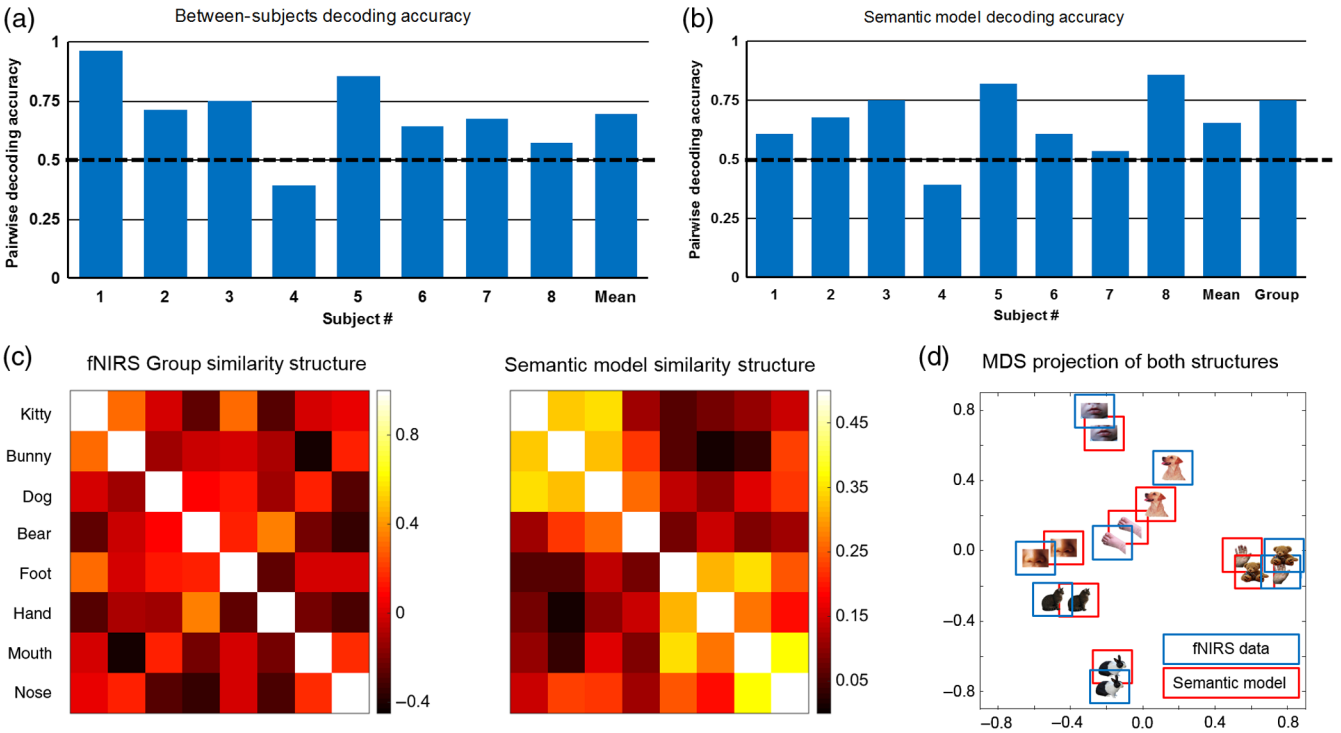

Fig. 3 (a) Between-subject decoding accuracy and (b) semantic model-based decoding accuracy for each of the eight participants. Chance level is 50%, and significance test is performed on the mean decoding accuracy across subjects. Semantic model-based decoding accuracy of the group-level model is also depicted. (c) Similarity matrices for the group-level fNIRS data and semantic model. (d) Classical multidimensional scaling of group-level fNIRS and semantic model (rotated, scaled, and overlaid).

These semantic model- and fNIRS response-based representational similarity structures are compared at the subject level, decoding each subject's fNIRS data using a group average similarity structure (excluding that subject, i.e., leave-one-out cross-validation) and using the semantic model. These comparisons are calculated based on the average accuracy of all possible two-alternative (pairwise) forced choice comparisons among the eight stimulus classes, as typically done in fMRI neural decoding studies.[20] Decoding accuracy provides a measure of the discriminability of the classes and the reliability of the similarity structures that support decoding across subjects or between the subjects and the model.

Figure 3(d) shows the strong similarities between a group-level fNIRS model (in blue) and the semantic model (in red) of the eight stimuli when these two structures are aligned in multidimensional space. The figure is generated using classical multidimensional scaling (MDS, MATLAB® function cmdscale) to translate each 8×8 similarity matrix [from Fig. 3(c)] into distances between each stimulus in seven-dimensional space. After each of these models is independently scaled to best capture the relative similarities of the eight stimuli to one another (within the model), the two models' structures can be compared for overall alignment. In this illustration, the semantic model is rotated through MDS space and scaled using Procrustes alignment to find the closest possible alignment with the neural data. The first two dimensions of these aligned structures are shown in Fig. 3(d). MATLAB® scripts and sample data necessary to reproduce these analyses are provided on our Github page.[24]

Importantly, this visualization illustrates the best possible alignment between the observed fNIRS and model-based structures when the correct matches for all eight classes are known and considered simultaneously. The pairwise decoding procedures rely only on the similarity data as shown in Fig. 3(c) and compare only two stimuli at a time without access to this structure-level alignment. Consequently, pairwise decoding is weaker than if this higher dimensional structural alignment was already known.

2.2 *Results*

2.2.1 *Between-subjects fNIRS response-based decoding*

Between-subjects decoding is performed with a leave-one-subject-out cross-validation procedure, wherein each subject is iteratively removed from the group and compared to the averaged group model. The group model is computed as the element-wise mean of the remaining subjects' representational similarity structures (yielding an 8×8 group matrix). The average pairwise decoding accuracy in this between-subjects analysis was 0.70 [$p = 0.01$, Fig. 3(a)], compared to chance level of 0.50. All significance tests are based on the empirical null distribution for randomly permuted stimulus labels.[14,16,20] Null distributions for each significance test are provided with the demonstration code and can be regenerated using the code available on Github. p-values are obtained by measuring the proportion of the null distribution that is greater than or equal to the observed accuracy.

Individual level performance was variable, ranging from 0.39 to 0.96. This variability suggests that either the individual subjects' data quality was inconsistent or some subjects' response patterns differed from the group. In the next analysis, we ask

whether the COMPOSES semantic model explains the subjects' fNIRS response patterns.

2.2.2 *Semantic model-based decoding*

The COMPOSES semantic model decoded the fNIRS data from the stable channels with a mean accuracy of 0.66 ($p = 0.03$) across subjects. Figure 3(b) shows individual subjects' decoding accuracy based on the COMPOSES model. We also investigated whether this result could be improved by averaging across the eight subjects' similarity structures to test the group model's semantic decoding. The semantic model decoded the group model data with an accuracy of 0.75, a considerable gain over the mean individual level performance, although this value did not achieve conventional significance when tested as a single-point observation. Instead, we return to this group model in the next experiment.

Finally, we performed some additional analyses that explore these results in greater depth, beyond the scope of this paper. We contrast the roles of within-category (e.g., foot versus nose) and between-category (e.g., foot versus kitty) comparisons to decoding accuracy, and we compare the performance of a second distributional semantic model to the COMPOSES model reported here. These supplemental materials are available for download on our Github page.

2.3 *Discussion*

Experiment 1 illustrates neural decoding of fNIRS data using a representational similarity-based method. The similarity structures used by this method allowed us to compare across different subjects' fNIRS data, to compare individual subjects' fNIRS data to a semantic model, and to combine subjects' patterns into a group model, despite individual variation in channel location and data quality. All of these comparisons are made possible by the fact that similarity-based multivariate pattern analyses abstract the individual's neural data out of their unique anatomical space and into similarity space.

The high between-subjects decoding accuracy illustrates the effectiveness of using these similarity structures to overcome the limitations of fNIRS imaging and to find reliable stimulus-dependent response patterns that are preserved among individual subjects. This finding indicates that a shared set of representations for the eight stimuli is encoded in the fNIRS data for these subjects, and the across-subject reliability is sufficient to predict the stimulus labels of a given subject based on a group model of the other seven subjects.

A significant portion of these fNIRS patterns was also explained by specific properties of the stimulus, as indicated by the success of the semantic model at decoding both individual subjects' neural response patterns and decoding the group-level model. In the next experiment, we ask whether this group-level fNIRS response pattern can be replicated with additional groups of participants, whether these groups can also be explained by the semantic model, and whether the two fNIRS arrays may carry different types of information.

3 Experiment 2

In the foregoing experiment, we demonstrated that fNIRS data contained neural response patterns that were (1) preserved across subjects, (2) decodable by a semantic model, and (3) can be meaningfully combined across subjects to improve signal-to-noise ratio in a group-level model. In this experiment, we build upon these findings to further test the reliability of the group-level fNIRS response patterns by using between-groups decoding (analogous to between-subjects decoding) and by applying the semantic model to these group models.

Further, we explore the possibility that different types of information are encoded in different fNIRS channels. Although exact channel placement varies across participants, the similarity-based metrics allow us to compare on a larger scale: array-level response patterns should be strongly preserved among subjects because the arrays were consistently placed over comparable anatomical regions and contain several channels each from which to perform decoding analyses.

3.1 *Method*

3.1.1 *Participants*

Two additional groups of eight adults each (group 2: five males and five females and group 3: two males and six females) participated in the same 15-min passive viewing and listening task as in experiment 1. Participants in group 2 were undergraduate students from the University of Rochester Brain and Cognitive Sciences subject pool, recruited approximately four months after group 1 (the participants in experiment 1). Participants in group 3 were graduate students and staff from the Brain and Cognitive Sciences Department, the University of Rochester, recruited approximately seven months after group 1.

3.1.2 *fNIRS measurement and preprocessing*

The fNIRS probes were arranged according to the same guidelines as experiment 1, with one array covering the posterior region and a second array over the left lateral region. Preprocessing of the fNIRS data also followed the same protocols as experiment 1. In this experiment, all individual subject-level fNIRS representational similarity structures were averaged to produce new group models for group 1 (the eight participants of experiment 1), group 2, and group 3. As demonstrated in experiment 1, this averaging amplifies the fNIRS patterns shared across individual subjects while reducing the influence of individual differences or measurement error. Further, no channel stability analysis was performed for the individual arrays to maintain the number of channels (posterior: 24 and lateral: 18) used in the analysis with both arrays.

3.1.3 *Analyses*

In this experiment, we compare group-level models in the combined arrays, as well as separately for each individual array. Each group-level model was estimated by averaging the representational similarity structures for the eight subjects into a single group-level structure (8×8 matrix). We performed both between-groups decoding and semantic model-based decoding on each group following the same procedures described in experiment 1, applied to the group-level data instead of the individual subject data. Because the same tests (between-groups decoding and model-based decoding) were performed three times over the various array configurations, results of these tests were adjusted using false discovery rate (FDR).[25,26]

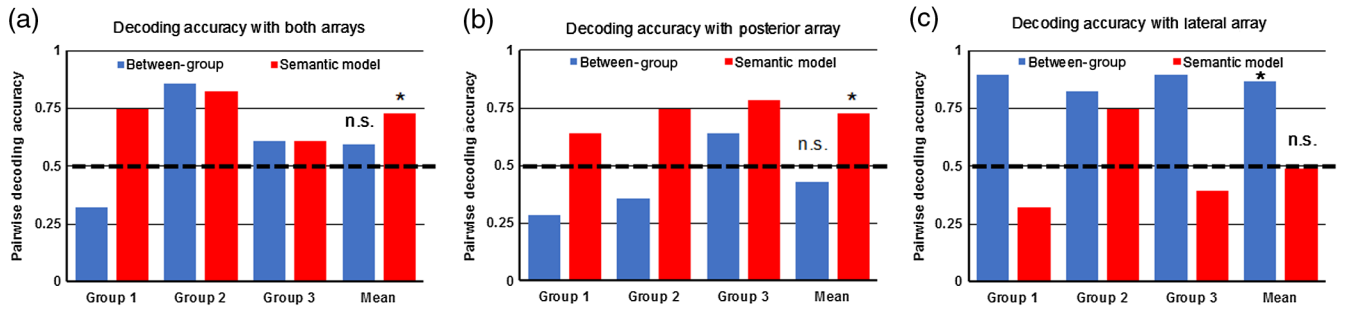

Fig. 4 Between-group and semantic model-based decoding for each of the three groups and mean performance across groups. Whole array, posterior array only, and lateral array only were tested separately to compare relative performance of each. n.s. means not significant and * means p-adj < 0.05.

3.2 *Results*

3.2.1 *Decoding with both arrays*

In the leave-one-out analysis between groups (1 versus 2 and 3, 2 versus 1 and 3, and 3 versus 1 and 2), mean decoding accuracy for the combined arrays was 0.60 (p-adj = 0.36). The semantic model, however, decoded the groups with a mean accuracy of 0.73 (p-adj = 0.045), as shown in Fig. 4(a).

3.2.2 *Posterior array decoding*

In the leave-one-out analysis between groups, mean decoding accuracy for the posterior array was 0.43 (p-adj = 0.75). The semantic model, however, decoded the groups with a mean accuracy of 0.73 (p-adj = 0.045), as shown in Fig. 4(b).

3.2.3 *Lateral array decoding*

In the leave-one-out analysis between groups, mean decoding accuracy was 0.87 (p-adj = 0.006). The semantic model did not show consistent performance, failing to decode the three groups above chance level, with a mean accuracy of 0.49 [p-adj = 0.54, see Fig. 4(c)].

3.2.4 *Group versus group decoding*

Finally, we return to the between-groups decoding task to clarify the contributions of each group to the leave-one-out analyses performed in the prior sections. We repeated the between-groups decoding for each possible pair of groups (1 versus 2, 1 versus 3, and 2 versus 3) to determine whether any one group stood out as divergent from the other two.

Table 1 shows these between-group decoding accuracies for each array. Accuracies are presented as individual observations (rather than estimates of a mean) and are thus not significance tested. All groups decoded each other with high accuracy in the lateral array, consistent with the foregoing leave-one-out analysis (Sec. 3.2.3). Results in the posterior array were more mixed, ranging from 0.21 (group 1 versus 2) to 0.64 (group 1 versus 3). We tentatively interpret these results below.

Table 1 Group versus group decoding accuracies for lateral and posterior arrays.

	Group 1	Group 2	Group 3	
Group 1	—	0.21	0.64	Posterior array
Group 2	0.82	—	0.50	
Group 3	0.79	0.75	—	
		Lateral array		

3.3 *Discussion*

This experiment aimed to replicate the strong group-level semantic decoding observed in experiment 1, where a group model successfully decoded fNIRS response patterns at the individual subject level, and compare the fNIRS response patterns across multiple groups. We used an approach analogous to the subject-level comparisons performed in experiment 1 but substituting the group-level averages. When using both the posterior and lateral arrays, between-group decoding did not significantly exceed chance, even though the average semantic decoding of each group was significant. There are some reasons that between-groups decoding may have failed in this case. First, the effective sample size of $n = 3$ sets a very high bar for statistical significance. Although this size would not work for parametric testing, cross validation can be performed and significance tested for three observations (that is, the three group-level models) using an empirical null distribution. In this case, the mean accuracy necessary for significance at $p < 0.05$ would have been 0.70, even before FDR correction.

Further, it is important to note that in leave-one-out cross-validation, the decoding accuracy of a given observation (in this case, a group) is a function of both the pattern in that observation and the average pattern from the remaining observations. Thus, it may be the case that group 1 produced fNIRS response patterns that uniquely differed from groups 2 and 3, but it is also plausible that groups 2 and 3 yielded lower-quality data themselves and in combination failed to decode group 1. We lean toward this latter explanation for a simple, practical reason: the posterior array on our fNIRS measurement device was recalibrated for another study after completion of group 1 data collection, and this recalibration appeared to have an adverse effect on the data quality. The follow-up group versus group analysis, which showed that groups 2 and 3 did not decode each other any better in either array than they performed against group 1, supports this explanation. Further, strong between-groups decoding was observed in the lateral array, which was not recalibrated among groups, in both the leave-one-out and the group versus group analyses.

Both between-groups and semantic model-based decoding were successful in individual subsets of channels (the two arrays) but not in the same subsets, highlighting a potentially important dissociation. Surprisingly, the lateral array's between-group reliability did not translate into significant above-chance decoding with the semantic model. In contrast, the posterior array was significantly decoded by the semantic model but did not decode among groups. This result raises the question of what type of information the semantic model is decoding, which is addressed at greater length in Sec. 4.

4 General Discussion

The experiments described in this study provide strong evidence for the representation of highly discriminable response patterns to multiple stimuli, encoded in about 20 channels of fNIRS data. Experiment 1 demonstrated that multivariate pattern analyses can be successfully applied to fNIRS data collected in adults to decode a multiclass, passive, and infant-friendly audiovisual paradigm. Specifically, we combined channel selection procedures with representational similarity-based models to decode stimulus categories from fNIRS signals collected over the posterior and left-lateral areas of the cortex, with relatively few repetitions of each stimulus category (12 presentations of each stimulus). Experiment 2 further provided evidence for the replicability and robustness of these findings. In particular, individual subject-level representations were successfully combined to produce group models that could be decoded with high accuracy. Experiment 2 also revealed a curious dissociation in the informational content of the posterior and lateral arrays. Signals captured by the left lateral probes (covering aspects of the left temporal cortex) discriminated among the stimuli from the three different participant groups with an average between-groups pairwise decoding accuracy of over 80%. Signals from the posterior probe (covering aspects of the occipital cortex) were decoded using the semantic model in each of the three groups with an average accuracy of over 70%.

It is unclear why a signal in the posterior array would be predictable by the semantic model but not shared among three groups of subjects that the semantic model explains. One explanation may be that the posterior array calibration errors discussed in experiment 2 (or additional measurement errors due to channel placement, hair density, or poor contact between the scalp and probes) would be doubled in any comparison of two groups' fNIRS data. Thus between-group decoding would be prohibitively difficult, while the group structures still sufficiently correlate with the (much stronger) semantic model to allow model-based decoding.

The very strong between-group decoding in the lateral array suggests that important representational information exists in this region, but this information may not be adequately explained by the present semantic model. Previous fMRI decoding research has found that both occipital and left temporal regions are significantly decoded by a distributional semantic model,[1] only the latter of which is replicated in the present study. One possibility is that semantic information in temporal cortical regions is represented at a spatial resolution smaller than the one available to fNIRS, as suggested by the high-density encoding of face-identity information in the anterior temporal lobe (in contrast to the coarse representations in fusiform gyrus) observed by Kriegeskorte et al.[27]

This latter finding, in particular, highlights the importance of exploring additional representational models to explain fNIRS data in future research. Neural decoding of printed words in fMRI has successfully combined distributional semantic data with vision-based models of objects to improve decoding accuracy.[5] An analogous approach could be applied to fNIRS to both enhance decoding accuracy and to help distinguish the contributions from these sources of information. In this study, variance in the fNIRS signal that is explained by the semantic model might be equally explained by a visual model. Simply put, words with similar meanings often refer to objects that also look similar. Particularly, for the eight words depicted in our experiments, the four animals (dog, kitty, bunny, and bear) were far more similar in meaning to each other than to the four body parts (mouth, nose, hand, and foot). A casual examination of the stimuli suggests that a model of visual similarity would reach similar conclusions. Studies integrating a range of different visual, semantic, and other explanatory models may help to disentangle these components of the fNIRS response.

Despite these limitations, the present results represent the first link between the functional brain response patterns measured by fNIRS and an extrinsic model (in this case, a model based on distributional semantics), and the first successful decoding of semantic information in the fNIRS signal. This demonstration establishes a few important precedents for multivariate pattern analyses applied to fNIRS data: (1) multivariate patterns describing the distributed fNIRS response are preserved across subjects, allowing decoding between individuals and combination of individuals into group-level models to improve signal-to-noise ratio when very few trials are available per subject (e.g., studies with infants). (2) These regularities can be explained, in part, by extrinsic representational models, such as the COMPOSES distributional semantic model. (3) Representational similarity-based approaches to fNIRS decoding provide one effective means to bridge the gap among subjects and between model and neural data, even when placement and quality of individual channels varies.

The extension of these methods to fNIRS expands the analytic tools available to the field and facilitates direct, theoretical inferences about the informational content of neural signals. Further, applying multivariate pattern analyses and model-based decoding to fNIRS data will allow comparisons that are currently impossible with analogous methods for fMRI. Because fNIRS is a motion-tolerant and silent measure, neural data can be obtained from infants, children, and adults using the same techniques and making direct comparisons of the representational content among these groups.

Disclosures

The authors declare no conflicts of interest, financial or otherwise.

Acknowledgments

We wish to thank Holly Palmeri, Zoe Pruitt, and Julia Evans of the Rochester Baby Lab for assistance with data collection. We are also grateful to Andrew Anderson from the University of Rochester for suggestions and feedback on statistical analyses. This work was supported in part by the grants National Institutes of Health (NIH) EY-001319 to the Center for Visual Science at the University of Rochester, NIH HD-088731 to R.N.A., National Institute of Child Health and Human Development R00 AWD1004538 to L.L.E., a Philippe Foundation grant to L.B., and National Science Foundation BCS-1514351 to R.N.A.

References

1. T. M. Mitchell et al., "Predicting human brain activity associated with the meanings of nouns," *Science* **320**(5880), 1191–1195 (2008).
2. K. A. Norman et al., "Beyond mind-reading: multi-voxel pattern analysis of fMRI data," *Trends Cognit. Sci.* **10**(9), 424–430 (2006).
3. J. T. Serences and S. Saproo, "Computational advances towards linking BOLD and behavior," *Neuropsychologia* **50**(4), 435–446 (2012).
4. J.-D. Haynes, "A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives," *Neuron* **87**(2), 257–270 (2015).
5. A. J. Anderson et al., "Reading visually embodied meaning from the brain: visually grounded computational models decode visual-object mental imagery induced by written text," *NeuroImage* **120**, 309–322 (2015).
6. L. Fernandino et al., "Concept representation reflects multimodal abstraction: a framework for embodied semantics," *Cereb. Cortex* **26**(5), 2018–2034 (2015).
7. A. G. Huth et al., "A continuous semantic space describes the representation of thousands of object and action categories across the human brain," *Neuron* **76**(6), 1210–1224 (2012).
8. G. Dehaene-Lambertz, S. Dehaene, and L. Hertz-Pannier, "Functional neuroimaging of speech perception in infants," *Science* **298**(5600), 2013–2015 (2002).
9. B. Deen et al., "Organization of high-level visual cortex in human infants," *Nat. Commun.* **8**, 13995 (2017).
10. L. Biagi et al., "BOLD response selective to flow-motion in very young infants," *PLoS Biol.* **13**(9), e1002260 (2015).
11. Y. Liu et al., "Measuring speaker-listener neural coupling with functional near infrared spectroscopy," *Sci. Rep.* **7**, 43293 (2017).
12. X. Cui, D. M. Bryant, and A. L. Reiss, "NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation," *NeuroImage* **59**(3), 2430–2437 (2012).
13. J. Jiang et al., "Neural synchronization during face-to-face communication," *J. Neurosci.* **32**(45), 16064–16069 (2012).
14. L. L. Emberson et al., "Decoding the infant mind: multichannel pattern analysis (MCPA) using fNIRS," *PLoS One* **12**(4), e0172500 (2016).
15. N. Kriegeskorte, M. Mur, and P. Bandettini, "Representational similarity analysis-connecting the branches of systems neuroscience," *Front. Syst. Neurosci.* **2**, 4 (2008).
16. R. D. S. Raizada and A. Connolly, "What makes different people's representations alike: neural similarity space solves the problem of across-subject fMRI decoding," *J. Cognit. Neurosci.* **24**(4), 868–877 (2012).
17. J. M. Correia et al., "EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations," *Front. Psychol.* **6**, 71 (2015).
18. A. M. Chan et al., "Decoding word and category-specific spatiotemporal representations from MEG and EEG," *NeuroImage* **54**(4), 3028–3039 (2011).
19. L. Su et al., "Spatiotemporal searchlight representational similarity analysis in EMEG source space," in *Int. Workshop on Pattern Recognition in NeuroImaging (PRNI '12)*, pp. 97–100 (2012).
20. A. J. Anderson, B. D. Zinszer, and R. D. S. Raizada, "Representational similarity encoding for fMRI: pattern-based synthesis to predict brain activity using stimulus-model-similarities," *NeuroImage* **128**, 44–53 (2016).
21. M. M. Plichta et al., "Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study," *NeuroImage* **35**(2), 625–634 (2007).
22. T. J. Huppert et al., "HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain," *Appl. Opt.* **48**(10), D280–D298 (2009).
23. M. Baroni, G. Dinu, and G. Kruszewski, "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors," in *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 238–247 (2014).
24. B. D. Zinszer et al., "Semantic_Decoding_2017," Github, http://teammcpa.github.io/Semantic_Decoding_2017.
25. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Stat. Soc. B* **57**(1), 289–300 (1995).
26. D. Yekutieli and Y. Benjamini, "Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics," *J. Stat. Plann. Inference* **82**, 171–196 (1999).
27. N. Kriegeskorte et al., "Individual faces elicit distinct response patterns in human anterior temporal cortex," *Proc. Natl. Acad. Sci. U. S. A.* **104**(51), 20600–20605 (2007).

Benjamin D. Zinszer is now a research associate at the University of Texas at Austin. He previously worked as a postdoctoral associate at the Rochester Center for Brain Imaging, the University of Rochester. He received his PhD in psychology from Penn State University, studying the effects of cross-language interaction in Chinese–English bilinguals. His work explores linguistic categories, neural semantic representations, and the development of these structures in monolingual and bilingual learners.

Laurie Bayet is a postdoctoral research fellow at Boston Children's Hospital and Harvard Medical School. She was a postdoctoral research associate at the University of Rochester, and received her PhD from the University of Grenoble. Her work uses behavioral methods, computational tools, EEG, and fNIRS to examine changes in visual representations unfolding from infancy to adulthood with an emphasis on face and facial emotion perception development.

Lauren L. Emberson is an assistant professor in the Psychology Department, Princeton University. She received her PhD from Cornell University and was a postdoctoral research associate at the University of Rochester. Her work uses fNIRS, as well as behavioral methods, to investigate perceptual development and learning in young infants. Her research consistently pushes both theoretical and methodological or technical boundaries with the ultimate goal of understanding how experience supports development.

Rajeev D. S. Raizada is an assistant professor in the Department of Brain and Cognitive Sciences, the University of Rochester. He works in the development of neural decoding approaches for fMRI data, with application in particular to decoding semantic information from words and sentences. In this collaboration, he is excited to be able to extend such work from fMRI to fNIRS.

Richard N. Aslin is a senior scientist at Haskins Laboratories and affiliated with the Psychology Departments at Yale and the University of Connecticut (formerly emeritus professor and director of the Rochester Center for Brain Imaging at the University of Rochester). He has conducted research on human infants and adults at the behavioral and neural levels for the past 40 years, including studies of statistical learning, spoken word recognition, and sensory-motor control. He led a consortium that pioneered the use of fNIRS with infants in the 2000s and is a member of the National Academy of Sciences.