

Self-Supervised Pretraining on Paired Sequences of fMRI Data for Transfer Learning to Brain Decoding Tasks

Sean Paulsen, Michael Casey

Abstract—In this work, we present a self-supervised pretraining framework for transformers on functional Magnetic Resonance Imaging (fMRI) data. First, we pretrain our architecture on two self-supervised tasks simultaneously to teach the model a general understanding of the temporal and spatial dynamics of human auditory cortex during music listening. Our pretraining results are the first to suggest a synergistic effect of multitask training on fMRI data. Second, we finetune the pretrained models and train additional fresh models on a supervised fMRI classification task. We observe significantly improved accuracy on held-out runs with the finetuned models, which demonstrates the ability of our pretraining tasks to facilitate transfer learning. This work contributes to the growing body of literature on transformer architectures for pretraining and transfer learning with fMRI data, and serves as a proof of concept for our pretraining tasks and multitask pretraining on fMRI data.

Keywords—Transfer learning, fMRI, self-supervised, brain decoding, transformer, multitask training.

I. INTRODUCTION

FUNCTIONAL MRI (fMRI) scans measure blood-oxygen-level-dependent (BOLD) responses that reflect changes in metabolic demand consequent to neural activity [1]–[3]. By measuring BOLD responses at specific combinations of spatio-temporal resolutions and coverages, fMRI data provide the means to study complex cognitive processes in the human brain [4]–[6]. In particular, task-based fMRI protocols include targeted stimuli or other task variables, such as question answering, during the scan. Researchers can then conclude associations between task features and the evoked responses across the brain [7]–[9]. Regions of activity that are correlated with the presence of a particular task feature are thus taken to be involved in the brain's representation of that feature [10], and they are considered to be functionally connected [11]. Even rest-state fMRI data, that is, data collected in the absence of external stimuli or task, contain characteristic multi-variate signals of the brain [12]–[16]. Such rest-state signals have been shown to be predictive of the diagnosis and characterization of multiple neurological diseases and psychiatric conditions [17]–[19].

fMRI researchers have thus adopted machine learning (ML) techniques to analyze the complex relationship between BOLD signal and the underlying task, disease, or biological information. More specifically, training an ML model to predict such information given the BOLD data as input is known

as *task-state decoding*, or *brain decoding*. Toward the goal of more powerful brain decoding models, many advances in modern *deep* machine learning have been applied to fMRI research. These include convolution-based models [12], [20], [21], recurrent neural networks (RNN) [22], and graph neural networks [23]. Most recently, transformer [24] based models have achieved state of the art results on several brain decoding tasks [1], [25], [26], having already grown to dominate the fields of time series forecasting [27], natural language processing [28], and computer vision [29], [30].

However, training deep models is data intensive, while fMRI scans are expensive with relatively little data obtained per scan. One strategy to somewhat alleviate the burden of data is to *pretrain* the model on a *self-supervised* task to acquire general knowledge inherent in the dataset. The pretrained model then has a head start, so to speak, on the task of interest, by leveraging its general understanding of the data [31]. This strategy is nearly ubiquitous in the domain of Natural Language Processing (NLP) [32] and has begun to appear in fMRI studies [25], [26]. As Kalyan et al. [32] note: “These models provide good background knowledge to downstream tasks which avoids training of downstream models from scratch.” This process is called *transfer learning*. In this paper we propose two self-supervised pretraining tasks on sequences of audio-evoked fMRI data to facilitate transfer learning to downstream auditory brain decoding tasks. We demonstrate our transformer architecture's ability to learn these tasks and “transfer” that knowledge to improve convergence time on a supervised auditory brain decoding task. Further, our results show that simultaneous training on both pretraining tasks achieves superior final performance than training on only one of the tasks.

Our contributions are: (1) we present self-supervised tasks for two-task simultaneous pretraining on sequences of fMRI data, (2) we report our transformer architecture's successful learning of those tasks and achieve, to the best of our knowledge, the first evidence of a synergistic benefit from multitask training on fMRI data, (3) we demonstrate transfer learning to a supervised brain decoding task, and thereby establish a proof of concept of our pretraining tasks' suitability and our framework's capacity for transfer learning on fMRI data.

II. RELATED WORK

Univariate approaches to fMRI data such as contrast subtraction can be useful for basic analysis, but such

S. Paulsen is with the Department of Computer Science, Dartmouth College, Hanover, NH, 03755 USA (e-mail: paulsen.sean@gmail.com).

M. Casey is with the Department of Music and Department of Computer Science, Dartmouth College.

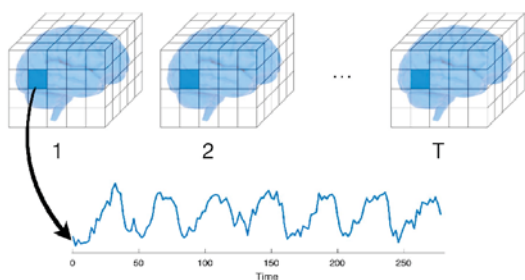


Fig. 1 The sequences of voxel data used in our experiments which are timeseries of neural activity measured by fMRI [51]

approaches struggle to isolate the densely overlapping patterns of multivariate signals which comprise neural activity [33], [34]. This challenge motivated the adoption of early ML architectures for multivariate fMRI analysis [35], [36], notably support vector machines for brain decoding classification [37]–[40]. Progression into *deep* ML models saw multilayer perceptrons [41], autoencoders [12], [42], convolutional neural networks (CNN) [43], [44], and graph neural networks (GNN) [23] for feature extraction and classification of single fMRI images. Time series analysis is perhaps more desirable due to the high degree of temporal correlation in BOLD responses, and indeed recurrent neural networks (RNN) and various long short-term memory (LSTM) models have been reported [45]–[49].

Most recently, the transformer [24] architecture has emerged as a superior alternative to recurrent methods for fMRI timeseries modeling. Bedel et al. [1] improved the state of the art for timeseries classification on multiple public fMRI datasets with a novel fused-window attention mechanism, but their work did not explore pretraining or transfer learning. Nguyen et al. [26] achieved state of the art classification accuracy for a task-state decoding task on the Human Connectome Project 7-task dataset [50]. Their analysis includes the explicit benefits of the transformer’s self-attention module when compared to previous recurrent architectures, as well as a demonstration of transfer learning when pretraining on held-out subsets of HCP 7-task. However, their pretraining task was supervised classification specific to HCP 7-task labelled data, and thus their pretrained models would be of little to no value toward transfer learning on different datasets or modalities [32]. Malkiel et al. [25] pretrain on a self-supervised fMRI reconstruction task by wrapping the transformer block in an encoder-decoder. They report that their pretraining was crucial for improved state of the art performance on a variety of fMRI tasks such as age and gender prediction, and schizophrenia recognition. We note that their downstream task uses the CLS token decoding method popularized by Devlin et al. [28], while their pretraining task does *not* incorporate the CLS token. This inconsistency between training phases does not obtain the full value of the transfer learning paradigm.

Extending the above work, we explore multitask pretraining and transfer learning with novel self-supervised pretraining tasks which include the CLS token, with all model inputs in a standardized geographical brain space, *without* passing through an embedding layer.

III. ARCHITECTURE AND TRAINING TASKS

A. Paired-Sequence Transformer

Our architecture is a modified stacked bidirectional-encoder design (Fig. 2) with two separate output blocks, one for each of two self-supervised pretraining tasks on which the model is trained simultaneously. We implemented our models from scratch with the pyTorch library. Our model does not include the standard embedding layer after positional encoding. We hypothesize that the composition of the fMRI scanner’s measurement of BOLD signal with the mapping of that measurement to MNI space constitutes a meaningful embedding of the physical, biological neural response. The data are *already* in a shared, distributed, representative space. Hence, we dispense with the embedding layer in our design.

A thorough explanation of the data preprocessing and the construction of the inputs to the model is given in the Data Preparation section below, which we summarize here. All training data in this work were built from the Music Genre fMRI Dataset (2021) [52]. The images were collected while five subjects listened to samples of music from ten different genres. Each input to the model is constructed by extracting a contiguous sequence of five fMRI images of a subject listening to music (Seq1), and pairing it with another (different) such sequence (Seq2). A *separator token* (SEP) [28] is inserted between the two sequences, and a *classification token* (CLS) [25], [26], [28] is inserted at the front. Before constructing the inputs, we reduced all fMRI images to only the left-side auditory cortex, resulting in 420 voxels, which we then flattened to 1-D. Thus each input x_i in the training set is a sequence of twelve 420-dimensional vectors:

$$x_i = [CLS, \vec{v}_0, \dots, SEP, \vec{v}_5, \dots, \vec{v}_{11}], \quad v_j \in \mathbb{R}^{420} \quad (1)$$

The implementation of these tokens without an embedding layer is explained in the last paragraph of the Materials section below.

B. Pretraining Tasks

We now present our two self-supervised pretraining tasks. First is *Next Thought Prediction* (NTP). The goal of this task is binary classification, predicting whether or not Seq2 follows immediately after Seq1 in the original data. From the output of the final transformer block, the transformed CLS token is sent to *Output Block 1*. This block consists of a linear layer projecting down from 420 dimensions to 210, then a linear layer projecting down from 210 to 2, and finally a softmax is applied to obtain probabilities for “No” (index 0) and “Yes” (index 1). The loss for NTP is calculated as the Cross-Entropy between the result of Output Block 1 and a one-hot encoding of the ground truth.

Our second pretraining task is *Masked Brain Modeling* (MBM). The goal of this task is to reconstruct a masked element or elements of the input sequence. When an input arrives at the model, before positional encoding, either one or two of the ten fMRI images are chosen uniformly at random (without replacement) for masking. It is a 50/50 chance whether one or two are chosen. When an image is chosen, there is an 80% chance to replace it with the *mask token* (MSK), a 10%

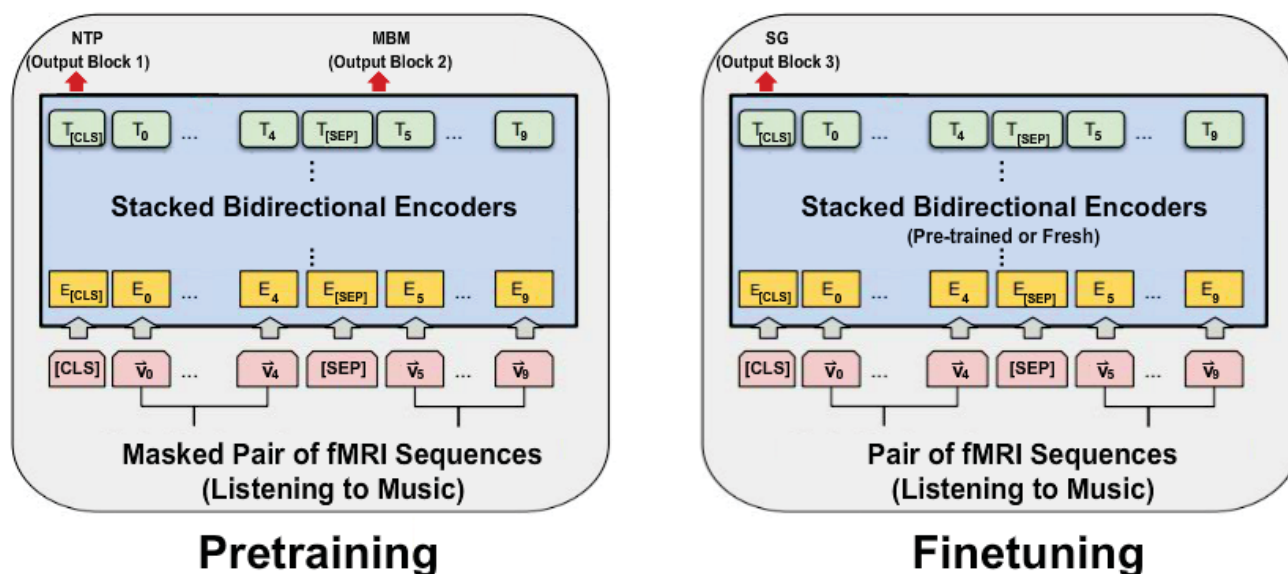


Fig. 2 Depiction of our Pretraining and Finetuning phases; output blocks are not pictured but are detailed in corresponding sections

chance to replace it with a random image sampled uniformly from the full dataset, and a 10% chance to leave it unchanged. The chosen indices are recorded, and the elements of the final transformer block's output at those indices are passed separately to *Output Block 2*. This block consists of a dense layer with ReLU activation, then a second dense layer with linear activation. The loss for MBM is calculated as the Mean Squared Error between the output and the original chosen fMRI image. In the case of two chosen images, the total MBM loss is the average of the two individual MBM losses.

Note the inherent data augmentation of the MBM task. There are ten fMRI images in each training sample, and the result of each possible masking configuration yields a distinct training sample. Thus MBM can effectively grow the size of the training set by an order of magnitude if the model is trained long enough. This gain is perhaps overlooked in domains such as natural language processing where billions of training samples are available. In fMRI studies, however, data poverty is a consistent concern due to the financial and time costs of scanning. While we make no specific claims about the effects of this augmentation in this work, this potential benefit built into the task is noteworthy.

C. Multitask Learning

Training on more than one task simultaneously, known as *Multitask Learning* (MTL), has been shown to improve downstream performance in several domains [53] by benefiting from the underlying *relationships* between tasks, but to our knowledge this has not previously been done when training on fMRI data. In their thorough treatment of the brain's musical reward system, Salimpoor et al. [54] comment "music pleasure is thought to rely on generation of expectations, anticipation of their development and outcome, and violation or confirmation of predictions." In other words, the notion of "what comes next" is intimately connected to the explicit values of voxel

activity. NTP and MBM embody these two concepts, so indeed our multitask pretraining scheme is aligned with the literature.

The raw loss value of NTP for a single training sample is, on average, at least an order of magnitude greater than the loss value of MBM when training begins. Therefore the parameter updates will certainly be dominated by NTP, stifling any learning from MBM. Deriving a theoretically optimal way to combine the two losses would be a significant endeavor, so our total training loss for a single sample is merely the weighted sum of our two loss values:

$$E_{multi} = \alpha_1 * E_{NTP} + \alpha_2 * E_{MBM} \quad (2)$$

$$\alpha_1 + \alpha_2 = 1. \quad (3)$$

These two weights are simply hyperparameters to be tuned. We explore these and other hyperparameters in the Experiments and Results section below.

D. Finetuning Task

Our supervised brain decoding task used for finetuning is the *Same Genre* (SG) task. The goal of this task is binary classification, predicting whether or not Seq1 corresponds to listening to the same genre of music as Seq2. From the output of the final transformer block, the transformed CLS token is sent to *Output Block 3*. This block consists of a single linear layer projecting down from 420 dimensions to 2, then a softmax is applied to obtain probabilities for "No" (index 0) and "Yes" (index 1). The loss for SG is calculated as the Cross-Entropy between the result of *Output Block 3* and a one-hot encoding of the ground truth. We made the finetuning output block as simple as possible to ensure that the brunt of the work is supported by the pretrained transformer blocks.

TABLE I
 BEST PERFORMING CONFIGURATION FOR THE TWO TRAINING REGIMENS

	MULTI	NTP
α_1, α_2	0.1, 0.9	N/A
LR	10^{-4}	10^{-5}
ATN HDS	2	2
F EXP	4	4

IV. EXPERIMENTS AND RESULTS

A. Hyperparameter Search

One of the most important questions to ask in the context of multitask learning is whether the model would have been better off with only one task. In particular, *how much* is the performance on NTP impeded by having to learn MBM at the same time? To explore this, we performed our hyperparameter grid search for training on the multitask regimen as well as NTP alone. Table I shows the best performing (i.e. achieved the highest validation accuracy on NTP at some point during training) configurations. We let the NTP task guide our search because its binary accuracy is simply more interpretable than any metric involving the MBM task. Nevertheless, the multitask models' performance on MBM is included in our analysis below.

All training during grid search held out run 0 from the dataset as a validation split. We applied a dropout rate of 0.1 in all transformer blocks. Models were trained for ten epochs via backpropagation with the Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$, and $\text{weight_decay} = 0.0001$.

In general, fewer attention heads with more layers outperformed the reverse. It is reassuring to obtain the same value for attention heads and forward expansion on both regimens. The best performing learning rate for NTP-only is an order of magnitude smaller than for multitask, but this is not surprising. NTP's contribution to the loss is scaled by $\alpha_1 = 0.1$, and in the most basic gradient descent, scaling the loss function by a constant is functionally the same as scaling the learning rate by that constant instead. The Adam optimizer is a bit more complex, but the general idea holds.

B. Pretraining

After identifying the best performing hyperparameters for both cases, we performed 12-fold cross validation for both multitask and NTP-only, where each fold holds out one of twelve runs from the dataset. It was unclear during hyperparameter search whether a 3 or 4 layer model was superior, so we considered both here. The same Adam specifications as hyperparameter search were used. The exact details of pretraining dataset construction can be found in the Materials section below, but we note here that each fold has 10,000 training samples and 800 validation samples.

Results are presented in Table II. For each fold, we saved the model's state after the epoch with the highest NTP accuracy on the validation split ("Best Val Acc" in the table). The "Best Epoch" column contains the epoch in which this accuracy was achieved. The MBM loss calculated on the validation split after the Best Epoch is also given to consider the relationship between the two tasks. We consider as a baseline that the MBM

training loss on the first training sample seen by a model is around 0.4. The averages of each column are given in the last row of the table.

Models with 3 layers outperformed on average on both accuracies of interest as well as MBM Validation Loss, so we proceeded to finetuning with the saved 3-layer models. The exact details of finetuning dataset construction can be found in the Materials section below, but as in the pretraining phase, each fold has 10,000 training samples and 800 validation samples.

C. Finetuning

We loaded the 12 3-layer models saved after their Best Epoch during the Multitask and NTP-only regimens and trained them for ten epochs on the Same Genre task described above. The training data for each model hold out the same run as were held out during pretraining as a validation split. Preliminary testing showed that freezing the pretrained weights and updating only the new output block was not a successful training strategy for this work. Therefore all parameters were updated during finetuning. To examine the benefit of transfer learning, we also trained 12 "fresh" models. The fresh models are identical to the other models used in finetuning but do not load any pretrained weights.

The Adam optimizer parameters were the same as during pretraining. We trained all 36 models with a Learning Rate of 10^{-4} and then again with 10^{-5} - the two learning rates used during pretraining. Table III gives the 10^{-5} results. These results outperformed the 10^{-4} results across the board so those are not reported.

D. Discussion

The first point of interest is that the pretraining phase was successful at all. fMRI data are a challenging domain and paired-sequence transformers have not previously been used in this domain, nor has multitask learning. Nevertheless, our implementation is conclusively capable of learning these tasks. The average best performance between the two regimens is not significantly different (87.6% vs. 88.2%), which alleviates concerns about MBM impeding the ability to learn NTP. Moreover, it does not impede the *speed* at which the multitask models achieve their best performance- about 8 epochs in both cases. The multitask models are more volatile, with lower lows but also higher highs. NTP-only achieves a highest validation accuracy of 90.75%, but multitask runs achieve 92%, 93.5%, and 94.875%, which is *our first evidence of a synergistic benefit from self-supervised multitask training on fMRI data*.

Our supervised brain decoding task, Same Genre, was also successful on both pretrained models and fresh models. The models pretrained on NTP-only significantly outperformed the fresh models, which is *our first significant evidence of the ability to perform transfer learning with our model from one of our self-supervised pretraining tasks to a supervised brain decoding task*. The models pretrained on Multitask almost exactly matched the baseline fresh models on average, but we note a similar volatility to the pretraining phase. The average of the Multitask models is dragged down by folds 4 (72.75%) and

TABLE II
RESULTS OF 12-FOLD CROSS VALIDATION FOR MULTITASK (NTP AND MBM) AND NTP-ONLY PRETRAINING REGIMENS, ON 3 AND 4 LAYERS

Heldout Run	N Layers	Multitask (NTP and MBM)			NTP Only	
		Best Val Acc	Best Epoch	MBM Val Loss	Best Val. Acc	Best Epoch
0	4	93.5%	8	0.00103	88.375%	6
	3	88.125%	8	0.00048	88.25%	9
1	4	87.375%	9	0.00088	87.375%	8
	3	90.6%	6	0.00051	88.375%	9
2	4	88.625%	4	0.00070	87.875%	9
	3	88.75%	9	0.00037	89.375%	8
3	4	86.875%	7	0.00043	87.375%	8
	3	89.5%	9	0.00118	87.75%	7
4	4	80.0%	3	0.00107	83.0%	8
	3	80.5%	8	0.00045	90.75%	9
5	4	88.375%	9	0.00080	87.0%	9
	3	90.75%	9	0.00040	87.75%	9
6	4	79.375%	8	0.00079	83.875%	6
	3	84.125%	8	0.00051	87.75%	9
7	4	79.875%	3	0.00259	85.375%	9
	3	85.625%	8	0.00071	89.25%	9
8	4	81.75%	6	0.00098	90.0%	9
	3	94.875%	8	0.00083	90.125%	8
9	4	82.25%	9	0.00102	85.0%	8
	3	85.0%	8	0.00076	84.75%	4
10	4	80.375%	5	0.00079	87.0%	9
	3	92.0%	9	0.00077	87.25%	9
11	4	72.278%	1	0.00070	88.734%	9
	3	82.152%	9	0.00032	87.468%	9
Average	4	83.388%	6	0.00098	86.749%	8.17
	3	87.613%	8.25	0.00061	88.237%	8.25

TABLE III
RESULTS OF 12-FOLD CROSS VALIDATION FOR THREE FINETUNING REGIMENS ON THE SAME GENRE TASK: MULTITASK-PRETRAINED MODELS, NTP-ONLY-PRETRAINED MODELS, AND FRESH MODELS

Heldout Run	Multitask (NTP and MBM)		NTP Only		Fresh	
	Best Val Acc	Best Epoch	Best Val Acc	Best Epoch	Best Val. Acc	Best Epoch
0	82.625%	7	94.75%	9	84.625%	8
1	86.625%	9	91.375%	9	88.75%	5
2	88.375%	9	94.0%	6	89.625%	6
3	93.0%	4	92.125%	8	89.5%	9
4	72.75%	9	93.25%	8	82.5%	6
5	89.5%	9	92.0%	5	86.5%	9
6	82.0%	9	91.0%	9	82.75%	9
7	98.25%	9	90.875%	5	82.5%	9
8	94.25%	9	95.375%	9	83.625%	6
9	78.125%	9	92.375%	9	83.125%	5
10	82.625%	7	91.875%	8	87.875%	8
11	81.392%	2	97.089%	8	83.291%	9
Average	85.793%	7.67	93.007%	7.75	85.389%	7.42

9 (78.125%). On the other hand, fold 7 achieves a staggering 98.25% validation accuracy, as well as 93% and 94.25%, all of which exceed the fresh models' best fold of 91.625%. NTP-only reached a maximum of 97.089%, which is also short of the Multitask maximum.

The relationship between pretraining performance and finetuning performance is unclear. For example, the second highest finetuning accuracy for Multitask was on folds 8, which was the highest performance of the 3-layer models during pretraining, indicating the positive relationship between the two phases that we would expect. On the other hand, fold 7 had the best performing Multitask finetuning accuracy, or rather the best finetuning accuracy of any regimen, while the pretraining accuracy and MBM loss were both below average on that fold. More work is required to properly identify a

relationship between the two phases.

V. CONCLUSIONS AND FUTURE WORK

In this work we presented two self-supervised tasks for pretraining on sequences of fMRI data- Next Thought Prediction, and Masked Brain Modeling. The results of our pretraining phase demonstrated our paired-sequence transformer architecture's successful learning of those tasks, as well as the first evidence of a synergistic benefit from multitask training on fMRI data. The results of our finetuning phase demonstrated transfer learning from NTP to a supervised brain decoding task, establishing a proof of concept of our framework's suitability for transfer learning on fMRI data, in particular, in the absence of an embedding layer.

We observe the similarity between the Same Genre task and a standard contrast analysis of genre in STG. Fundamentally, both are asking if STG encodes different genres in different ways. Our implementation answers this question in the affirmative without the need for a contrast analysis. More generally, any contrast analysis interested in the difference between two conditions could be substituted by our model. Our upcoming work aims to show that our model can detect such differences which univariate approaches, such as contrast, fail to recognize.

Looking elsewhere, the experiments in this work do not exploit a particular power of transformers for learning long term dependencies. Increasing the sequence length would of course reduce the size of the training set, and this was our primary motivation for using a small sequence length in these experiments, but there were also some restrictions in the data collection protocol preventing us from wielding longer sequences. This is discussed in detail in the Materials section below. The Human Connectome Project (2013) [50] is a prevailing candidate for future on long sequences of fMRI data due to its overwhelming size and well-established benchmarks.

As stated above, we believe this paired-sequence framework has potential for replacing or supplementing contrast analysis, but there is also a great deal of work with fMRI data, for example basic brain decoding, that cannot use the paired-sequence structure. Our upcoming work has a self-supervised pretraining task for only *one* sequence, which generalizes immediately to common brain decoding tasks and datasets.

VI. MATERIALS

A. Data Preprocessing

The training data for this work were built from the Music Genre fMRI Dataset (2021) [52] available on the OpenNeuro database. This dataset contains whole-brain images of five subjects listening to 540 music pieces from 10 music genres. We refer to the original paper [55] for the full details of data collection, but much of it has been covered in this work as it became relevant. We used the Brainlife (2017) [56] application to conduct our suite of preprocessing for the Music Genre dataset. The fMRIPrep (2019) [57] preprocessing package performed motion correction, field unwarping, normalization, bias field correction, and brain extraction. Next, fMRIPrep mapped the 3-D images of each participant's brain into a standardized 3-D (96x114x96) vector space (MNI space) [58] such that the various structural components of each brain are at the same coordinates. This enables direct comparison and analysis of BOLD signals in physically different brains. In our case, MNI space facilitates among-subject training of machine learning models. We refer to this standardized vector space throughout the paper as the *voxel space*, where the 3-D coordinates correspond to $1x1x1mm^3$ cubes of brain called *voxels* and the value at each coordinate is the BOLD signal in that volume. To summarize, the *voxel data* used in model training are the timeseries of 3-D BOLD signals after preprocessing and mapping to MNI space. (Fig. 1).

B. Region of Interest

The full MNI space is several orders of magnitude too large for our purposes, but more importantly we are only interested in regions of the brain recruited during attentive listening to music. The Superior Temporal Gyrus (STG) is the site of the auditory cortex, which processes auditory information. Angulo-Perkins et al. [59] showed preferential involvement of STG in processing music in both musicians and non-musicians, which fits our goal of learning from the Music Genre dataset. STG has also been used to learn decoding models of complex natural sounds [60], language [61], and even imagined sound [37] from fMRI data. Therefore we chose to extract STG from each subject's voxel data. FSLEyes (2022) [62] is a free application for viewing fMRI images and includes several **atlases** for isolating structural ROIs in the brain with respect to MNI space. We used the Harvard-Oxford cortical structure atlas (HO atlas), some regions of which are shown as an example in Fig. 4.

The HO atlas assigns a probability to each voxel of belonging to each ROI. Therefore in order to extract STG, we needed to choose a minimum probability threshold for inclusion in STG. This threshold is a hyperparameter to be tuned in future work, but in this work all datasets have a threshold of 23%. We obtained our threshold by visual inspection of the resulting regions. In their seminal work, Craddock et al. [63] used a threshold of 25% with the HO atlas, so our visual inspection method is only slightly more lenient.

The HO atlas labels anterior and posterior STG separately, so we applied the threshold to both regions and concatenated them. Voxels which met the threshold for inclusion in *both* anterior and posterior are only included once and the greater of the two probabilities is preserved. Fig. 3 shows the heatmap corresponding to this union.

We proceed with only one of the two lateral STGs for reduced model complexity and thus lower resource demand for training. In our previous work [12], we successfully decoded a particular auditory stimulus information from left-STG but not right-STG, and this was our primary basis for choosing left rather than right for these experiments. The number of voxels in left-STG with inclusion threshold 23% is 413. However, to perform multi-head attention, the dimension of the input to a transformer must be evenly divisible by the number of attention heads. Thus our inputs should have a dimension (i.e. number of voxels) with a "nice" factorization to evaluate different numbers of attention heads. The choice was between *removing* the voxels with the *lowest* probabilities of inclusion *above* 23% to reach 400, or inserting voxels with the highest probabilities below 23% to reach 420. We chose the latter for its more diverse factorization, that is, allowing for all of 2, 3, 4, 5, 6, and 7 attention heads. We then performed linear detrending across the full scan to each voxel and, finally, standardized each voxel to mean zero across the full scan.

C. Creating the Training Data

The original dataset was collected by scanning while 15-second clips of music played with a 1.5 second repetition time (TR). So each clip corresponds to 10 consecutive images in

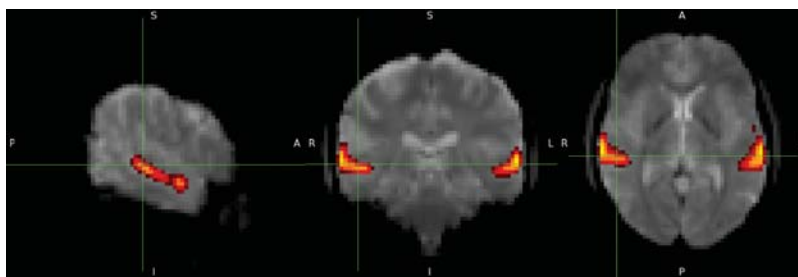


Fig. 3 Heatmaps for probability of voxel inclusion in STG; only probabilities greater than or equal to 23% are shown

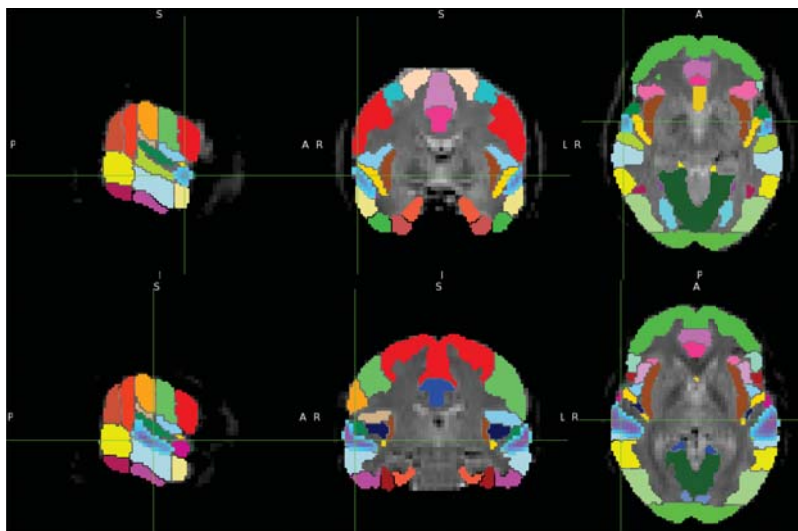


Fig. 4 Heatmaps for probability of a particular voxel being included in a chosen region of interest; anterior STG (top) and Posterior STG (bottom) maps with a minimum inclusion probability of 23% are overlaid on the Harvard-Oxford atlas

Open Science Index, Cognitive and Language Sciences Vol:18, No:2, 2024 publications.waset.org/10013511.pdf

the dataset. There were no rest periods between clips. Therefore, with the Same Genre task in mind, the most natural sequence length for each of the inputs to the model is 10, but we chose 5 for more training data. Thus each sequence of 5 (5-seq) for the SG task is either timesteps 1 through 5 or 6 through 10 of a music clip. All 5-seqs were extracted. For each 5-seq, a positive and negative training sample was created. That is, each 5-seq was inserted to the left of the SEP token, and a 5-seq with the same music genre was sampled uniformly at random from the same subject's data and placed to the right of the SEP token to create a positive sample, similarly for a negative sample. We have not yet trained this task when the pairs are among-subject, but the models in this work are trained on the collection of pairs from all subjects, resulting in a within-among-subject hybrid. The size of our SG datasets could be multiplied by making more positive and negative samples, but that would of course multiply the training time, and our results were already strong on this task.

We recall that the masking process for MBM is done when an input arrives at the model, so creating a dataset for multitask training reduces to creating a dataset for NTP-only. We needed to use 5-seqs to make training data for the NTP task to have the correct model specifications for finetuning on SG. All 5-seqs were extracted and both a positive and negative within-subject sample were created, as with SG. These 5-seqs could have arbitrary start and end points and possibly overlap, but we chose

the same start and end points as for SG. This was because it sets up an interesting challenge for the model to overcome. When the 5-seq is images 1 through 5 for a music clip, the positive sample for NTP is necessarily a positive sample for SG. But when the 5-seq is images 6 through 10, the positive sample for NTP is necessarily a negative sample for SG- that is, the 5-seq immediately following it must be for a different genre (a genre never followed itself during scanning). In simple terms, for half the samples where "yes" is the correct answer during pretraining, "no" would be the correct answer during finetuning. Thus this construction presented a meaningful hurdle for the models to overcome during finetuning, they could not simply follow the same pathways they learned during pretraining.

The training data for each fold of the cross-validation were constructed by holding out one of the 12 runs labelled "Training Run" in the original dataset. The inputs to the model contain two 5-seqs, so inputs in the validation splits were constructed by sampling both 5-seqs from the heldout run, and the training splits did not sample from the heldout runs. Note that while the dataset construction process was the same for pretraining and finetuning, the datasets themselves are *not* the same. We constructed a dataset for each fold of pretraining and finetuning. Within each fold of pretraining, Multitask and NTP-only used the same dataset in order to directly compare performance. Similarly, for each fold of finetuning, the three regimens used the same dataset.

In addition to the 12 runs labelled “Training Run” in the original dataset, there are six “Test Runs,” which were constructed slightly differently. We emphasize here that the words “training” and “test” in the original run labels have no relation to our own training and validation splits. Each “Training Run” corresponds to 40 different music clips, while each “Test Run” corresponds to a sequence of 10 music clips (one from each genre) repeated four times. When creating our datasets, we only included 5-seqs from the *first* instance of each clip in the “Test Runs,” and the others were discarded. The result for each fold was 10000 total training samples and 400 validation samples.

Finally, we recall that our architecture does not have an embedding layer. In NLP, the tokens are added as word indices to the vocabulary and the embedding layer learns their distributed representations [24]. Malkiel et al. [25] prepend a CLS token to sequences of fMRI images and pass that sequence through a learned embedding layer. But the original form of the CLS token must have the same dimension as the fMRI images in the sequence in order for the embedding layer to accept it. They do not report what this original form was. Logically, the tokens ought to be “far away” from the rest of the data in the distributed space. Thus we simply reserved the first three of the 420 dimensions for our tokens. The CLS, SEP, and MSK tokens have a 1 in the zeroth, first, and second dimensions respectively, and are zero elsewhere. Each fMRI image has zero in those dimensions. Indeed, we had to remove the three voxels with lowest probability of inclusion from each image to make room for the tokens, and thus in practice only had 417 voxels in each image rather than 420. The success of our training validates our implementation of these tokens without an embedding space.

REFERENCES

- [1] H. A. Bedel, I. Şıvgın, O. Dalmaz, S. U. H. Dar, and T. Çukur, “BoLT: Fused Window Transformers for fMRI Time Series Analysis,” Feb. 2023, arXiv:2205.11578 [cs, eess]. [Online]. Available: <http://arxiv.org/abs/2205.11578>
- [2] E. M. C. Hillman, “Coupling mechanism and significance of the BOLD signal: a status report,” *Annual Review of Neuroscience*, vol. 37, pp. 161–181, 2014.
- [3] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D. Y. von Cramon, “Modeling hemodynamic response for analysis of functional MRI time-series,” *Human Brain Mapping*, vol. 6, no. 4, pp. 283–300, 1998.
- [4] M. Kubicki, R. W. McCarley, P. G. Nestor, T. Huh, R. Kikinis, M. E. Shenton, and C. G. Wible, “An fMRI study of semantic processing in men with schizophrenia,” *NeuroImage*, vol. 20, no. 4, pp. 1923–1933, Dec. 2003.
- [5] D. Wang, L. Shi, D. S. Yeung, P.-A. Heng, T.-T. Wong, and E. C. C. Tsang, “Support vector clustering for brain activation detection,” *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 8, no. Pt 1, pp. 572–579, 2005.
- [6] J. M. Papma, M. Smits, M. de Groot, F. U. Mattace Raso, A. van der Lugt, H. A. Vrooman, W. J. Niessen, P. J. Koudstaal, J. C. van Swieten, F. M. van der Veen, and N. D. Prins, “The effect of hippocampal function, volume and connectivity on posterior cingulate cortex functioning during episodic memory fMRI in mild cognitive impairment,” *European Radiology*, vol. 27, no. 9, pp. 3716–3724, Sep. 2017.
- [7] K. Li, L. Guo, J. Nie, G. Li, and T. Liu, “Review of methods for functional brain connectivity detection using fMRI,” *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, vol. 33, no. 2, pp. 131–139, Mar. 2009.
- [8] A. Venkataraman, K. R. A. Van Dijk, R. L. Buckner, and P. Golland, “EXPLORING FUNCTIONAL CONNECTIVITY IN FMRI VIA CLUSTERING,” *Proceedings of the ... IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP (Conference)*, vol. 2009, pp. 441–444, Apr. 2009.
- [9] S. Nishimoto, A. Vu, T. Naselaris, Y. Benjamini, B. Yu, and J. Gallant, “Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies,” *Current Biology*, vol. 21, no. 19, pp. 1641–1646, Oct. 2011. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0969082211009377>
- [10] O. Simon, F. Kherif, G. Flandin, J.-B. Poline, D. Rivière, J.-F. Mangin, D. Le Bihan, and S. Dehaene, “Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number,” *NeuroImage*, vol. 23, no. 3, pp. 1192–1202, Nov. 2004. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S105381190400549X>
- [11] B. P. Rogers, V. L. Morgan, A. T. Newton, and J. C. Gore, “Assessing functional connectivity in the human brain by fMRI,” *Magnetic Resonance Imaging*, vol. 25, no. 10, pp. 1347–1357, Dec. 2007.
- [12] S. Paulsen, L. May, and M. Casey, “Decoding imagined auditory pitch phenomena with an autoencoder based temporal convolutional architecture,” in *BRAININFO*. Nice, France: IARIA, July 2021.
- [13] C. Niu, A. D. Cohen, X. Wen, Z. Chen, P. Lin, X. Liu, B. H. Menze, B. Wiestler, Y. Wang, and M. Zhang, “Modeling motor task activation from resting-state fMRI using machine learning in individual subjects,” *Brain Imaging and Behavior*, vol. 15, no. 1, pp. 122–132, Feb. 2021. [Online]. Available: <http://link.springer.com/10.1007/s11682-019-00239-9>
- [14] B. T. T. Yeo, F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, B. Fischl, H. Liu, and R. L. Buckner, “The organization of the human cerebral cortex estimated by intrinsic functional connectivity,” *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, Sep. 2011.
- [15] K. R. A. Van Dijk, T. Hedden, A. Venkataraman, K. C. Evans, S. W. Lazar, and R. L. Buckner, “Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization,” *Journal of Neurophysiology*, vol. 103, no. 1, pp. 297–321, Jan. 2010.
- [16] Z. Hu and P. Shi, “Interregional functional connectivity via pattern synchrony,” 01 2007, pp. 1 – 6.
- [17] X. Zhan and R. Yu, “A Window into the Brain: Advances in Psychiatric fMRI,” *BioMed Research International*, vol. 2015, p. 542467, 2015.
- [18] N. D. Woodward and C. J. Cascio, “Resting-State Functional Connectivity in Psychiatric Disorders,” *JAMA psychiatry*, vol. 72, no. 8, pp. 743–744, Aug. 2015.
- [19] C. H. Xia, Z. Ma, R. Ciric, S. Gu, R. F. Betzel, A. N. Kaczkurkin, M. E. Calkins, P. A. Cook, A. García de la Garza, S. N. Vandekar, Z. Cui, T. M. Moore, D. R. Roalf, K. Ruparel, D. H. Wolf, C. Davatzikos, R. C. Gur, R. E. Gur, R. T. Shinohara, D. S. Bassett, and T. D. Satterthwaite, “Linked dimensions of psychopathology and connectivity in functional brain networks,” *Nature Communications*, vol. 9, no. 1, p. 3003, Aug. 2018.
- [20] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, “3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI,” *IEEE Access*, vol. 5, pp. 23 626–23 636, 2017. [Online]. Available: <http://ieeexplore.ieee.org/document/8067637/>
- [21] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, “BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment,” *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017.
- [22] J. Dakka, P. Bashivan, M. Gheiratmand, I. Rish, S. Jha, and R. Greiner, “Learning Neural Markers of Schizophrenia Disorder Using Recurrent Neural Networks,” Dec. 2017, arXiv:1712.00512 [cs]. [Online]. Available: <http://arxiv.org/abs/1712.00512>
- [23] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, “BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis,” *Medical Image Analysis*, vol. 74, p. 102233, Dec. 2021.
- [24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” Dec. 2017, arXiv:1706.03762 [cs]. [Online]. Available: <http://arxiv.org/abs/1706.03762>
- [25] I. Malkiel, G. Rosenman, L. Wolf, and T. Hendler, “Self-Supervised Transformers for fMRI representation,” Aug. 2022, arXiv:2112.05761 [cs, eess]. [Online]. Available: <http://arxiv.org/abs/2112.05761>
- [26] S. Nguyen, B. Ng, A. D. Kaplan, and P. Ray, “Attend and Decode: 4D fMRI Task State Decoding Using Attention

- Models,” Jan. 2021, arXiv:2004.05234 [cs]. [Online]. Available: <http://arxiv.org/abs/2004.05234>
- [27] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan, “Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting,” Jan. 2020, arXiv:1907.00235 [cs, stat]. [Online]. Available: <http://arxiv.org/abs/1907.00235>
- [28] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” May 2019, arXiv:1810.04805 [cs]. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Jun. 2021, arXiv:2010.11929 [cs]. [Online]. Available: <http://arxiv.org/abs/2010.11929>
- [30] L. H. Li, M. Yatskar, D. Yin, C.-J. Hsieh, and K.-W. Chang, “VisualBERT: A Simple and Performant Baseline for Vision and Language,” Aug. 2019, arXiv:1908.03557 [cs]. [Online]. Available: <http://arxiv.org/abs/1908.03557>
- [31] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio, “Why does unsupervised pre-training help deep learning?” *Journal of Machine Learning Research*, vol. 11, no. 19, pp. 625–660, 2010. [Online]. Available: <http://jmlr.org/papers/v11/erhan10a.html>
- [32] K. S. Kalyan, A. Rajasekharan, and S. Sangeetha, “AMMUS : A Survey of Transformer-based Pretrained Models in Natural Language Processing,” Aug. 2021, arXiv:2108.05542 [cs]. [Online]. Available: <http://arxiv.org/abs/2108.05542>
- [33] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.
- [34] M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith, “Temporal autocorrelation in univariate linear modeling of FMRI data,” *NeuroImage*, vol. 14, no. 6, pp. 1370–1386, Dec. 2001.
- [35] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, “Beyond mind-reading: multi-voxel pattern analysis of fMRI data,” *Trends in Cognitive Sciences*, vol. 10, no. 9, pp. 424–430, Sep. 2006.
- [36] J. V. Haxby, “Multivariate pattern analysis of fMRI: the early beginnings,” *NeuroImage*, vol. 62, no. 2, pp. 852–855, Aug. 2012.
- [37] L. May, A. R. Halpern, S. D. Paulsen, and M. A. Casey, “Imagined Musical Scale Relationships Decoded from Auditory Cortex,” *Journal of Cognitive Neuroscience*, vol. 34, no. 8, pp. 1326–1339, 07 2022. [Online]. Available: https://doi.org/10.1162/jocn_a_01858
- [38] X. Song and N.-k. Chen, “A SVM-based quantitative fMRI method for resting-state functional network detection,” *Magnetic Resonance Imaging*, vol. 32, no. 7, pp. 819–831, Sep. 2014.
- [39] Z. Wang, A. R. Childress, J. Wang, and J. A. Detre, “Support vector machine learning-based fMRI data group analysis,” *NeuroImage*, vol. 36, no. 4, pp. 1139–1151, Jul. 2007.
- [40] S. H. Hojjati, A. Ebrahimzadeh, A. Khazaei, A. Babajani-Feremi, and Alzheimer’s Disease Neuroimaging Initiative, “Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM,” *Journal of Neuroscience Methods*, vol. 282, pp. 69–80, Apr. 2017.
- [41] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, “State-space model with deep learning for functional dynamics estimation in resting-state fMRI,” *NeuroImage*, vol. 129, pp. 292–307, Apr. 2016.
- [42] H. Huang, X. Hu, Y. Zhao, M. Makkie, Q. Dong, S. Zhao, L. Guo, and T. Liu, “Modeling Task fMRI Data Via Deep Convolutional Autoencoder,” *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1551–1561, Jul. 2018.
- [43] X. Wang, X. Liang, Z. Jiang, B. A. Nguchu, Y. Zhou, Y. Wang, H. Wang, Y. Li, Y. Zhu, F. Wu, J.-H. Gao, and B. Qiu, “Decoding and mapping task states of the human brain via deep learning,” *Human Brain Mapping*, vol. 41, no. 6, pp. 1505–1519, Apr. 2020, arXiv:1801.09858 [q-bio]. [Online]. Available: <http://arxiv.org/abs/1801.09858>
- [44] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, “Performance-optimized hierarchical models predict neural responses in higher visual cortex,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8619–8624, Jun. 2014. [Online]. Available: <https://pnas.org/doi/full/10.1073/pnas.1403112111>
- [45] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, “Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks,” *Machine learning in medical imaging. MLMI (Workshop)*, vol. 10541, pp. 362–370, Sep. 2017.
- [46] W. Li, X. Lin, and X. Chen, “Detecting Alzheimer’s disease Based on 4D fMRI: An exploration under deep learning framework,” *Neurocomputing*, vol. 388, pp. 280–287, May 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0925231220301041>
- [47] C. Zhao, H. Li, Z. Jiao, T. Du, and Y. Fan, “A 3D Convolutional Encapsulated Long Short-Term Memory (3DConv-LSTM) Model for Denoising fMRI Data,” *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 12267, pp. 479–488, Oct. 2020.
- [48] A. W. Thomas, H. R. Heekeren, K.-R. Müller, and W. Samek, “Analyzing Neuroimaging Data Through Recurrent Deep Learning Models,” Apr. 2019, arXiv:1810.09945 [cs, q-bio, stat]. [Online]. Available: <http://arxiv.org/abs/1810.09945>
- [49] M. Pominova, A. Artemov, M. Sharaev, E. Kondrateva, A. Bernstein, and E. Burnaev, “Voxelwise 3D Convolutional and Recurrent Neural Networks for Epilepsy and Depression Diagnostics from Structural and Functional MRI Data,” in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. Singapore, Singapore: IEEE, Nov. 2018, pp. 299–307. [Online]. Available: <https://ieeexplore.ieee.org/document/8637478/>
- [50] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, and WU-Minn HCP Consortium, “The WU-Minn Human Connectome Project: an overview,” *NeuroImage*, vol. 80, pp. 62–79, Oct. 2013.
- [51] P. Sidén, *Scalable Bayesian spatial analysis with Gaussian Markov random fields*, 09 2020.
- [52] T. Nakai, N. Koide-Majima, and S. Nishimoto, ““music genre fmri dataset”,” 2021.
- [53] F. F. Labs, “Supercharging classification - the value of multi-task learnin,” June 2018.
- [54] V. N. Salimpoor, D. H. Zald, R. J. Zatorre, A. Dagher, and A. R. McIntosh, “Predictions and the brain: how musical sounds become rewarding,” *Trends in Cognitive Sciences*, vol. 19, no. 2, pp. 86–91, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364661314002538>
- [55] T. Nakai, N. Koide-Majima, and S. Nishimoto, “Correspondence of categorical and feature-based representations of music in the human brain,” *Brain and Behavior*, vol. 11, no. 1, p. e01936, 2021. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/brb3.1936>
- [56] F. Pestilli and L. Kitchell, “Brainlife,” 2017. [Online]. Available: <https://brainlife.io/>
- [57] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, M. Snyder *et al.*, “fmriprep: a robust preprocessing pipeline for functional mri,” *Nature methods*, vol. 16, no. 1, pp. 111–116, 2019.
- [58] V. Fonov, A. C. Evans, K. Botteron, C. R. Almlri, R. C. McKinstry, and D. L. Collins, “Unbiased average age-appropriate atlases for pediatric studies,” *NeuroImage*, vol. 54, no. 1, pp. 313–327, Jan 2011.
- [59] A. Angulo-Perkins, W. Aubé, I. Peretz, F. A. Barrios, J. L. Armony, and L. Concha, “Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians,” *Cortex; a journal devoted to the study of the nervous system and behavior*, vol. 59, pp. 126–137, Oct. 2014, place: Italy.
- [60] V. D. Angelis, F. D. Martino, M. Moerel, R. Santoro, L. Hausfeld, and E. Formisano, “Cortical processing of pitch: Model-based encoding and decoding of auditory fMRI responses to real-life sounds,” *NeuroImage*, vol. 180, pp. 291–300, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1053811917309278>
- [61] N. Staeren, H. Renvall, F. De Martino, R. Goebel, and E. Formisano, “Sound categories are represented as distributed patterns in the human auditory cortex,” *Current biology : CB*, vol. 19, no. 6, pp. 498–502, Mar. 2009, place: England.
- [62] P. McCarthy, “Fsleyes,” Aug. 2022. [Online]. Available: <https://doi.org/10.5281/zenodo.7038115>
- [63] R. C. Craddock, G. A. James, P. E. Holtzheimer, 3rd, X. P. Hu, and H. S. Mayberg, “A whole brain fMRI atlas generated via spatially constrained spectral clustering,” *Hum Brain Mapp*, vol. 33, no. 8, pp. 1914–1928, Jul. 2011.