Hyper-HMM: A novel computational neuroscientific approach in studying neuroplasticity

Philip Tham
Advisor: Prof. Christopher Baldassano

Cognitive Science Senior Thesis 2025

Contents

1. Introduction

Neuroplasticity refers to the ability of the brain to undergo functional reorganization, and sensory loss studies—wherein participants who have diminished sensory capabilities are compared against a control group—provide insights into the specifics of this reorganization. In the deaf, neuroplasticity has been well-documented in the auditory cortex. Studies have shown evidence of the auditory cortex being recruited to perform visual tasks in deaf subjects but not hearing subjects [1, 2, 3]. Despite clear evidence of such neuroplasticity, the extent and specificities of visual system functional reorganization remain unresolved.

In particular, Zimmermann et al. [3] found that, when exposed to the same visual naturalistic stimuli with no audio, differences between the deaf and hearing were observed in secondary and higher-order auditory cortices. They used two analytical methods—an intersubject synchronization for an intact animated movie and gradually distorted variants of the same stimulus, and a data-driven Hidden Markov Model (HMM) to derive underlying temporal structures of neural responses to the movie—to analyze neuroplasticity. The first approach was used to test whether different auditory areas occupied different positions in the cognitive processing hierarchy in the deaf. They observed that secondary (but not primary) auditory cortices synchronized significantly more in deaf than hearing individuals, and became less synchronized as the meaning of the stimulus was distorted by scrambling. In the second approach, they used the HMM to measure the lengths of semantically-related groupings of time points (“events”) across brain regions. The HMM revealed a coherent event structure in the secondary auditory cortices of the deaf at slow and intermediate timescales. Taken together, the results of the synchronization and HMM suggested that deaf secondary auditory cortices get recruited to some higher-level semantically-related visual function.

Neuroplasticity studies, such as Zimmermann et al., employ intersubject correlation (ISC) to assess neural synchrony. However, ISC measures alignment at a fast time point-by-time point scale rather than looking for alignment within longer clusters of time points that might be a single semantic unit (i.e. an “event”). Moreover, standard ISCs are unable to test what kinds of stimulus features are driving synchronization

This problem with ISC analyses extends to a broader challenge in neuroscientific research of how to preserve heterogeneous dynamics (either between individuals or between groups) while capturing shared cognitive processes. While exposed to the same stimuli and undergoing the same cognitive processes, group-level differences can result in the same semantic concept being represented at varying points in time or by different spatial patterns of neural activity [4]. In other words, while measuring correlation across brains, two other exogenous variables are in flux—first, the exact timepoints in which neural activity is to be expected in a particular region, and second, the exact contours of the brain region in which activity is expected to occur.

One line of research has focused on controlling for *temporal* differences across participants. In particular, the use of Hidden Markov Models (HMMs) abstracts individual timepoints into broader semantic units (“events”) [5], where the boundaries between events may not temporally align

precisely across different individuals. Comparisons between participants based on events, however, assumes that event-specific spatial activity patterns across voxels are identical across participants [4]. Another approach aims to control for *spatial* variability by learning a functional alignment (ie a hyperalignment) across subjects by projecting neural data from two sources onto a lower dimensional shared latent space, accounting for differences in their brain's functional organization [6]. Yet, such approaches assume temporal synchrony in neural responses [4].

A Hyper-Hidden Markov Model (HMM) introduced by Lee et al. [4] stands as one possible solution to this issue. It combines aspects of HMMs and hyperalignment techniques to simultaneously account for both temporal and spatial differences across participants, allowing for more robust correlation comparison across subjects. While ISC analyses assume fixed temporal correlations, a H-HMM models brain activity as a sequence of latent states (semantically loaded “events”) that change over time, and it can model such sequences jointly across subjects. However, the H-HMM is still novel and its use cases have not fully been explored.

In this study, we test the applicability of H-HMM for neuroplasticity studies. Using H-HMM, we seek to determine how well activity in the deaf auditory cortex can be explained by encoding models that generate predictions of neural activity in visual brain regions based on a visual stimulus. High correlation between the deaf auditory cortex and the visual encoding models would, firstly, cohere with existing literature on auditory cortex neuroplasticity and recruitment for visual tasks, and second, validate the use of H-HMM for other intersubject correlation tasks in computational neuroscience. To this end, this study computes three sets of z-scores: first, correlation between neural activity of the whole deaf brain and visual encoding models; second, correlation between neural activity of the whole hearing brain and visual encoding models as a baseline; and third, the difference in correlation between the deaf and hearing to examine which areas of the deaf brain are better explained by the visual encoding models than the hearing brain. Correlations are measured across three sub-regions of the visual cortex, namely, the early visual cortex (EV), the parahippocampal place area (PPA), and the lateral occipital complex (LOC). Neural activity in these three regions of interest (ROIs) are generated using img2fmri, a visual system neural encoding model.

2. Methods overview

2.1 Participants and fMRI results

fMRI data from Zimmerman et al. [2] was used for data analysis. In that study, 21 early deaf participants and 22 hearing participants were exposed to a 35 minute extract of the animated cartoon *Triplets of Belleville* with no audio. MRI structural and functional data of the whole brain were collected on a 3 Tesla Siemens MAGNETOM Tim Trio scanner with minimal preprocessing being conducted using fmriprep. The brain data was split into 100 parcels according to the parcellation of the cerebral cortex based on functional connectivity identified in Schaefer et al. [7].

2.2 Generation of stimulus feature embeddings using img2fmri encoding model

img2fmri is a python package encoding model for predicting group-level fMRI responses to visual stimuli using deep neural networks [8]. It uses an artificial deep neural network that first learns to extract features (shapes, textures etc) from naturalistic visual data that allows the model to classify objects, and then uses those extracted features from the input to predict cortical responses. The img2fmri package models neural activity in five ROIs in the visual cortex, namely the early visual (EV) region, the retrosplenial cortex (RSC), the occipital place area (OPA), the parahippocampal place area (PPA), and the lateral occipital complex (LOC) in the typical (ie hearing) visual cortex.

The same 35 minute extract of the *Triplets of Belleville* naturalistic stimuli that Zimmerman et al. (2024) exposed subjects to was input into the img2fmri model for the EV, PPA, and LOC ROIs under the same preprocessing conditions as the Zimmerman et al. study (ie assuming 1035 time points of 1.4 seconds each) [4]. This process generated a total of three 1035 by n matrices (one for each ROI), where n is the number of voxels for each ROI preset by the img2fmri package. The correlation between fMRI neural activity from actual subjects in the Zimmerman et al. study and the three img2fmri encoding models was then examined using the H-HMM model.

Instead of directly comparing the neural activity between the deaf and hearing participants from the Zimmerman et al. study, we chose to generate stimulus feature embeddings of the typical hearing individual using the img2fmri encoding model and then correlated these embeddings with the neural activity of participants. We did so for two reasons: first, to remove further sources of individual variability and use img2fmri encoding models as a standard baseline to compare the participants to, and two, to individually isolate ROIs and examine which specific parts of the visual cortex were correlating well with the deaf participants.

2.3 Hyper-Hidden Markov Model (H-HMM)

2.3.1 H-HMM overview

The H-HMM was introduced in Lee et al. as a method to simultaneously capture temporal and spatial differences across participants, thereby removing the assumptions of spatial or temporal alignment across individuals [4]. The method is a variant of a Hidden Markov Model (HMM) as it assumes that a hidden structure—a particular sequence of semantic states (i.e. events) composed of clusters of time points with some relation to each other—exists beneath surface observations of neural activity at specific time points, and attempts to find the hidden sequence of transitions between event states. In so doing, the H-HMM mitigates the temporal alignment problem as it does not assume that the specific time points between the fMRI neural activity and the stimulus feature embeddings are exactly aligned. Rather, it merely assumes that all subjects and stimulus models proceed through the same sequence of events. See Figure 1 below for an example from our results on how event segmentation works. To mitigate the spatial alignment problem, the H-HMM also employs hyperalignment techniques that map different neural data—in this case, participant fMRI data and the img2fmri stimulus feature embeddings—onto a shared low-dimension representational latent space. In other words, instead of looking at specific brain areas, the H-HMM looks only at principal components of the neural data, which capture the most significant variance in activity across individuals, and aligns these components into a common space.

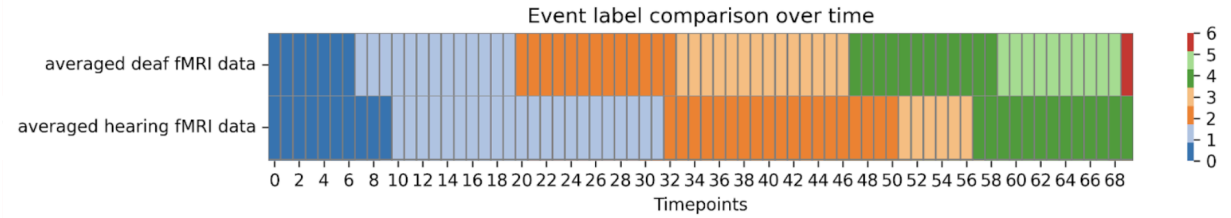

Figure 1: Event segmentation of the first 70 timepoints for deaf and hearing subjects in parcel 63 (mid-superior temporal sulcus, a higher-order auditory region). Each time point is 1.4 seconds. In the H-HMM, the forward-backward algorithm segments the time points into events of variable length based on how similar neural activity at the time point is to other timepoints in the fMRI data and the visual encoding model predictions.

Put simply, the H-HMM finds an event-based alignment across fMRI data and the stimulus feature embeddings. In this study, **the fMRI data consists of whole brain fMRI activity averaged across participants in each deaf/hearing category,** while **the stimulus feature embeddings are the predictions of visual system neural activity generated by the EV, PPA, and LOC img2fmri encoding models.**

For each brain parcel in the subjects' fMRI data, the H-HMM iteratively estimates a temporal alignment between the subject fMRI neural data and the predictions, and updates the shared latent space representation (G, intialized as the mean across all projected data) and the

transformation matrix for each subject and stimulus feature embedding (W_i, intialized randomly) to project the fMRI and embedding data onto the latent space. Specifically:

- First, the fMRI (averaged by deaf/hearing category) and embedding data (the EV, PPA, and LOC predictions) are projected onto a latent space using the W from the previous iteration (hyper-alignment technique).
- Second, the forward-backward algorithm from Baldassano et al. [5] is used to segment each timeseries into events corresponding to patterns in G (HMM technique).

- Third, new event patterns are estimated based on the average of fMRI activity across timepoints predicted to be in the same event. These new event patterns are then concatenated together and PCA is performed to reduce the large matrix to the provided shared dimensionality, forming an updated G.

- Fourth, update the transformation matrices W_i for the subject and stimulus feature embedding using a ridge regression (with alpha as a hyperparameter to be tuned) to predict the updated G from the subject and stimulus embedding patterns.

These steps are iterated until the model's log-likelihood stops improving, yielding the learnt transformation matrices, that is, the matrices that project the fMRI data and stimulus embeddings into a shared latent space where event patterns are defined.

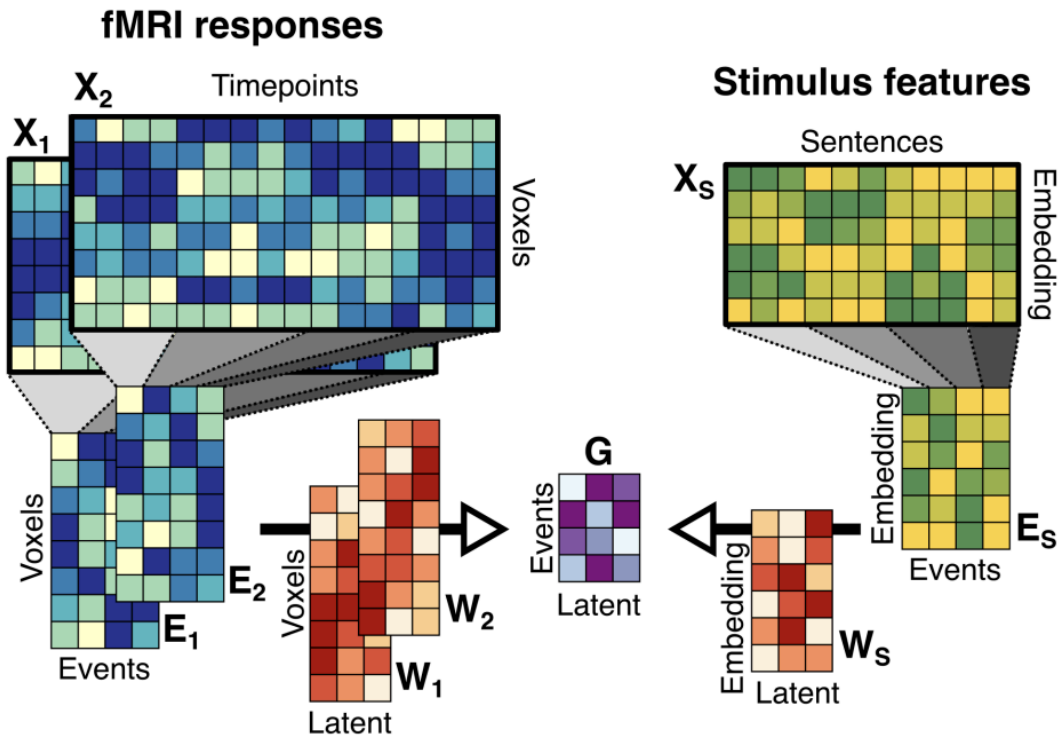

Figure 2: Hyper-HMM alignment across multiple brains and stimulus features, adapted from Lee et al. [4].
For this study, the subjects' fMRI data was averaged out across deaf/hearing categories and each

img2fmri encoding model was input to the H-HMM at a time, so there is only one pair of fMRI response and stimulus feature embedding. In this diagram, the H-HMM temporally divides each subject's brain data into discrete events with subject-specific patterns E_i and temporally divides the stimulus embedding into E_s. The event patterns from the subjects and the stimulus are linearly projected through matrices W_i (one matrix per subject or stimulus feature embedding) to a shared low-dimensional latent space representation G. In this paper, the stimulus feature embedding refers to the predictions of visual system neural activity generated by the img2fmri EV, PPA, and LOC encoding models.

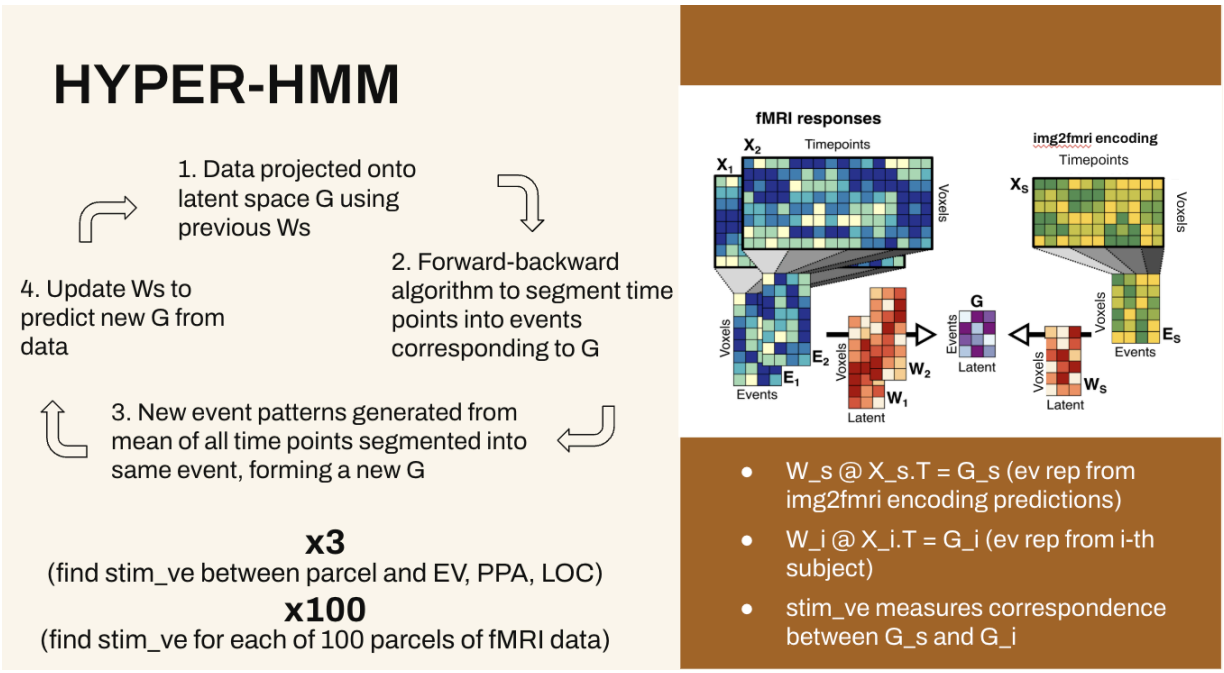

Figure 3: H-HMM fitting loop across all encoding models and brain parcels. The averaged deaf/hearing subject fMRI data and visual encoding model predictions are input into the H-HMM and the four steps are looped through until convergence. At convergence, a stim_ve score is obtained that measures correspondence between event representations generated by the fMRI subject data and that generated by the encoding model predictions. Higher stim_ve means greater correlation, i.e. that the visual models are better able to predict neural activity in the parcel. This process is repeated for each visual encoding model (EV, PPA, and LOC), and for all 100 brain parcels.

To compare the correlation between the subjects' fMRI neural activity in the brain parcel and the stimulus feature embeddings, a variance explained measure (stim_ve) is calculated for each parcel of the subjects' brain. **This measure quantifies how well the stimulus event representations match the corresponding event representations derived from the fMRI data.** In other words, the corresponding transform matrices are applied to the stimulus embeddings and fMRI data, and the resulting lower-dimensional latent space event representations are compared. See Figure 8 for a comparison of event representations in three-dimensionality latent space (but with only the first two dimensions shown) for parcel 63 of deaf and hearing subjects and the img2fmri EV encoding model. The variance explained by the variability between clusters is calculated using the formula

$$stim\_ve = 1 - \frac{within\_cluster\_variance}{total\_variance}$$

The total variance reflects the difference between each stimulus event representation and the overall average of all fMRI event representations, while the within-event variance reflects the difference between each stimulus event representation and the average fMRI event representation for that same event. Higher stim_ve values indicate that the clustering of fMRI event representations more closely aligns with the stimulus embedding event representations, suggesting a stronger correlation between the two and more meaningful clustering.

2.3.2 H-HMM hyperparameter selection

Several H-HMM hyperparameters were determined based on the Zimmermann et al. study. Based on data from the study, the 35 minute *Triplets of Belleville* extract contained 70 events spread across 1035 timepoints of 1.4 seconds each, with each event lasting around 14 timepoints (1035//70). The number of events and event length were used as hyperparameters in the H-HMM.

A key hyperparameter to be tuned in the H-HMM model was the ridge regression alpha value that penalizes non-zero coefficients when updating the transform matrices W_i. To find the optimal alpha value, the stim_ve values generated from correlating a toy brain parcel—parcel 86 in the auditory cortex of deaf (d86) and hearing (h86) participants—and the img2fmri stimulus feature embeddings were calculated across a range of alpha values. The alpha value that generated the maximum stim_ve across all tested alpha values was selected as the optimal value for the H-HMM. Although different embeddings produced different optimal values, alpha = 1 was determined by visual inspection as the optimal value for the H-HMM across the three encoding models.

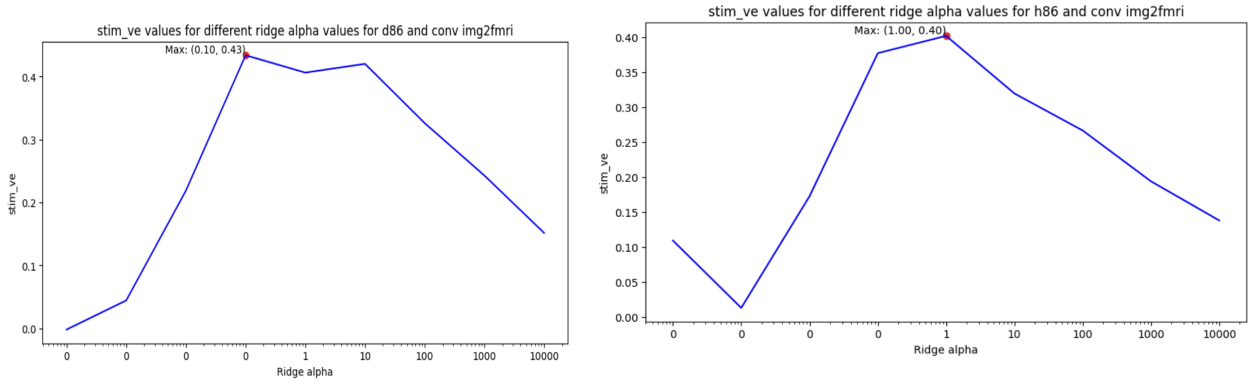

Figure 4: Hyperparameter tuning for alpha value for ridge regularization for the img2fmri EV model on toy parcel 86 (d86 for the deaf, h86 for the hearing). While the peaks for d86 and h86 differ, visual inspection of the graphs indicates that an alpha value of 1 would work well for both the deaf and hearing.

Another hyperparameter to configure was whether to maximize the model's log probability for each individual iteration or the combined log probability across all iterations as the stopping condition for the H-HMM learning algorithm. Maximizing individual log probability would process each iteration independently with no assumption about temporal alignment, while accumulating log probabilities across iterations assumes some temporal alignment between the fMRI data and stimulus feature embeddings. Ultimately, we decided to assume temporal alignment between the fMRI and embedding data and use combined log probability because there was greater stability in stim_ve values across cross validation folds when using combined as opposed to individual log probabilities. The pseudocode for how the combined log probability was calculated is below.

```
Unset
#EventSegment.fit() in event.py

while step <= max_num_iterations:
	segmentation_probs = []
	combined_logprob = None

	for stim_video in stim_videos:
		for subject in subjects:
			calculate logprob_i
			if combined_logprob is None:
				combined_logprob = logprob_i
			else:
				combined_logprob = np.logaddexp(combined_logprob,
	logprob_i)
				#combine multiple log prob vectors in log space
			log_gamme, log_likelihood =
	forward_backward(combined_logprob)
	...
```

The dimensionality of the shared latent space representation was another hyperparameter to be chosen. After preliminary analysis, we realized that the optimum dimensionality that produced the highest stim_ve value might differ from parcel to parcel, which would make sense as different ROI in the brain might represent events at varying levels of complexity. We thus decided to run the H-HMM across a range of nine dimensionalities (from 2D to 10D) and then select the dimension that produced the highest stim_ve score as the optimal dimension.

2.3.3 Calculation of stim_ve value for each parcel

To calculate the stim_ve value for each parcel, a k-fold cross validation algorithm was carried out on the fMRI data from the deaf and hearing data separately. This method was used to obtain

a more stable and robust estimate of stim_ve by averaging across four folds for easier interpretation.

At each fold, the subject fMRI data was split into a training and validation set, with the fMRI data in each set being averaged out across participants. The H-HMM model was first fit on the averaged training set and the stimulus feature embeddings, yielding transformation matrices that project fMRI data and stimulus features into a shared low-dimensional latent space. These learned transformation matrices were then applied to the averaged validation set, projecting the held-out fMRI and stimulus embedding data into the same latent space. Note that new event boundaries for the validation data were computed using Brainiak's EventSegment function. The resulting latent-space representations were then evaluated using the stim_ve variance explained measure to assess the alignment between stimulus representations and fMRI event representations. The average stim_ve value across the cross validation folds was taken to be the stim_ve for the run. This cross validation was repeated across a range of nine dimensionalities, and the dimensionality which produced the highest stim_ve value was selected to be the stim_ve for the parcel. Below is a pseudocode of the general steps of this algorithm.

```
Unset
#fmri_parcellated_data as 100 (number of brain parcels) x num_subj x 1035
(number of timepoints) x num_voxels array either for deaf or hearing group
#img2fmri_stim_embeddings as the stimulus feature embeddings generated by
img2fmri for either EV, PPA, or LOC ROI

def evaluate_hhmm(fmri_parcellated_data, img2fmri_stim_embedding):
	highest_stim_ve_per_parcel = []
	dimensions = range(2,11) #dimensions range from 2 to 10

	for parcel in fmri_parcellated_data:
		avg_score_per_dim = []

		for dimension in dimensions:
			#perform splitting, cross validation, evaluation
			cv_scores = custom_cv(parcel, img2fmri_stim_embeddings)

			#get avg score across cv folds
			avg_score_across_folds = mean(cv_scores)
			avg_score_per_dim.append(avg_score_across_folds)

		#get highest stim ve across dimensions
		highest_stim_ve = max(avg_score_per_dim)

		highest_stim_ve_per_parcel.append(highest_stim_ev)

	return highest_stim_ve_per_parcel
```

2.4 Data analysis and statistical testing

The raw stim_ve scores of each parcel were converted into z-scores, with the z-scores thresholded at 1.645 (5% significance level) to determine the significance of correlation between parcels in the subjects' brains and the img2fmri encoding models. To generate these z-scores, the mean and standard deviation of stim_ve values for each parcel from a random null distribution were obtained.

To generate a random null distribution for the z-scores of the stim_ve values between deaf or hearing participants and each img2fmri ROI encoding model (EV, PPA, LOC), the alignment between subject fMRI data and the img2fmri embeddings was jumbled up, deliberately misaligning the temporal alignment between the fMRI and embedding data. Random cyclic rotations of the fMRI data across timepoints were used to create this misalignment. The jumbled up fMRI data and stimulus embeddings were then fed into the H-HMM model and cross validated to generate stim_ve values. This process was repeated across 50 runs to generate 50 stim_ve values, from which a null mean and standard deviation could be derived. Each parcel's z-score was calculated using the classic z-score formula:

$$Z = \frac{parcel\_stim\_ve - null\_mean}{null\_std}$$

Second, we also generated z-scores to measure the difference in stim_ve between deaf and hearing subjects across brain parcels. This was done to determine which brain parcels exhibited the greatest difference in correlation with the img2fmri encoding models between the deaf and the hearing, or in other words, how well the img2fmri models explained deaf neural activity as opposed to the hearing.

To generate the random null distribution in this case, we randomly shuffled the labels of the deaf and hearing participants, creating new "deaf" and "hear" groups of subjects that did not necessarily fit their category. For each shuffled group, the difference between the "deaf" and "hearing" labelled groups were calculated using the H-HMM and cross validation algorithm. This process was also repeated with 50 random shuffles to create the null distribution of the difference of stim_ve values under the assumption that there is no real difference between the deaf and hearing participants. Each parcel's z-score was similarly calculated using the following formula:

$$Z = \frac{(parcel\_stim\_ve\_deaf - parcel\_stim\_ve\_hear) - null\_mean}{null\_std}$$

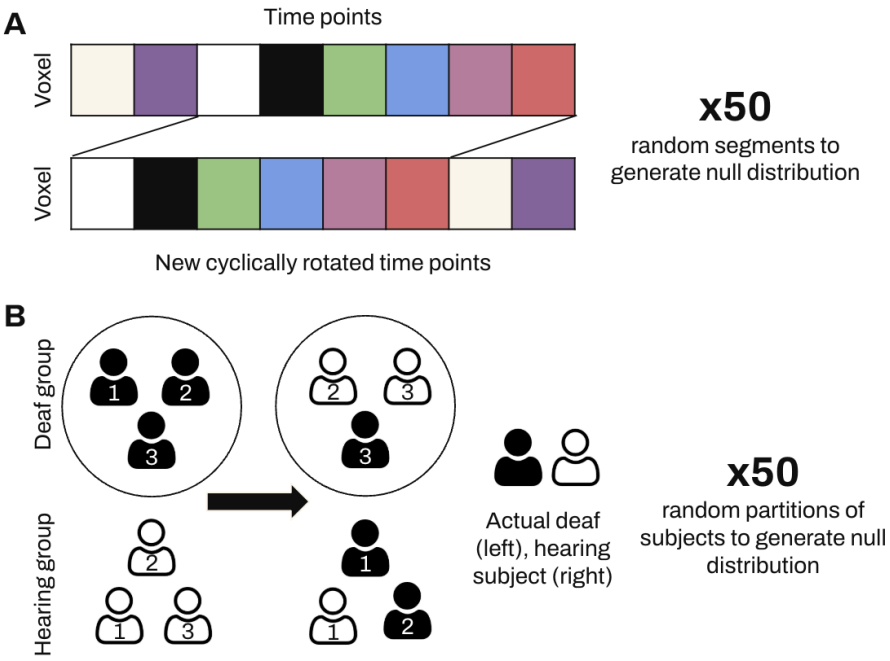

Figure 5: Schematic of cyclic rotation and shuffle logic to generate null distributions. A illustrates the logic behind a cyclic rotation to deliberately misalign fMRI data and img2fmri-generated visual model predictions to create the null distribution. B shows the random partitioning of deaf and hearing subjects into new "deaf" and "hearing" groups to generate a null distribution for the deaf minus hearing stim_ve measurement.

2.5 Projection onto brain

Overall, nine brain projections were made—for each of the three img2fmri encoding models investigated (EV, PPA, LOC), z-scores of the correlation between the whole deaf brain and the encoding model, z-scores for the correlation between the whole hearing brain and the encoding model, and z-scores of the difference in correlation between fMRI and encoding data between the deaf and hearing brains for each of the three img2fmri encoding models investigated. These brain projections were made using nilearn's plotting functionality (view_img and plot_img_on_surf). Only z-scores above the 1.645 (p=0.05 uncorrected) threshold were projected onto the brain.

2.6 Compute resources

To generate the null distributions for z-score calculation, Columbia University's Ginsburg High Performance Computer cluster was used. For each parcel, generating the null distribution for the deaf and hearing participants with 50 cyclic rotations involved 100 Slurm jobs in total and took around three hours and around 10GB of memory. Generating the null distribution for the difference between deaf and hearing stim_ve values required assigning an individual Slurm job

for each of the 50 random shuffles and 100 parcels, resulting in 5000 Slurm jobs. Each job took around 25 minutes and required 25GB of memory.

3. Results

Data analysis was split into two parts. First, a **sanity check** was conducted to check for correspondence[1] between the img2fmri models and actual brain data from the deaf and hearing subjects. Since the encoding models were trained to predict neural activity in the EV, PPA, and LOC of hearing individuals, it remained an open question as to whether neural activity in the three same regions of deaf subjects could also be predicted by the img2fmri encoding models, and whether the H-HMM would be able to pick up on such correlations. Second, **regions where the visual predictions were able to explain activity better in the deaf than the hearing were identified**. Regions where visual predictions more effectively explained activity in deaf individuals may suggest a reorganization of visual processing functions to those areas. This was done by measuring the relative differences between deaf and hearing subjects in their correlations with the three visual cortex encoding models. Several other non-auditory parts of the brain might have correlated activity with the visual cortex. In order to specifically identify which parts of the deaf brain showed significantly *more* correlation with the visual cortex than the hearing brain, the relative difference between the two categories in img2fmri correlation was thus calculated. Areas of high relative difference indicate that these regions were significantly more correlated to the visual encoding models in the deaf versus the hearing. A key barometer of the success of the H-HMM model in identifying neuroplasticity was whether the auditory cortex was a region exhibiting high relative difference in correlation since existing literature indicates that the deaf auditory cortex is recruited to some visual functions.

In the first part of the analysis (the “sanity check”), the H-HMM was able to pick up significant correlations (p=0.05 uncorrected) between predictions of the EV, PPA, and LOC encoding models and the corresponding ROIs in both deaf and hearing subjects. This is illustrated in Figure 6 below where, across both categories of subjects, areas of high correlation with the three visual encoding models overlap with the actual ROI areas. This result indicates that, first, the img2fmri predictions—trained to model typical visual cortex activity—were also generalizable to deaf subjects; and second, the H-HMM was indeed able to pick up correlations between encoding model predictions and fMRI data.

While the correlations between the EV, PPA, and LOC of both classes of subjects and the encoding models were to be expected, the encoding models were also able to explain activity in other regions of the subjects. First, the encoding models were able to explain activity in brain regions commonly associated with the processing of visual scenes. One such region is the medial place area [9], which exhibits correlation with the EV and PPA models in both the deaf and hearing (see medial views of the deaf and hearing brain correlations with EV and PPA predictions). These correlations were also to be expected and served as another sanity check for the H-HMM and img2fmri encoding models. Second, correlations were found between the models and the auditory cortices of both the hearing and the deaf. Further analysis on the

[1] Note that “correspondence” and “correlation” refer to significant stim_ve values that measure the variance explained or alignment between event representations generated from the img2fmri visual encoding model predictions and those generated from the subject fMRI data.

relative difference in strength of such correlations between the deaf and the hearing is done below.

Standardized z-scores of correlations between deaf/hearing subject fMRI data and img2fmri EV/PPA/LOC encoding model predictions

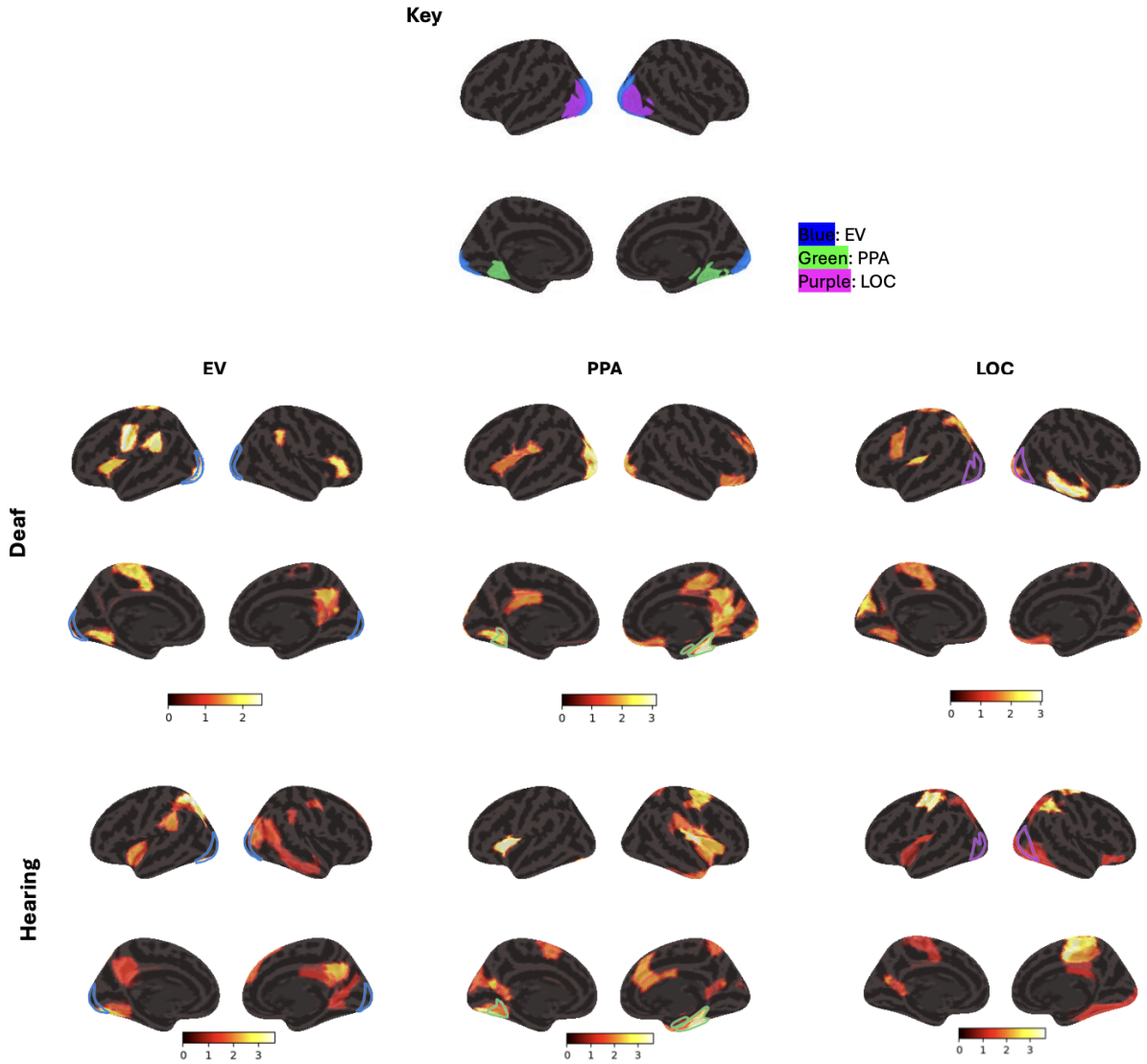

Figure 6: Standardized z-scores of correlations between deaf/hearing subject fMRI data and img2fmri EV/PPA/LOC encoding model predictions. Note that "correlation" here refers to the standardized stim_ve score which measures the alignment of representations generated by the img2fmri visual predictions and those generated by the fMRI subject data. Only regions with z-scores above 1.645 (5% significance level) are shown. The EV, PPA, and LOC are traced in blue, green, and purple respectively on the brain images. Overlaps indicate that the img2fmri visual encoding model predictions correspond to actual observed brain activity. Top brain projections show a lateral view and bottom projections show a medial view.

In the second part of the analysis, we identified regions wherein the img2fmri visual encoding model predictions were able to explain neural activity better in the deaf than the healing. This

was done by measuring the relative difference between the deaf and hearing in terms of correlation with the img2fmri visual encoding models. Regions of the auditory cortex were defined in the same way as Zimmermann et al., with the transverse temporal gyrus (Te, also commonly referred to as Heschl's gyrus) being considered the primary auditory cortex, the secondary regions of the Te being considered the secondary auditory cortex, and the mid-superior temporal sulcus (STS) being considered the higher order auditory cortex.

Based on the results of the relative differences in correlation generated by the H-HMM, only the EV and LOC visual encoding models were able to explain auditory cortex activity significantly better in the deaf than the hearing. Note that the z-score threshold for significance was also 1.645 (p=0.05 uncorrected). For the EV, the secondary and higher auditory cortices of the right hemisphere exhibited higher correlation with the visual predictions in the deaf versus the hearing. For the LOC, the secondary and higher auditory cortices of both hemispheres exhibited higher correlation. Apart from auditory regions, the LOC encoding model predictions also correlated more strongly with deaf brain activity in the prefrontal cortex. In terms of investigation visual function neuroplasticity in the deaf auditory cortex, these findings suggest: first, activity in the EV corresponds to activity in the right secondary and higher auditory cortices significantly more in the deaf than the hearing; and second, activity in the LOC corresponds to activity in both left and right secondary and higher auditory cortices significantly more in the deaf than the hearing. Curiously, the H-HMM was unable to pick up any significantly high relative difference in correlations between any part of the subjects' auditory cortex and the encoding model generated PPA predictions. The PPA model predictions, however, correlated strongly with deaf brain activity in the angular gyrus region versus the hearing brain [10]. These findings are illustrated in Figure 7 below.

Standardized z-scores of difference between deaf and hearing subjects in terms of correlation with img2fmri visual encoding model predictions

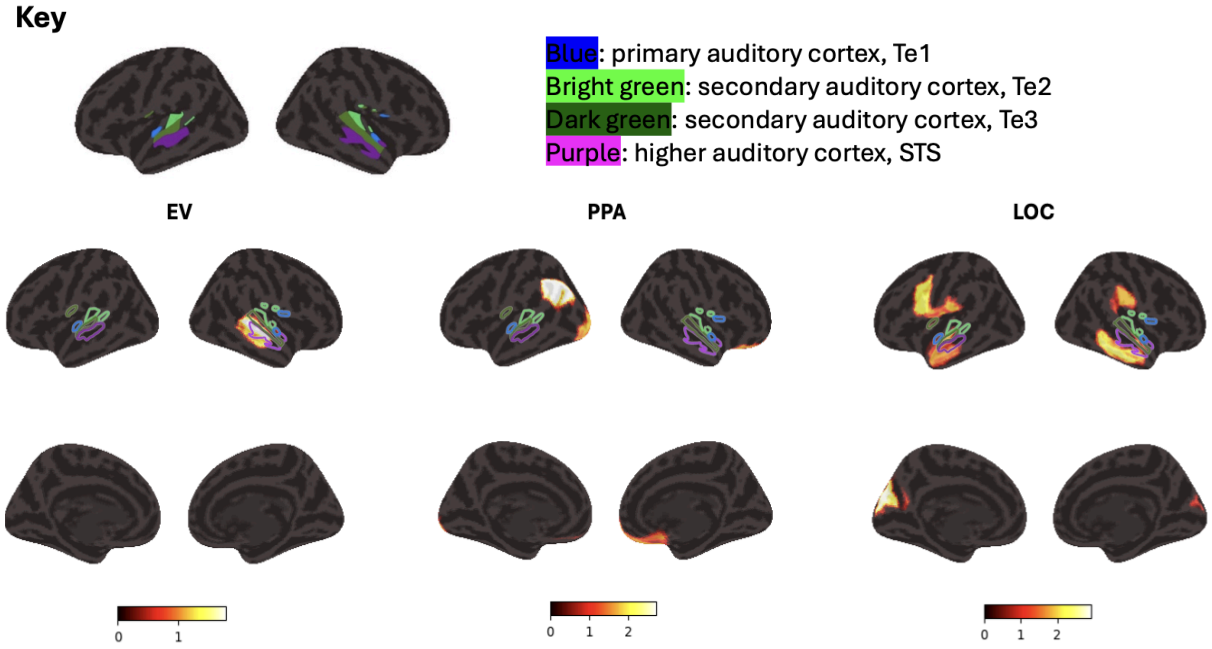

Figure 7: Relative difference in correlation with img2fmri encoding models between deaf and hearing subjects. Regions shown are those that visual models can explain better in deaf participants vs hearing participants. In the key, the auditory regions cannot be seen from a medial view and thus these views are omitted. The primary, secondary, and higher auditory cortices are traced out in the brain projections below. Top projections show a lateral view and bottom projections show a medial view.

From the results above, there seems to be significantly greater correlation between the higher-order STS and EV encoding model in the deaf than the hearing (Figure 7, purple traced outline in the EV brain projection). Event representations generated by the H-HMM from the STS (parcel 63) of both deaf and hearing fMRI data and those generated from encoding model predictions were plotted in Euclidean space in Figure 8 below to further illustrate this observation. While the shared latent space representation in the illustration below was three-dimensional (the optimal dimensionality for both hearing and deaf subjects was three based on a sweep across a range of dimensionality values from two to ten), only the first two dimensions of this are shown for simplicity. The Euclidean distance (across all three dimensions) between the event representations generated from the fMRI data and the encoding models were calculated for each event, and eight events with the lowest Euclidean distance for the deaf and hearing subjects were plotted below. The average Euclidean distance across all events (i.e. the average distance between the event representation generated from the fMRI data and the encoding mode) for the deaf was 1129.64, while the stim_ve value across all events was 0.16. The average Euclidean distance across all events in the hearing was 1166.18, and the stim_ve value across all events was 0.09. Taken together, the H-HMM captures the observation that the EV encoding model is able to better explain fMRI activity in the deaf STS

(d63) as opposed to the hearing STS (h63). This suggests that the deaf STS is recruited to visual functions in ways that the hearing STS is not, which could be interpreted as evidence supporting neuroplasticity.

Comparison of eight H-HMM generated event representations when correlating parcel 63 with img2fmri EV encoding model predictions

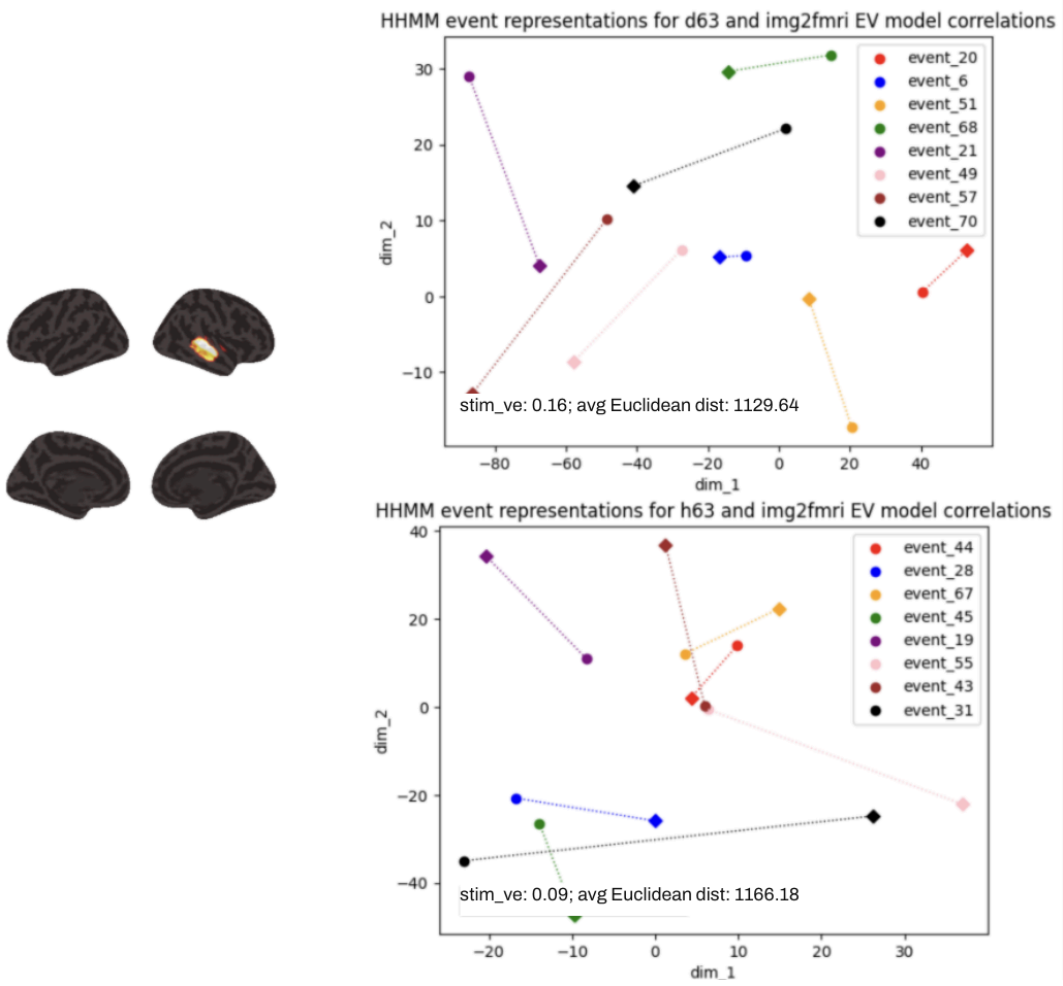

Figure 8: Comparison of eight H-HMM-generated event representations when correlating parcel 63 of the deaf (d63) and hearing (h63) with img2fmri EV encoding model. The location of parcel 63 is shown on the left. Circle markers indicate event representations generated from subject fMRI data, while diamond markers indicate representations generated from the img2fmri encoding models. The stim_ve and the average Euclidean distance between the event representations across all events are shown in the figure as well.

4. Discussion

H-HMM and img2fmri data-driven approach

Based on the first part of data analysis above (the “sanity check”), deaf EV, PPA, and LOC activity can still be modeled by encoding models like img2fmri, which are trained to predict brain activity in neurotypical individuals. **This indicates that these models generalize reasonably well even when applied to non-neurotypical populations. Importantly, H-HMM is also able to detect correlations between these predicted activations and the actual fMRI data in deaf individuals.** The successful application of H-HMM, in conjunction with encoding models, to conduct this neuroplasticity study acts as a proof-of-concept that such novel computational neuroscientific techniques can control for temporal and spatial alignment assumptions while shedding light on the specificities of functional reorganization that are still hitherto unclear.

Selective neuroplasticity across visual function and auditory cortex levels

On the second part of data analysis investigating the use of H-HMM in modelling neuroplasticity, the H-HMM was able to detect correlations between visual neural activity and specific regions of the auditory cortex that are significantly stronger in the deaf than in the hearing population. These “neuroplastic correlations” suggest that parts of the deaf auditory cortex are recruited to perform visual functions. The H-HMM detected such correlations specifically between: (i) the EV visual predictions and the right STS (a higher-order auditory region) and right Te3 (secondary auditory region), and (ii) the LOC visual predictions and both left and right STS and Te3. In other words, the EV explains activity in the right secondary and higher auditory cortices better in the deaf than the hearing, while the LOC explains activity in both the left and right secondary and higher auditory cortices better in the deaf as well. This suggests that the visual functions performed by the EV (low-level visual feature representation) could be supported by the right secondary and higher auditory cortices in the deaf, and that those performed by the LOC (object recognition) could also be supported by the left and right secondary and higher auditory cortices. Interestingly, no auditory regions were explained by the PPA significantly better in the deaf than the hearing, suggesting that PPA visual functions (scene feature representation) might not be delegated to the auditory cortex in the deaf.

No neuroplastic correlations were found between any visual ROI and Te1 (primary auditory cortex). This finding aligns with Zimmermann et al., which also reported that the primary auditory cortex does not appear to engage in visual processing, with neuroplasticity restricted to secondary and higher-order auditory areas [2].

Taken as a whole, **our findings similarly suggest that only the secondary and higher auditory cortices are engaged in visual system neuroplasticity. Further, it seems that only functions associated with the EV and LOC are reorganized into the deaf auditory cortex. Visual processing performed by the PPA does not seem to be reassigned.** Additional studies are needed to confirm this pattern and delineate its underlying mechanisms.

Hemispheric heterogeneity of neuroplasticity

Another interesting question that warrants further study is whether visual neuroplasticity in the deaf auditory cortex occurs bilaterally, is lateralized, or varies depending on the functional demands of the visual task. Zimmermann et al. observed hemispheric asymmetries in their HMM analyses, finding longer processing time-scales in the left hemisphere with slower event transitions [2]. Such an observation suggested that higher-order visual processing in left auditory areas in the deaf. In our data, correlations between the EV (lower level visual area) and regions in the deaf auditory cortex seem to appear only in the right hemisphere. Correlations between the LOC (mid level visual area) and the auditory cortex occur bilaterally. **Taken together, these findings suggest that neuroplasticity in the deaf auditory cortex may show different patterns of lateralization depending on the specific visual function it is recruited to support.** Broader analyses incorporating more visual ROIs, perhaps starting with other ROIs available in the img2fmri model like the RSC and OPA, and a larger participant pool could help test this hypothesis.

Further investigation

While these analyses demonstrate how H-HMM can be used to model and investigate neuroplasticity, further research is needed to evaluate its generalizability and robustness in other contexts. First, our current approach **assumes temporal alignment across events** between the visual encoding models and fMRI data. Although this assumption allows for flexible alignment at the level of individual timepoints (i.e. individual time points did not need to match exactly, rather the time points were grouped into broader events with semantic meanings, and these events were them aligned) and yielded stable results, it limits flexibility in modeling across events. Second, our approach also **pools and averages fMRI data from multiple subjects** and uses that as a single input matrix to compare against the img2fmri-generated visual encoding models. This means that we were only looking at group-level correlations with the encoding models, even though the H-HMM architecture allows for multi-subject and multi-stimuli comparisons. Future work could relax these two constraints and perform sanity checks on the H-HMM under conditions without fixed temporal alignment and without pooling/averaging, thereby assessing the model's adaptability and reliability in less structured settings.

H-HMM results also showed that **visual encoding models were also able to explain activity in some non-auditory regions better in the deaf than the hearing**, specifically in the correlation between the PPA model and the angular gyrus, and between the LOC model and some prefrontal cortex regions. Such an analysis suggests that these regions correspond to visual cortex activity more strongly in the deaf than the hearing, which could also be indicative of potential regions of visual functional reorganization, although more research is needed to investigate this.

Another area that could be ripe for further investigation is **whether visual system neuroplasticity in the deaf auditory cortex is hierarchical**. The visual system processes information in a hierarchy fashion, with the EV processing low-level features, and the PPA and LOC processing higher-level, more abstract features. Zimmermann et al. and this paper have found that the primary auditory cortex seems not to be recruited for any visual function in the

deaf. Perhaps secondary and higher-level visual functions could be reorganized into specific regions of the deaf auditory cortex, replicating the hierarchical structure of the visual system in the post-sensory auditory cortex. Our findings suggest that EV and PPA functions might be supported by the secondary and higher auditory regions, but more research is needed to back up the hypothesis that visual system neuroplasticity is hierarchical.

5. References

[1] Bavelier D, Dye MW, Hauser PC. Do deaf individuals see better? Trends Cogn Sci. 2006 Nov;10(11):512-8. doi: 10.1016/j.tics.2006.09.006. Epub 2006 Oct 2. PMID: 17015029; PMCID: PMC2885708.

[2] Zimmermann M, Mostowski P, Rutkowski P, Tomaszewski P, Krzysztofiak P, Jednoróg K, Marchewka A, Szwed M. The Extent of Task Specificity for Visual and Tactile Sequences in the Auditory Cortex of the Deaf and Hard of Hearing. J Neurosci. 2021 Nov 24;41(47):9720-9731. doi: 10.1523/JNEUROSCI.2527-20.2021. Epub 2021 Oct 18. PMID: 34663627; PMCID: PMC8612642.

[3] Zimmermann, M., Cusack, R., Bedny, M. et al. Auditory areas are recruited for naturalistic visual meaning in early deaf people. Nat Commun 15, 8035 (2024). https://doi.org/10.1038/s41467-024-52383-6.

[4] C. Lee, J. Han, M. Feilong, G. Jiahui, J. Haxby, C. Baldassano. Hyper-hmm: aligning human brains and semantic features in a common latent event space
Adv. Neural Inf. Process. Syst., 36 (2024).

[5] C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, and K. A. Norman. Discovering event structure in continuous narrative perception and memory. Neuron, 95(3):709–721, 2017.

[6] J. V. Haxby, J. S. Guntupalli, A. C. Connolly, Y. O. Halchenko, B. R. Conroy, M. I. Gobbini, M. Hanke, and P. J. Ramadge. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron, 72(2):404–416, Oct. 2011. doi: 10.1016/j.neuron.2011.08.026. URL https://doi.org/10.1016/j.neuron.2011.08.026.

[7] Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb. Cortex 28, 3095–3114 (2018).

[8] Bennett M, Baldassano C. img2fmri: a python package for predicting group-level fMRI responses to visual stimuli using deep neural networks. Aperture Neuro. 2023;3. doi:10.52294/001c.87545.

[9] Epstein RA, Baker CI. Scene Perception in the Human Brain. Annu Rev Vis Sci. 2019 Sep 15;5:373-397. doi: 10.1146/annurev-vision-091718-014809. Epub 2019 Jun 21. PMID: 31226012; PMCID: PMC6989029.

[10] Humphreys GF, Lambon Ralph MA, Simons JS. A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. Trends Neurosci. 2021 Jun;44(6):452-463. doi: 10.1016/j.tins.2021.01.006. Epub 2021 Feb 18. PMID: 33612312.