Voxel-wise Encoding of Visual and Social Meaning During Silent Movie Viewing in Deaf and Hearing Participants.

Maria Zimmermann[1,2], Marcin Szwed[3], Leyla Isik[1], Marina Bedny[1]
(1) Johns Hopkins University, (2) University of Warsaw, (3) Jagiellonian University

Abstract

In deaf individuals, higher auditory regions such as the superior temporal cortex are thought to be repurposed for visual processing. In previous study we showed that these regions are recruited for processing rich visual meaning during silent films (Zimmermann et al., 2024). To investigate which specific features drive this reorganization, we applied a voxel-wise encoding model to fMRI data from deaf and hearing participants as they watched a silent animated movie.

The model included a range of visual, social, and affective features. It significantly explained variance across the whole brain in both groups, revealing patterns consistent with prior findings. In regions that showed differential intersubject synchronization between groups, we observed higher prediction performance scores (R^2) in deaf participants. This was particularly evident for social interaction features in the right superior temporal sulcus (STS) and Theory of Mind features in the right posterior STS. These findings suggest that reorganization in the temporal cortex in deaf individuals may reflect an expansion of nearby visual and social feature representations into formerly auditory regions.

Background

Deaf individuals rely on vision to interpret the world, drawing from a rich and dynamic visual environment. To support this shift, the brain appears to reorganize, with auditory regions adapting to process visual input. Prior work has shown that parts of the temporal cortex—typically involved in hearing—can support low-level visual perception such as motion tracking (Benetti et al., 2021) as well as higher-level functions like visual working memory (Cardin et al., 2018). However, it remains unclear what types of visual information these reorganized regions represent.

In a previous study (Zimmermann et al., 2024) we used a silent film to explore this question in naturalistic conditions. They found that higher-order auditory areas in deaf individuals responded to complex visual content, with stronger right-hemisphere engagement for low-level features. Building on this, we applied a voxel-wise encoding model to fMRI data from deaf and hearing participants viewing a silent film, modeling brain responses to a range of visual, social, and affective features.

Method

Participants.

Twenty-five deaf participants and twenty-two hearing participants watched a 35-minute silent animated film (*The Triplets of Belleville*) in the fMRI scanner. The film contained no language, and the soundtrack was removed.

Feature Annotation.

We automatically extracted low-level visual properties, including motion energy and HSV color components. Three raters labeled each 4-second (3 TRs) segment for the presence of faces, social interactions, number of people, arousal, and valence. Theory of mind content was assessed by three additional raters on longer (~20 TR) segments by indicating whether the events elicited reasoning about others' mental states.

Encoding Model.

To relate brain activity to specific movie features, we trained a voxel-wise encoding model to estimate beta weights linking the annotated features to fMRI BOLD responses. We used linear banded ridge regression with a nested cross-validation scheme (Nunez-Elizalde et al.,2019). The model was implemented using Himalaya package (Dupré La Tour et al., 2022)

Results

Visual, social, and affective features accounted for variance in brain responses across both deaf and hearing participants. In both groups, each feature engaged regions associated with its representation in prior work. Specifically, consistent with prior research, social interaction features were represented in the right superior temporal sulcus (STS) (Lee Masson & Isik, 2021), Theory of Mind in the right temporoparietal junction (TPJ) and precuneus (Saxe & Kanwisher, 2003).

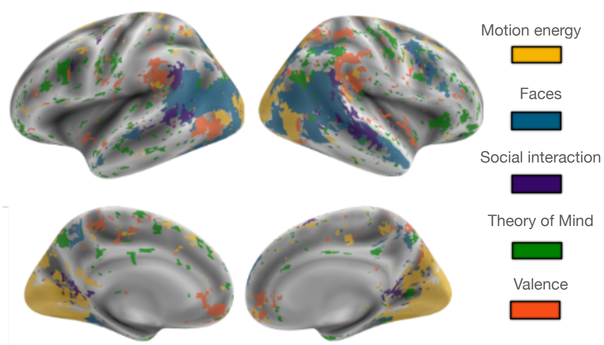

Figure 1. Winner-takes-all maps showing the top predictive feature (visual, social, affective) in each voxel in both groups. Maps include only voxels with significant prediction scores (R^2) in the full model **(**$p < 0.001$**,** uncorrected).

We next examined whether higher-order 'auditory' regions showing differential synchronization between deaf and hearing individuals—identified in a previous ISC analysis (Figure 2)—also showed differences in feature representation across groups.

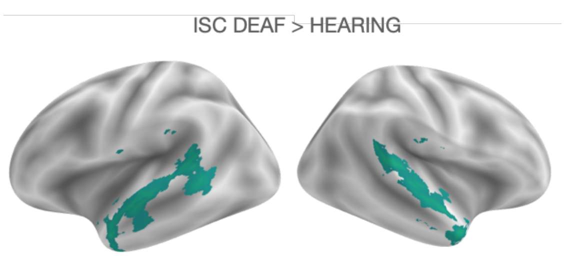

Figure 2. Group differences in intersubject correlation (ISC) -Deaf > Hearing during silent movie watching ($p < 0.001$, uncorrected).

In the anatomical location of higher order 'auditory' regions, for the deaf group, we observed higher explained variance for features typically represented in neighboring superior temporal cortex. Deaf individuals showed greater variance explained by social interaction features in the right superior temporal sulcus (STS), and stronger representation of Theory of Mind in the posterior STS.

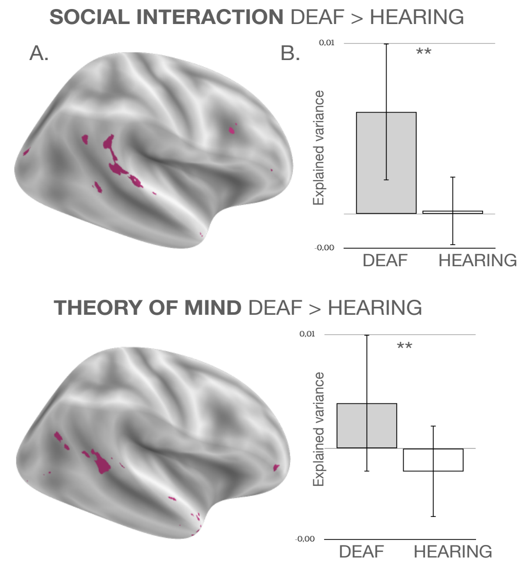

Figure 3. Features showing differential representation in the deaf and hearing groups**:** social interaction and **T**heory of Mind (A) Voxels showing significantly higher R^2 in the deaf group than in the hearing group across the whole brain, $p<0.01$ uncorrected. (B) Group and ROI averaged performance scores (R^2) for social interaction and Theory of Mind within voxels showing between group difference in ISC (Fig.2)

Conclusion

Using a voxel-wise encoding model during silent movie viewing, we found higher prediction performance (R^2) in deaf individuals within regions that showed differential intersubject synchronization between groups. Specifically, social interaction and Theory of Mind features explained more variance in deaf participants within higher-order auditory regions. These findings are consistent with the possibility that, in the absence of auditory input, visual and social processing may extend into nearby regions of the superior temporal cortex that are typically associated with auditory functions.

References

Benetti, S., Zonca, J., Ferrari, A., Rezk, M., Rabini, G., & Collignon, O. (2021). Visual motion processing recruits regions selective for auditory motion in early deaf individuals. *NeuroImage*, *230*, 117816. https://doi.org/10.1016/j.neuroimage.2021.117816

Cardin, V., Rudner, M., De Oliveira, R. F., Andin, J., Su, M. T., Beese, L., Woll, B., & Rönnberg, J. (2018). The Organization of Working Memory Networks is Shaped by Early Sensory Experience. *Cerebral Cortex*, *28*(10), 3540–3554. https://doi.org/10.1093/cercor/bhx222

Dupré La Tour, T., Eickenberg, M., Nunez-Elizalde, A. O., & Gallant, J. L. (2022). Feature-space selection with banded ridge regression. *NeuroImage*, *264*, 119728. https://doi.org/10.1016/j.neuroimage.2022.119728

Lee Masson, H., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage*, *245*, 118741. https://doi.org/10.1016/j.neuroimage.2021.118741

Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). *Voxelwise encoding models with non-spherical multivariate normal priors*.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in “theory of mind.” *NeuroImage*, *19*(4), 1835–1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Zimmermann, M., Cusack, R., Bedny, M., & Szwed, M. (2024). Auditory areas are recruited for naturalistic visual meaning in early deaf people. *Nature Communications*, *15*(1), 8035. https://doi.org/10.1038/s41467-024-52383-6