Finding Music Genre Classification Mechanism in Human Brain

Baekho Jang, Jiwoo Seo | Supervisor: Prof. Sungshin Kim
Major in Cognitive Science, Department of Data Science, Hanyang University

Motivation

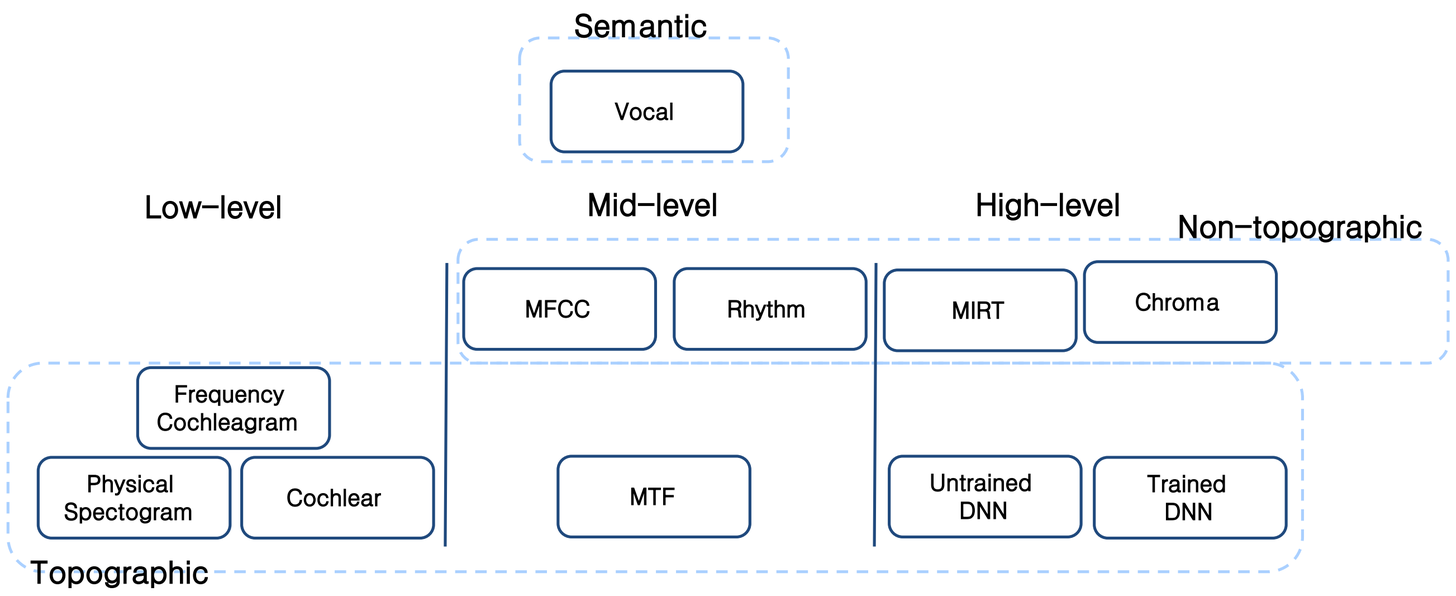
We aim to identify the criteria the brain uses to classify music genres by examining whether neural representations are driven by physical “Sound” or perceptual “Hearing,” through comparisons of timbre, modulation, and rhythm.

Introduction

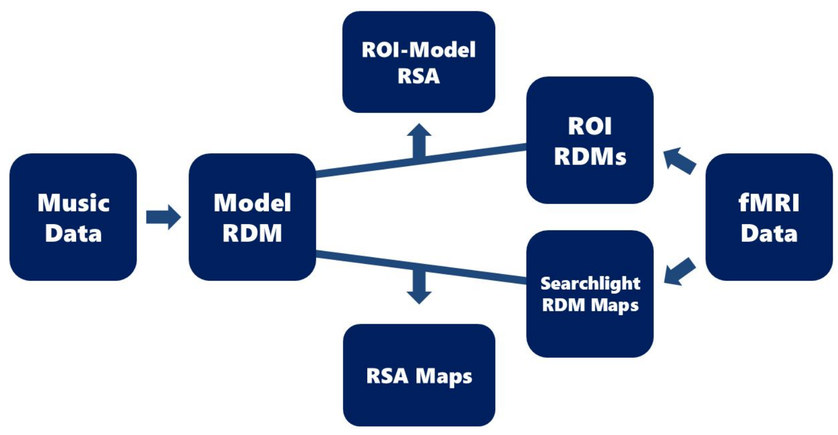
Dataset: Music Genre Neuroimaging Dataset (Nakai et al., 2021). fMRI data were collected from 5 participants listening to 540 music clips across 10 genres (e.g., Rock, Jazz, Classical).

Related Work: Nakai et al. (2022) identified the Modulation Transfer Function (MTF) as the superior encoding model, highlighting the Superior Temporal Gyrus (STG) as a key region for genre processing. Additionally, we implemented a Task-Optimized DNN (based on Kell et al., 2018) to compare hierarchical deep features with biological auditory models.

Models



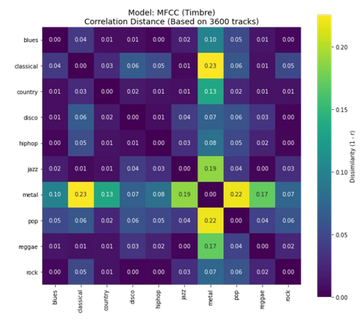
Analysis Pipeline



The pipeline extracts feature-based RDMs and correlates them with brain activity patterns via RSA.

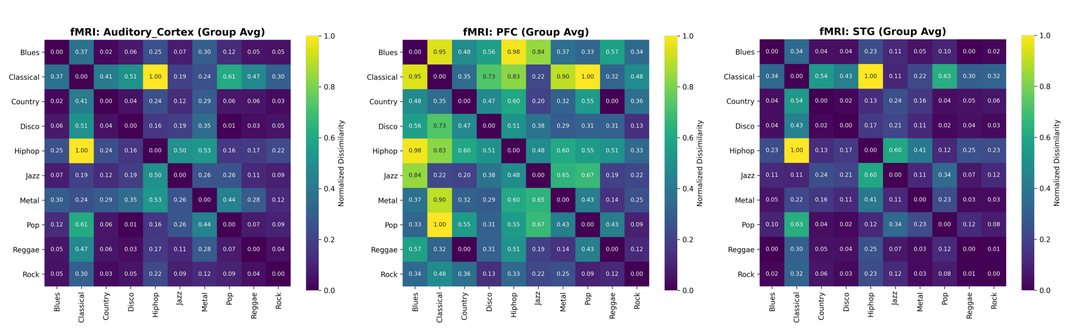
RDMs

Model RDM



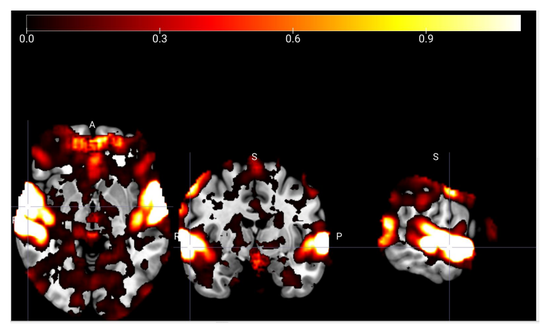
Theoretical dissimilarity matrices constructed by calculating acoustic feature distances.

ROI RDMs



Neural representations quantified by analyzing fMRI BOLD signals within specific Regions of Interest (ROIs).

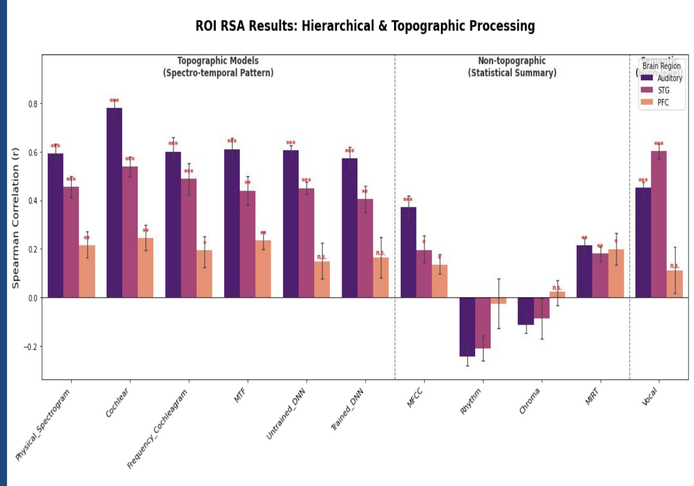
Searchlight RDM Maps



Spatial maps generated by whole-brain searchlight analysis to visualize the global distribution of neural patterns.

RSA Results

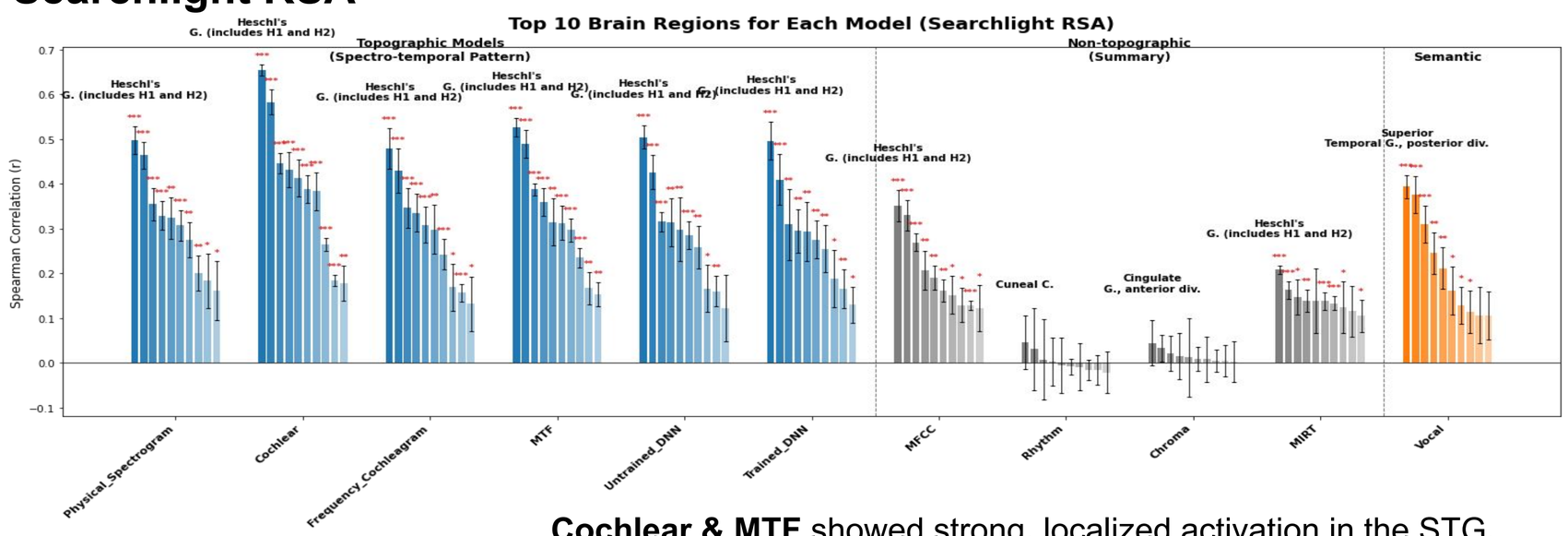
ROIs RSA



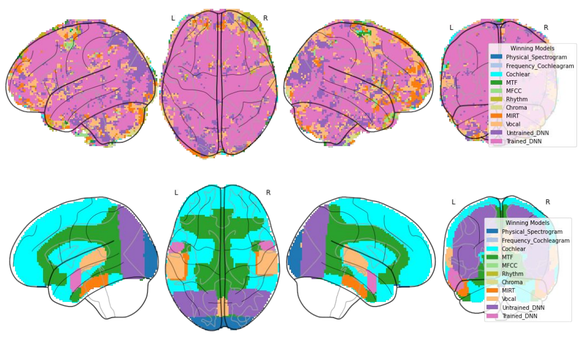
Spectro-temporal models (Cochlear, MTF, DNNs) significantly outperformed statistical summary models in the Auditory Cortex and STG.

The **Vocal model** showed high correlations in the STG, comparable to acoustic models, confirming specialized vocal processing.

Searchlight RSA



Winner Map



Cochlear & MTF showed strong, localized activation in the STG, reflecting sensitivity to **spectro-temporal modulations**.**Trained DNN & Vocal** extended into association areas (Precuneus, Frontal), indicating high-level **semantic processing**.

Concentric Hierarchy: Core (Topographic-Acoustic) => Belt (Topographic-Feature) => Association (Semantic/Learned). **Trained DNN** dominated cortical coverage, suggesting functional convergence between deep features and neural processing.

Previous Study Comparison

Consistent with **Nakai et al. (2022)**, results support the importance of pattern-based representations over statistical summaries in the STG.

Aligning with **Kell et al. (2018)**, the structural similarity of Untrained DNNs and the broad cortical expansion of Trained DNNs were observed.

Conclusion

Results suggest the brain may encode music as **dynamic spectro-temporal patterns** rather than static statistics.

The processing pipeline appears to evolve from early acoustic analysis to high-level **semantic inference**, where voice and learned features play key roles.

Future Work

Time-Resolved RSA: Future studies could use MEG/EEG to investigate millisecond-scale dynamics of music processing.

Layer-wise Analysis: Mapping specific DNN layers to cortical regions is needed to verify the direct correspondence between artificial and biological hierarchies.