



Aix*Marseille
université



Maastricht
University

Intermediate acoustic-to-semantic representations link behavioural and neural responses to natural sounds

*BL Giordano¹, M Esposito², G Valente², and E Formisano²
Nat Neurosci (2023)*

¹ Inst. Neurosci. La Timone, CNRS and Aix-Marseille University

² Dept. Cog. Neurosci., Maastricht University, Maastricht, the Netherlands.

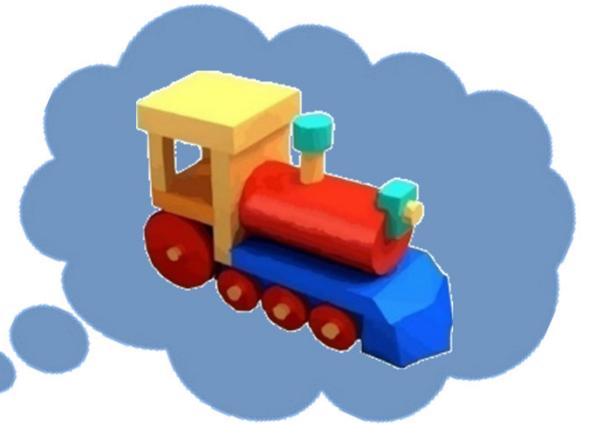
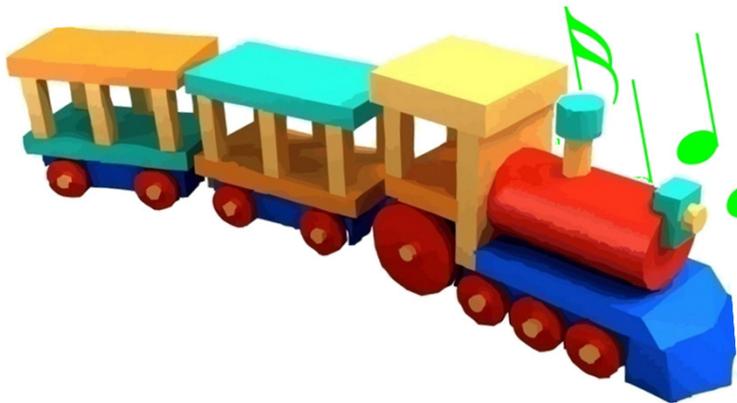
bruno.giordano@univ-amu.fr

e.formisano@maastrichtuniversity.nl



08/06/2023 @ CERIMED





Computational models: from acoustics to semantics

A

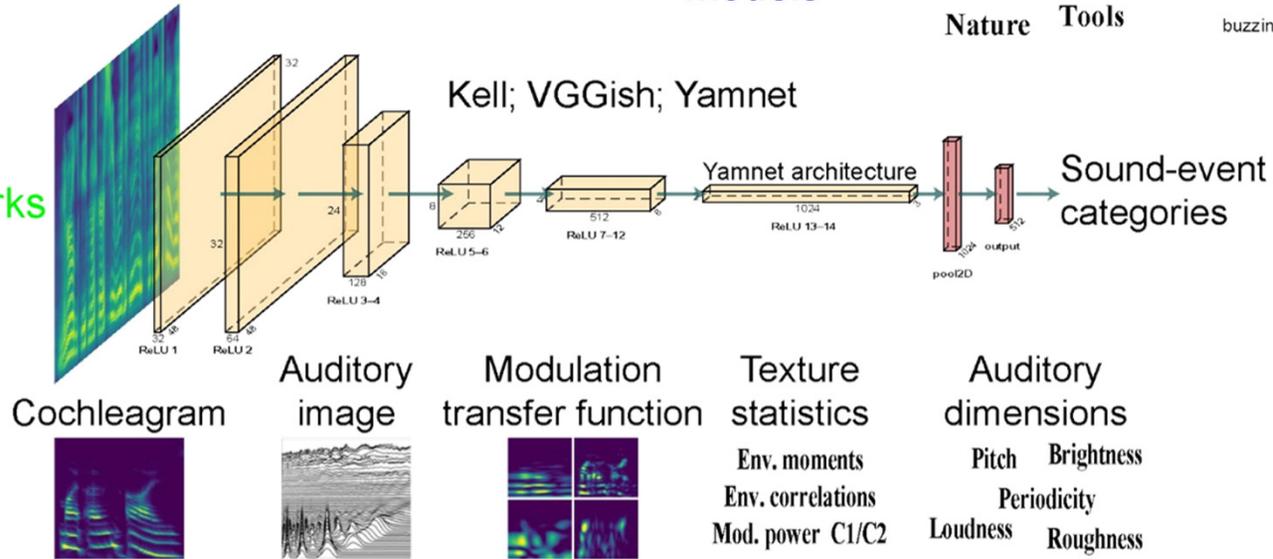
Verbal sound descriptors

Semantic models

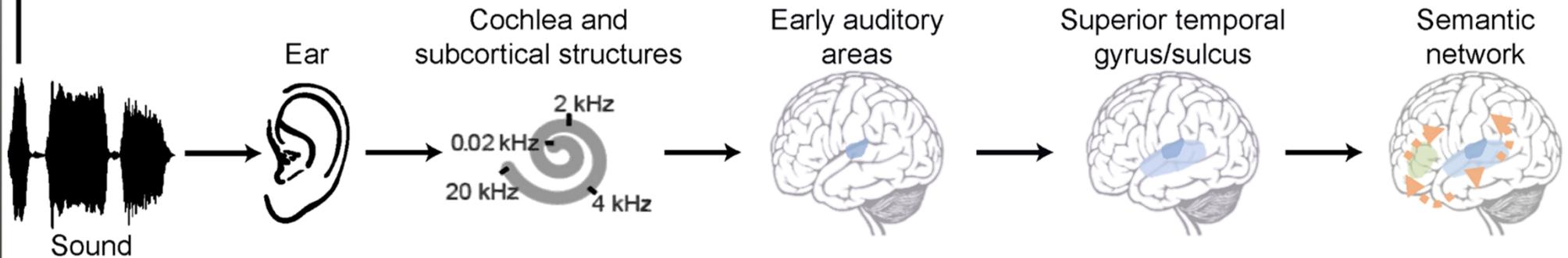
Categories
Speech Voices
Music Animals
Nature Tools

NLP embeddings:
w2v; GloVe; GUSE
meowing cat → [0.83;0.55;...;0.20]
buzzing bee → [0.43;0.64;...;0.23]

Sound-to-event deep neural networks



Acoustic models



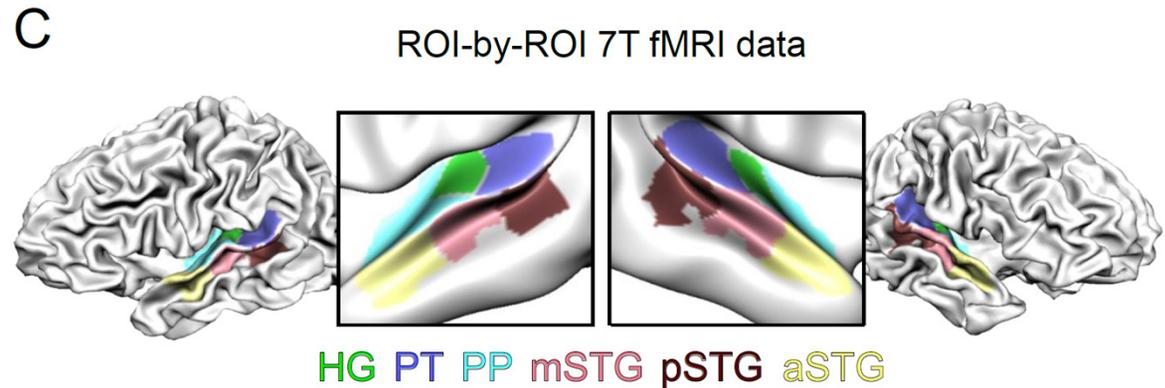
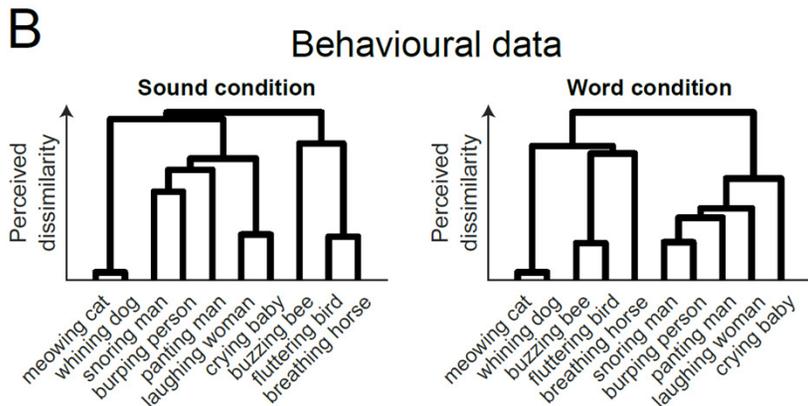
Datasets

Behaviour: Giordano et al. (2010)
80 stimuli, 20 participants

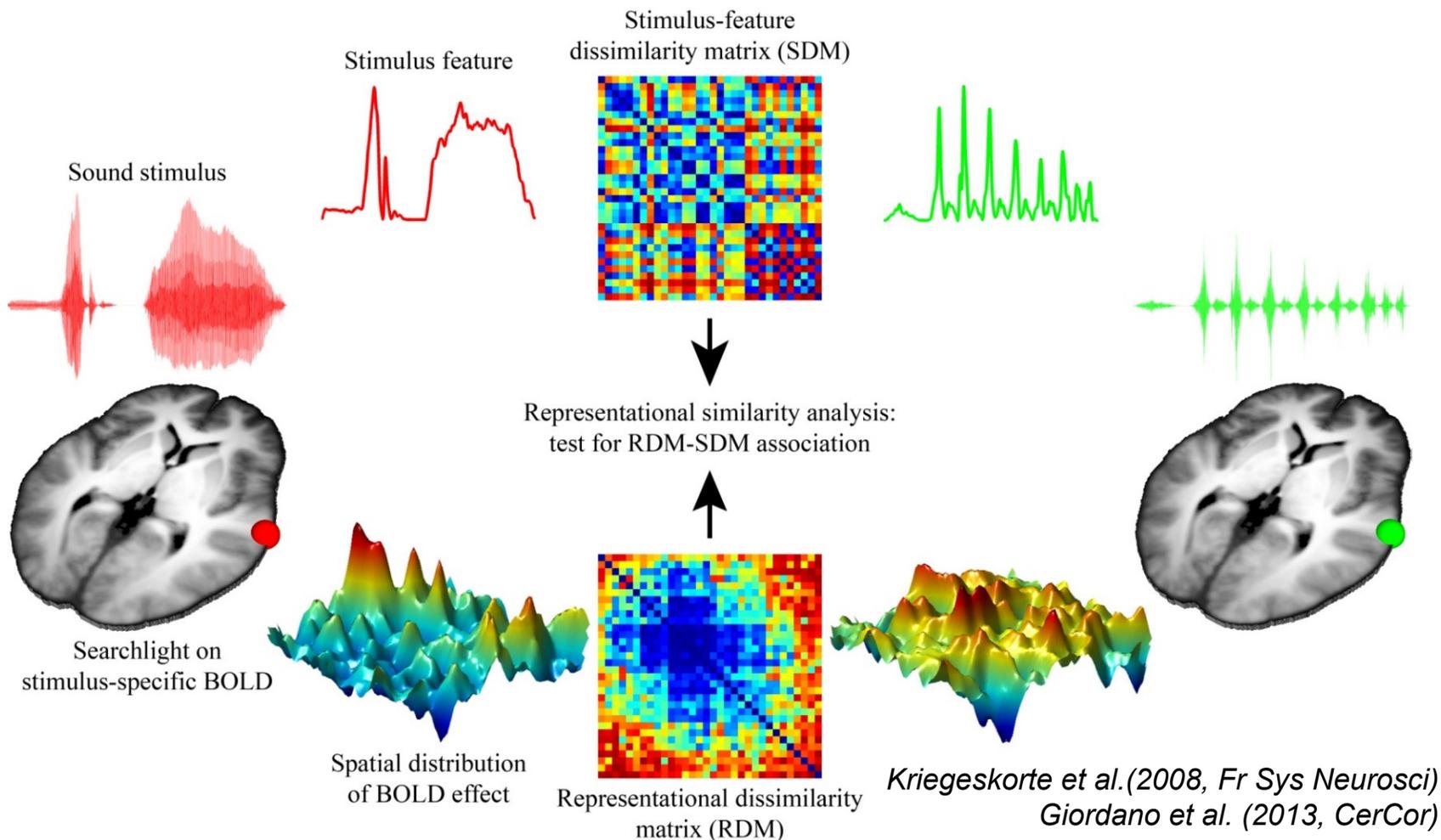
Task: Hierarchical sorting
Perceived dissimilarity
(sound/word)

7T fMRI: Santoro et al. (2017)
288 stimuli, 5 participants

Task: one-back repetition
detection



Cross-validated Representational Similarity Analysis



Training step

$$\begin{matrix}
 \text{fMRI/behaviour} \\
 \text{training distance}
 \end{matrix}
 = \beta_1 \begin{matrix} \text{model distance 1} \\ \text{(training stimuli)} \end{matrix} + \dots + \beta_n \begin{matrix} \text{model distance n} \\ \text{(training stimuli)} \end{matrix}$$

Testing/internal validation step

$$\begin{matrix}
 \text{fMRI/behaviour} \\
 \text{test-set prediction}
 \end{matrix}
 = \beta_1 \begin{matrix} \text{model distance 1} \\ \text{(test stimuli)} \end{matrix} + \dots + \beta_n \begin{matrix} \text{model distance n} \\ \text{(test stimuli)} \end{matrix}$$

$$R^2_{cv} = 1 - \frac{SSE_{test}}{SST_{test}}$$

Variance partitioning



COMMONALITY ANALYSIS: A METHOD FOR DECOMPOSING EXPLAINED VARIANCE IN MULTIPLE REGRESSION ANALYSES

DAVID R. SEIBOLD, ROBERT D. MCPHEE

First published: June 1979 | <https://doi.org/10.1111/j.1468-2958.1979.tb00649.x> | Citations: 14

Two Variables:

$$U(i) = R_{Y \cdot ij}^2 - R_{Y \cdot j}^2$$

$$C(ij) = R_{Y \cdot i}^2 + R_{Y \cdot j}^2 - R_{Y \cdot ij}^2$$

Three Variables:

$$U(i) = R_{Y \cdot ijk}^2 - R_{Y \cdot jk}^2$$

$$C(ij) = R_{Y \cdot ik}^2 + R_{Y \cdot jk}^2 - R_{Y \cdot k}^2 - R_{Y \cdot ijk}^2$$

$$C(ijk) = R_{Y \cdot i}^2 + R_{Y \cdot j}^2 + R_{Y \cdot k}^2 - R_{Y \cdot ij}^2 - R_{Y \cdot ik}^2 - R_{Y \cdot jk}^2 + R_{Y \cdot ijk}^2$$

Four Variables:

$$U(i) = R_{Y \cdot ijkm}^2 - R_{Y \cdot jkm}^2$$

$$C(ij) = R_{Y \cdot ikm}^2 + R_{Y \cdot jkm}^2 - R_{Y \cdot km}^2 - R_{Y \cdot ijkm}^2$$

$$C(ijk) = R_{Y \cdot im}^2 + R_{Y \cdot jm}^2 + R_{Y \cdot km}^2 - R_{Y \cdot m}^2 - R_{Y \cdot ijm}^2 - R_{Y \cdot ikm}^2 - R_{Y \cdot jkm}^2 + R_{Y \cdot ijkm}^2$$

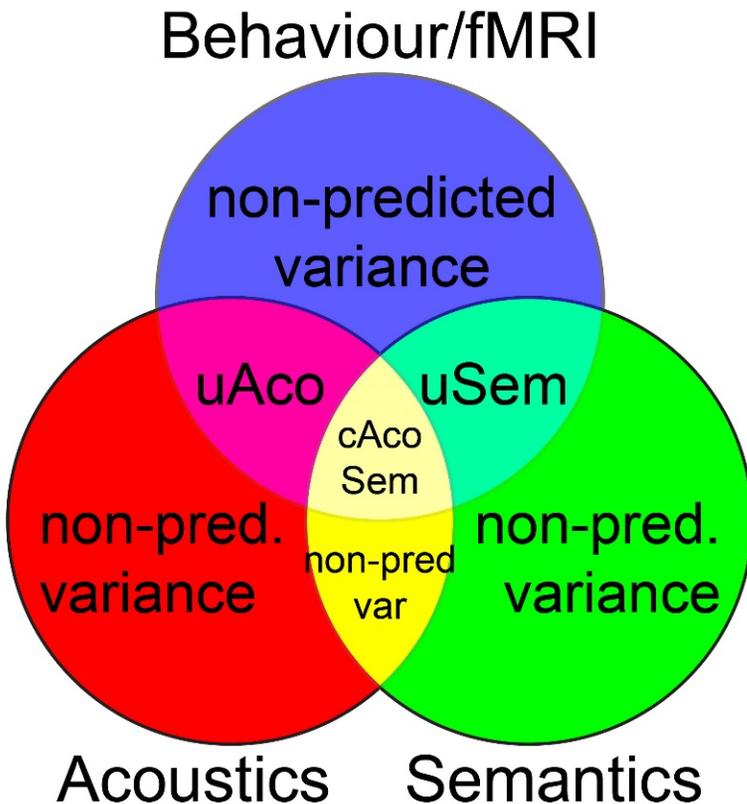
$$C(ijkm) = R_{Y \cdot i}^2 + R_{Y \cdot j}^2 + R_{Y \cdot k}^2 + R_{Y \cdot m}^2 - R_{Y \cdot ij}^2 - R_{Y \cdot ik}^2 - R_{Y \cdot im}^2 - R_{Y \cdot jk}^2 - R_{Y \cdot jm}^2 - R_{Y \cdot km}^2 + R_{Y \cdot ijk}^2 + R_{Y \cdot ijm}^2 + R_{Y \cdot ikm}^2 + R_{Y \cdot jkm}^2 - R_{Y \cdot ijkm}^2$$

Five Variables:

$$U(i) = R_{Y \cdot ijkmn}^2 - R_{Y \cdot jkmn}^2$$

$$C(ij) = R_{Y \cdot ikmn}^2 + R_{Y \cdot jkmn}^2 - R_{Y \cdot kmn}^2 - R_{Y \cdot ijkmn}^2$$

$$C(ijk) = R_{Y \cdot imn}^2 + R_{Y \cdot jmn}^2 + R_{Y \cdot kmn}^2 - R_{Y \cdot mn}^2 - R_{Y \cdot ijm}^2 - R_{Y \cdot ikm}^2 - R_{Y \cdot jkm}^2 - R_{Y \cdot jmn}^2 + R_{Y \cdot ijkmn}^2$$



$$R^2_{aco} = u_{Aco} + c_{AcoSem}$$

$$R^2_{sem} = u_{Sem} + c_{AcoSem}$$

$$R^2_{AcoSem} = u_{Aco} + u_{Sem} + c_{AcoSem}$$

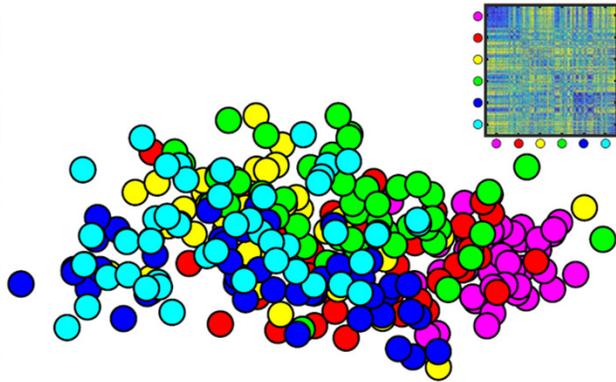
$$u_{Aco} = R^2_{AcoSem} - R^2_{Sem}$$

$$u_{Sem} = R^2_{AcoSem} - R^2_{Aco}$$

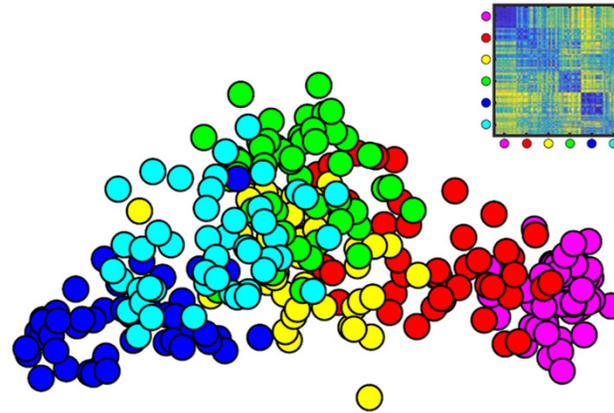
$$c_{AcoSem} = R^2_{Aco} + R^2_{Sem} - R^2_{AcoSem}$$

Acoustics to semantics in models and brain

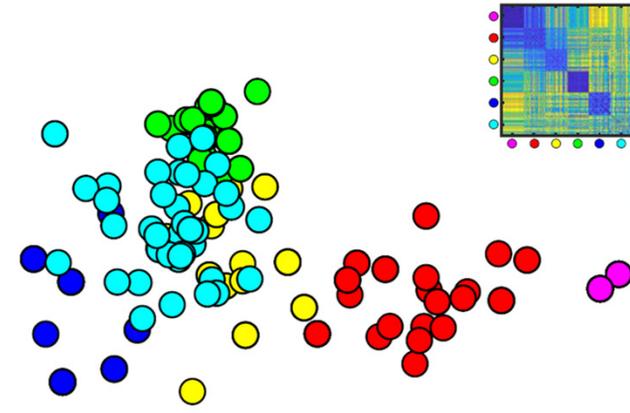
Acoustics



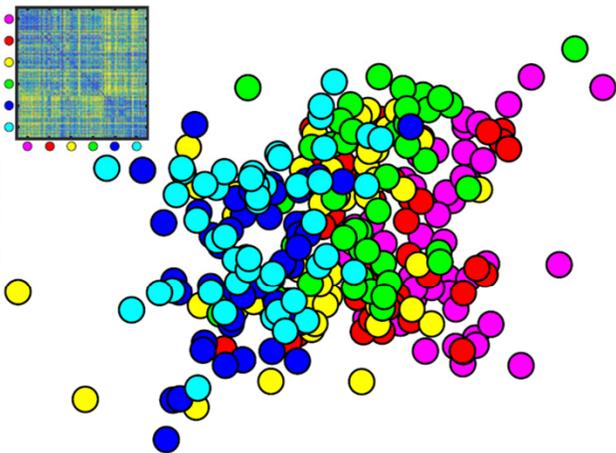
Sound-to-event DNNs



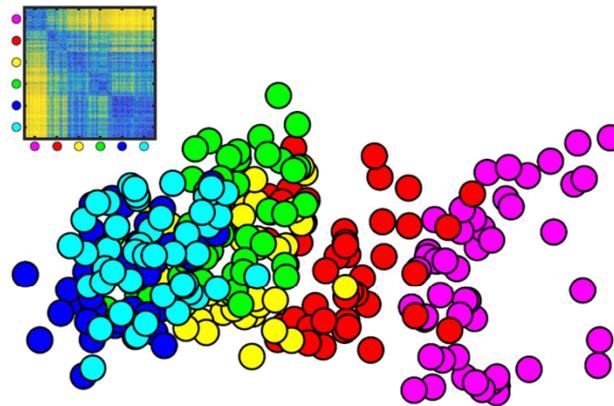
NLP embeddings



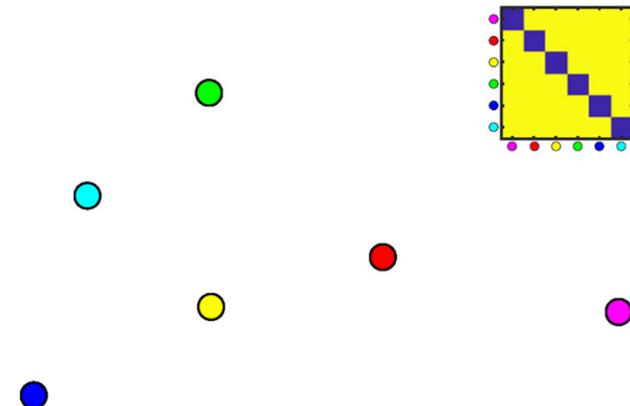
HG



pSTG



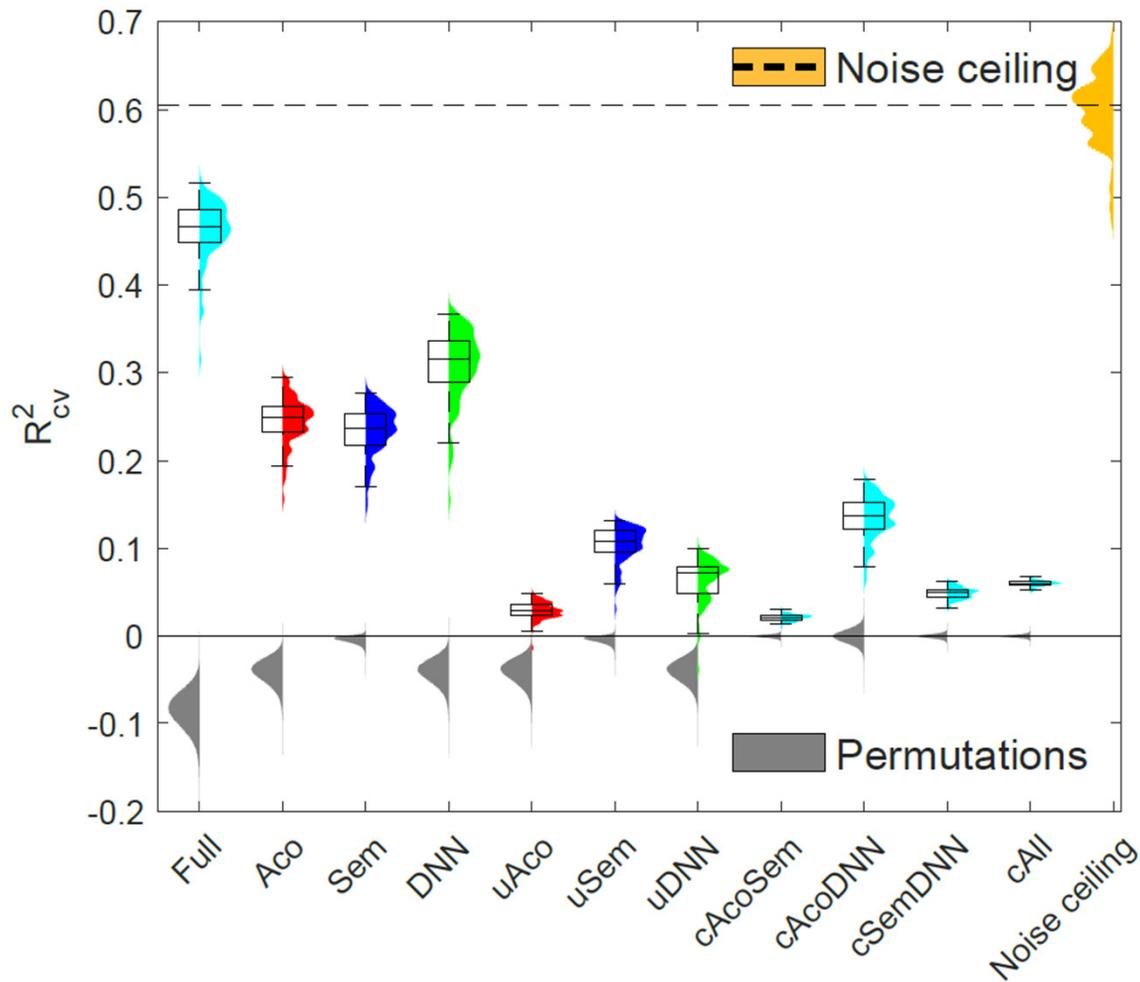
Categories



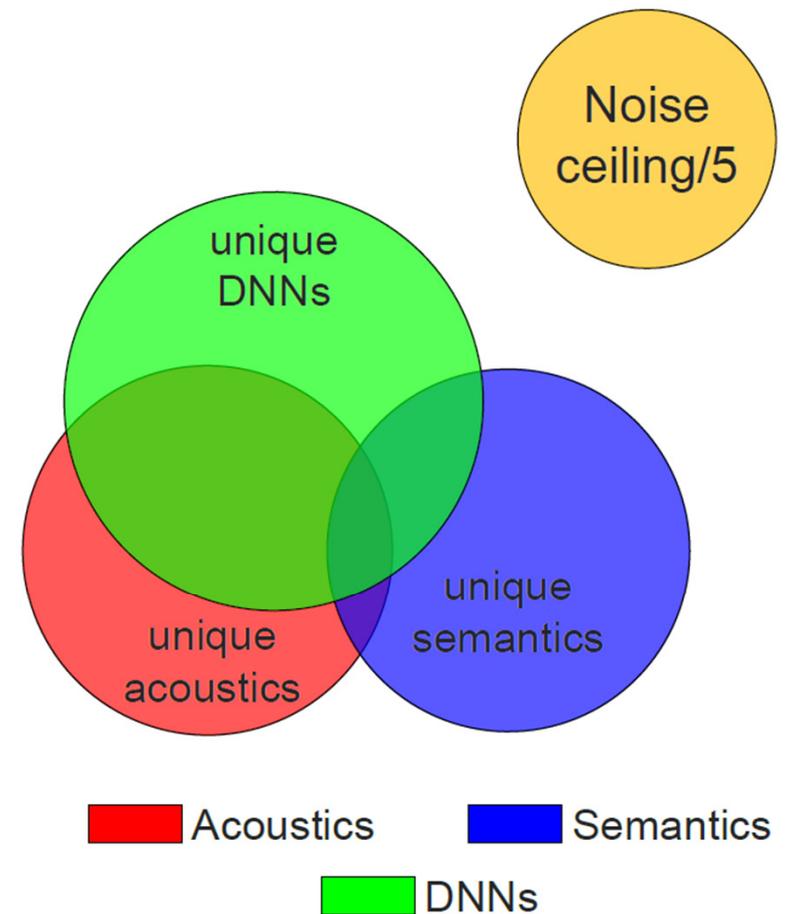
● speech ● voice ● animal ● music ● nature ● tools

min max
Distance rank

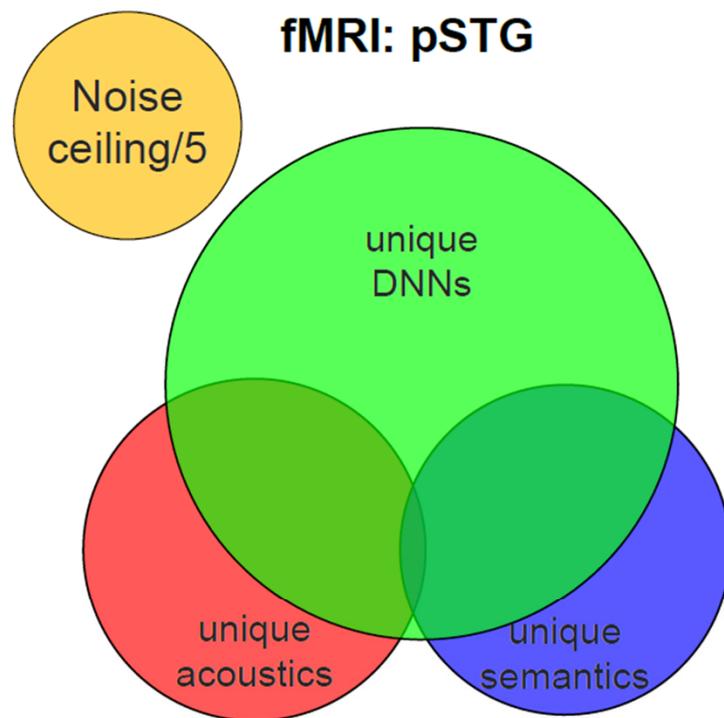
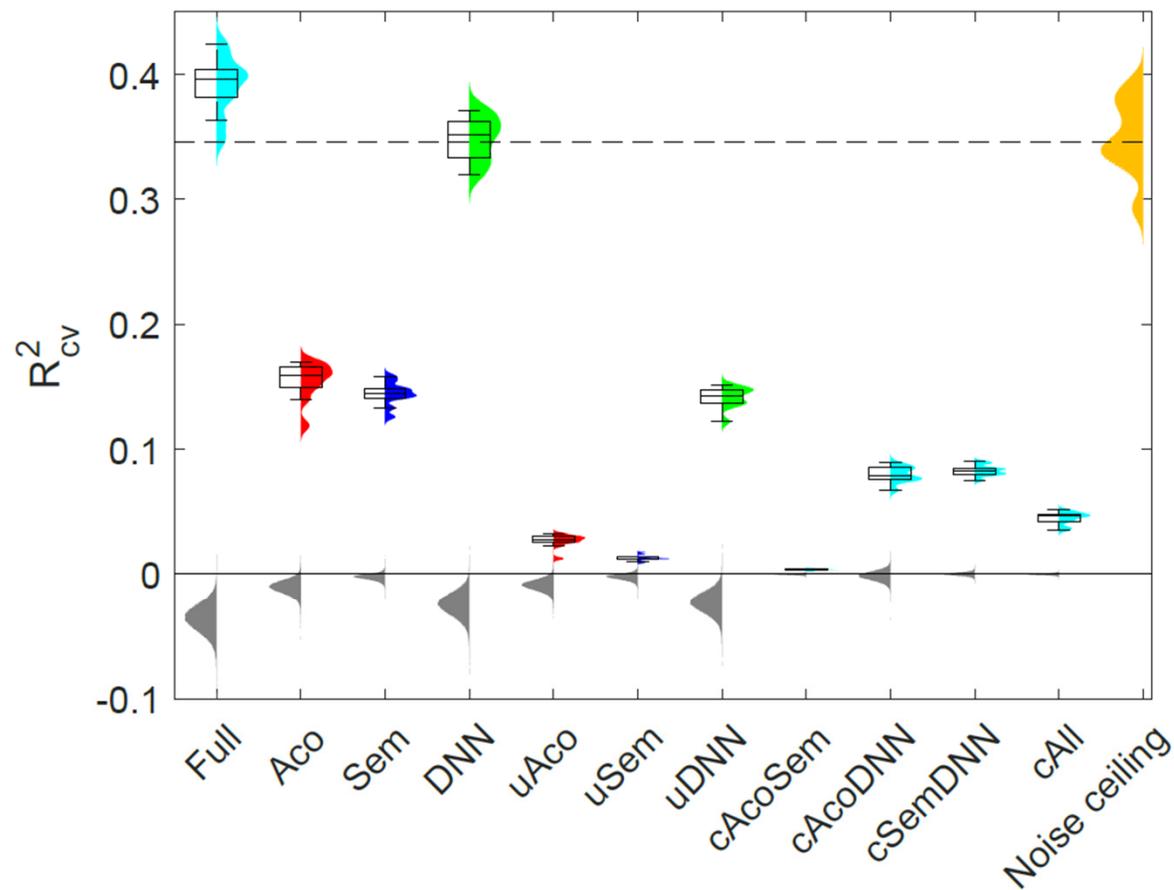
Behavioural results: Perceived sound dissimilarity



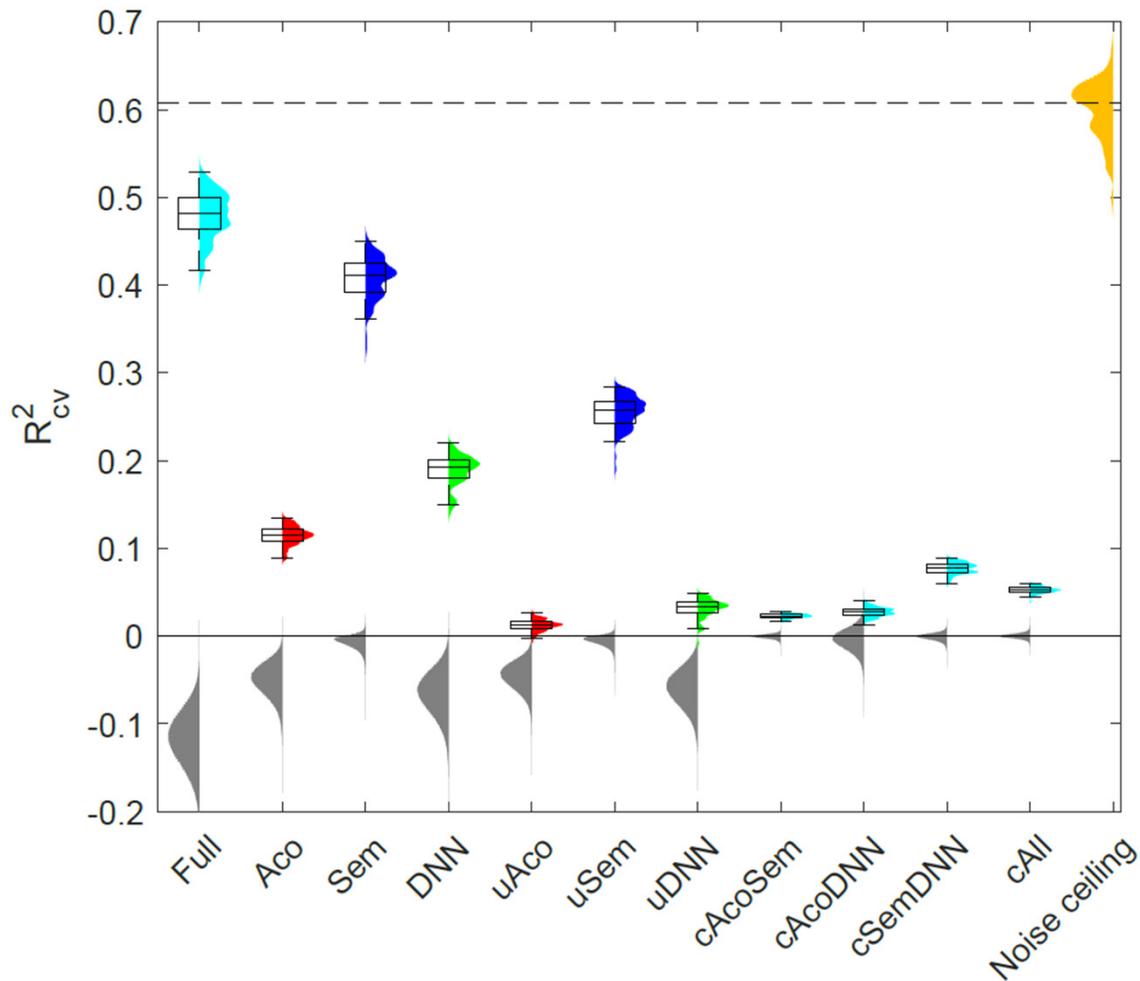
Behaviour: sound dissimilarity



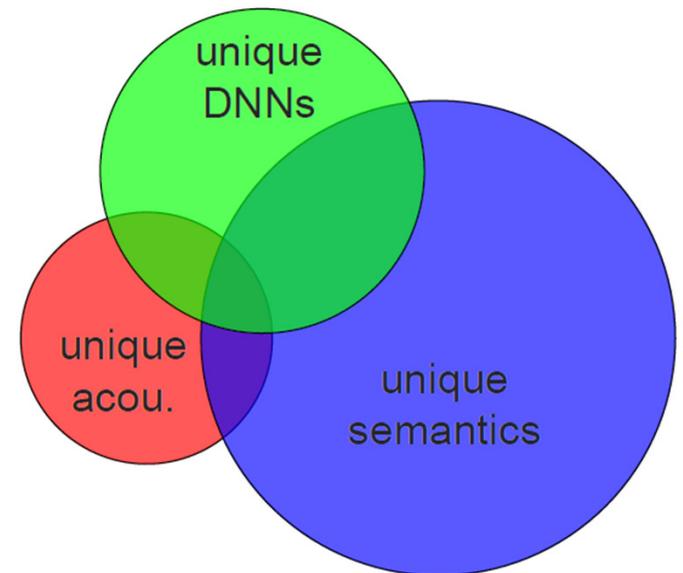
7T fMRI results: Posterior Superior Temporal Gyrus (pSTG)



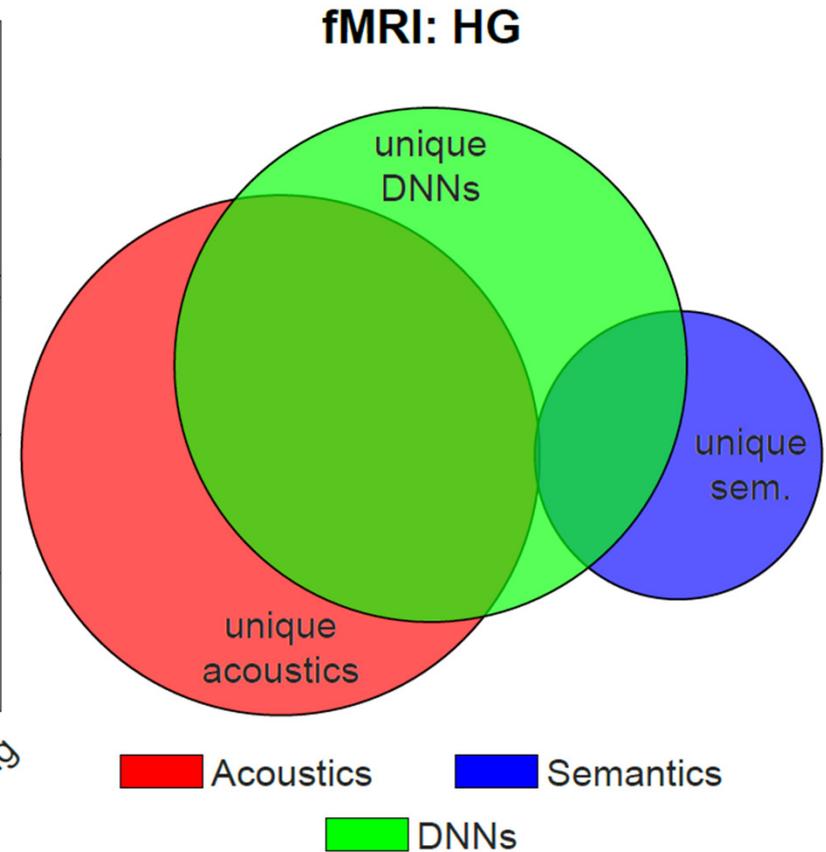
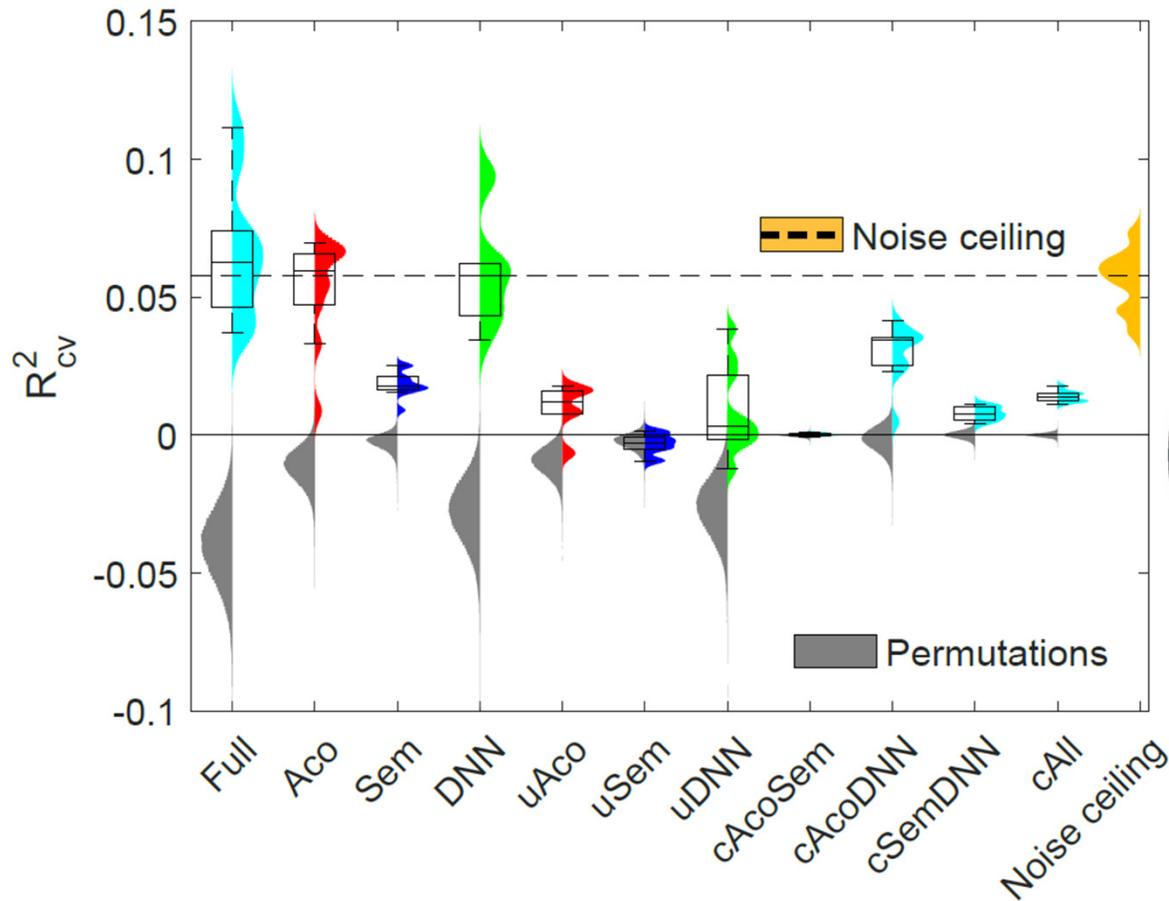
Behavioural results: Perceived word dissimilarity



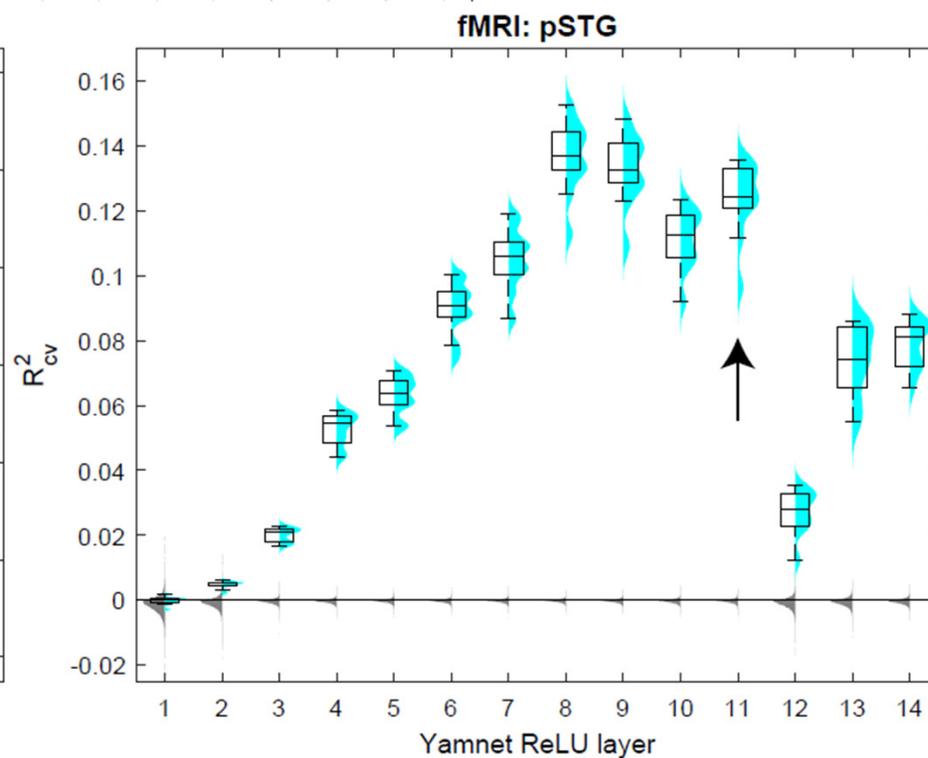
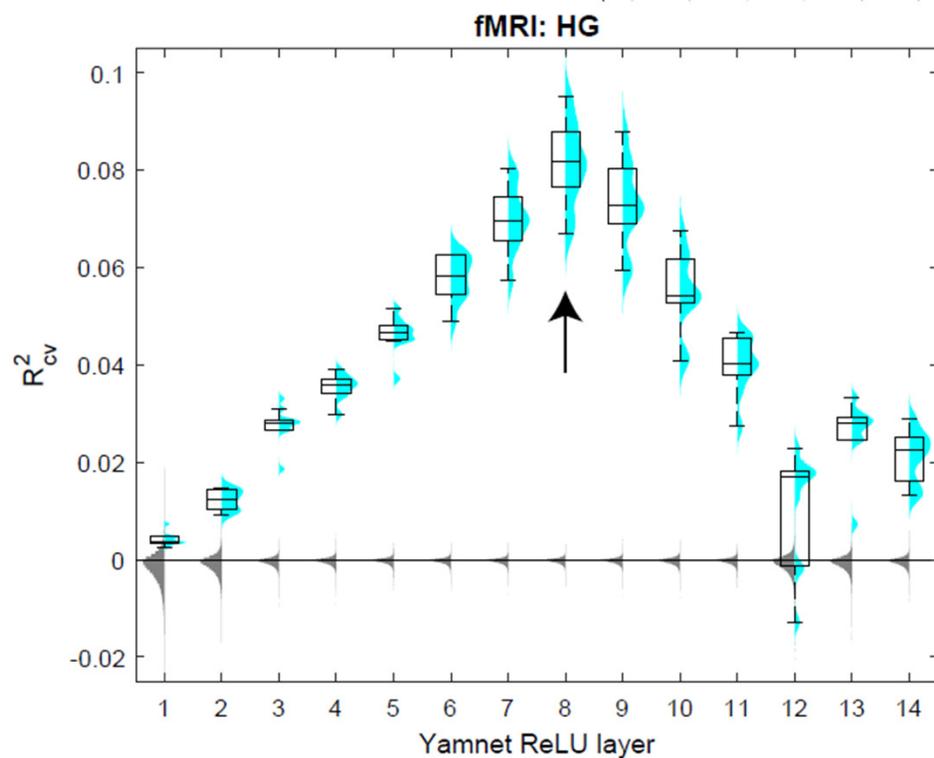
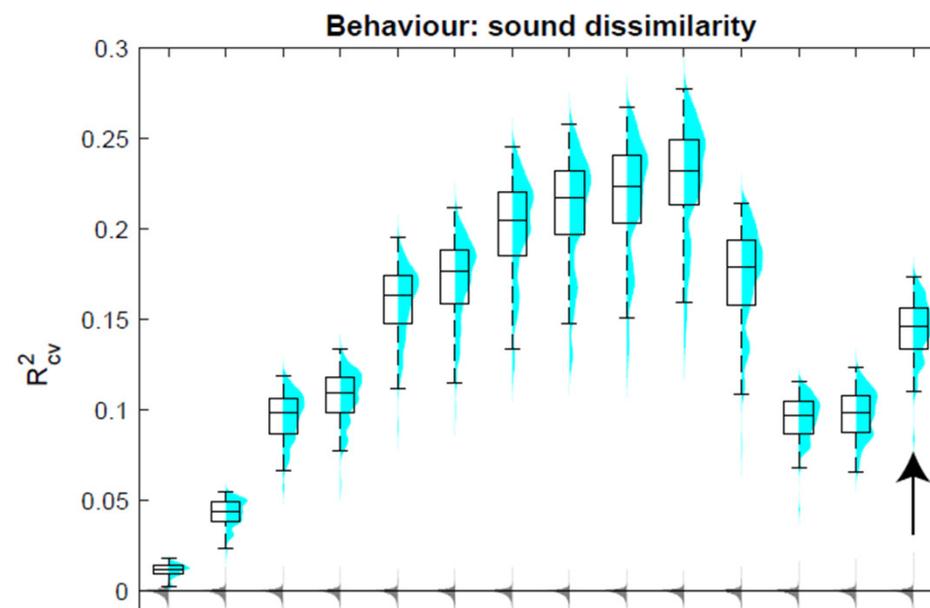
Behaviour: word dissimilarity



7T fMRI results: Heschl's gyrus (HG, early auditory cortex)



Sound2event DNNs: intermediate layers





M Esposito



G Valente



E Formisano

anr



Data &
Code



DRYAD

SEMANTICALLY-INFORMED DEEP NEURAL NETWORKS FOR SOUND RECOGNITION

*Michele Esposito¹, Giancarlo Valente¹, Yenisel Plasencia-Calaña²,
Michel Dumontier³, Bruno L. Giordano⁴, Elia Formisano^{1,2}*

¹ Department of Cognitive Neuroscience - Maastricht University, Maastricht, The Netherlands

² BISS institute - Maastricht University, Maastricht, The Netherlands

³ Institute of Data Science - Maastricht University, Maastricht, The Netherlands

⁴ Institut des Neurosciences de La Timone - Université Aix-Marseille, Marseille, France

