A.

Audiovisual naturalistic movie

B.

Extract model embeddings of first frame of every TR

Collect human judgments

Extract model embeddings of spoken language

vision

motion

social interaction

speech

word

sentence

Time

C.

	vision	motion	social interaction	speech	word
motion	0.33				
social interaction	0.11	0.31			
speech	0.15	0.08	0.62		
word	0.12	0.13	0.62	0.93	
sentence	0.11	0.14	0.54	0.82	0.77

Canonical correlation coefficient

1.00

0.75

0.50

0.25

0.00

D.

Fit encoding model

x

$\beta_1\ \beta_2\ \beta_3\ \cdots\ \beta_{n-2}\ \beta_{n-1}\ \beta_n$

Predicted fMRI time series

R^2

Actual fMRI time series