

Supplementary Tables

Table S1. Summary of study participants, including handedness, sex and age.

Subject	Handedness	Sex
Participant 1	right	female
Participant 2	right	male
Participant 3	right	female
Participant 4	right	male
Participant 5	right	male
Participant 6	left	female
Participant 7	left	male
Participant 8	left	male
Participant 9	right	female
Participant 10	right	female
Participant 11	left	female
Participant 12	right	male
Participant 13	right	male
Participant 14	right	male

Table S2. Channel distribution across brain areas.

area	counts
Cerebral-white-matter	576
Middle temporal	177
Hippocampus	142
Rostral middle frontal	115
Superior temporal	106
Amygdala	78
Lateral orbitofrontal	78
Caudal middle frontal	55
Putamen	46
Thalamus	45
Inferior temporal	37
Precentral	33
Pars triangularis	33
Caudate	31
Insula	29
Inferior parietal	28
Supramarginal	26
Superior frontal	26
Caudal anterior cingulate	24
Parahippocampal	23
Fusiform	22
Pars orbitalis	20
Medial orbitofrontal	20
Precuneus	19
Isthmus cingulate	17
Lingual	16
Superior parietal	14
Pallidum	12
Lateral occipital	12
Rostral anterior cingulate	9
Transverse temporal	7
Postcentral	6
Accumbens-area	6
Posterior cingulate	5
Pericalcarine	5
Cerebellum-cortex	5

Entorhinal	4
Parsopercularis	2
Cuneus	1

Table S3a. Example 1 of the natural dialogue between experimenter (E) and participant (P).

Speaker	Sentences
E	Okay Batman would make a great dad
P	I just saw the new Batman movie the Lego Batman movie
E	Any good
P	Yeah it's pretty good
E	I think like the thing has there been one before

Table S3b. Example 2 of the natural dialogue between experimenter (E) and participant (P).

Speaker	Sentences
E	I was gonna ask when's the last time a joke made you laugh but technically
P	Oh I laugh all the time at jokes
E	Okay
P	Even the dumb ones honestly
E	Do you have a favorite comedian or
P	I think ah not really I like Seinfeld a lot
E	Okay
P	this one guy uh Louis CK
E	Louis CK is awesome
P	He's pretty funny
E	I think he actually comes from Boston
P	Yeah he he comes from around Boston yeah

Table S3c. Example 3 of the natural dialogue between experimenter (E) and participant (P).

Speaker	Sentences
P	I watch like Top Chef uh a few times and like Hell's kitchen and stuff
E	What's Top Chef
P	Those are pretty good
E	So
P	It's fun where like they get a bunch of people together and they give them a bunch of like challenges and then they vote them off one by one
E	No okay it's like survivor style
P	Yeah basically

Supplementary Figures

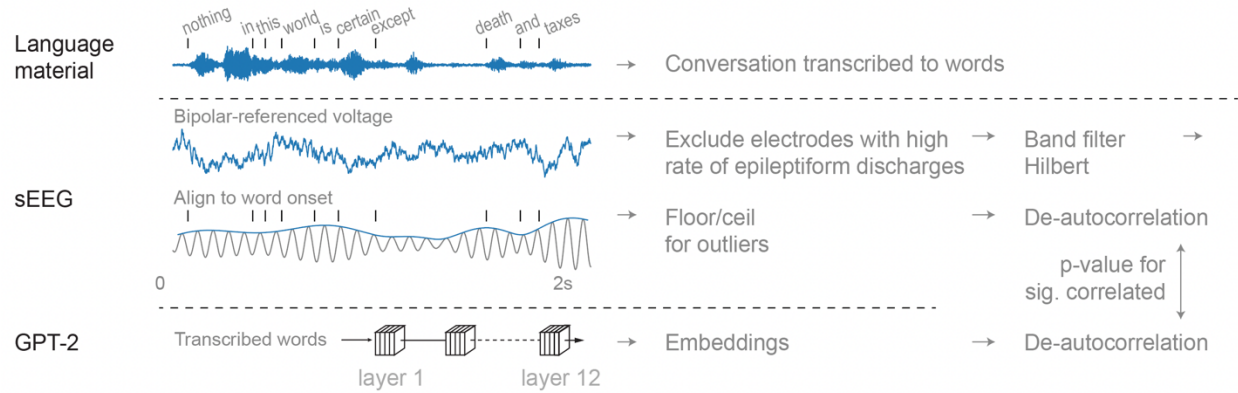
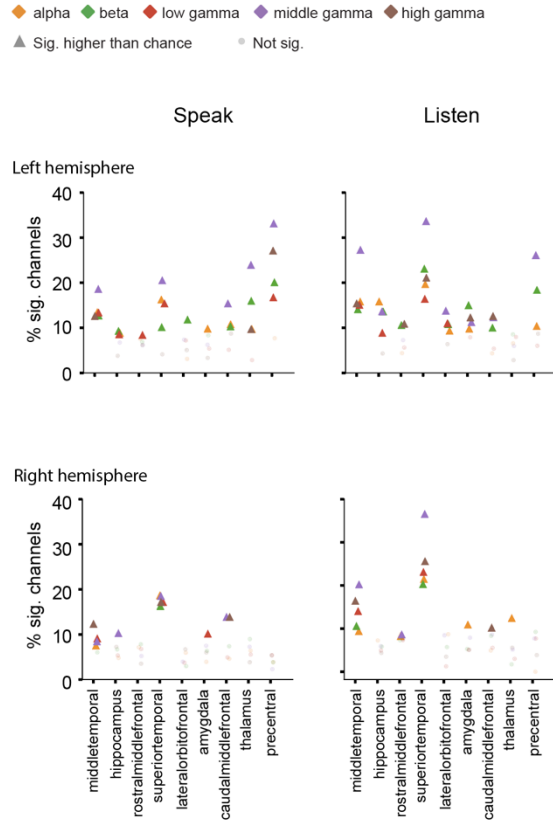


Figure S1. Illustration of neural activity pre-processing and comparison to the GPT-2 model. Audios of language materials and neural channels were simultaneously recorded and were transcribed to words. Bipolar-referenced LFPs were processed for envelopes at the frequency bands of alpha (8-13 Hz), beta (13-30), low gamma (30-55), mid-gamma (70-110), high gamma (130-170). Meanwhile, the transcribed words were input to a pre-trained GPT-2 model. We quantified the degree to which the artificial activities of specific nodes in the language model were predictive of neural activities obtained through depth electrode recordings by examining the correlation between the de-autocorrelated embeddings and the de-autocorrelated LFPs at each frequency band.

a. Correlated channel by area and frequency



b. Correlated channels grouped by hemisphere and handedness

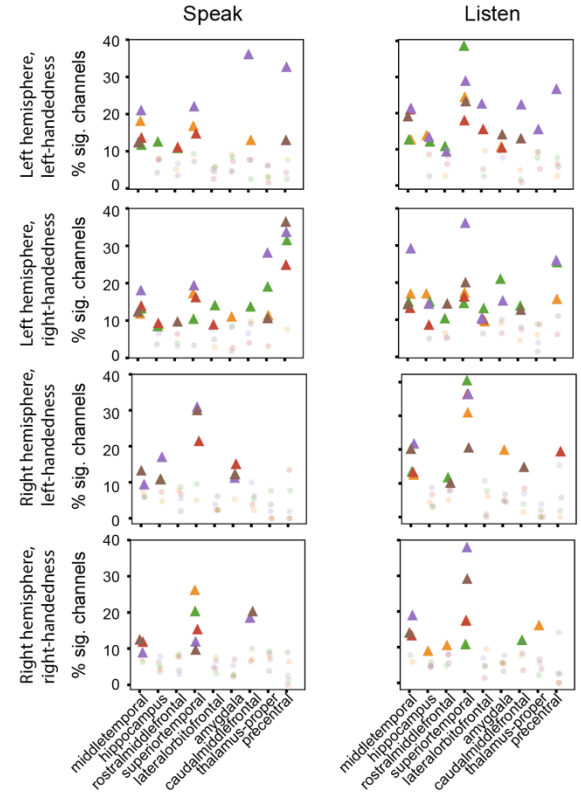


Figure S2. Artificial-neural activity correlations by brain areas. **a.** The percentage of channels with significant correlation to the NLP model were illustrated by frequency bands, hemispheres, and brain areas. **b.** These channels were further grouped based on the handedness of the participants. Among left-handed participants, there were higher percentage of correlated channels in the left hemisphere compared to the right (2-side permutation test, $n = 10^4$, $p < 10^{-4}$). Source data are provided as a Source Data file.

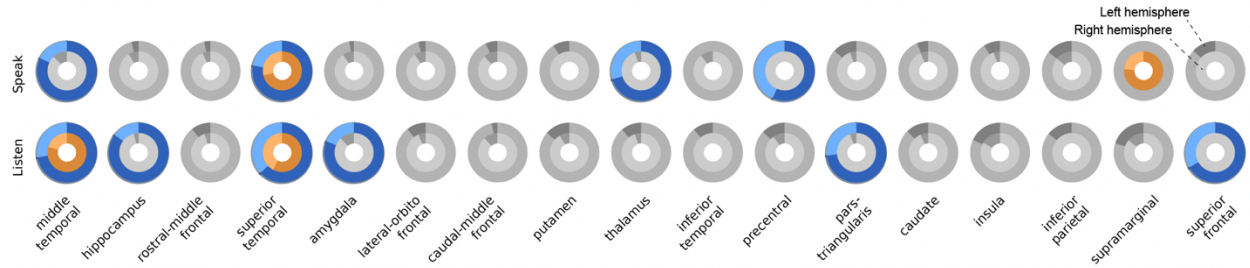


Figure S3. Brain areas with significant high percentage of NLP-correlating channels calculated with Bonferroni corrections. The percentage of NLP-correlating channels were illustrated in each subplot across different brain areas, hemispheres, and modalities. To examine whether there were areas with significant higher percentage of NLP-correlating channels, we performed Bonferroni correction of the brain areas (17 areas), after the Bonferroni correction of NLP nodes, layers and frequency bands. Areas with significantly higher percentage of channels were shown in non-gray colors (blue for left hemisphere and orange for right hemisphere). Source data are provided as a Source Data file.

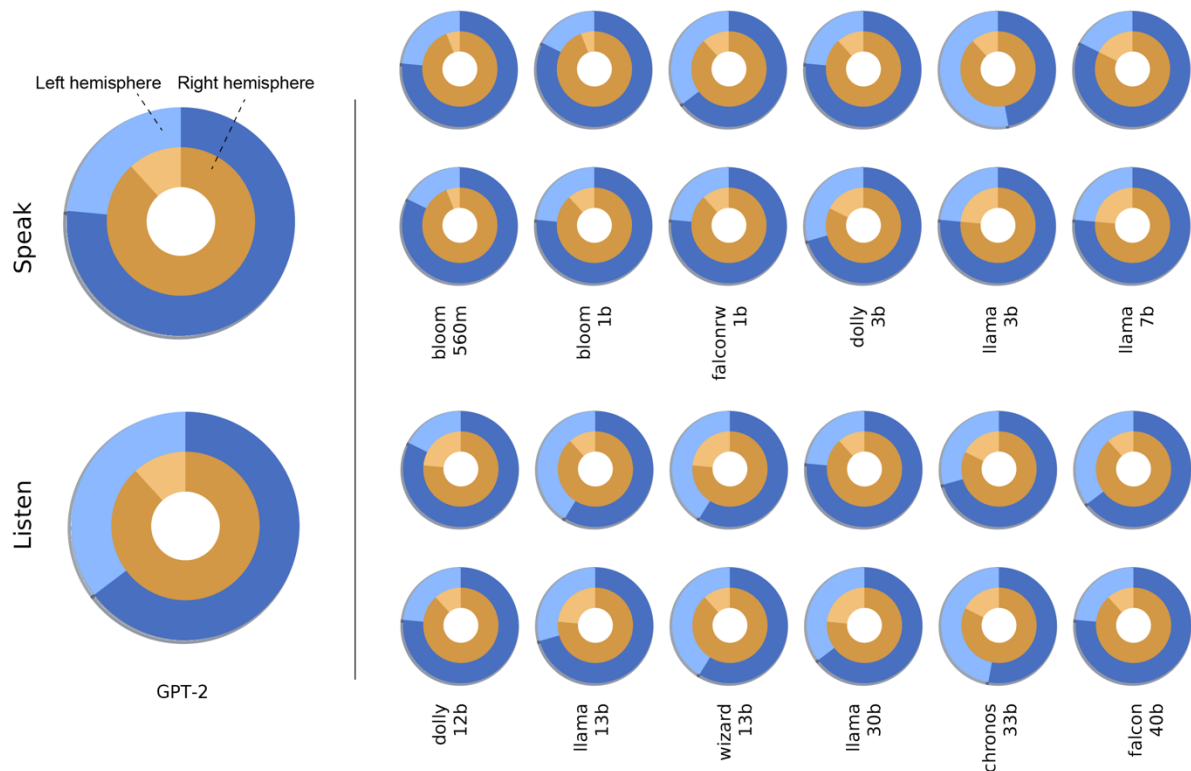


Figure S4. Percentage of NLP-responding brain areas across different NLP models after Bonferroni corrections on multiple comparison of brain regions. *Left*, using GPT-2 model, the percentage of brain areas grouped by hemispheres is plotted during speech production and comprehension. The percentage of brain areas were calculated based on whether they showed a significantly higher ratio of NLP-correlated channels with Bonferroni correction for multiple brain areas (17 areas), after Bonferroni correction for NLP nodes, layers, and channel frequency bands. The colors label the hemispheres. *Right*, same analyses were performed on a range of pre-trained state-of-the-arts large language / NLP models with varying sizes, layers, and training datasets. The percentages of NLP-responding brain areas were consistent when using these different NLP models. Source data are provided as a Source Data file.

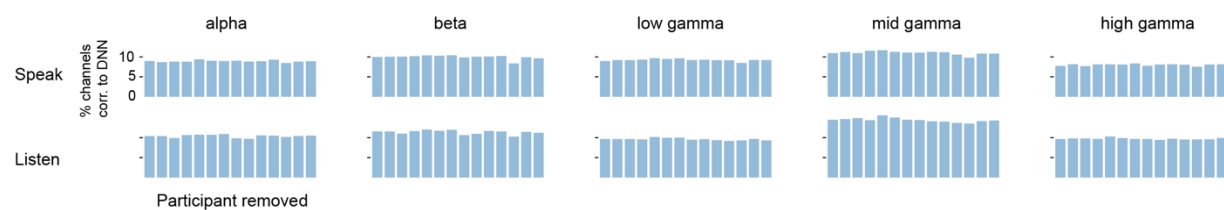


Figure S5. Percentage of channels correlating to NLP model embeddings by removing a participant. To ensure the ratio of channels significantly correlated to NLP was not caused by a single participant, we removed each participant (x-axis) and plot the percentage for both production (top row) and perception (bottom row) at different frequency bands. An one-way ANOVA was applied to examine whether removing a patient caused a significant change of the ratio ($n = 14$), and all p-values were above 0.75. Individual data points are not applicable, and are not plotted because bar plots represent the percentage of responding channels relative to all recorded channels from all but one participants. Source data, p-values, and the sex of the participants are provided as a Source Data file.

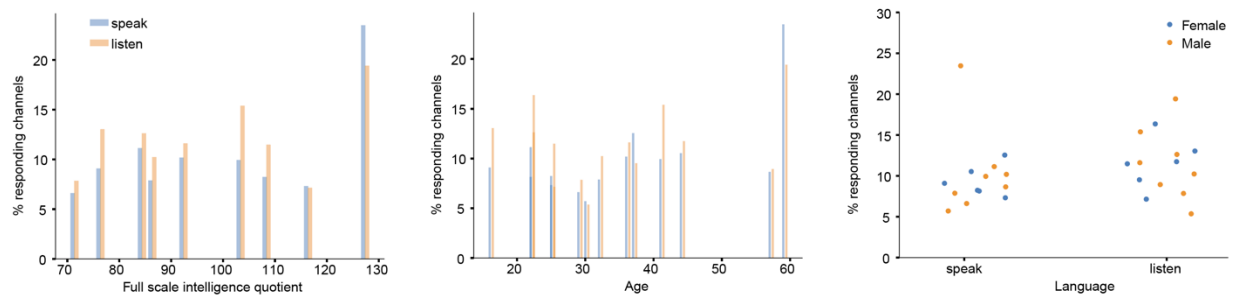


Figure S6. Responding channel by participant's intelligence quotient (IQ), age and sex. To examine whether the neural response is correlated to the IQ score for the participants, we performed full scale intelligence quotient tests for 9 out of 14 participants and calculated the percentage of responding channels as a function of their IQ (*left*). Each pair of bars represent a participant, and the location of these bars are plotted based on their IQ values. There is no significant correlation between the percentage of channels and participants IQ (Kendall Tau test, $n = 9$, statistics = 0.167, $p = 0.61$ for speak; statistics = 0.11, $p = 0.76$ for listen). Similar, the percentage of channels is not correlated to the age of the participants (*middle*, Kendall Tau test, $n = 14$, statistics = 0.26, $p = 0.21$ for speak; statistics = 0.01, $p = 0.96$ for listen). For the plots on the *left* and in the *middle*, individual data points are not applicable, and are not plotted because bar plots represent the percentage of responding channels relative to recorded channels from one participant. Lastly, the percentage channels are not dependent on the sex of the participants (*right*, $n_{\text{female}} = 6$ and $n_{\text{male}} = 8$, One-way ANOVA, statistics = 0.23, $p = 0.64$ for speak; statistics = 0.003, $p = 0.96$ for listen). Source data are provided as a Source Data file.

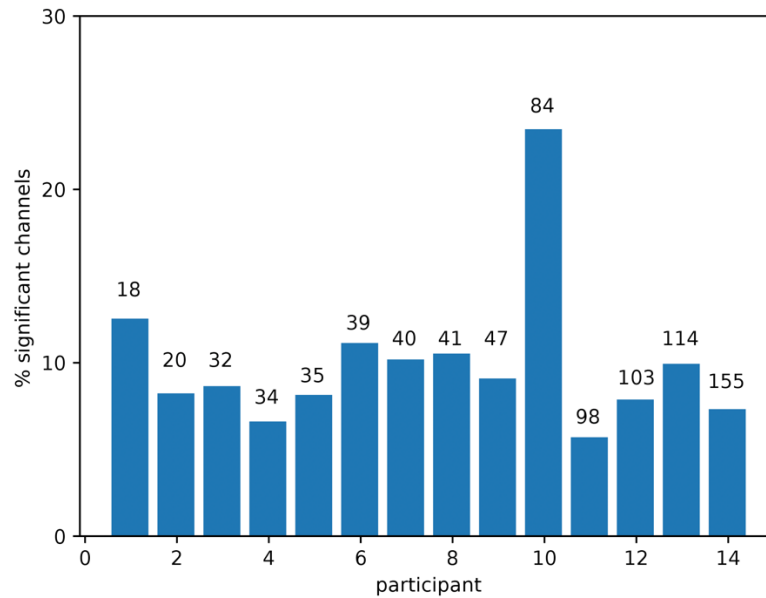


Figure S7. Percentage of significant channels is not impacted by the number of words. Participants were sorted based on the total number of words they speak and listen, and the percentage of responding channels are plotted. The numbers on top of each bar indicate the total number of words (x 100). There is no correlation with the percentage of channels to the rank of participants ($n = 14$, regression test, $p = 0.95$). This suggests that the sample size is not a factor contributing to our results, hence having a longer conversation does not seem to have an impact on the percentage of the responding channels. Individual data points are not applicable, and are not plotted because bar plots represent the percentage of responding channels relative to recorded channels from one participant. Source data are provided as a Source Data file.

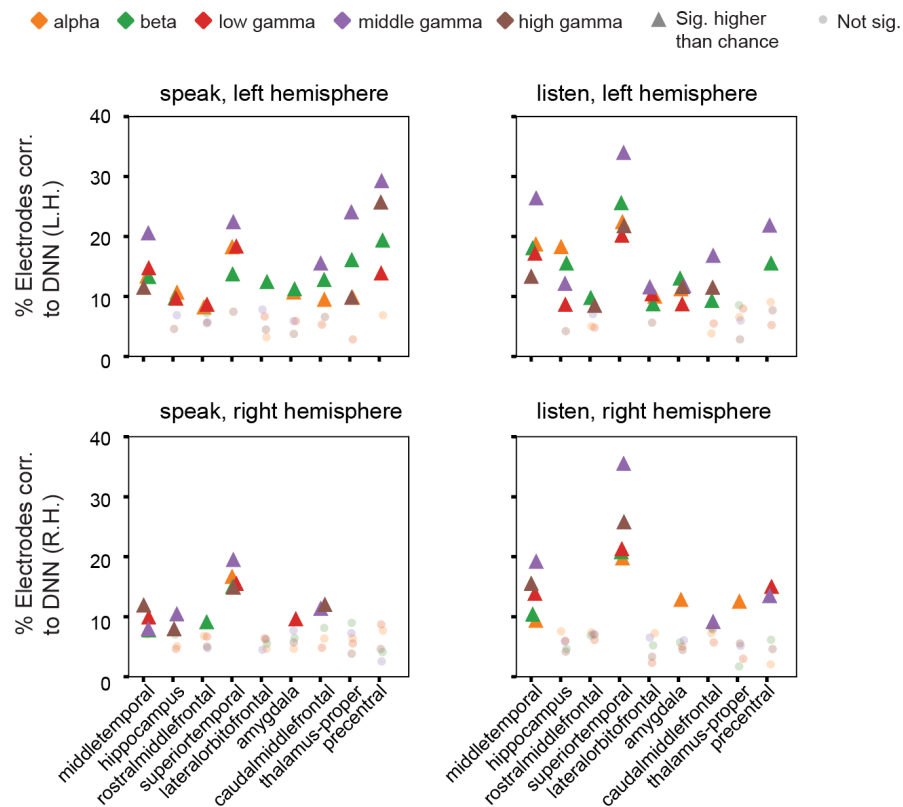


Figure S8. Controls for the categorization of white matter. In the main analysis, brain areas identified as white matter were categorized as ‘Other’ and included in all analyses. Since nearby gray matters contribute more to EEG, another way to categorize brain areas is to assign white matter locations based on the nearest gray matter. Here we found that the new categories without white matter show very similar ratios of channels that were significantly correlated to the NLP embeddings compared to excluding these channels. Source data are provided as a Source Data file.

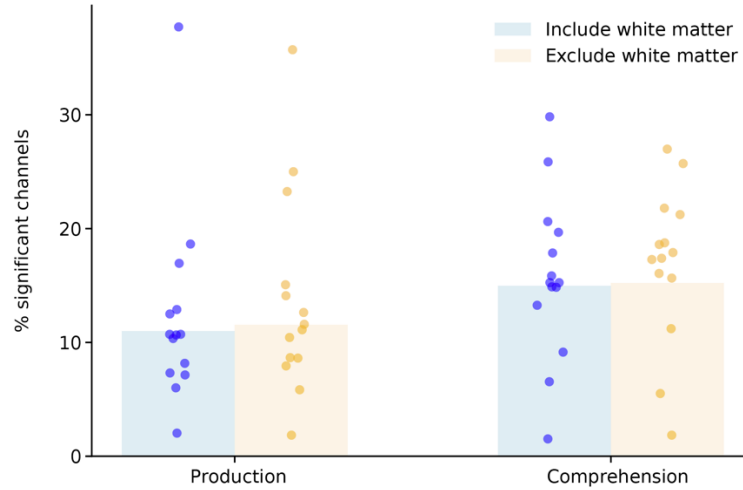
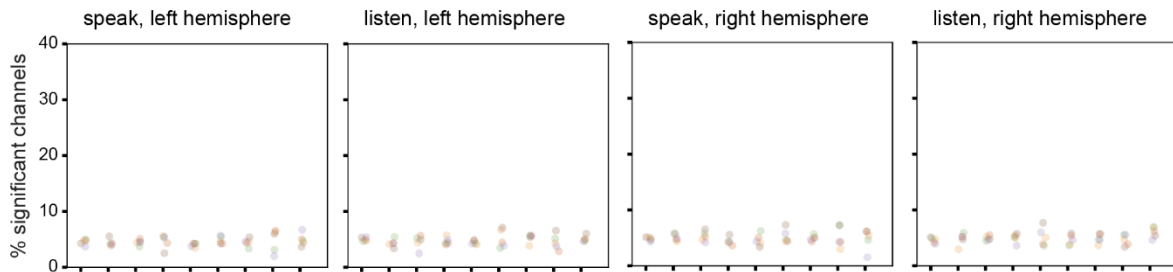


Figure S9. Control analysis to exclude channels from white matter areas. By excluding white matter channels, we found that on average, there were 11.5% channels responding to NLP embeddings during language production and 15.2% channels during comprehension. Both were significantly higher than random fluctuations (Chi-square proportion test: statistics = 37.7, $p = 8.2 \times 10^{-10}$ for speaking; statistics = 127.4, $p = 1.5 \times 10^{-29}$ for listening). These percentages were very similar to those when white matters were included (11.0% and 15.0% for production and comprehension). Therefore, the neural-NLP correlations were present regardless of whether white matter channels were included. Individual data points represent the percentage of individual participants ($n = 14$ participants). Source data are provided as a Source Data file.

a. Correlated channels after random permutation

alpha beta low gamma middle gamma high gamma ▲ Sig. higher than chance ● Not sig.



b. Correlated channels to BERT model

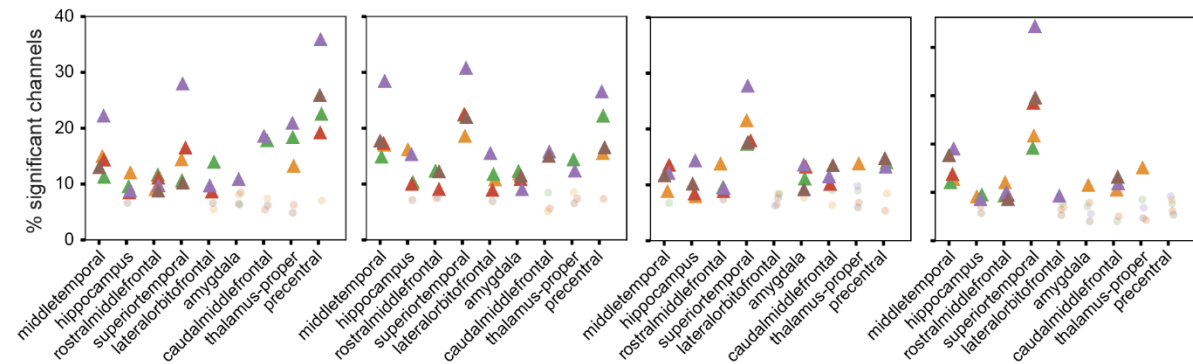


Figure S10. Controls for robustness and generalization. **a.** A control with the permuted NLP embeddings showed that correlated channels were homogeneously distributed at chance-level (5%), which were significantly lower than those without permutation (Chi-square test, statistic = 15278, $p < 10^{-100}$). **b.** Percentage of correlated channels to another pre-trained NLP model (BERT) showed similar or higher percentage to the pre-trained GPT-2 model, suggesting the neural-artificial convergence is generalizable. Source data are provided as a Source Data file.

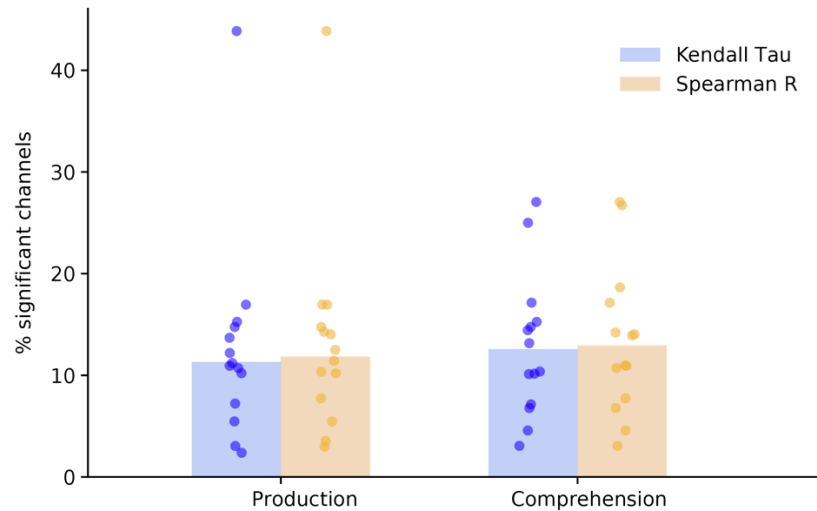


Figure S11. Control analysis with non-parametric rank correlations. We used Kendall Tau test and Spearman R to examine the correlations between channels and NLP embeddings. Using Kendall Tau test, we found there were 11.3% channels responding to NLP during language production, whereas there were 12.6% to comprehension (Chi-square proportion test, statistics = 150.0, $p = 1.1 \times 10^{-36}$ for speaking; statistics = 230.1, $p = 5.5 \times 10^{-52}$ for listen). Similar percentage of channels were selected using the Spearman R test (speaking: 11.8%, statistics = 187.7, $p = 1.0 \times 10^{-42}$; listening: 12.9%, statistics = 253.0, $p = 5.8 \times 10^{-57}$). Taken together, these indicate that the channel-NLP correlations were not limited to specific statistical methods, and were generalizable to various methods for correlation tests. Individual data points represent the percentage of individual participants ($n = 14$ participants). Source data are provided as a Source Data file.

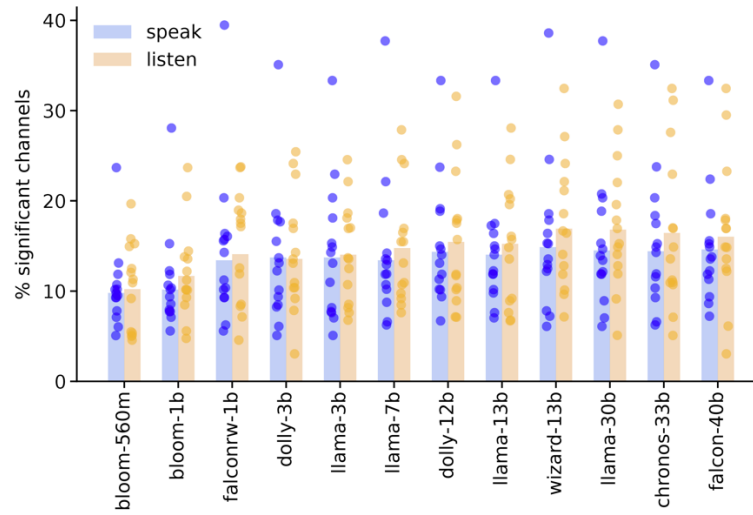


Figure S12. Percentage of significant channels examined by various NLP models. By applying NLP models of various sizes, number of layers, and families, we found that there were significant percentage of channels correlating to all models for both speech production and comprehension (Chi square proportion test, $p < 1 \times 10^{-23}$), with larger models showing higher percentage of correlated channels ($R > 0.80$ and $p < 2 \times 10^{-3}$). Individual data points represent the percentage of individual participants ($n = 14$ participants). Source data are provided as a Source Data file.

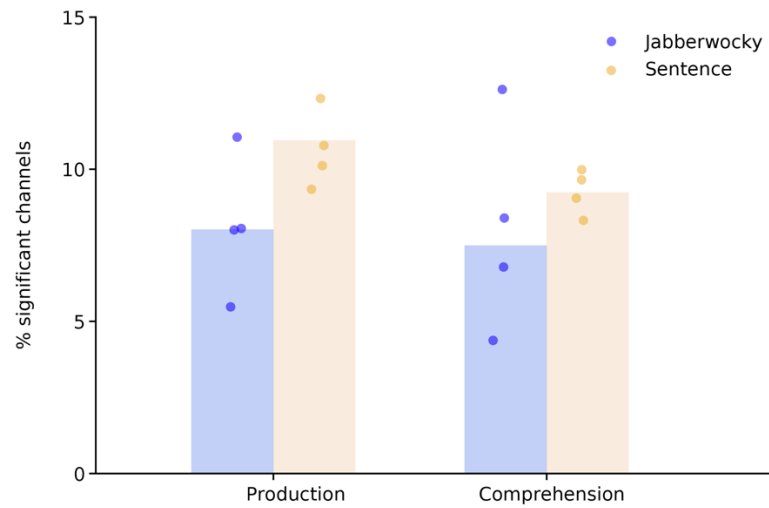
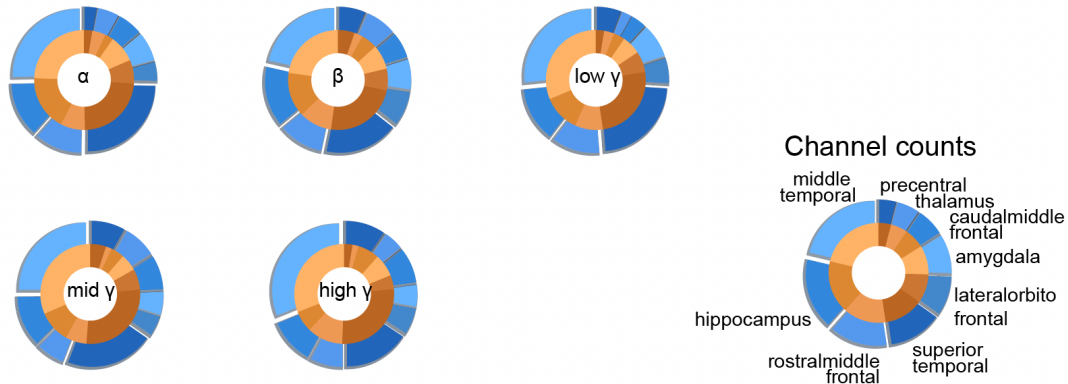


Figure S13. Channels responding to passively repeating fake or real sentences. We performed control analysis by instructing the participants to passively repeat fake sentences that sounded like English, but do not convey any actual meanings (i.e. jabberwocky, for example: “Dorty of the pikap incaged a flup of mool.”). These tasks were performed with the same four participants who performed the passive repeating and listening to sentences control, and each participants spoke and heard the same number of sentences and fake sentences. In this way, we can directly compare the percentage of channels under these two control tasks to identify the contribution of meaningful language materials encoded in channels. We found that there was a significantly higher percentage of channels responding to real sentences compared to jabberwocky for both comprehension and production (ANOVA Repeated measurement, $n = 4$, $F = 34.1$, $p = 0.0043$ for speaking; $F = 18.5$, $p = 0.012$ for listening). This showed that coherent, meaningful sentences are processed with a higher percentage of channels when compared to meaningless fake sentences, which indicated that the contribution of semantic feature in neural processing. Individual data points represent the percentage of individual participants ($n = 4$ participants). Source data are provided as a Source Data file.

a. Correlated electrodes distribution by brain areas



b. Correlated electrodes distribution by frequency band

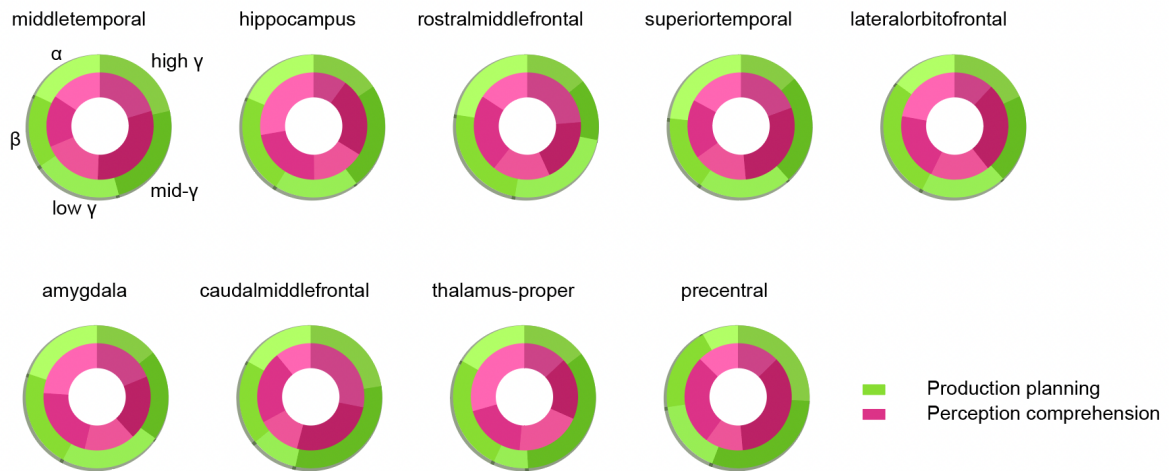


Figure S14. NLP-correlated channels across brain areas and frequency bands. a. Channels that were significantly correlated to NLP were grouped by brain areas and plotted across different frequency bands. **b.** Channels that were significantly correlated to NLP were grouped by frequency bands and plotted by different brain areas. Source data are provided as a Source Data file.

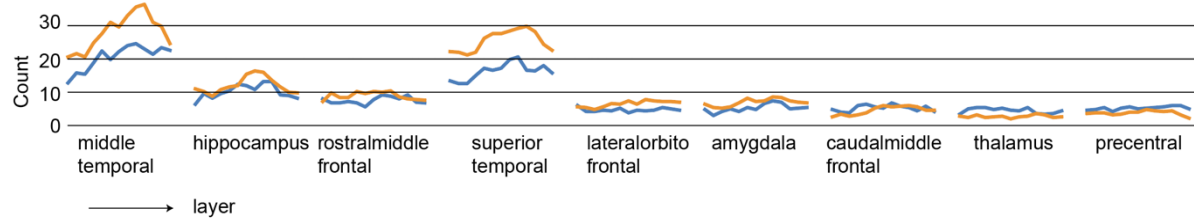


Figure S15. Correlated channels across various NLP layers grouped by brain areas. Channels correlated to NLP were plotted based on the NLP layers from various areas in natural conversation. Source data are provided as a Source Data file.

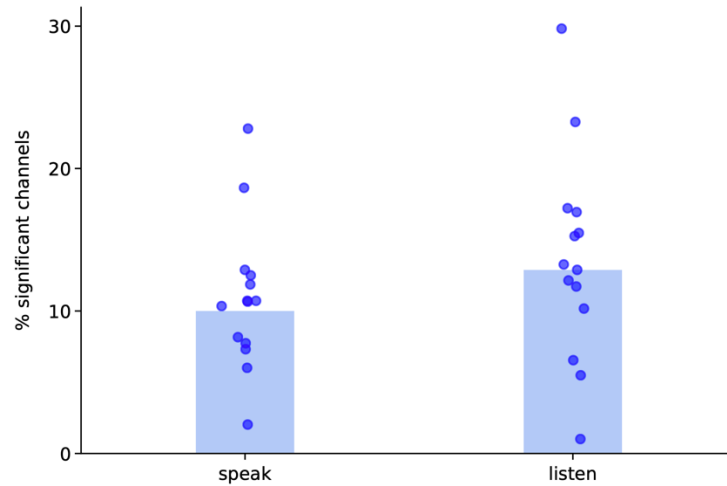
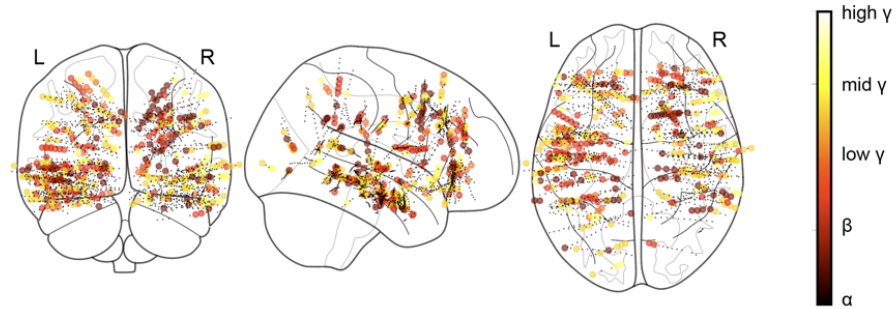


Figure S16. Control analysis with matching word number for speech production and comprehension. With same number of words for speech production and comprehension, we still found more channels responding to speech comprehension (Chi-square proportion test, statistics = 14.1, $p = 1.7 \times 10^{-4}$). Both production and comprehension involved significant channels higher than chance (chi-square proportion test, statistics > 100, $p < 2 \times 10^{-23}$ for both conditions). Individual data points represent the percentage of individual participants ($n = 14$ participants). Source data are provided as a Source Data file.

Transition from comprehension to articulation



Transition from articulation to comprehension

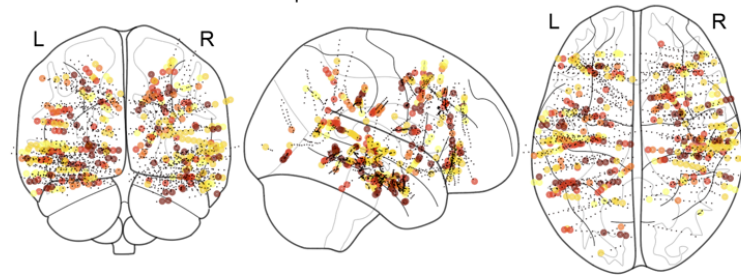


Figure S17. Channels with frequency bands significantly changed during speaker-listener transitions.

Mapping of channels by their frequency bands that showed most significant changes during transitions. Glass brain plots were generated using Nilearn package (Nilearn contributors, Chamma, A., et al. (2025). *nilearn* (TestRelease). Zenodo. <https://doi.org/10.5281/zenodo.14697221>). Source data are provided as a Source Data file.

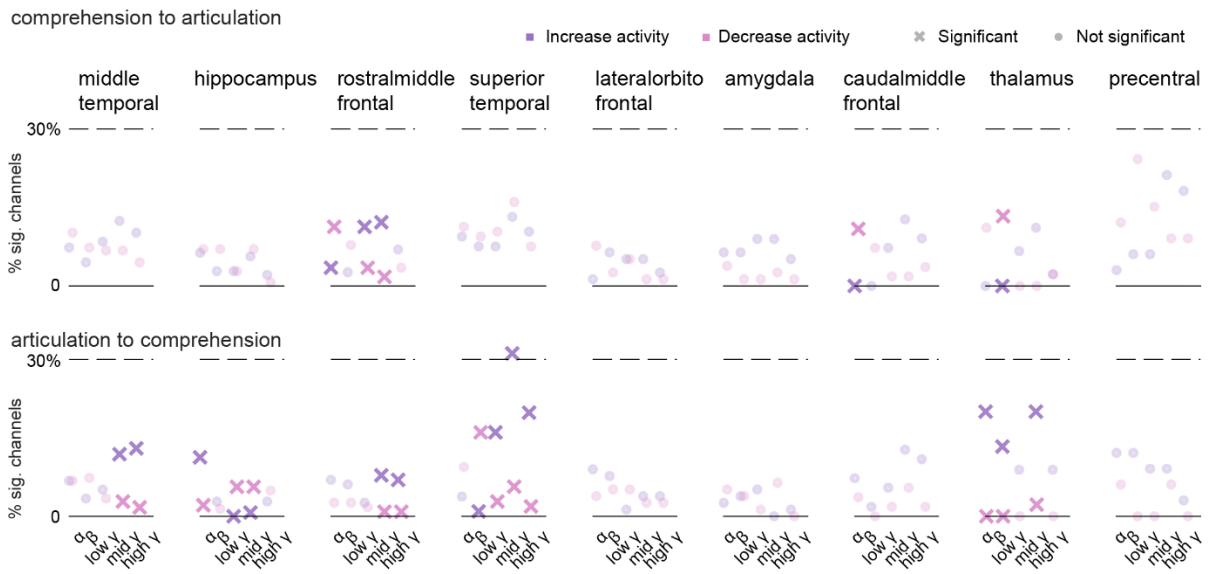


Figure S18. Neural activity changes with turn-takings during conversation. Percentage of channels that show increased activity (purple) and decreased activity (pink) by different the brain areas and the frequency bands. Source data and p-values are provided as a Source Data file.