

Supplementary information

List of Tables

S1	Total count and rate of missing values for phenotypic variables	7
S2	Summary statistics of signal per fMRI sequence depending on repetition time.	8
S3	Data dictionary for the phenotypic variables.	8
S4	Descriptive statistics for phenotypic variables.	8
S5	Phenotypic dataset.	8

List of Figures

S1	Correlation matrix of linguistic environment variables.	2
S2	Summary statistics of the fMRI time series as a function of repetition time (RT).	3
S3	Gray matter volume relationship with age.	3
S4	Visual language comprehension functional topography.	4
S5	Auditory language comprehension functional topography.	5
S6	Distribution of mean T values across fMRI runs for the top 10 regions of the AAL3 atlas.	6

Supplementary figures

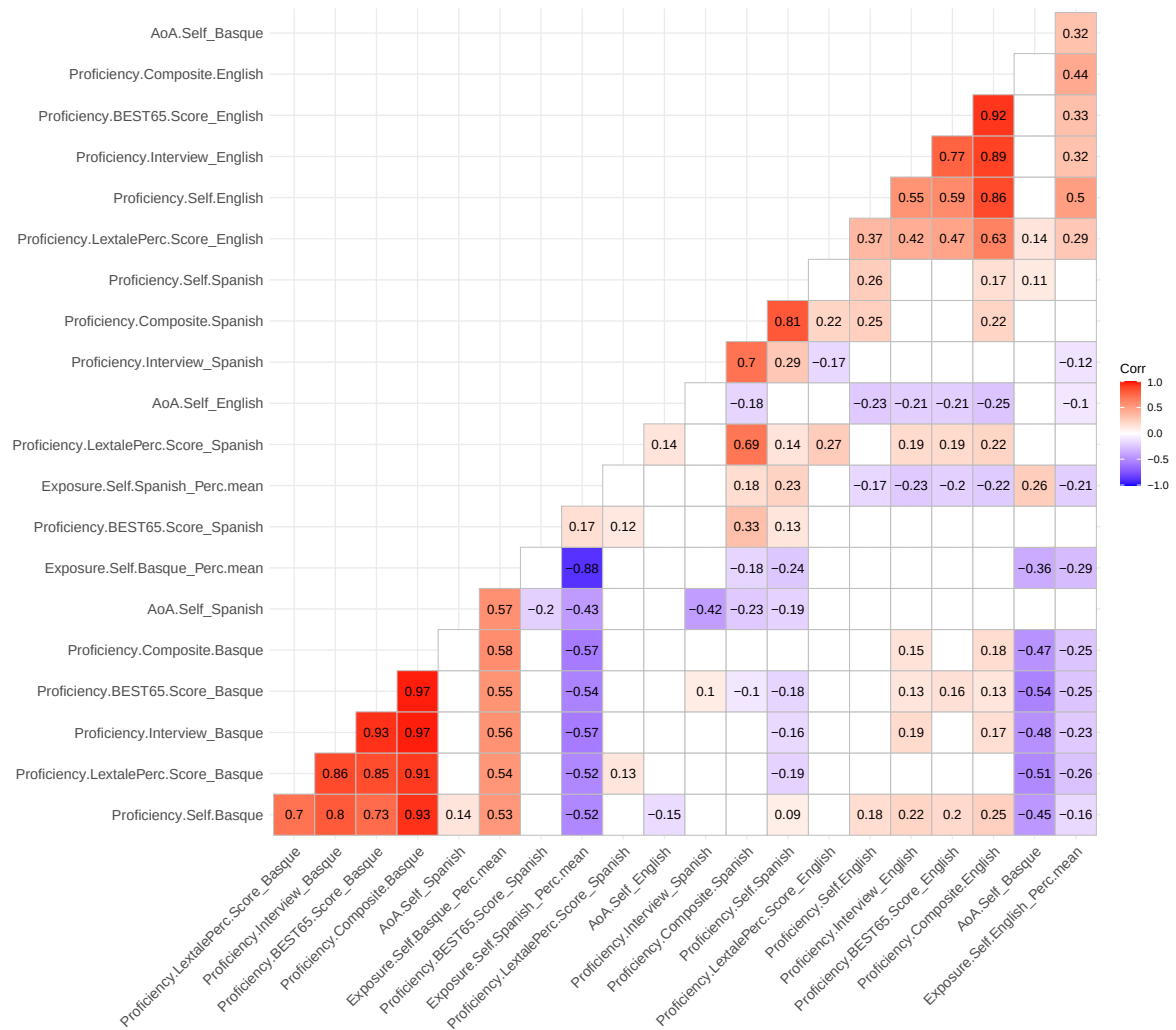


Figure S1: Correlation matrix of linguistic environment variables (Pearson's r). Non-significant values are blanked.

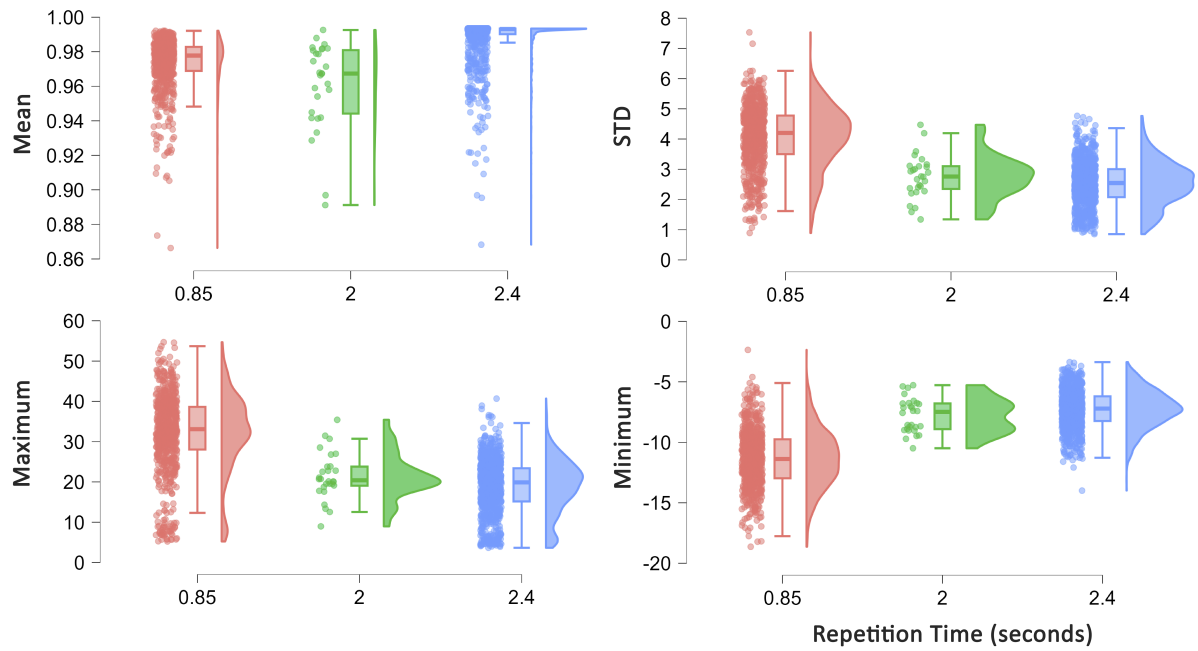


Figure S2: Summary statistics of the fMRI time series as a function of repetition time (RT). Violin plots combined with boxplots show the distribution of the mean, standard deviation, maximum, and minimum signal values for sequences acquired at RT = 0.85 s, 2 s, and 2.4 s.

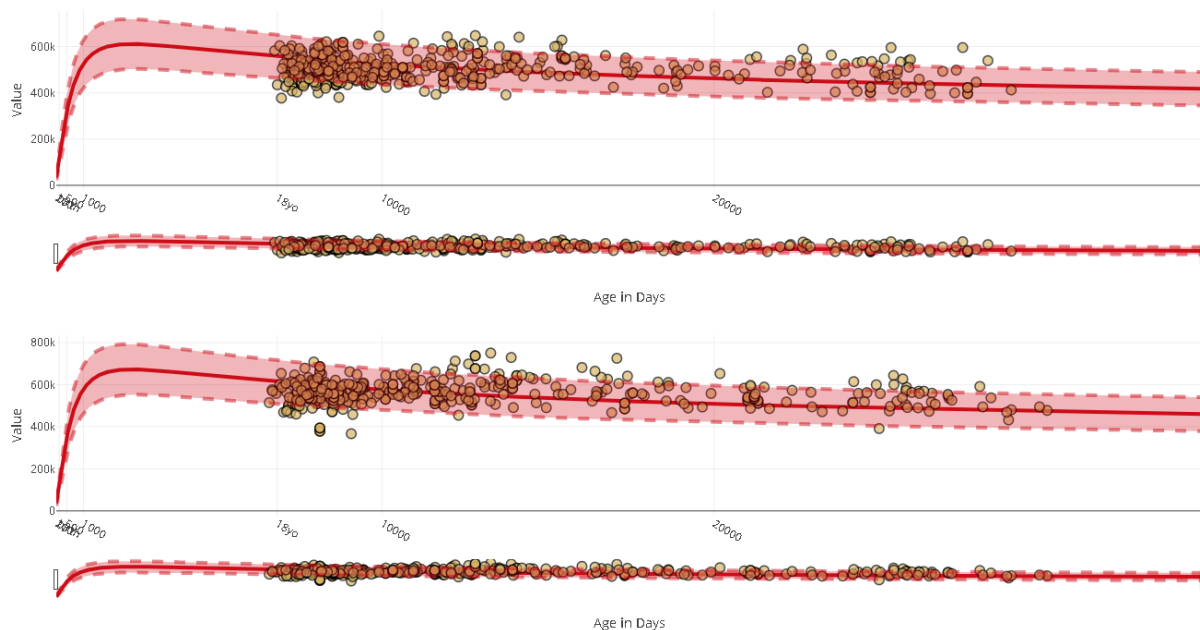


Figure S3: Gray matter volume relationship with age. Value= Gray matter volume. Dots= Gray matter volume (derived with CAT12) of the NeuroLingua participants. Red lines= normative brain-charts (derived with Freesurfer) through the brainchart interactive application <https://brainchart.shinyapps.io/brainchart/> Bethlehem et al., 2022. Upper panels represent data for females, and lower panels for males.

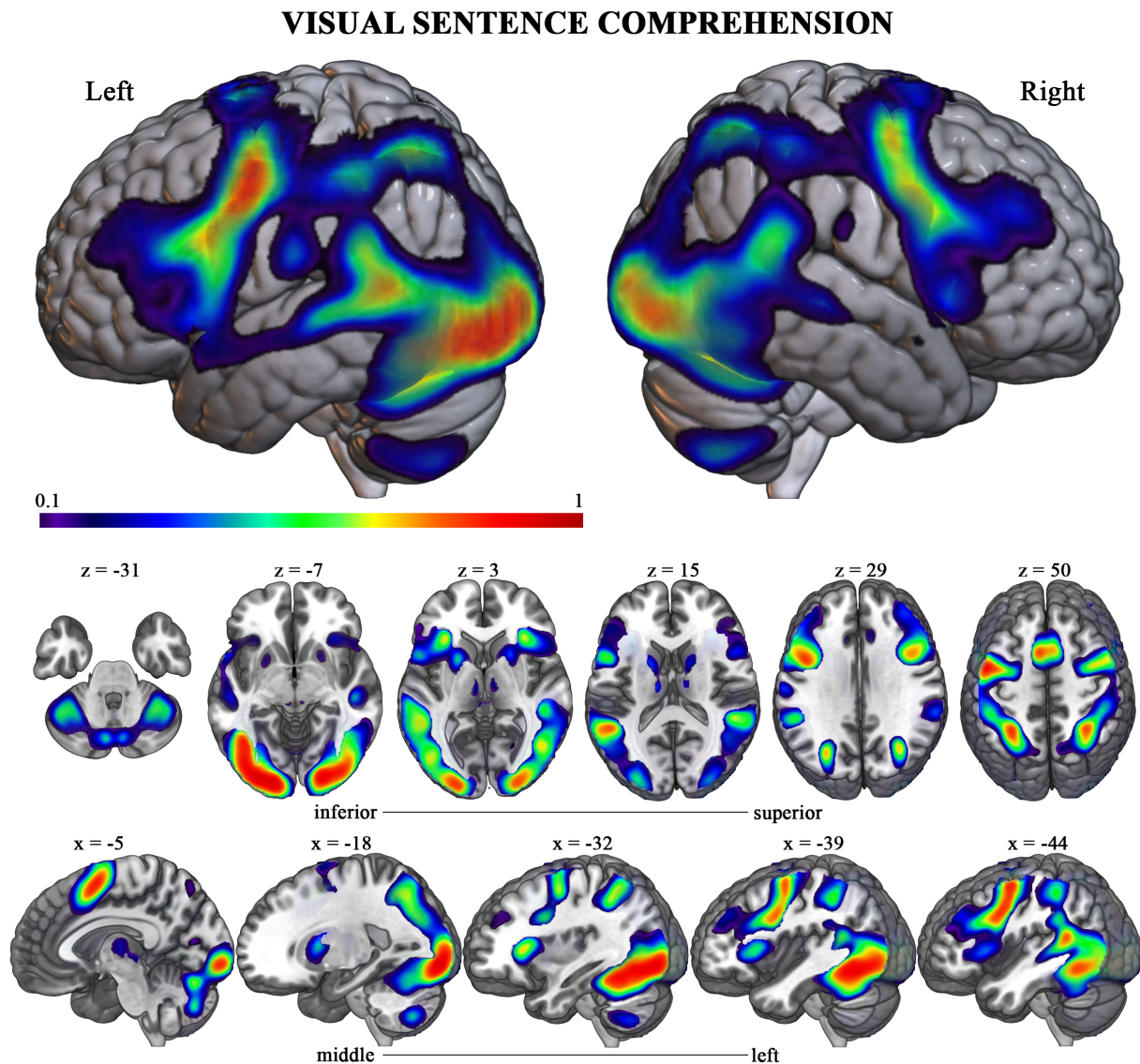


Figure S4: Visual language comprehension functional topography. Brain renderings and sagittal and axial slices showing probabilistic functional activation for the visual language comprehension > baseline contrast. Activation maps are based on overlaid individual binarized maps, with p-values computed from the same t-maps and combined using Stouffer's method. FDR correction was applied to establish the activation threshold ($q = 0.1$). This method yielded an alternative atlas that provides a more statistically robust selection of activated voxels.

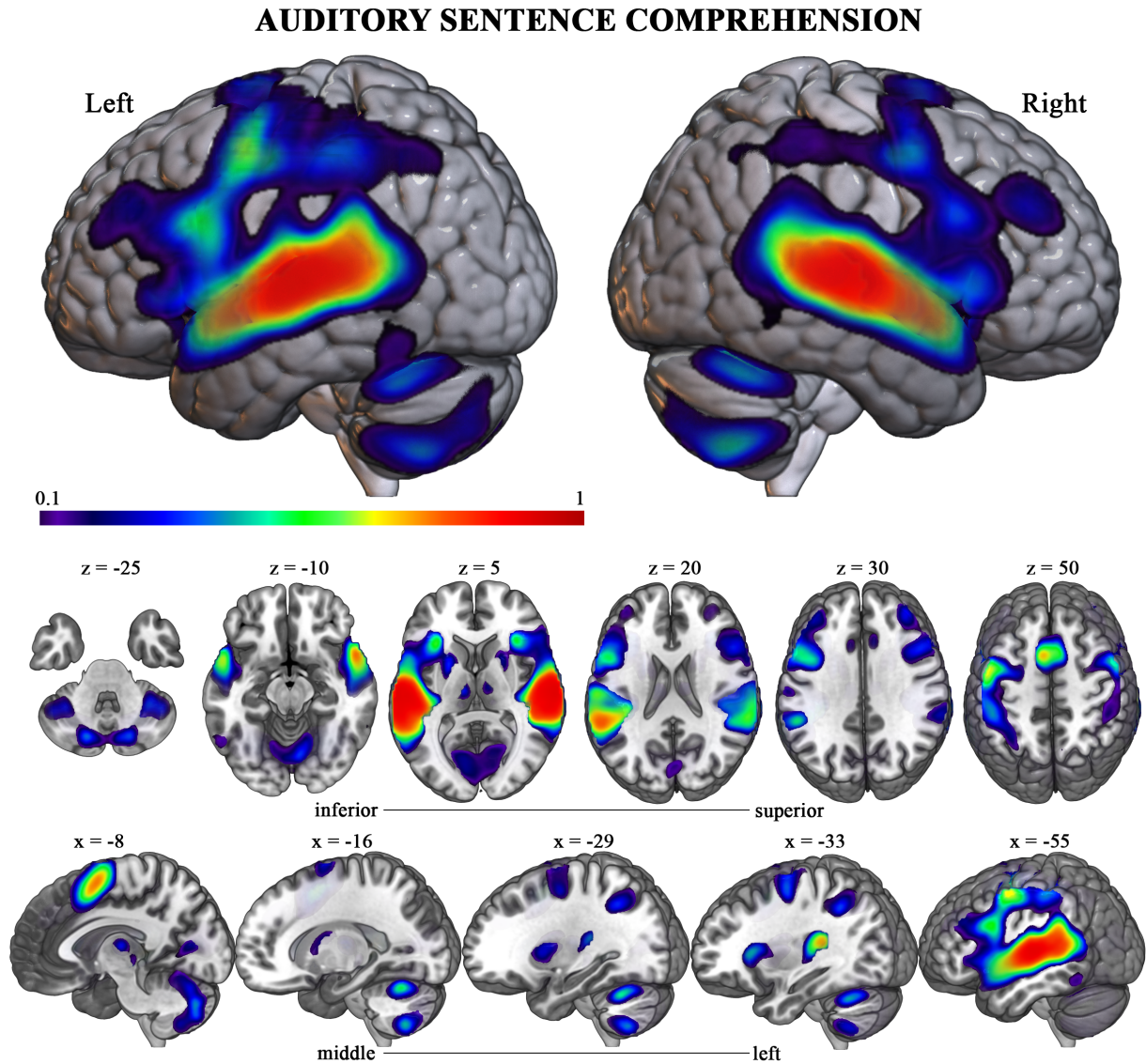


Figure S5: Auditory language comprehension functional topography. Brain renderings and sagittal and axial slices showing probabilistic functional activation for the auditory language comprehension > baseline contrast. Activation maps are based on overlaid individual binarized maps, with p-values computed from the same t-maps and combined using Stouffer's method. FDR correction was applied to establish the activation threshold ($q = 0.1$). This method yielded an alternative atlas that provides a more statistically robust selection of activated voxels.

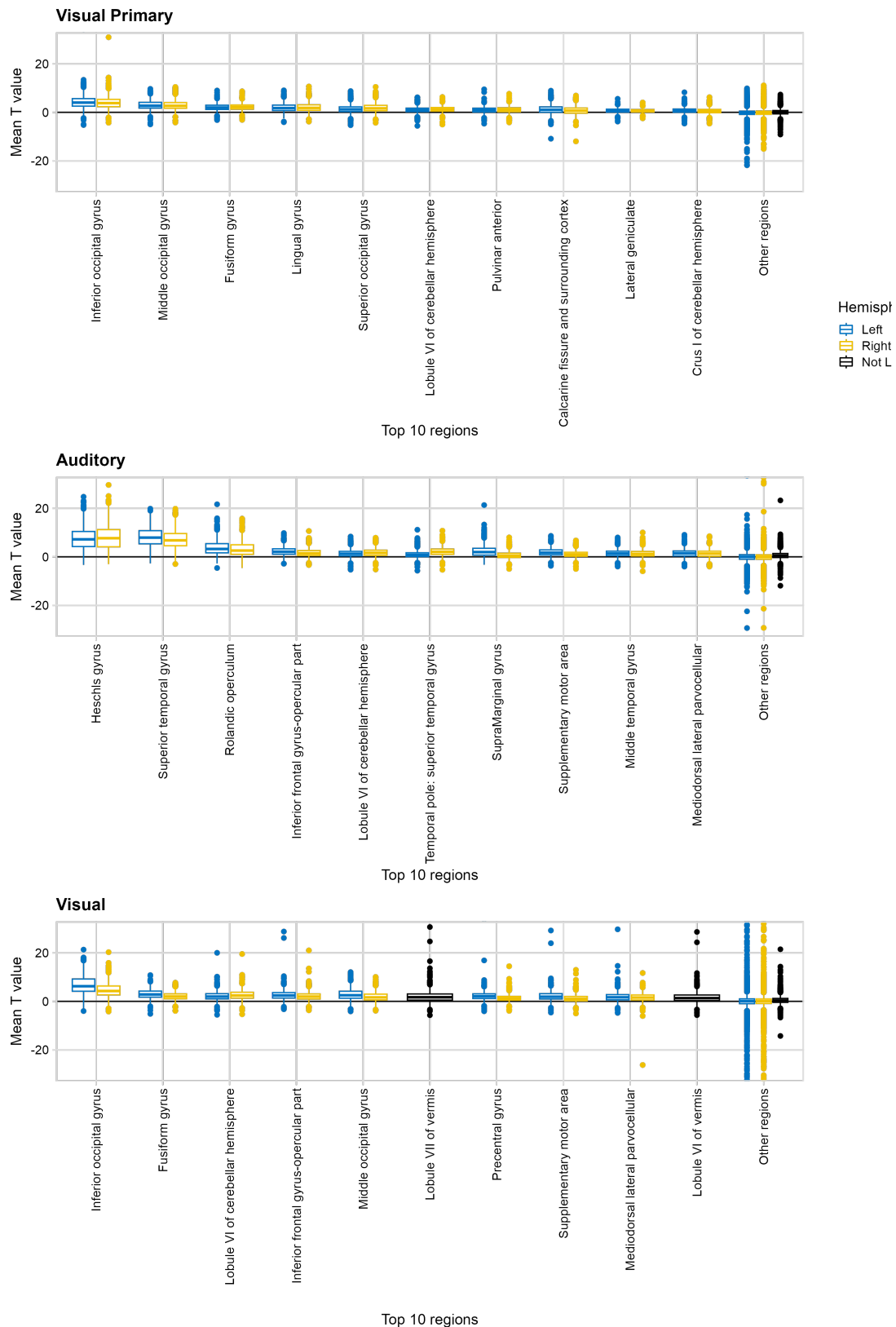


Figure S6: Distribution of mean T values across fMRI runs for the top 10 regions of the AAL3 atlas in the **Visual Primary**, **Auditory** and **Visual** contrasts. The top 10 regions were selected based on the mean T value. For visualization purposes, data were censored to exclude outliers with absolute T-values greater than 30. **Visual Primary**: checkerboards *versus* baseline. **Auditory**: all Auditory runs *versus* baseline. **Visual**: all Visual runs *versus* baseline.

Supplementary tables

Table S1: Total count and rate of missing values for phenotypic variables. N (non missing): number of subjects with non-missing data for that variable. Missing rate: N (non missing)/ N (total). A full data dictionary including Variable definition is provided as Table S3.

Category	Variable	N (non missing)	Missing rate
MRI	acqYear1	725.00	0.00
MRI	ImageCount_Spanish	717.00	1.10
MRI	ImageCount_Basque	294.00	59.45
MRI	ImageCount_English	57.00	92.14
demographics	age.Acq1	725.00	0.00
demographics	birthyear	725.00	0.00
demographics	gender	725.00	0.00
demographics	educational.attainment	610.00	15.86
demographics	eduYears	610.00	15.86
laterality	handedness	710.00	2.07
laterality	ehi_1	502.00	30.76
laterality	EHF_Writing	502.00	30.76
cognitive	KBIT.matrices	382.00	47.31
cognitive	KBIT.composite	346.00	52.28
cognitive	KBIT.vocabulary	346.00	52.28
cognitive	mini.mental	55.00	92.41
L1/L2/L3	L1	703.00	3.03
L1/L2/L3	Spanish	695.00	4.14
L1/L2/L3	Basque	538.00	25.79
L1/L2/L3	English	467.00	35.59
age of acquisition	AoA.Self_Spanish	694.00	4.28
age of acquisition	AoA.Self_Basque	563.00	22.34
age of acquisition	DiffEsEu.AoA	555.00	23.45
age of acquisition	AoA.Self_English	510.00	29.66
exposure	Exposure.Self.L1_Perc.mean	540.00	25.52
exposure	Exposure.Self.Spanish_Perc.mean	539.00	25.66
exposure	Exposure.Self.Spanish	528.00	27.17
exposure	Exposure.Self.Basque	477.00	34.21
exposure	DiffEsEu.Exposure	452.00	37.66
exposure	Exposure.Self.Basque_Perc.mean	452.00	37.66
exposure	Exposure.Self.English	435.00	40.00
exposure	Exposure.Self.English_Perc.mean	430.00	40.69
proficiency	Proficiency.Composite.Spanish	629.00	13.24
proficiency	Proficiency.Self.Spanish	591.00	18.48
proficiency	DiffEsEu.Proficiency	588.00	18.90
proficiency	Proficiency.Composite.Basque	588.00	18.90
proficiency	Proficiency.Composite.English	545.00	24.83
proficiency	Proficiency.Interview_Spanish	484.00	33.24
proficiency	Proficiency.Self.Basque	484.00	33.24
proficiency	Proficiency.Interview_Basque	470.00	35.17
proficiency	Proficiency.Self.English	460.00	36.55

proficiency	Proficiency.Interview_English	444.00	38.76
proficiency	Proficiency.BEST65.Score_Basque	441.00	39.17
proficiency	Proficiency.BEST65.Score_English	441.00	39.17
proficiency	Proficiency.BEST65.Score_Spanish	441.00	39.17
proficiency	Proficiency.LextalePerc.Score_Spanish	411.00	43.31
proficiency	Proficiency.LextalePerc.Score_Basque	399.00	44.97
proficiency	Proficiency.LextalePerc.Score_English	388.00	46.48
bilingualism	IndexTriLingA	703.00	3.03
bilingualism	IndexTriLingB	703.00	3.03
bilingualism	bilingualism	672.00	7.31
bilingualism	Lang.N	672.00	7.31
bilingualism	IndexBilingEsEu.composite	624.00	13.93
bilingualism	LingEnv.Self_Perc.multiling	408.00	43.72

Table S2: Summary statistics of signal per fMRI sequence depending on repetition time (sequence RT, seconds). Mean and standard deviation (SD) across runs within each RT sequence, for the mean, minimum (min), maximum (max) and SD signal per run.

sequence RT (seconds)	across runs within runs	Mean				SD	
		Mean	Min	Max	SD	Mean	SD
0.85		0.97	-13.09	37.73	4.68	0.04	16.04
2.00		0.96	-7.76	21.53	2.74	0.04	0.71
2.40		0.98	-7.22	18.88	2.53	0.06	0.73

Supplementary Dataset

The "Supplementary Dataset" (xlsx file) contains the following tables:

Table S3: Data dictionary for the phenotypic variables.

Table S4: Descriptive statistics for phenotypic variables.

Table S5: Phenotypic dataset.

References

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., ... Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, 604(7906), 525–533.