Auditory cortex anatomy reflects multilingual phonological experience

Olga Kepinska[1,2], Josue Dalboni da Rocha[3], Carola Tuerk[4], Alexis Hervais-Adelman[5], Florence Bouhali[6], David Green[7], Cathy J. Price[7], Narly Golestani[1,2,4]

1. Brain and Language Lab, Cognitive Science Hub, University of Vienna, Austria;
2. Department of Behavioral and Cognitive Biology, Faculty of Life Sciences, University of Vienna, Austria;
3. Department of Diagnostic Imaging, St Jude Children's Research Hospital, Memphis, USA;
4. Brain and Language Lab, Department of Psychology, Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland;
5. Neurolinguistics Division, Psychology Institute, University of Zurich, Switzerland;
6. Aix Marseille University, CNRS, LPC, Marseille, France;
7. Wellcome Trust Centre for Neuroimaging, University College London, UK.

ABSTRACT. This study explores the relationship between the anatomy of the auditory cortex and multilingual experience, shedding light on the complex mechanisms of auditory processing in humans. Integrating previous research on auditory information processing and the impact of bi- and multilingualism on brain structure, we investigate how the morphology of auditory brain regions reflects individuals' language experience and, more specifically, their phonological repertoire. Leveraging two distinct samples comprising over 200 participants, each exposed to between 1 and 7 languages encompassing 36 different languages, we explore whether the morphological variability of auditory brain regions reflects individual language experience, specifically focusing on the phonological repertoire. Additionally, we examine the potential influence of typological distances between languages spoken by multilingual individuals on the neural signatures of multilingualism within the auditory cortex. Our findings revealed a relationship between the average thickness of the left and right second transverse temporal gyrus (TTG) and participants' language experience. Model comparisons demonstrated that the inclusion of phoneme-level information in the language experience index accounted for the greatest variance in average TTG thickness (bilaterally). The direction of this effect was negative, indicating that a more extensive and more phonologically diverse language experience was associated with thinner cortices in the second TTG. Across two independent datasets, we provide robust evidence linking the degree of language experience and typological distance between languages to cortical thickness in early auditory brain regions. We propose that this phenomenon may reflect experience-driven pruning and neural efficiency, warranting further investigation through longitudinal studies of language acquisition. Moreover, our findings support the notion that experiences with typologically similar languages differ from those with typologically distant languages, offering valuable insights into the intricate relationship between language diversity and brain structure. Our findings also indicate that early auditory regions seem to represent phoneme-level cross-linguistic information, contrary to the most established models of language processing in the brain, which suggest that phonological processing happens in more lateral posterior STG and STS.

ABBREVIATIONS: STG: superior temporal gyrus, STS: superior temporal sulcus, PT: planum temporale, TTG: transverse temporal gyrus, HG: Heschl's gyrus (first transverse temporal gyrus), HS: Heschl's sulcus, PP: planum polare, L1: first language, AoA: age of onset of acquisition, ROI: region of interest

1 Introduction

Processing of complex auditory stimuli is fundamental to vocal communication. In humans with typical development, auditory processing has been shown to depend on multiple regions within the temporal cortex spreading across the superior and lateral aspects of the superior temporal gyrus (STG). The superior aspect of the STG hosts tonotopically organized primary auditory regions, which have been shown to be located in regions including the first transverse temporal gyrus (TTG), i.e., Heschl's gyrus (HG; Dick et al., 2012; Moerel et al., 2014). The planum polare (PP) is immediately anterior to HG, while the planum temporale (PT) borders it posteriorly. Of note, TTG shows high variability in shape and size, some individuals having one single gyrus (HG), others presenting duplications (with a common stem or fully separated) or multiplication of TTG (Geschwind and Levitsky, 1968; Marie et al., 2015). The superior temporal sulcus (STS) lies directly below the STG and separates it from the middle temporal gyrus. Understanding functional contributions of these different brain areas to processing of sound and vocal stimuli has been advanced by lesion studies (Hillis et al., 2017), ablation, and resection case studies (Hamilton et al., 2021; Hullett et al., 2022), functional magnetic resonance imaging (Booth et al., 2002; da Costa et al., 2011; Humphries et al., 2014), and intracranial recordings (Hamilton et al., 2021; Mesgarani et al., 2014). To date, however, no single functional parcellation of the human auditory cortical areas is generally agreed upon (Hamilton et al., 2021; Moerel et al., 2014). For example, in contrast to hierarchical models of auditory processing (e.g., Binder et al., 2000; Humphries et al., 2014; Scott and Johnsrude, 2003; Wessinger et al., 2000), recent proposals of sound processing postulate that processing of speech is distributed, and happens in parallel in different regions of the auditory cortex (Hamilton et al., 2021). According to parallel processing models, the posterior STG has been shown to be a crucial and essential locus for language and phonological processing (Bhaya-Grossman and Chang, 2022; Hillis et al., 2017), and to encode acoustic-articulatory features of speech sounds (Lakertz et al., 2021; Mesgarani et al., 2014). The primary auditory regions, including Heschl's gyrus, have been argued not to be necessary for speech perception (Hamilton et al., 2021; Hullett et al., 2022).

Structural brain imaging provides another route for gaining insight into the functional roles of cortical subregions, albeit at a different timescale. Quantifying individual differences in behavioral skill and/or language experience and relating them to cortical morphology can inform us about the relative influences of experience-dependent plasticity and of potential predisposition, in different domains. Of note, distinct influences (environmental *versus* genetic) tend to be reflected by different underlying anatomical characteristics of the brain. Indeed, a recent large-scale genome-wide association meta-analysis suggests that cortical surface area is relatively more influenced by genetics and that cortical thickness tends to reflect environmentally-driven neuroplasticity (Grasby et al., 2020). In the context of multilingualism, structural brain imaging can offer insights into whether and how the brain accommodates experience with different languages, and into what type of linguistic information is encoded, stored, and processed by particular brain structures. Previous findings point to an influence of experience-dependent plasticity on thickness of regions associated with speech processing, including the left posterior STS and left PT (Hervais-Adelman et al., 2017), as well as the STG (Mårtensson et al., 2012). Moreover, Ressel and colleagues (2012) associated lifelong bilingualism with increased

volume of the Heschl's gyrus, a finding interpreted as reflecting experience-related plasticity related to the acquisition of novel phonology.

Languages differ from each other in many ways and to different degrees: language similarities vary per linguistic domain, with distinct typological distances that can be computed between the languages' phonological, lexical, and syntactic systems. The present study investigates whether there are brain structural signatures of multilingualism in the auditory cortex, and whether typological distances between spoken languages further modulate those signatures. Given our focus on auditory brain regions, we quantified how the languages spoken by the participants differed from each other in terms of their phonological systems. Specifically, we investigated whether the neuroanatomical indices describing the auditory cortex regions were related to cross-linguistic phonological information at different levels: acoustic and articulatory feature-level, phoneme-level, or (more abstract) counts of phonological classes.

To investigate the relationship between the individual variability in the morphology of the auditory cortex and variability in language experience, we analyzed the anatomical brain scans of 204 healthy, right-handed participants from the PLORAS database (Seghier et al., 2016). The sample was split into two groups according to the date of data acquisition. The main sample included 136 participants speaking between 1 to 7 languages (2.65 languages on average), and representing a relatively wide range of linguistic diversity (34 different languages in total), see Figure 1 for a visual representation of the sample's language experience. The replication sample included 68 participants speaking to up to 5 languages, 2.44 languages on average. All MRI anatomical images were processed with FreeSurfer's brain structural pipeline (Fischl et al., 2004). Multilingualism was operationalized in a continuous way, without artificially dichotomizing the sample into multi-, bi-, and monolinguals (cf. DeLuca et al., 2019; Luk and Bialystok, 2013). Given the focus on auditory brain regions and on potential associations with language distance at the phonological level, and the fact that phonological perceptual space is known to be shaped relatively early in life (e.g., Werker and Hensch, 2015), we used information regarding the age of onset(s) of acquisition (AoA) of the different languages spoken by participants, and weighed each language by its AoA. For this, we operationalized the language experience in a continuous and quantitative measure for each participant, using Shannon's (1948) entropy equation (see Materials and Methods, Section 4.2), with high entropy values indicating more diverse language experience.

Our analyses aimed at a fine localization of the structural effects of multilingualism. We predicted (see Bhaya-Grossman and Chang, 2022; Hamilton et al., 2021; Lakertz et al., 2021; Mesgarani et al., 2014) that the cortical thickness of posterior STG would be related to the composite measures of language experience in our multilingual participants, and to the differences between their languages at the acoustic and articulatory feature level. In terms of the predicted direction of these effects, according to most current views, multilingualism is hypothesized to be a dynamic process reshaping brain structure and inducing both increases (in initial stages) and decreases (in peak efficiency) in brain morphological indices as a function of quantity of the multilingual experience (Pliatsikas, 2020). Furthermore, acquisition of multiple languages is thought to have additive effects on brain structure, and to induce cortical and subcortical adaptations in order to accommodate the additional languages (see e.g., Hervais-Adelman et al., 2018). The exact nature of such adaptations for multiple languages of phonologically diverse repertories on auditory brain regions is to date unknown. A positive relationship between the thickness of auditory regions and the

degree of multilingualism (weighted by greater phonological distance) would indicate that the auditory regions need to expand in order to accommodate the variability of phonological input in the environment; a negative one would suggest a specialization of the auditory system in relation to such variability (Pliatsikas, 2020).

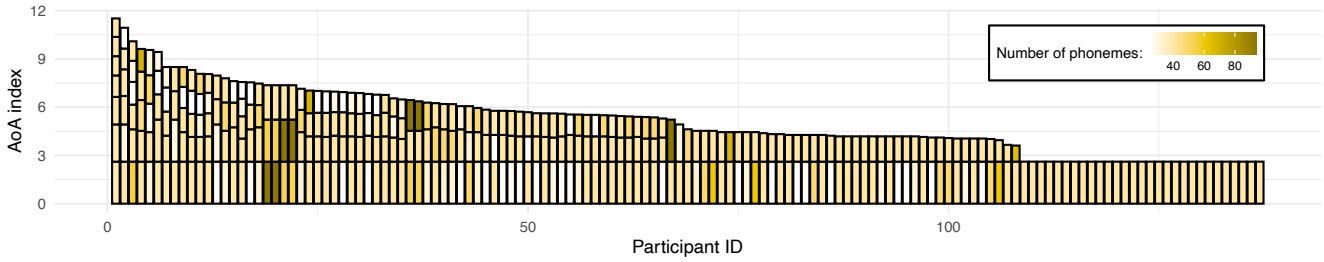

Figure 1. Illustration of the sample's language experience. Each bar represents a single participant's overall language experience; the height of the stacked bars within each bar represents the AoA index for individual languages (the taller the bar, the earlier in life a given language was acquired). The color of each stacked bar refers to the number of phonemes in each language's phonological inventory. For reference, English phonological inventory has 40 phonemes (Stanford Phonology Archive, 2019a). Prior to plotting, data was sorted by the overall language experience based on a sum of AoA index for participants' individual languages; consequently, data of participants with most diverse language experience can be found on the left-hand side of the figure, and the right-hand side includes data from monolinguals (i.e., exposed only to one language).

2 Results

Table 1 provides an overview of the performed analyses and of results.

	ANALYSIS	REGION(S)	MEASURE(S)	MOTIVATION	RESULTS
1.	Language experience and auditory regions (Section 2.1)	Auditory regions segmented by FreeSurfer (anterior STG, posterior STG, anterior STS, posterior STS, HG, HS, PT and PP)	cortical volume, surface area and thickness	Exploratory first step	Relationship between the average thickness of the PT (bilaterally) and language experience (p = .01)
2.	Language experience and TTG (Section 2.2)	Transverse temporal gyri segmented by TASH	cortical volume, surface area and thickness and number of gyri	To elucidate above exploratory results	Relationship between the average thickness of the 2nd TTG (bilaterally) and language experience (p = .01)
3.	Vertex-wise analysis (Section 2.2)	Whole-brain	cortical thickness	To confirm above exploratory results	Cluster of vertices negatively related to participants' language experience at p < .0001 (uncorrected), located in the superior aspect of the left STG, corresponding to the location of the second TTG
4.	Effects of language proficiency (Section 2.2)	2nd TTG	cortical thickness	To confirm the results obtained from AoA-based language experience index	Relationship between the average thickness of the 2nd TTG (bilaterally) is confirmed using a different metric to calculate language experience (p = .018 for left, and p = .065 for right hemisphere)
5.	Language typology and 2nd TTG (Section 2.3)	2nd TTG	cortical thickness	To assess the effect of language typology within the identified region	Quantifying the distance between languages based on phoneme-level information explains the most variance in the average thickness of the bilateral second TTG
6.	Language experience expressed as 'cumulative phoneme inventory' (Section 2.3)	2nd TTG	cortical thickness	To account for language experience in participants with different L1s	Relationship between the thickness of the left 2nd TTG and the specific characteristics of languages at the phoneme-level of their phonological inventories (p = .004)

7.	Language experience in participants with a single TTG (Section 2.4)	Right HG (1st TTG) and right PT	cortical thickness	To assess the effect of language experience in participants without multiple TTGs	Positive relationship between the thickness of right HG and language experience (p = .007); no relationship between the thickness of right PT and language experience (p = .81)
8.	Replication analysis (Section 2.5)	2nd TTG	cortical thickness	To replicate the main result (analysis #2 and #5) in an independent sample	Relationship between the average thickness of the 2nd TTG and language experience partially replicated (for left hemisphere, p = .047)

Table 1. Overview of the performed analyses.

2.1 Auditory cortex and language experience

In a first, exploratory analysis, we investigated relationships between the composite language experience measure and the volume, surface area, and average thickness of the following auditory subregions: STG, STS, HG, HS, PT and PP, as parcellated by FreeSurfer (Destrieux et al., 2010). Given the large size of STG and STS, we refined their segmentations by dividing the STG into anterior and posterior part, and the STS into anterior, middle, and posterior parts (see Materials and Methods, Section 4.6 for details). Using the `lme4` R package (Bates et al., 2015), we fit three linear mixed models to the extracted anatomical measures (volume, surface area, average thickness), with participants modeled as random effects, language experience, region of interest (ROI), and hemisphere as fixed effects, controlling for the covariates of age, sex and whole-brain volume, area, or mean thickness. Interaction terms for language experience, ROI, and hemisphere were included in the models to determine whether language experience would differentially affect any of the segmented regions. Significance was calculated using the `lmerTest` package (Kuznetsova et al., 2017), which applies Satterthwaite's method to estimate degrees of freedom and generate p-values for mixed models. Out of all investigated cortical measures, only average thickness of the PT, bilaterally, was related to participants' language experience at p = .01, see Table S 1. Specifically, participants with more extensive and diverse language experience had a significantly thinner PT.

2.2 Superior temporal plane and language experience

Given that full TTG duplications, triplications etc. by definition belong to the PT, and that previous work has shown relationships between TTG duplications patterns and language aptitude (Turker et al., 2017), we wanted to better pinpoint the location of the above effects within PT subregions. To this end, and to gain more detailed insight into the localization of the relationship between the thickness of the PT and participants' language experience, we ran three follow-up analyses, an ROI-based one, a vertex-wise based one, as well as an analysis investigating the effect of language proficiency.

First, we used an automatic toolbox (TASH; Dalboni da Rocha et al., 2020) to segment gyri along the superior temporal plane (HG, or 1st TTG, and additional TTG(s), when present), and to extract their cortical thickness, surface area and volume. Note that FreeSurfer does not segment additional TTGs specifically, but provides measures for the PT as a whole. We included HG in these follow-up analyses (i.e., even though we had included the Freesurfer HG ROI in

the above, broader analyses), given that the default FreeSurfer pipeline is not fine-tuned for the segmentation of this small and variable region, and is thus error-prone for this ROI (Dalboni da Rocha et al., 2020). As expected from previous work, the TTG showed large individual variability in overall duplication patterns in the sample. Segmentations revealed different numbers of gyri across hemispheres and participants, ranging from a single gyrus to four identified TTGs, see Table 2. On average, participants had 2.375 gyri in the left hemisphere (*SD* = 0.63), and 1.84 in the right (*SD* = 0.64). The number of gyri was not related to participants' language experience, either in the left (β = 0.10, t = 0.97, p = .335), or in the right hemisphere (β = -0.01, t = -0.13, p = .90), according to a linear model with number of gyri as dependent and language experience as independent variables (controlling for participants' age and sex).

	left				*right*				**whole sample:**
total number of gyri:	1	2	3	4	1	2	3	4	
N	6	78	47	5	40	78	18	0	**136**
Male/Female	0/6	28/50	23/24	3/2	10/30	33/45	11/7	-	**54/82**
Age (mean)	42.95	35.95	36.29	28.48	35.12	36.49	36.63	-	**36.11**
Language experience index (mean)	0.69	0.79	0.83	0.93	0.85	0.80	0.76	-	**0.807**
Language experience index (min)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	**0.000**
Language experience index (max)	1.09	1.89	1.76	1.56	1.76	1.89	1.34	-	**1.893**

Table 2. Number of participants, their demographic and language experience characteristics displaying different overall shapes of the TTG (i.e., total number of identified gyri in the left and right hemisphere). Last column lists whole sample's descriptive statistics.

To localize the effect of language experience on the cortical thickness of the segmented gyri, we fit another linear mixed model with participants modeled as random effects, language experience, gyrus (first, second, third), and hemisphere as fixed effects, controlling for the covariates of age, sex, and mean thickness (including interaction terms for language experience, gyrus, and hemisphere). We also ran two additional models on the volume and surface area values of the segmented regions, to confirm that the results were specific to the average thickness values, as in the model above (including all auditory regions, see Section 2.1). Out of all investigated cortical measures, only average thickness of the second TTG (bilaterally) was related to participants' language experience at $p < .01$, see Table S 2.

Second, to confirm the above result, we also ran a whole-brain vertex-wise analysis in FreeSurfer. We fit a general linear model to the cortical thickness surface maps of all subjects (smoothed with a 5-mm kernel), testing for a significant effect of the language experience index across the whole brain. We found one cluster of vertices negatively related to participants' language experience at $p < .0001$ (uncorrected), located in the superior aspect of the left STG, corresponding to the location of the second TTG, see Figure S 4. At the same threshold, no vertices were found in the right hemisphere, possibly due to greater known anatomical variability in right compared to left auditory regions (cf. Dalboni da Rocha et al., 2020).

We also explored whether the proficiency attained in the spoken languages was related to the thickness of the 2nd TTG. Analyses calculating cumulative language experience based on the participants' proficiency instead of AoA yielded similar results as above, with a significant negative relationship between the proficiency-based cumulative language experience index and the thickness of the left 2nd TTG, and a trend towards significance for the right 2nd TTG (see Section 7.6 for details).

2.3 Second Transverse Temporal Gyrus and effects of Language Typology

Building on the above result, we investigated whether the thickness of participants' 2nd TTG was related to cross-linguistic phonological information describing the languages they spoke, *above and beyond* being related to mere experience of speaking several languages. The analysis below was therefore performed on a sub-sample of participants who had multiple TTGs (n = 130 in the left and n = 96 in the right hemisphere). We constructed three measures of typological distance between the languages spoken by our participants: (1) overlaps in distinctive acoustic and articulatory features describing the phonemes of each language (e.g., "short", "long"); (2) overlaps in phoneme inventories (meaning-distinguishing sounds) of each language; and (3) similarity in counts of phonological classes that share certain features (e.g., counts of "front rounded vowels", "clicks"), see Material and Methods for details. Next, using Rao's quadratic entropy equation (1982), we weighted the distance for each pair of languages by the AoA of each language (with higher weights for lower AoA), and the resulting values were summed, per participant. This calculation resulted in three indices of language experience accounting for different typological relations between languages (see Kepinska et al., 2023 for a similar approach applied to lexical distances). We used these indices as dependent variables in a set of multiple regression analyses fit to the average thickness values of the left and right second TTG, and performed a model comparison procedure to gauge which of the typological measure was most related to the thickness of the second TTG. Model comparison was performed by computing differences in explained variance (ΔR^2 *Adjusted*) compared to the baseline model (i.e., with language experience without typological information), and using the Bayesian information criterion (BIC) (Schwarz, 1978), where the difference between two BICs was converted into a Bayes factor using the below equation, following Wagenmakers (2007):

$$BF_{10} = \exp\frac{\Delta BIC_{01}}{2}$$

The model without typological information served as baseline for the comparisons. The analyses showed that out of the three language experience indices (feature-level, phoneme-level and counts of phonological classes), the model containing **phoneme** level information explained the most variance in the average thickness of the bilateral second TTG. It also outperformed the model containing the language experience index not accounting for any typological information, see Table 3. None of the other phonological distance measures combined with language experience performed better than language experience alone. The direction of the relationship between language experience weighted by typological distance was negative, and had a small effect size (β = -0.35, t = -2.96, p = .004, f^2 = 0.07 and β = -0.26, t = -1.98, p = .05, f^2 = 0.04 for left and right hemisphere, respectively), showing that the more extensive one's language experience, and the more varied at the phoneme level one's languages are, the thinner the second TTG cortex (see Figure 2). The statistics for the overall regression models are presented in the Supplementary Materials, Section 7.4.

		left				*right*			
Language experience models:		No typology	Features	Phonemes	Phonological classes	No typology	Features	Phonemes	Phonological classes
(Intercept)	β	0.00	-0.01	0.00	-0.01	0.01	0.01	0.01	0.01
	SE	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
Language Experience	β	-0.12**	-0.33	-0.35**	-0.60*	-0.10+	-0.25	-0.26+	-0.44
	SE	(0.05)	(0.23)	(0.12)	(0.29)	(0.05)	(0.27)	(0.13)	(0.33)
Age	β	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	SE	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean thickness (left/right)	β	1.40***	1.41***	1.37***	1.41***	0.94*	0.99*	0.92*	0.99**
	SE	(0.30)	(0.31)	(0.30)	(0.30)	(0.37)	(0.38)	(0.37)	(0.37)
Sex	β	0.00	0.01	0.00	0.01	-0.01	-0.01	-0.01	0.00
	SE	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Num.Obs.		*130*	*130*	*130*	*130*	*96*	*96*	*96*	*96*
R2		*0.20*	*0.17*	*0.21*	*0.18*	*0.16*	*0.13*	*0.16*	*0.14*
R2 Adj.		*0.17*	*0.14*	*0.18*	*0.16*	*0.12*	*0.09*	*0.12*	*0.10*
AIC		*18.5*	*23.5*	*16.9*	*21.2*	*11.9*	*14.9*	*11.8*	*14.0*
BIC		*35.7*	*40.7*	*34.1*	*38.4*	*27.3*	*30.3*	*27.2*	*29.4*
Log.Lik.		*-3.23*	*-5.75*	*-2.43*	*-4.62*	*0.04*	*-1.47*	*0.10*	*-0.99*
F		*7.80*	*6.32*	*8.29*	*6.98*	*4.30*	*3.46*	*4.33*	*3.72*
RMSE		*0.25*	*0.25*	*0.25*	*0.25*	*0.24*	*0.25*	*0.24*	*0.24*
ΔR^2 Adjusted		-	*-0.03*	*0.01*	*-0.02*	-	*-0.03*	*0.00*	*-0.02*
BF_{10}		-	*0.08*	*2.23*	*0.25*	-	*0.22*	*1.08*	*0.36*

. $p < 0.1$, + $p = 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Left and right second transverse temporal gyri and language experience. Multiple regression model parameters (parameter estimates and standard errors, in brackets; *p*-values are listed according to the coding presented underneath the table) for the average cortical thickness values of the second TTG, as predicted by the four language experience indices: (1) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (2) acoustic/articulatory features, (3) phonemes, and (4) counts of phonological classes. Last two rows present model comparison results (additional variance explained and BF_{10} values). *NB*. all models including typological information were compared against the 'No typology' model.

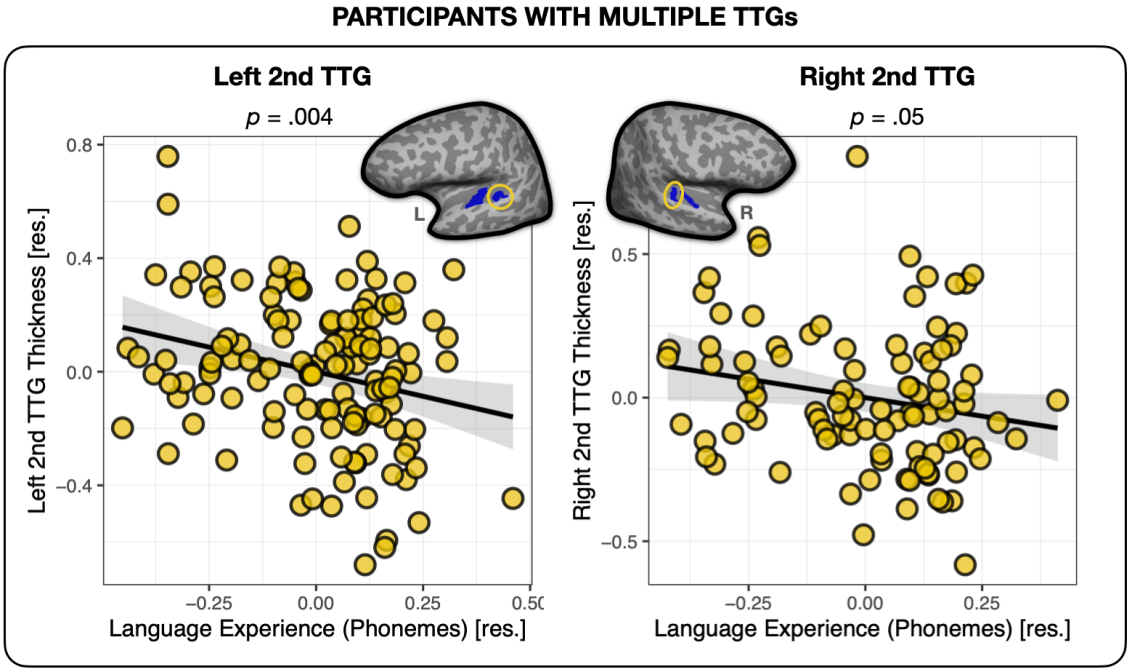

Figure 2. Multilingual language experience and thickness of the second TTG. Average thickness of the second TTG in the left and right hemisphere were negatively related to the multilingual language experience index weighted by their phoneme-level phonological distances. Plots show residuals, controlling for age, sex and mean hemispheric thickness.

To further probe how cross-linguistic sound inventories may modulate the effect of language experience on the brain, and to account for language experience in participants with different mother tongues (L1s), we constructed a cross-linguistic description of individual languages (see Methods, Section 4.3.2) and used it to calculate a measure of cumulative phoneme inventory for each participant. Here, we counted how many unique phonemes each participant was exposed to across all their languages (i.e., phonemes that overlapped between languages were counted only once), irrespective of when these languages were acquired. We used this measure as an independent variable in another linear model fit to the average thickness values of the second TTG (controlling for age, sex, hemispheric thickness, and accounting for language experience irrespective of language typology) (see Section 4.3.2, Supplementary Materials, Section 7.4.1 and Table S 3). In the left hemisphere, this measure explained variance in the thickness of the second TTG *above and beyond* language experience ($\beta = -0.004$, $t = -2.910$, $p = .004$), with a small effect size ($f^2 = 0.06$). In the right hemisphere, the thickness of the second TTG was not significantly related to the 'cumulative phoneme inventory' measure ($p = .76$), when also accounting for overall language experience irrespective of typology. This result shows that the thickness of the left second TTG in these participants was related to the specific characteristics of languages at the phoneme-level of their phonological inventories, irrespective of when these languages were acquired, while the thickness of the right second TTG was relatively more related to <u>when and how many</u> languages were learned (even though the typological distance information did improve the model fit, see Table 3), see Figure S 5 and Table S 3, and Supplementary Materials, Section 7.4.

2.4 Language experience in participants with a single TTG

So far, we revealed associations between phonological language experience and the cortical thickness of the 2nd TTG bilaterally in multilinguals. Yet, 40 participants out of 136 (29.4%) had a single TTG (Heschl's gyrus) in the right hemisphere (*versus* only 6 participants, i.e., 4.4% in the left; see Table 2). We therefore repeated the above analysis in the 40 participants with a single TTG in the right hemisphere in order to investigate the relationship between TTG anatomy and language experience in this sample. The analysis revealed that in participants without posterior TTGs in the right hemisphere, the thickness of the first TTG was indeed related to their language experience. The nature of this relation proved however to be different than that for multilingual participants with multiple TTGs. First, the language experience index accounting for most variance in the thickness of the first TTG was the index <u>not</u> accounting for typological relations between multilinguals' languages, see Table S 4. Second, the direction of this effect was inverse to that found for participants with multiple TTGs: the more extensive one's language experience, the thicker the HG ($\beta = 0.15$, $t = 2.89$, $p = .007$; medium effect size: $f^2 = 0.237$), see Figure 3. Of note, we did not find any significant relationship between the thickness of the PT and language experience in this sub-group of participants ($\beta = 0.02$, $t = 2.43$, $p = .81$). In addition, for PT thickness, compared to a model including only covariates of no-interest, the model including the language experience index offered moderate evidence against it ($BF_{10} = 0.16$), suggesting that the observed effects between cortical thickness and language experience in the superior temporal plane are specific to gyri only, see Table S 4, and Supplementary Materials, Section 7.5.

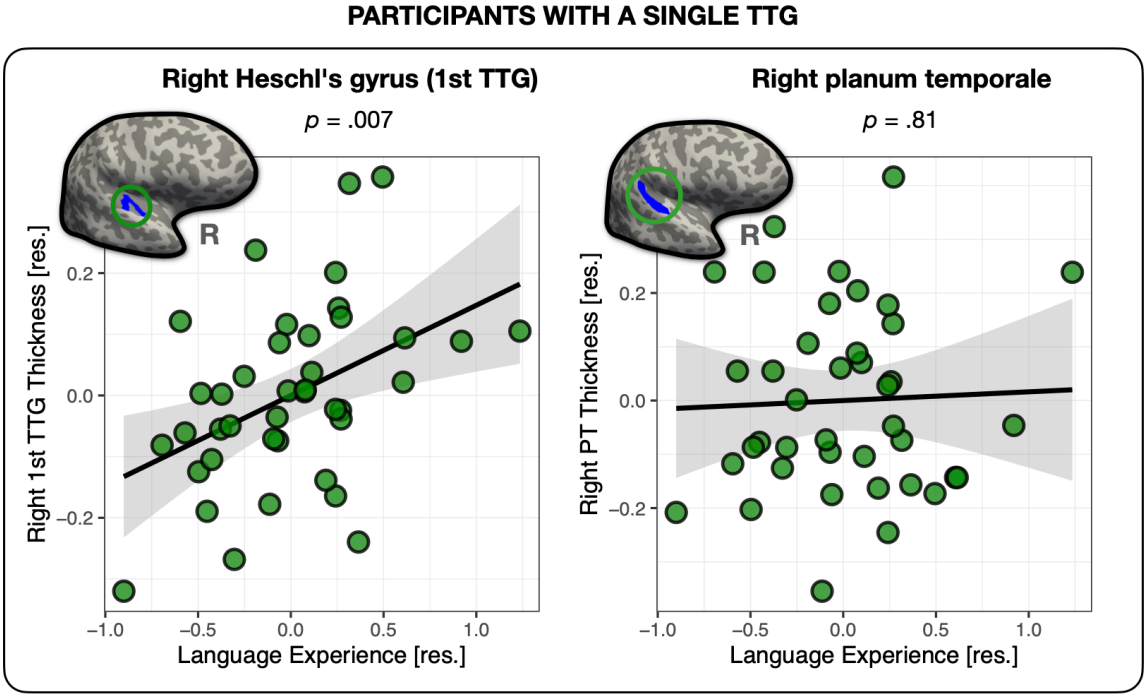

Figure 3. Thickness of the right first TTG and PT in participants with a single TTG in the right hemisphere. Average thickness of HG in the right hemisphere was positively related to the amount of multilingual experience, irrespective of typological relations between languages (left panel). The average thickness of the right PT was not related to language experience (right panel). Plots show residuals, controlling for age, sex, and mean hemispheric thickness.

2.5 Replication analysis

In an effort to strengthen the inferences that can be made from the above results, we tested whether the relationship between multilingual language experience and thickness of the second TTG could be replicated in an independent sample of 68 participants. Using linear regression, we fit models to the average thickness values of their left and right second TTG. As in the analyses on the original sample, we first fit a model using the language experience index irrespective of typology, and then we fit models that additionally accounted for the three types of typological relations between languages (features, phonemes, and counts of phonological classes), controlling for covariates of age, sex, scanner model, and mean hemispheric thickness. We again observed that language experience did significantly predict the average thickness of the second TTG, but only in the right hemisphere (β = -0.13, t = -2.05, p = .046; small effect size: f^2 = 0.087), see Table 4. Furthermore, similarly to the analysis on the main sample (see Section 2.3), the language experience index weighed by phonological information at the phoneme level, showed a better fit than the model without typological information, explaining 3% of additional variance in the average thickness of the right second TTG. In contrast to the results of the main sample, however, this time the model including phonological feature-level information had a better fit to the thickness data and explained an additional 14% of the variance (11% more variance than the model with language experience weighted by phoneme-level distances), see Table 4 and Figure 4C. No language experience indices were significantly related to the average thickness values of second TTG in the

left hemisphere (β = -0.03, t = -0.53, p = .60, and β = -0.09, t = -0.58, p = .57, for language experience index without and with typological phoneme-level information, respectively), see Figure 4 and Table 4.

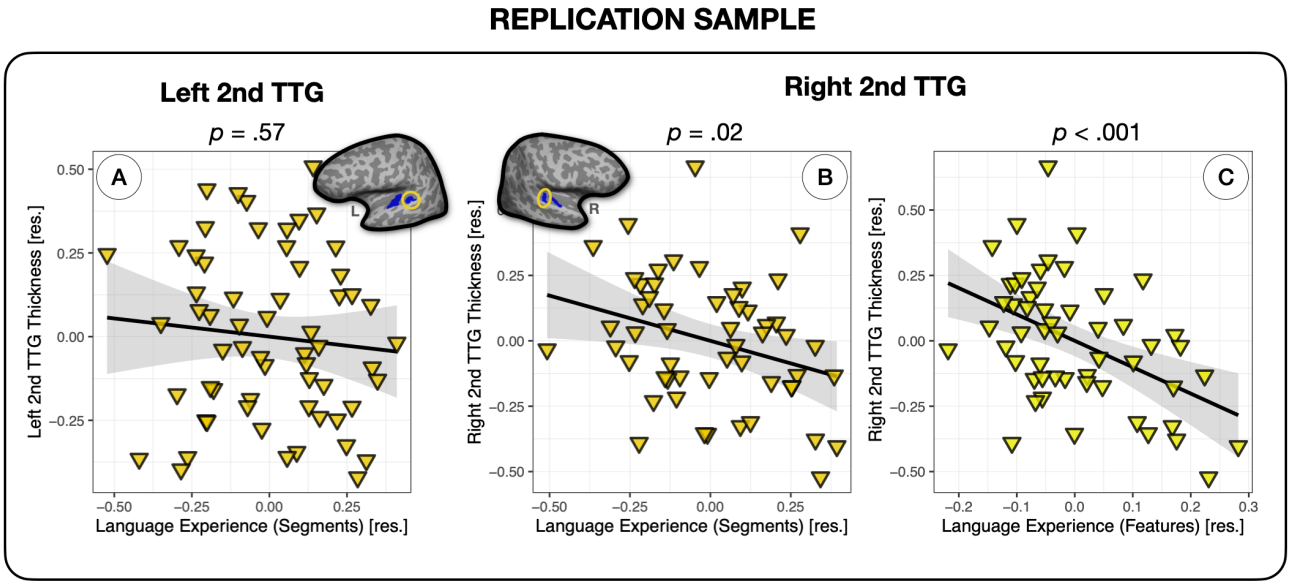

Figure 4. Multilingual language experience and thickness of the second TTG in an independent sample of participants. (A) Average thickness of the left second TTG was not significantly related to the language experience index; average thickness of the right second TTG was significantly related to the language experience indices accounting for phoneme-level phonological overlaps between multilinguals' languages (B) and feature level information (C). The model including phonological feature level information presented in panel (C) had the best fit to the average thickness data of the right second TTG.

		left				*right*			
Language experience models:		No typology	Features	Phonemes	Phonological classes	No typology	Features	Phonemes	Phonological classes
(Intercept)	β	1.00	1.12	0.99	1.24	-1.85	-2.20.	-2.09	-1.21
	SE	(1.14)	(1.10)	(1.13)	(1.09)	(1.34)	(1.21)	(1.33)	(1.30)
Language Experience	β	-0.03	-0.09	-0.09	0.32	-0.13*	-1.02***	-0.38*	-0.75
	SE	(0.06)	(0.28)	(0.15)	(0.49)	(0.06)	(0.27)	(0.15)	(0.49)
Scanner	β	-0.08	-0.09	-0.08	-0.09	0.10	0.09	0.11	0.09
	SE	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)
Age	β	0.00	0.01	0.00	0.01	0.01*	0.01**	0.01*	0.01*
	SE	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sex	β	0.09	0.09	0.09	0.07	0.00	0.03	0.01	-0.02
	SE	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)
Mean thickness (left/right)	β	0.57	0.52	0.58	0.42	1.72**	1.86***	1.81***	1.52**
	SE	(0.44)	(0.42)	(0.44)	(0.42)	(0.52)	(0.46)	(0.51)	(0.50)
Num.Obs.		*61*	*61*	*61*	*61*	*54*	*54*	*54*	*54*
R2		*0.092*	*0.089*	*0.092*	*0.094*	*0.225*	*0.350*	*0.253*	*0.196*
R2 Adj.		*0.009*	*0.006*	*0.010*	*0.012*	*0.144*	*0.282*	*0.175*	*0.112*
AIC		*13.9*	*14.1*	*13.8*	*13.7*	*7.6*	*-1.9*	*5.6*	*9.5*
BIC		*28.7*	*28.9*	*28.6*	*28.5*	*21.5*	*12.0*	*19.5*	*23.5*
Log.Lik.		*0.047*	*-0.046*	*0.077*	*0.130*	*3.205*	*7.947*	*4.196*	*2.229*
F		*1.109*	*1.072*	*1.120*	*1.142*	*2.781*	*5.158*	*3.244*	*2.342*
RMSE		*0.24*	*0.24*	*0.24*	*0.24*	*0.23*	*0.21*	*0.22*	*0.23*
ΔR^2 Adjusted		-	*> -0.01*	*< 0.01*	*< 0.01*	-	*0.14*	*0.03*	*-0.03*
BF_{10}		-	*0.91*	*1.03*	*1.09*	-	*114.68*	*2.70*	*0.38*

. $p < 0.1$, + $p = 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Thickness of left and right second TTGs and language experience in an independent sample of participants. Multiple regression model parameters (parameter estimates and standard errors, in brackets; *p*-values are listed according to the coding presented underneath the table) for the average cortical thickness of the second TTG, as predicted by the four language experience indices: (1) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (2) features, (3) phonemes, and (4) counts of phonological classes. The last two rows present model comparison results (additional variance explained and BF_{10} values). NB. all models including typological information were compared against the 'No typology' model.

3 Discussion

Multilingual language experience, quantified in a continuous manner, is related to the structure of the auditory cortex. Using two independent samples of participants, we find replicable effects showing that multilingual language experience, weighed by the age of acquisition of the respective languages, is specifically related to the thickness of gyri (TTGs) of the superior temporal plane. Moreover, we show that the typological distances between multilinguals' languages, in particular typological distance measures based on phonemes of the different languages (and to some degree on their acoustic and articulatory features), explains variation in TTG thickness *above and beyond* the effect of AoA alone.

The effects we find are specific to the second transverse temporal gyrus in people having at least one full posterior duplication, and to the first TTG in people having only one gyrus. The anterior-most TTG, i.e. HG, is the first cortical relay station for auditory input. HG and the TTG as a whole exhibit large individual variation in shape and size: some individuals have a single TTG, but duplications (complete or partial) and multiplications of the gyrus are also common, and have been observed both *ex-vivo* (Geschwind and Levitsky, 1968) and *in-vivo* (Marie et al., 2015). The TTG are known to be formed before birth, *in utero* (Chi et al., 1977). Further, genes known to be involved in speech processing,

in the development of the nervous system, and in X-linked deafness, have been related to the surface area and thickness of the left and right HG (Cai et al., 2014). A body of work has shown relationships between both HG volume and TTG shape (i.e., duplication patterns) and auditory and language abilities that are thought to be relatively stable and thus possibly arising from predisposition, such as speech sound processing and learning abilities (Golestani et al., 2011, 2007), linguistic tone learning (Wong et al., 2008), and general language aptitude (Turker et al., 2017). Studies have also shown larger HG and more TTG duplications in professional musicians (Schneider et al., 2005), and longitudinal work in children undergoing musical training shows that the grey matter volume of this region is relative stable over the course of a year (Seither-Preisler et al., 2014). To date, however, little evidence has been shown supporting experience-dependent plasticity of TTG. Ressel et al. (2012) showed that lifelong, non-elective bilingual language experience with Spanish and Catalan was associated with a larger bilateral HG volume in comparison to monolingual experience. Importantly, the bilinguals in their study were not self-selected, so unlikely to have learned a second language because of a special talent for languages. The findings of the Ressel et al. study thus suggested that the HG volume differences arose from experience with a different phonological system, since Spanish and Catalan are quite similar at the lexical level but have different phonologies. This finding was partially replicated in our study, in a sub-sample of participants with a single gyrus in the right hemisphere. More generally, our results show that experience with multiple languages is associated with differences in the thickness of the TTG, and that increasing degrees of multilingualism are linearly related to the structure of the transverse temporal gyrus. Establishing that non-pathological environmental experiences are related to the morphology of TTG offers important insights into the adaptability of this early sensory region, and has important implications for our understanding of the neural underpinnings of auditory processing in general, which we outline below.

We find an effect of multilingualism on the cortical thickness of the second TTG in individuals with a second such gyrus, but on the thickness of the first gyrus in individuals without a full TTG duplication. The result aligns with recent findings of cortical surface area being relatively more associated with genetic factors and of cortical thickness being relatively more associated with environmental factors (e.g., Eyler et al., 2012; Grasby et al., 2020). Individual differences in the shape (i.e., duplication patterns) of the TTG seem to, however, be related to *how* multilingualism will be accommodated by the thickness of the auditory brain regions: increasing thickness of the first gyrus when no duplications are present, and decreasing thickness of the second gyrus for participants with more than one TTG.

The fact that the most multilingual participants in the current sample had thinner cortex in their second TTG might not automatically point to the conclusion that multilingual language *experience* is associated with this specific neural marker. In fact, polyglotism and multilingualism could partly result from innate predispositions (i.e., language aptitude), which in turn have been associated with brain structural markers within the TTG (Turker et al., 2021, 2017). However, we observed that accounting for typology (i.e., including information regarding how languages spoken by the participants differed from each other in terms of their phonological systems) explained more variance in the cortical thickness data of the 2nd TTG than an index of language experience accounting for the age of exposure to different languages alone. The contribution of typological distance to our results is important, since it strengthens our interpretation that the observed results are more likely to be due to experience-induced plasticity and not to predispositions: it

is far less likely that specific brain anatomical features would induce individuals to learn specific language combinations.

Importantly, we show that, within the auditory cortex, language experience is particularly related to gyri in the superior temporal plane, and not to the rest of the planum temporale (as shown in the analysis investigating the sub-sample of participants with a single gyrus in the right hemisphere), nor to other auditory regions such as posterior STG, as shown in the first, exploratory analysis across broader auditory cortex sub-regions. The posterior STG, has been previously shown to be an essential site for language and phonological processing (Bhaya-Grossman and Chang, 2022; Hamilton et al., 2021; Hillis et al., 2017), and for representing acoustic-phonetic features of speech sounds (Lakertz et al., 2021; Mesgarani et al., 2014). We had predicted that we would find effects of differences between our multilingual participants' languages at the acoustic and articulatory feature level within posterior STG, however indices related to the degree of multilingualism in our sample were not related to its structure. Tuning to the different acoustic-phonetic features might be an experience-independent feature of the posterior STG (i.e., it might be sensitive to features both present and absent from the languages a person knows), and therefore multilingualism might not induce any neuroplastic changes to this region. Alternatively, our findings may suggest that current language models might underestimate the role of early auditory regions (i.e., of HG and 2nd TTG, when present) in phonological processing, and support their role beyond an initial spectro-temporal analysis of speech sounds. Future research, using both structural and functional mapping, is needed to clarify the respective roles of TTG(s) and STG in phonological processing.

With regards to the specific typological features that did best explain our results, we found that accounting for overlaps between languages in terms of their phoneme-level sound structure (i.e., discrete sound categories present in the individual languages' phonological inventories) was the best predictor of the thickness of the second TTG, and that a measure of how many unique phonemes each participant was exposed to across all their languages explained variance in the thickness of the left second TTG *above and beyond* language experience (in the larger main sample). Functional neuroimaging work by Fisher and colleagues (2018) has shown that speech sounds, and specifically vowel formants, are encoded in tonotopic auditory regions including HG and the STG, suggesting that even early auditory regions such as HG and second TTG (when present) are involved in speech sound processing. Similarly, Bonte et al. (2017) found that vowels can be decoded from fMRI signal of (among others) the temporal plane, Heschl's gyrus and sulcus (see also Jäncke et al., 2002; Obleser and Eisner, 2009; van Atteveldt et al., 2004 for findings of processing of isolated phonemes in the HG and PT regions), and Rutten et al. (2019) showed that attending to speech sounds (stop consonants) in pseudowords results in an increase in the neural processing of higher temporal modulations in regions as early as HG and also in the PT. It is therefore plausible that our finding of an association between greater multilingual language experience and decreased thickness of the second TTG arises from increased processing demands (and associated cortical pruning, see below) inherent to the acquisition and mastery of many non-overlapping phonological inventories. Moreover, to the best of our knowledge, previous studies did not account for fine variation in the exact individual anatomy of the TTG (i.e., single *versus* multiplicated gyri). Our results may thus call for a re-

evaluation of the specific anatomical site within the superior temporal plane in which phonemes are preferentially processed: based on our results, the second TTG (when present) could be a possible candidate.

Apart from phoneme-based typological distance measure explaining more variance in cortical thickness of the 2nd TTG in the main sample, the data from the replication sample showed that phonological feature-level language distance measure explained additional variance (to the composite language experience measure) in the thickness values. We believe that this inconsistency across samples might be related to the smaller sample size in the replication analysis and the particular phonological characteristics of participants' languages. Specifically, the replication sample included 31 individuals (i.e., 46% of participants) who spoke German and English, a language combination that has a phonological feature distance of zero. In contrast, the main sample included a more heterogenous mix of 34 different languages. Notably, however, in the replication sample, the phonemes-based typological distance measure numerically explained more variance in the thickness of the 2nd TTG than a composite language experience measure based on AoA of different languages, as observed in the main sample.

The decreased thickness of the second TTG in relation to greater multilingualism may arise from different, non-exclusive microstructural and physiological mechanisms, including functional remapping, experience-driven pruning and neural efficiency, or learning-related increased myelination. Primary auditory cortex is known to lie within the anterior-medial part of HG (i.e., the first TTG; Rademacher et al., 1993), although there is no exact correspondence between cytoarchitectural and macrostructural boundaries (Rademacher et al., 2001). *In-vivo* markers for human primary auditory cortex have been identified, including tonotopy, increased myelination (Dick et al., 2017), and decreased cortical thickness (Zoellner et al., 2019). Also, there is evidence that in professional musicians, functional activation extends to posterior TTGs during listening to musical sounds, while activation is confined to HG in non-musicians (Schneider et al., unpublished findings). Our findings of thinning in relation to extensive multilingual language experience with a wider diversity of speech sounds could be speculatively interpretated as arising from spatial expansion of primary-like functionality to the second TTG, when present. This spreading of primary-like processing may subsequently induce specialization and pruning of the more posterior gyrus. Alternatively, it could be that the secondary auditory regions beyond HG may be specialised for the processing of (cross-linguistic) phonological information, and this might result in experience-driven pruning. Why this occurs in the second TTG and not in non-gyral portions of the superior temporal plane (i.e., in the PT more generally) remains to be explored in future work, but one possibility is that gyri are more likely to host such a specialised function due to their greater neuronal density or to specific micro-anatomical features and connectivity properties (i.e., greater proximity and thus more direct, local functional and structural connectivity of neurons within the gyrus, as opposed to with neurons within the sulcal parts). Either way, the neural efficiency account would be in line with the Dynamic Restructuring Model of bilingualism (Pliatsikas, 2020), according to which the "peak efficiency" stage of multilingual functioning comes hand in hand with reductions in local volume. Of note, our supplementary analysis (see Section 7.6) investigating second TTG thickness in relation to proficient *versus* non-proficient languages points to the tentative conclusion that the observed reduced thickness results might be driven by experience with proficiently spoken languages (more so than with non-proficiently spoken ones), aligning with the predictions of the Dynamic Restructuring Model. Another potential microstructural mechanism underlying

the thinner cortex result could be increased myelination of 2nd TTG; this would again be in line with the idea that multilingual language experience leads to modifications such that the of 2nd TTG becomes more 'primary-like', since HG is characterized by higher myelination. It has been suggested that the apparent thinning of the ventral temporal cortex during childhood (as argued by e.g., Sowell et al., 2004) is actually driven by increased myelination, since increased myelination in deep cortical layers can shift the apparent gray–white boundary in MR images (Natu et al., 2019). Given the cross-sectional design of the current study, the above explanations remain hypotheses to be tested in further longitudinal studies of language acquisition, preferably ones using myelin mapping (Lutti et al., 2014; Marques et al., 2010) to explore the underlying physiological mechanisms in more details.

Interestingly, in individuals who have only one TTG (i.e., no duplications), we find that multilingualism is related to thickness but in the opposite direction. We speculate that there might be different mechanisms underlying plasticity in regions more likely to contain primary *versus* secondary auditory cortex. In other words, it may be more likely that experience-induced pruning occurs in secondary compared to primary regions, and that in primary regions, greater efficiency comes at the cost of a neuronal 'bulking'. Alternately but non-exclusively, it could be that in people who do not have a second TTG, HG needs to accommodate both lower-level auditory processing and also the processing of more complex speech sounds. Further research, especially with larger samples of participants with a single TTG, is needed to elucidate the opposite direction of the relationship between multilingualism and cortical thickness in HG *versus* the second TTG in our results.

Different methodological tools are available for capturing the variability of TTG. Most previous studies used either manual labelling, whole-brain approaches involving normalizing structural images to a common template space in order to assess regional variation in grey or white matter probability, or automated pipelines serving to segment and label different regions of interest (e.g. VBM or FreeSurfer). Both approaches have important advantages, the former allowing for high precision in capturing details of individual variation, the latter two being far less time-consuming and labor-intensive, and therefore more applicable to large datasets. In the current study, we capitalized on recent developments in cortical segmentation efforts and used an automatic toolbox (TASH; Dalboni da Rocha et al., 2020) specifically designed for fine delineation of auditory cortex gyri, and extraction measures describing their anatomy. Coupled with visual inspection of the data, TASH provides a "best-of-both-worlds" approach, by being anatomically precise while involving little manual involvement with the data (which also tends to be error-prone). The measures we obtained align with the high degree of variability reported in previous literature, but they also seem to be more sensitive, since we identify more multiplications of the TTG than previously reported. For example, in Marie et al. (2015), out of 232 subjects, 204 (88%) had one gyrus (including partially duplicated) in the left hemisphere, 28 (12%) had a complete posterior duplication; the prevalence of further multiplications was not assessed in that study. In our data, a minority of participants (4%) presented only one gyrus in the left hemisphere, and most of them had either two (57%) or three (35%) separate gyri. TASH identifies gyri based on anatomical landmarks and curvature values, and therefore might be more sensitive to shallow cortical folds that might otherwise not be discernable in manual labelling. The precise functional relevance of these multiplications remains to be uncovered in future research.

Notably, such variability in shape of the auditory regions is absent from other, evolutionarily older, species: macaques' auditory cortex is flat, and some chimpanzees have only a single gyrus (Hackett et al., 2001). By showing that a human-specific experience, i.e., speaking multiple languages, is reflected in the anatomy of cortical regions that seem to be evolutionarily particular to humans, we contribute to informing both models of neuroplasticity and of language evolution.

A large body of work has provided insights into both how the brain processes auditory information and speech signals (see Bhaya-Grossman and Chang, 2022; Moerel et al., 2014 for overviews), and into how bi- and multilingualism are related to the structure of the brain (see García-Pentón et al., 2016; Li et al., 2014; Pliatsikas, 2020 for overviews). In the current study, we brought together both strands of research in an effort to investigate how the auditory cortex accommodates multilingual experience, thereby providing insights into the intricate functioning of auditory processing in humans. Apart from showing that individual anatomy appears to constrain the architecture of phonological representations, our findings also support the idea that experience with typologically similar languages might be different from experience with typologically distant languages (Antoniou, 2018; Berthele, 2020; Li et al., 2014). Indeed, across two independent datasets, we show, for the first time, that cortical thickness of early auditory brain regions is related to the degree of one's language experience coupled with typological distance between the languages one knows. These findings indicate that early auditory regions seem to represent (or be shaped by) phoneme-level cross-linguistic information, contrary to the most established models of language processing in the brain, which suggest that phonological processing happens in more lateral posterior STG and STS.

4 Materials and Methods

4.1 Participants

Main sample. The full sample consisted of N = 146 participants (M_{age} = 35.79, SD = 15.77; 85 females; 31 monolinguals) exposed to up to 7 languages (2.65 languages on average), with n = 136 complete cases. **Replication sample.** The replication sample consisted of N = 69 participants (M_{age} = 32.04, SD = 11.68; 38 females; 29 monolinguals) exposed to up to 5 languages (2.44 languages on average), with n = 68 complete cases.

4.2 Language experience

To describe the diverse language background of our participants, we expressed age of onset(s) of acquisition (AoA) of different languages in a continuous quantitative measure. First, AoA of individual languages was log-transformed to minimize differences between values for languages learned later in life, and the values were inverted to express early AoAs as the highest values; a constant value of 1 was added to the index before log-transformation and after inverting the values, each time to avoid values equal to zero. Next, the AoAs of participants' different languages were combined into one "language experience" index per participant. It was computed using Shannon's entropy equation (Shannon, 1948):

$$H = -\sum_{i=0}^{n} p_i \, log_2 \, i$$

where, n stands for the total number of languages a participant has been exposed to and p_i is the AoA index calculated as above. High entropy values indicated more diverse language experience.

4.3 Typological distance measures

Across the two samples, the participants spoke a total of N = 36 different languages (n = 34 in the main sample, and n = 20 in the replication sample): Arabic, Bengali, Cantonese, Creole (Antiguan), Czech, Danish, Dutch, English, Farsi, Finnish, French, German, Greek, Gujarati, Hakka, Hebrew, Hindi, Hokkien, Hungarian, Irish, Italian, Japanese, Korean, Luxembourgish, Malay, Mandarin, Portuguese, Punjabi, Russian, Sindhi, Spanish, Swahili, Swedish, Swiss German, Urdu and Vietnamese. The PHOIBLE database (Moran et al., 2019) and open-source software (Dediu and Moisik, 2016) were used to construct three measures of typological distance between the languages: distances between the distinctive phonological features describing the phonemes of each phonological inventory, distances between the sets of phonemes of the phonological inventories, and similarity in counts of phonological classes that share certain features, as detailed below. Figure S 2 provides a visual representation of the three distance matrices between all the languages.

4.3.1 Acoustic and articulatory features

For each documented phonological inventory, PHOIBLE includes their phonemes (see Section 4.3.2 below) and the lower-level distinctive acoustic and articulatory features describing the phonemes. These features offer a qualitative description of differences between discrete sound categories, and are based on articulatory phonology (Hayes, 2011). Every phoneme in every language is described by the following 37 binary features: tone, stress, syllabic, short, long, consonantal, sonorant, continuant, delayed release, approximant, tap, trill, nasal, lateral, labial, round, labiodental, coronal, anterior, distributed, strident, dorsal, high, low, front, back, tense, retracted tongue root, advanced tongue root, periodic glottal source, epilaryngeal source, spread glottis, constricted glottis, fortis, raised larynx ejective, lowered larynx implosive, click. Individual languages were described in terms of presence or absence of these features across all phonemes belonging to the given phonological inventory, and represented as vectors. For a given pair of language feature vectors, a pairwise distance was computed as cosine distance using the `scipy.spatial.distance` function in Python.

4.3.2 Phonemes

Phonemes are individual speech sounds such as vowels and consonants, representing the smallest units of sound that can change the meaning of a word when substituted. Lists of phonemes belonging to each of the languages represented in the data-base were exported from PHOIBLE (Moran et al., 2019) and represented as strings. For a given pair of languages A and B, a pairwise string distance between phonemes of both languages was computed with the Jaccard distance method:

$$1 - J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

representing the inverse of the ratio between the size of the intersection and the size of the union of the two sets of phonemes, and varying between 0 (both languages having exactly the same phonemes) and 1 (no phonemes in common).

In addition, using the language-specific phoneme-level information, we constructed a cross-linguistic description of individual languages and used it to compute a measure of cumulative phoneme inventory for each participant. We counted how many unique phonemes each participant was exposed to across all their languages (i.e., phonemes that overlapped between languages were counted only once). For example, the cumulative phoneme inventory of a participant who was exposed only to English was equal to 40, while an English-French bilingual's cumulative phoneme inventory would be equal to 64 (both English and French have 40 unique phonemes (Stanford Phonology Archive, 2019b, 2019a), with 16 phonemes overlapping across the two inventories).

4.3.3 Counts of phonological classes

Both the lists of phonemes and phonological features describing them can be used to derive a more abstract, higher-level description of individual languages in terms of which and how many *phonological classes* their phonological inventories contain. Using the PHOIBLE inventories and feature system, Dediu and Moisik (2016) proposed a comprehensive list of 167 phonological classes, along with a method for selecting a language's phonemes belonging to given classes and computing their counts. The system describes how many "segments", "consonants", "vowels", "diphthongs" etc. a given language has, and counts further classes that share certain features (e.g., "bilabial consonants", "front vowels", "clicks", see Supplementary Information, Section 7.8 for the full list). Per language, we created a vector representing the counts of all 167 phonological classes, and computed pairwise distances between the languages spoken by the participants in our database with cosine distance (as above).

4.4 Combining typology and language experience (indexed by AoA)

Quantifying diverse language experience with Shannon's entropy equation does not account for similarities and differences between the languages. Therefore, using Rao's quadratic entropy equation (*QE*) (De Bello et al., 2007; Pavoine and Bonsall, 2011; Rao, 1982), the language experience index was combined with each of the above typological distance measures, separately, to generate weighted measures accounting for the three different types of cross-linguistic phonological distances. The summed phonological distances (feature-level, phoneme-level or based on counts of phonological classes) between all language pairs were weighted by the log-transformed and inverted AoA index for each language, as follows:

If the AoA index for *i*-th language in a participant's repertoire is p_i and the dissimilarity between language *i* and *j* is d_{ij}, then language experience index accounting for typological information has the form:

$$QE = \sum_{i=1}^{S} \sum_{j=1}^{S} d_{ij}\, p_i\, p_j$$

where d_{ij} varies from 0 (i.e., where the two languages have exactly the same phonological systems) to 1 (i.e., the two languages have completely different phonological systems). The `SYNCSA` R package (Debastiani and Pillar, 2012) (https://rdrr.io/cran/SYNCSA/) was used to perform calculations.

For a similar methodology, but applied to lexical distance measures, see Kepinska et al. (2023).

4.5 Neuroimaging data acquisition and processing

For all participants, structural MRIs with 176 sagittal slices were acquired using a T1-weighted three-dimensional (3D) Modified Driven Equilibrium Fourier Transform (MDEFT) sequence. The resulting images had a matrix size of 256 × 224, yielding a final resolution (voxel size) of 1 × 1 × 1 mm. Repetition time (TR)/echo time (TE)/inversion time (TI) was: [12.24/3.56/530 ms] for the main sample (from a 1.5T Siemens Sonata scanner), and 7.92/ 2.48/910 ms for the replication sample (from a 3T Siemens Trio scanner). The T1 images were denoised using the spatially adaptive non-local means filter (SANLM; Manjón et al., 2010) in MATLAB within the CAT12 toolbox. The images were then processed with FreeSurfer's (version 7.2) brain structural pipeline (Fischl et al., 2004), which consists of motion correction, intensity normalization, skull stripping, and reconstruction of the volume's voxels into white and pial surfaces.

4.6 Parcellation of the auditory regions

The reconstructed surfaces were parcellated into regions using an atlas-based procedure (Destrieux et al., 2010). The auditory regions were delineated using the following labels from the Destrieux atlas: planum polare of the superior temporal gyrus (`G_temp_sup-Plan_polar`; PP), anterior transverse temporal gyrus (of Heschl) (`G_temp_sup-G_T_transv`; referring to HG, or to the anterior-most TTG), transverse temporal sulcus (`S_temporal_transverse`; HS), planum temporale, or temporal plane of the superior temporal gyrus (`G_temp_sup-Plan_tempo`; PT), superior temporal sulcus (parallel sulcus) (`S_temporal_sup`; STS), and lateral aspect of the superior temporal gyrus (`G_temp_sup-Lateral`; STG). The STG and the STS labels were further divided into smaller regions using FreeSurfer's freeview software and FreeSurfer command line tools. For this, on the inflated surface of the fsaverage template, we first drew additional ROIs, dividing the STG into anterior and posterior parts, with the transverse temporal sulcus as the dividing landmark (Hamilton et al., 2021). We further divided the STS into anterior, middle and posterior parts. The anterior versus mid-STS were divided by again aligning the boundary with the posterior border of the anterior TTG, and the mid- versus posterior STS were divided by aligning the boundary with the posterior border of STG. Figure S 3 shows an example of the parcellation in the native space of one of the participants. Next, we transformed the new ROIs into the participants native space (using `mri_surf2surf`), and intersected them with individual subjects' unaltered STG and STS labels. Cortical volume, surface area and average thickness were computed and extracted from the final labels (with `mris_anatomical_stats`), and were used in the linear mixed models reported in Section 2.1.

For analysis on the TTG along the superior temporal plane (i.e., on the superior aspect of the STG) (see Results, Section 2.2), we used an automated toolbox (TASH; Dalboni da Rocha et al., 2020). TASH runs on the output of the FreeSurfer Destrieux atlas structural segmentation, and provides a finer segmentation not only of HG – whether a

single gyrus or a common stem duplication (default version of TASH) – but also of additional TTGs when present (extended version of TASH, called 'TASH_complete'; Dalboni da Rocha et al., 2023). Volume, surface area and average thickness of the resulting labels were extracted. To explore the significant interaction between average thickness of PT and language experience in the exploratory analysis on all auditory regions (Section 2.1), we used 'TASH_complete'; this allowed us to explore possible relationships with additional TTGs, when present. We performed a visual selection of the gyri segmented by TASH_complete, excluding from the analysis gyri that lay along the portion of the superior temporal plane that curved vertically (i.e., within the parietal extension, Honeycutt et al., 2000), when present. The volume, surface area and average thickness of the resulting labels were extracted, and used as dependent variables in the statistical analysis (see Results, see Section 2.2).

5 Acknowledgments

This work was supported by the Wellcome Trust (grant numbers 203147/Z/16/Z and 205103/Z/16/Z) to CJP. The authors gratefully acknowledge support by the NCCR Evolving Language, Swiss National Science Foundation (SNSF) Agreement #51NF40_180888, and by the SNSF grant #100014_18238. OK was funded by the Marie Jahoda Stipendium from University of Vienna.

6 References

Antoniou, M., 2018. The advantages of bilingualism debate. Annu. Rev. Linguist.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using {lme4}. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Berthele, R., 2020. The Extraordinary Ordinary: Re-engineering Multilingualism as a Natural Category. Lang. Learn. 1–41. https://doi.org/10.1111/lang.12407

Bhaya-Grossman, I., Chang, E.F., 2022. Speech Computations of the Human Superior Temporal Gyrus. Annu. Rev. Psychol. 2022 73, 2021. https://doi.org/10.1146/annurev-psych-022321

Binder, J., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528. https://doi.org/10.1093/cercor/10.5.512

Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2002. Modality independence of word comprehension. Hum. Brain Mapp. 16, 251–261. https://doi.org/10.1002/hbm.10054

Cai, D.C., Fonteijn, H., Guadalupe, T., Zwiers, M., Wittfeld, K., Teumer, A., Hoogman, M., Arias-Vásquez, A., Yang, Y., Buitelaar, J., Fernández, G., Brunner, H.G., van Bokhoven, H., Franke, B., Hegenscheid, K., Homuth, G., Fisher, S.E., Grabe, H.J., Francks, C., Hagoort, P., 2014. A genome-wide search for quantitative trait loci affecting the cortical surface area and thickness of Heschl's gyrus. Genes, Brain Behav. 13, 675–685. https://doi.org/10.1111/gbb.12157

Chi, J.G., Dooling, E.C., Gilles, F.H., 1977. Gyral development of the human brain. Ann. Neurol. 1, 86–93. https://doi.org/https://doi.org/10.1002/ana.410010109

da Costa, S., van der Zwaag, W., Marques, J.P., Frackowiak, R.S.J., Clarke, S., Saenz, M., 2011. Human primary auditory cortex follows the shape of Heschl's Gyrus. J. Neurosci. 31, 14067–14075. https://doi.org/10.1523/JNEUROSCI.2000-11.2011

Dalboni da Rocha, J.L., Kepinska, O., Schneider, P., Benner, J., Degano, G., Schneider, L., Golestani, N., 2023. Multivariate Concavity Amplitude Index (MCAI) for characterizing Heschl's gyrus shape. Neuroimage 272, 120052. https://doi.org/10.1016/j.neuroimage.2023.120052

Dalboni da Rocha, J.L., Schneider, P., Benner, J., Santoro, R., Atanasova, T., Van De Ville, D., Golestani, N., 2020. TASH: Toolbox for the Automated Segmentation of Heschl's gyrus. Sci. Rep. 10, 3887. https://doi.org/10.1038/s41598-020-60609-y

De Bello, F., Lepš, J., Lavorel, S., Moretti, M., 2007. Importance of species abundance for assessment of trait composition: An example based on pollinator communities. Community Ecol. 8, 163–170. https://doi.org/10.1556/ComEc.8.2007.2.3

Debastiani, V., Pillar, V., 2012. S{YNCSA} - {R} tool for analysis of metacommunities based on functional traits and phylogeny of the community components. Bioinformatics 28, 2067–2068.

Dediu, D., Moisik, S.R., 2016. Defining and counting phonological classes in cross-linguistic segment databases. Proc. 10th Int. Conf. Lang. Resour. Eval. Lr. 2016 1955–1962.

DeLuca, V., Rothman, J., Bialystok, E., Pliatsikas, C., 2019. Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. Proc. Natl. Acad. Sci. 116, 7565–7574. https://doi.org/10.1073/pnas.1811513116

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Dick, F., Tierney, A.T., Lutti, A., Josephs, O., Sereno, M.I., Weiskopf, N., 2012. In vivo functional and myeloarchitectonic mapping of human primary auditory areas. J. Neurosci. 32, 16095–16105. https://doi.org/10.1523/JNEUROSCI.1712-12.2012

Dick, F.K., Lehet, M.I., Callaghan, M.F., Keller, T.A., Sereno, M.I., Holt, L.L., 2017. Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally directed attention and systematically related to cortical myeloarchitecture. J. Neurosci. 37, 12187–12201. https://doi.org/10.1523/JNEUROSCI.1436-17.2017

Eyler, L.T., Chen, C.H., Panizzon, M.S., Fennema-Notestine, C., Neale, M.C., Jak, A., Jernigan, T.L., Fischl, B., Franz, C.E., Lyons, M.J., Grant, M., Prom-Wormley, E., Seidman, L.J., Tsuang, M.T., Fiecas, M.J.A., Dale, A.M., Kremen, W.S., 2012. A comparison of heritability maps of cortical surface area and thickness and the influence of adjustment for whole brain measures: A magnetic resonance imaging twin study. Twin Res. Hum. Genet. 15,

304–314. https://doi.org/10.1017/thg.2012.3

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically Parcellating the Human Cerebral Cortex. Cereb. Cortex 14, 11–22. https://doi.org/10.1093/cercor/bhg087

Fisher, J.M., Dick, F.K., Levy, D.F., Wilson, S.M., 2018. Neural representation of vowel formants in tonotopic auditory cortex. Neuroimage 178, 574–582. https://doi.org/10.1016/J.NEUROIMAGE.2018.05.072

García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J.A., Carreiras, M., 2016. The neuroanatomy of bilingualism: how to turn a hazy view into the full picture. Lang. Cogn. Neurosci. 31, 303–327. https://doi.org/10.1080/23273798.2015.1068944

Geschwind, N., Levitsky, W., 1968. Human brain: left-right asymmetries in temporal speech region. Science 161, 186–187. https://doi.org/10.1126/SCIENCE.161.3837.186

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., Pallier, C., 2007. Brain structure predicts the learning of foreign speech sounds. Cereb. Cortex 17, 575–582.

Golestani, N., Price, C.J., Scott, S.K., 2011. Born with an ear for dialects? Structural plasticity in the expert phonetician brain. J. Neurosci. 31, 4213–4220. https://doi.org/10.1523/JNEUROSCI.3891-10.2011

Grasby, K.L., Jahanshad, N., Painter, J.N., Colodro-Conde, L., Bralten, J., Hibar, D.P., Lind, P.A., Pizzagalli, F., Ching, C.R.K., McMahon, M.A.B., Shatokhina, N., Zsembik, L.C.P., Thomopoulos, S.I., Zhu, A.H., Strike, L.T., Agartz, I., Alhusaini, S., Almeida, M.A.A., Alnæs, D., Amlien, I.K., Andersson, M., Ard, T., Armstrong, N.J., Ashley-Koch, A., Atkins, J.R., Bernard, M., Brouwer, R.M., Buimer, E.E.L., Bülow, R., Bürger, C., Cannon, D.M., Chakravarty, M., Chen, Q., Cheung, J.W., Couvy-Duchesne, B., Dale, A.M., Dalvie, S., de Araujo, T.K., de Zubicaray, G.I., de Zwarte, S.M.C., den Braber, A., Doan, N.T., Dohm, K., Ehrlich, S., Engelbrecht, H.-R., Erk, S., Fan, C.C., Fedko, I.O., Foley, S.F., Ford, J.M., Fukunaga, M., Garrett, M.E., Ge, T., Giddaluru, S., Goldman, A.L., Green, M.J., Groenewold, N.A., Grotegerd, D., Gurholt, T.P., Gutman, B.A., Hansell, N.K., Harris, M.A., Harrison, M.B., Haswell, C.C., Hauser, M., Herms, S., Heslenfeld, D.J., Ho, N.F., Hoehn, D., Hoffmann, P., Holleran, L., Hoogman, M., Hottenga, J.-J., Ikeda, M., Janowitz, D., Jansen, I.E., Jia, T., Jockwitz, C., Kanai, R., Karama, S., Kasperaviciute, D., Kaufmann, T., Kelly, S., Kikuchi, M., Klein, M., Knapp, M., Knodt, A.R., Krämer, B., Lam, M., Lancaster, T.M., Lee, P.H., Lett, T.A., Lewis, L.B., Lopes-Cendes, I., Luciano, M., Macciardi, F., Marquand, A.F., Mathias, S.R., Melzer, T.R., Milaneschi, Y., Mirza-Schreiber, N., Moreira, J.C. V., Mühleisen, T.W., Müller-Myhsok, B., Najt, P., Nakahara, S., Nho, K., Olde Loohuis, L.M., Orfanos, D.P., Pearson, J.F., Pitcher, T.L., Pütz, B., Quidé, Y., Ragothaman, A., Rashid, F.M., Reay, W.R., Redlich, R., Reinbold, C.S., Repple, J., Richard, G., Riedel, B.C., Risacher, S.L., Rocha, C.S., Mota, N.R., Salminen, L., Saremi, A., Saykin, A.J., Schlag, F., Schmaal, L., Schofield, P.R., Secolin, R., Shapland, C.Y., Shen, L., Shin, J., Shumskaya, E., Sønderby, I.E., Sprooten, E., Tansey, K.E., Teumer, A., Thalamuthu, A., Tordesillas-Gutiérrez, D., Turner, J.A., Uhlmann, A., Vallerga, C.L., van der Meer, D., van Donkelaar, M.M.J., van Eijk, L.,

van Erp, T.G.M., van Haren, N.E.M., van Rooij, D., van Tol, M.-J., Veldink, J.H., Verhoef, E., Walton, E., Wang, M., Wang, Y., Wardlaw, J.M., Wen, W., Westlye, L.T., Whelan, C.D., Witt, S.H., Wittfeld, K., Wolf, C., Wolfers, T., Wu, J.Q., Yasuda, C.L., Zaremba, D., Zhang, Z., Zwiers, M.P., Artiges, E., Assareh, A.A., Ayesa-Arriola, R., Belger, A., Brandt, C.L., Brown, G.G., Cichon, S., Curran, J.E., Davies, G.E., Degenhardt, F., Dennis, M.F., Dietsche, B., Djurovic, S., Doherty, C.P., Espiritu, R., Garijo, D., Gil, Y., Gowland, P.A., Green, R.C., Häusler, A.N., Heindel, W., Ho, B.-C., Hoffmann, W.U., Holsboer, F., Homuth, G., Hosten, N., Jack, C.R., Jang, M., Jansen, A., Kimbrel, N.A., Kolskår, K., Koops, S., Krug, A., Lim, K.O., Luykx, J.J., Mathalon, D.H., Mather, K.A., Mattay, V.S., Matthews, S., Mayoral Van Son, J., McEwen, S.C., Melle, I., Morris, D.W., Mueller, B.A., Nauck, M., Nordvik, J.E., Nöthen, M.M., O'Leary, D.S., Opel, N., Martinot, M.-L.P., Pike, G.B., Preda, A., Quinlan, E.B., Rasser, P.E., Ratnakar, V., Reppermund, S., Steen, V.M., Tooney, P.A., Torres, F.R., Veltman, D.J., Voyvodic, J.T., Whelan, R., White, T., Yamamori, H., Adams, H.H.H., Bis, J.C., Debette, S., Decarli, C., Fornage, M., Gudnason, V., Hofer, E., Ikram, M.A., Launer, L., Longstreth, W.T., Lopez, O.L., Mazoyer, B., Mosley, T.H., Roshchupkin, G. V., Satizabal, C.L., Schmidt, R., Seshadri, S., Yang, Q., Alvim, M.K.M., Ames, D., Anderson, T.J., Andreassen, O.A., Arias-Vasquez, A., Bastin, M.E., Baune, B.T., Beckham, J.C., Blangero, J., Boomsma, D.I., Brodaty, H., Brunner, H.G., Buckner, R.L., Buitelaar, J.K., Bustillo, J.R., Cahn, W., Cairns, M.J., Calhoun, V., Carr, V.J., Caseras, X., Caspers, S., Cavalleri, G.L., Cendes, F., Corvin, A., Crespo-Facorro, B., Dalrymple-Alford, J.C., Dannlowski, U., de Geus, E.J.C., Deary, I.J., Delanty, N., Depondt, C., Desrivières, S., Donohoe, G., Espeseth, T., Fernández, G., Fisher, S.E., Flor, H., Forstner, A.J., Francks, C., Franke, B., Glahn, D.C., Gollub, R.L., Grabe, H.J., Gruber, O., Håberg, A.K., Hariri, A.R., Hartman, C.A., Hashimoto, R., Heinz, A., Henskens, F.A., Hillegers, M.H.J., Hoekstra, P.J., Holmes, A.J., Hong, L.E., Hopkins, W.D., Hulshoff Pol, H.E., Jernigan, T.L., Jönsson, E.G., Kahn, R.S., Kennedy, M.A., Kircher, T.T.J., Kochunov, P., Kwok, J.B.J., Le Hellard, S., Loughland, C.M., Martin, N.G., Martinot, J.-L., McDonald, C., McMahon, K.L., Meyer-Lindenberg, A., Michie, P.T., Morey, R.A., Mowry, B., Nyberg, L., Oosterlaan, J., Ophoff, R.A., Pantelis, C., Paus, T., Pausova, Z., Penninx, B.W.J.H., Polderman, T.J.C., Posthuma, D., Rietschel, M., Roffman, J.L., Rowland, L.M., Sachdev, P.S., Sämann, P.G., Schall, U., Schumann, G., Scott, R.J., Sim, K., Sisodiya, S.M., Smoller, J.W., Sommer, I.E., St Pourcain, B., Stein, D.J., Toga, A.W., Trollor, J.N., Van der Wee, N.J.A., van 't Ent, D., Völzke, H., Walter, H., Weber, B., Weinberger, D.R., Wright, M.J., Zhou, J., Stein, J.L., Thompson, P.M., Medland, S.E., 2020. The genetic architecture of the human cerebral cortex. Science (80-.). 367. https://doi.org/10.1126/science.aay6690

Hackett, T.A., Preuss, T.M., Kaas, J.H., 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. 441, 197–222. https://doi.org/10.1002/cne.1407

Hamilton, L.S., Oganian, Y., Hall, J., Chang, E.F., 2021. Parallel and distributed encoding of speech across human auditory cortex. Cell 184, 4626-4639.e13. https://doi.org/10.1016/j.cell.2021.07.019

Hayes, B., 2011. Introductory phonology. John Wiley & Sons.

Hervais-Adelman, A., Egorova, N., Golestani, N., 2018. Beyond bilingualism: multilingual experience correlates with

caudate volume. Brain Struct. Funct. 223, 3495–3502. https://doi.org/10.1007/s00429-018-1695-0

Hervais-Adelman, A., Moser-Mercer, B., Murray, M.M., Golestani, N., 2017. Cortical thickness increases after simultaneous interpretation training. Neuropsychologia 98, 212–219. https://doi.org/10.1016/j.neuropsychologia.2017.01.008

Hillis, A.E., Rorden, C., Fridriksson, J., 2017. Brain regions essential for word comprehension: Drawing inferences from patients. Ann. Neurol. 81, 759–768. https://doi.org/10.1002/ana.24941

Honeycutt, N.A., Musick, A., Barta, P.E., Pearlson, G.D., 2000. Measurement of the planum temporale (PT) on magnetic resonance imaging scans: Temporal PT alone and with parietal extension. Psychiatry Res. - Neuroimaging 98, 103–116. https://doi.org/10.1016/S0925-4927(00)00043-3

Hullett, P.W., Kandahari, N., Shih, T.T., Kleen, J.K., Knowlton, R.C., Rao, V.R., Chang, E.F., 2022. Intact speech perception after resection of dominant hemisphere primary auditory cortex for the treatment of medically refractory epilepsy: illustrative case. J. neurosurgery. Case lessons 4, 1–6. https://doi.org/10.3171/CASE22417

Humphries, C., Sabri, M., Lewis, K., Liebenthal, E., 2014. Hierarchical organization of speech perception in human auditory cortex. Front. Neurosci. 8, 1–12. https://doi.org/10.3389/fnins.2014.00406

Jäncke, L., Wüstenberg, T., Scheich, H., Heinze, H.J., 2002. Phonetic perception and the temporal cortex. Neuroimage 15, 733–746. https://doi.org/10.1006/nimg.2001.1027

Kepinska, O., Caballero, J., Oliver, M., Marks, R.A., Haft, S.L., Zekelman, L., Kovelman, I., Uchikoshi, Y., Hoeft, F., 2023. Language combinations of multilinguals are reflected in their first-language knowledge and processing. Sci. Rep. 13, 1947. https://doi.org/10.1038/s41598-023-27952-2

Kolde, R., 2019. pheatmap: Pretty Heatmaps.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. {lmerTest} Package: Tests in Linear Mixed Effects Models. J. Stat. Softw. 82, 1–26. https://doi.org/10.18637/jss.v082.i13

Lakertz, Y., Ossmy, O., Friedmann, N., Mukamel, R., Fried, I., 2021. Single-cell activity in human STG during perception of phonemes is organized according to manner of articulation. Neuroimage 226, 117499. https://doi.org/10.1016/j.neuroimage.2020.117499

Li, P., Legault, J., Litcofsky, K.A., 2014. Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. Cortex 1–24. https://doi.org/10.1016/j.cortex.2014.05.001

Luk, G., Bialystok, E., 2013. Bilingualism is not a categorical variable: Interaction between language proficiency and usage. J. Cogn. Psychol. 25, 605–621. https://doi.org/10.1080/20445911.2013.795574

Lutti, A., Dick, F., Sereno, M.I., Weiskopf, N., 2014. Using high-resolution quantitative mapping of R1 as an index of cortical myelination. Neuroimage 93, 176–188. https://doi.org/10.1016/j.neuroimage.2013.06.005

Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D.L., Robles, M., 2010. Adaptive non-local means denoising of

MR images with spatially varying noise levels. J. Magn. Reson. Imaging 31, 192–203. https://doi.org/10.1002/JMRI.22003

Marie, D., Jobard, G., Crivello, F., Perchey, G., Petit, L., Mellet, E., Joliot, M., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2015. Descriptive anatomy of Heschl's gyri in 430 healthy volunteers, including 198 left-handers. Brain Struct. Funct. 220, 729–743. https://doi.org/10.1007/s00429-013-0680-x

Marques, J.P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.F., Gruetter, R., 2010. MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. Neuroimage 49, 1271–1281. https://doi.org/10.1016/j.neuroimage.2009.10.002

Mårtensson, J., Eriksson, J., Bodammer, N.C., Lindgren, M., Johansson, M., Nyberg, L., Lövdén, M., 2012. Growth of language-related brain areas after foreign language learning. Neuroimage 63, 240–244. https://doi.org/10.1016/j.neuroimage.2012.06.043

Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science (80-.). 343, 1006–1010. https://doi.org/10.1126/science.1245994

Moerel, M., De Martino, F., Formisano, E., 2014. An anatomical and functional topography of human auditory cortical areas. Front. Neurosci. 0, 225. https://doi.org/10.3389/FNINS.2014.00225/BIBTEX

Moran, S., McCloy, D., Wright, R. (Eds.), 2019. PHOIBLE Online. Leipzig.

Natu, V.S., Gomez, J., Barnett, M., Jeska, B., Kirilina, E., Jaeger, C., Zhen, Z., Cox, S., Weiner, K.S., Weiskopf, N., Grill-Spector, K., 2019. Apparent thinning of human visual cortex during childhood is associated with myelination. Proc. Natl. Acad. Sci. U. S. A. 116, 20750–20759. https://doi.org/10.1073/pnas.1904931116

Obleser, J., Eisner, F., 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn. Sci. 13, 14–19. https://doi.org/10.1016/j.tics.2008.09.005

Pavoine, S., Bonsall, M.B., 2011. Measuring biodiversity to explain community assembly: A unified approach. Biol. Rev. 86, 792–812. https://doi.org/10.1111/j.1469-185X.2010.00171.x

Pliatsikas, C., 2020. Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. Bilingualism 23, 459–471. https://doi.org/10.1017/S1366728919000130

Rademacher, J., Caviness Jr, V.S., Steinmetz, H., Galaburda, A.M., 1993. Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cereb. Cortex 3, 313–329.

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H.J., Zilles, K., 2001. Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13, 669–683. https://doi.org/10.1006/nimg.2000.0714

Rao, R., 1982. Diversity and Dissimilarity Coefficients: A Unified Approach. Theor. Popul. Biol. 21, 24–43. https://doi.org/10.13140/RG.2.1.3901.9924

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., Sebastián-Gallés, N., 2012. An effect of Bilingualism on the auditory cortex. J. Neurosci. 32, 16597–16601. https://doi.org/10.1523/JNEUROSCI.1996-12.2012

Rutten, S., Santoro, R., Hervais-Adelman, A., Formisano, E., Golestani, N., 2019. Cortical encoding of speech enhances task-relevant acoustic information. Nat. Hum. Behav. 3, 974–987. https://doi.org/10.1038/s41562-019-0648-9

Santoro, R., Moerel, M., De Martino, F., Valente, G., Ugurbil, K., Yacoub, E., Formisano, E., 2017. Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. Proc. Natl. Acad. Sci. U. S. A. 114, 4799–4804. https://doi.org/10.1073/pnas.1617622114

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H.J., Dosch, H.G., Bleeck, S., Stippich, C., Rupp, A., 2005. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. Nat. Neurosci. 8, 1241–1247. https://doi.org/10.1038/nn1530

Schwarz, G., 1978. Estimating the Dimension of a Model. Ann. Stat. 6, 461–464. https://doi.org/10.2307/2958889

Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107. https://doi.org/Pii s0166-2236(02)00037-1 10.1016/s0166-2236(02)00037-1

Seghier, M.L., Patel, E., Prejawa, S., Ramsden, S., Selmer, A., Lim, L., Browne, R., Rae, J., Haigh, Z., Ezekiel, D., Hope, T.M., Leff, A.P., Price, C.J., 2016. The PLORAS Database: A data repository for Predicting Language Outcome and Recovery After Stroke. Neuroimage 124, 1208–1212. https://doi.org/10.1016/j.neuroimage.2015.03.083

Seither-Preisler, A., Parncutt, R., Schneider, P., 2014. Size and synchronization of auditory cortex promotes musical, literacy, and Attentional skills in children. J. Neurosci. 34, 10937–10949. https://doi.org/10.1523/JNEUROSCI.5315-13.2014

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. https://doi.org/10.1145/584091.584093

Sowell, E.R., Thompson, P.M., Leonard, C.M., Welcome, S.E., Kan, E., Toga, A.W., 2004. Longitudinal Mapping of Cortical Thickness and Brain Growth in Normal Children. J. Neurosci. 24, 8223–8231. https://doi.org/10.1523/JNEUROSCI.1798-04.2004

Stanford Phonology Archive, 2019a. English sound inventory (SPA), in: Moran, S., McCloy, D. (Eds.), PHOIBLE 2.0. Max Planck Institute for the Science of Human History, Jena.

Stanford Phonology Archive, 2019b. French sound inventory (SPA), in: Moran, S., McCloy, D. (Eds.), PHOIBLE 2.0. Max Planck Institute for the Science of Human History, Jena.

Turker, S., Reiterer, S.M., Seither-Preisler, A., Schneider, P., 2017. "When Music Speaks": Auditory Cortex Morphology as a Neuroanatomical Marker of Language Aptitude and Musicality. Front. Psychol. 8, 2096.

https://doi.org/10.3389/fpsyg.2017.02096

Turker, S., Seither-Preisler, A., Reiterer, S.M., 2021. Examining Individual Differences in Language Learning: A Neurocognitive Model of Language Aptitude. Neurobiol. Lang. 1–27. https://doi.org/10.1162/nol_a_00042

van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of Letters and Speech Sounds in the Human Brain. Neuron 43, 271–282.

Wagenmakers, E.J., 2007. A practical solution to the pervasive problems of p values. Psychon. Bull. Rev. https://doi.org/10.3758/BF03194105

Werker, J.F., Hensch, T.K., 2015. Critical Periods in Speech Perception: New Directions. Annu. Rev. Psychol. 66, 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Wessinger, C.M., Vanmeter, J., Tian, B., Lare, J. Van, Pekar, J., Rauschecker, J.P., 2000. Hierarchical Organization of the Human Auditory Cortex Revealed by Functional Magnetic Resonance Imaging S7 1–7.

Wong, P.C.M., Warrier, C.M., Penhune, V.B., Roy, A.K., Sadehh, A., Parrish, T.B., Zatorre, R.J., 2008. Volume of left Heschl's gyrus and linguistic pitch learning. Cereb. Cortex 18, 828–836. https://doi.org/10.1093/cercor/bhm115

Zoellner, S., Benner, J., Zeidler, B., Seither-Preisler, A., Christiner, M., Seitz, A., Goebel, R., Heinecke, A., Wengenroth, M., Blatow, M., Schneider, P., 2019. Reduced cortical thickness in Heschl's gyrus as an in vivo marker for human primary auditory cortex. Hum. Brain Mapp. 40, 1139–1154. https://doi.org/10.1002/hbm.24434

7 Supplementary Information

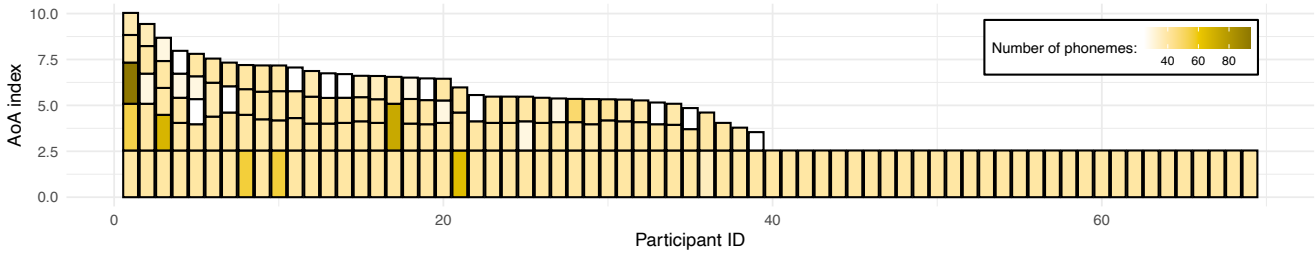

Figure S 1. Illustration of the replication sample's language experience. As in Figure 1, each bar here represents a single participant's overall language experience; the height of the stacked bars within each bar represents the AoA index for individual languages (the taller the bar, the earlier in life a given language was acquired). The color of each stacked bar refers to the number of phonemes in each language's phonological inventory.

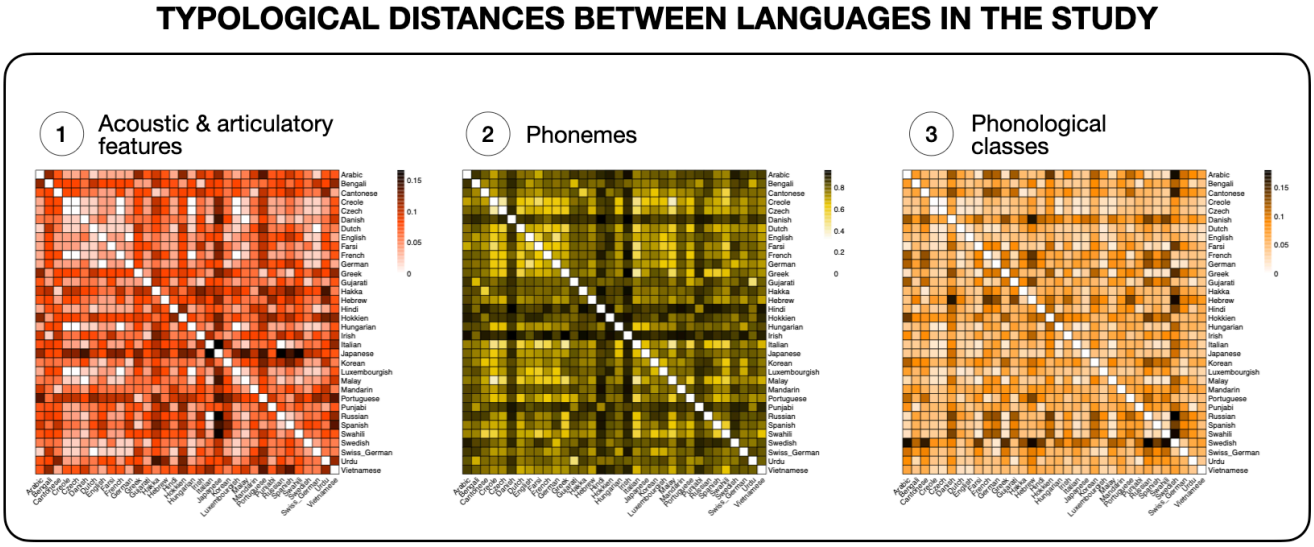

Figure S 2. Similarity matrices of typological distances between all languages represented in the study (N = 36) based on: (1) distances in distinctive acoustic and articulatory features describing the phonemes of each language (e.g., "short", "long"); (2) distances in sets of phonemes belonging to each language; and (3) distances based on counts of phonological classes that share certain features (e.g., “consonants”, "front rounded vowels", "clicks"). Data for individual languages were collected from the PHOIBLE database (Moran et al., 2019) and open-source software (Dediu & Moisik, 2016). The figure was generated in R, with the package `pheatmap` (Kolde, 2019), version 1.0.12.

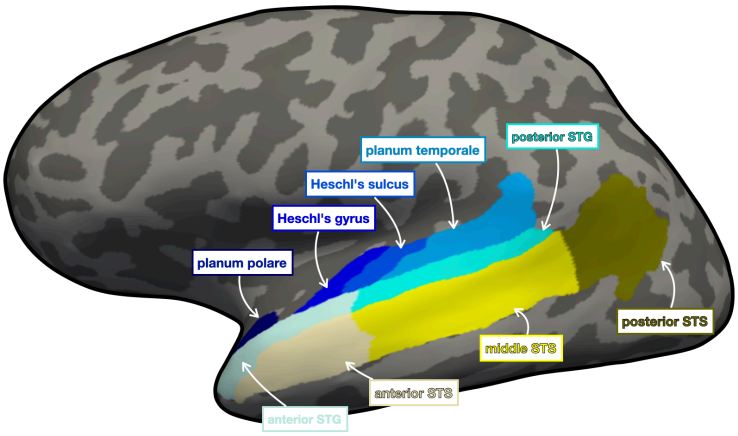

Figure S 3. Auditory ROIs used in the analysis. The ROIs are overlaid on an inflated surface in the native space of one of the participants.

7.1 Auditory cortex regions and language experience

		VOLUME	AREA	THICKNESS
(Intercept)	β	1711.89***	58.10***	0.58***
	SE	(35.28)	(11.85)	(0.02)
Age	β	1.01	0.07	0.00
	SE	(0.88)	(0.24)	(0.00)
Sex	β	0.63	-7.41	0.02
	SE	(26.07)	(8.07)	(0.01)
Whole-brain: Volume / Area / Thickness	β	0.00***	0.01***	0.77***
	SE	(0.00)	(0.00)	(0.09)
Language experience	β	41.50	7.45	0.03
	SE	(66.09)	(22.33)	(0.03)
Posterior STG	β	-1548.51***	-257.01***	-0.34***
	SE	(46.38)	(15.85)	(0.02)
Anterior STS	β	-2465.10***	-190.57***	-0.77***
	SE	(46.38)	(15.85)	(0.02)
Middle STS	β	-688.17***	646.88***	-1.01***
	SE	(46.38)	(15.85)	(0.02)
Posterior STS	β	-185.92***	861.89***	-1.01***
	SE	(46.38)	(15.85)	(0.02)
PT	β	-2064.07***	-163.35***	-0.80***
	SE	(46.38)	(15.85)	(0.02)
HG	β	-2886.10***	-482.40***	-0.78***
	SE	(46.38)	(15.85)	(0.02)
HS	β	-3416.84***	-552.82***	-0.97***
	SE	(46.38)	(15.85)	(0.02)
PP	β	-1870.15***	-316.73***	0.19***
	SE	(46.38)	(15.85)	(0.02)
Hemisphere	β	-478.00***	-103.17***	0.02
	SE	(46.38)	(15.85)	(0.02)
Language experience x Posterior STG	β	-54.04	-20.53	-0.02
	SE	(90.37)	(30.88)	(0.05)
Language experience x Anterior STS	β	-104.26	-28.61	-0.01
	SE	(90.37)	(30.88)	(0.05)
Language experience x Middle STS	β	23.56	28.91	-0.02
	SE	(90.37)	(30.88)	(0.05)
Language experience x Posterior STS	β	40.34	38.03	-0.05
	SE	(90.37)	(30.88)	(0.05)
Language experience x PT	β	-172.20.	-35.12	-0.12*
	SE	(90.37)	(30.88)	(0.05)
Language experience x HG	β	-39.06	-4.31	-0.02
	SE	(90.37)	(30.88)	(0.05)
Language experience x HS	β	-82.13	-25.18	-0.01

	SE	(90.37)	(30.88)	(0.05)
Language experience x PP	β	-91.67	-9.65	-0.07
	SE	(90.37)	(30.88)	(0.05)
Language experience x Hemisphere	β	41.88	-17.55	0.02
	SE	(90.37)	(30.88)	(0.05)
Posterior STG x Hemisphere	β	407.62***	99.63***	0.02
	SE	(65.60)	(22.42)	(0.03)
Anterior STS x Hemisphere	β	371.93***	67.74**	-0.05
	SE	(65.60)	(22.42)	(0.03)
Middle STS x Hemisphere	β	440.40***	43.39.	0.11**
	SE	(65.60)	(22.42)	(0.03)
Posterior STS x Hemisphere	β	1186.10***	435.79***	0.01
	SE	(65.60)	(22.42)	(0.03)
PT x Hemisphere	β	222.12***	7.78	0.04
	SE	(65.60)	(22.42)	(0.03)
HG x Hemisphere	β	255.17***	19.78	0.09**
	SE	(65.60)	(22.42)	(0.03)
HS x Hemisphere	β	363.43***	23.36	0.16***
	SE	(65.60)	(22.42)	(0.03)
PP x Hemisphere	β	486.42***	146.40***	-0.15***
	SE	(65.60)	(22.42)	(0.03)
Language experience x Posterior STG x Hemisphere	β	-36.81	32.36	-0.02
	SE	(127.81)	(43.68)	(0.07)
Language experience x Anterior STS x Hemisphere	β	75.05	52.32	0.01
	SE	(127.81)	(43.68)	(0.07)
Language experience x Middle STS x Hemisphere	β	15.40	16.32	0.01
	SE	(127.81)	(43.68)	(0.07)
Language experience x Posterior STS x Hemisphere	β	50.84	52.48	-0.03
	SE	(127.81)	(43.68)	(0.07)
Language experience x PT x Hemisphere	β	31.46	33.64	0.03
	SE	(127.81)	(43.68)	(0.07)
Language experience x HG x Hemisphere	β	-71.22	8.23	0.00
	SE	(127.81)	(43.68)	(0.07)
Language experience x HS x Hemisphere	β	-13.56	24.11	0.02
	SE	(127.81)	(43.68)	(0.07)
Language experience x PP x Hemisphere	β	-37.13	13.98	0.02
	SE	(127.81)	(43.68)	(0.07)
SD (Intercept id)		*91.84*	*24.92*	*0.06*
SD (Observations)		*382.49*	*130.71*	*0.20*
Num.Obs.		*2448*	*2448*	*2448*
R2 Marg.		*0.904*	*0.937*	*0.793*
R2 Cond.		*0.910*	*0.939*	*0.812*
AIC		*35864.3*	*30661.7*	*-580.1*
BIC		*36102.2*	*30899.7*	*-342.2*
ICC		*0.1*	*0.0*	*0.1*
RMSE		*374.14*	*128.27*	*0.19*

STG: superior temporal gyrus, STS: superior temporal sulcus, PT: planum temporale, HG: Heschl's gyrus, HS: Heschl's sulcus, PP: planum polare

. $p < 0.1$, + $p = 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S 1. Results of linear mixed models (parameter estimates and standard errors, in brackets; *p*-values are listed according to the coding presented underneath the table) testing the effect of language experience on the structure (volume, area, and average thickness) of the auditory regions: planum polare, Heschl's gyrus, Heschl's sulcus, planum temporale, anterior and posterior superior temporal gyrus, and anterior, middle, and posterior superior temporal sulcus. Anterior STG was used as the reference level.

7.2 Superior temporal region and language experience

		VOLUME	AREA	THICKNESS
(Intercept)	β	-86.44***	-31.09***	0.05**
	SE	(14.45)	(4.14)	(0.01)
Language experience	β	-4.83	0.92	-0.06*
	SE	(32.25)	(9.24)	(0.03)
Age	β	-1.33	-0.30	0.00
	SE	(0.86)	(0.22)	(0.00)

Sex	β	14.78	2.72	0.01
	SE	(12.69)	(3.66)	(0.01)
Whole-brain: volume / area / thickness	β	0.00***	0.00***	0.98***
	SE	(0.00)	(0.00)	(0.16)
2nd gyrus	β	351.19***	119.93***	-0.13***
	SE	(17.05)	(4.89)	(0.01)
3rd gyrus	β	-169.80***	-50.83***	0.01
	SE	(17.55)	(5.03)	(0.02)
Hemisphere	β	51.04***	19.58***	-0.04**
	SE	(14.42)	(4.13)	(0.01)
Language experience x 2nd gyrus	β	-30.18	-14.89	0.09**
	SE	(36.03)	(10.33)	(0.03)
Language experience x 3rd gyrus	β	-5.66	3.50	-0.05
	SE	(36.88)	(10.57)	(0.03)
Language experience x hemisphere	β	-27.16	-7.25	0.00
	SE	(31.46)	(9.02)	(0.03)
2nd Gyrus x hemisphere	β	8.34	8.32.	-0.02
	SE	(17.01)	(4.88)	(0.01)
3rd Gyrus x hemisphere	β	-18.79	-6.34	-0.01
	SE	(17.55)	(5.03)	(0.02)
Language experience x 2nd gyrus x hemisphere	β	-11.32	-2.86	-0.02
	SE	(36.04)	(10.34)	(0.03)
Language experience x 3rd gyrus x hemisphere	β	30.25	11.01	-0.03
	SE	(36.88)	(10.57)	(0.03)
SD (Intercept id)		*0.01*	*0.00*	*0.07*
SD (Observations)		*257.90*	*73.94*	*0.22*
Num.Obs.		*567*	*567*	*567*
R2 Marg.		*0.521*	*0.596*	*0.254*
R2 Cond.		*0.521*		*0.324*
AIC		*7836.9*	*6455.5*	*66.9*
BIC		*7910.7*	*6529.3*	*140.7*
ICC		*0.0*		*0.1*
RMSE		*254.47*	*72.95*	*0.21*

. $p < 0.1$, + $p = 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S 2. Results of the linear mixed models testing the effect of Language Experience on the structure (volume, area and average thickness) of the gyri in the superior temporal region: first, second and third TTG. Anterior TTG was used as the reference level.

7.3 Vertex-wise analysis

Figure S 4. Results of a whole-brain vertex-wise analysis, aimed at establishing relations between the language experience index and whole-brain cortical thickness. Overlaid on the inflated surface of the fsaverage template brain is the thresholded at $p < .0001$ (uncorrected) significance map from the conducted *F*-test showing a negative relationship between cortical thickness in the highlighted region and the degree of multilingual language experience.

7.4 Second transverse temporal gyrus and effects of language typology

Kolmogorov-Smirnov normality tests were run on all models' residuals, revealing that parametric linear models were appropriate for the present data (all *ps* > .4). According to a frequentist analysis of the data, all eight overall regression models were statistically significant: (1) Left hemisphere: $F(4,125) = 7.802$, $p < .001$, $F(4,125) = 6.32$, $p < 0.001$, $F(4,125) = 8.287$, $p < 0.001$, $F(4,125) = 6.98$, $p < 0.001$, respectively for the models with the four language experience indices: (a) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (b) acoustic/articulatory features, (c) phonemes, and (d) counts of phonological classes; and (2) Right hemisphere: $F(4,91) = 4.295$, $p = 0.003$, $F(4,91) = 3.461$, $p = 0.01$, $F(4,91) = 4.331$, $p = 0.002$, $F(4,91) = 3.722$, $p = 0.007$, respectively for the models with the four language experience indices: (a) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (b) acoustic/articulatory features, (c) phonemes, and (d) counts of phonological classes.

7.4.1 Cumulative phoneme inventory and the second TTG

The two models including the 'cumulative phoneme inventory' measure were statistically significant: (1) Left hemisphere: $F(5,124) = 8.308$, $p < .001$, and (2) Right hemisphere: $F(5,90) = 3.42$, $p = 0.007$, see further Table S 3. Figure S 5 presents the relationship between the thickness of the second TTG (left and right) and the cumulative phoneme inventory.

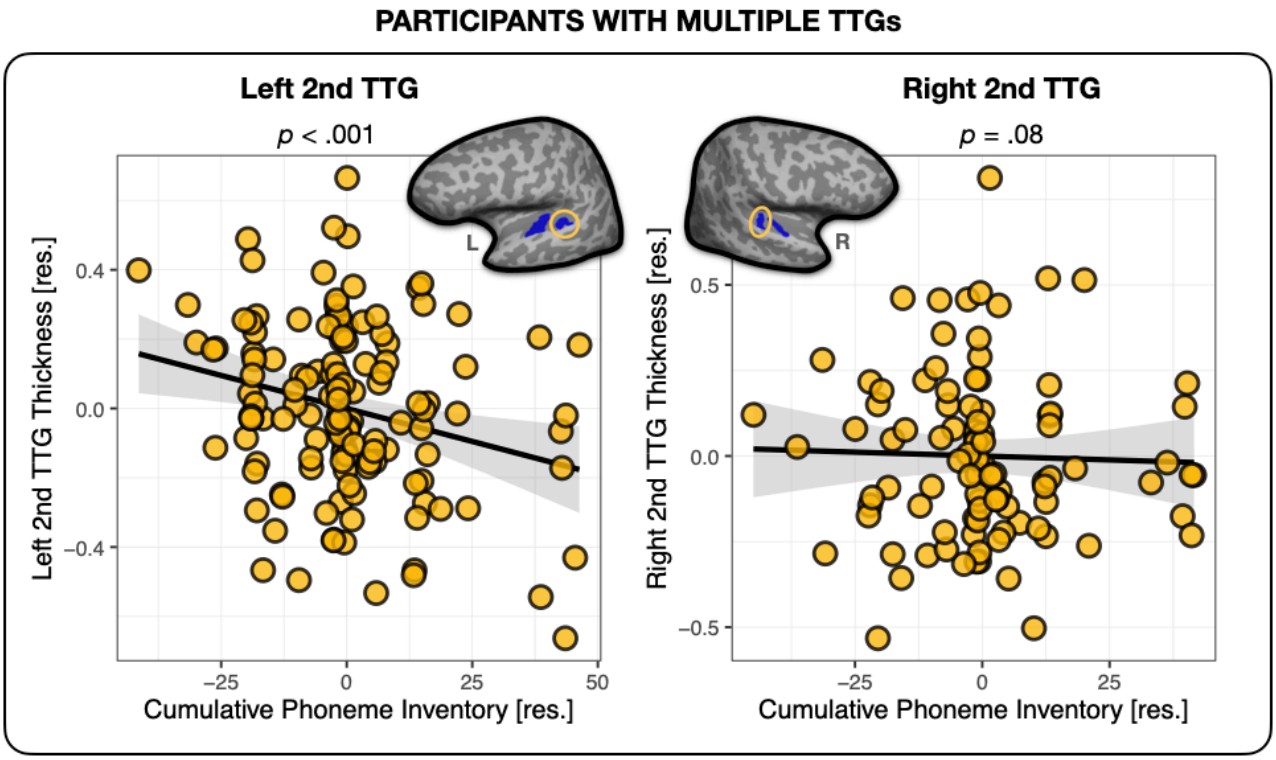

Figure S 5. Cumulative phoneme inventory and thickness of the second TTG. Average thickness of the second TTG in the left and right hemisphere in relation to the number of unique phonemes each participant was exposed to across all their languages (the plotted values are residuals controlled for age, sex, mean hemispheric thickness and the language experience index irrespective of typology).

		left			*right*		
Language experience Models:		Cumulative Phoneme inventory	Language experience	Language experience (Phonemes)	Cumulative phoneme inventory	Language experience	Language experience (Phonemes)
(Intercept)	β	0.00	0.00	0.00	0.00	0.00	0.00
	SE	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Language experience	β	0.05	-0.12**	-	-0.08	-0.10+	-
	SE	(0.07)	(0.05)	-	(0.09)	(0.05)	-
Phonemes	β	0.00**	-	-	0.00	-	-
	SE	(0.00)	-	-	(0.00)	-	-
Language experience (Phonemes)	β	-	-	-0.35**	-	-	-0.26+
	SE	-	-	(0.12)	-	-	(0.13)
Age	β	0.00.	0.00	0.00	0.00	0.00	0.00
	SE	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean thickness (left/right)	β	1.43***	1.40***	1.37***	0.95*	0.94*	0.92*
	SE	(0.29)	(0.30)	(0.30)	(0.37)	(0.37)	(0.37)
Sex	β	0.00	0.00	0.00	0.01	0.00	0.01
	SE	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Num.Obs.		*130*	*130*	*130*	*96*	*96*	*96*
R2		*0.251*	*0.200*	*0.210*	*0.160*	*0.159*	*0.160*
R2 Adj.		*0.221*	*0.174*	*0.184*	*0.113*	*0.122*	*0.123*
AIC		*11.9*	*18.5*	*16.9*	*13.8*	*11.9*	*11.8*
BIC		*31.9*	*35.7*	*34.1*	*31.8*	*27.3*	*27.2*
Log.Lik.		*1.062*	*-3.232*	*-2.430*	*0.085*	*0.036*	*0.101*
F		*8.308*	*7.802*	*8.287*	*3.420*	*4.295*	*4.331*
RMSE		*0.24*	*0.25*	*0.25*	*0.24*	*0.24*	*0.24*
ΔR2 Adjusted		-	*-0.04*	*-0.03*	-	*0.01*	*0.01*
BF_{10}		-	*0.15*	*0.34*	-	*9.33*	*9.96*
BF_{01}		-	*6.43*	*2.88*	-	*0.11*	*0.10*

. $p < 0.1$, + $p = 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S 3. left and right second transverse temporal gyri and cumulative phoneme inventory. Multiple regression model parameters (parameter estimates and standard errors, in brackets; *p*-values are listed according to the coding presented underneath the table) for the average cortical thickness of the second TTG (left and right), as predicted by the "cumulative phoneme inventory" index. For comparison, models with the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of phonemes are also reported. Last two rows present model comparison results (additional variance explained and BF_{10} and BF_{01} values are also reported).

7.5 Language experience in participants with a single TTG

Kolmogorov-Smirnov normality tests were run on all models' residuals, revealing that parametric linear models were appropriate for the present data (all *p*s > .29). According to a frequentist analysis of the data, all eight overall regression models were statistically significant: (1) Right Heschl's gyrus: $F(4,35) = 6.365$, $p < .001$, $F(4,35) = 4.851$, $p = 0.003$, $F(4,35) = 5.403$, $p = 0.002$, $F(4,35) = 5.72$, $p = 0.001$, respectively for models with the four language experience indices: (a) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (b) acoustic/articulatory features, (c) phonemes, and (d) counts of phonological classes; and (2) Right planum temporale: $F(4,35) = 2.655$, $p = 0.049$, $F(4,35) = 2.639$, $p = 0.050$, $F(4,35) = 2.66$, $p = 0.048$, $F(4,35) = 2.685$, $p = 0.049$, respectively for models with the four language experience indices: (a) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (b) acoustic/articulatory features, (c) phonemes, and (d) counts of phonological classes.

		Right Heschl's gyrus				*Right planum temporale*			
		No typology	Features	Phonemes	Phonological classes	No typology	Features	Phonemes	Phonological classes
(Intercept)	β	-0.01	-0.01	-0.01	-0.13*	2.52***	2.52***	2.52***	2.50***
	SE	(0.03)	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)	(0.03)	(0.07)
Language experience indices	β	0.15**	0.48+	0.35*	0.82*	0.02	0.03	0.06	0.11
	SE	(0.05)	(0.24)	(0.15)	(0.32)	(0.07)	(0.30)	(0.19)	(0.41)
Age	β	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00
	SE	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sex	β	0.05	0.03	0.05	0.04	0.06	0.05	0.06	0.06
	SE	(0.05)	(0.06)	(0.06)	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
Mean Thickness (right)	β	0.93**	0.99**	0.96**	0.93**	0.90*	0.90*	0.91*	0.90*
	SE	(0.30)	(0.31)	(0.30)	(0.30)	(0.38)	(0.39)	(0.38)	(0.38)
Num.Obs.		40	40	40	40	40	40	40	40
R2		*0.42*	*0.36*	*0.38*	*0.40*	*0.23*	*0.23*	*0.23*	*0.23*
R2 Adj.		*0.36*	*0.28*	*0.31*	*0.33*	*0.15*	*0.14*	*0.15*	*0.15*
AIC		*-36.6*	*-32.4*	*-34.0*	*-34.9*	*-15.6*	*-15.5*	*-15.6*	*-15.6*
BIC		*-26.5*	*-22.3*	*-23.8*	*-24.7*	*-5.4*	*-5.4*	*-5.5*	*-5.4*
Log.Lik.		*24.30*	*22.19*	*22.99*	*23.43*	*13.79*	*13.76*	*13.81*	*13.79*
RMSE		*0.13*	*0.14*	*0.14*	*0.13*	*0.17*	*0.17*	*0.17*	*0.17*
ΔR2 Adjusted		-	*-0.07*	*-0.04*	*-0.03*	-	*> -0.01*	*< 0.01*	*< 0.01*
BF_{10}		-	*0.12*	*0.27*	*0.42*	-	*0.97*	*1.02*	*1.01*

. $p < 0.1$, + $p = 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S 4. Right superior temporal plane (Heschl's gyrus and planum temporale) and language experience in participants with one TTG. Multiple regression model parameters (parameter estimates and standard errors, in brackets; *p*-values are listed according to the coding presented underneath the table) for the average cortical thickness of the right Heschl's gyrus, and the right planum temporale, as predicted by the four language experience indices: (1) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (2) phonemes, (3) acoustic/articulatory features and (4) counts of phonological classes. Last two rows present model comparison results (additional variance explained and BF_{10} values); NB. all models including typological information were compared against the 'No typology' model.

7.6 Effects of language proficiency

The findings of thinner cortex in the second TTG (when present) were further followed up by an analysis investigating the effect of proficiency attained by the participants in their individual languages. Here, we again constructed cumulative language experience measures per participant using the Shannon's entropy equation (Shannon, 1948), but this time from the self-reported proficiency ratings (on a scale from 0 to 10) of each of participants' languages. First, we computed a general proficiency score and used it in two linear models to establish whether the AoA-derived cumulative language experience measure could be replicated by the proficiency-derived measure, since AoA and proficiency are known to be highly correlated. This analysis therefore served as a 'sanity check' for the results obtained from the analyses reported in Section 2.2. Indeed, we observed that the thickness of the second TTG (when present) was negatively related to the cumulative language experience measure derived from proficiency ratings. This effect was significant for the left hemisphere 2nd TTG thickness, and showed a trend towards significance for the right ($\beta = -0.11$, $t = -2.41$, $p = .018$, small effect size: $f^2 = 0.05$; and $\beta = -0.10$, $t = -1.87$, $p = .065$, small effect size: $f^2 = 0.05$, for left and right 2nd TTG respectively).

Given the hypothesis that multilingualism is a dynamic process reshaping brain structure and inducing both increases (in initial stages) and decreases (in peak efficiency) in brain morphological indices (Pliatsikas, 2020), we further explored the effect of proficiency of different languages by investigating the role of languages spoken proficiently and

non-proficiently on the structure of the 2nd TTG. Here, we computed two further different indices per participant: (1) derived only from languages spoken proficiently (rated at 7 and higher out of 10 on a self-reported proficiency scale), and (2) derived from languages spoken at a low level of proficiency (rated lower than 7 out of 10). We subsequently used these indices in two further analyses to test the hypothesis that decreases in brain morphological indices (here in cortical thickness) are driven by peak efficiency of language functioning (operationalized here as high proficiency in many languages). We did observe that the thickness of the 2nd TTG was related to the measure of language experience derived from languages spoken proficiently, however this effect was only significant in the right hemisphere (β = -0.18, t = -2.89, p = .005, small effect size: f^2 = 0.11; and β = -0.02, t = -0.44, p = .66, for right and left 2nd TTG respectively). 2nd TTG's thickness was not related to the language experience measure derived from languages spoken at low level of proficiency (β = -0.01, t = -0.11, p = .91; and β = -0.12, t = -1.69, p = .097 for right and left 2nd TTG respectively), see Figure S 6.

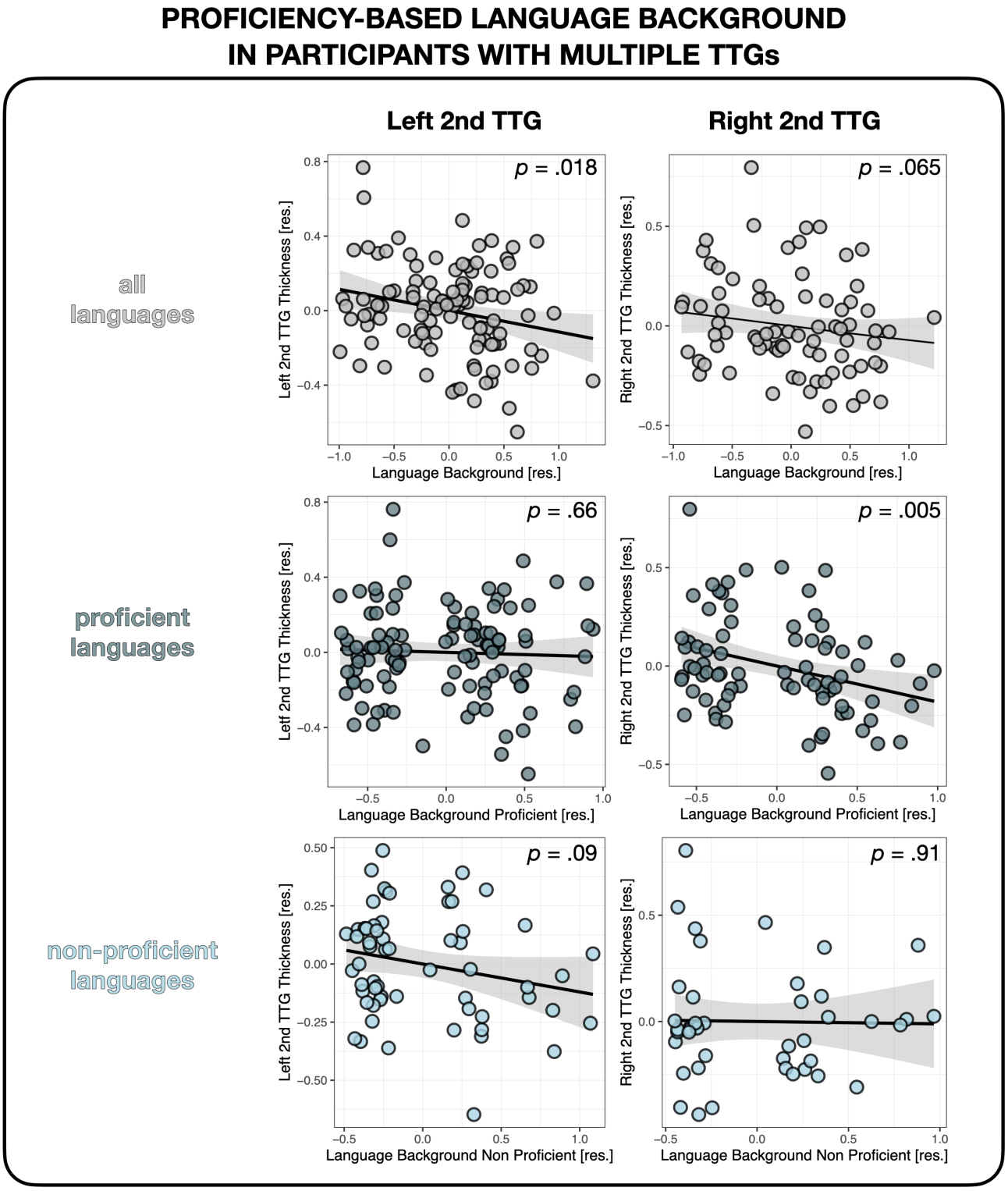

Figure S 6. Multilingual proficiency and thickness of the second TTG. Average thickness of the second TTG in the left and right hemisphere as a function of language proficiency in all languages of each participant (top panel), only their proficient languages (middle panel) and only their non-proficient languages (bottom panel). Plots show residuals, controlling for age, sex and mean hemispheric thickness.

7.7 Replication analysis

Kolmogorov-Smirnov normality tests were run on all models' residuals, revealing that parametric linear models were appropriate for the present data (all *p*s > .5). According to a frequentist analysis of the data, all four overall regression models for left hemisphere data were statistically not significant: $F(5,55) = 1.109$, $p < .001$, $F(5,55) = 1.12$, $p = .36$, $F(5,55) = 8.287$, $p = .37$, $F(5,55) = 1.072$, $p = .38$, respectively for models with the four language experience indices: (a) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (b) acoustic/articulatory features, (c) phonemes, and (d) counts of phonological classes. Models for the right hemisphere data were all statistically significant, apart from the counts of phonological classes model: $F(5,48) = 2.781$, $p = 0.03$, $F(5,48) = 5.158$, $p < 0.001$, $F(5,48) = 3.244$, $p = 0.013$, $F(5,48) = 2.342$, $p = 0.06$, respectively for models with the four language experience indices: (a) the cumulative language experience measure not accounting for typology, and cumulative language experience weighted by overlaps between languages at the level of (b) acoustic/articulatory features, (c) phonemes, and (d) counts of phonological classes.

7.8 Phonological classes as defined by Dediu and Moisik (2016)

(v = vowels, c = consonants) segments, vowels, monophtongs, diphtongs, triphtongs, heights.v, lengths.v, long.v, nasal.v, round.v, high.v, mid.v, low.v, front.v, back.v, tense.v, lax.v, atr.v, rtr.v, raised.v, retracted.v, fronted.v, glottalized.v, unique.v, unique.nasal.v, heights_mono.v, heights_di.v, heights_tri.v, lengths_mono.v, lengths_di.v, lengths_tri.v, long_mono.v, long_di.v, long_tri.v, nasal_mono.v, nasal_di.v, nasal_tri.v, round_mono.v, round_di.v, round_tri.v, high_mono.v, high_di.v, high_tri.v, mid_mono.v, mid_di.v, mid_tri.v, low_mono.v, low_di.v, low_tri.v, front_mono.v, front_di.v, front_tri.v, back_mono.v, back_di.v, back_tri.v, tense_mono.v, tense_di.v, tense_tri.v, lax_mono.v, lax_di.v, lax_tri.v, atr_mono.v, atr_di.v, atr_tri.v, rtr_mono.v, rtr_di.v, rtr_tri.v, raised_mono.v, raised_di.v, raised_tri.v, retracted_mono.v, retracted_di.v, retracted_tri.v, fronted_mono.v, fronted_di.v, fronted_tri.v, glottalized_mono.v, glottalized_di.v, glottalized_tri.v, unique_mono.v, unique_di.v, unique_tri.v, unique.nasal_mono.v, unique.nasal_di.v, unique.nasal_tri.v, consonants, places.c, bilabial.c, labiodental.c, dental.c, alveolar.c, dental_alveolar.c, palatoalveolar.c, alveolopalatal.c, postalveolar.c, true_retroflex.c, palatal.c, velar.c, uvular.c, pharyngeal_epiglottal.c, glottal.c, labial.c, coronal.c, dorsal.c, guttural.c, manners.c, obstruent.c, voiced_obstruent.c, voiceless_obstruent.c, aspirated_obstruent.c, glottalized_obstruent.c, stop.c, voiced_stop.c, voiceless_stop.c, aspirated_stop.c, glottalized_stop.c, fricative.c, voiced_fricative.c, voiceless_fricative.c, affricate.c, sonorant.c, voiced_sonorant.c, voiceless_sonorant.c, glottalized_resonant.c, nasal.c, approximant.c, tapflap.c, trill.c, trill_tap.c, coronal_trill_tap.c, glottalized.c, uvt.c, uvt.stops.c, uvt.fricatives.c, uvt.affricates.c, uvt.nasals.c, uvt.approximants.c, lvt.c, lvt.stops.c, lvt.fricatives.c, lvt.affricates.c, lvt.nasals.c, lvt.approximants.c, ratio.voiced.voiceless.obstruents, ratio.voiced.voiceless.stops, ratio.obstruents.sonorants, egressive, implosive, ejective, click, voiceless, voiced, breathy, creaky, tones, bilabial.fricatives, labiodental.fricatives, alveolar.fricatives, nonsibilant.dental.fricatives, sibilant.dental.fricatives, bilabiallabiodental.affricates, bilabial.affricates, retroflex.stops, retroflex.fricatives, retroflex.affricates, retroflex.nasals, retroflex.approximants