MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

Language Aptitude and the Morphology and Shape of the Auditory Cortex Transverse Temporal Gyri

verfasst von / submitted by

Carmen Ramoser BSc BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 878
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Verhaltens-, Neuro-, und Kognitionsbiologie
Betreut von / Supervisor:	Assoz. Prof. Narly Golestani, BSc PhD
Mitbetreut von / Co-Supervisor:	lic. Olga Kepinska, MA MA PhD

Content

Abstract

Why is it that some people seem to learn new languages faster and easier than others? The present study investigates the neuroanatomical basis of language learning aptitude, with a focus on the shape of the transverse temporal gyrus/gyri (TTG/TTGs) within the Sylvian fissure of the auditory cortex. The size and shape of the first TTG (i.e., Heschl's gyrus; HG) and of additional posterior TTG(s), when present, are highly variable both between brain hemispheres and individuals. Previous work has shown the shape of the TTG to be related to musical and linguistic abilities. Specifically, one study found that high language learning aptitude correlated with more TTGs in the right hemisphere, even though language functions are generally left-lateralized. In this study, we used TASH (Toolbox for the Automated Segmentation of Heschl's Gyrus) and the newly developed MCAI (Multivariate Concavity Amplitude Index) toolbox, to automatically extract TTG's morphology and shape from 82 MRI scans and relate it to participants' language aptitude scores. In contrast to the previous results, we found that what sets average form high language aptitude apart was less TTGs in the right hemisphere, with more surface area in the first right TTG and second left TTG. Furthermore, high executive working memory function and many languages learned in life were associated with high language learning aptitude. This corroborates previous findings of the importance of left lateralization, secondary auditory cortex and working memory in language learning and it opens questions on how aptitude relates to seeking experiences.

Zusammenfassung

Warum lernen manche Menschen neue Sprachen scheinbar schneller und leichter als andere? Die vorliegende Studie untersucht die Neuroanatomie der Begabung zum Sprachenlernen, mit einem Fokus auf die Form des *Gyrus temporales transversus/Gyri temporales transversi* (TTG/TTGs), die in der *Fossa Sylvii* des Hörzentrums liegen. Die Morphologie und Form des ersten TTG (genannt Heschl'scher Gyrus; HG) sowie, sofern vorhanden, zusätzlicher TTG(s) variieren sowohl zwischen Gehirnhälften als auch zwischen Individuen stark. Frühere Arbeiten haben Zusammenhänge zwischen der Morphologie und Form der TTGs und musikalischen und sprachlichen Fähigkeiten gefunden. In einer Studie wurde festgestellt, dass Personen mit hoher Sprachlernbegabung mehr TTGs in der rechten Gehirnhälfte aufweisen, obwohl Sprachfunktionen generell in der linken Gehirnhälfte verortet sind. In dieser Studie nutzten wir TASH (Toolbox for the Automated Segmentation of Heschl's Gyrus) und die neu entwickelte MCAI (Multivariate Concavity Amplitude Index) Toolbox, um die Morphologie und Form der TTGs automatisiert aus 82 MRT-Scans zu extrahieren und mit der Sprachlernbegabung in Beziehung zu setzen. Im Gegensatz zu der früheren Studie fanden wir, dass hohe Sprachlernbegabung im Vergleich zu durchschnittlicher sich durch weniger TTGs in der rechten Gehirnhälfte auszeichnet, dafür aber größerer Oberfläche im ersten

rechten TTG und im zweiten linken TTG. Darüber hinaus waren hohe Werte des exekutiven Arbeitsgedächtnisses und viele erlernte Sprachen mit hoher Sprachlernbegabung korreliert. Diese Ergebnisse unterstützen frühere Erkenntnisse zu der Wichtigkeit der linken Gehirnhälfte, des sekundären Hörzentrum und des Arbeitsgedächtnisses beim Sprachenlernen und sie werfen Fragen darüber auf, wie Begabung mit dem Streben nach Erfahrungen zusammenhängt.

Acknowledgement

I want to thank Narly Golestani for giving me the opportunity to write my master thesis in her lab. I think I found the perfect place to combine my background in biology with my interest in language and I learned a lot about cognitive neuroscience and its research. It was a pleasure to get to know everybody of the lovely brain and language lab, be it in Vienna or Geneva, and be surrounded by smart and warm-hearted people.

The biggest thanks go to Olga Kepinska for being the best co-supervisor and role model that I could have. The clear structure from the beginning and the help with all my questions made it all manageable. I am incredibly grateful for all the support and encouragement in my first steps in research, be it R, poster presentations or applications and interviews for PhD positions. I feel really honored to publish our findings in a paper as a co-author. I enjoyed immensely sitting with you and Sevil Maghsadhagh in the office and talking about life in research and everything else.

Also, thank you, Sevil, for all the tea breaks and enjoyable chats. The time since September flew by so fast!

I want to thank also my friend, Chiara Angori, for always believing in me and cheering me on. You are my greatest fan!

Special thanks go to my boyfriend Daniele Pretolesi for always being there for me and listening to my enthusiastic talks about my project or my despair when having to write about it, taking care that I eat and rest. You always support and encourage me with everything I do and I am eternally grateful.

Of course, I thank my parents as well, who supported me financially and made it possible for me to go my own way, even if the name of my master is very complicated.

1 Theoretical Background

1.1 Language Aptitude

Individuals differ in their rate and success in foreign language (L2) learning and apart from age and motivation, this is mostly determined by a certain "talent" for learning languages (Skehan, 2002; Dörnyei & Skehan, 2003; Abrahamsson & Hyltenstam, 2008; Biedroń & Pawlak, 2016). Since the last century, this "talent" has been formalized and studied under the name of foreign language aptitude or simply language aptitude (Carroll, 1981; Dörnyei & Skehan, 2003; Wen et al., 2017; Turker et al., 2019). The 'founding father' of aptitude research, John Carroll, saw aptitude as the capacity to learn fast and with facility, this ability being relatively stable and innate (Carroll, 1981). He conceptualized language aptitude as four distinct and measurable abilities: phonemic coding ability (capacity to code unfamiliar sound so that it can be retained), grammatical sensitivity (capacity to identify the functions that words fulfil in sentences), inductive language ability (capacity to extrapolate from a given corpus to create new sentences) and associative memory (capacity to form associative links in memory) (Carroll, 1981; Dörnyei & Skehan, 2003). Since then, these categories held up during empirical investigations, but due to their similarity, 'grammar sensitivity' and 'inductive language learning' were combined to the concept 'language analytic ability' (Skehan 2002, Biedroń, 2015; Biedroń & Pawlak, 2016; Wen et al., 2017; Turker et al., 2017, 2019). Currently, language aptitude is also considered to be more dynamic than it was in Carroll's time, meaning it is recognized to change for example in regard to age (Robinson, 2007; Z. Wen et al., 2017).

Carroll and his colleague Stanley Sapon (1959) developed the first comprehensive aptitude test battery, the Modern Language Aptitude Test (MLAT). While the MLAT was made for English speakers and had to be adapted for different mother tongues (L1s), Meara (2005) developed on its basis the language-independent LLAMA test with four subtests: vocabulary acquisition (LLAMA_B), sound recognition (LLAMA_D), sound-symbol correspondence (LLAMA_E) and grammatical inferencing (LLAMA_F). The test gained popularity, as it is free, quick to administer and easily available (Rogers et al., 2017), which is why it was used in this study. A meta-analysis by Li (2015) showed that language aptitude test scores were indeed positively correlated with ultimate L2 proficiency and independent of factors like motivation. Also, Rogers et al. (2017) found that LLAMA test scores are robust to background variables with the only limitations that participants with prior L2 instructions receive higher scores compared to monolinguals, and younger children receive lower scores compared to adults (Rogers et al. 2017).

Since the beginning of the 2000s there has been a new wave of theorizing in language aptitude (see Wen et al., 2017 for review). Special emphasis was given to the importance of working memory, as better working memory correlates with higher language aptitude, faster language learning and greater L2 final proficiency (Skehan, 2002; Baddeley, 2003; Biedroń, 2015; Linck et al., 2014; Wen et al., 2017; Wen & Skehan, 2011). For example, Turker et al. (2019) found that from 17 behavioral variables, those relying on language aptitude and working memory formed a joint component in a principal component analysis.

There is also a relationship between language aptitude and musicality, especially for pronunciation (Christiner & Reiterer, 2013; Milovanov & Tervaniemi, 2011; Turker et al., 2017, 2019; Vangehuchten et al., 2015). An explanation for this can be the shared neurological underpinnings of language and musical abilities (Turker et al., 2021; Turker & Reiterer, 2021).

1.2 Cognitive Abilities and Neuroanatomy

Neurolinguistic research of language aptitude has been growing in the past years and especially phonological abilities have been thoroughly analyzed (Biedroń, 2015). Brain structures that are most often related to language aptitude are grey and white matter volumes of the left inferior parietal lobe (IPL), the left inferior frontal cortex and the auditory cortices (Turker et al., 2021). For example, Golestani et al. (2007) found that faster learning of a foreign consonant correlated with higher volume in the left IPL and Novén et al. (2019) found that higher cortical thickness in Broca's area (left inferior frontal gyrus/IFG) was related to higher language analytic abilities. Reiterer found that less talented learners have greater activation in language related areas due to increased "control effort" and highly talented individuals show enhanced grey matter volume (Christiner & Reiterer, 2013; Reiterer et al., 2013). She suggests that high language aptitude is due to reduced effort in speech production as well as increased cortical efficiency.

This study focuses on the auditory cortex, specifically the transverse temporal gyrus/gyri (TTG/TTGs) in the Sylvian fissure. The auditory cortex (Brodmann's area 41) is where auditory input like speech, music and environmental sounds are processed (Moerel et al., 2014). The primary auditory cortex is located at the medial end of the first TTG, also called Heschl's Gyrus (HG) after its first descriptor Richard Heschl (Rademacher et al., 1993). The TTG(s) are located on the superior plane of the superior temporal gyrus (STG) in the Sylvian fissure. Posterior to HG lies the planum temporale (PT), which is known for its hemispheric asymmetry consistent with the left-hemisphere specialization for language (Geschwind & Levitsky, 1968; Moerel et al., 2014). Also the left HG has consistently more volume than the right due to more cortical white matter (Marie et al., 2015; Penhune et al., 1996). HG and PT are separated by the first Heschl's sulcus (HS) and posterior to HS additional TTGs can be located, which are considered part of PT as they do not contain the primary auditory cortex (Marie et al., 2015; Penhune et al., 1996; Rademacher et al., 1993). TTG morphology (i.e. surface area, thickness and volume) and shape (i.e. gyrification patterns) is highly variable both between individuals and between hemispheres (Marie et al., 2015; Rademacher et al., 1993). In the past, additional TTGs have often been referred to as duplicated (or triplicated) HGs. Furthermore, a sulcus intermedius (SI) can also divide HG incompletely. It has become practice to categorize TTG shape into a single gyrus, a common stem duplication (CSD), defined as a SI that divides one-third of HG at the lateral, but not the medial end, a complete posterior duplication (CPD), defined by a first Heschl's sulcus (HS), that splits HG at its medial end, and multiple duplications (MD), which include all possible variations with two HS or combined CSD/CPD structures (Benner et al., 2017; Takahashi et al., 2022; Turker et al., 2019). The study of Marie et al. (2015) revealed that 64% of the right-handed population have at least one additional TTG besides bilateral single HG. More people have an additional TTG in the right (49%) than in the left (37%) hemisphere. The most frequent hemisphere configuration is bilateral single HG (36%), followed by two right TTGs and a single left HG (27%) (Marie et al., 2015).

Studies of the relationship between musical experience and TTG morphology and shape found that 90% of the professional musicians exhibited additional TTGs (Benner et al., 2017) and that professional musicians, compared to non-musicians have higher gray matter volume at the medial end of HG (Schneider et al., 2002) and higher overall volume of HG bilaterally (Schneider et al., 2005). High musical aptitude in primary school children correlated with the intensity of musical practice and higher HG volume overall, particularly in the right HG, while musical practice alone did not influence volume as much (Seither-Preisler et al., 2014). Grey matter volume of the right HG correlated also with absolute pitch perception proficiency (Wengenroth et al., 2014).

In the domain of language functions, more successful learners of foreign speech sounds and linguistic pitch patterns exhibited larger volumes in left HG and more left TTGs (Golestani et al., 2007; Wong et al., 2008). The same was found for phonetic transcription expertise (Golestani et al., 2011). However, it is of note that even though expert phoneticians were more multilingual than controls, foreign language experience was not correlated with structural brain measures (Golestani et al., 2011). In contrast, Ressel et al. (2012) showed that language experience may be related to auditory cortex morphology after birth, given that early Spanish-Catalan bilinguals were found to have larger HGs, especially in the left hemisphere, and Kepinska et al. (*in preparation*) found a relationship between the thickness of the second TTGs and the languages that people knew. In contrast to language ability, dyslexia in boys was found to be correlated with additional right TTGs (Altarelli et al., 2014). Another disorder connected with a prevalence of additional right TTGs is schizophrenia, specifically the deficit syndrome subtype (Takahashi et al., 2022)

The findings of specific relationships of left and right HG's morphology with language and music, respectively, are in accordance with the spectrotemporal model of acoustic processing of Zatorre et al. (2002) and the theory of asymmetric sampling in time of Poeppel (2003). According to these models, the left HG is specialized for handling fast temporal information like the rapid acoustic changes in phonemes and the right HG is more apt at analyzing fine frequency distinctions like the complex spectral information of music due to the integration of relative longer periods of time (Poeppel, 2003; Warrier et al., 2009; Zatorre et al., 2002). Warrier et al. (2009) corroborated this by correlating participants HG volume with their functional activation depending on an increasing number of spectral components and temporal change rate. Also, volume of left or right lateral HG predicted preference for fundamental or spectral pitch perception, respectively (Schneider et al., 2005).

Thus, it could be expected, that language aptitude would benefit from left hemisphere processing. However, recent research into the neurobiology of language aptitude revealed a different picture: Subjects with high language aptitude and high phonetic coding abilities exhibited predominately additional right TTGs (Turker et al., 2017, 2019). Notably, Turker et al. (2017) found a positive relationship between language aptitude scores and number of instruments played, but not with the number of foreign languages spoken. Also, in an artificial grammar learning task, participants with high language analytical abilities (based on LLAMA_F) showed in general more widespread activation and specifically increased activity in the right hemisphere compared to those with average scores (Kepinska, de Rover, et al., 2017). An EEG study showed that highly skilled learners exhibited stronger local synchronization within the right hemisphere during an artificial grammar learning task (Kepinska, Pereda, et al., 2017). A recent overview article relating individual differences in the right hemisphere to language learning proposed a model, in which the right hemisphere integrates a model of what is happening with memory and detects novelty, which subsequently the left hemisphere uses for learning feature-processing (Prat et al., 2023). As a result, language learning should be facilitated by the right hemisphere in the early stages, while activation in later stages "might reflect poorer performance, more reliance on context, or a slower transition to feature-specific LH systems".

In general, the shape of the TTG is thought to develop in utero (Chi et al., 1977), be stable during development (Seither-Preisler et al., 2014) and be heritable as shown in studies of di- and monozygotic twins (Peper et al., 2007). The unique folding pattern of human brains characterized by convex (gyral) and concave (sulcal) regions serves to increase the surface area (Grasby et al., 2020; Rakic, 1988). Its development is explained by the radial unit model (Rakic, 1988): after migration to the cortex stacks of neurons, called ontogenetic columns, become basic processing units in the adult cortex. The number of these columns determines the surface area of each cytoarchitectonic region, while the thickness is determined by the number of cell divisions produced by them. During evolution, an increase in the number of radial units results in increased surface area and gyrification. This model is corroborated by a modern genome wide association study (GWAS), that showed a higher genetic component to cortical surface area, while cortical thickness was found to be more influenced by regulatory elements in adult brains (Grasby et al., 2020). Therefore, the relationship between TTG's shape and language aptitude could point towards potential innate, genetic influences on this cognitive ability.

The correlation between multiple right TTGs in high language aptitude could explain the close correlation of musicality with language aptitude. Turker et al. (2021) argued that a reason for the interplay between musicality and language aptitude could be a shared genetic basis, that gives rise to distinct abilities that develop side-by-side. They presented a neurocognitive model of language aptitude, where a largely genetically predetermined language aptitude profile develops into a language competence profile. The language aptitude profile possesses a neural basis, visible in the anatomy of the auditory cortex and other language-related regions like the left IPL or the left IFG, and a cognitive basis like auditory and musical capacities, fluid reasoning abilities and memory-related abilities. During development environmental factors interact with neuroanatomical and cognitive predispositions to give rise to a language competence profile.

However, to date the high variability of TTG's shape made manual classification necessary (Benner et al., 2017; Golestani et al., 2007, 2011; Marie et al., 2015; Schneider et al., 2002, 2005; Turker et al., 2017, 2019), which is slow, work-intensive and dependent on somewhat inconsistent definitions. There are also instances where classifications fail (Benner et al., 2017). Due to this need, the Toolbox for the Automated Segmentation of Heschl's Gyrus (TASH) (Dalboni da Rocha et al., 2020) and Multivariate Concavity Amplitude Index (MCAI) toolbox were developed. They segment and characterize the shape of HG and/or additional TTGs in a continuous manner (Dalboni da Rocha et al., 2023), which allows a fully automated and reproducible assessment of shape. For this, MCAI uses structural T1-weighted MRI images, which are segmented by Freesurfer and then by TASH that extracts volume, surface area and shape (Dalboni da Rocha et al., 2020). The output of TASH is used to calculate a lateral MCAI value either of HG alone or of HG with potential additional TTGs (Dalboni da Rocha et al., 2023). The MCAI toolbox was able to replicate the finding of a higher amount of bilateral gyrification in professional musicians compared to non-musicians (Dalboni da Rocha et al., 2023). It revealed, that the shape of HG alone predicted musicianship status better than of all TTGs together.

1.3 Research Questions and Hypotheses

In this study, we will analyze data from participants who completed the LLAMA test and subsequently underwent MRI scanning. We will extract the shape of HG and potential additional TTGs with the TASH and MCAI toolboxes and investigate their relation to language aptitude and determine if this relationship is limited to only HG or if it exists across additional TTGs when present. Finally, self-reported measures of musical experience and language experience and working memory scores will be related to language aptitude and auditory cortex anatomy.

Our hypothesis is that (1) participants with higher language aptitude scores will have higher lateral MCAI values in the right hemisphere and that this will be due to the shape of all TTGs. Furthermore, (2) we expect better working memory to be related to higher language aptitude. We have no prediction for the relationship between working memory and the anatomy of the auditory cortex. Additionally, (3) language experience is expected not to be related to language aptitude, but to be related to TTG morphology. Finally, (4) we expect that participants with more musical experience will have higher language aptitude scores and higher volume of bilateral HG.

2 Methods

2.1 Language Aptitude

Language aptitude was measured with the LLAMA test (Meara, 2005). The LLAMA test is free, computer-based and language-independent relying only on pictures and a made-up language based on Native American languages. It consists of four subtests:

1) LLAMA_B tests vocabulary learning, where participants have 2 minutes to learn 20 pairs of words and imaginary figures. Afterwards, they are asked to identify the correct figure for each word.
2) LLAMA_D tests phonetic memory. Participants hear a stream of unknown words and then the same words interspersed with new words. They have to decide for every word if they already heard it or not.
3) LLAMA_E tests sound-symbol correspondence. Participants have 2 minutes to learn associations between a consonant-vowel syllable and a written symbol consisting of a simple digit-letter combination (e.g. /pa/ may be written as 0í). Afterwards, they are presented with new words and have to choose the correct spelling from two possible ones.
4) LLAMA_F tests grammatical inferencing. Participants are provided with 20 sentences and corresponding imaginary figures in a picture for 5 minutes. They have to infer syntax and semantics of the sentences and in the testing phase they are presented with two sentences for every picture, and they have to decide, which one describes the picture in a grammatically correct way.

For LLAMA_D, scores from 0 to 75 can be obtained. For the other subtests, scores from 0 to 100 can be obtained (Meara, 2005). The total LLAMA score ranges between 0 and 375.

2.2 Participants

307 participants were recruited for the LLAMA test at Leiden University through posters, email invitations and word of mouth advertising from February 2013 until November 2014. They completed it on computers in a computer lab at the Faculty of Humanities. 239 of them completed all parts of the test and the biographical information sheet, had Dutch as first language and were not early bilinguals (i.e., acquired a second language after age of four).

2.3 MRI Data Acquisition

Of those, 82 (59 female, 18-43 years old, $M = 22.83$, $SD = 4.12$) were invited for MRI scanning between January and December of 2014. They were all right-handed. They were chosen for either high or average vocabulary (LLAMA_B) or grammar learning (LLAMA_F) subtest scores (as per criteria of the larger study the data was collected for). Imaging data were acquired using a Philips 3T MR-system (Best, The Netherlands) located at the Leiden University Medical Centre (LUMC) equipped with a SENSE-32 channel head coil. For each subject, an anatomical image including a 3D gradient-echo T1-weighted sequence was acquired (TR = 9.755 ms, TE = 4.59 ms; matrix 256 x 256; voxel size: 1.2 x 1.2 x 1.2 mm; 140 slices).

2.4 Multilingual Experience and Musicality

Participants filled out an online questionnaire where they listed their foreign languages and the age of acquisition for each language. They also answered questions on their musicality, including “Do you play an instrument, or do you sing?”, "Which instruments do you play?" and “How active do you sing or play an instrument?” with a choice between "I'm taking lessons", "I used to take lessons", "I play in an orchestra/ensemble/band", "I sing in a choir/band", "I used to sing in a choir/band", "None of this, I only play/sing at home for myself".

To assess the multilingual experience with between 1 and 5 different foreign languages, we created a single "language entropy" score per participant (following Kepinska et al. 2023 and *in preparation*). To do this, we used age of acquisition (AoA) of each language which were log-transformed (to minimize the differences between values for languages learned later in life) and inverted (to express early AoAs as the highest values). To avoid values equal to zero, a constant value of 1 was added before each step. The language entropy score was computed with Shannon's entropy (H) equation (Shannon, 1948), where n stands for the number of languages participants reported and p_i for the AoA index.

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

It was calculated using the R package entropy (v1.3.1; Hausser & Strimmer, 2021). More extensive multilingual experience is expressed by higher language entropy values.

Two participants that were invited for MRI scanning had incomplete questionnaires and were excluded from analyses of multilingualism and musicality.

2.5 Working Memory

Additionally, participants were invited to perform the Operation Span Task (OSPAN). The OSPAN tests working memory by engaging participants in solving simple mathematical equations and verify the solution offered while at the same time presenting them with to-be-remembered letters (Đokić et al., 2018). After presentation of the letter the next equation follows. A set contains three to seven such sequences and after each set participants have to click on the letters in correct serial order on a computer screen. Each set is presented three times in random order for a total of 75 equations and letters. In a training phase beforehand, the time available for solving the processing tasks is individualized as the average time plus 2.5 SD. The to-be-remembered letters in correct serial order give a measure of recollection versus interference by the processing task. Scores are calculated in two ways: the absolute score counts the to-be-remembered letters only from sets that are completely correctly reproduced, while the partial score counts all to-be-remembered letters in the correct serial position even if the entire set was not completely correctly reproduced. Here, the partial score is used since it has been argued to be a more complete and continuous measure of a participant's performance and has been proven to be more normally distributed and to have higher reliability (Friedman & Miyake, 2005).

60 participants completed the OSPAN (45 female, 18-43 year old, M = 23.18, SD = 4.19).

2.6 Neuroanatomical Analyses

T1-weighted MRI images were processed with FreeSurfer, version 7.2 (Fischl et al., 2004). Freesurfer output was then further segmented by the Toolbox for the Automated Segmentation of Heschl's Gyrus (TASH). For the current study the extension of TASH, TASH_complete was used, which extracts a numerical output for surface area, average thickness, volume, and mean curvature for every single TTG and all TTG of one hemisphere together (Dalboni et al., 2020). The TTG as segmented by TASH were visually inspected and those that were fully located in the parietal extension of the PT (Honeycutt et al., 2000), or if the majority was located there, were removed. If the majority of it was located in the superior temporal gyrus, its corresponding volume and area was recalculated accordingly. Those regions mostly corresponded to the third or fourth regions selected by TASH. In two subjects, the first left hemisphere TTG selected by TASH extended too far into the anterior part of the temporal lobe, so they were removed. All TTG with their bottom borders on the part of the superior temporal sulcus which borders on the medial temporal gyrus were kept. This classification was performed by three classifiers. Discrepancies were resolved in a separate discussion.

The visual inspection of TASH output leaned towards removing right hemisphere gyri as seen in Table 1.

	Original		After visual inspection	
	Left hemisphere	Right hemisphere	Left hemisphere	Right hemisphere
mean number of TTG	2.88	2.66	2.49	1.98

Table 1: Original and adjusted after visual inspection mean number of TTG per hemisphere.

On the TASH output, the Multivariate Concavity Amplitude Index (MCAI) was calculated (Dalboni da Rocha et al., 2023). MCAI outputs a concavity score for every gyrus in four directions: anterior, posterior, medial and lateral. It was calculated for every gyrus separately. In this study, lateral MCAI values were used, since most sulci in CSD occur laterally (Dalboni da Rocha et al., 2023). From here on they will be meant when referring to MCAI scores. Total MCAI score per hemisphere was calculated by adding the sum of MCAI scores to the number of TTG according to TASH.

We furthermore calculated asymmetry indices of TTG volume, area, thickness, MCAI scores and number per hemisphere (TTG volume, area, MCAI scores and number per hemisphere were summed up and TTG thickness per hemisphere was averaged). The following formula was used to calculate an asymmetry index.

$$\frac{(Left - Right)}{(Left + Right)}$$

All statistical analyses were performed using R Statistical Software (v4.1.1; R Core Team, 2021).

3 Results

3.1 Language Aptitude

Total LLAMA scores ranged between 135 and 340 (*M* = 258.60, *SD* = 46.00). Total LLAMA scores summarize LLAMA_B (*M* = 64.45, *SD* = 20.89), LLAMA_D (*M* = 38.66, *SD* = 13.13), LLAMA_E (*M* = 87.20, *SD* = 20.08) and LLAMA_F scores (*M* = 68.29, *SD* = 23.08). In general, scores for subtests and total LLAMA were high (Figure 1). This is probably due to the high education level of the participants. As university students, they are already a sample selected for academic achievements which also involve language skills. This can be seen especially for LLAMA_E, the sound-symbol correspondence task, which shows a ceiling effect. LLAMA_B and LLAMA_F are bimodal, because participants were chosen for average and high scores in those tests. For this reason, in the subsequent analyses we focused on total LLAMA score as an average of different language learning subskills with subsequent exploration of the subtests.

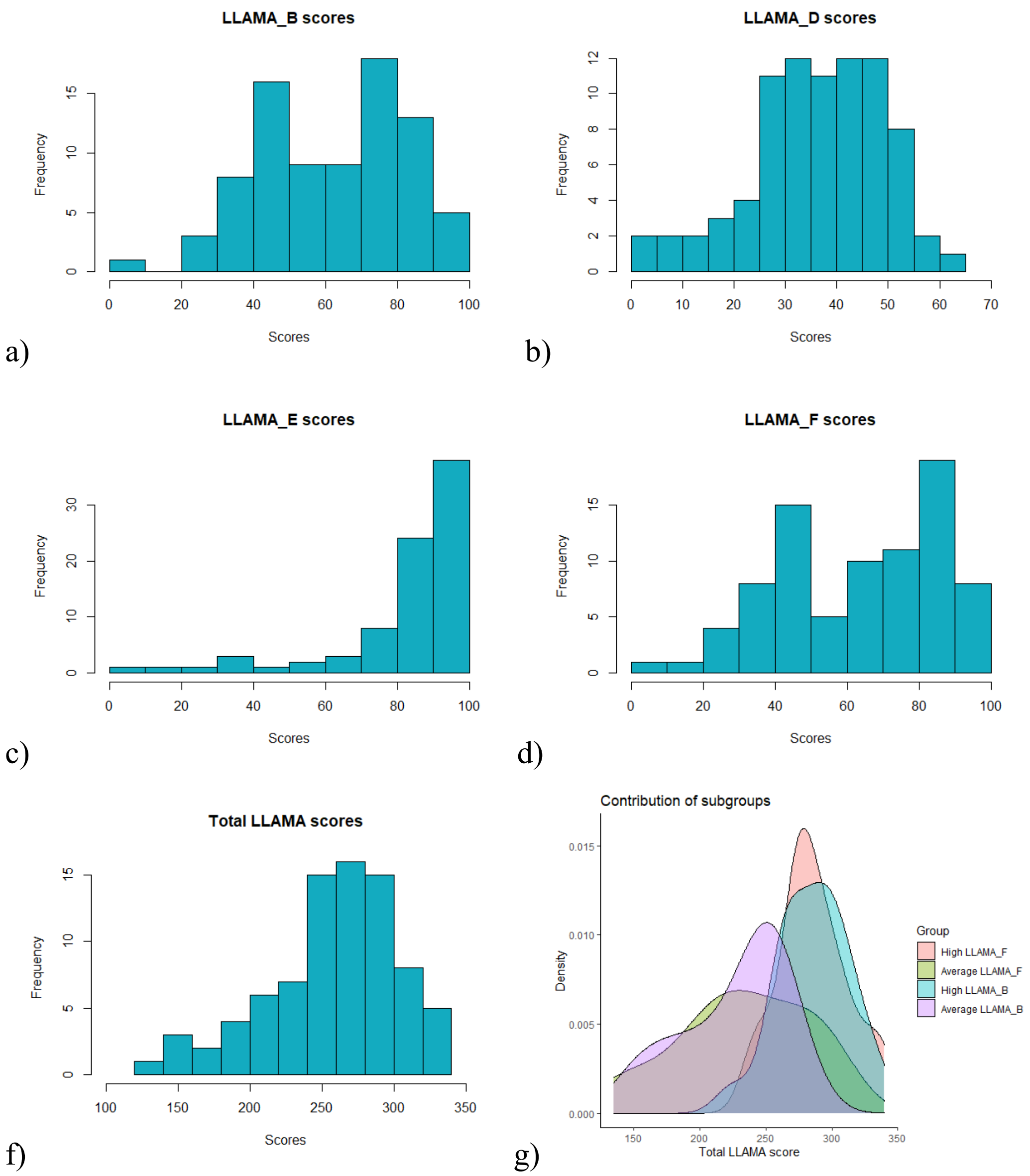


Figure 1: a-f) LLAMA subtests and total LLAMA score, g) density plot of subgroup contribution to total LLAMA score

Table 2 presents an overview of all tests done on the neuroanatomy of the auditory cortex.

Analysis	**Region**	**Measure(s)**	**Motivation**	**Results**
Mixed effect model of total LLAMA scores	All TTG	Volume Surface area Thickness	Exploratory first step	Negative correlation between total LLAMA score and second right hemisphere TTG volume (*t* = -2.11, *p* = .04) and surface area (*t* = -2.64, *p* = .009)
Linear models between TTG volume and total LLAMA scores	First and second, left and right TTG	Volume	To further confirm mixed effect model results	No significant correlation
Linear models between TTG surface area and total LLAMA score	First and second, left and right TTG	Surface area	To further confirm mixed effect model results	Significant positive correlation between total LLAMA score and surface area of first right TTG (*t* = 2.38, *p* = .02) and second left TTG (*t* = 2.22, *p* = .03)
Linear models between number of TTG and total LLAMA score	Left and right TTG	Number of TTG	To further confirm mixed effect model results	Significant negative correlation between total LLAMA score and right hemisphere number of TTG (*t* = -2.23, *p* = .03)
Mixed effect models between TTG measures and LLAMA subtests	First and second, left and right TTG	Surface area	To elucidate above total LLAMA score results	LLAMA_B, LLAMA_E and LLAMA_F are associated to the surface area of the second right hemisphere TTG
Linear models between MCAI scores and total LLAMA scores	Left and right first TTG, all TTG	MCAI scores	Test finding of relationship between right hemisphere TTG shape and language aptitude	Significant negative correlation between total right hemisphere MCAI score and total LLAMA score (*t* = -2.09, *p* = .04)
t-test on total LLAMA score between single HG and common stem duplicaton	First right TTG	MCAI scores	Test finding of increase in LLAMA score between single HG and common stem duplication	No significant effect
Linear models between asymmetry indices and total LLAMA score	All TTG	Volume, Surface area, thickness, TTG number and MCAI asymmetry indices	Exploration of asymmetry indices	No significant correlation

Table 2: Overview of neuroanatomy analyses. Dark colours indicate primary analyses, that were corrected for multiple testing, and the light colours indicate follow-up analyses

3.2 Mixed effect models of language aptitude and auditory cortex morphology

In an exploratory search, we used the R package lme4 (Bates et al., 2015) to perform three linear mixed effects analyses of the relationship between TTG volume, surface area, average thickness and language aptitude. As fixed effects, we entered age, gender, whole brain measurements and the interaction between total LLAMA scores, gyrus number and hemisphere. As random effects, we had intercepts for subjects. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Results are seen in Table 3.

		Volume	Area	Thickness
(Intercept)	β	366.587***	0.931***	-0.086**
	SE	(33.978)	(0.078)	(0.026)
Age	β	6.675	0.077*	0.002
	SE	(4.616)	(0.038)	(0.004)
Gender 1	β	-33.346	0.006	-0.074*
	SE	(47.737)	(0.103)	(0.033)
LLAMA_TOTAL	β	-0.580	-0.002	0.000
	SE	(0.692)	(0.002)	(0.001)
Gyrus 2	β	-516.765***	-1.363***	0.190***
	SE	(42.715)	(0.102)	(0.033)
Gyrus 3	β	-614.602***	-1.644***	0.147***
	SE	(54.801)	(0.130)	(0.042)
Hemisphere 1	β	-154.816***	-0.343***	-0.029
	SE	(42.428)	(0.101)	(0.032)
LLAMA_TOTAL : Gyrus 2	β	2.229*	0.007**	-0.001
	SE	(0.932)	(0.002)	(0.001)
LLAMA_TOTAL : Gyrus 3	β	0.705	0.002	0.001
	SE	(1.227)	(0.003)	(0.001)
LLAMA_TOTAL : Hemisphere 1	β	1.897*	0.006**	0.000
	SE	(0.928)	(0.002)	(0.001)
Gyrus 2 : hemisphere1	β	38.727	0.020	0.071
	SE	(62.173)	(0.148)	(0.048)
Gyrus 3 : hemisphere1	β	28.221	-0.026	0.245**
	SE	(103.002)	(0.242)	(0.079)
LLAMA_TOTAL : Gyrus 2 : Hemisphere 1	β	-2.851*	-0.008**	0.000
	SE	(1.350)	(0.003)	(0.001)
LLAMA_TOTAL : Gyrus 3 : Hemisphere 1	β	-1.959	-0.007	0.000
	SE	(2.027)	(0.005)	(0.002)
Cortex Volume	β	0.002***		
	SE	(0.000)		
Cortex White Surface Area	β		0.192***	
	SE		(0.044)	
Cortex Mean Thickness	β			1.746***
	SE			(0.177)
SD (Intercept id)		88.735	0.139	0.083
SD (Observations)		271.591	0.647	0.208
Num.Obs.		360	360	360
R^2 Marg.		0.489	0.569	0.426
R^2 Cond.		0.538	0.588	0.505
AIC		5042.4	835.4	75.3
BIC		5108.5	901.4	141.4
ICC		0.1	0.0	0.1
RMSE		256.35	0.62	0.19

. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 3: Results of linear mixed effect models of TTG volume, surface area and average thickness with total LLAMA scores

Benjamini–Hochberg correction for multiple comparisons (Benjamini & Hochberg, 1995) was used on all comparisons with total LLAMA score of this model, as well as on comparisons between total LLAMA score and auditory cortex shape and asymmetry. All comparisons of this model with surface area remained significant, while those with volume did not. Linear models between language aptitude and the volume of the first and second TTGs left and right were not significant (Table 1 and Figure 1 in the Appendix).

3.2.1 Language Aptitude and TTG surface area

In the mixed model between TTG area and total LLAMA scores, there was a significant interaction between total LLAMA score and the second TTG, total LLAMA score and the right hemisphere and total LLAMA score and the second right hemisphere TTG.

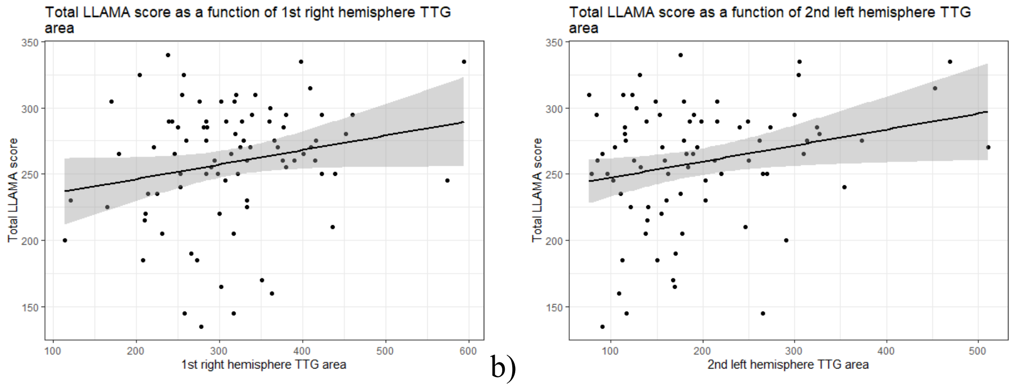
For this reason, we ran linear models between the first and second right and left hemisphere TTG area and total LLAMA score. Results are reported in Table 4.

		Total LLAMA score			
		1st TTG		2nd TTG	
		Left	Right	Left	Right
(Intercept)	β	293.262***	276.979***	254.686**	291.897**
	SE	(81.435)	(79.755)	(82.756)	(92.426)
Estimated Total Intracranial Volume	β	0.000	0.000	0.000	0.000
	SE	(0.000)	(0.000)	(0.000)	(0.000)
Gender	β	0.791	6.196	6.017	2.356
	SE	(13.890)	(13.776)	(13.900)	(15.927)
Age	β	-0.523	-1.585	-0.827	-0.738
	SE	(1.338)	(1.298)	(1.288)	(1.636)
Area	**β**	**-0.078**	**0.150***	**0.127***	**0.074**
	SE	**(0.063)**	**(0.063)**	**(0.057)**	**(0.087)**
Num.Obs.		82	82	80	66
R^2		0.028	0.077	0.069	0.023
R^2 Adj.		-0.023	0.029	0.020	-0.041
AIC		869.2	865.0	845.6	704.6
BIC		883.6	879.4	859.9	717.8
Log.Lik.		-428.597	-426.485	-416.798	-346.319
RMSE		45.05	43.91	44.30	45.99

Table 4: Results of linear models with total LLAMA score as the dependent variable and left and right first and second TTG surface area as the explanatory variable. The highlighted variable is the variable of interest.

The overall regression between the first right hemisphere TTG area and total LLAMA score was not significant (Adj. $R^2 = 0.03$, $F(4, 77) = 1.60$, $p = .183$) (Figure 2a). However, the variable of interest, first right hemisphere TTG area, had a statistically significant relationship with the total LLAMA score ($\beta = 0.15$, $SE = 0.06$, $t = 2.38$, $p = .020$).

The overall regression between the second left hemisphere TTG area was not significant (Adj. $R^2 = 0.02$, $F(4, 75) = 1.40$, $p = .242$) (Figure 2b). However, the variable of interest, second left hemisphere TTG area, had a statistically significant relationship with the total LLAMA score (β = 0.13, SE = 0.06, t = 2.22, p = .030).



a) b)

Figure 2: Scatterplots of total LLAMA score and a) first right hemisphere TTG area and b) second left hemisphere TTG

The other linear models were not significant (Figure 2 in the appendix).

3.2.2 Language Aptitude and TTG number

As significant effects of the mixed models involve also the entire right hemisphere, which is not captured by correlations with individual TTG, linear models with age and gender as covariates were used to determine the effect of the number of gyri on language aptitude. Results are reported in Table 5.

The linear model for the left hemisphere number of gyri was not significant (Adj. R^2 = 0.02, F(4, 77) = 0.25, p = .907) (Figure 4 in the appendix).

Also the overall regression of the right hemisphere number of gyri was not significant (Adj. R^2 = 0.02, F(4, 77) = 1.42, p = .235). However, total LLAMA scores had a statistically significant negative relationship with the right hemisphere number of gyri (β = -18.21, SE = 8.18, t = -2.23, p =.029) (Figure 3a). Figure 3b shows how different LLAMA scores are distributed over the number of right hemisphere TTG, showing that people with one TTG have more frequently a high LLAMA score.

		Total LLAMA score	
		Left hemisphere	Right hemisphere
(Intercept)	β	287.435**	349.035***
	SE	(85.764)	(82.138)
Estimated Total Intracranial Volume	β	0.000	0.000
	SE	(0.000)	(0.000)
Gender	β	0.976	-0.276
	SE	(14.109)	(13.575)
Age	β	-0.768	-1.407
	SE	(1.334)	(1.291)
Number of TTG	**β**	**4.680**	**-18.213***

	SE	(8.288)	(8.181)
Num.Obs.		82	82
R^2		0.013	0.069
R^2 Adj.		-0.038	0.020
AIC		870.4	865.7
BIC		884.9	880.1
Log.Lik.		-429.220	-426.832
RMSE		45.40	44.09

Table 5: Results of linear models with total LLAMA score as the dependent variable and left and right hemisphere TTG number as the explanatory variable. The highlighted variable is the variable of interest.

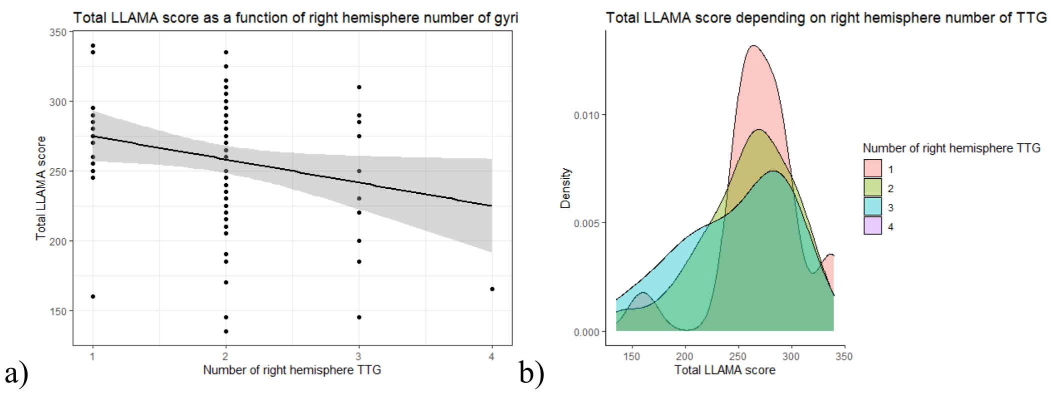


Figure 3: a) Scatterplot of total LLAMA score and number of right hemisphere TTG b) Density plot of total LLAMA score and right hemisphere number of TTG

3.2.3 LLAMA subtests and Auditory Cortex Morphology

To further explore the relationship between total LLAMA score and TTG surface area, we performed a linear mixed effect analyses of TTG surface area for every subtest individually. Table 6 shows which interactions were significant.

		LLAMA_B	LLAMA_D	LLAMA_E	LLAMA_F
(Intercept)	β	0.923***	0.927***	0.932***	0.930***
	SE	(0.079)	(0.080)	(0.079)	(0.078)
Age	β	0.069.	0.072.	0.077*	0.070.
	SE	(0.039)	(0.040)	(0.039)	(0.038)
Gender	β	0.034	0.009	-0.001	0.003
	SE	(0.106)	(0.107)	(0.104)	(0.103)
Surface Area	β	0.194***	0.186***	0.194***	0.178***
	SE	(0.045)	(0.046)	(0.045)	(0.045)
LLAMA_Subtest	β	-0.005	0.001	-0.004	-0.002
	SE	(0.004)	(0.006)	(0.004)	(0.003)
Gyrus 2	β	-1.361***	-1.355***	-1.359***	-1.358***
	SE	(0.103)	(0.104)	(0.103)	(0.103)
Gyrus 3	β	-1.643***	-1.641***	-1.644***	-1.642***
	SE	(0.131)	(0.133)	(0.132)	(0.131)
Hemisphere 1	β	-0.338**	-0.336**	-0.340***	-0.340***
	SE	(0.102)	(0.103)	(0.102)	(0.102)
LLAMA_Subtest : Gyrus 2	β	0.012*	-0.003	0.011*	0.009.
	SE	(0.005)	(0.008)	(0.005)	(0.004)
LLAMA_Subtest : Gyrus 3	β	0.008	-0.001	0.005	-0.001
	SE	(0.006)	(0.012)	(0.007)	(0.006)
LLAMA_Subtest : Hemisphere 1	β	0.009.	0.001	0.010*	0.008.

	SE	(0.005)	(0.008)	(0.005)	(0.004)
Gyrus 2 : Hemisphere 1	β	0.012	0.004	0.014	0.011
	SE	(0.149)	(0.151)	(0.149)	(0.149)
Gyrus 3 : Hemisphere 1	β	-0.013	-0.003	-0.004	-0.031
	SE	(0.238)	(0.241)	(0.245)	(0.252)
LLAMA_Subtest: Gyrus 2 : Hemisphere 1	β	-0.013.	-0.001	-0.014.	-0.013*
	SE	(0.007)	(0.011)	(0.007)	(0.006)
LLAMA_Subtest : Gyrus 3 : Hemisphere 1	β	-0.012	-0.013	-0.010	-0.007
		(0.010)	(0.020)	(0.011)	(0.011)
SD (Intercept id)		0.143	0.149	0.143	0.135
SD (Observations)		0.653	0.659	0.653	0.654
Num.Obs.		360	360	360	360
R^2 Marg.		0.560	0.550	0.560	0.561
R^2 Cond.		0.580	0.572	0.580	0.579
AIC		833.3	834.0	832.8	833.0
BIC		899.4	900.0	898.9	899.0
ICC		0.0	0.0	0.0	0.0
RMSE		0.63	0.63	0.63	0.63

. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 6: Mixed effect model of LLAMA subtest scores depending on surface area measures

3.3 Language Aptitude and Auditory Cortex Shape

Linear models with age and gender as covariates were used to determine the effect of MCAI scores on LLAMA scores. Results can be seen in Table 7.

		Total LLAMA score			
		1st TTG		All TTG	
		Left	Right	Left	Right
(Intercept)	β	271.457***	280.269***	255.176***	315.503***
	SE	(39.119)	(39.068)	(50.103)	(43.401)
Gender	β	3.187	3.314	4.222	2.788
	SE	(12.087)	(11.853)	(12.240)	(11.676)
Age	β	-0.769	-1.479	-0.670	-1.112
	SE	(1.286)	(1.351)	(1.299)	(1.252)
MCAI score	**β**	**-11.140**	**62.354**	**4.322**	**-17.224***
	SE	**(55.240)**	**(44.264)**	**(8.169)**	**(8.236)**
Num.Obs.		82	82	82	82
R^2		0.007	0.031	0.010	0.059
R^2 Adj.		-0.031	-0.006	-0.028	0.023
AIC		868.9	866.9	868.7	864.5
BIC		881.0	878.9	880.7	876.5
Log.Lik.		-429.463	-428.454	-429.337	-427.247
RMSE		45.53	44.98	45.46	44.32

Table 7: Results of linear models with total LLAMA score as the dependent variable and left and right first and total TTG MCAI score as the explanatory variable. The highlighted variable is the variable of interest.

The regression of right hemisphere first TTG MCAI on total LLAMA score was not significant (Adj. R^2 = -0.01, $F(3, 78)$ = 0.84, p = .477) (Figure 4a). The overall regression of right hemisphere all TTG MCAI on total LLAMA scores was also not significant (Adj. R^2 = 0.02, $F(3, 78)$ = 1.64, p = .187). However, the variable of interest, total right hemisphere MCAI score, had a statistically significant negative relationship with the total LLAMA score (β = -17.22, *SE* = 8.24, *t* = -2.09, *p* = 0.040) (Figure 4b).

This relationship is not significant after correction for multiple testing. It is however in accordance with the negative correlation of the number of right hemisphere TTG on total LLAMA score, as all TTG MCAI scores are strongly driven by the number of TTG.

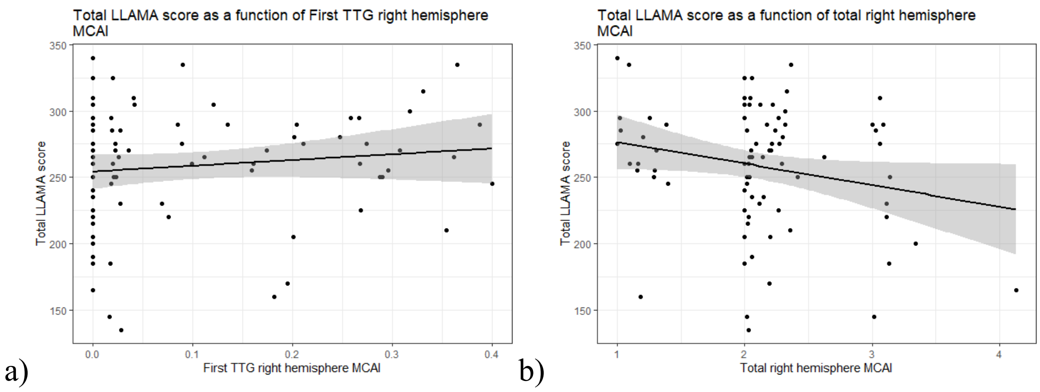


Figure 4: Scatterplot of total LLAMA score and a) first right hemisphere MCAI b) total right hemisphere MCAI

MCAI does not distinguish categorically between single HG and common stem duplication. To explore if this made a difference, after a visual inspection, MCAI scores below 0.12 were classified as a single gyrus and MCAI scores equal or above 0.12 were classified as common stem duplication. A *t*-test on total LLAMA scores between these groups was not significant (*t*(64.63) = 0.72, *p* = .474) (Figure 5).

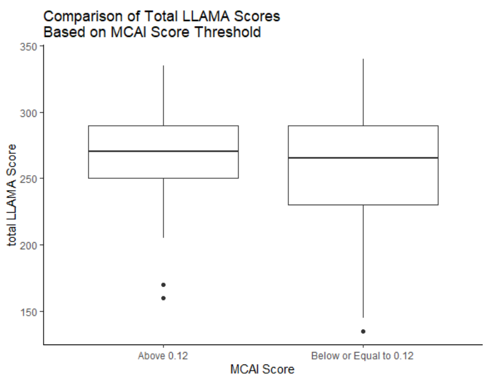


Figure 5: Boxplot of total LLAMA score depending on the shape of the HG (single gyrus or CSD determined by MCAI threshold)

3.4 Language Aptitude and Auditory Cortex Asymmetry

We ran linear models with age and gender as covariates between language aptitude and the asymmetry indices of auditory cortex measures with a previous significant correlation. Results are reported in Table 8.

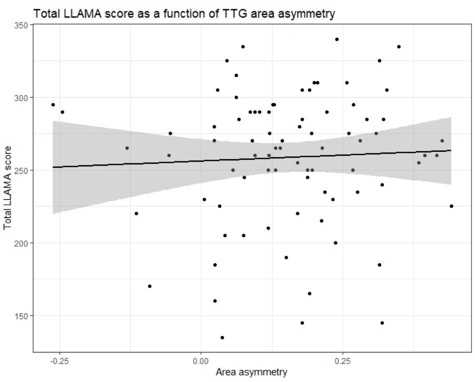
		Total LLAMA score		
		Area	Number of TTG	All TTG MCAI
(Intercept)	β	266.636***	255.836***	259.104***
	SE	(40.471)	(38.874)	(38.994)
Gender	β	3.934	7.603	6.072
	SE	(12.175)	(11.946)	(11.901)
Age	β	-0.774	-0.651	-0.714
	SE	(1.277)	(1.257)	(1.254)
Asymmetry index	**β**	**18.165**	**44.883.**	**46.556.**
	SE	**(37.585)**	**(24.962)**	**(26.292)**
Num.Obs.		82	80	82
R^2		0.010	0.046	0.045
R^2 Adj.		-0.029	0.009	0.008
AIC		868.7	844.5	865.7
BIC		880.8	856.4	877.8
Log.Lik.		-429.361	-417.237	-427.868
RMSE		45.48	44.55	44.65

Table 8: Results of linear models with total LLAMA score as the dependent variable and asymmetry measures of TTG surface area, number and total MCAI score as the explanatory variable. The highlighted variable is the variable of interest.

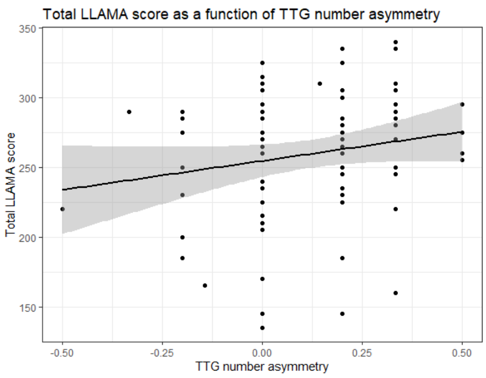
The linear model for asymmetry of surface area was not significant before correction (Adj. R^2 = -0.03, $F(3, 78) = 0.25$, $p = .861$) (Figure 6a).

The linear model for asymmetry of number of gyri was not significant before correction (Adj. R^2 = 0.01, $F(3, 76) = 1.23$, $p = .305$). However, there was a tendency towards a relationship with the value of interest ($\beta = 44.88$, $SE = 24.96$, $t = 1.80$, $p = .076$) (Figure 6b).

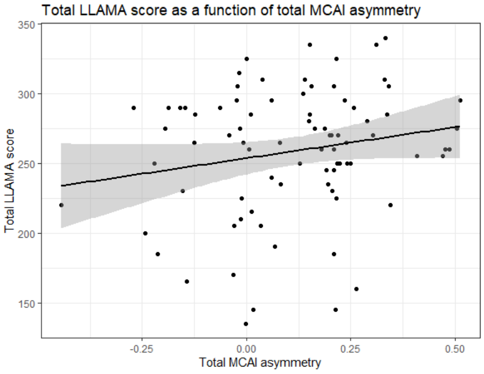
The overall regression of total MCAI asymmetry on total LLAMA scores was not significant before correction (Adj. $R^2 = 0.01$, $F(3, 78) = 1.23$, $p = .307$). However, the variable of interest showed a tendency towards a relationship with total LLAMA score ($\beta = 46.56$, $SE = 26.29$, $t = 1.77$, $p = .081$) (Figure 6c).



a)



b)



c)

Figure 6: Scatterplot of total LLAMA score and a) TTG area b) TTG number asymmetry c) total MCAI asymmetry

3.5 Language Aptitude and Related Measures

Table 9 presents an overview of all analyses of language aptitude with related behavioral measures.

Analysis	**Region**	**Measure(s)**	**Motivation**	**Results**
Linear model between total LLAMA score and working memory	Behavioral	Ospan Partial Score	Known relationship between language aptitude and working memory	Significant positive correlation between working memory and total LLAMA score ($F(3,56)$= 3.98, p = .01)
Linear models between LLAMA subtests and working memory	Behavioral	Ospan Partial Score	To further explore above result	Significant positive correlation between working memory and LLAMA_F ($F(3, 56) = 3.30$, p = .03)
Mixed effect model of partial Ospan scores	All TTG	Volume Surface area Thickness	To explore the relationship between working memory and the auditory cortex	No significant correlation
Linear model between total LLAMA score and language experience measures	Behavioral	Number of languages, Language entropy	To explore the relationship between language experience and language aptitude	Significant positive correlation between total LLAMA score and number of languages ($F(3, 75) = 4.06$, p = 0.010) and language entropy ($F(3, 75) = 4.41$, p = .017)
Mixed effect model of language entropy	All TTG	Volume Surface area Thickness	Test finding of relationship between language entropy and second left hemisphere TTG thickness	No significant correlation
t-tests on total LLAMA score between musicality categories	Behavioral	Absolute musical experience, active musical involvement	Known relationship between language aptitude and musicality	No significant effect of musicality on language aptitude
Linear models between number of instruments and total LLAMA score and LLAMA subtests	Behavioral	Number of instruments	Test finding of relationship between number of instruments and language aptitude	No significant correlation between instrument number and language aptitude

Anova on TTG measures dependent on active musical involvement	Left and right TTG	Volume, Surface Area, Thickness	Test finding of relationship between musical experience and auditory cortex measures	No significant effect of active musical involvement on auditory cortex morphology

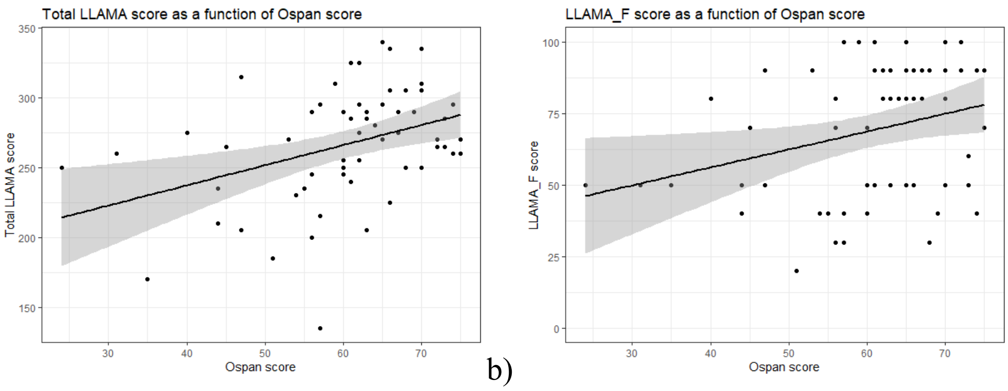
Table 9: Overview of related aptitude analyses. Dark colours indicate primary analyses, that were corrected for multiple testing, and the light colours indicate follow-up analyses.

3.5.1 Working Memory

A linear model was used to determine the effect of partial OSPAN score on total LLAMA score. The regression was statistically significant before correction (Adj. $R^2 = 0.13$, $F(3, 56) = 3.98$, $p = .012$) (Figure 7a), but not if all analyses on behavioral tests were corrected for multiple comparison. To explore the effect, we ran linear models between partial OSPAN score and LLAMA subtests. Of all subtests, only LLAMA_F was statistically significant (Adj. $R^2 = 0.11$, $F(3, 56) = 3.30$, $p = .027$) (Figure 7b, Figure 5 in the appendix). Results can be seen in Table 10.

		LLAMA_TOTAL	LLAMA_B	LLAMA_D	LLAMA_E	LLAMA_F
(Intercept)	β	234.671***	32.946	40.936*	106.963***	53.827.
	SE	(48.313)	(22.327)	(17.376)	(22.555)	(27.541)
Gender	β	-13.893	10.328.	-4.345	-6.630	-13.246.
	SE	(11.834)	(5.469)	(4.256)	(5.525)	(6.746)
Age	β	-1.278	0.179	-0.460	-1.003.	0.007
	SE	(1.168)	(0.540)	(0.420)	(0.545)	(0.666)
Ospan Partial Score	**β**	**1.405****	**0.233**	**0.272**	**0.270**	**0.631***
	SE	**(0.460)**	**(0.212)**	**(0.165)**	**(0.215)**	**(0.262)**
Num.Obs.		60	60	60	60	60
R^2		0.176	0.079	0.078	0.095	0.150
R^2 Adj.		0.131	0.030	0.029	0.047	0.105
AIC		613.2	520.6	490.5	521.8	545.8
BIC		623.7	531.0	501.0	532.3	556.2
Log.Lik.		-301.602	-255.286	-240.243	-255.898	-267.881
RMSE		36.88	17.04	13.26	17.22	21.03
. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001						

Table 10: Results of the linear models of partial OSPAN scores with LLAMA scores before correction. The highlighted variable is the variable of interest.



a) b)

Figure 7: Scatterplot of partial OSPAN score and a) total LLAMA score and b) LLAMA_F score

In an exploratory search, three mixed models were run for the relationship between TTG volume, surface area, average thickness and partial OSPAN score. As fixed effects, we entered age, gender, whole brain measurements and the interaction between partial OSPAN score, gyrus number and hemisphere. As random effects, we had intercepts for subjects. No interaction with partial OSPAN score was significant (Table 11). However, while visual inspection of residual plots for average thickness did not reveal any obvious deviations from homoscedasticity or normality, for volume and surface area residuals formed two distinct clouds, probably related to the high and average LLAMA_F groups.

		Volume	Area	Thickness
(Intercept)	β	353.308***	0.943***	-0.103***
	SE	(42.185)	(0.097)	(0.030)
Age	β	6.482	0.066	0.007.
	SE	(5.491)	(0.046)	(0.004)
Gender 1	β	7.288	0.104	-0.107**
	SE	(62.168)	(0.134)	(0.037)
Ospan Partial Score	β	2.190	0.004	0.002
	SE	(3.642)	(0.008)	(0.003)
Gyrus 2	β	-485.836***	-1.338***	0.213***
	SE	(52.058)	(0.124)	(0.039)
Gyrus 3	β	-618.063***	-1.690***	0.126*
	SE	(70.170)	(0.167)	(0.052)
Hemisphere 1	β	-125.813*	-0.284*	-0.022
	SE	(51.815)	(0.124)	(0.038)
Ospan Partial Score : Gyrus 2	β	-1.987	-0.001	-0.003
	SE	(4.817)	(0.012)	(0.004)
Ospan Partial Score : Gyrus 3	β	-3.540	-0.007	-0.002
	SE	(5.613)	(0.013)	(0.004)
Ospan Partial Score : Hemisphere 1	β	-4.391	-0.004	-0.006
	SE	(4.815)	(0.012)	(0.004)
Gyrus 2 : Hemisphere 1	β	-9.417	-0.073	0.039
	SE	(76.474)	(0.183)	(0.057)
Gyrus 3 : Hemisphere 1	β	23.016	-0.041	0.252*
	SE	(135.998)	(0.321)	(0.100)
Ospan Partial Score : Gyrus 2 : Hemisphere 1	β	1.761	0.000	0.005
	SE	(7.015)	(0.017)	(0.005)
Ospan Partial Score : Gyrus 3 : Hemisphere 1	β	-0.855	-0.014	0.019
	SE	(16.333)	(0.039)	(0.012)
Cortex Volume	β	0.002**		
	SE	(0.001)		
Cortex White Surface Area	β		0.167**	
	SE		(0.054)	
Cortex Mean Thickness	β			1.839***
	SE			(0.195)
SD (Intercept id)		105.823	0.181	0.063
SD (Observations)		283.801	0.678	0.210
Num.Obs.		257	257	257
R^2 Marg.		0.451	0.544	0.444
R^2 Cond.		0.518	0.574	0.490
AIC		3589.8	633.1	67.1
BIC		3650.2	693.4	127.4

ICC	0.1	0.1	0.1
RMSE	263.64	0.64	0.20

. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 11: Results of linear mixed effect models of TTG volume, surface area and average thickness with partial OSPAN scores

3.5.2 Multilingual experience

The number of foreign languages reported to be known by participants ranged from one to five, with most participants having learned three (*M* = 2.75, *SD* = 1.07).

A linear model with age and gender as covariates was used to determine the effect of number of languages learnt in life on language aptitude. The overall regression was statistically significant before correction (Adj. $R^2 = 0.11$, $F(3, 75) = 4.06$, $p = .010$) (Figure 9a), but lost significance after correction of behavioral tests for multiple comparisons.

Furthermore, a linear model with age and gender as covariates was used to determine the effect of language entropy on language aptitude. The overall regression was also statistically significant (Adj. $R^2 = 0.09$, $F(3, 75) = 4.41$, $p = .017$) (Figure 9b) and lost it after correction.

Results can be seen in Table 12.

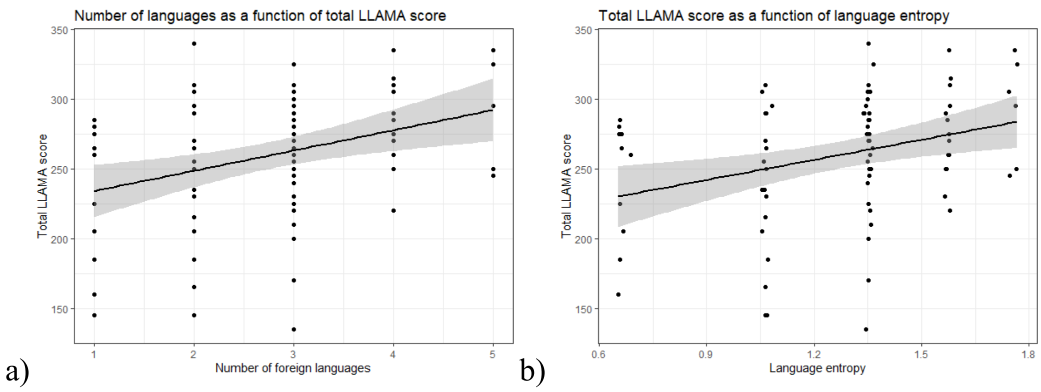


Figure 9: Linear model of a) number of languages and b) language entropy as a function of total LLAMA score

		Total LLAMA score	
		Number of languages	Language entropy
(Intercept)	β	212.497***	187.506***
	SE	(40.110)	(44.559)
Gender	β	16.546	15.631
	SE	(11.717)	(11.819)
Age	β	-1.380	-1.287
	SE	(1.218)	(1.228)
Measure	**β**	**17.650*** **	**57.863*** **
	SE	**(4.784)**	**(16.804)**
Num.Obs.		80	80
R^2		0.160	0.143
R^2 Adj.		0.126	0.109
AIC		834.4	835.9
BIC		846.3	847.9

Log.Lik.	-412.180	-412.971
RMSE	41.82	42.23
. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001		

Table 12: Results of linear models with total LLAMA score as the dependent variable and number of languages and language entropy as the explanatory variables.

We performed three linear mixed effects analyses on the relationship between TTG volume, surface area, average thickness and language entropy. As fixed effects, we entered age, gender, whole brain measurements and the interaction between language entropy, gyrus number and hemisphere. As random effects, we had intercepts for subjects. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. No interaction with language entropy was significant (Table 13).

		Volume	Area	Thickness
(Intercept)	β	370.393***	0.937***	-0.083**
	SE	(34.082)	(0.079)	(0.026)
Age	β	5.743	0.071.	0.002
	SE	(4.591)	(0.038)	(0.004)
Gender 1	β	-49.549	-0.026	-0.085*
	SE	(48.313)	(0.105)	(0.034)
Language Entropy	β	94.219	0.061	0.009
	SE	(89.879)	(0.073)	(0.069)
Gyrus 2	β	-513.935***	-1.354***	0.189***
	SE	(42.876)	(0.103)	(0.033)
Gyrus 3	β	-612.649***	-1.640***	0.149***
	SE	(54.947)	(0.131)	(0.042)
Hemisphere 1	β	-152.260***	-0.335**	-0.029
	SE	(42.587)	(0.102)	(0.032)
Language Entropy : Gyrus 2	β	-17.944	0.003	-0.019
	SE	(119.013)	(0.100)	(0.090)
Language Entropy : Gyrus 3	β	-289.838.	-0.243.	0.079
	SE	(157.813)	(0.132)	(0.120)
Language Entropy : Hemisphere 1	β	95.252	0.078	0.071
	SE	(118.563)	(0.100)	(0.090)
Gyrus 2 : Hemisphere 1	β	29.480	-0.006	0.074
	SE	(62.482)	(0.150)	(0.047)
Gyrus 3 : Hemisphere 1	β	21.963	-0.046	0.265***
	SE	(101.956)	(0.242)	(0.078)
Language Entropy : Gyrus 2 : Hemisphere 1	β	-58.163	-0.049	-0.058
	SE	(178.673)	(0.150)	(0.136)
Language Entropy : Gyrus 3 : Hemisphere 1	β	87.433	-0.013	0.449.
	SE	(330.191)	(0.274)	(0.252)
Cortex Volume	β	0.002***		
	SE	(0.000)		
Cortex White Surface Area	β		0.179***	
	SE		(0.045)	
Cortex Mean Thickness	β			1.727***
	SE			(0.174)
SD (Intercept id)		86.935	0.134	0.079
SD (Observations)		272.668	0.654	0.207
Num.Obs.		360	360	360
R^2 Marg.		0.487	0.560	0.435
R^2 Cond.		0.534	0.578	0.507

AIC	4984.7	795.4	11.4
BIC	5050.8	861.5	77.5
ICC	0.1	0.0	0.1
RMSE	257.69	0.63	0.19

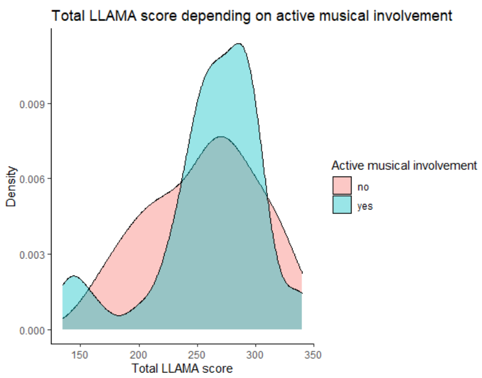
. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

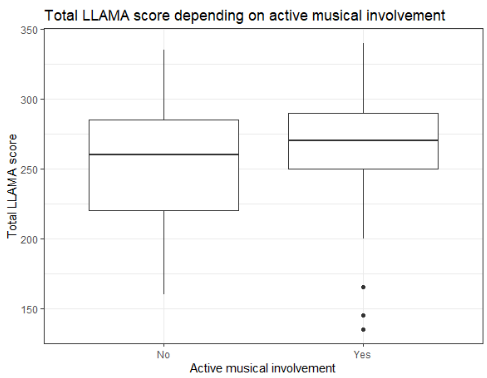
Table 13: Results of linear mixed effect models of TTG volume, surface area and average thickness with language entropy

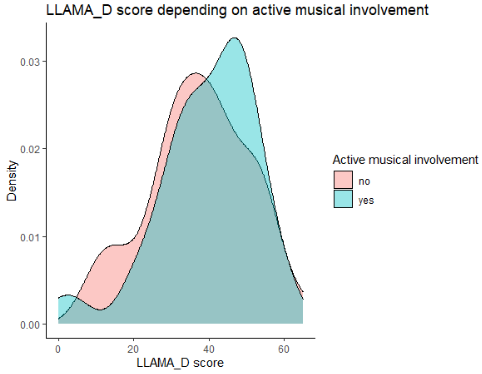
3.5.3 Musicality

37.8% of participants reported to neither sing nor play an instrument. A *t*-test revealed, that participants who neither sung nor played an instrument (*M* = 256.94, *SD* = 45.10) and those who did (*M* = 259.61, *SD* = 46.91), did not differ in their total LLAMA score, *t*(80) = -0.25; *p* = .8 (before correction).

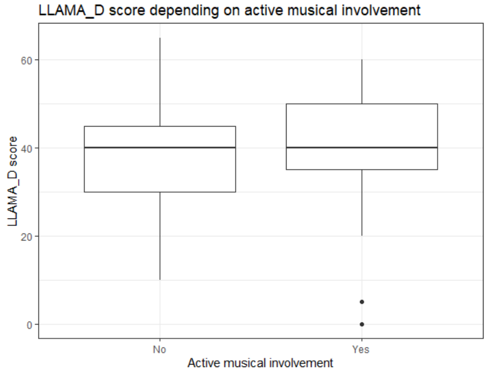
According to the theory that aptitude drives experience seeking (Golestani et al., 2011; Seither-Preisler et al., 2014; Turker et al., 2019) we chose from the group of participants who sung or played an instrument, those that actively sought out musical involvement. This included participants who reported to receive music lessons, play or sing in an orchestra, ensemble, band or choir or did so in the past. Those that had received lessons in the past and only sung or played for themselves were excluded, because it did not provide evidence of active niche construction. By these criteria 54.9% sought out musical experiences and 45.1% did not. A density plot reveals that those who did (*M* = 261.33, *SD* = 45.64), had a tendency for higher total LLAMA scores, while those who did not (*M* = 255.27, *SD* = 46.78) had more widely distributed scores (Figure 10a). However, the difference was not statistically significant in a *t*-test, *t*(80) = -0.59; *p* = .6 (Figure 10b). We tested if LLAMA_D, the subtest for phonetic memory, was influenced by active musical involvement. However, the difference was not statistically significant in a *t*-test, *t*(80) = -0.77; *p* = .4 (before correction) (Figure 10d).

a) 

b) 



c)



d)

Figure 10: a,c) Density plot and b,d) Boxplot of LLAMA score depending on musical activities

Turker et al. (2017) found a correlation between number of instruments and LLAMA score. Here, we found no significant relationship between the number of instruments and LLAMA scores in a model with age and gender as covariates before correction. See Table 14 and Figure 11.

Model		LLAMA_TOTAL	LLAMA_B	LLAMA_D	LLAMA_E	LLAMA_F
(Intercept)	β	272.126***	35.807.	56.730***	98.462***	81.127***
	SE	(40.775)	(17.997)	(11.525)	(17.690)	(20.561)
Gender	β	3.393	13.360*	-4.547	-0.301	-5.119
	SE	(12.175)	(5.374)	(3.441)	(5.282)	(6.139)
Age	β	-1.007	0.219	-0.472	-0.566	-0.189
	SE	(1.313)	(0.579)	(0.371)	(0.569)	(0.662)
Instrument number	**β**	**3.597**	**0.681**	**0.320**	**2.264**	**0.331**
	SE	**(4.800)**	**(2.119)**	**(1.357)**	**(2.083)**	**(2.421)**
Num.Obs.		82	82	82	82	82
R^2		0.018	0.074	0.038	0.031	0.010
R^2 Adj.		-0.019	0.038	0.001	-0.006	-0.028
AIC		868.0	733.9	660.8	731.0	755.7
BIC		880.0	745.9	672.8	743.1	767.7
Log.Lik.		-428.990	-361.925	-325.383	-360.513	-372.847
RMSE		45.27	19.98	12.80	19.64	22.83
. p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001						

Table 14: Results of the linear models of instrument number with LLAMA scores. The highlighted variable is the variable of interest.

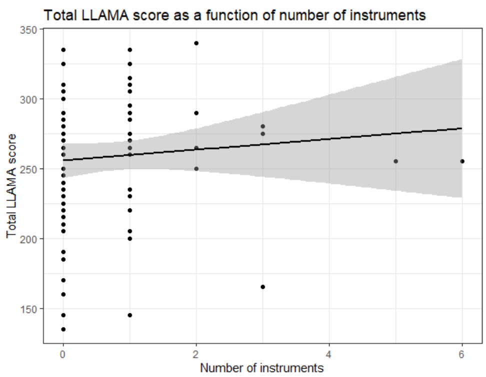


Figure 11: Scatterplot of total LLAMA score and the number of instruments

Benner et al. (2017) found a significant difference in right hemisphere HG volume between non-musicians and amateur musicians. Also Dalboni da Rocha et al. (2020) found this using TASH. However, their criteria for amateur musicians were much stricter than ours. Benner et al.'s (2017) criteria included musical training intensity of 17.7 ± 2.2 weekly hours averaged over the past 3 years and a minimum of 5 years of instrumental practice beyond standard school education.

For the participants of this study, a boxplot reveals that those with no active musical involvement have a slightly lower mean volume in comparison to those with it (Figure 12a). A mixed ANOVA on first TTG volume, with musicianship status as between-subjects factors and hemisphere as within-subject factor revealed no statistically significant effect on musicianship status ($F(1,80) = 2.40$, $p = .125$). Exclusion of the outliers did not affect the result.

The same is true for first TTG area and average thickness (Figure 12bc). Their tendency is in accordance with the findings of Benner at al. (2017) and Dalboni et al. (2020), but the effect is not strong enough to be significant. The values for the mixed ANOVA for HG area were $F(1,80) = 0.764$, $p = .385$ and for average thickness $F(1,80) = 1.458$, $p = .231$. Exclusion of the outliers did not affect the results.

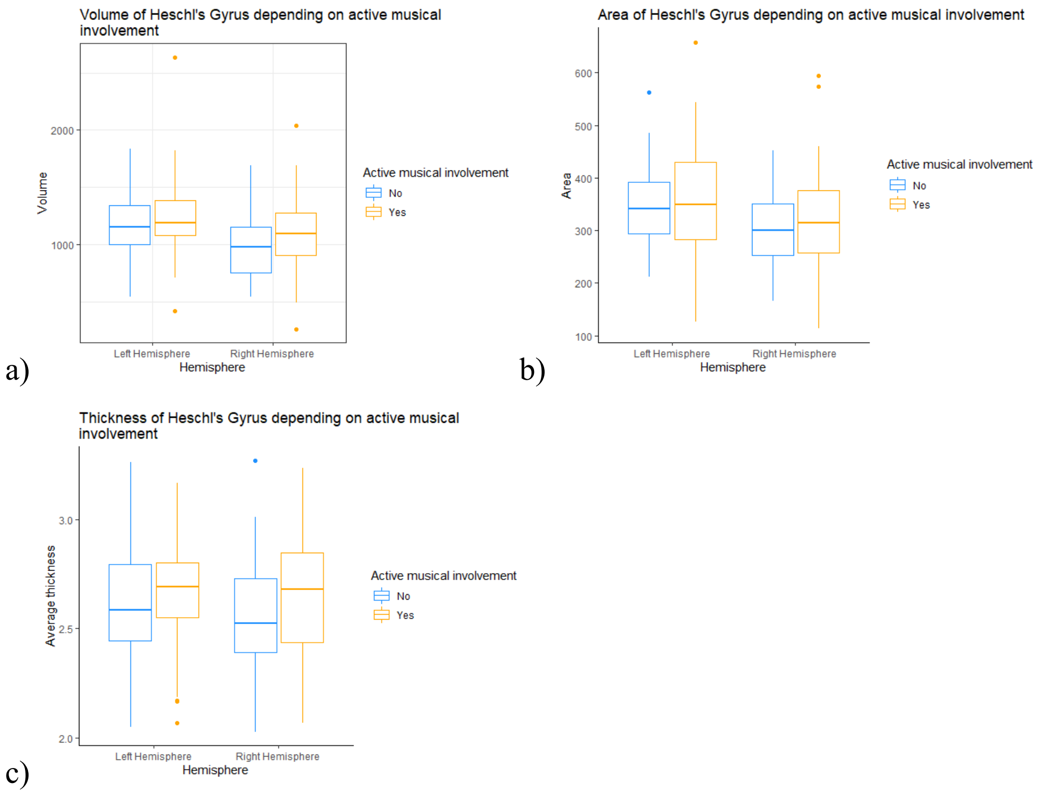


Figure 12: a) Volume b) Area and c) Thickness of the first TTG depending on musical involvement

4 Discussion

In this study, previous findings on the relationships between (1) the morphology and shape of auditory cortex TTGs and language aptitude, (2) multilingual experience, (3) working memory and (4) musicality were tested in a new dataset. It was found that higher language aptitude was related to less TTGs in the right auditory cortex as well as bigger surface area in the first right and second left TTG.

4.1 Language Aptitude and Auditory Cortex Neuroanatomy

The main question was about the relationship between neuroanatomy and language aptitude as measured by four subtests of the LLAMA test. We used both the total LLAMA score as the main measure, as it gives a mean between specific language learning skills, as well as individual tests in the follow-up analyses.

Hypothesis (1) stated that participants with higher language aptitude would have more complex shape in the right auditory cortex. However, we found the opposite: subjects with high language aptitude had less complex shape in the right auditory cortex. This is in line with our second hypothesis, that language aptitude is influenced not by the shape of the first TTG alone, but by all TTG. The negative relationship between language aptitude and TTG shape (in the right hemisphere) was indeed significant only for all TTGs and not for the first TTG.

This relationship was also found when investigating the number of TTG, a measure related to the shape of all TTG per hemisphere. The number of TTG drives the total MCAI score, our measure for shape, and it is also significantly negatively related to language aptitude.

We hypothesized that we would find a positive relationship between language aptitude and more complex TTG shape due to Turker et al. (2019), who found a positive relationship between language aptitude and TTG number and volume. However, our results are rather in line with previous studies, supporting left lateralization of language skills in the auditory cortex: Phoneticians are more likely to have multiple or split TTGs in the left auditory cortex (Golestani et al., 2011) and dyslexic boys are more likely to have more than one TTG in the right auditory cortex in comparison to controls (Altarelli et al., 2014).

However, direct comparisons with either study may be difficult, as there may be a difference in the distribution of TTG number between studies that performed visual inspection and studies that use TASH. While Altarelli et al. (2014) had a similar sample size as this study, only five participants had two TTGs in the right hemisphere. Marie et al. (2015) report in their study with a sample size of 430, that the most common combination is bilateral single TTG, with single left and two right TTGs being second most common (while not investigating additional TTGs). Also the data of Turker et al. (2019) follows this distribution, while other studies using TASH follow our distribution of more TTGs on the left and most participants having at least two gyri bilaterally (n=650, unpublished data, Arato 2023). Reasons for the difference may be that in visual inspection shallow gyri are not picked up or no additional TTG are expected (or even investigated, as in Marie et al. 2015).

Regarding the shape of the first TTG, Turker et al. (2019) found already higher language aptitude scores in participants with a common stem duplication in the first right TTG (in comparison to a single gyrus). In our data, we did not find a relationship between the shape of the first right TTG and language aptitude, even though there was a positive correlation between the first right TTG surface area and language aptitude. As MCAI ranges per gyrus continuously from 0 to 0.4, it does not categorize between single HG and common stem duplication. However, even when MCAI scores were binarized into single HG and common stem duplication, higher language aptitude for duplicated HG could not be replicated.

That language aptitude is positively correlated with surface area of the first TTG and negatively with the number of TTG in the right hemisphere might be explained by the difference between the first TTG, which houses the primary auditory cortex, and additional TTG, which are structurally part of the PT and functionally part of the secondary auditory cortex. Altarelli et al. (2014) found that only when additional TTG were included was the size asymmetry between left and right PT in dyslexia significant. When both dyslexia and language aptitude are seen on a spectrum spanning from a disability to extraordinary ability, it could be understood, that a higher predisposition towards dyslexia could be negatively related to language aptitude, because more TTGs in the right PT may negatively impact optimal secondary auditory processing of language. The exact mechanism behind such explanation is still an open question, though it could be speculated that there is a difference in microstructure underlying continuous surfaces and surfaces that are split by sulci (as when there are multiple TTG) and that processing in the continuous surface of a big first right TTG could be beneficial for language learning and processing in split surfaces could be disadvantageous.

In addition to the positive correlation between language aptitude and the first right TTG, there was also a positive correlation between language aptitude and the surface area of the second left TTG. This is in accordance with left-lateralized processing of language, which may benefit language learning. While previously correlations have been found between phonetic coding and surface area of the first left TTG (Golestani et al., 2011), our more general test for language aptitude may draw more on secondary auditory cortex.

We found no evidence for a significant correlation of language aptitude with asymmetry of TTG surface area, number or shape. The reason why we find relationships between language aptitude and the right hemisphere instead of the left hemisphere or the asymmetry between left and right hemisphere might be because the right hemisphere is more variable and its variability has a bigger influence on language aptitude in this sample of average to high language learners (see Dalboni da Rocha et al., 2020; Penhune et al., 1996; Westbury et al., 1999 for accounts of greater morphological variability in right HG and STG).

When exploring the relation of language aptitude with surface area, all underlying cognitive abilities related to surface area in a similar way as general language aptitude, except for the test for phonetic memory (LLAMA_D). It is surprising that of all the tests, the phonetic memory test does not relate to surface area of the auditory cortex. It may be that LLAMA_D, while testing the memory of unfamiliar words, does not draw as much on phonetic processing as on memory. Subsequently, we tested the relationship between language aptitude and working memory.

4.2 Language Aptitude and Working Memory

In accordance with the literature (Baddeley, 2003; Linck et al., 2014; Z. Wen & Skehan, 2011), we hypothesized (2) that individuals with higher working memory would also have higher language aptitude. Indeed, we found a positive correlation between language aptitude and working memory.

First and second language acquisition have been connected to phonological and executive working memory (Wen, 2015). The phonological loop supports vocabulary acquisition in the first and second language and possibly long-term grammar acquisition (Baddeley, 2003; Baddeley et al., 1998; Duyck et al., 2003). Polyglots indeed show expanded phonological loop capacity, while no other comparable cognitive task proved different to controls (Papagno & Vallar, 1995). However, the domain-general operation span task draws more on the executive function of working memory, which is implicated in conscious monitoring of language processing activities and noticing corrective feedback (Wen, 2015). This is corroborated by our study, where out of all LLAMA subtests only LLAMA_F, the test for grammatical inferencing, was related to the partial OSPAN measures.

Studies on executive control function of working memory indicate a crucial role of the prefrontal cortex, though also distributed networks including sensory systems and sub-cortical areas are involved (Mansouri et al., 2015). We did not find a correlation between executive working memory functions and auditory cortex morphology, thus while executive working memory might draw on auditory processing, this might not be reflected by auditory cortex anatomy.

4.3 Language Aptitude and Multilingual Experience

On the basis that language aptitude should be an innate trait, we hypothesized (3) that there would be no correlation between the number of languages learned in life and language aptitude. However, we did find a positive relationship between the number of languages learned in life and language aptitude. Previously, Turker et al. (2017) reported no correlation between the number of languages spoken by adults and their language aptitude, while Turker et al. (2019) found a correlation between the number of languages learned by the children and their language aptitude.

The effect of learned languages on language aptitude can either be due to participants scoring higher who learned more languages in their life, or participants who have high language aptitude seeking out more opportunities to use this skill. Though intuitively it makes sense that people who learned many languages are better at language learning, also the second option seems intriguing as these are Dutch people, who besides English are under no pressure of learning more languages.

We hypothesized, that there would be a relationship between auditory cortex morphology and languages learned, because Kepinska et al. (*in preparation*) found a relationship between language entropy and the thickness of the second TTG. We could also not corroborate this hypothesis, as we did not find a relationship between participants' language experience measure and their auditory cortex thickness or any other measure. This could be, because in our sample early bilinguals were excluded, so nobody had to integrate two or more languages in their still highly plastic brain under the age of four. Another reason could be that our relatively homogenous Dutch university-student participants had not experienced as much language diversity as the sample of Kepinska et al. (*in preparation*).

4.4 Language Aptitude and Musicality

We hypothesized (4), that higher musicality would correlate with higher language aptitude, due to previous findings in the literature: Milovanov and Tervaniemi (2011) reviewed findings of how individuals with higher musical aptitude were better in phonetic discrimination and second language pronunciation tasks, which is supported by the findings of Vangehuchten et al. (2015). Christiner and Reiterer (2013) showed a correlation between singing abilities and speech imitation and Turker et al. (2017) showed a correlation between number of instruments played and language aptitude and between a musicality test and speech imitation ability. Turker et al. (2021) proposed that the correlation between musical skills and language learning skills are due to shared neuroanatomical basis. However, we could not find an effect of active musical involvement or instrument number on language aptitude or phonetic memory. This can be explained with the fact that here only general self-reported measures were collected and there was low variance in participant's musicality. Also, previous findings related musicality with specific phonetic tasks like pronunciation, speech imitation and discrimination of phonemic minimal pair contrasts. This is not tested in LLAMA_D, our test for phonetic memory, or any other of the LLAMA subtests. Furthermore, in the study of Christiner and Reiterer (2013) singing abilities explained much less variance in speech imitation tasks than working memory. Thus, the effect of musicality on language aptitude is probably only small and limited to phonetic abilities, which cannot be captured well with the LLAMA test. While this explains why no effect can be found, there is a tendency towards higher language aptitude in participants who actively engage in musical activities. This tendency is not as strong for phonetic memory, as it is for general language aptitude, which may support that LLAMA_D does not test for phonetic abilities well and there may be other factors, that both contribute to language aptitude and musicality, like working memory.

It has been established, that musical aptitude correlates with the volume of anteromedial HG (Schneider et al., 2002), overall volume bilaterally (Schneider et al., 2005) and volume of right HG in school children (Seither-Preisler et al., 2014). This is why we hypothesized that participants with active musical involvement would exhibit higher volumes of HG. While our results do not support this, they also do not contradict it. Our university-student participants with active musical involvement had bilaterally a higher mean for volume, surface area and thickness of HG, even though the difference was not significant. This fits with the notion of musicality and linked auditory cortex structure as being on a spectrum, on which only expert populations (who were absent from the present sample) differ significantly from the general population.

4.5 Summary and Outlook

What constitutes a talent for language learning? According to our results, what distinguishes high language aptitude is less TTGs in the right hemisphere, high executive working memory function and many languages learned in life. Furthermore, also higher surface area in the first right and second left TTG are beneficial.

Future studies with TASH and MCAI should provide large datasets, with which diverse populations can be studied and compared, to more finely resolve relationships between the anatomy of the auditory cortex and behavioral measures. As language aptitude can be measured relatively easy with the LLAMA test, it could be included as a behavioral measure in large MRI studies, to refine our understanding of the relationship between total LLAMA scores, LLAMA subtests and auditory cortex anatomy in populations with high variability. To answer the question if language aptitude is indeed relatively stable or can be improved by language learning, studies on a diverse sets of language learners are necessary. In countries with a large population of monolinguals like the USA, the number of languages learned may indeed be influenced by the ease in learning languages. In contrast, it would be interesting to test whether a relationship between language aptitude and the number of languages persists in countries or regions where it is a necessity to learn multiple languages, like Switzerland or South Tyrol. What is more, the number of foreign languages does not reflect how well they are spoken. Thus, future studies addressing this question should assess objectively the level of proficiency of each foreign language in relation to the time of its exposure. To answer the question if musicality is beneficial for language aptitude, the language aptitude of professional musicians could be measured.

Finally, unravelling the relationships between language ability and disability variation, brain anatomy, function and genetics can aid in our understanding of how these abilities evolved in our past. Selection works on variation and by understanding how genetics, brain anatomy and the experiences with the environment give rise to language ability variations, we can try to understand how in our past higher and higher abilities were selected for.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. https://doi.org/10.1017/S027226310808073X

Altarelli, I., Leroy, F., Monzalvo, K., Fluss, J., Billard, C., Dehaene-Lambertz, G., Galaburda, A. M., & Ramus, F. (2014). Planum temporale asymmetry in developmental dyslexia: Revisiting an old question. *Human Brain Mapping*, *35*(12), 5717–5735. https://doi.org/10.1002/hbm.22579

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. In *Source: Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 57, Issue 1).

Benner, J., Wengenroth, M., Reinhardt, J., Stippich, C., Schneider, P., & Blatow, M. (2017). Prevalence and function of Heschl's gyrus morphotypes in musicians. *Brain Structure and Function*, *222*(8), 3587–3603. https://doi.org/10.1007/s00429-017-1419-x

Biedroń, A. (2015). Neurology of foreign language aptitude. *Studies in Second Language Learning and Teaching*, *5*(1), 13–40. https://doi.org/10.14746/ssllt.2015.5.1.2

Biedroń, A., & Pawlak, M. (2016). New conceptualizations of linguistic giftedness. *Language Teaching*, *49*(2), 151–185. https://doi.org/10.1017/S0261444815000439

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (117th ed., Vol. 83, pp. 867–873). Newbury House.

Carroll, J. B., & Sapon, S. M. (1959). Modern language aptitude test (MLAT). In *Psychological Corporation*.

Chi, J. G., Dooling, E. C., & Gilles, F. H. (1977). Gyral development of the human brain. *Annals of Neurology*, *1*(1), 86–93. https://doi.org/10.1002/ana.410010109

Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, *4*(NOV). https://doi.org/10.3389/fpsyg.2013.00874

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. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-60609-y

Đokić, R., Koso-Drljević, M., & Đapo, N. (2018). Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics. *PLOS ONE*, *13*(10), e0205169. https://doi.org/10.1371/journal.pone.0205169

Dörnyei, Z., & Skehan, P. (2003). Individual Differences in Second Language Learning. In *The Handbook of Second Language Acquisition* (pp. 589–630). wiley. https://doi.org/10.1002/9780470756492.ch18

Duyck, W., Szmalec, A., Kemps, E., & Vandierendonck, A. (2003). Verbal working memory is involved in associative word learning unless visual codes are available. In *Journal of Memory and Language* (Vol. 48, Issue 3, pp. 527–541). Academic Press Inc. https://doi.org/10.1016/S0749-596X(02)00533-8

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. *Cerebral Cortex*, *14*(1), 11–22. https://doi.org/10.1093/cercor/bhg087

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*(4), 581–590. https://doi.org/10.3758/BF03192728

Geschwind, N., & Levitsky, W. (1968). References and Notes. In *BroaJhbent Quart. J. Exls. Psy'chol* (Vol. 90). Academic Press.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex*, *17*(3), 575–582. https://doi.org/10.1093/cercor/bhk001

Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *Journal of Neuroscience*, *31*(11), 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., … Medland, S. E. (2020). The genetic architecture of the human cerebral cortex. *Science*, *367*(6484). https://doi.org/10.1126/science.aay6690

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. In *Psychiatry Research: Neuroimaging Section* (Vol. 98).

Jean Hausser, & Korbinian Strimmer. (2021). *entropy: Estimation of Entropy, Mutual Information and Related Quantities* (1.3.1.). R package.

Kepinska, O., Bouhali, F., Tanaka, H., Hoeft, F., & Golestani, N. (n.d.). Auditory cortex accommodates multilingualism. *In Preparation*.

Kepinska, O., de Rover, M., Caspers, J., & Schiller, N. O. (2017). On neural correlates of individual differences in novel grammar learning: An fMRI study. *Neuropsychologia*, *98*, 156–168. https://doi.org/10.1016/j.neuropsychologia.2016.06.014

Kepinska, O., Pereda, E., Caspers, J., & Schiller, N. O. (2017). Neural oscillatory mechanisms during novel grammar learning underlying language analytical abilities. *Brain and Language*, *175*, 99–110. https://doi.org/10.1016/j.bandl.2017.10.003

Li, S. (2015). The Associations between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research. In *Applied Linguistics* (Vol. 36, Issue 3, pp. 385–408). Oxford University Press. https://doi.org/10.1093/applin/amu054

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Mansouri, F. A., Rosa, M. G. P., Atapour, N., Sigala, N., Pardo-Vazquez, J. L., Champalimaud, F., & Funahashi, S. (2015). Working Memory in the Service of Executive Control Functions SHORT-TERM STORAGE OF INFORMATION REQUIRED TO GUIDE ONGOING OR UPCOMING BEHAVIOR. *Frontiers in Systems Neuroscience | Www.Frontiersin.Org*, *9*, 166. https://doi.org/10.3389/fnsys.2015.00166

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 Structure and Function*, *220*(2), 729–743. https://doi.org/10.1007/s00429-013-0680-x

Meara, P. (2005). *LLAMA Language Aptitude Tests The Manual*. http://www.swan.ac.uk/cals/calsres/lognostics.htm

Milovanov, R., & Tervaniemi, M. (2011). The interplay between musical and linguistic aptitudes: A review. In *Frontiers in Psychology* (Vol. 2, Issue NOV). https://doi.org/10.3389/fpsyg.2011.00321

Moerel, M., de Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. In *Frontiers in Neuroscience* (Issue 8 JUL). Frontiers Research Foundation. https://doi.org/10.3389/fnins.2014.00225

Novén, M., Schremm, A., Nilsson, M., Horne, M., & Roll, M. (2019). Cortical thickness of Broca's area and right homologue is related to grammar learning aptitude and pitch discrimination proficiency. *Brain and Language*, *188*, 42–47. https://doi.org/10.1016/j.bandl.2018.12.002

Papagno, C., & Vallar, G. (1995). Verbal Short-term Memory and Vocabulary Learning in Polyglots. *The Quarterly Journal of Experimental Psychology Section A*, *48*(1), 98–107. https://doi.org/10.1080/14640749508401378

Penhune, V. B., Zatorre, R. J., Macdonald, J. D., & Evans, A. C. (1996). *Interhemispheric Anatomical Differences in Human Primary Auditory Cortex: Probabilistic Mapping and Volume Measurement from Magnetic Resonance Scans*. https://academic.oup.com/cercor/article/6/5/661/497934

Peper, J. S., Brouwer, R. M., Boomsma, D. I., Kahn, R. S., & Hulshoff Pol, H. E. (2007). Genetic influences on human brain structure: A review of brain imaging studies in twins. In *Human Brain Mapping* (Vol. 28, Issue 6, pp. 464–473). https://doi.org/10.1002/hbm.20398

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, *41*(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Prat, C. S., Gallée, J., & Yamasaki, B. L. (2023). Getting language right: Relating individual differences in right hemisphere contributions to language learning and relearning. *Brain and Language*, *239*, 105242. https://doi.org/10.1016/j.bandl.2023.105242

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rademacher, J., Caviness, V. S., Steinmetz, H., & Galaburda, A. M. (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex*, *3*(4), 313–329. https://doi.org/10.1093/cercor/3.4.313

Rakic, P. (1988). Specification of Cerebral Cortical Areas. *Science*, *241*(4862), 170–176. https://doi.org/10.1126/science.3291116.

Reiterer, S. M., Hu, X., Sumathi, T. A., & Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology*, *4*(OCT). https://doi.org/10.3389/fpsyg.2013.00782

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. *Journal of Neuroscience*, *32*(47), 16597–16601. https://doi.org/10.1523/JNEUROSCI.1996-12.2012

Robinson, P. (2007). Aptitudes, abilities, contexts, and practice. In *Practice in a Second Language* (pp. 256–286). Cambridge University Press. https://doi.org/10.1017/CBO9780511667275.015

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the LLAMA aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49–60. https://doi.org/10.22599/jesla.24

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–694. https://doi.org/10.1038/nn871

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. *Nature Neuroscience*, *8*(9), 1241–1247. https://doi.org/10.1038/nn1530

Seither-Preisler, A., Parncutt, R., & Schneider, P. (2014). Size and synchronization of auditory cortex promotes musical, literacy, and Attentional skills in children. *Journal of Neuroscience*, *34*(33), 10937–10949. https://doi.org/10.1523/JNEUROSCI.5315-13.2014

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). John Benjamins. https://doi.org/https://doi.org/10.1075/lllt.2.06ske

Takahashi, T., Sasabayashi, D., Takayanagi, Y., Furuichi, A., Kobayashi, H., Noguchi, K., & Suzuki, M. (2022). Different Heschl's Gyrus Duplication Patterns in Deficit and Non-deficit Subtypes of Schizophrenia. *Frontiers in Psychiatry*, *13*. https://doi.org/10.3389/fpsyt.2022.867461

Turker, S., & Reiterer, S. M. (2021). Brain, musicality, and language aptitude: A complex interplay. In *Annual Review of Applied Linguistics* (Vol. 41, pp. 95–107). Cambridge University Press. https://doi.org/10.1017/S0267190520000148

Turker, S., Reiterer, S. M., Schneider, P., & Seither-Preisler, A. (2019). Auditory cortex morphology predicts language learning potential in children and teenagers. *Frontiers in Neuroscience*, *13*(JUL). https://doi.org/10.3389/fnins.2019.00824

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. *Frontiers in Psychology*, *8*(DEC). 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. *Neurobiology of Language*, *2*(3), 389–415. https://doi.org/10.1162/nol_a_00042

Vangehuchten, L., Verhoeven, V., & Thys, P. (2015). Pronunciation proficiency and musical aptitude in Spanish as a foreign language: Results of an experimental research project. *Revista de Linguistica y Lenguas Aplicadas*, *10*, 90–100. https://doi.org/10.4995/rlyla.2015.3372

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl’s gyrus and acoustic processing. *Journal of Neuroscience*, *29*(1), 61–69. https://doi.org/10.1523/JNEUROSCI.3489-08.2009

Wen, Z., Biedroń, A., & Skehan, P. (2017). Foreign language aptitude theory: Yesterday, today and tomorrow. In *Language Teaching* (Vol. 50, Issue 1, pp. 1–31). Cambridge University Press. https://doi.org/10.1017/S0261444816000276

Wen, Z. (Edward). (2015). 3. Working Memory in Second Language Acquisition and Processing: The Phonological/ Executive Model. In *Working Memory in Second Language Acquisition and Processing* (pp. 41–62). Multilingual Matters. https://doi.org/10.21832/9781783093595-007

Wen, Z., & Skehan, P. (2011). A new perspective on foreign language aptitude research: building and supporting a case for “working memory as language aptitude.” *Ilha Do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, *0*(60). https://doi.org/10.5007/2175-8026.2011n60p015

Wengenroth, M., Blatow, M., Heinecke, A., Reinhardt, J., Stippich, C., Hofmann, E., & Schneider, P. (2014). Increased volume and function of right auditory cortex as a marker for absolute pitch. *Cerebral Cortex*, *24*(5), 1127–1137. https://doi.org/10.1093/cercor/bhs391

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. *Cerebral Cortex*, *18*(4), 828–836. https://doi.org/10.1093/cercor/bhm115

Zatorre, R., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. In *Physiol. Ser. A* (Vol. 175, Issue 4). MIT Press 38 Urzelai. https://doi.org/10.1016/S1364-6613(00)01816-7

Appendix

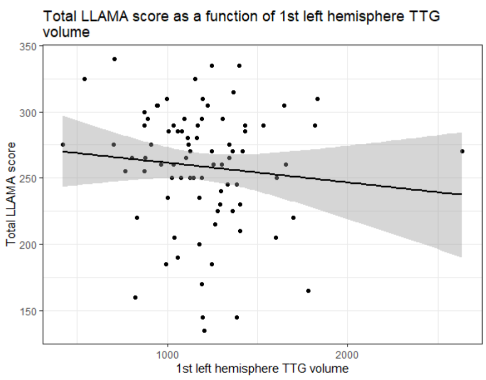
Language Aptitude and TTG volume

In the mixed model between TTG volume and total LLAMA scores, there was a significant interaction before correction between total LLAMA score and the second TTG, total LLAMA score and the right hemisphere and total LLAMA score and the second right hemisphere TTG.

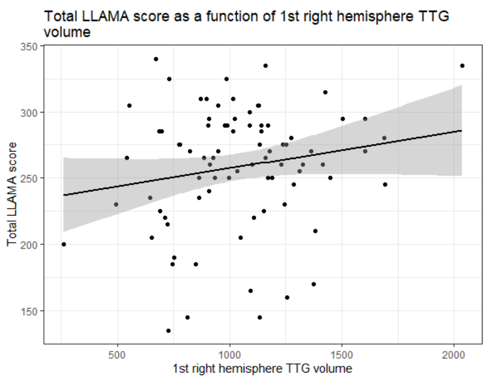
For this reason, we ran linear models between the first and second right and left hemisphere TTG volume and total LLAMA score while controlling for age and gender and whole brain volume. No model was significant (Table 1 in the appendix; Figure 1 in the appendix).

		Total LLAMA score			
		1st TTG		2nd TTG	
		Left	Right	Left	Right
(Intercept)	β	300.524***	285.915***	273.412**	289.819**
	SE	(81.641)	(80.386)	(81.295)	(92.637)
Estimated Total Intracranial Volume	β	0.000	0.000	0.000	0.000
	SE	(0.000)	(0.000)	(0.000)	(0.000)
Gender	β	0.694	3.104	4.319	2.840
	SE	(13.998)	(13.772)	(13.859)	(15.911)
Age	β	-0.787	-1.202	-0.919	-0.718
	SE	(1.318)	(1.290)	(1.281)	(1.640)
Volume	**β**	**-0.013**	**0.035.**	**0.029.**	**0.017**
	SE	**(0.017)**	**(0.018)**	**(0.015)**	**(0.023)**
Num.Obs.		82	82	82	66
R^2		0.016	0.055	0.056	0.020
R^2 Adj.		-0.035	0.006	0.006	-0.044
AIC		870.2	866.8	866.8	704.8
BIC		884.6	881.3	881.3	718.0
Log.Lik.		-429.099	-427.417	-427.413	-346.419
RMSE		45.33	44.41	44.41	46.05

Table 1: Results of linear models of total LLAMA scores depending on left and right first and second TTG volumes. The highlighted variable is the variable of interest.



a)



b)

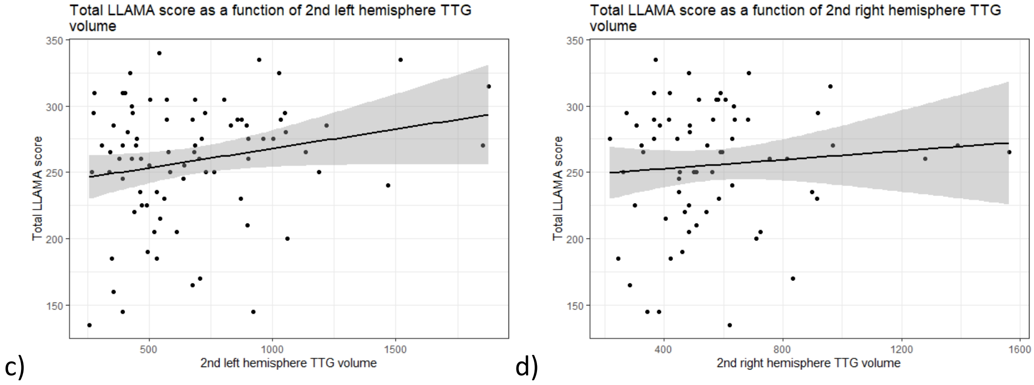


Figure 1: a-d Scatterplots of total LLAMA score and auditory cortex volumes

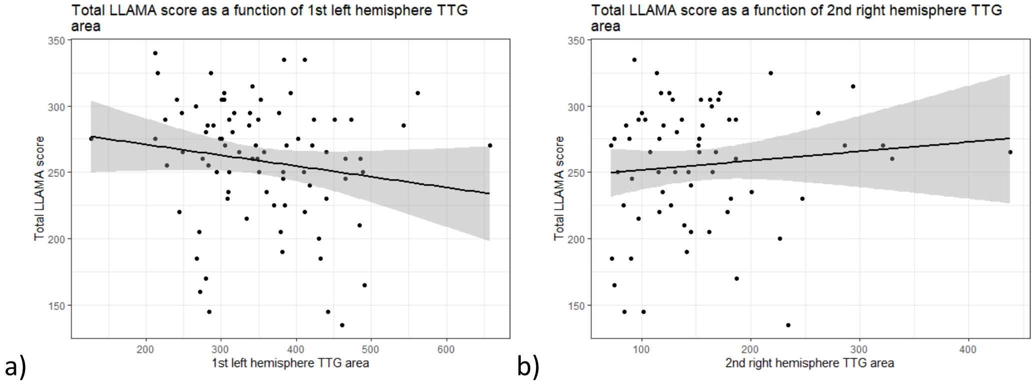


Figure 2: Scatterplot of total LLAMA score and a) first left hemisphere area b) second right hemisphere TTG area

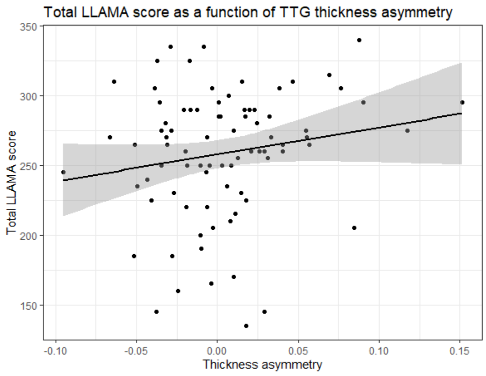


Figure 3: Scatterplot of total LLAMA score and whole TTG thickness asymmetry

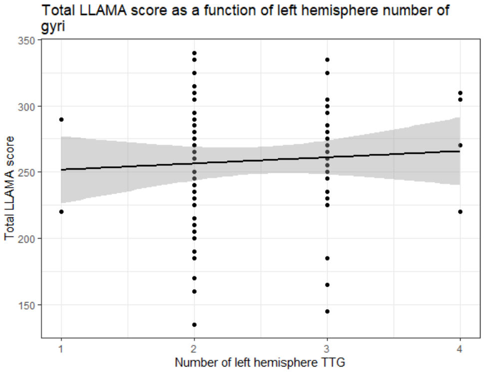
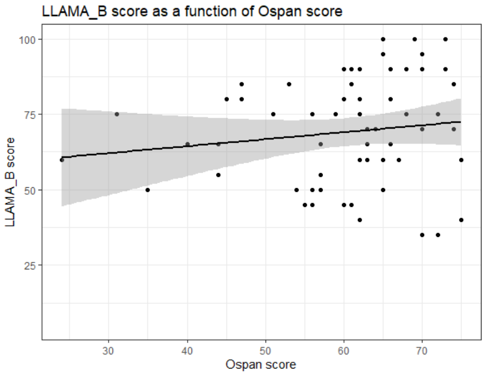
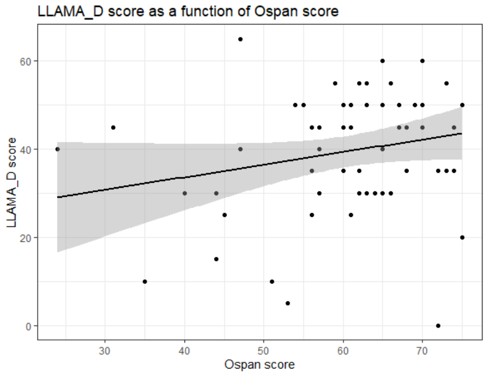


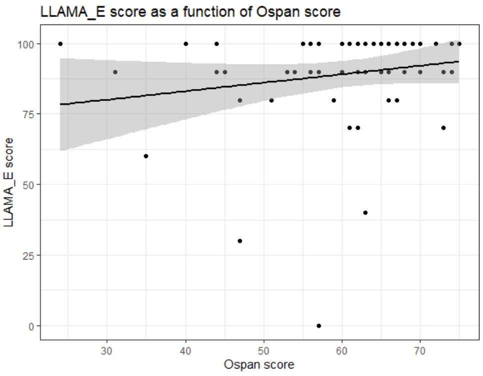
Figure 4: Scatterplot of total LLAMA score and number of left hemisphere TTG



a)



b)



c)

Figure 5: a-c Scatterplot of partial OSPAN score and LLAMA subtests