The heterogeneous engagement of the language network during auditory statistical learning: an fMRI study

Julie M. Schneider[1,2], Terri L. Scott[3], Jennifer Legault[2] and Zhenghan Qi[2,4]

Louisiana State University[1], University of Delaware[2], University of California San Franscisco[3], Northeastern University[4]

Corresponding Author:

Julie M. Schneider
Louisiana State University
77 Hatcher Hall
Field House Dr.
Baton Rouge, LA 70803
Phone: 225-578-9054
Email: juschnei@lsu.edu

Declarations of interest: None.

Highlights:
- We ask whether the core language network is recruited for statistical learning (SL)
- Subject-specific language regions (bilateral MTG and STG) were engaged during SL.
- Individuals vary in the most activated regions engaged during SL
- Little spatial convergence in co-activated brain regions for SL and language.
- SL and language comprehension were not similarly activated at the voxel level.

Keywords:
fMRI; Statistical Learning; Language Processing; Individual Differences

Abstract.
Statistical learning (SL) refers to the ability to detect and learn regularities from input and is foundational to language acquisition. Recent behavioral research has demonstrated a reciprocal relationship between prior language experiences and performance on linguistic measures of SL. Despite the dominant role of SL as a theoretical construct for language development, there has been a lack of direct evidence supporting shared neural substrates underlying language and SL. The current study tests whether the brain regions involved in language processing are similarly recruited during a speech segmentation SL task by probing the brain activation patterns of auditory SL within each learner's language network. Twenty-two adults performed an auditory SL task and a language localizer task as their brain activities were monitored with functional magnetic resonance imaging. Within the subject-specific language network, the bilateral superior temporal gyri (STG) and middle temporal gyri (MTG) showed greater hemodynamic responses while participants were processing structured versus random syllable sequences. However, due to substantial spatial heterogeneity of brain regions engaged during SL, there was no convergence across participants in the conjunction of SL and language tasks. We further ask whether neural representation of the SL and language tasks is similar within language regions and outside of the core language network by using a multi-voxel pattern similarity analysis approach. We found no evidence of task similarity either within the language regions or outside of the classic language network. Our findings suggest that while parts of the language network are recruited during auditory SL, the specific subregions vary substantially across individuals. Despite the close relationship between language processing and SL, they do not share a similar neurocomputational basis within the mature language network.

1. Introduction

Statistical learning (SL), a robust ability to rapidly detect regularities and variabilities from inputs, is a central theoretical account for first language acquisition (Aslin & Newport, 2012; Saffran et al., 1996). However, whether linguistic SL shares a neurocomputational basis with mature language processing remains an open question. On one hand, successful SL has been demonstrated across age groups (e.g., infants and adults), species (e.g. monkeys: Meyer & Olson, 2011); rats: Toro & Trobalón, 2005), modalities (e.g., non-linguistic auditory: Gebhart, Newport, review, & 2009, 2009; Saffran, Johnson, Aslin, & Newport, 1999); visual: Fiser & Aslin, 2002), and linguistic hierarchies (e.g., phoneme discrimination: Maye, Werker, & Gerken, 2002; syntax discrimination: Gomez & Gerken, 1999; Thompson & Newport, 2007), advocating for a domain-general SL mechanism. On the other hand, learning outcomes systematically vary across sensory modalities (e.g., Conway & Christiansen, 2005; Emberson, Conway, & Christiansen, 2011) suggesting SL performance is constrained by domain-specific mechanisms. Recent behavioral studies emphasize individual differences in SL and hint at a domain-specific relationship between linguistic SL and language processing (Arnon, 2019; Erickson et al., 2016; Schneider et al., 2020; Siegelman et al., 2018; Siegelman & Frost, 2015). Indeed, the modern neurobiological framework of SL proposes a dual-stream model that includes both domain-specific and domain-general learning components (Conway, 2020; Frost, Armstrong, Siegelman, & Christiansen, 2015). Limited studies investigating the neural mechanisms of SL have reported brain activation in both domain-general and language-related areas during learning of speech sound sequences (Cunillera et al., 2009; Finn et al., 2018; Finn & Hudson Kam, 2015; McNealy et al., 2006; Plante et al., 2017). However, domain-general and language-specific brain regions can often be found in close proximity to one another despite having contrasting functional profiles (Fedorenko et al., 2012),

confounding claims that SL is supported by one type of region or another. Elucidating the relationship between language and SL in the brain can provide key insights into the nature of domain-specific constraints on the process of SL. The current study aims to clarify whether auditory linguistic SL and auditory language processing share similar neurocomputational processes in adults by comparing the neural processes that support both language processing and SL within individual participants.

The potential functional overlap between SL and language processing has received indirect support from a limited set of functional neuroimaging studies. For example, in studies of SL involving speech stimuli, the inferior frontal gyrus (IFG), superior temporal gyrus (STG), and supramarginal gyrus (SMG) in the left hemisphere demonstrate sensitivity to statistical regularities (Cunillera et al., 2009; McNealy et al., 2006; Karuza et al., 2013; Plante et al., 2017; Orpella et al., 2021, 2022). For example, these regions show greater activation in processing structured versus random speech sequences (Cunillera et al., 2009; McNealy et al., 2006; Orpella et al., 2021, 2022), and increasing activation as a function of exposure (Plante et al., 2017; McNealy et al., 2006; Karuza et al, 2013). Due to the anatomical similarity between these regions and regions within the core-language network, which is a network of regions engaged specifically during comprehension of semantic and syntax information (Bautista & Wilson, 2016; Blank et al., 2016; Julian et al., 2012; Keller et al., 2001) but not during processing of nonlinguistic information (Fedorenko et al., 2011; Fedorenko & Varley, 2016), the activation of these regions during SL is often associated with language, although they have never directly been compared (e.g., Park et al., 2022; Finn et al., 2018). Existing evidence for the association between SL and language hinges on the assumption that the neural substrates found in SL have functions related to language. Despite providing early insights into the link between SL and language processing, this practice, known as

reverse inferencing, is limited by the weak functional selectivity of these brain regions/markers, as well as the remarkable inter-subject variabilities in the locations of the activation (Fedorenko et al., 2010; Poldrack, 2011). The lack of direct comparison between SL and language processing is problematic as the cortical regions found in these studies can be functionally heterogeneous themselves. The only known study to investigate common neural activation during spoken language processing and SL within the same individual utilized an artificial grammar learning (AGL) task and found similar P600 responses were elicited for structural incongruencies during spoken language and AGL. The authors interpreted this to indicate that the same neural mechanisms were recruited during syntactic processing of both language stimuli and SL of sequential patterns; however, it remains possible that both tasks are subserved by domain-general neural generators (Christiansen et al., 2012).

A more thorough investigation of the functional overlap between SL and language processing relies on the quantitative association between two cognitive processes. Despite substantial research documenting the behavioral relationships between SL and language processing skills (Arciuli & Simpson, 2012; Arciuli & Torkildsen, 2012; Conway, Bauernschmidt, Huang, & Pisoni, 2010; Conway & Pisoni, 2008; Kirkham, Slemmer, & Johnson, 2002; Misyak, Christiansen, & Tomblin, 2010; Qi, Sanchez Araujo, Georgan, Gabrieli, & Arciuli, 2018), only a few studies have demonstrated neurobiological evidence linking the two. For example, greater activation in the left middle frontal gyrus (MFG) during an auditory linguistic SL task was positively associated with rapid auditory processing speed (McNealy et al., 2006). Greater activation at left frontal insular cortex and left STG was associated with greater SL of non-adjacent dependencies (Orpella et al., 2021). In children with autism, less activation in the left inferior parietal lobule (IPL) and putamen was associated with more severe communicative deficits (Scott-

Van Zeeland et al., 2010). While brain-behavior correlational approaches facilitate the interpretation of complex patterns of neural activation (Reber, 2013), it is unclear whether the functional organization of SL in the brain is similar to that of language processing.

In addition to language-specific regions, many studies have highlighted the domain-generality of neural activation during SL tasks. For example, domain-general learning and memory systems, including the basal ganglia (e.g., Finn et al., 2019; Orpella et al., 2022) and hippocampus (e.g., Schapiro et al., 2014; 2017), are considered important for SL, though the necessity of their engagement is still greatly debated (Covington et al., 2018; Schapiro et al., 2014; 2017). Intriguingly, activation at the left, right, or bilateral inferior frontal gyrus (IFG) during processing of both linguistic and non-linguistic stimuli has been reported in most neuroimaging studies using SL paradigms (Milne et al., 2018; Karuza et al., 2013, Schapiro et al., 2013; Turk-Browne, Scholl, Chun, & Johnson, 2009). Activation in the LIFG has been positively correlated with familiarity ratings of shape sequences (Turk-Browne et al., 2009) and with segmentation of structured tone sequences (Abla & Okanoya, 2008). From these studies, researchers have concluded that because SL across modalities engages the LIFG, it is tapping into a domain-general mechanism. Therefore, contrary to the interpretation above, where activation of the LIFG is involved in processing linguistic stimuli, activation of the LIFG has been viewed as a generic online sequence processor, which drives the computation of statistical regularities (Petersson et al., 2012; Karuza et al., 2013). These strikingly contradictory interpretations for the same anatomical region (e.g., LIFG) emphasizes how additional research is needed to establish whether or not extraction of abstract statistical patterns engages linguistic processing.

While SL tasks have been shown to activate regions often engaged during language processing, others suggest these same regions underlie a domain-general SL mechanism. Three

reasons may account for the paradox. First, as mentioned, many studies associating SL with language processing rely on correlation and reverse inferencing, oversimplifying activation within these regions which are known to be highly heterogenous. Second, SL is highly heterogeneous across individuals. Group-level activation patterns may mask voxels that are activated at the individual level, and critical for SL. Third, activation during SL within the core-language network does not necessarily mean this region is processing regularities in an artificial language the same way as natural language processing. Taken together, there is a critical need to directly investigate similarities in cortical architecture among individual subjects during language processing and SL.

The current study therefore seeks to identify the functional neural profiles of auditory SL among individuals, and whether these patterns of activity are similar to those engaged during language processing. To answer these questions, we implement a Group-Constrained Subject-Specific (GCSS) analysis, which is designed to identify functional regions of interest engaged in high-level language processing (Scott et al., 2019; Julian et al., 2012; Fedorenko, et al., 2010) and address intersubject variability. We examine how individually defined language regions are activated in response to structured versus random speech sequences in an auditory linguistic SL task. We also seek to elucidate similarities in activation patterns within an individual during learning of structured versus random speech in an auditory linguistic SL task and during processing of intact versus degraded speech in an auditory language comprehension task using both univariate conjunction analyses and multivariate similarity analyses. We hypothesize that if linguistic SL is constrained by an existing language system, then we should expect to see at least a part of the mature language network actively engaged during linguistic SL. In addition, a shared functional organization of the language-related brain regions underlying SL and language processing will support a stronger version of this hypothesis.

2. Methods

2.1. Participants

Twenty-seven adults (M_{age} = 20.8 years, SD_{age} = 3.53 years, 20 Females) from the mid-Atlantic region of the United States were recruited via public flyers and websites. All participants were right-handed, monolingual English speakers, with no history of neurological disorder or developmental delay based on self-report. Participants were screened for MRI compatibility. The study was approved by the University Institutional Review Board and is in compliance with the Declaration of Helsinki. All participants provided informed consent and were compensated at a rate of $20/hour in gift cards for their time. Two participants were missing data from the first auditory SL run, and three participants were missing data from the second auditory SL run, yielding insufficient data points for analysis. Therefore, the final sample size for this study is twenty-two (M_{age} = 19.87, SD_{age} = 1.25, 15 females).

2.2. Language Localizer fMRI Task

2.2.1. Stimuli

To identify brain areas sensitive to spoken language, participants underwent functional neuroimaging while they passively listened to engaging audio clips of natural speech taken from the *Alice in Wonderland* story. To control for low-level acoustic features of speech, these intact clips were contrasted with clips of degraded speech, which had been processed to be unintelligible. The design and validation of this task is detailed in Scott et al. (2017). Briefly, the degradation procedure consisted of creating low-pass (cutoff-frequency 500Hz) filtered copies of a set of clips of speech similar to the intact set. Then, for each clip, amplitude envelope-modulated white noise was added where the envelope was copied from the original clip. This white noise was then "softened" with low-pass filtering at pass-band frequency 8,000Hz with stop frequency of

10,000Hz. The resulting clips are recognizable as human speech but lack lexical and phonemic content, and sound similar to an unintelligible radio broadcast. These stimuli, their transcriptions, and a presentation script are available online[1].

2.2.2. Procedure

Across two separate runs, 16 blocks of intact speech and 16 blocks of degraded speech were presented. Each run lasted 5 min and 58 s. Blocks consisted of one continuous 18s audio clip. In addition to the 16 stimulus blocks per run, five 14-s blocks of rest were included, in which participants heard no stimuli. A fixation cross was presented on the screen continuously during the entire run. Participants were only instructed to listen attentively, and no overt responses were recorded.

2.3. Auditory Statistical Learning Task

A common practice in fMRI studies of SL is to compare brain responses during a familiarization phase between sequences containing regular patterns versus those containing the same stimuli which are ordered randomly. However, implementing a random control condition often interferes with learning of structured sequences. Previous research found no evidence for behavioral learning, despite neural differences between structured and random sequences (McNealy et al., 2006, 2011). We address this issue in the current study by presenting structured and random sequences containing different types of stimuli in the same run (speech syllables and non-speech tones) and counterbalancing these conditions across runs. This approach has led to successful learning of structured sequences, as indicated by significantly quicker response times in the structured than the random condition (Schneider et al., 2020).

2.3.1. Stimuli

---

[1] https://evlab.mit.edu/aliceloc/

Syllable stimuli were constructed from twelve English consonant vowel (CV) syllables (pi/pa/pu/ti/ta/tu/di/da/du/bi/ba/bu). All syllable stimuli were made using an artificial speech synthesizer in a female voice. The mean duration of each syllable was 350 ms (SD = 40 ms) with a stimulus onset asynchrony (SOA) set at 480 ms. Tone stimuli were constructed from twelve musical pure tones within the same octave (a full chromatic scale starting from middle C). The mean duration of each tone was 460ms with a 20ms inter-stimulus interval.

Participants were exposed to experimental blocks containing sequences with embedded regularities (i.e., structured blocks), sequences with no statistical regularities (i.e. random blocks), and silence (i.e., resting blocks; see Figure 1). Structured blocks involved the presentation of sequences containing four triplets, which were created by concatenating the syllables into tri-syllabic "words" (pi-tu-bi, bu-pa-da, di-ba-pu, and ta-ti-du) and the tones into four target triplets (F#DE, ABC, C#A#F, and GD#G#). Random blocks contained the same 12 stimuli as presented in the structured blocks but were ordered pseudo-randomly, so that no combinations of any three stimuli were repeated more than once. A silent blank screen was presented during the resting blocks. Within each block, each triplet was repeated 8 times for a total of 32 triplets per block. Three structured blocks, 3 random blocks, and 6 resting blocks were concatenated in a pseudorandom order to create one run of auditory stimuli. Among the four experimental runs, half contained the syllable structured blocks intermixed with the tone random blocks and the other half contained the syllable random blocks intermixed with the tone structured blocks. The order of the structured, random, and resting blocks was pseudo-randomized, and the randomization was independent for each run and for each participant. Adjustment was made to ensure that no more than two blocks of the same stimulus type were adjacent to each other.
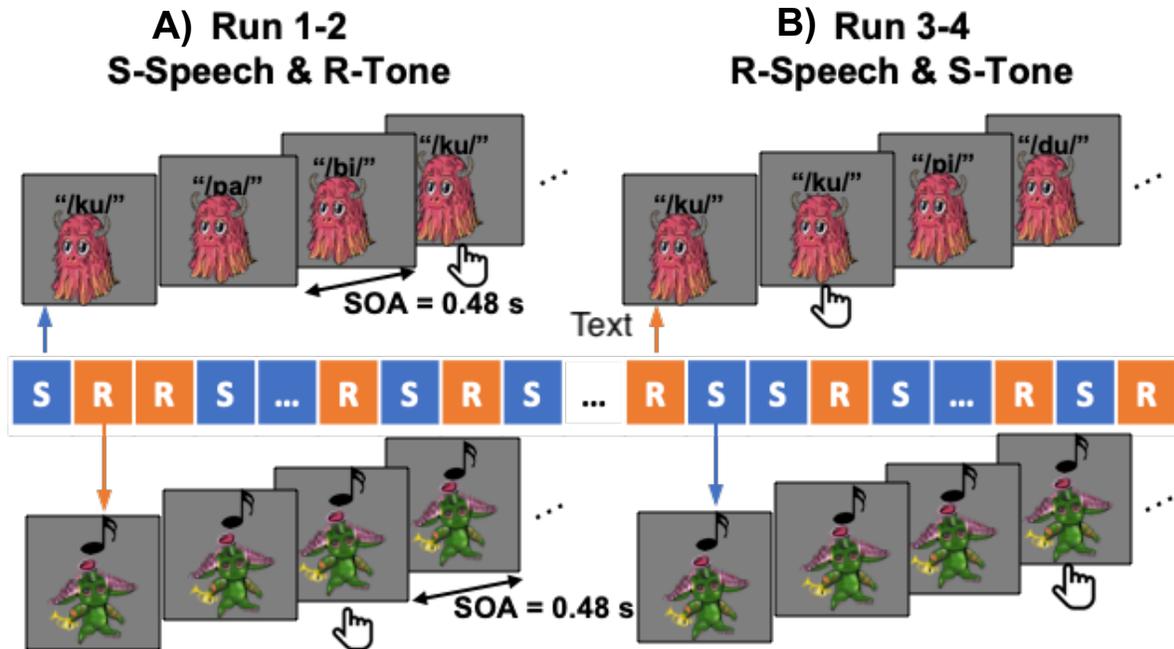
Figure 1. Overview of Auditory Linguistic Statistical Learning fMRI task. Participants were exposed to A) Structured blocks each containing a 96-syllable sequence with the embedded triplets (S-Speech), intermixed with three random blocks of tone sequences (R-Tone) and B) Random blocks each containing a 96-syllable randomly ordered sequence (R-Speech), intermixed with three structured blocks of tone sequences (S-Tone). Participants were instructed to alternate between tracking a specific syllable (e.g., /ku/) and tracking a specific tone in these sequences by pressing a button. Note: only the alien images, but not the spelling of the syllables, were present on the screen.

2.3.2. Procedure

Participants listened to sequences of sounds and viewed a screen with an alien cartoon character in the center while performing an auditory target detection task with a button response pad (Cambridge Research Systems). Before each block, a cue for the target was presented briefly with an auditory instruction "now listen for the sound /ku/" (Figure 1). The cue was a stimulus chosen from the last position of a triplet and was randomly assigned across each participant. Throughout the task, in each participant, the same cue was used for all the structured and random blocks of the same stimulus type. Because the order of stimuli and that of blocks were randomized independently for each participant, the target location among stimuli followed no systematic pattern either within- or cross-participants. Participants completed four experimental runs, each

lasted 6 minutes and 33 seconds. Because the focus of the current study is on the intersection between linguistic SL and language processing, only the syllable SL task was included in the current analysis (see Schneider et al., 2020 for more information on the tone SL task).

2.4. Neuroimaging Data

2.4.1. MRI acquisition

Data were acquired on a Siemens 3T Magnetom Prisma scanner with a 64-channel phased array head coil at the Center for Brain and Biomedical Imaging at the University of Delaware. Prior to functional imaging, whole-head, high resolution structural images, including a T1-weighted, magnetizations-prepared rapid gradient-echo (MPRAGE) anatomical volume (TR = 2500 ms, TE = 2.9 ms, TI = 1070 ms, flip angle = 8.0°, voxel resolution = 1.0mm isotropic, FOV = 256 x 256, 176 sagittal slices) and a T2-weighted anatomical volume (TR = 3200 ms, TE = 565 ms, flip angle = 2.0°, voxel resolution = 1.0 mm isotropic, FOV = 256 x 256, 32 sagittal slices) were collected.

Functional data were acquired using simultaneous T2*-weighted echo-planar imaging with multi-slice scans acquisition (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012) with the following acquisition parameters: TR= 800 ms, TE = 32 ms, flip angle = 61°, FOV = 210 x 210 mm, acceleration factor = 6. For the SL task, we required sixty adjacent slices in an interleaved order with 2.5 mm slice thickness resulting in an in-plane resolution of $2.5 \times 2.5 \times 2.5$ mm^3. For the language localizer task, we required 45 adjacent slices in an interleaved order with 3.0 mm slice thickness resulting in a an in-plane resolution of 3.0 x 3.0 x 3.0 mm^3. After fMRI preprocessing, the language localizer data was re-sampled to the same space as the SL data within FSL (FSL 5.0.9; Zhang, Brady, & Smith, 2001).

2.4.2. fMRI preprocessing

Functional and anatomical data was first converted using HeuDiConv (Halchenko et al., 2020) and then organized to cohere with the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016). fMRIPrep 1.3.1 (Esteban, Markiewicz et al., 2018; Esteban et al., 2018; RRID:SCR_016216), which is based on Nipype 1.1.9 (Gorgolewski et al., 2011; 2018), was utilized for all pre-processing steps. Many internal operations of fMRIPrep use Nilearn 0.5.0 (Abraham et al., 2014), mostly within the functional processing workflow. This automated preprocessing pipeline combines methodology from AFNI (Cox & Hyde, 1997), ANTs 2.2.0 (Avants et al., 2018), Freesurfer 6.0.1 (Dale, Fischl, & Sereno, 1999), FSL (FSL 5.0.9; Zhang, Brady, & Smith, 2001), and Mindboggle (Klein et al., 2005; 2017) to provide scientifically rigorous and reproducible data for use in data analysis. fMRIPrep was first conducted on anatomical data only, described in more detail below. All functional data was then pre-processed based on these corrected T1 images.

2.4.3. Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped. Brain surfaces were reconstructed using Freesurfer. Nonlinear spatial normalization (warping) of brain-extracted anatomical volumes to the ICBM 152 Nonlinear asymmetrical template version 2009c (Fonov et al., 2009) was performed. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w.

2.4.4. Functional data preprocessing

For each of the functional runs per subject (four auditory SL runs and two language localizer runs), the following preprocessing was performed. First, a brain-masked functional reference volume was generated using a custom methodology of fMRIPrep. A deformation field

to correct for susceptibility distortions was estimated based on a co-registered field map, using a custom workflow of fMRIPrep derived from D. Greve's epidewarp.fsl script [2] and further improvements from HCP Pipelines (Glasser et al., 2013). Based on the estimated field distortion, a distortion-corrected BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then used to compute the linear transformation between the subject's functional data and anatomical volume using boundary-based registration (bbregister; Greve & Fischl, 2009). Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters for each volume in the time-series were estimated with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters). Time-series data were then high-pass filtered and underwent slice-timing correction, recommended for multiband data with short TRs (Parker & Razlighi, 2019). Head-motion and field distortion corrections were then applied to the time-series with a single, composite transformation. The BOLD time-series were resampled to MNI 152.

2.5. Whole-brain univariate analysis

First level statistical analysis was carried out using FEAT (FMRI Expert Analysis Tool;(Woolrich, Ripley, Brady, & Smith, 2001). For each individual run, parameter estimates for structured and random speech relative to baseline, as well as for contrasts of interest (e.g., structured > random) were calculated. To combine data across runs for each participant, we averaged runs 1 and 2, and runs 3 and 4, together within FEAT. Parameter estimates for speech structured > random, speech random > structured were constructed. Group-level means for each contrast of interest were computed using FLAME (FMRIB's Local Analysis of Mixed Effects;

---

[2] http://www.nmr.mgh.harvard.edu/~greve/fbirn/b0/epidewarp.fsl

(Beckmann, Jenkinson, & Smith, 2003). All *z*-statistic (Gaussianised T/F) images were thresholded at cluster-forming threshold at $z > 3.1$ and a corrected cluster-level threshold of $p = 0.05$ (Worsley, 2012).

2.6. Group-constrained subject-specific (GCSS) analyses

Because understanding whether there are shared neural mechanisms underlying auditory SL and language processing depends critically on investigating functional activation within individual subjects, we performed GCSS analyses designed to take into account intersubject variability (Scott et al., 2019; Julian et al., 2012; Fedorenko, et al., 2010). With these analyses, we defined probabilistic regions of interest ("parcels") which we use to constrain our individual subject analyses. We used these parcels to measure patterns of similarity between tasks within individuals, as well as to constrain individual functional regions of interest (fROIs).

We first generated a set of parcels for the language localizer task. In each case we thresholded the data at $p < 0.001$ and binarized each subjects' contrast maps (intact > degraded speech in the language localizer) and then summed each set of maps to obtain a probability map. The more conservative threshold of $p < 0.001$ was chosen to capture the most statistically significant common areas of activation across individuals. The probability map was smoothed with a Gaussian kernel of 6mm FWHM and further thresholded to include only voxels in which at least two participants had significant activity. The results of this analysis are presented in Analysis section 3.3. This smoothed map was then segmented into parcels algorithmically, by first identifying all local maxima in the map, and then "growing" the parcels around these local maxima using a watershed algorithm (Meyer, 1991) implemented in the SPM-SS toolbox (Fedorenko et al., 2010, available at http://web.mit.edu/evlab/funcloc/) that extends the parcel to adjacent voxels until the edge reaches a zero-valued voxel or a local minimum. We imposed a constraint that local

maxima must be at least 4 voxels (or 1cm) apart. For the language localizer task, we chose to report parcels within which 80% or more participants had at least one significant voxel in line with previous work (Scott et al 2017, Federenko et al 2010; the results of which are included in section 3.2). For each subject, the top 10% of voxels in each parcel are designated as their individual fROI. From within each individuals' language fROI, the contrast between structured speech during SL and rest and the contrast between random speech during SL and rest were extracted and compared (Analysis 3.3).

2.7. Univariate conjunction analysis

In order to identify regions within individuals that show significant activation for both auditory SL and language, we use the same approach to identify parcels based on individual conjunction maps. First, each contrast maps for both tasks (intact > degraded speech in the language localizer and structured > random in the auditory SL task) were voxel-wise thresholded at p < 0.001 and binarized. Then, for each subject, we computed the conjunction of their two binarized contrast maps. This resulted in one conjunction map per subject, which was then used to locate common conjunction areas across participants using a similar parcellation technique as described above (Analysis 3.4): by first identifying all conjunction voxels (greater than 4 voxels apart) in the map, and then "growing" the parcels around these areas until the edge reaches a zero-valued voxel. The Harvard-Oxford cortical and subcortical structural atlases were used to identify brain regions.

2.8. Multi-voxel pattern analyses

To ensure thorough investigation of neural patterns shared across auditory SL and language processing, we used local pattern similarity analysis (LPSA). This technique is designed to identify brain regions that support similar functions during different tasks, regardless of the exact level of

activation (Scott & Perrachione, 2019). For this analysis, unsmoothed functional contrast maps from each task are compared. These maps have undergone the same preprocessing and first-level analysis as the data discussed so far, including warping to common MNI space, except that these maps have not been spatially smoothed and thresholded. To first determine whether the pattern of activity in each language parcel reflects similar task engagement during linguistic SL and language processing, we computed Pearson correlation coefficients between the structured vs. random contrast images from structured > random SL and intact > degraded language processing across all voxels in each parcel, within individual subjects. We assessed the significance of these correlations across our participants under a null hypothesis in which unrelated patterns of activity had a correlation of zero.

Next, correlations between the contrast maps of structured > random SL and intact > degraded language processing were computed across the whole brain using a 3-voxel radius searchlight for each subject. The searchlight is centered on each voxel in the brain and the Pearson correlation is computed between contrast maps for voxels falling within the sphere. The center voxel is then assigned the value of the resulting correlation coefficient, thus constructing a map of local correlations between tasks for each subject. The correlation maps were then Fisher-transformed and then normalized resulting in a z-score correlation map for each subject. These maps were then combined across subjects using GCSS (described above) to form parcels representing common regions with high pattern similarities across subjects. Z-scored correlation maps were thresholded at $z = 2.3$ ($p < 0.01$) and parcels with 50% or greater participants were chosen for this study. The use of liberal thresholds in this approach is to maximize the power to detect a common region. The Harvard-Oxford cortical and subcortical structural atlases were used to identify brain regions.

3. Results

3.1. Auditory SL performance

All behavioral analyses for the current study were conducted in R (R Core Team, 2012). Mean reaction time (RT) was computed as the average time it took for a participant to press the button for each target syllable. To be considered a valid response to the target syllable, the button press had to occur in the time window of one stimulus before and one stimulus after the target syllable; that is, -480 ms to +960 ms relative to the onset of the target syllable. A button press prior to the target syllable indexed anticipation, resulting in a negative reaction time. One participant was removed from the RT analysis, as they did not have enough valid key presses (< 6 trials) during the exposure phase ($M_{age} = 19.90$ years, $SD_{age} = 1.26$ years, Number of females = 14). Consistent with what we have reported in the complete sample in Schneider et al. (2020), the average reaction time to the target syllable across the entire exposure phase was significantly quicker in the structured ($M = 591.50$ ms, $SD = 200.27$ ms) versus random condition ($M = 678.71$ ms, $SD = 48.57$ ms; $t(20) = -1.93$, one-tailed $p = 0.03$).

3.2. Group-level neural engagement during auditory SL occurs in language-related brain regions.

Before taking a subject-specific approach to clarify the functional overlap between auditory SL and language processing, a whole-brain analysis was performed in order to investigate the brain regions sensitive to embedded regularities in speech. This analysis resulted in 23 significant clusters (Figure 2; Table 1). Similar to previous studies contrasting structured and random speech sequences (Karuza et al., 2014; Cunillera et al., 2009; McNealy et al., 2006), we obtained greater activation in bilateral temporal gyri for the structured speech stream compared to the random control. However, at the same z-score threshold ($z > 3.1$), there were no parcels in which activation

was observed consistently across more than 80% of participants. GCSS analysis revealed that the regions with the most individuals showing significant activation were left frontal pole ($n = 13$, 59%), right frontal pole ($n = 12$, 54%), and left lateral occipital cortex ($n = 12$, 54%).
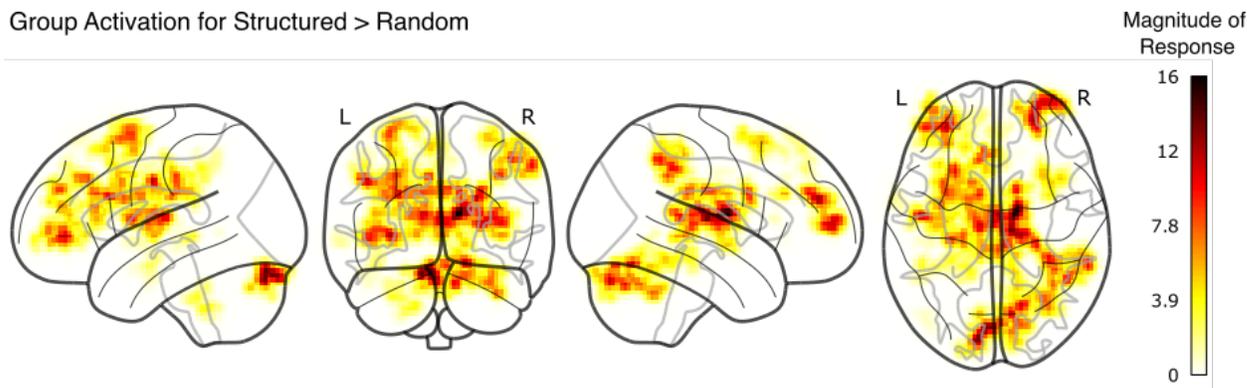

Figure 2. Group mean probability maps depicting neural engagement during processing of structured versus random speech sequences. The color bar represents the z-normed (z-stat) mean activity (β) across participants at each voxel.

Table 1.
Regions of significant clusters with peak coordinates in the speech SL task (structured > random). Regions labelled according to Harvard-Oxford Cortical and Subcortical Structural Atlases.

Cluster	(voxels)	Region	x	y	z	Z max	Uncorrected p-value
23	428	R Superior Temporal	39	-39.5	15.5	9.22	8.12E-14
22	339	L Anterior Cingulate	-6	28	21	9.12	1.05E-11
21	237	R Cerebellum (Visual Association Area)	6.5	-74.5	-28.5	14	4.89E-09
20	185	L Frontal Pole	-31	53	34.8	8.55	1.79E-07
19	170	R Frontal Pole	41.5	60.5	7.25	8.6	4.77E-07
18	160	R Supramarginal	61.5	-44.5	43	11	9.54E-07
17	150	R Posterior Cingulate	1.5	-12	29.2	9.16	1.97E-06
16	123	L Premotor	-18.5	13	73.2	7.98	1.57E-05
15	94	R Fusiform	34	-69.5	-23	8.78	1.75E-04
14	75	R Fusiform	14	-52	-25.8	10.7	9.78E-04
13	65	R Medial Prefrontal	39	13	59.5	10.4	0.003
12	57	L Medial Prefrontal	-33.5	35.5	34.8	10.5	0.006
11	57	L Fusiform	-13.5	-47	-45	9.1	0.006
10	52	L Insula	-23.5	30.5	4.5	7.72	0.010
9	49	L Supp Motor Area	-6	20.5	59.5	8.97	0.013

8	47	L Inferior Parietal	-48.5	-39.5	43	7.03	0.017
7	46	R Medial Prefrontal	4	45.5	34.8	8.57	0.018
6	45	L Inferior Temporal	-41	-9.5	-20.2	7.34	0.021
5	44	L Medial Cingulate	-6	-2	37.5	9.03	0.023
4	43	L Middle Inferior	-63.5	-4.5	-6.5	9.83	0.026
3	41	L Inferior Temporal	-43.5	-29.5	-12	6.6	0.032
2	40	L Cerebellum/Occipital Cortex	-41	-74.5	-25.8	7.69	0.036
1	40	R Caudate	9	23	4.5	8.07	0.036

Note: L: left; R: right.

3.3. Subject-specific language regions are engaged during auditory SL

In order to identify the brain regions commonly activated during language processing, a set of language parcels were generated from the language localizer task by contrasting intact and degraded conditions. Statistical maps were thresholded at $p < 0.001$, and only parcels with 80% or more participants are reported. From this parcellation of the language localizer maps, eleven parcels emerged: two in the left superior temporal gyrus – posterior (LSTG (posterior); $N = 21$, 95%) and anterior (LSTG (anterior); $N = 22$; 100%), left middle temporal gyrus (LMTG; $N = 21$; 95%), left temporal pole ($N = 21$; 95%), left precentral gyrus ($N = 19$; 86%), the left inferior frontal gyrus pars opercularis (LIFG; $N = 19$; 86%), two in the right superior temporal gyrus – posterior (RSTG (posterior); $N = 21$; 95%) and anterior (RSTG (anterior); $N = 21$; 95%), right middle temporal gyrus (RMTG; $N = 21$; 95%), right temporal pole ($N = 20$; 91%) and right precentral gyrus ($N = 20$; 91%; Figure 3).
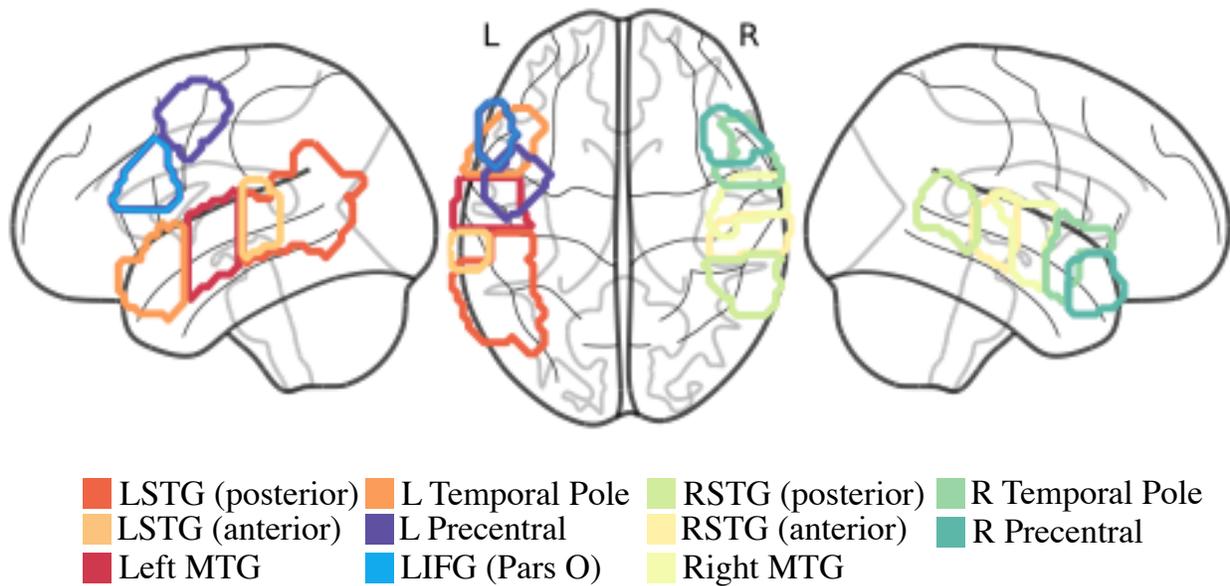
LSTG (posterior)	L Temporal Pole	RSTG (posterior)	R Temporal Pole
LSTG (anterior)	L Precentral	RSTG (anterior)	R Precentral
Left MTG	LIFG (Pars O)	Right MTG	

Figure 3. Parcels representing significant activation during language processing (intact > degraded) across more than 80% of participants ($z > 3.09$).

We next asked whether subject-specific language regions are engaged during SL. To statistically evaluate this functional overlap, we defined subject-specific language fROIs as the top 10% of language responsive voxels within each language parcel and then extracted the mean activation during SL for each condition (structured and random vs. rest) across these predefined voxels. This results in a measure of how sensitive language areas are to the SL contrast. Bonferroni corrections were applied to control for 11 comparisons. Listening to structured sequences of speech syllables resulted in significantly greater hemodynamic responses than listening to random sequences of speech syllables in bilateral MTG (left $t(34.46) = -3.87, p < 0.001$; right $t(41.44) = -2.85, p < 0.01$), left posterior STG ($t(41.90) = -2.13, p < 0.05$) and left anterior STG ($t(34.82) = -3.85, p < 0.001$) (Figure 4). Greater hemodynamic responses for structured versus random sequences in these fROIs were not significantly associated with learning outcome, as measured by mean reaction time difference between both conditions (left MTG, $R = 0.38, p = 0.09$; right MTG, $R = 0.32, p = 0.15$; left STG-posterior, $R = 0.21, p = 0.37$; left STG-anterior, $R = 0.32$, one-tailed $p = 0.15$). No

other language regions showed sensitivity to differences in statistical regularities of speech (p's > 0.05; see Supplementary Table 1 for full statistics).
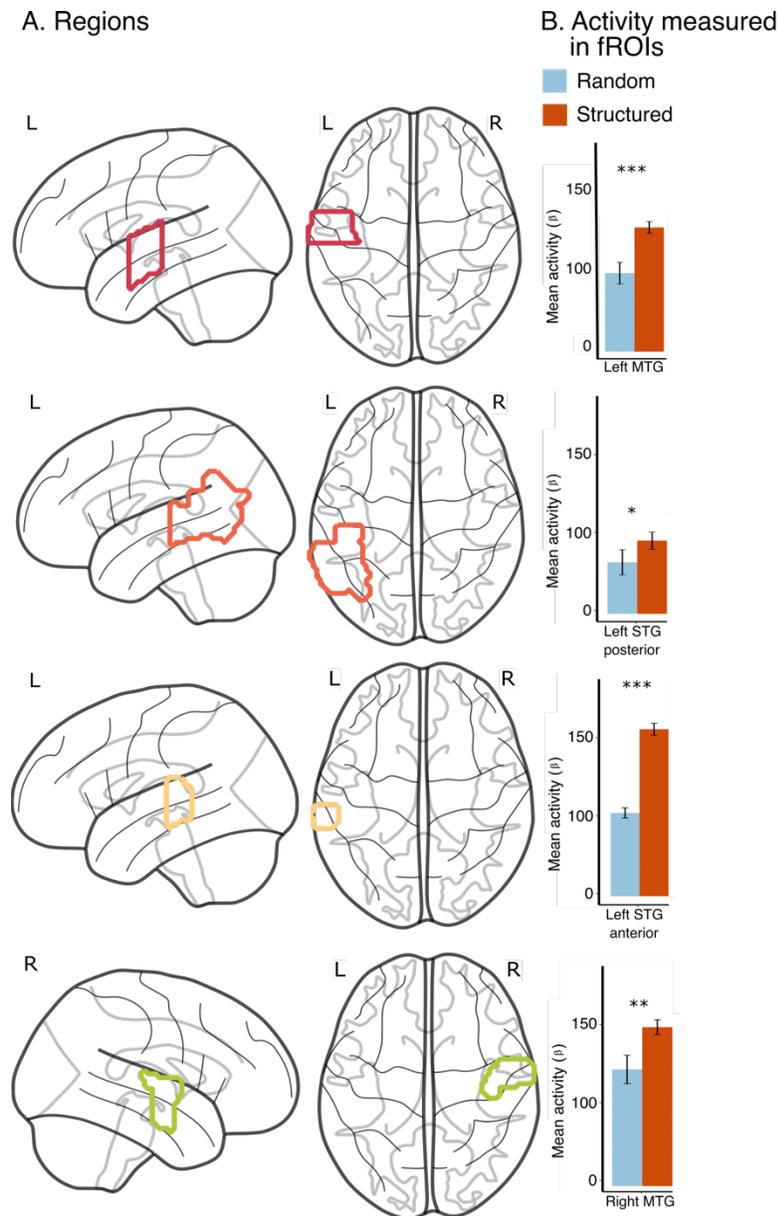

Figure 4. Mean activation during processing of structured speech and random speech, within each subject-specific language fROI. A) Greater activation for structured speech sequences compared to random speech sequences was observed in four fROIs: the bilateral MTG and LSTG (anterior and posterior). B) Mean activity (β) for random (blue) and structured (orange) speech sequences within each fROI. Error bars represent the within-subject standard-error. Asterisks represent Bonferroni-corrected values: *p < .05, **p < .01, ***p< .001.

3.4. Lack of spatial convergence in the neural basis of language processing and auditory SL

In order to characterize the degree of spatial heterogeneity in the functional organization of SL and language processing, we depicted the probability of activation (proportion of subjects showing statistically significant response at each voxel) within each task and the probability of conjunction (proportion of subjects showing co-activation at each voxel) across the two tasks. The probabilistic maps were smoothed and thresholded at 2 subjects (10%). As expected, participants' activation patterns showed a strong spatial consistency in the language localizer task especially in regions identified as language parcels in Figure 3 (Figure 5a). However, during the processing of structured versus random speech sequences, there was substantial individual variability in the patterns of neural activation, with evidently weak-to-modest overlap across participants in the left temporal region, right parietal region, and medial prefrontal region (Figure 5b). The probability map summarizing individual conjunctions (Figure 5c) shows that although many similar regions are recruited during SL and language, few common regions showing co-activation patterns emerged across subjects. As a result, we did not locate any significant conjunction area using this parcellation technique. This analysis demonstrates that, although subject-specific language parcels showed neural sensitivity for embedded speech regularities, there was no common brain region consistently activated by both tasks across all participants.
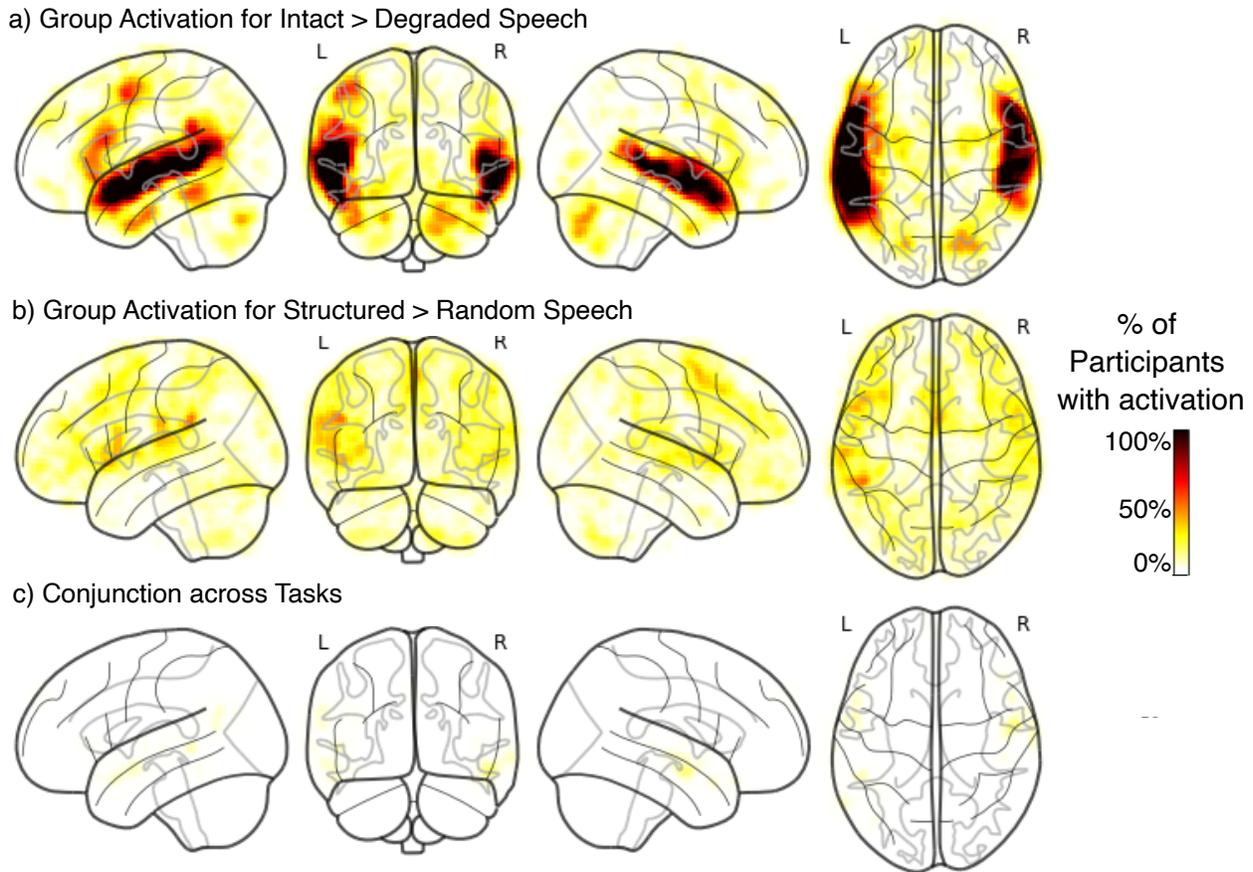
Figure 5. Probability maps within and between tasks depicting the high degree of inter-subject variability present during statistical learning. a) Group activation across voxels for intact > degraded speech. b) Group activation across voxels during processing of structured versus random speech. c) Lack of conjunction in activation between language processing and SL across voxels. The color bar represents the percent of participants with significant activation or conjunction at each voxel.

3.5. SL and language processing do not share common activation patterns.

Univariate conjunction analysis applies a voxel-wise activation threshold and searches for voxels consistently co-activated by both tasks across participants. This approach does not necessarily indicate similarities in multi-voxel activation patterns across both tasks. It may be the case that, despite greater activation for structured speech sequences compared to random speech sequences within fROIs (top 10% of voxels activated) defined by the language localizer task, the voxels which are most activated within a given parcel differs across tasks. To determine whether the pattern of activity in each parcel reflects similar engagement across tasks, we masked each

task's contrast map within each language parcel (bilateral MTG and LSTG (anterior and posterior); as shown in Figure 4), extracted the voxel values from within the whole parcel for each task independently, and then computed Pearson correlation coefficients between both tasks across all voxels in each parcel, within individual subjects. Correlations across tasks within each parcel did not yield any significant results (see Figure 5 and Supplementary Table 2). These findings point to a lack of representational similarity between tasks within individual subjects. This suggests that even though some language regions are sensitive to the SL contrast, substantial heterogeneity exists.
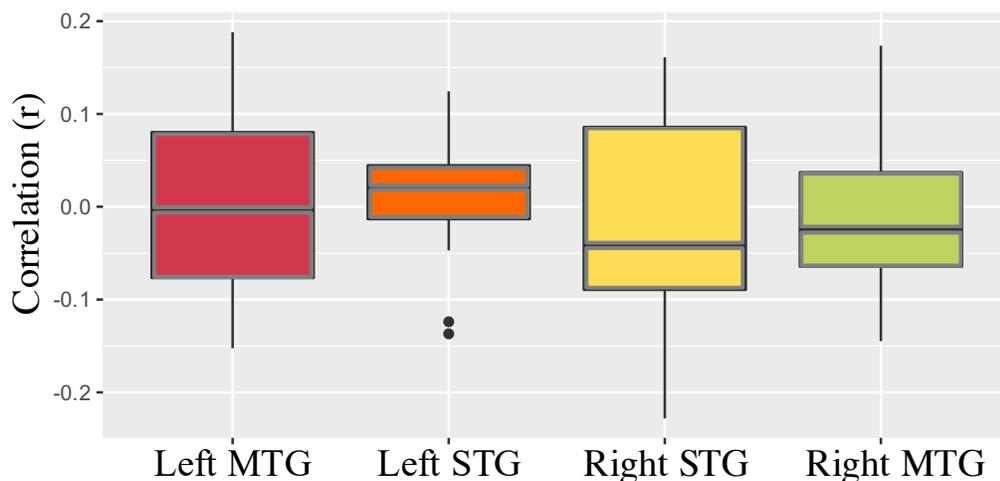
Figure 5. Lack of representational similarity between language and auditory SL within language parcels that were sensitive to differences in statistical regularities of speech during structured versus random sequences. Cross-task correlation values within the bilateral MTG and STG did not significantly differ from zero.

We next implemented a whole-brain local pattern similarity analysis to assess the voxel-wise correlations across the two tasks within each individual participant. Using a 3-voxel radius spherical searchlight, we calculated the local correlation coefficient between the two tasks centered on each voxel. This technique is designed to identify brain regions that support similar functions during different tasks, regardless of the exact level of activation. Since this analysis aims to reveal

more subtle correlations between patterns of activation across tasks in the whole-brain, rather than identifying the most highly activated regions within a parcel, we chose a more liberal threshold that is still appropriate for voxel-wise tests ($p < 0.01$) and not the most active regions. Despite this more liberal threshold, we did not uncover any region in which patterns of activation across tasks were correlated in more than 50% of participants (individual subject cross-task correlations can be seen in Supplementary Figure 1).

4. Discussion

The current study investigates the relationship between SL and language processing by testing whether the two tasks share similar neural substrates. Processing of structured speech in the auditory SL task engaged a number of brain regions that are traditionally part of the language network, including the bilateral STG and MTG. The brain functional overlap between tasks was supported by the sensitivity of language-related regions to embedded regularity in syllable sequences. Importantly, these language regions are defined individually using the GCSS approach in a language localizer task. However, further investigation revealed substantial variability across individuals in the spatial patterns during auditory SL. As a result, participants did not converge in the subregions of their language parcels recruited for auditory SL. Moreover, the multi-voxel activation patterns within each language parcel were not associated across tasks, suggesting distinct computations undertaken for each task within the same brain region. Lastly, our whole brain local-pattern similarity analysis expanded upon this finding, revealing no regions in which neural activity was correlated across tasks in regions outside of the core-language network. Taken together, the current findings suggest that, while SL does in fact engage language regions, the neural processes underlying language comprehension and SL of an artificial language are likely different.

Our approaches, relying on rigorous functional localization techniques, confirmed the involvement of language regions during auditory SL. We determined that the bilateral MTG and the left STG, parts of the frontotemporal core-language network (Fedorenko & Thompson-Schill, 2014; Price, 2010), are sensitive to embedded regularities in a stream of meaningless syllables. These findings are consistent with previous reports on the activation of temporal cortices in various auditory SL paradigms (Cunillera et al., 2009; Karuza et al., 2013; McNealy et al., 2006; Plante et al., 2015). Human temporal regions are sensitive to processing of auditory sequential patterns in artificial languages, in addition to the classical frontal regions (Petersson et al., 2012; Wang et al., 2015), and are reliably activated during syntactic processing of natural languages (Fedorenko et al., 2020; Snijders et al., 2009).

In addition to these important confirmatory results, our study contributes to the recent emerging literature on individual differences during SL (Erickson et al., 2016; Siegelman, Bogaerts, Christiansen, & Frost, 2017; Siegelman & Frost, 2015). We provide neural evidence of heterogeneous learning patterns across individuals. When identifying the percent of participants with overlapping neural activity at the voxel level for each task (see Figure 5), we found substantial convergence in neural engagement during the language processing task, similar to past research that highlighted the homogeneity of the core-language network across participants (Fedorenko, et al., 2010). In contrast, agreement across participants at the voxel-level during auditory SL was greatly reduced. These findings suggest that while some individuals engage language regions during auditory SL, which regions are recruited remains highly variable. Our exploratory analyses add to these findings, as the relationship between the magnitude of language-region engagement and learning behavior yielded null results, suggesting individual learning behavior might vary more with the functions of other brain regions, such as learning, memory, and executive

functioning (Conway, 2020). Further investigation utilizing neuromodulation should test whether successful auditory SL requires the functional involvement of language regions to provide better causal certainty (Uddén et al., 2008; Uddén, Ingvar, Hagoort, & Petersson, 2017).

Our whole-brain task correlation searchlight analysis further indicated a lack of spatial consistency between tasks across individuals. This overall lack of inter-task similarity between SL and language processing in adult brains is intriguing, especially in the face of emerging psycholinguistic frameworks aiming to unify language processing and acquisition. For example, part of the P-chain framework (Chang & Dell, 2014) proposes implicit learning as the mechanism for how a mature language system adapts to changing inputs during processing. However, our findings indicate that important future steps are required to elucidate the nature of the SL-language relationship in the brain. First, a developmental approach comparing children and adults in their recruitment of the language network during SL is necessary to test whether the maturity of the language system is associated with greater or lesser statistical sensitivity of an artificial language. Second, implementing a task that requires processing statistical regularities in natural language (Isbilen, McCauley, & Christiansen, 2022) will offer a better opportunity for us to discover language-SL functional similarities in the brain. Similarly, altering the language localizer task utilized may uncover additional functional similarities between language-SL, as the language parcels identified in the current study have been shown to be biased towards semantic, as opposed to syntactic, processing (Fedorenko et al., 2020). Finally, the current study has limited statistical power to uncover critical brain-behavior relationships. Although our sample size is sufficient to demonstrate group-level activation patterns in SL, a larger sample size with careful subject characterization will lead to a clearer understanding of how the involvement of language-related regions during SL is associated with SL outcome.

Taken together, our findings provide evidence that SL recruits brain regions specific to language with high spatial variability across individuals. As a result, we did not find commonly activated brain regions across processing of sentences versus structured sequences of speech without meaning. Moreover, multi-voxel pattern analyses suggest that distinct neural processes are undertaken by language regions during auditory comprehension and auditory SL. Our study is the first to specifically examine whether SL engages a core language network. Future research should be careful to not draw assumptions about the similarity of SL and language processing on the basis of activation in a given region alone, as our findings indicate that there are substantial regional differences between tasks when examined at the level of individual subjects.

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952–964.

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*(FEB). https://doi.org/10.3389/FNINF.2014.00014/FULL

Arciuli, J., & Simpson, I. C. (2012). Statistical Learning Is Related to Reading Ability in Children and Adults. *Cognitive Science*, *36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Psychology*, *3*, 324.

Arnon, I. (2019). Statistical Learning, Implicit Learning, and First Language Acquisition: A Critical Evaluation of Two Developmental Predictions. *Topics in Cognitive Science*, *11*(3), 504–519. https://doi.org/10.1111/TOPS.12428

Aslin, R. N., & Newport, E. L. (2012). What statistical learning can and can't tell us about language acquisition. In *Infant Pathways to Language: Methods, Models, and Research Directions* (pp. 15–30). https://doi.org/10.4324/9780203809907

Avants, B., Tustison, N. J., & Song, G. (2022). Advanced Normalization Tools: V1.0. *The Insight Journal*. https://doi.org/10.54294/uvnhin

Bautista, A., & Wilson, S. M. (2016). Neural responses to grammatically and lexically degraded speech. *Language, Cognition and Neuroscience*, *31*(4), 567–574. https://doi.org/10.1080/23273798.2015.1123281/SUPPL_FILE/PLCP_A_1123281_SM7002.WAV

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage*, *20*(2), 1052–1063. https://doi.org/10.1016/S1053-8119(03)00435-X

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. In *Trends in Cognitive Sciences* (Vol. 15, Issue 11, pp. 527–536). https://doi.org/10.1016/j.tics.2011.10.001

Blank, I., Balewski, Z., Mahowald, K., & Fedorenko, E. (2016). Syntactic processing is distributed across the language system. *NeuroImage*, *127*, 307–323. https://doi.org/10.1016/j.neuroimage.2015.11.069

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Http://Dx.Doi.Org/10.1080/01690965.2011.606666*, *27*(2), 231–256. https://doi.org/10.1080/01690965.2011.606666

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. In *Neuroscience and Biobehavioral Reviews* (Vol. 112, pp. 279–299). https://doi.org/10.1016/j.neubiorev.2020.01.032

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. https://doi.org/http://dx.doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(1), 24–39. https://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, *1145*, 113–131. https://doi.org/10.1196/annals.1416.009

Covington, N. V., Brown-Schmidt, S., & Duff, M. C. (2018). The necessity of the hippocampus for statistical learning. *Journal of Cognitive Neuroscience*, *30*(5), 680–697. https://doi.org/10.1162/jocn_a_01228

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., & Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*(3), 541–553. https://doi.org/10.1016/J.NEUROIMAGE.2009.06.069

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, *64*(5), 1021–1040. https://doi.org/10.1080/17470218.2010.538972

Erickson, L. C., Kaschak, M. P., Thiessen, E. D., & Berry, C. A. S. (2016). Individual Differences in Statistical Learning: Conceptual and Measurement Issues. *Collabra*, *2*(1). https://doi.org/10.1525/collabra.41

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). FMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111. https://doi.org/10.1038/S41592-018-0235-4

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), 16428–16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., Blank, I. A., Siegelman, M., & Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition*, *203*. https://doi.org/10.1016/j.cognition.2020.104348

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062. https://doi.org/10.1016/j.cub.2012.09.011

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. https://doi.org/10.1152/JN.00032.2010

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. In *Trends in Cognitive Sciences* (Vol. 18, Issue 3, pp. 120–126). https://doi.org/10.1016/j.tics.2013.12.006

Fedorenko, E., & Varley, R. (2016). Language and thought are not the same thing: Evidence from neuroimaging and neurological patients. *Annals of the New York Academy of Sciences*, *1369*(1), 132–153. https://doi.org/10.1111/NYAS.13046

Finn, A. S., & Hudson Kam, C. L. (2015). Why segmentation matters: Experience-driven segmentation errors impair "morpheme" learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(5), 1560–1569. https://doi.org/10.1037/xlm0000114

Finn, A. S., Kharitonova, M., Holtby, N., & Sheridan, M. A. (2018). Prefrontal and hippocampal structure predict statistical learning ability in early childhood. *Journal of Cognitive Neuroscience*, *31*(1), 126–137. https://doi.org/10.1162/jocn_a_01342

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327. https://doi.org/10.1016/j.neuroimage.2010.07.033

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. In *Trends in Cognitive Sciences* (Vol. 19, Issue 3, pp. 117–125). Elsevier Ltd. https://doi.org/10.1016/j.tics.2014.12.010

Gebhart, A., Newport, E., review, R. A.-P. bulletin &, & 2009, undefined. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Springer*, *16*(3), 486–490. https://doi.org/10.3758/PBR.16.3.486

Ghosh, S., Burns, C., Clark, D., Gorgolewski, K., Halchenko, Y., Madison, C., Tungaraza, R., & Millman, K. J. (2010). Nipype: Opensource platform for unified and replicable interaction with existing neuroimaging tools. *16th Annual Meeting of the Organization for Human Brain Mapping*.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135. https://doi.org/10.1016/S0010-0277(99)00003-7

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data 2016 3:1*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.44

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Halchenko, Y., Goncalves, M., Castello, M. V. di O., Ghosh, S., Salo, T., Hanke, M., Velasco, P., Dae, Kent, J., Brett, M., Amlien, I., Gorgolewski, C., Lukas, D. C., Markiewicz, C., Tilley, S., Kaczmarzyk, J., Stadler, J., Kim, S., Kahn, A., … Meyer, K. (2021). *nipy/heudiconv:* https://doi.org/10.5281/ZENODO.5557588

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123. https://doi.org/10.1016/J.COGNITION.2022.105123

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. https://doi.org/10.1016/J.NEUROIMAGE.2012.02.055

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*(1), 46–54.

Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, *11*(3), 223–237. https://doi.org/10.1093/cercor/11.3.223

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/http://dx.doi.org/10.1016/S0010-0277(02)00004-5

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology*, *13*(2). https://doi.org/10.1371/journal.pcbi.1005350

Klein, A., Mensh, B., Ghosh, S., Tourville, J., & Hirsch, J. (2005). Mindboggle: Automated brain labeling with multiple atlases. *BMC Medical Imaging*, *5*. https://doi.org/10.1186/1471-2342-5-7

Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., Li, P., Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., & Li, P. (2019). Immersive Virtual Reality as an Effective Tool for Second Language Vocabulary Learning. *Languages*, *4*(1), 13. https://doi.org/10.3390/languages4010013

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3). https://doi.org/10.1016/S0010-0277(01)00157-3

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience*, *26*(29), 7629–7639. https://doi.org/10.1523/JNEUROSCI.5501-05.2006

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2011). Age and experience shape developmental changes in the neural basis of language-related learning. *Developmental Science*, *14*(6), 1261–1282. https://doi.org/10.1111/j.1467-7687.2011.01075.x

Meyer, F. (1991). Un algorithme optimal pour la ligne de partage des eaux. 8me congrès de reconnaissance des formes et intelligence artificielle. Lyon, France.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19401–19406. https://doi.org/10.1073/PNAS.1112895108

Milne, A. E., Wilson, B., & Christiansen, M. H. (2018). Structured sequence learning across sensory modalities in humans and nonhuman primates. In *Current Opinion in Behavioral Sciences* (Vol. 21, pp. 39–48). https://doi.org/10.1016/j.cobeha.2017.11.016

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*, 31. https://doi.org/10.3389/fpsyg.2010.00031

Orpella, J., Mas-Herrero, E., Ripollés, P., Marco-Pallarés, J., & de Diego-Balaguer, R. (2021). Language statistical learning responds to reinforcement learning principles rooted in the striatum. *PLoS biology*, *19*(9), e3001119.

Orpella, J., Assaneo, M. F., Ripollés, P., Noejovich, L., López-Barroso, D., Diego-Balaguer, R. D., & Poeppel, D. (2022). Differential activation of a frontoparietal network explains population-level differences in statistical learning from speech. *PLoS biology*, *20*(7), e3001712.

Parker, D. B., & Razlighi, Q. R. (2019). The benefit of slice timing correction in common fMRI preprocessing pipelines. *Frontiers in Neuroscience*, *13*(JUL), 821. https://doi.org/10.3389/FNINS.2019.00821/BIBTEX

Petersson, K. M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95. https://doi.org/10.1016/J.BANDL.2010.08.003

Plante, E., Patterson, D., Gómez, R., Almryde, K. R., White, M. G., & Asbjørnsen, A. E. (2015). The nature of the language input affects brain activation during learning from a natural language. *Journal of Neurolinguistics*, *36*, 17–34. https://doi.org/10.1016/J.JNEUROLING.2015.04.005

Plante, E., Patterson, D., Sandoval, M., Vance, C. J., & Asbjørnsen, A. E. (2017). An fMRI study of implicit language learning in developmental language impairment. *NeuroImage: Clinical*, *14*, 277–285. https://doi.org/10.1016/j.nicl.2017.01.027

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. In *Neuron* (Vol. 72, Issue 5, pp. 692–697). https://doi.org/10.1016/j.neuron.2011.11.001

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. In *Annals of the New York Academy of Sciences* (Vol. 1191, pp. 62–88). Blackwell Publishing Inc. https://doi.org/10.1111/j.1749-6632.2010.05444.x

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2018). Hearing Matters More Than Seeing: A Cross-Modality Study of Statistical Learning and Reading Ability . *Scientific Studies of Reading*. https://doi.org/10.1080/10888438.2018.1485680

Reber, P. J. (2013). The neural basis of implicit learning and memory: a review of neuropsychological and neuroimaging research. *Neuropsychologia*, *51*(10), 2026–2042. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2013.06.019

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/http://dx.doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747. https://doi.org/10.1162/jocn_a_00578

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017a). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0049

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017b). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0049

Schneider, J. M., Hu, A., Legault, J., & Qi, Z. (2020). Measuring statistical learning across modalities and domains in school-aged children via an online platform and neuroimaging techniques. *Journal of Visualized Experiments*, *2020*(160), 1–21. https://doi.org/10.3791/61474

Scott, T. L., Gallée, J., & Fedorenko, E. (2017). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience*, *8*(3), 167–176. https://doi.org/10.1080/17588928.2016.1201466

Scott, T. L., & Perrachione, T. K. (2019). Common cortical architectures for phonological working memory identified in individual brains. *NeuroImage*, *202*(April), 116096. https://doi.org/10.1016/j.neuroimage.2019.116096

Scott-Van Zeeland, A. A., McNealy, K., Wang, A. T., Sigman, M., Bookheimer, S. Y., & Dapretto, M. (2010). No neural evidence of statistical learning during exposure to artificial languages in children with autism spectrum disorders. *Biological Psychiatry*, *68*(4), 345–351. https://doi.org/10.1016/j.biopsych.2010.01.011

Sheppard, J. P., Wang, J. P., & Wong, P. C. M. (2012). Large-scale cortical network properties predict future sound-to-word learning success. *Journal of Cognitive Neuroscience*, *24*(5), 1087–1103. https://doi.org/10.1162/jocn_a_00210

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/J.COGNITION.2018.04.011

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105–120.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: an FMRI study using word-category ambiguity. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(7), 1493–1503. https://doi.org/10.1093/CERCOR/BHN187

Thompson, S. P., & Newport, E. L. (2007). Statistical Learning of Syntax: The Role of Transitional Probability. *Language Learning and Development*, *3*(1), 1–42. https://doi.org/10.1080/15475440709336999

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875. https://doi.org/10.3758/BF03193539

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.

Uddén, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P., & Petersson, K. M. (2008). The inferior frontal cortex in artificial syntax processing: An rTMS study. *Brain Research*, *1224*, 69–78. https://doi.org/10.1016/j.brainres.2008.05.070

Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2017). Broca's region: A causal role in implicit processing of grammars with crossed non-adjacent dependencies. *Cognition*, *164*, 188–198. https://doi.org/10.1016/j.cognition.2017.03.010

Wang, L., Uhrig, L., Jarraya, B., & Dehaene, S. (2015). Representation of numerical and sequential patterns in macaque and human brains. *Current Biology : CB*, *25*(15), 1966–1974. https://doi.org/10.1016/J.CUB.2015.06.035

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Worsley, K. J. (2012). Statistical analysis of activation images. *Functional Magnetic Resonance Imaging: An Introduction to Methods*. https://doi.org/10.1093/ACPROF:OSO/9780192630711.003.0014

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*, *33*, 29–49. https://doi.org/10.1016/j.jneuroling.2014.09.004

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Supplementary Table 1.
Sensitivity to differences in statistical regularities of speech during structured versus random sequences within each fROI.

fROI	t-statistic	df	p-value
LMTG	-3.87	34.46	4.67E-4***
LSTG (posterior)	-2.13	41.90	3.93E-2*
L Temporal Pole	0.09	40.60	0.93
LSTG (anterior)	-3.85	34.82	4.88E-4***
RSTG (anterior)	-1.66	35.40	0.11
RMTG	-2.85	41.44	6.76E-3**
RSTG (posterior)	-0.93	41.79	0.36
R Temporal Pole	-0.16	35.13	0.87
R Precentral	0.79	41.87	0.44
L Precentral	-0.71	40.15	0.48
LIFG (Pars O)	-0.31	41.69	0.76

*Note: *Bonferroni-corrected $p < .05$, **$p < .01$, ***$p < .001$*

Supplementary Table 2.
Lack of representational similarity between language and auditory SL within language parcels that were sensitive to differences in statistical regularities of speech during structured versus random sequences. Median correlation values (R) across tasks within each parcel computed using unsmoothed data, as well as the result of one-sample, two-tailed t-tests on the Fisher transformed correlation values to determine if correlations across subjects differed significantly from zero. All correlations were performed within subjects comparing intact > degraded and structured > random contrasts across the tope 10% of activated voxels within each parcel.

ROI	R	t-value	p-value	CI
LMTG	-0.004	0.20	0.84	-0.04, 0.05
LSTG (posterior)	0.02	0.76	0.45	-0.02, 0.04
LSTG (anterior)	-0.04	-0.75	0.46	-0.07, 0.03
RMTG	-0.02	-0.20	0.84	-0.04, 0.03
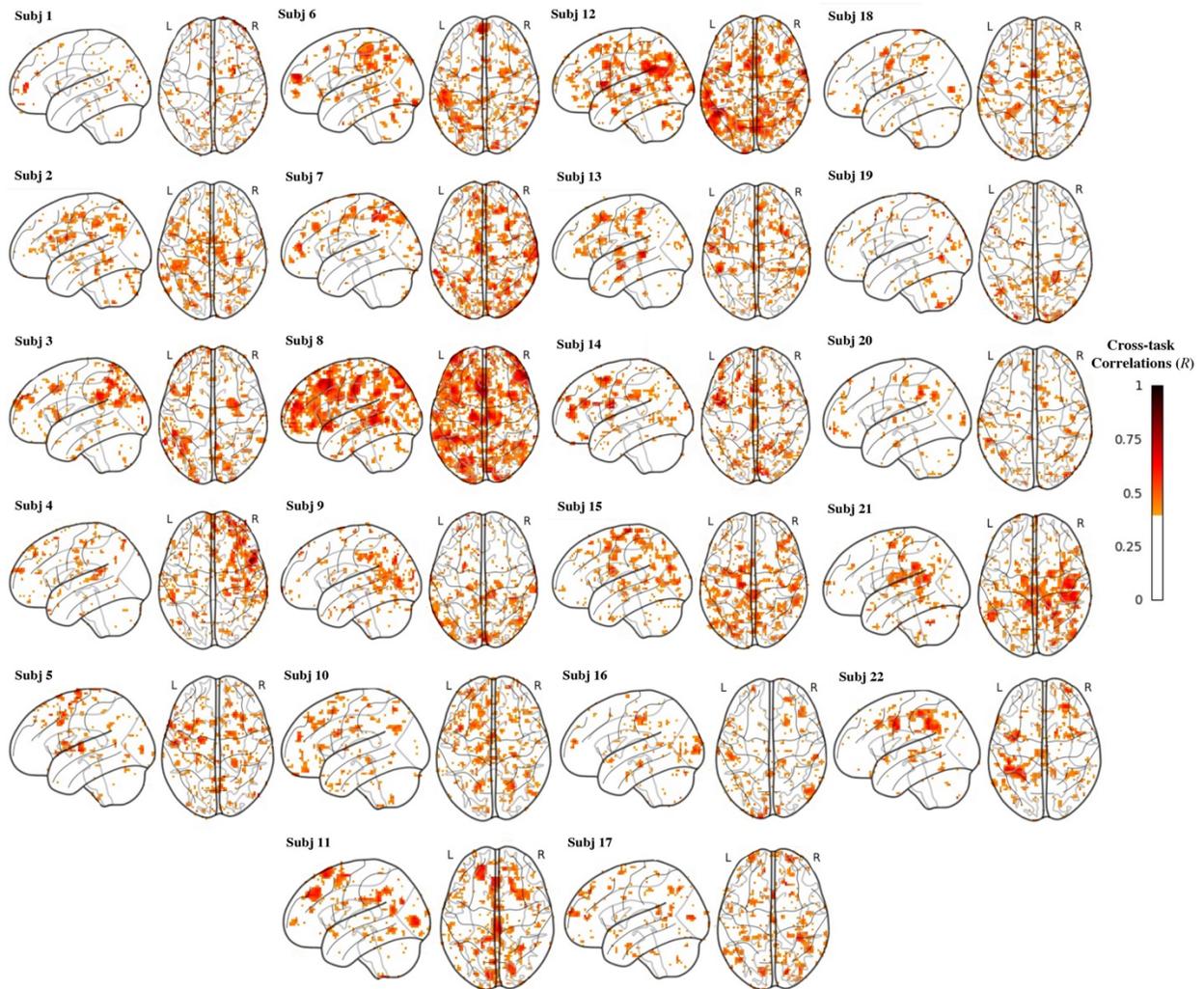
Supplementary Figure 1. Individual subject cross-task correlation maps. Using a 3-voxel radius spherical searchlight, we calculated the local correlation coefficient between the language processing and SL task centered on each voxel ($p < 0.01$). No parcels emerged as significant in at least 50% of participants, and individual maps suggest these null results may be attributed to substantial heterogeneity. Plots indicate correlation coefficients (R) at the voxel-level within individuals.