Brain & Language 215 (2021) 104919

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Structural neural correlates of individual differences in categorical perception

Pamela Fuhrmeister *, Emily B. Myers

Department of Speech, Language, and Hearing Sciences, University of Connecticut, 2 Alethia Drive, Storrs, CT 06269, United States

A R T I C L E I N F O

Keywords:
Speech perception
Categorical perception
Brain structure
Individual differences

A B S T R A C T

Listeners perceive speech sounds categorically. While group-level differences in categorical perception have been observed in children or individuals with reading disorders, recent findings suggest that typical adults vary in how categorically they perceive sounds. The current study investigated neural sources of individual variability in categorical perception of speech. Fifty-seven participants rated phonetic tokens on a visual analogue scale; categoricity and response consistency were measured and related to measures of brain structure from MRI. Increased surface area of the right middle frontal gyrus predicted more categorical perception of a fricative continuum. This finding supports the idea that frontal regions are sensitive to phonetic category-level information and extends it to make behavioral predictions at the individual level. Additionally, more gyrification in bilateral transverse temporal gyri predicted less consistent responses on the task, perhaps reflecting subtle variation in language ability across the population.

1. Introduction

One of the most well-known findings in the field of speech perception is that listeners perceive speech sounds categorically (e.g. Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Evidence for this phenomenon comes from studies in which listeners categorize and discriminate speech sounds taken from a synthetically modified continuum from, for example, /dɑ/ to /tɑ/ in which (in this case) voice onset time is increased in equal steps to form the continuum. Categorization data typically reveal a sharp increase in the proportion of /tɑ/ responses near the category boundary, rather than a gradual increase as voice onset time increases. Complementing categorization data, discrimination of tokens within a phonetic category tends to be poor but relatively good for tokens that span a category boundary. This behavioral pattern suggests that there is decreased sensitivity to distinctions within well-established phonetic categories. However, as discussed in the following sections, there is also ample evidence that listeners maintain sensitivity to the internal category structure of speech sounds and that listeners differ in how sensitive they are to that structure.

1.1. Graded vs. categorical perception

Many early studies assessed categorical perception using two-alternative forced choice tasks (e.g. Liberman et al., 1957, 1967). Although findings of categorical perception of speech using this method are robust, we have known for some time from studies using more sensitive measures that listeners do not completely discard within-category acoustic-phonetic variation and in fact maintain sensitivity to subtle within-category differences. For example, studies utilizing reaction time data (Pisoni & Tash, 1974), goodness judgments (Drouin, Theodore, & Myers, 2016; Miller, 1997), eye tracking (Clayards, Tanenhaus, Aslin, & Jacobs, 2008; McMurray, Danelz, Rigler, & Seedorff, 2018; McMurray, Tanenhaus, & Aslin, 2002), and visual analogue scaling tasks (a task in which participants move a slider between two options on a visual scale, Kapnoula, Winn, Kong, Edwards, & McMurray, 2017; Kong & Edwards, 2016) have found that listeners can indeed distinguish subtle within-category differences in speech stimuli. Notably, in contrast to binary categorization tasks, all of these measures allow a graded response. This gives listeners the opportunity to demonstrate their sensitivity to variation among tokens along the continuum. The ability to distinguish between subtle variants of a speech sound has some theoretical advantage and might be beneficial for understanding spoken language. For example, the ability to detect subtle

* Corresponding author.
E-mail address: pamela.fuhrmeister@uconn.edu (P. Fuhrmeister).

https://doi.org/10.1016/j.bandl.2021.104919
Received 26 August 2020; Received in revised form 18 November 2020; Accepted 12 January 2021

acoustic detail in the speech signal can help a listener recognize words by anticipating coarticulation (Gow, 2001). A common example of this is the difference in the /s/ sound in the words "see" and"sue." The acoustic properties of these two /s/ sounds differ depending on the following vowel, and sensitivity to these acoustic differences may help a listener predict upcoming speech sounds and in turn, recognize words more quickly. Sensitivity to subtle differences in sounds can also help a listener track the particular distribution of a speech sound for a given talker (Clayards et al., 2008) or resolve ambiguities in the speech signal when needing to revise an initial interpretation of a word (McMurray, Tanenhaus, & Aslin, 2009). A study by Kapnoula et al. (2017) found that participants who showed more graded perception of speech contrasts were more successful at integrating a secondary acoustic cue to distinguish voiced and voiceless stop consonants (F0), and secondary cue integration is important for perceiving many distinctions among speech sounds. All of these findings provide evidence that graded perception of speech sounds might confer certain advantages to a listener.

1.2. Individual and group differences in categorical perception

Until recently, most studies examining differences in how categorically or graded a listener could perceive speech sounds were focused on group-level differences, i.e., in different populations, such as children and adults or individuals with reading or language disorders and typically developing individuals. Many of these studies used two-alternative forced choice tasks and have found, for example, that children have shallower categorization slopes than adults (Burnham, Earnshaw, & Clark, 1991; Hazan & Barrett, 2000) and that individuals with reading or language disorders have shallower categorization slopes than a typically developing control group (Joanisse, Manis, Keating, & Seidenberg, 2000; Manis et al., 1997; Werker & Tees, 1987). The traditional interpretation of this data is that children or individuals with reading or language impairment show more graded patterns of perception of speech contrasts. In other words, the interpretation was that as tokens on a continuum change from one speech sound to another, the proportion of responses of one of the sounds changes accordingly. However, recent evidence using eye tracking (a more sensitive measure than a two-alternative forced choice task) suggests that children's perception of speech sounds actually becomes more *graded* throughout development (McMurray et al., 2018). Specifically, McMurray et al. (2018) assessed how categorically children perceive speech sounds by measuring lexical competition for speech sounds that varied along voice onset time and fricative continuua using a visual world paradigm. In their study, participants saw pictures that corresponded to two phonetic alternatives, e.g. "beach" and "peach" and were asked to select the picture that matched spoken words sampled along a continuum between the two endpoints. Considering the picture that the child chose as matching the input, they found that younger children indeed had shallower categorization slopes (a pattern that had traditionally been taken as evidence of more graded representations). At the same time, eye-tracking data revealed the opposite pattern: older children looked to the competitor item more often as tokens came closer to the category boundary, where younger children did not.

McMurray et al. (2018) argue that shallower categorization slopes from a two-alternative forced choice task may be more indicative of noisy representations or noisy encoding of the sounds rather than graded representations. A related possibility is that shallower categorization slopes reflect the reliability of responses (i.e., young children may be less reliable responders), which is likely influenced by the precision (or noisiness) of the representation. In other words, if younger children are less sensitive to subtle within-category differences, they may show some variation in their responses to tokens around the boundary, and this would lead to shallower categorization slopes in a two-alternative forced choice task that have historically been taken as evidence of graded perception (e.g., Burnham et al., 1991; Hazan & Barrett, 2000). Further support for this notion comes from the study by Kapnoula et al. (2017). They found no relationship between an individual's categorization slope from a visual analogue scaling task (which allows for more graded responses) and the slope from a two-alternative forced choice task, suggesting that these two tasks measure different aspects of speech perception. In that same study, they also found that a measurement of noise from responses on the visual analogue scaling task was more closely related to the slope of the two-alternative forced choice task, in that shallower slopes in the two-alternative forced choice task predicted noisier responses on the visual analogue scaling task. This adds to evidence that shallower slopes in two-alternative forced choice categorization tasks may be more reflective of noisy responses or representations than of true sensitivity to within-category distinctions. On the whole, it seems that most adults perceive speech sounds less categorically than originally thought and that graded perception may reflect a mature category representation that supports spoken language processing (McMurray et al., 2002). Thus, the studies reviewed above provide evidence that methods other than two-alternative forced choice tasks should be used in investigating individual or group differences in how categorically speech sounds are perceived.

Only a few studies that have found more graded speech representations among adults have specifically looked at *variability* in how graded or categorically an individual perceives speech. Two recent studies using visual analogue scaling tasks (a task in which participants are asked to move a slider along a continuum to indicate where they think a token lies between, in this case, two speech sounds) suggest that even typically developing adults vary in how categorically they perceive speech sounds (Kapnoula et al., 2017; Kong & Edwards, 2016). The earlier study by Kong and Edwards (2016) provided some of the first evidence that a visual analogue scaling task could measure individual differences in category gradiency, and the study by Kapnoula et al. (2017) validated this technique with a substantially larger sample size of over 100 participants. In addition, Kapnoula et al. (2017) used a novel statistical approach to measure integration of a secondary acoustic cue. Their results showed that individuals vary substantially in how categorically they perceive consonants along a voice onset time continuum and (as mentioned above) that more graded listeners are better at integration of secondary acoustic cues.

1.3. Sources of individual variability in categorical perception

As with any question of individual differences, it is of interest to understand where these differences originate. Testing relationships between individual variation in brain structure and behavior can shed light on the mechanism involved in that behavior, and it can also hint at whether differences arise from experience or whether they are innate differences (see Golestani, 2014; Golestani, Price, & Scott, 2011; Zatorre, Fields, & Johansen-Berg, 2012).

1.4. Structural and functional architecture of phonological category structure

To our knowledge, no studies have directly examined relationships between brain structure and individual differences in native-language speech perception (specifically categorical perception of native-language speech categories); however, we can make some predictions about where these differences might emerge from the functional activation literature. Well-established findings from functional MRI studies indicate that the left superior temporal gyrus and left inferior frontal gyrus are some of the primary brain regions involved in processing native-language speech (e.g., Damasio & Geschwind, 1984; Price, 2012). The brainstem encodes stimuli with high fidelity (Bidelman, Moreno, & Alain, 2013; Skoe & Kraus, 2010), but at some point in the auditory processing stream, these sounds are perceived categorically. Using a variety of methods, evidence from several studies suggests that frontal and temporal regions underlie representations of phonetic category structure.

Many studies have found that more posterior regions are involved in categorical perception. Some evidence suggests that categorical perception emerges in secondary auditory cortex including the posterior superior temporal gyrus (Bidelman et al., 2013; Chang et al., 2010). A study by Chang et al. (2010) used electrocorticography (a technique in which electrodes are placed directly onto the cortical surface in patients undergoing brain surgery) and found that parts of the superior temporal gyrus respond invariantly to specific acoustic-phonetic features. In contrast, an fMRI study by Myers (2007) found that the superior temporal gyrus responds to speech category structure in a graded manner. Specifically, greater activation was found in bilateral superior temporal gyri when tokens from a stop continuum that were poor members of the category (either exaggerated stimuli that were not competitive with another category or near-boundary tokens that were competitive with another category) were heard. This suggests that this region is not only sensitive to the category boundaries of speech sounds, but it is also sensitive to how prototypical a given exemplar is of its category. Functional activation in left temporal areas has also been found to predict individual differences in categorization of phonemic and non-phonemic stimuli. An fMRI study by Desai, Liebenthal, Waldron, and Binder (2008) suggests that a region encompassing the left posterior superior temporal gyrus and left posterior superior temporal sulcus is more active in response to sine wave speech when participants perceive the tokens as speech as compared to before participants are aware of the phonemic properties of the stimuli. In addition, activation in this region predicted how categorically participants perceived the speech and non-speech continua. Therefore, we may see that individual differences in how graded or categorically sounds are perceived may be correlated with differences in brain structure or morphology in left temporal areas.

In addition to temporal regions, several studies of native-language speech perception and non-native speech sound learning have suggested a role for frontal regions in categorical perception. This is typically indicated by changes in activation for members of different phonetic categories but no change in activation for acoustically distinct members of the same category. fMRI studies using univariate and multivariate approaches have found that the left inferior frontal gyrus and left middle frontal gyrus have been shown to respond more categorically, or invariantly, to speech categories (Lee, Turkeltaub, Granger, & Raizada, 2012; Myers, 2007; Myers, Blumstein, Walsh, & Eliassen, 2009). Myers (2007) found that bilateral inferior frontal gyri show greater activation for stimuli near a category boundary, suggesting these regions may help resolve competition among competing alternatives. Myers et al. (2009) found that the left inferior frontal sulcus responded invariantly to speech sounds. Using a multivariate analysis approach, Lee et al. (2012) found a similar pattern of results showing that Broca's area of the left inferior frontal gyrus showed patterns of activation consistent with categorical representations in two different data sets. In addition, several studies show evidence that the bilateral middle frontal gyri show categorical-like responses in perceptual learning tasks (Myers & Mesite, 2014) or newly learned phonetic categories (Luthra et al., 2019; Myers & Swan, 2012). Because brain function and brain structure are often related, it is likely that individual differences in brain structure that relate to behavioral differences in categoricity will be found in these frontal areas (inferior and middle frontal gyri) or nearby, functionally related areas.

One study of individual differences in brain structure and morphology hints at some relationships between the brain and individual differences in the perception of phonetic category structure. Golestani et al. (2011) looked for anatomical differences between a group of expert phoneticians and a group of non-expert controls and found that expert phoneticians were more likely to have multiple or split transverse temporal gyri compared to the controls. Additionally, gray matter volume of the pars opercularis, a region in the inferior frontal gyrus, was predicted by years of phonetic training. Although this study did not test the participants' perception of native-language phonetic category structure, it is nonetheless interesting that both frontal and temporal regions predicted phonetic expertise. Therefore, it is possible that brain structure may differ as a function of native-language speech ability or perception of native-language speech sounds. If graded perception of speech sounds indeed represents a mature or optimal representation of speech sounds, and phonetic expertise is predicted by differences in structure and morphology of the transverse temporal gyrus and the inferior frontal gyrus, we expect those regions to also be related to individual differences in gradedness of speech categories.

1.5. *Current study*

In the current study, our goal is to establish whether certain measurements of brain structure (surface area, cortical thickness, volume, or gyrification) predict individual differences in categorical perception and how consistently listeners respond to tokens on a phonetic continuum. To our knowledge, this is the first study to test relationships between brain structure and individual differences in categorical perception of native-language speech sounds. This is of interest because how categorically or graded an individual perceives speech sounds has been found to be related to language and reading disorders (e.g., Joanisse et al., 2000; Manis et al., 1997; Werker & Tees, 1987), and brain structure can often suggest whether abilities are learned or innate due to the developmental trajectory of different aspects of brain structure (e.g., gyrification patterns, cortical thickness). Because (to our knowledge) no previous studies have tested which regions' structural metrics predict individual differences in categorical perception of speech sounds, we mainly rely on the functional MRI literature to make predictions.

As reviewed above, the extant literature suggests that the following regions are involved in categorical perception or speech perception more generally: frontal regions, including the inferior and middle frontal gyri (Blumstein, Myers, & Rissman, 2005; Lee et al., 2012; Myers, 2007; Myers et al., 2009; Myers & Mesite, 2014; Myers & Swan, 2012; see also Golestani et al., 2011, for a structural MRI study about phonetic expertise), the superior temporal gyri (Bidelman et al., 2013; Blumstein et al., 2005; Chang et al., 2010; Desai et al., 2008; Feng, Gan, Wang, Wong, & Chandrasekaran, 2018; Myers, 2007), the planum temporale (Schremm et al., 2018), and the transverse temporal gyri (Golestani et al., 2011; Turker, Reiterer, Seither-Preisler, & Schneider, 2017). Frontal regions have been shown to be sensitive to category boundaries or phonetic competition (e.g., Blumstein et al., 2005; Myers, 2007; Myers et al., 2009) or show categorical-like responses (e.g., more sensitivity to between-category changes than within-category changes to stimuli; Luthra et al., 2019; Myers & Mesite, 2014), whereas temporal regions show sensitivity to the internal category structure of phonemes (Blumstein et al., 2005; Myers, 2007). Therefore, we expect to find relationships with structural measures from these regions and individual measures of categoricity and response consistency. More specifically, we predict that we will find relationships with brain structure and gradiency of perception in auditory/temporal regions, but individuals who are more categorical will show structural differences in frontal regions, such as the inferior frontal gyrus or middle frontal gyrus.

2. Method

2.1. *Participants*

We recruited 58 native speakers of English (43 female, 15 male) from the broader University of Connecticut community. We excluded data from one participant from all analyses because of an equipment error. Another participant did not complete the MRI session of the experiment, so that participant's data is included in the descriptive statistics of the behavioral data but excluded from the MRI analyses. Participants reported having no history of speech or language disorders and gave informed consent according to the guidelines of the University of Connecticut Institutional Review Board. Participants received $10 per hour for behavioral tasks and $30 per hour for the MRI.

2.2. Stimuli and Materials

Behavioral tasks were presented using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA). To obtain a measure of how categorically and consistently individuals perceive native-language speech categories, we asked participants to complete a (modified) visual analogue scale task. On each trial, participants heard one token from a seven-step continuum and were asked to move a slider to one of seven points on a line between two speech sounds to indicate where that speech sound belonged on the continuum (see Fig. 1). Having discrete response options (one response option per point on the continuum) allowed participants to respond completely consistently with the input. This way, we could obtain a measure of how consistently a participant responded each time a particular token was played, in addition to how categorically the sounds were perceived. Participants rated stimuli from a fricative continuum embedded in real words (sign-shine) and a synthetic stop contrast of consonant-vowel syllables (ba-da). The ba-da continuum was made at Haskins Laboratories with a Klatt synthesizer. The sign-shine stimuli were recorded from a native speaker of English (a female), and the continuum was created by waveform averaging in Praat (Boersma & Weenink, 2013). The tokens consisted of blends from 20% /s/ to 80% /s/ in 10% steps. We chose both a stop and fricative continuum because stop consonants are typically perceived more categorically than other classes of sounds (e.g., Eimas, 1963; Healy & Repp, 1982; Repp, 1981), and we wanted to ensure that we would see enough variability in our sample to test individual differences.

2.3. Procedure

The current study is a portion of a larger study, which consisted of two behavioral sessions and one MRI session. The two behavioral sessions took place on consecutive days. In the first session, participants gave informed consent and completed a non-native phonetic training task (data are not reported here). In the second session, participants completed two tasks to measure perception of native-language speech sounds (reported here, see next section for details). Measures of cognitive and language ability were obtained, as well (data not reported here). In a third session, we obtained structural MRI images from participants. Scanning was done with a 3-T Siemens Prisma with a 64-channel head coil. T1-weighted images were acquired sagitally by an MPRAGE sequence (TR = 2300 ms, TE = 2.98 ms, FOV = 256 mm, flip angle = 9 degrees, voxel size = 1 × 1 × 1 mm^3).

2.4. Categorical perception analysis

We analyzed the rating data from the visual analogue scale to extract two measures that relate to an individual's sensitivity to phonetic variability, **categoricity** (how categorically an individual perceived the speech sounds and **response consistency** (a measure of noisiness or how consistently a participant responded to a particular token on the continuum).

To obtain the measure of categoricity, we ran a mixed effects non-linear regression model that fit responses to a 3-parameter logistic function (3-parameter because the 4-parameter model never converged) for data from the ba-da continuum (correlations between random effects set to zero for convergence reasons). For the sign-shine continuum, we fit a mixed effects non-linear 2-parameter logistic model because the 3-parameter model did not converge. These models were run in R (R Core Development Team, 2008) using the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2019). The 3-parameter model estimates coefficients for the maximum asymptote, the inflection point (conceptually understood here as the category boundary), and the slope of the function (higher slope values indicate more categorical responses). The 2-parameter model estimates the inflection point and slope.

Response consistency for each participant was obtained by taking the mean of the squared (to avoid negative values) residuals from each model for each participant. This means that larger values represented *less* consistent responses because they were derived from the residuals. To make interpretation of the results more intuitive, however, we changed the sign of the response consistency measure so that larger values would represent *more* consistent responses on the task. The measures of categoricity and response consistency were entered into further analyses described below. All raw data and analysis scripts can be found at https://osf.io/7hak4/.

2.5. Analysis approach

2.5.1. Preprocessing

FreeSurfer's automated preprocessing pipeline was used to preprocess structural MRI data (Dale, Fischl, & Sereno, 1999; Fischl, 2012). FreeSurfer reconstructs cortical surfaces into a two-dimensional triangular mesh and estimates the pial surface (boundary between gray matter and cerebral spinal fluid) and white matter surface (boundary between white matter and gray matter), from which surface area, cortical thickness, and volume can be calculated. For region of interest analyses, each vertex of the triangular mesh is probabilistically assigned to a region according to an atlas[1].

2.5.2. Region of interest analyses

Regions of interest were selected from the Destrieux atlas in Freesurfer (Destrieux, Fischl, Dale, & Halgren, 2010). We identified the following bilateral regions of interest for our analyses based on the studies reviewed above: the pars opercularis region of the inferior frontal gyrus (Lee et al., 2012; Myers, 2007; Myers et al., 2009), the superior temporal gyrus (Myers, 2007), the transverse temporal gyrus and the planum temporale (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007; Golestani et al., 2011; Schremm et al., 2018; Turker et al., 2017; Wong et al., 2008), and the middle frontal gyrus (Luthra et al., 2019; Myers & Mesite, 2014; Myers & Swan, 2012). The FreeSurfer labels for these regions are included in Table 1 and can be seen in Fig. 2.

2.5.3. Gyrification

Because of previous work relating gyrification of the bilateral transverse temporal gyri to speech abilities (e.g., Golestani et al., 2007, 2011; Leonard et al., 2001; Turker et al., 2017), we tested the relationship between gyrification of this region and categoricity and response consistency. To maximize statistical power, we calculated a continuous measure of gyrification, the local gyrification index, using Freesurfer's -localGI flag in the recon -all command. As explained in more detail in Schaer et al. (2012), the local gyrification index is a ratio of the smoothed pial surface to the cortical surface, and it is calculated at each vertex of the two-dimensional cortical surface. To calculate the local gyrification index for a region of interest, as was done in the present study, the mean of the local gyrification indices at each vertex in that region of the cortical parcellation is calculated.

3. Results

3.1. Behavioral analyses of categoricity and response consistency

Descriptive statistics on these measures are included in Table 2. In general, there was substantial variability among individuals for both categoricity and response consistency measures from both continua (Fig. 3). Representative psychometric functions are shown in Fig. 3A for two participants illustrating graded vs. categorical and consistent vs. inconsistent (Fig. 3B) responses to phonetic variability. Of interest,

[1] An exploratory whole-brain analysis can be found in supplementary materials.

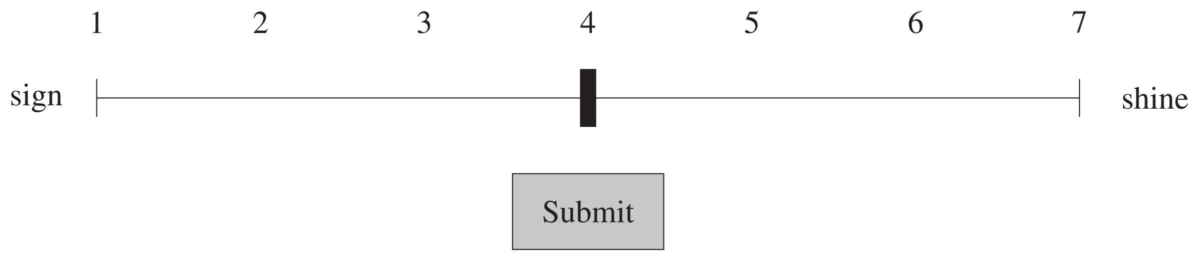

Fig. 1. Sample trial of the visual analogue scaling task.

Table 1
Regions of interest and Freesurfer Destrieux atlas labels. All regions were tested bilaterally.

Region of interest	Destrieux atlas label
Middle frontal gyrus	G_front_middle
Inferior frontal gyrus (pars opercularis region)	G_front_inf-Opercular
Transverse temporal gyrus	G_temp_sup-G_T_transv
Planum temporale	G_temp_sup-Plan_tempo
Superior temporal gyrus	G_temp_sup-Lateral

within each continuum, categoricity and response consistency measures did not correlate with one another (Fig. 3C & D), suggesting that these metrics measure separable aspects of phonetic perception. The categoricity measure (slope) did show a modest correlation between phonetic continua, r =.32, p =.02 (Fig. 3E), whereas the response consistency measure did not correlate between continua (Fig. 3F).

3.2. Region of interest analyses

For each of the following analyses, we fit a series of linear regression models that predicted categoricity (slope coefficient) or consistency (mean of squared residuals with the sign changed to facilitate interpretation) for the ba-da and sign-shine continuua. We included all bilateral regions of interest as predictors in each regression model but fit separate models for each dependent variable (ba-da categoricity, sign-shine categoricity, ba-da consistency, sign-shine consistency), as these are testing different questions. For each dependent variable, we fit three models: one for surface area, one for cortical thickness, and one for volume. To account for differences in head size, total intracranial volume was included as a predictor in models with surface area or volume measurements as predictors.

3.2.1. Categoricity

Continuum: ba-da. No structural measurements of the regions of interest predicted the ba-da slope.

Continuum: s-sh. Surface area of the right middle frontal gyrus positively predicted categoricity when holding other predictors constant, β =.09, SE =.04, t = 2.43, p =.02 (see Fig. 4), suggesting that individuals with more surface area in this region showed more categorical patterns of perception. The analyses reported here are largely exploratory, as (to our knowledge) this is the first study to address this particular question. Therefore, full exploration of our data resulted in a number of statistical tests being done. To better estimate the confidence or uncertainty around this effect, we computed non-parametric bootstrapped confidence intervals for the predictors of this model using the Boot function in the car package in R (Fox & Weisberg, 2019) with 1000 bootstrap samples. This technique resamples the data by taking random samples of the data and fitting the model for each random sample. This allows us to get a better estimate of the distribution of effects and the confidence around them. Confidence intervals for this estimate did not include zero: β =.09 (95% CI [.01, .17]).

3.2.2. Response consistency

Continuum: ba-da. No structural metrics from our regions of interest predicted response consistency on the ba-da continuum.

Continuum: s-sh. No structural metrics from our regions of interest predicted response consistency on the sign-shine continuum.

3.3. Gyrification

3.3.1. Categoricity

To test whether gyrification of the transverse temporal gyri predicted measures of categoricity in the native-language, we fit two linear regression models that predicted categoricity (slope coefficients from visual analogue scaling tasks). Fixed effects included only the

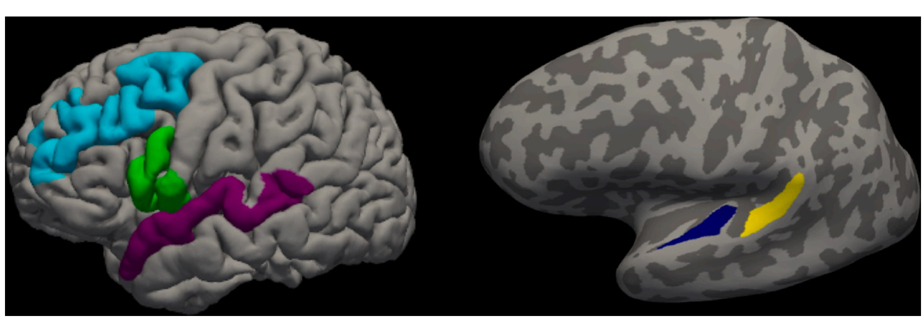

Fig. 2. Regions of interest and Freesurfer Destrieux atlas labels. All regions were tested bilaterally. Labels can be found in Table 1.

Table 2
Descriptive statistics for categoricity and response consistency measures for each continuum (N = 57).

Continuum	Categoricity mean	Categoricity SD	Categoricity min.	Categoricity max.	Consistency mean	Consistency SD	Consistency min.	Consistency max.
ba-da	.57	.31	.17	1.95	−1.11	.39	−2.21	−.28
sign-shine	.80	.15	.44	1.06	−1.11	.39	−1.99	−.24

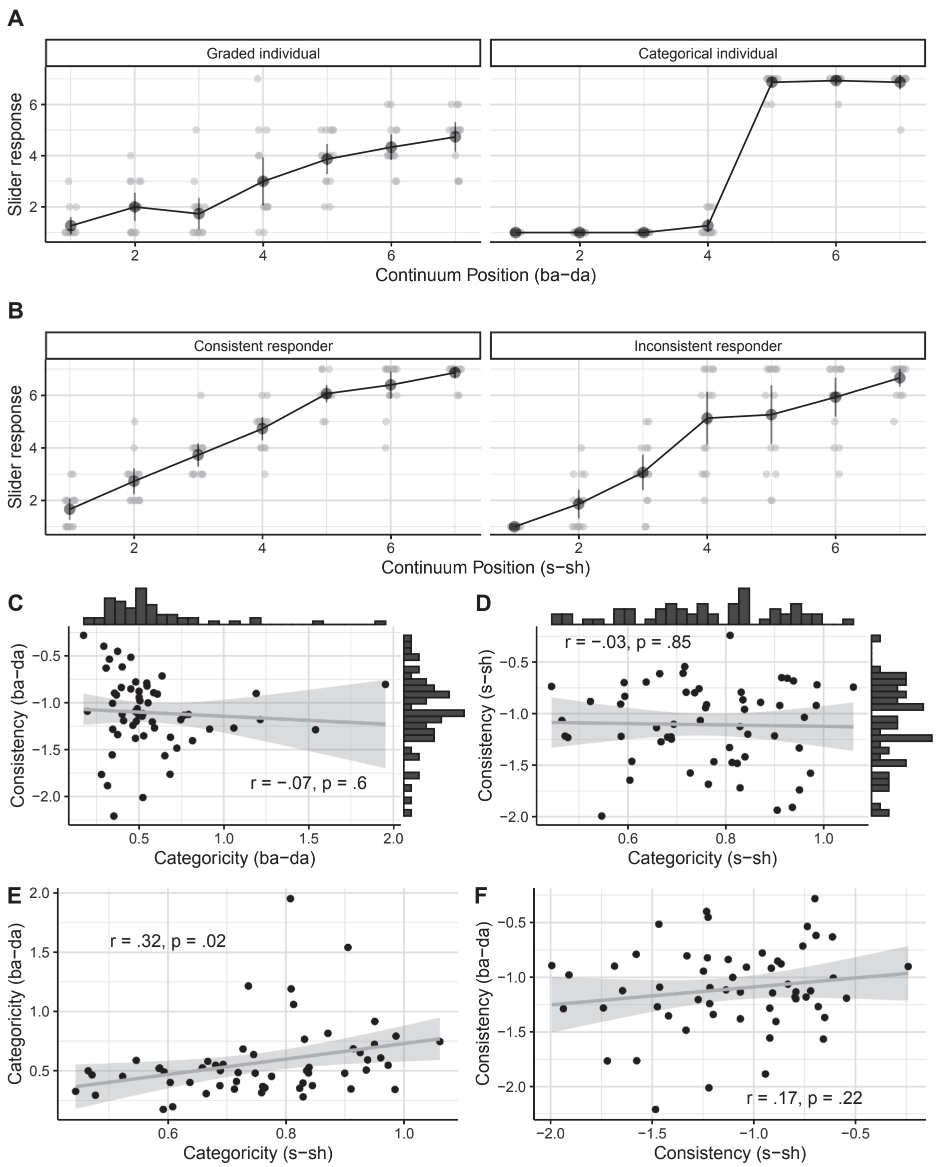

Fig. 3. A. Examples of individual participant data for graded and categorical response patterns. B. Examples of individual participant data for consistent and inconsistent response patterns. For plots A and B, gray points are responses on each trial, and black points are mean responses for each continuum point with error bars showing 95% confidence intervals. C. Categoricity and response consistency of the ba-da continuum were not correlated with each other. Distribution of values are shown as marginal histograms. D. Categoricity and response consistency of the sign-shine continuum were not correlated with each other. Distribution of values are shown as marginal histograms. E. Categoricity measures (slope coefficients) for ba-da and sign-shine were correlated with each other, so that more categorical responses on one continuum were related to more categorical responses on the other. F. Response consistency measures for ba-da and sign-shine were not correlated with each other. Shaded regions for plots C-F represent 95% confidence intervals. (N = 57).

interaction of local gyrification and hemisphere. This allowed us to test the simple effects of the local gyrificaion index on the dependent variable in each hemisphere separately and also allowed us to uncorrelate the measure between the two hemispheres. Hemisphere was deviation coded as in the previous models. The local gyrification index of either hemisphere did not significantly predict categoricity in either continuum (ba-da or sign-shine).

3.3.2. Response consistency

To test whether gyrification of the transverse temporal gyri predicted measures of response consistency on the native categorization task, we fit two linear regression models that predicted response consistency (the mean of the squared residuals, again with the sign changed to facilitate interpretation). Fixed effects in both models included only the interaction of the local gyrification index and hemisphere. Hemisphere was deviation coded as in the previous models. The first model predicted response consistency on the ba-da continuum. Local gyrification index in the left hemisphere negatively predicted response consistency, $\beta = -.290$ (95% CI [-0.55, −0.07]), $SE = .113$, $t = -2.573$, $p = .011$, as well as in the right hemisphere, $\beta = -.286$ (95% CI [-0.56, −0.07]), $SE = .112$, $t = -2.567$, $p = .012$, suggesting that individuals with more gyrification in the transverse temporal gyri are less consistent (or more variable) in their responses on the visual analogue scaling task. The second model predicted response consistency on the sign-shine continuum. Local gyrification in the left hemisphere negatively predicted response consistency, $\beta = -.294$ (95% CI [-0.49, −0.09]), $SE = .118$, $t = -2.485$, $p = .015$, as well as in the right hemisphere, $\beta = -.292$ (95% CI [-0.48, −0.09]), $SE = .117$, $t = -2.489$, $p = .014$. This also suggests that participants with more gyrification in this region were less consistent on the categorization task (see Fig. 5).

3.3.3. Exploratory analyses

The measure of response consistency is difficult to interpret on its own, and part of the motivation for including it was to better explain patterns of graded or categorical perception (i.e., are participants' categorization slopes shallower because they are less consistent in their responses or are they shallower because they can consistently perceive within-category differences in speech sounds?). Therefore, we divided participants into two groups of categorical and graded perceivers by a median split on the categoricity score for each continuum separately (ba-da continuum: categorical n = 28, graded n = 29, sign-shine continuum: categorical n = 28, graded n = 29). We first fit two exploratory models (one for each continuum) that predicted response consistency as the dependent variable and included fixed effects of categoricity group

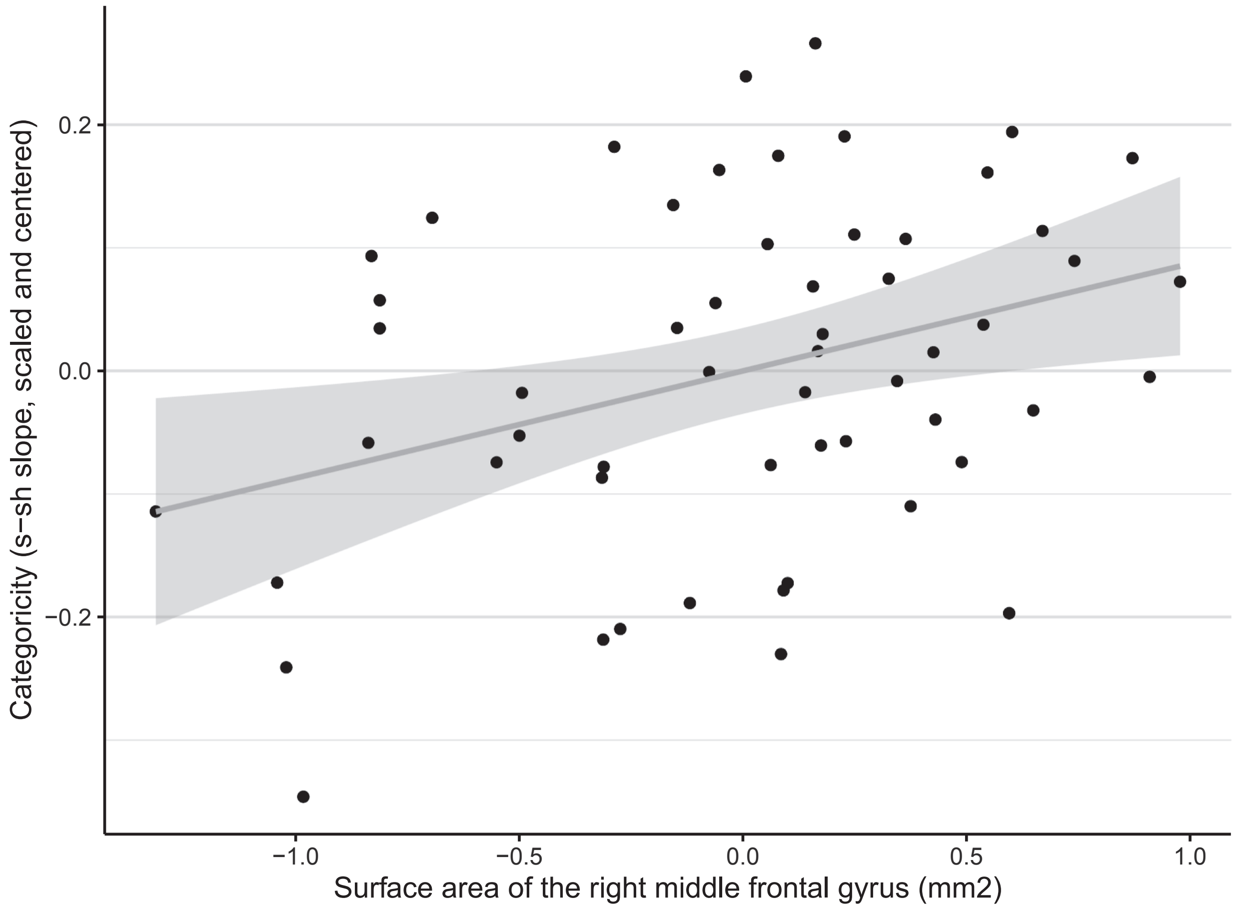

Fig. 4. Surface area of the right middle frontal gyrus predicted s-sh slope (categoricity) when holding other predictors constant.

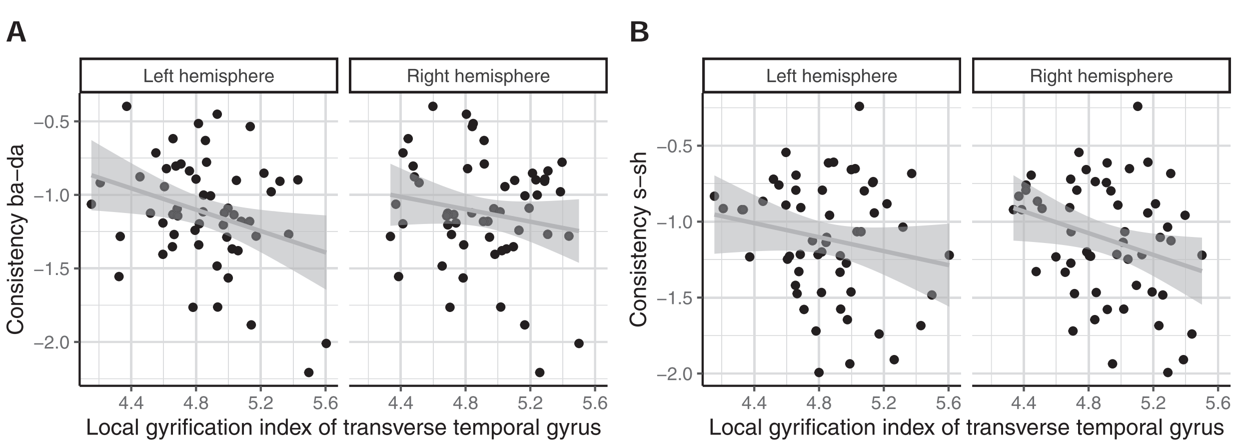

Fig. 5. Local gyrification index of the bilateral transverse temporal gyri negatively predicts response consistency on the A. ba-da and B. s-sh categorization tasks.

(deviation coded: categorical = -.5, graded =.5) and the interaction of hemisphere and gyrification index (as before), and their interaction. The interaction did not reach significance suggesting that the groups of categorical and graded perceivers did not significantly differ with respect to the relationships between gyrification and response consistency. However, to get an idea of the effect sizes of these relationships for each group, we fit two exploratory models predicting response consistency as the dependent variable with fixed effects of categoricity group (deviation coded: categorical = -.5, graded =.5) and the interaction of hemisphere and gyrification index nested within categoricity group. This allowed us to test simple effects of the relationship between gyrification and response consistency for each categoricity group separately (see e.g., Schad, Vasishth, Hohenstein, & Kliegl, 2020 for a detailed explanation of nested fixed effects). In the model predicting ba-da response consistency, we found that the local gyrification index in the left hemisphere negatively predicted response consistency for the categorical group, β = -.363 (95% CI [-0.71, −0.04]), SE =.178, t = −2.040, p =.04, as did the local gyrification index in the right hemisphere, β = -.358 (95% CI [-0.71, −0.05]), SE =.176, t = −2.037, p =.04. Effects went in the same direction for the graded group, but these effects were smaller. The model predicting response consistency on the sign-shine continuum additionally showed negative relationships between the local gyrification index and response consistency for both the categorical and graded group, but these effects did not reach significance. It is possible that we did not have enough power to detect those effects when grouping participants by categorical/graded perception.

4. Discussion

In the current study, we tested whether individual variability in brain structure is related to behavioral performance on native-language speech categorization tasks. Differences in the perception of speech sound categories have been seen in populations with speech and language disorders (Joanisse et al., 2000; Manis et al., 1997; Werker & Tees, 1987), but here we demonstrate that even within a typical population, there are substantial individual differences in sensitivity to variability along the phonetic continuum. Notably, some listeners show a graded pattern and were able to accurately rate differences between successive tokens along the continuum (Fig. 3E). In contrast, others show a more classical categorical pattern, treating all members of that category the same. We also introduced a new behavioral metric of sensitivity, "response consistency," which measures the degree to which listeners assign the same rating to the same token over successive trials. Response consistency allows us to distinguish between the listener who shows a shallower categorization slope because of inconsistent or stochastic responses to each token from a participant who shows a flatter slope but can accurately rate each token. Of particular interest is whether individual differences in brain *structure* associated with these perceptual profiles overlap with prior work on *functional* correlates of graded vs. categorical perception of speech.

4.1. Categoricity

We found that surface area of the right middle frontal gyrus predicted more categorical responses on a fricative continuum. No relationships were found, however, between brain structure and categoricity of the stop continuum. The findings from the fricative continuum parallel previous findings from the functional MRI literature, namely, that the middle frontal gyri (or adjacent regions) show categorical-like responses to native and non-native speech sounds (Chevillet, Jiang, Rauschecker, & Riesenhuber, 2013; Lee et al., 2012; Luthra et al., 2019; Myers, 2007; Myers et al., 2009). The fact that we saw this relationship with a right-hemisphere structure is consistent with theories that have proposed hemispheric specialization for shorter vs. longer integration windows (Boemio, Fromm, Braun, & Poeppel, 2005). Specifically, the ba/da distinction is marked by rapid (<40 ms) spectral sweeps, which would be predicted to rely more strongly on left-hemisphere systems, whereas the s/sh distinction is marked by longer (>150 ms) steady state spectral information, which would be predicted to recruit right hemisphere structures.

Though the findings from the fricative continuum are fairly straightforward to interpret, a unified interpretation is more difficult because we did not find similar relationships with brain structure for the stop continuum. There are a few potential explanations for this. First, speech sounds appear to differ in the degree to which they are perceived categorically, with stop sounds being reported as more categorical than vowels or fricatives (Eimas, 1963; Healy & Repp, 1982; Repp, 1981, see also Kronrod, Coppess, & Feldman, 2012 for discussion). In theory, a continuum that is more continuously perceived might offer more opportunities to explore individual differences in perception. Notably, however, no studies to our knowledge have directly compared the categoricity of stop vs. fricative continua using the methods described here. It is also possible that with a more sensitive measure of categoricity, we would see relationships between brain structure and the stop continuum. A more sensitive or automatic measure of categoricity could potentially be measured by eye tracking or a traditional visual analogue scale (without discrete points on the line as was used in the current study). In addition, we used a typical population in the current study, and it is possible that including listeners with reading or language disorders would have increased the between-participant variability on either or both continua. We suggest that for speech sound continua for which graded perception is more commonly observed, individual differences in brain structure can predict how graded or categorically an individual perceives the sounds. Nonetheless, it is an open question whether the failure to find brain relationships with the stop continuum is meaningfully related to neural differences in processing stop vs. fricative sounds, or whether the lack of a relationship was due to our specific task or stimuli.

We should also caution that this relationship was found in a multiple regression model with several other a priori selected regions of interest, which means this relationship was found when holding the other predictors constant. It is therefore entirely possible that, had we chosen slightly different regions of interest, we would have found different relationships. Nonetheless, surface area of the right middle frontal gyrus predicted unique variance in categoricity. Finally, we ran a number of tests in a somewhat exploratory manner, so we encourage future work to replicate this finding in a more confirmatory way.

The current findings suggest that frontal structures are not only involved in categorical perception of speech perception, but *variation* in their surface area can predict how categorically an individual perceives certain sounds. Our results are consistent with the view that innate or experience-driven differences in brain structure may drive differences in speech perception in the typical population. Furthermore, these differences may exist along a continuum with the structural differences that have been associated with developmental language and reading differences (e.g., Leonard et al., 2001; Romeo et al., 2018; Williams, Juranek, Cirino, & Fletcher, 2018).

4.2. Response consistency

We tested whether gyrification of the bilateral transverse temporal gyri predicted behavioral measures of response consistency. Our metric of response consistency provides an estimate of the reliability of the perception of a given token. This *behavioral* metric is conceptually similar to measures of *neural* response consistency found in electrophysiological studies of sound processing (e.g., Hornickel & Kraus, 2013; Omote, Jasmin, & Tierney, 2017). In these studies, neural response consistency is quantified by measuring the similarity of the evoked neural response to the same sound or speech token across presentations, and poor response consistency has been found in poorer readers (Lam, White-Schwoch, Zecker, Hornickel, & Kraus, 2017). In this sense, it may not be the precise nature of the brain's response to speech sounds that predicts larger differences in language and reading behavior, but rather the consistency or stability of the perceptual response to the same token over time. In the current study, we showed that response consistently was negatively related to gyrification in the bilateral transverse temporal gyri, and exploratory analyses suggest that this relationship was primarily driven by individuals who showed more categorical response patterns. In other words, increased gyrification in the transverse temporal gyri predicted less consistent responses, and this was especially so for more categorical perceivers. Several previous studies found that split or duplicate transverse temporal gyri were related to phonetic expertise (Golestani et al., 2011), faster phonetic learning (Golestani et al., 2007), and better non-native speech sound imitation (Turker et al., 2017). Based on these findings, we predicted that more gyrification of the transverse temporal gyrus in either hemisphere would predict more graded perception of native-language speech sounds and more consistent responses on the visual analogue scaling tasks. Instead, we found that gyrification negatively predicted response consistency on the discrete visual analogue scaling task, in that participants with more gyrification were less consistent with their responses. More generally, it is interesting that we found this *negative* relationship between brain structure and behavior because it suggests that "less is more" for certain tasks. This apparent discrepancy could have come about from the differences in methodology (i.e., using a continuous measure of gyrification rather than morphological differences in number of gyri), or it may instead be that the global gyrification measure does not closely relate to differences in gyral morphology (i.e., whether an individual has a single, duplicate, or split transverse temporal gyrus). As we discuss below, our findings may be relevant to work on the role of categorical perception in reading and language disorders.

Older studies in the field used two-alternative forced choice tasks to assess categorical perception (e.g., Liberman et al., 1957). Many of these studies interpreted shallower slopes on an identification task as evidence for more graded perception, a pattern found in younger children (Burnham et al., 1991) and individuals with reading or language disorders (e.g., Joanisse et al., 2000; Manis et al., 1997; Werker & Tees, 1987). However, more recent evidence using other methods or tasks suggest that adults actually show more graded perception than perhaps originally thought (Kapnoula et al., 2017; Kong & Edwards, 2016; McMurray et al., 2002) and that this perceptual gradiency develops slowly through adolescence (McMurray et al., 2018). This suggests that previous behavioral findings of shallower categorization slopes may be more indicative of noisy representations, which result in less reliable or less consistent responses (Hornickel & Kraus, 2013; Kapnoula et al., 2017; McMurray et al., 2018). In other words, it is possible that the findings from earlier studies showing shallower categorization slopes in individuals with reading and language disorders were actually measuring noisy representations or inconsistent response patterns on speech categorization tasks, rather than truly graded speech category representations.

If our measure of response consistency is indeed tapping into how noisy an individual's phonological representations are, our results may actually be compatible with previous research. First, there is evidence that having split or duplicate transverse temporal gyri is related to phonological dyslexia (Leonard et al., 2001), though this finding seems difficult to reconcile with those showing that split or duplicate transverse temporal gyri predict various measurements of phonetic or auditory expertise (Golestani et al., 2011; Golestani et al., 2007; Turker et al., 2017). Another study that used the local gyrification index that was used in the present study found that individuals with dyslexia had increased gyrification in certain brain regions, though not in temporal regions as we found here (Williams et al., 2018). Therefore, it is not unprecedented to find this pattern of "less is more" for gyrification. In light of new evidence (McMurray et al., 2018), we interpret previous behavioral work showing shallower phonetic categorization slopes in individuals with language or reading disorders as evidence of noisy phonological representations (Joanisse et al., 2000; Manis et al., 1997; Werker & Tees, 1987). Our exploratory analyses may be relevant to this discussion: We found that the inverse relationship between gyrification and response consistency was more driven by categorical perceivers than graded perceivers. If graded perception is optimal for adults, it is possible that the relationship between gyrification in auditory areas and response consistency is stronger for the suboptimal pattern of perception (i.e., more categorical perception). An interesting question for future research to address is whether gyrification in auditory areas predicts categorical perception or response consistency in individuals with reading or language disorders. These results should be interpreted with caution, however, as they were exploratory and the interaction that was originally tested did not reach significance. Assuming our measurement of response consistency used in the current study captures the degree of noise in an individual's phonological representations, we speculate that gyrification of the transverse temporal gyri may predict precision of speech category representations, even in a typical population. Skoe, Brody, and Theodore (2017) observed variation in the auditory brainstem response that was related to reading ability even among individuals with no history of reading or language disorders, and our results may be reflective of a similar pattern, in which gyrification of the transverse temporal gyri predicts subtle variation in reading or language ability in the typical population or the broader population more generally. We acknowledge, however, that our interpretations of these findings are speculative, and we hope that future research can more definitively answer the question of whether gyrification is related to categorical or consistent perception of speech.

4.3. Innate vs. experience-driven differences in brain structure

An interesting question is whether brain-behavior relationships are innate or whether they arise because of experience. Brain structure is influenced by both genetic factors and experience (Zatorre et al., 2012), but it is often difficult to know whether observed individual variation in brain structure is innate or experience-dependent. This is further complicated by the fact that until recently, many studies examining the relationship between brain structure and behavior have not distinguished among various measures of brain structure, such as gyrification, surface area, cortical thickness, and volume, and these might be differentially susceptible to environmental influences. Many previous studies have used volumetric procedures such as voxel-based morphometry to measure cortical volume, but because volume is the product of surface area and cortical thickness, it is not clear whether relationships found between behavior and volume were due to cortical thickness or surface area. The radial unit hypothesis posits that surface area and cortical thickness result from different genetic processes: Specifically, surface area is a result of the number of columns in the cerebral cortex, whereas cortical thickness results from the number of cells in the columns (Rakic, 1988). Several recent MRI and genetic studies support this hypothesis as well, showing that the genetic processes responsible for the development of surface area and cortical thickness are independent (Panizzon et al., 2009; Wierenga, Langen, Oranje, & Durston, 2014; Winkler et al., 2010).

In the current study, we found that surface area and gyrification predicted performance on the behavioral measures, but cortical thickness did not. It is clear that the development of both surface area and cortical thickness is susceptible to genetic influences, but it is less clear whether one or the other is more susceptible to environmental influences. Some preliminary evidence suggests that cortical thickness is less heritable than surface area (see preprint by Hofer et al., 2019), and other studies support this notion as well. For example, Piccolo et al. (2016) found that cortical thickness but not surface area was related to socio-economic status, suggesting environmental influences on cortical thickness. Other studies have found increases in cortical thickness after intensive foreign language learning (Mårtensson et al., 2012) and spatial navigation training (Wenger et al., 2012). Thus, it is possible that cortical thickness is more reflective of experience than surface area or gyrification. The present results are consistent with this idea: Our participants were all native speakers of English, so we assume they have had relatively similar amounts of experience with their native language. Therefore, the absence of a relationship between behavioral measures and cortical thickness is perhaps unsurprising if we assume that cortical thickness is indeed more experience-driven than surface area or gyrification. The malleability of surface area and cortical thickness is a complex issue, and ultimately future work will need to clarify more definitively whether one or the other is more affected by language experience. Taken together, our findings suggest that individual differences in categorical perception of speech likely arise because of neural variation that emerges very early in the neurodevelopmental timeline, rather than from experience-related factors.

5. Conclusion

The current study explored the structural neural correlates of categorical perception and consistency of responses on a categorical perception task at an individual level. Findings reported here complement the functional literature, in that structural measures of frontal regions positively predicted how categorically individuals perceived tokens on a fricative continuum. Gyrification of the bilateral transverse temporal gyri negatively predicted how consistently listeners responded on the categorization task, and we speculate that this may be related to subtle variation in reading or language ability in the population.

Funding

This material is based upon work supported in part by the National Science Foundation under grant DGE-1747486 to the University of Connecticut, and NSF BCS 1554510 to EBM. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to Dr. Betsy McCoach for help with the curve fitting analyses and to Dr. Peter Molfese for help with the gyrification preprocessing.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.bandl.2021.104919.

References

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage, 79*, 201–212.
Blumstein, S. E., Myers, E. B., & Rissman, J. (2005). The perception of voice onset time: An fmri investigation of phonetic category structure. *Journal of Cognitive Neuroscience, 17*(9), 1353–1366.
Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience, 8*(3), 389–395.
Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [computer program]. version 5.3. 51. Online: http://www.praat.org, 2.
Burnham, D. K., Earnshaw, L. J., & Clark, J. E. (1991). Development of categorical identification of native and non-native bilabial stops: Infants, children and adults. *Journal of Child Language, 18*(2), 231–260.
Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience, 13*(11), 1428.
Chevillet, M. A., Jiang, X., Rauschecker, J. P., & Riesenhuber, M. (2013). Automatic phoneme category selectivity in the dorsal auditory stream. *Journal of Neuroscience, 33*(12), 5208–5215.
Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804–809.
Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179–194.
Damasio, A. R., & Geschwind, N. (1984). The neural basis of language. *Annual Review of Neuroscience, 7*(1), 127–147.
Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience, 20*(7), 1174–1188.
Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage, 53*(1), 1–15.
Drouin, J. R., Theodore, R. M., & Myers, E. B. (2016). Lexically guided perceptual tuning of internal phonetic category structure. *The Journal of the Acoustical Society of America, 140*(4), EL307–EL313.
Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech, 6*(4), 206–217.
Feng, G., Gan, Z., Wang, S., Wong, P. C., & Chandrasekaran, B. (2018). Task-general and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral Cortex, 28*(9), 3241–3254.
Fischl, B. (2012). Freesurfer. *Neuroimage, 62*(2), 774–781.
Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.
Golestani, N. (2014). Brain structural correlates of individual differences at low-to high-levels of the language processing hierarchy: A review of new approaches to imaging research. *International Journal of Bilingualism, 18*(1), 6–34.
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.
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.
Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*(1), 133–159.
Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377–396.
Healy, A. F., & Repp, B. H. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 8*(1), 68.
Hofer, E., Roshchupkin, G. V., Adams, H. H. H., Knol, M. J., Lin, H., Li, S., & Medland, S. E. (2019). Genetic determinants of cortical structure (thickness, surface area and volumes) among disease free adults in the charge consortium. bioRxiv. Retrieved from https://www.biorxiv.org/content/early/2019/06/19/409649. doi: 10.1101/409649.
Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience, 33*(8), 3500–3504.
Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77*(1), 30–60.
Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1594.
Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics, 59*, 40–57.
Kronrod, Y., Coppess, E., & Feldman, N. (2012). A unified model of categorical effects in consonant and vowel perception. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 34).
Lam, S. S. Y., White-Schwoch, T., Zecker, S. G., Hornickel, J., & Kraus, N. (2017). Neural stability: A reflection of automaticity in reading. *Neuropsychologia, 103*, 162–167.
Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in broca's area: an fmri study using multivariate pattern-based analysis. *Journal of Neuroscience, 32*(11), 3942–3948.
Leonard, C. M., Eckert, M. A., Lombardino, L. J., Oakland, T., Kranzler, J., Mohr, C. M., & Freeman, A. (2001). Anatomical risk factors for phonological dyslexia. *Cerebral Cortex, 11*(2), 148–157.
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431.
Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358.
Luthra, S., Fuhrmeister, P., Molfese, P. J., Guediche, S., Blumstein, S. E., & Myers, E. B. (2019). Brain-behavior relationships in incidental learning of non-native phonetic categories. *Brain and Language, 198*, 104692.
Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Child Psychology, 66*(2), 211–235.
Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage, 63*(1), 240–244.
McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology, 54*(8), 1472.
McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42.
McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category vot affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language, 60*(1), 65–91.
Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes, 12*(5–6), 865–870.
Myers, E. B. (2007). Dissociable effects of phonetic competition and category typicality in a phonetic categorization task: An fmri investigation. *Neuropsychologia, 45*(7), 1463–1473.
Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science, 20*(7), 895–903.
Myers, E. B., & Mesite, L. M. (2014). Neural systems underlying perceptual adjustment to non-standard speech tokens. *Journal of Memory and Language, 76*, 80–93.
Myers, E. B., & Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience, 24*(8), 1695–1708.
Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex, 93*, 146–154.
Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., … Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex, 19*(11), 2728–2735. ISO 690.
Piccolo, L. R., Merz, E. C., He, X., Sowell, E. R., Noble, K. G., Pediatric Imaging, G. S., & Neurocognition. (2016). Age-related differences in cortical thickness vary by socioeconomic status. *PloS One, 11*(9), e0162511.
Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2019). R core team. 2019. nlme: linear and nonlinear mixed effects models. r package version 3.1-141. Available at h Ttp:// CRAN. R-Project. Org/Package= Nlme.
Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*(2), 285–290.

Price, C. J. (2012). A review and synthesis of the first 20 years of pet and fmri studies of heard speech, spoken language and reading. *Neuroimage, 62*(2), 816–847.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science, 241*(4862), 170–176.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics, 30*(3), 217–227.

Romeo, R. R., Christodoulou, J. A., Halverson, K. K., Murtagh, J., Cyr, A. B., Schimmel, C., & Gabrieli, J. D. (2018). Socioeconomic status and reading disability: Neuroanatomy and plasticity in response to intervention. *Cerebral Cortex, 28*(7), 2297–2312.

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*, 104038.

Schaer, M., Cuadra, M. B., Schmansky, N., Fischl, B., Thiran, J. P., & Eliez, S. (2012). How to measure cortical folding from mr images: A step-by-step tutorial to compute local gyrification index. *JoVE (Journal of Visualized Experiments), 59*, e3417.

Schremm, A., Novén, M., Horne, M., Söderström, P., van Westen, D., & Roll, M. (2018). Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain and Language, 176*, 42–47.

Skoe, E., Brody, L., & Theodore, R. M. (2017). Reading ability reflects individual differences in auditory brainstem function, even into adulthood. *Brain and Language, 164*, 25–31.

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing, 31*(3), 302.

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*, 2096.

Wenger, E., Schaefer, S., Noack, H., Kühn, S., Mårtensson, J., Heinze, H. J., & Lövdén, M. (2012). Cortical thickness changes following spatial navigation training in adulthood and aging. *Neuroimage, 59*(4), 3389–3397.

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology/Revue canadienne de psychologie, 41*(1), 48.

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *Neuroimage, 87*, 120–126.

Williams, V. J., Juranek, J., Cirino, P., & Fletcher, J. M. (2018). Cortical thickness and local gyrification in children with developmental dyslexia. *Cerebral Cortex, 28*(3), 963–973.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage, 53*(3), 1135–1146.

Wong, P. C., 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.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nature Neuroscience, 15*(4), 528–536.