

Neural sensitivity to syllable frequency and mutual information in speech perception and production

P. Tremblay¹, I. Deschamps¹, M. Baoni², & Uri Hasson²

¹Université Laval, Département de Réadaptation, Québec City, Qc., Canada

²Center for Mind/Brain Sciences (CIMEC), University of Trento, Trento Italy

Correspondence to: Pascale.tremblay@fmed.ulaval.ca

Author copy of manuscript accepted for publication in *NeuroImage*

Abstract—Many factors affect our ability to decode the speech signal, including its quality, the complexity of the elements that compose it, as well as their frequency of occurrence and co-occurrence in a language. Syllable frequency effects have been described in the behavioral literature, including facilitatory effects during speech production and inhibitory effects during word recognition, but the neural mechanisms underlying these effects remain largely unknown. The objective of this study was to examine, using functional neuroimaging, the neurobiological correlates of three different distributional statistics in simple 2-syllable nonwords: the frequency of the first and second syllables, and the mutual information between the syllables. We examined these statistics during nonword perception and production using a powerful single-trial analytical approach. We found that repetition accuracy was higher for nonwords in which the frequency of the first syllable was high. In addition, brain responses to distributional statistics were widespread and almost exclusively cortical. Importantly, brain activity was modulated in a distinct manner for each statistic, with the strongest facilitatory effects associated with the frequency of the first syllable and mutual information. These findings show that distributional statistics modulate nonword perception and production. We discuss the common and unique impact of each distributional statistic on brain activity, as well as task differences.

Keywords—

I. INTRODUCTION

Human verbal communication is accomplished through a complex series of neurocognitive and neuromotor processes involving the planning and production of speech sounds to form syllables, words and sentences. Several neurobiological and computational models assume that the syllable is a basic sublexical unit for speech processing and production (e.g., Bohland et al., 2010; Guenther & Vladusich, 2012; Guenther et al., 2006; Levelt, 1999; MacNeilage, 1998). Levelt (1999)

suggested that motor programs for the most frequent syllables are stored in a “mental syllabary”, whereas less frequent syllables are compiled from smaller units such as phonemes or diphones. Consistent with the notion of a role for syllables in speech comprehension and production, neuroimaging studies suggest an abstract representation of syllables that is independent of their particular acoustic features (Brendel et al., 2011; Evans & Davis, 2015; Otaka et al., 2008; Peeva et al., 2010; Siok et al., 2003). Neurobiological studies have documented sensitivity to the syllable or to syllabic structure manipulations in several brain areas including the supratemporal plane, the inferior frontal gyrus and precentral gyrus/sulcus (premotor cortex) during passive speech perception (e.g., Binder et al., 1994; Deschamps & Tremblay, 2014; Giraud & Price, 2001; Zatorre et al., 1992) and speech production (e.g., Bohland & Guenther, 2006; Riecker et al., 2000; Riecker et al., 2008; Riecker et al., 2005; Tremblay & Small, 2011).

Based on a lifetime of experience, adult speakers likely acquire a representation of the distributions of spoken single syllable frequency and of the degree of association between syllable pairs in their native language. Specifically, a large body of behavioral work has shown sensitivity to the frequency of single syllables (Alario et al., 2004; Carreiras & Perea, 2004; Cholin et al., 2011; Cholin et al., 2006; Laganaro & Alario, 2006; Levelt & Wheeldon, 1994). A separate body of work has documented a strong capacity, from childhood, to acquire such statistics (Newport & Aslin, 2004; Pelucchi et al., 2009a, 2009b; Pena et al., 2002; Saffran et al., 1996; Saffran et al., 1999). Interestingly, it has also been shown that, in adults with apraxia of speech, a disorder of motor programming, speech production errors reflect the influence of syllable-based processing mechanisms, with more errors in words containing a less frequent first syllable, consistent with impairment to the mental syllabary (Aichert & Ziegler, 2004; Staiger & Ziegler, 2008), and supporting the notion that sublexical distributional information is learned (we refer to this as distributional knowledge). Neurobiological studies have shown that adults are highly sensitive to distributional knowledge. Specifically, recent Magnetoencephalography (MEG) and Electroencephalography (ECoG) studies have used auditory speech as input and documented sensitivity to transition

This study was supported by a research grant from the European Research Council under the 7th framework starting grant program (ERC-STG #263318) to U. Hasson and by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC #1958126) and Social Sciences and Humanities Research Council of Canada (SSHRC #430-2013-1084) to P. Tremblay. The authors declare no competing financial interests.

probabilities (TP) between single phonemes (Leonard et al., 2015), sensitivity to phoneme “surprisal” given the cohort of phonemes heard to that point (Ettinger et al., 2014), and sensitivity to the range of potential upcoming phonemes (Cibelli et al., 2015; Ettinger et al., 2014). These studies focused on the temporal lobe via pre-defined ROIs (Ettinger et al., 2014) or a limited ECoG sampling space (Cibelli et al., 2015; Leonard et al., 2015). In a whole-brain fMRI study of spoken word comprehension, Vaden et al. (2011b) examined the neural correlates of biphone frequency. Their results showed that participants’ performance during a nonword detection task was poorer for sequences containing high frequency biphones. Higher BOLD signal was found for high frequency biphones only in the left inferior frontal gyrus (IFG), a finding that was replicated in a subsequent study by the same group (Vaden et al., 2011a). The authors took the absence of sensitivity to biphone frequency in lower level auditory regions to suggest that these regions are not involved in sublexical processes. Carreiras et al. (2006) using fMRI, found that low-frequency syllables were associated with greater brain activation as compared to high-frequency syllables in the left anterior insula during a reading aloud task (but see also Riecker et al., 2008 for null findings).

This neurobiological literature is consistent with pivotal behavioral studies that documented sensitivity to phonotactic probability. In these studies (Vitevitch et al., 1997; Vitevitch et al., 1999; Vitevitch & Luce, 1998), it was shown that participants process nonwords with higher phonotactic probability more fluently compared to those with lower probability. Interestingly, the opposite was found for words, suggesting that neighborhood effects are responsible for increased competition among lexical representations, which slows down word recognition. In addition, it was found that having participants consider nonwords as words reverses the facilitatory effect of frequency (Vitevitch et al., 1999), which is consistent with the inhibitory effect documented by Vaden et al. (Vaden et al., 2011a; 2011b).

Given this prior work, it is still unclear how distributional knowledge of syllable statistics impact processing at the whole brain level, and particularly whether there are brain regions that are sensitive to mutual constraints between syllables. Speaking to this issue, Leonard et al. (2015), using ECoG, found sensitivity to biphone transition probabilities in nonwords in left temporal regions (the right hemisphere was not examined). Interestingly, in some electrodes, higher TPs were associated with lower activity, consistent with both predictive coding accounts (Friston & Kiebel, 2009; Rao & Ballard, 1999), and the behavioral literature, but in others they were linked to higher activity. Furthermore, Leonard et al. (2015) found that not only forward TP (the probability of the next phoneme given the current) but also backward TP (the probability of the previous phoneme given the current) were tracked in temporal electrodes. This retrospective process is consistent with the notion that the temporal cortex integrates information over relatively long time periods and can use accumulating information for re-interpreting prior inputs (Shahin et al., 2009). In summary, the literature suggests that some brain regions are sensitive to language-specific sublexical distributional probabilities, with distinct responses depending on the type of stimuli used (words vs. nonwords). However, it is unclear whether distinct distributional statistics engage spatially distinct or overlapping brain networks, and whether

these effects (facilitatory vs. inhibitory) differ as a function of the task. There is reason to think that distributional knowledge of base-rate (marginal) frequencies and co-occurrence (associative) frequencies are mediated by partially separate systems. In a study of artificial auditory streams where marginal frequencies and co-occurrence frequencies were manipulated orthogonally, largely separate areas within the lateral temporal cortex were identified (Tobia et al., 2012). However, prior work within the speech domain has focused on either “surprisal effects” reflecting transitional probabilities (Ettinger et al., 2014; Leonard et al., 2015), or on marginal frequency effects (Cibelli et al., 2015). It is therefore unclear whether, during speech processing, different systems are sensitive to marginal frequencies and associative constraints.

The specific objectives of the current study were therefore twofold: (1) to compare the neurobiological correlates of three different sublexical distributional probabilities: first syllable frequency, second syllable frequency and mutual information; and (2) to determine whether the nature of sublexical distributional-probability effects (facilitation, inhibition) varies with task demands (here implemented via speech production vs. perception). Because we were interested in identifying sublexical processes that are not subservient to word-level effects, and since lexical status is known to interact with phonotactic-level frequency manipulations (Alvarez et al., 2001; Vitevitch, 2003; Vitevitch et al., 1999), we used legal nonwords (i.e., pseudowords consistent with the phonotactic constraints of Italian) throughout the study. We manipulated three syllable-level distributional statistics as estimated from a corpus: the log-transformed frequency of the first and second syllable in a disyllable nonword (e.g., /bi/ - /da/), and the pointwise mutual information (MI) of the syllable pairs, which is expressed as $\log [P(\text{syllable1} \mid \text{syllable2}) / P(\text{syllable1})]$. We used MI as it is a symmetric measure that captures the extent to which two events are mutually constraining, thereby allowing us to compare the correlates of predictability for single and multi-unit syllables. The MI metric differs from conditional probability, $P(\text{syl2} \mid \text{syl1})$ as it effectively normalizes the conditional probability by $P(\text{syl2})$. This is important because in the distributions of natural language there might be strong correlations between these two statistics (more frequent syllables are more likely to come after others; see Methods). Finally, we used a powerful single-trial fMRI analytic approach (Pernet et al., 2011), which allow for the examination of variation in brain activations in relation to parametric variations in stimuli properties.

Based on prior work, we expected that the correlation between brain activity and first syllable frequency (FS1) would be mainly negative (i.e., reduced activity with increased frequency) for speech production, reflecting the known behavioral facilitatory effect of syllable frequency on tasks involving nonword production. We expected to find these effects in regions known to be involved in sublexical processing, including the anterior insula, the supratemporal cortex, the premotor cortex, and inferior frontal gyrus and sulcus. For the second-syllable frequency effect (FS2), we expected a much more limited spatial distribution compared to FS1, as given any predictive framework to speech perception, the responses to the second syllable should depend not only on its base-rate frequency but also on the relative transition constraints between the two syllables which was quantified here via MI. For this reason, we expected that higher MI

should be strongly linked to lower activity reflecting a facilitatory effect. Because there is little previous research on the neural processing of FS2 and MI, we did not have specific hypotheses about the neural network that would be involved, though we expected the supratemporal cortex would be involved because of its known involvement in processing other distributional statistics such as transition probabilities at the phonemic level (Ettinger et al., 2014; Leonard et al., 2015). To determine whether there are task differences in the processing of distributional statistics, we also compared the neural processing of distributional statistics in nonword perception and production. Though this is the first study, to our knowledge, to examine task effects in statistical information processing, we expected similar networks to be engaged for perception and production, based on prior neuroimaging studies showing overlap in the brain activity for perception and production at the level of the cerebral cortex, in both premotor and auditory areas of the frontal and temporal lobes (Aziz-Zadeh et al., 2010; Buchsbaum et al., 2001; Okada & Hickok, 2006; Tremblay et al., 2013b; Tremblay et al., 2013c; Tremblay & Small, 2011; Zheng et al., 2010).

II. METHODS

A. Participants

Twenty-three healthy native Italian speakers were recruited for this study. Data of one participant were excluded due to technical failure, and data for two others were excluded due to artifact in the MRI data, leaving a final group of 20 adults (8 males; 24.4 ± 4.6 years, education: 15.7 ± 2.8 years), with normal self-reported hearing, and no history of language or neurological or neuropsychological disorders. 18 participants were right-handed and two were ambidextrous (mean \pm SD = $+71 \pm 26$) (Oldfield, 1971). The study was approved by the Ethical Review Board of the University of Trento in Italy.

B. Stimuli

The stimuli were 450 meaningless disyllabic nonwords that were controlled for both syllable complexity (only CV syllables were used) and suprasyllabic complexity (all nonwords contained two different CV syllables). The nonwords were created from a set of 119 simple consonant-vowel (CV) syllables selected from a database derived from itWaC, a large corpus (~ 1.5 billion words) of Italian online content (Baroni et al., 2009). The syllable database was obtained by automatically transcribing and syllabifying the corpus using a phonetic lexicon containing transcriptions of about 400,000 Italian words (Cosi & Avesani, 2001). Words that were present in the online content but not present in this lexicon were discarded. The resulting transcribed corpus contained more than 3 billion syllables. We used the syllable frequency information from this corpus to select syllables for the experiment. A total of 5405 unique CV syllables were coded in the database. For the present study, we selected a set of 119 CV syllables, which were composed of combinations of a subset of all Italian phonemes, including seven different vowels (/a, e, i, o, u, ε, ɔ/) and nineteen consonants (/b, d, f, g, j, k, l, λ, m, n, ɲ, p, r, s, ʃ, t, v, w, z/).

From this syllable set, we constructed 450 meaningless disyllabic nonwords (the lexical status of the nonwords was verified by two native Italian speakers). Each of the 119 different syllables was used at least once, with a mean \pm SD of 7 ± 4 occurrences. For each nonword, the frequency of the first

syllable (FS1), the frequency of the second syllable (FS2) and the pointwise mutual information (MI) were calculated from the corpus. We calculated these frequencies based on the overall distribution of a syllable in the corpus, independent of positional information within a word. While this may not appear sophisticated from a psycholinguistic perspective, from a neurobiological perspective, it loads directly on the notion of Hebbian learning of more frequently encountered syllables. In this initial study of the topic, we wanted to see if we would find a neural signature for this property. Because the frequency distributions of FS1 and FS2 were not normal, they were log transformed. Pointwise Mutual Information (MI) is a measure of the quantity of shared information by two variables (here FS1 and FS2) that is based on their joint probability distribution ($MI = \log[P(\text{syl}2 | \text{syl}1) / P(\text{syl}2)]$; or alternatively, $\log[P(\text{syl}1, \text{syl}2) / (P(\text{syl}1) * P(\text{syl}2))]$, and can be used to describe the statistical dependence between two elements (Gueguen et al., 2014). The nonwords were created via a repeated sampling procedure to construct a set of nonwords such that across the entire set, MI, $\log(\text{FS}1)$ and $\log(\text{FS}2)$ were uncorrelated. Note that in Italian, the correlation between $P(\text{syl}2)$ and $P(\text{syl}2 | \text{syl}1)$, as estimated from the corpus is positive ($r = 0.28$). The correlation between $P(\text{syl}1)$ and $P(\text{syl}1 | \text{syl}2)$ is also positive ($r = 0.27$). Put simply, more frequent syllables are more likely to both follow and precede others. Thus, it is difficult to dissociate base-rate correlates from those of transition probability. Instead, our use of MI obviates this problem, (dividing transition probability by base rate simply produces the MI expression).

To generate the auditory nonword stimuli, a native male Italian speaker from the North of Italy recorded the 119 syllables in a sound-attenuated booth. The speaker produced each syllable at least five times, always within a carrier sentence (“*adesso dico ...*”; translation: “*now I say ...*”). The best token of each syllable was used in the experiment, such that each syllable in the study was represented with a single token, following the procedure developed by Buiatti et al. (2009) and McNealy et al. (2006). The syllables were recorded at 44 KH using a unidirectional microphone connected to a professional amplifier, saved directly to disk using Sound Studio 3.5.4 (Felt Tip Software, NY, USA), edited offline using Wave Pad Sound Editor 4.53 (NHC Software, Canberra, Australia). All syllables were equalized in duration (275 ms), envelope (275 ms, ± 15 ms fade in, ± 15 ms fade out) and root mean square (RMS) intensity. After establishing the set of single-syllable sound files, syllable pairs were concatenated into disyllabic nonwords. It should be noted that this procedure prevented nonwords from being perceived as words in the sense that they contained no trace of coarticulation that could provide cues into the following syllable. Coarticulation can provide cues to upcoming phonemes/syllables, oftentimes well in advance of a future phoneme (e.g., Amerman et al., 1970; Daniloff & Moll, 1968; Goffman et al., 2008; Lehiste & Shockey, 1972; Ohala, 1993). However, in our study coarticulation only existed within a syllable, but not across. Thus, the typical coarticulation patterns in the VCV section of the bi-syllable CVCV were absent, which removed any predictive information that the first vowel typically carries about the second consonant. This allowed us to ensure that any effect of predictability was entirely related to statistics and not to acoustical cues in the signal (coarticulation). We provide a list of all stimuli in Supplementary Table 1 along with FS1, FS2 and MI statistics.

C. Tasks

The experiment included two experimental tasks: (1) passive speech perception (Perception), and (2) speech production (Production). Task instructions were pre-recorded by the same speaker who recorded the experimental stimuli and played back to all participants before beginning the fMRI session to aid in normalizing to speaker-specific pronunciation patterns. During Perception, participants listened to 225 disyllabic nonwords (CV-CV nonwords) passively presented through a high quality, digital passive noise-attenuation MRI-compatible stereo headset (Serene Sound, Resonance Technology Inc.). To ensure that participants maintained vigilance during Perception, they were instructed to press a response-box button each time the same syllable was repeated within a pair ("catch trials"; e.g., /baba/). There were a total of 45 catch trials, representing approximately 20% of all experimental trials.

During Production, participants were instructed to listen to and repeat the 225 disyllabic nonwords. As this task requires vigilance and a response, no catch trials were included. Two research assistants who were blind to the stimuli independently transcribed all recorded responses to the international phonetic alphabet (IPA). There was agreement between the two judges in 87.7% of all trials. In case of a disagreement, a third research assistant transcribed the response.

Each task was presented over the course of 3 functional runs. Each of the three Perception runs was composed of 75 experimental trials interleaved with 65 "jittered" short silence (rest) intervals and 15 Catch trials. Each of the Production runs was composed of 75 experimental trials interleaved with 45 short silence (rest) intervals. Within each run, the order of the conditions and the number of rest trials were optimally randomized for event-related designs using OPTseq2 (<http://surfer.nmr.mgh.harvard.edu/optseq/>). To avoid item-specific effects, two versions of the experiment were created (A, B). The stimuli used in Perception runs in Version A were used in Production in experiment B, and vice versa. Half participants underwent experiment version A and the other half underwent experiment version B.

D. Image acquisition

A 4T 8-coil Bruker system at the University of Trento was used to acquire high-resolution anatomical and functional data for each participant. Structural scans were acquired with a 3D T1-weighted MP-RAGE sequence (TR/TE=2700/4 ms, flip angle = 7°, isotropic voxel size = 1 mm³, matrix=256x224; 176 sagittal slices). Two structural volumes were obtained for all but two participants and averaged to allow more accurate image processing.

Single-shot EPI BOLD functional images were acquired using the point-spread-function distortion correction method (Zaitsev et al., 2004) and a standard sparse sampling acquisition whereby each volume was followed by 2 sec during which the gradients are turned off (the "silent period"). (We do note however, that even in this relative silence, the scanner environment still contains background noises, which could impact performance). Each functional EPI run began with six dummy scans to allow the magnetization to stabilize to a steady state. 825 functional images were acquired across 6 runs (TR/TE=3740/33 ms, 37 interleaved slices parallel to AC/PC, voxel size= 3x3x3.45, gap = 0.2 mm; matrix = 64x64; 1740 ms

of scan time followed by 2000 ms of silence). All stimuli were presented during the silent period, beginning 260 ms after the end of the volume acquisition to avoid temporal masking. There were three runs of Perception, and three runs of Production. In the Perception runs, 150 volumes were acquired in ~9 min, which included 75 experimental trials and 15 catch trials. These 90 trials were interleaved with rest intervals of 3.74 to 14.96 sec with an average of 5.4 sec. In the Production runs, 125 volumes were acquired in ~8 minutes, which included 75 experimental trials interleaved with rest intervals of 3.74 to 14.96 sec with an average of 5.3 sec. The trial timing is presented in Figure 1. The Perception runs were always carried out first to avoid priming production mechanisms. A relatively long TR was used to allow presentation of the auditory stimuli within a silent period (gradients off), and to record the responses.

E. Data analysis

2.5.1 Behavioral data analysis

In the Perception task, participants were asked to press a button whenever the syllable pair was identical. The number of correct responses, as well as the bias (d') and sensitivity (c) (Macmillan & Creelman, 1991) was calculated for each run and entered in a repeated-measure ANOVA with Run (1, 2, 3) as the within subject-factor to examine whether accuracy varied as a function of time.

In the Production task, accurate responses were defined as those in which participants repeated the correct nonword. However, we did not consider as errors those cases in which participants substituted open/closed /ə/ and open/closed /o/ vowels, as there is some freedom in the usage of these vowels in Italian (Bertinetto & Loporcaro, 2005). To examine the distribution of errors, we examined the phonemes that were mispronounced in a subset of the recordings (run 2). Phonemes were divided into 6 classes: vowels, plosives (/p/, /b/, /t/, /k/, /b/, /d/), fricatives (/f/, /v/, /s/, /ʃ/, /v/, /z/), nasals (/m/, /n/, /ɲ/), approximants (/l/, /j/, /w/, /ʎ/) and trills (/r/). A repeated-measure ANOVA was conducted on the number of errors with Run (1, 2, 3) as the within subject-factor to examine whether accuracy varied as a function of time.

To determine whether the syllable statistics affected speech production accuracy, we examined the relation between FS1, FS2 and M1 and response accuracy. The percentage of accurate responses was calculated for low and high FS1 nonwords, low and high FS2 nonwords, and low and high M1 nonwords using a median split procedure. The percentage of accurate responses was then compared, using single-tailed paired-sample t-tests. Unilateral tests were used because we hypothesized, based on the literature, that higher syllable frequency or MI would result in increased accuracy.

2.5.2 MRI data analysis

2.5.2.1. Pre-processing. All data were converted to the AFNI file format, and visually inspected for artefacts. All time series were spatially registered to the end of the third functional run (occurring half-way through the experiment), motion-corrected, time-shifted, de-spiked and mean-normalized using AFNI (Cox, 1996). All functional volumes acquired during excessive motion, defined as >1 mm, were excluded from the regression model using AFNI's censor function.

2.5.2.2 Individual-level analysis. For the Perception task, separate regressors were created for each participant for the experimental and catch trials. For Production, separate regressors were created for each participant for the correct and incorrect trials (regressors for incorrect trials were not considered in this work). To detect areas where the BOLD signal magnitude was modulated by the statistical properties of the stimuli, three amplitude-modulated parametric regressors were included in the statistical model. This analysis method identifies brain regions where BOLD amplitude varies with a continuous independent variable and can be used in event related designs (implemented via AFNI programs 1dMarry, and 3dDeconvolve with the `stim_times_AM2` option). The independent variables that were included to model BOLD responses were the log frequency of the first syllable in the nonword (FS1), the log frequency of the second syllable (FS2) and the pointwise mutual information of the syllable-pair (MI). Additional regressors were the mean, linear, and quadratic trend components, and the 6 motion parameters (*x*, *y*, *z* and roll, pitch and yaw). The basis function used to fit our statistical model and BOLD signal was a 1-parameter (fixed-shape regression) block shape (AFNI model BLOCK of length 3.74 sec [corresponding to trial duration]). The regression model was fit after concatenating the time series of all runs into a single time series (i.e., we did not estimate regression coefficients per run and then average those by condition).

We used SUMA (Saad et al., 2004) to import each participant's cortical surface representations created with the Freesurfer software package (Dale et al., 1999; Fischl et al., 1999; Fischl et al., 2004) into the AFNI 3D space and to project the results of the first-level (single-subject) analysis from the 3D volumes onto the 2D surfaces. Results of the first-level analysis were smoothed on the surface to achieve a target smoothing value of 6 mm using a Gaussian FWHM filter, and group-level analyses of the entire cortical surface were conducted on this 2D surface representation, as analyses in surface space achieve better cortical alignment particularly in perisylvian regions (Argall et al., 2006; Desai et al., 2005).

To analyze functional data within subcortical structures, in parallel to the surface analyses, we conducted a volume analysis. Prior to the regression analysis, a moderate spatial smoothing was applied to the functional data (3 mm FWHM Gaussian filter), in order to prevent reduction of signal in small subcortical regions by volume averaging with larger surrounding regions of inactivity such as the white matter (Crosson et al., 2003). The anatomical and functional datasets were then spatially normalized to the MNI TT_N27 template using the 12-parameter affine transform implemented in AFNI (@auto_t1rc program). The T1 image was first normalized to the template, and then the T2 images were normalized to the normalized T1 images. The same regression procedure was used to generate within-subject statistical images for each of the conditions for both the surface and volume data.

2.5.2.3. Group-level voxel-wise analyses. As a validity check, we examined the core network for the Perception and Production of nonwords by conducting a one-way repeated measures ANOVA on the BOLD signal, with task as the within-subject factor (Perception, Production correct trials only). This analysis constituted a validity check for the

paradigm and data and was expected to reveal well-established lateral temporal, pre-motor and inferior frontal regions.

The first analysis examined the relationship of each of the statistics of interest (FS1, FS2, MI) and BOLD signal magnitude using a series of contrasts. These were examined by first averaging data (Beta coefficients) from the Perception and Production tasks to maximize power, and then separately for each of the two tasks (for Production, only correct responses were included). These analyses generated six statistical maps identifying brain regions that track these probabilistic features, in the two tasks. Second, to identify differences in neural systems that track FS1 and FS2, within each task we compared the regression values of FS1 and FS2. This analysis addressed the theoretical question presented in the introduction of whether the purported impact of FS1 would qualitatively differ than that of FS2, as the latter may reflect responses already conditioned on the information provided by FS1. Finally, we computed a series of contrasts evaluating the impact of task on the differences between the FS1, FS2 and MI metrics. These task effects were directly compared across statistics using two high level interaction contrasts [((ProductionFS1 – ProductionFS2) – (PerceptionFS1 – PerceptionFS2)) and ((ProductionFS1 – ProductionMI) – (PerceptionFS1 – PerceptionMI))]. The purpose of these was to evaluate whether language statistics have different impact depending on the task used.

All the analyses were corrected for multiple comparisons to limit false-positive (Type 1) errors. For the surface results, the Monte Carlo simulation procedure implemented in Freesurfer was used to identify significant clusters of activated vertices, with a relatively-liberal individual vertex threshold of $p < .01$, corrected for multiple comparisons to achieve a family-wise error (FWE) rate of $p < .05$ (representing a cluster size of ≥ 96 vertices). For the volume analyses, the AFNI program 3dClustSim was used to achieve the same level of correction (clusters ≥ 9 voxels [243 mm]). Given a sample size of 20 participants, a voxel-level correction of .01 provides us with a power of .8 to detect medium to large effect sizes ($d = .76$), thereby balancing out Type I and Type II errors (for a discussion on this topic, see Bennett et al., 2009).

III. RESULTS

A. Catch Trials

The results for the Catch trials in Perception revealed a high mean accuracy ($92.07 \pm 6.31\%$; Run 1: 91.5%, Run 2: 91%, and Run 3: 90%), with a range of 69-97.8%, which shows that all participants were awake and alert during the Perception task. The distribution of responses (in terms of hits, false alarms and misses) is illustrated in Inline Supplementary Material S2A. We also calculated measures of sensitivity (overall $d' = 3.88 \pm .47$, Run 1: 3.53, Run 2: 3.59 and Run 3: 3.53) and bias (overall $c = .44 \pm .17$; Run 1: .32, Run 2: .33 and Run 3: .34). We examined if these two indexes varied as a function of Run. For sensitivity, we found that it did not ($F(2,44) = .177$, $p = .839$, $\eta^2 = .008$), and the same was found for bias ($F(2,44) = .14$, $p = .87$, $\eta^2 = .006$).

B. Speech production errors

The overall accuracy level for nonword repetition in Production condition was $68.2 \pm 9\%$ (Run 1: 67.36%, Run 2:

69.96%; Run 3 65.33%). To examine these errors in more detail, we coded a subset of the data (Run 2) for error types. This analysis revealed that 35% of all errors occurred on vowels and 65% on consonants. 62% of all errors occurred on the first syllable and 38% on the second. The distribution of errors is illustrated in Supplementary Material S2B, C and D.

To examine whether distributional statistics (FS1, FS2, MI) had an effect on nonword repetition accuracy, a series of t-tests was conducted, which revealed a facilitatory impact of FS1 (high > low) on accuracy ($t(21) = 3.15$, $p = .0048$; Cohen's $d = 1.3$), with a mean difference of 2.9%. No effect of FS2 or MI on accuracy was found. In summary, consistent with our hypothesis, there were fewer errors for the production of nonwords when FS1 was higher.

C. Core speech network

The core network for listening to and repeating disyllabic nonwords (as well as the overlap for listening and repeating) is shown in Inline Supplementary Material S3. Bilateral supratemporal, dorsal anterior insula (AI), lateral and medial premotor areas, cerebellum (mainly lobules V, VI, VII) and striatum were engaged during both Perception and Production, with more widespread activation for production than perception. Importantly, both Production and Perception were associated with activity not only in the supratemporal plane bilaterally, but also the left ventral premotor region and SMA.

D. Independent impact of FS1, FS2 and MI on the BOLD signal

Our main goal was to examine the impact of the three different syllable distributional statistics on brain activity during the perception and the production of nonwords. Most activations were found at the level of the cortex. As can be seen in Figure 2, most regions were sensitive to only one of the distributional statistics.

Among the three distributional statistics that we investigated, FS1 was associated with the strongest and most widespread effects on the BOLD signal, with cortical regions showing a significant effect extending over a total surface of 4077 mm² in the overall analysis (i.e. collapsing over task). In the vast majority of the regions identified, higher FS1 was associated with lower activity – that is, most effects reflected the expected negative correlation between BOLD amplitude and (log) syllable frequency. As shown in Inline Supplementary Materials S4 and S5 (which detail FS1 effects), negative correlations were found in the left ventral precentral gyrus, in the bilateral transverse temporal gyrus and sulcus (TTG, TTS), as well as in several parietal regions, including the bilateral precuneus, the bilateral dorsal postcentral sulcus/gyrus (PCG) and the bilateral superior parietal lobule. Interestingly, other brain regions showing negative correlations with FS1 included medial areas including the posterior, middle-section and anterior cingulate gyrus not typically associated with speech processing or production.

Insert Inline Supplementary Material S4 about here

In contrast to FS1 (see Inline Supplementary Materials 6 and 7), the overall FS2 effects on BOLD signal were extremely circumscribed, extending over a cortical area of only 108 mm² at the group level (same correction threshold as the FS1

analysis). Some of the correlations were some positive and some negative correlations. The overall analysis (shown in Figure 2) identified negative correlations in the left supratemporal plane and positive correlations in few other brain regions.

Insert Inline Supplementary Material S6 about here

The overall MI effects were negative (with an exception of a small cluster on the right superior frontal sulcus) (see Inline Supplementary Materials 8 and 9). That is, as expected, higher MI was linked to lower BOLD signal. These effects were less spatially circumscribed than those of FS2, but not as widespread as the FS1 effects, covering a surface area of 1771 mm² at the group level corrected maps. Areas showing this pattern in the overall analysis included parts of the bilateral ventral central sulcus/precentral gyrus/postcentral gyrus, left posterior STS, right ventral precentral sulcus, right PT, and bilateral posterior insula.

Insert Inline Supplementary Material S8 about here

An examination of the relative overlap of these effects (see Figure 2) points to the left and right supratemporal plane as convergence centers for sensitivity to these distributional statistics. In particular, in right superior temporal plane (Transverse Temporal Sulcus; TTS) we find overlapping effects of FS1 and FS2, as well as overlaps of FS1 and MI. In the left hemisphere, overlap of FS1 and MI effects were found in the medial TTG/posterior insula and PT. No other brain region showed overlap of effects. Focusing on these areas that tracked FS1 and MI in the supratemporal cortex, we derived for each participant the mean Beta for FS1 and MI (separately for the left and right supratemporal cortex). For the right hemisphere we found that participants that were more sensitive to FS1 were also more sensitive to MI (Pearson's $r = 0.54$, $p < .05$). However, this correlation was not significant for the left hemisphere. This might suggest a common computation in right supratemporal cortex underlying both FS1 and MI effects.

E. Differences between FS1 vs. FS2 within and across task

To quantify the differences between FS1 and FS2, we first contrasted the overall FS1 and FS2 beta values (see Figure 3 and Table 1A). Next, we examined the overlap between the statistics separately for each task (Figure 4AB), and, finally we contrasted the FS1 and FS2 beta values within each task (Figure 5AB and Table 1BC).

When collapsing across the two tasks (Figure 3) in all cases of significant differences, correlations were negative for FS1 and positive for FS2 (the opposite patterns was not found). Large clusters were found bilaterally in the precuneus, left TTG, right PT and the central sulcus bilaterally.

When examining each task separately, for speech perception (Figures 4A and 5A), we found several areas showed more negative correlations for FS1 than FS2, including the IPS bilaterally, left pericallosal sulcus, left posterior insula and left PCG. For speech production (Figure 4B and 5B), we

found several regions where correlations were more negative for FS1 than FS2, including the right posterior lateral fissure, right STG, left angular gyrus and posterior STS, left precuneus and the left central sulcus. The reverse pattern was also found (positive correlations with FS1 and negative correlations with FS2) but only in three regions, the right SMG, the right STS and the left ventral portion of the posterior IFG.

Supplementary Materials 4BC, 6BC and 8BC illustrate the effect of each of the statistics separately for Speech perception and Production.

F. Task modulation of the effects of FS1, FS2 and MI

Because one of our aims was to understand whether the impact of distributional statistics on brain activity is modulated by the task, we also examined whether there were task differences in the relationship between FS1, FS2 and MI with BOLD. As shown in Figure 6A and detailed in Table 2A, for FS1, task differences in magnitudes or signs of correlations were found in several cortical regions, but none in subcortical areas or in the cerebellum. In several regions, a negative relationship between FS1 and BOLD signal existed during Production, while a significant positive relationship (or no relationship) was found during Perception. This pattern was found in the bilateral dorsal central sulcus, left anterior medial frontal gyrus, right Sylvian fissure, and right anterior cingulate gyrus/sulcus. Thus, several regions showed facilitatory effects of FS1 but only during speech production. A task effect in the opposite direction (i.e. with a significant positive correlation between BOLD and FS1 in Production and a significant negative correlation in Perception, or no relationship) was found in the right anterior dorsal insula (Figure 6A), the right IPS, right angular gyrus and right ventral IFG (not shown in the figure).

For FS2, we found very circumscribed task differences (not shown in Figure, but see Table 2B). These were limited to the left superior occipital sulcus, where the beta values for Production were more negative than those for Perception.

Finally, for MI (Figure 6B and Table 2C), we found task differences in the bilateral medial prefrontal cortex/anterior cingulate gyrus and right PT. In all regions, MI values during Production were significantly negative, while they were not significantly different from zero during Perception (see Figure 6B for details of sample cluster in anterior cingulate). Thus, for the production task we found greater sensitivity to MI, in the predicted facilitatory direction, within these regions.

As discussed above, we found evidence suggesting that FS1 effects were more widespread than FS2 effects when collapsing across tasks. In a final analysis, we evaluated whether the differences between the statistics were themselves impacted by task. We compared task effects across statistics (FS1 vs. FS2, and FS1 vs. MI) using appropriate contrast terms (see Methods). As shown in Figure 6C, significant task effects on the difference between FS1 and FS2 were found in the bilateral IPS, right anterior insula and right SMG. In all regions, the difference was positive for Production (FS1>FS2) but negative or non-significant for Perception. Significant interactions were also found between FS1 and MI in the right anterior insula (ventrally and dorsally) and left temporal pole (not shown in the Figure). As can be seen in Figure 6D, in the insular regions, the difference was positive for Production (FS1>MI) but negative or non-significant for Perception. In the

temporal pole (not shown in the figure), the difference between FS1 and MI was negative in Production (FS1<MI) and non-significant during Perception.

To conclude, we found evidence of task differences on the processing of FS1 and MI, with a dominant pattern being greater sensitivity to statistics during speech Production. In addition, we found evidence that the difference between FS1 and FS2 correlates could themselves be task-related; such effects were found outside the supratemporal plane (IPS, bilaterally, left SMG, left insula).

IV. DISCUSSION

In the present study we used fMRI to conduct the first examination of the neurobiological correlates of three distributional statistics. We addressed this question using two different speech tasks and a trial-by-trial fMRI analytical approach. We found that the adult brain is sensitive to each of these distributional statistics. Furthermore, relatively distinct patterns of activation were found for each statistic, with the patterns found for FS1 and MI strongly supportive of mechanisms of distributional knowledge in the superior temporal cortex. Beyond these core findings, we identify three other important results: (1) though sensitivity was found for each statistic, the frequency of the first syllable (FS1) had the strongest facilitatory impact on BOLD signal, regardless of task, (2), most of the areas showing sensitivity to statistics were found within the ‘core’ speech system, but the majority of regions within that system were not sensitive to statistics, (3) sensitivity to syllable statistics showed task specificity within and outside the speech network.

A. Distributional knowledge in the syllabic domain

As reviewed in the introduction, there has been relatively limited neuroimaging work focusing on the brain regions or activity patterns linked to syllable distributional statistics as they occur in natural language (i.e. extracted from language corpora). Neuroimaging studies of biphone distributions have implicated the left IFG, with increased activity for more frequent phoneme pairs in a study using auditory words and a nonword detection task (Vaden et al., 2011b). However, an ECoG study has implicated the left supratemporal cortex in representing backward and forward transition probabilities between biphonemes (Leonard et al., 2015), and sensitivity to phoneme predictability has also been shown in MEG work focusing on left temporal cortex (Ettinger et al., 2014). Neuroimaging work examining syllable level coding (Carreiras & Perea, 2004; Riecker et al., 2008) has mainly used written words, where syllable-frequency effects may reflect a mix of orthographic and phonemic processes, and returned inconsistent findings. An exception is a study that evaluated the correlates of co-occurrence frequency of syllable and phonemes in spoken nonwords during covert rehearsal (Papoutsi et al., 2009), which found lower activity for high-frequency stimuli in left precentral gyrus, SMA and bilateral IFG, with no regions showing the opposite pattern.

In light of this prior work, the current findings speak to several issues. First, they demonstrate that natural syllable base-rate frequencies are in fact tracked within the supratemporal cortex bilaterally (even when other probabilistic cues, such as coarticulation, are not present), particularly for FS1. Importantly, using whole-brain imaging, we showed that frequency effects are not limited to the supratemporal cortex,

but are distributed across the cortex, including anterior and posterior medial regions, and sensorimotor cortices. Third, the strong MI effects, and relatively weak FS2 effects, suggest that processing of simple 2-syllable nonwords relies extensively on the knowledge of bi-syllable statistics, which is an outcome of long-term experience with a specific language. We refer to this as distributional knowledge. Below we develop each of these themes in detail.

4.1.1 Representations of syllable base-rates in the supratemporal cortex and beyond

One of the most striking patterns we found was the almost uniform negative correlation between FS1 and brain activity (we discuss task differences in section 4.2). The fact that more frequently encountered stimuli are more easily processed in the brain is a relatively well understood phenomenon (Dehaene et al., 2015) and has been examined via paradigms that manipulate frequency experimentally (Strange et al., 2005) as well as in studies examining responses to more vs. less frequent language stimuli (Cibelli et al., 2015; Frank et al., 2015; Fruchter et al., 2015; Papoutsi et al., 2009). While this demonstration of the FS1 effect is, in and of itself, important for functional/computational models of language (and consistent with our behavioral data), what is particularly important for neurobiological theories underlying speech comprehension is the set of regions showing sensitivity to FS1 in the current paradigm, and the difference between the FS1 and FS2 effects. FS1 effects were found in the supratemporal cortex bilaterally (posterior TTG, TTS and nearby PT) as well as middle frontal and posterior midline regions, with more modest involvement of the IFG and PCG (though we found effects in both; see Figure 2, and Supplementary Materials 4 and 6). These statistical maps subsume but are more extensive than those reported in prior findings, which reported either inferior frontal regions (Papoutsi et al., 2009; Vaden et al., 2011b), or left temporal regions (Cibelli et al., 2015) (the latter study examining only left temporal cortices).

A number of prior neuroimaging studies have suggested that abstract coding of syllable or phonetic identity is performed in left IFG, as well as lateral temporal regions. Using electrical stimulation mapping, Boatman et al. (Boatman et al., 1995) showed that direct stimulation of posterior-STG, mid-STG and left IFG introduce phonetic discrimination errors. Other work has shown that phonetic features such as place of articulation or voicing are tracked in the supratemporal cortex and nearby insula (Arsenault & Buchsbaum, 2015). Using a categorical-perception paradigm, Lee et al. (2012) reported evidence for categorical phonetic processing in left IFG and pre-SMA. Evans and Davis (2015), using a multivariate analysis, showed that somatomotor cortex (pre- and post-central gyri), but also bilateral mid-posterior STG and middle temporal gyrus (MTG), tracked syllable identity. However, in that last study areas around the primary auditory cortex did not track syllable identity but were driven by lower level acoustic features. In contrast, here we identified FS1 and MI effects around medial TTG/TTS bilaterally.

Other evidence for involvement of supratemporal regions in tracking frequency of auditory segments comes from a recent ECoG study by Cibelli et al. (2015). While limiting its scope to left temporal regions, that study showed that during the unfolding of auditory stimuli (words and nonwords), neural activity at each point in time tracks the cohort size in the

lexicon that matches the input until that point (i.e., the number of words [unique entries] in the lexicon matching the phonemes heard up to a given point). These regions also tracked the average cohort frequency (the mean frequency of words in the lexicon matching the phonemes heard up to a given point). Importantly, in both mid- and anterior-STG electrodes, stimuli with greater cohorts were associated with lower activity, suggesting less activation for more frequently encountered inputs. This was taken by the authors to suggest that “lexical competition and selection are encoded along the pathway of the auditory ventral stream, and do not exclusively rely on connections from other brain regions.” (p. 73). Similar findings were reported in an MEG study by Ettinger et al. (Ettinger et al., 2014). Our findings for FS1 are strongly consistent with these conclusions.

We also found FS1 effects in the posterior cingulate (pCing) – typically not reported in speech perception tasks (Vigneau et al., 2006) and even when reported, its activity is typically not interpreted (e.g., Chang et al., 2009; Deschamps & Tremblay, 2014). Yet, the pCing has been implicated in phoneme discrimination, along with posterior left IFG (Zatorre et al., 1996). Given that this region is a core node of the default mode network (Mazoyer et al., 2001; Raichle et al., 2001), its involvement in speech-related manipulations (e.g., processing of speech in noise) has been interpreted as reflecting a general function such as a switch in allocation of resources (Golestani et al., 2013). However, its involvement in the current study is of a different nature as we capture a parametric manipulation that is independent of gross sensory manipulations.

Whereas the findings for FS1 are largely consistent with prior ECoG and neuroimaging studies, we found a different pattern of results when examining FS2 (which should be considered together with the MI effects described below). In only one region, the right TTS, did we find a negative correlation between brain activity and FS2, whereas several regions showed a positive correlation. A direct contrast between FS1 and FS2 identified several regions, all of which showed more negative correlations for FS1 than FS2. When probing for mean regression slope values in several of these clusters we found statistically significant negative BOLD/frequency slopes for FS1, but not FS2.

These differences between the neurobiological correlates of FS1 and FS2 suggest that the extensive frequency-related effects found for FS1 do not reflect a bottom-up, context-independent correlate of syllable frequency. If that were the case, similar effects would have been found for FS2 as well. Rather, such effects are best thought of as utilization of prior (lifelong) distributional knowledge, (which is acquired through prolonged experience with a specific language), during processing of FS1, in the absence of any constraining prior phonetic context. In contrast, once the first syllable has effectively established a context, then the long-term mutual constraints between syllables play an increasingly important role, impacting the processing of the second syllable. Our account is consistent with prior explanations for the increased processing associated with word-initial syllables. In EEG studies, word-initial syllables are typically responded to more strongly than later ones (e.g., Astheimer & Sanders, 2011), and this has been interpreted in terms of paying attention to less-predictable segments of speech stimuli.

4.1.2 Representation of constraints

Our findings for pointwise Mutual Information (MI) speak to the neurobiological representation of mutual constraints between syllable-pairs that is independent of their base rate. MI reflects the (log) ratio between two events' joint probability (their collocation) and the product of their probabilities [$P(a,b)/(P(a)*P(b))$]. Note that MI can be formulated as $[P(a|b) / P(a)] = [P(b|a) / P(b)]$; i.e., it can be interpreted as indicating the additional information provided by one stimulus about the other, independent of the latter's base rate.

Several prior neuroimaging studies that have used simple artificial grammars have implicated perisylvian regions in coding statistical constraints between tonal, non-speech or non-lexical auditory stimuli (Karuza et al., 2013; McNealy et al., 2006; Tobia et al., 2012; Tremblay et al., 2013a). In several of those, the manipulation of transition probability was implemented while keeping the base-rates identical, meaning that greater transition probability amounted to greater MI (McNealy et al., 2006; Tobia et al., 2012; Tremblay et al., 2013a). Other studies have more specifically targeted long-term distributional knowledge, identifying many of the same regions. Leonard et al. (2015) used ECoG to evaluate correlates of phonotactic transition probabilities (TP) as estimated from corpora (in the left hemisphere only) and found signatures for both forward and backward TP in the supratemporal cortex. Similarly, Ettinger et al. (2014) used MEG to examine the neural correlates of phonotactic conditional probability [$P(\text{current phoneme} | \text{prior phonemes})$], and found such effects in three predefined ROIs in the left hemisphere (MTG, STG, TTG). In a study of adjective-noun combinations (Fruchter et al., 2015), activity in left lateral-temporal regions was negatively correlated with the frequency of both adjectives and nouns, and with the conditional probability, $P(\text{noun} | \text{adjective})$. We note, however, that in that latter study, $P(\text{noun} | \text{adjective})$ was positively correlated with $P(\text{noun})$ and negatively correlated with $P(\text{adjective})$, meaning that frequent nouns are ones more likely to follow prior adjectives, and conversely, more frequent adjectives are those followed by a more diverse cohort and therefore less predictive of any specific noun. As we noted in the Methods, this makes it difficult to interpret conditional probability metrics in relation to base-rate frequencies, and for this reason we used MI rather than TP.

Our findings for MI revealed a relatively well-defined set of lateral temporal, medial prefrontal and sensory-motor regions that negatively correlated with MI. This suggests that these regions benefit from knowledge of the syllable's co-occurrence patterns specifically, in a way that is independent of their base rates. This network consisted of posterior TTG bilaterally, right PT/TTS, right medial superior frontal gyrus and sulcus, and bilateral sensorimotor regions (pre- and post-central gyri). In addition, for regions showing sensitivity to FS1 and MI (in right supratemporal cortex) we found that participants who showed stronger sensitivity to FS1 also showed stronger sensitivity to MI. This might point to a common system that can take advantage of distributional knowledge in the auditory domain. It is interesting that "surprisal effects" in the lexical domain (i.e. unexpectedness of a word given prior context) do not appear to share this spatial distribution. In recent work, Willems et al. (2015) modeled the surprisal effect of lexical items given the recent past (defined using a corpus-derived Markov model $-\log(P(W_t | W_1 \dots W_{t-1}))$), and found that correlations between brain activity and

surprisal was found in bilateral lateral temporal regions largely excluding the TTG/TTS, and with only minor activity in the left PT (and no activity in left IFG). Thus, it may be that MI effects found here speak more specifically to sublexical processing, but this should be investigated in detail in future work.

B. Task modulation of the effects of FS1, FS2 and MI

As part of this study we examined the neurobiological correlates of syllable distributional statistics during two speech tasks (passive listening of clear speech [nonwords] vs. listening to and repeating nonwords). As was expected, both tasks engaged large bilateral segments of the frontal, temporal and insular cortices, as well as the striatum and cerebellum, with significant overlap across tasks in the supratemporal cortices bilaterally as well as differences (see supplementary Material S3). This is consistent with prior neuroimaging studies that provided evidence supporting the notion that the mechanisms for perception and production overlap at the level of the cerebral cortex, in both premotor and auditory areas (Aziz-Zadeh et al., 2010; Buchsbaum et al., 2001; Okada & Hickok, 2006; Tremblay et al., 2013b; Tremblay et al., 2013c; Tremblay & Small, 2011; Zheng et al., 2010), consistent with the notion of a bidirectional influence of sensory and motor systems on perception and production, and in line with a number of theoretical accounts (Lieberman et al., 1967; Lieberman & Mattingly, 1985; Schwartz et al., 2010; Skipper et al., 2007).

Perhaps surprisingly given the overlap between perceptual and motor processes, sensitivity to syllable statistics, particularly FS1 and MI, showed partially unique task-specificity. It is possible that differences in task demands were responsible for some of these effects, with the speech production task being more difficult than the speech perception task, requiring high concentration given the noisiness of the environment and the required fast response rate. In contrast, the speech perception task only required monitoring the repetition of a syllable within a nonword. Interestingly, task-sensitive statistical processing was found within and outside the "core" speech network. This may appear to challenge the notion of the perception/production unity. However, these findings are actually consistent with the behavioral literature on lexical processing and production, which shows that the effect of syllable frequency is dependent upon the nature of the task. Such studies have shown that for word production, the first syllable frequency effect is facilitatory, meaning that it is negatively correlated with response time (frequent syllables are produced faster than infrequent syllables) (Levelt, 2001; Levelt et al., 1999). This has been taken to reflect the presence of a mental syllabary that would contain the motor programs of the most frequent syllables in a language. In contrast, for speech perception and word recognition, syllable frequency effects are interfering, meaning that words containing frequent syllables are less rapidly recognized (e.g., Alvarez et al., 2001; Alvarez et al., 2000). And, as mentioned in the Introduction, different tasks can impact the extent to which sublexical frequency facilitates or interfered with task performance. It is therefore clear that, behaviorally, at least the first syllable frequency shows task-specific effects.

The finding of task differences in syllable distributional statistics is also consistent with the finding that syllable complexity affects speech perception and production in different ways. Tremblay and Small (2011) found that syllable

complexity, operationalized in terms of presence/absence of a consonant cluster, modulated activation in the premotor cortex during speech production but not during speech perception. It has also been shown that stimulating the premotor cortex only has an effect on complex speech perception task (Sato et al., 2009), suggesting that, though shared, brain regions involved in speech production may not be involved to the same degree during speech perception, especially simple ones such as the one used in the present study. In line with this idea, a recent meta-analysis has shown that distorted speech processing overlaps with speech production in the left pre-supplementary motor area and left anterior STS, but that regions of overlap are circumscribed (Adank, 2012). Thus, together with previous data, the current results suggest that the factors modulating sensitivity to distributional statistics are, at least in part, task-specific, and that, though perception and production appear to share at least part of their neural circuits, those circuits may not be engaged to the same extent or for the same purpose.

Alternatively, it is possible that access to underlying representations and distributional knowledge is not implemented to the same extent during Perception and Production. Correct performance on the speech repetition task that we used required a highly precise processing of complicated input (e.g., Trude & Brown-Schmidt, 2005), as speech is a complex and highly ambiguous signal. In contrast, correct performance on the perception task necessitated less effort, only requiring evaluating whether the stimulus contained a repeating syllable, which could be correctly achieved with less detailed acoustic/phonetic processing of the stimulus. It is possible that a principle of least effort may apply, whereby people seek the strategy that requires the minimal amount of work. In this view, only necessary operations would be conducted on the speech signal, meaning that the detail of motor/phonological representations may not be assessed in a simple passive perception task. This still speaks to differences between perception and production, but at a level that is not speech-specific but which pertains to the level of attention/manipulation required by the task. Finally, in the current study, it is very likely that stimuli in the perception task were misperceived on occasion, and therefore inaccurately modeled in our analysis. This may have reduced the sensitivity of the Perception task and could be, in part at least, responsible for task differences.

Based on previous studies, we expected to find syllable frequency effects in the ventral premotor cortex during speech production, a region in which speech sound maps are thought to be represented (Guenther, 2006; Guenther et al., 2006) and which is sensitive to syllable complexity during speech production (Tremblay & Small, 2011). Specifically, we expected the premotor cortex to show reduced activity for the production of highly frequent syllables reflecting facilitated access to more frequently used syllable motor programs. However, no effects for FS1, FS2 or MI were found in this region. Instead, we found sensitivity to FS1 in the anterior insula and IPS, as well as in the dorsal central sulcus. For MI we found task effects in the medial prefrontal cortex bilaterally as well as the in right PT. Though the insula is not usually associated with the processing of statistical information, in a recent study from our group, we found that for the left anterior insula, cortical thickness was positively associated with greater sensitivity, across individuals, to the statistical structure of syllable sequences (Deschamps et al., 2016). A few functional

neuroimaging studies have shown increased activation in this region during speech perception in noise as compared to non-degraded speech (i.e. without noise) (Adank et al., 2012; Poldrack et al., 2001), potentially reflecting greater monitoring needs for less intelligible speech. A role for the anterior insula as a cognitive/executive hub has been suggested (Chang et al., 2013; Menon & Uddin, 2010) in which the anterior insula is involved in detecting salient events for processing (Seeley et al., 2007). Consistent with this idea, in the current study, the modulation of activation within the right anterior insula could reflect different monitoring demands for speech perception and production, as well as for different statistics, with some possibly more salient than others. Further studies investigating the role of the insula during speech perception and production under different conditions (i.e. varying intelligibility and task difficulty systematically) are required in order to clarify the role of this region in speech and statistical information processing.

C. Future Directions

The current study suggests high sensitivity to syllable frequency and mutual information, but the way in which language is organized offers multiple perspectives on this issue. This is due to the fact that in language, there exist interesting relationships between statistical properties of single elements, conditional properties of pairs of elements, as well as even more complex statistics or groups of elements (e.g., triads). For instance, a more frequent syllable is defined by having a higher base rate frequency, $P(\text{syl})$, but more frequent syllables may also be less restrictive of the cohort of syllables that tend to appear after them (cohort entropy; Ettinger et al., 2014; Willems et al., 2015). For instance, in the Italian corpus that we used, the correlation between the (log) syllable frequency and cohort entropy (Shannon entropy of all immediately following syllables) is moderate but positive (Pearson's $r = 0.58$). While cohort effects appear to be weaker for nonwords than words (Cibelli et al., 2015), differentiating the relation between frequency and other information-theoretic features for sublexical stimuli is an important direction for future work (see Willems et al., (2015) for a related neurobiological study at the word level, and Cibelli et al. (2015) for modeling approach at the phonemic level).

Another issue that could be explored in future work is whether there are brain systems that are specifically sensitive to position information. As mentioned in the Methods section, we quantified overall syllable frequency from corpus, but it is definitely the case that certain syllables occur more frequently in certain positions within a word. To illustrate, when quantifying how frequently the syllables we used appear as the first or second in Italian words, we found only a moderate correlation between these values (Pearson's $r = 0.40$), speaking to relative positional asymmetry. It would therefore be interesting to know whether, when controlled for base-rate frequency, there is additional sensitivity specifically to positional frequency.

Finally, the role of coarticulation as providing predictive information is an interesting future investigation. Natural speech production includes anticipatory planning, often reflecting anticipation that is driven by quite remote future demands (Amerman et al., 1970; Daniloff & Moll, 1968; Goffman et al., 2008; MacNeilage & DeClerk, 1969). This phonetic information may offer additional information, which

may or may not be coded in the brain systems identified in the current study.

D. Summary and Conclusions

We examined the neurobiological correlates of three different syllable distributional statistics during the perception and production of nonwords: the first and second syllable frequency and mutual information of syllable pairs. Increased frequency of the first syllable and greater mutual information were associated with lower brain activity within and outside cortical regions typically involved in language processing, reflecting a facilitatory effect. By and large, most of the core speech network was insensitive to sublexical statistics. Nevertheless, our findings indicate that long-term distributional knowledge of syllable statistics, learned through prolonged exposition to language, is represented throughout a distributed set of regions, which includes the supratemporal cortex, where the most overlap between the different statistics was observed, as well as other frontal and parietal regions. This suggests that future work examining both lifelong and short-term distributional knowledge in the language domain could benefit from examining regions outside the perisylvian regions..

REFERENCES

- Adank, P. (2012). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain Lang*, 122(1), 42-54. doi: 10.1016/j.bandl.2012.04.014
- Adank, Patti, Davis, Matthew H, & Hagoort, Peter. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, 50(1), 77-84.
- Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and language*, 88(1), 148-159.
- Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behav Res Methods Instrum Comput*, 36(1), 140-155.
- Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: contrasting frequency effects in Spanish. *J Exp Psychol Learn Mem Cogn*, 27(2), 545-555.
- Alvarez, C.J., Carreiras, M., & de Vega, M. (2000). Syllable-frequency effect in visual word recognition: Evidence of sequential-type processing. *Psicologia*, 21, 341-374.
- Amerman, J. D., Daniloff, R., & Moll, K. L. (1970). Lip and jaw coarticulation for the phoneme (ae). *J Speech Hear Res*, 13(1), 147-161.
- Argall, Brenna D., Saad, Ziad S., & Beauchamp, Michael S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. *Human brain mapping*, 27(1), 14-27.
- Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *J Neurosci*, 35(2), 634-642. doi: 10.1523/JNEUROSCI.2454-14.2015
- Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia*, 49(12), 3512-3516. doi: 10.1016/j.neuropsychologia.2011.08.014
- Aziz-Zadeh, L., Sheng, T., & Gheytauchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS One*, 5(1), e8759. doi: 10.1371/journal.pone.0008759
- Baroni, M.S., Bernardini, A., Ferraresi, A., & Zanchetta, E. (2009). The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *Journal of Language Resources and Evaluation* 43(3), 209-226.
- Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Soc Cogn Affect Neurosci*, 4(4), 417-422. doi: 10.1093/scan/nsp053
- Bertinetto, P.M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, 35(2), 131-151.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., . . . et al. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology*, 35(6), 662-672. doi: 10.1002/ana.410350606
- Boatman, D., Lesser, R. P., & Gordon, B. (1995). Auditory speech processing in the left temporal lobe: an electrical interference study. *Brain Lang*, 51(2), 269-290. doi: 10.1006/brln.1995.1061
- Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *J Cogn Neurosci*, 22(7), 1504-1529. doi: 10.1162/jocn.2009.21306
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, 32(2), 821-841. doi: 10.1016/j.neuroimage.2006.04.173
- Brendel, B., Erb, M., Riecker, A., Grodd, W., Ackermann, H., & Ziegler, W. (2011). Do we have a "mental syllabary" in the brain? An FMRI study. *Motor Control*, 15(1), 34-51.
- Buchsbaum, B. R., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25, 663-678.
- Buiatti, M., Pena, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*, 44(2), 509-519. doi: 10.1016/j.neuroimage.2008.09.015
- Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Hum Brain Mapp*, 27(12), 963-972. doi: 10.1002/hbm.20236
- Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: effects of syllable frequency. *Brain Lang*, 90(1-3), 393-400. doi: 10.1016/j.bandl.2003.12.003
- Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. *Cereb Cortex*, 23(3), 739-749. doi: 10.1093/cercor/bhs065
- Chang, S. E., Kenney, M. K., Loucks, T. M., Poletto, C. J., & Ludlow, C. L. (2009). Common neural substrates support speech and non-speech vocal tract gestures. *Neuroimage*, 47(1), 314-325. doi: 10.1016/j.neuroimage.2009.03.032
- Cholin, J., Dell, G. S., & Levelt, W. J. (2011). Planning and articulation in incremental word production: syllable-frequency effects in English. *J Exp Psychol Learn Mem Cogn*, 37(1), 109-122. doi: 10.1037/a0021322
- Cholin, J., Levelt, W. J., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99(2), 205-235.
- Cibelli, E. S., Leonard, M. K., Johnson, K., & Chang, E. F. (2015). The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. *Brain Lang*, 147, 66-75. doi: 10.1016/j.bandl.2015.05.005
- Cosi, P., & Avesani, C. (2001). FESTIVAL Speaks Italian. Paper presented at the Proceedings Eurospeech 2001, Aalborg, Denmark.
- Cox, Robert W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research*, 29(3), 162-173.
- Crosson, B., Benefield, H., Cato, M. A., Sadek, J. R., Moore, A. B., Wierenga, C. E., . . . Briggs, R. W. (2003). Left and right basal ganglia and frontal activity during language generation: contributions to lexical, semantic, and phonological processes. *J Int Neuropsychol Soc*, 9(7), 1061-1077. doi: 10.1017/S135561770397010X
- Dale, Anders M., Fischl, Bruce, & Sereno, Martin I. (1999). Cortical surface-based Analysis: I. Segmentation and surface reconstruction. *Neuroimage*, 9(2), 179-194.
- Daniloff, R., & Moll, K. (1968). Coarticulation of lip rounding. *J Speech Hear Res*, 11(4), 707-721.

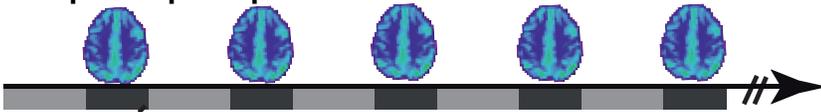
34. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron*, 88(1), 2-19. doi: 10.1016/j.neuron.2015.09.019
35. Desai, R., Liebenthal, E., Possing, E. T., Waldron, E., & Binder, J. R. (2005). Volumetric vs. surface-based alignment for localization of auditory cortex activation. *Neuroimage*, 26(4), 1019-1029. doi: 10.1016/j.neuroimage.2005.03.024
36. Deschamps, I., Hasson, U., & Tremblay, P. (2016). The Structural Correlates of Statistical Information Processing during Speech Perception. *PLoS One*, 11(2), e0149375. doi: 10.1371/journal.pone.0149375
37. Deschamps, I., & Tremblay, P. (2014). Sequencing at the syllabic and supra-syllabic levels during speech perception: an fMRI study. *Front Hum Neurosci*, 8, 492. doi: 10.3389/fnhum.2014.00492
38. Ettinger, A., Linzen, T., & Marantz, A. (2014). The role of morphology in phoneme prediction: evidence from MEG. *Brain Lang*, 129, 14-23. doi: 10.1016/j.bandl.2013.11.004
39. Evans, S., & Davis, M. H. (2015). Hierarchical Organization of Auditory and Motor Representations in Speech Perception: Evidence from Searchlight Similarity Analysis. *Cereb Cortex*. doi: 10.1093/cercor/bhv136
40. Fischl, Bruce, Sereno, Martin I., & Dale, Anders M. (1999). Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, 9(2), 195-207.
41. Fischl, Bruce, van der Kouwe, Andre, Destrieux, Christophe, Halgren, Eric, Segonne, Florent, Salat, David H., . . . Dale, Anders M. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1), 11-22. doi: 10.1093/cercor/bhg087
42. Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang*, 140, 1-11. doi: 10.1016/j.bandl.2014.10.006
43. Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1521), 1211-1221. doi: 10.1098/rstb.2008.0300
44. Fruchter, J., Linzen, T., Westerlund, M., & Marantz, A. (2015). Lexical preactivation in basic linguistic phrases. *J Cogn Neurosci*, 27(10), 1912-1935. doi: 10.1162/jocn_a_00822
45. Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *J Cogn Neurosci*, 13(6), 754-765. doi: 10.1162/08989290152541421
46. Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The breadth of coarticulatory units in children and adults. *J Speech Lang Hear Res*, 51(6), 1424-1437. doi: 10.1044/1092-4388(2008/07-0020)
47. Golestani, N., Hervais-Adelman, A., Obleser, J., & Scott, S. K. (2013). Semantic versus perceptual interactions in neural processing of speech-in-noise. *Neuroimage*, 79, 52-61. doi: 10.1016/j.neuroimage.2013.04.049
48. Gueguen, L., Velasco-Forero, S., & Soille, P. (2014). Local Mutual Information for Dissimilarity-Based Image Segmentation. *J Math Imaging Vis*, 48, 625-644. doi: DOI 10.1007/s10851-013-0432-9
49. Guenther, F. H., & Vladusich, T. (2012). A Neural Theory of Speech Acquisition and Production. *J Neurolinguistics*, 25(5), 408-422. doi: 10.1016/j.jneuroling.2009.08.006
50. Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. *J Commun Disord*, 39(5), 350-365. doi: 10.1016/j.jcomdis.2006.06.013
51. Guenther, F.H., Ghosh, S.S., & Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain Lang*, 96(3), 280-301. doi: S0093-934X(05)00115-X [pii] 10.1016/j.bandl.2005.06.001
52. 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 Lang*, 127(1), 46-54. doi: 10.1016/j.bandl.2012.11.007
53. Laganaro, M., & Alario, F. X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, 55, 178-196.
54. 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. *J Neurosci*, 32(11), 3942-3948. doi: 10.1523/JNEUROSCI.3814-11.2012
55. Lehiste, I., & Shockey, L. (1972). On the perception of coarticulation effects in English VCV syllables. *J Speech Hear Res*, 15(3), 500-506.
56. Leonard, M. K., Bouchard, K. E., Tang, C., & Chang, E. F. (2015). Dynamic encoding of speech sequence probability in human temporal cortex. *J Neurosci*, 35(18), 7203-7214. doi: 10.1523/JNEUROSCI.4100-14.2015
57. Levelt, W. J. (1999). Models of word production. *Trends Cogn Sci*, 3(6), 223-232.
58. Levelt, W. J. (2001). Spoken word production: a theory of lexical access. *Proc Natl Acad Sci U S A*, 98(23), 13464-13471.
59. Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav Brain Sci*, 22(1), 1-38; discussion 38-75.
60. Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50(1), 239-269.
61. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol Rev*, 74(6), 431-461.
62. Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36. doi: 0010-0277(85)90021-6 [pii]
63. Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.
64. MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behav Brain Sci*, 21(4), 499-511; discussion 511-446.
65. MacNeilage, P. F., & DeClerk, J. L. (1969). On the motor control of coarticulation in CVC monosyllables. *J Acoust Soc Am*, 45(5), 1217-1233.
66. Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houde, O., . . . Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res Bull*, 54(3), 287-298.
67. McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci*, 26(29), 7629-7639. doi: 10.1523/JNEUROSCI.5501-05.2006
68. Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Struct Funct*, 214(5-6), 655-667. doi: 10.1007/s00429-010-0262-0
69. Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn Psychol*, 48(2), 127-162. doi: S0010028503001282 [pii]
70. Ohala, J. J. (1993). Coarticulation and phonology. *Lang Speech*, 36 (Pt 2-3), 155-170.
71. Okada, K., & Hickok, G. (2006). Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain Lang*, 98(1), 112-117.
72. Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.
73. Otaka, Yohei, Osu, Rieko, Kawato, Mitsuo, Liu, Meigen, Murata, Satoshi, & Kamitani, Yukiyasu. (2008). Decoding Syllables from Human fMRI Activity. In M. Ishikawa, K. Doya, H. Miyamoto & T. Yamakawa (Eds.), *Neural Information Processing* (Vol. 4985, pp. 979-986): Springer Berlin Heidelberg.
74. Papoutsis, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. *Cereb Cortex*, 19(9), 2156-2165. doi: 10.1093/cercor/bhn239
75. Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J. L., Nazarian, B., & Alario, F. X. (2010). Distinct

- representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage*, 50(2), 626-638. doi: S1053-8119(09)01354-8 [pii] 10.1016/j.neuroimage.2009.12.065
76. Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009a). Learning in reverse: eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244-247. doi: S0010-0277(09)00182-6 [pii] 10.1016/j.cognition.2009.07.011
 77. Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009b). Statistical learning in a natural language by 8-month-old infants. *Child Dev*, 80(3), 674-685. doi: CDEV1290 [pii] 10.1111/j.1467-8624.2009.01290.x
 78. Pena, M., Bonatti, L. L., Nespore, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298(5593), 604-607. doi: 10.1126/science.1072901
 79. Pernet, C. R., Sajda, P., & Rousselet, G. A. (2011). Single-trial analyses: why bother? *Front Psychol*, 2, 322. doi: 10.3389/fpsyg.2011.00322
 80. Poldrack, Russell, Temple, Elise, Protopapas, Athanassios, Nagarajan, Srikantan, Tallal, Paula, Merzenich, Michael M, & Gabrieli, John DE. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. *Cognitive Neuroscience, Journal of*, 13(5), 687-697.
 81. Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A*, 98(2), 676-682. doi: 10.1073/pnas.98.2.676
 82. Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79-87. doi: 10.1038/4580
 83. Riecker, A., Ackermann, H., Wildgruber, D., Meyer, J., Dogil, G., Haider, H., & Grodd, W. (2000). Articulatory/phonetic sequencing at the level of the anterior perisylvian cortex: a functional magnetic resonance imaging (fMRI) study. *Brain Lang*, 75(2), 259-276. doi: 10.1006/brln.2000.2356
 84. Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain Lang*, 107(2), 102-113. doi: S0093-934X(08)00012-6 [pii] 10.1016/j.bandl.2008.01.008
 85. Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., & Ackermann, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, 64, 700-706.
 86. Saad, ZS, Reynolds, RC, Argall, B, Japee, S , & Cox, RW. (2004). SUMA: An interface for surface-based intra-and inter-subject analysis with AFNI. *IEEE Int Symp Biomed Imaging*, 1510-1513.
 87. Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.
 88. 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. doi: S0010-0277(98)00075-4 [pii]
 89. Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and language*, 111(1), 1-7. doi: 10.1016/j.bandl.2009.03.002
 90. Schwartz, J., Basirat, A., L., Ménard, & Sato, M. (2010). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception *Journal of NeuroLinguistics*.
 91. Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci*, 27(9), 2349-2356. doi: 10.1523/JNEUROSCI.5587-06.2007
 92. Shahin, A. J., Bishop, C. W., & Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage*, 44(3), 1133-1143. doi: 10.1016/j.neuroimage.2008.09.045
 93. Siok, W. T., Jin, Z., Fletcher, P., & Tan, L. H. (2003). Distinct brain regions associated with syllable and phoneme. *Human brain mapping*, 18(3), 201-207. doi: 10.1002/hbm.10094
 94. Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17(10), 2387-2399. doi: bh1147 [pii] 10.1093/cercor/bhl147
 95. Staiger, A. , & Ziegler, W. (2008). Syllable frequency and syllable structure in the spontaneous speech production of patients with apraxia of speech. *Aphasiology*, 22(11), 1201-1215.
 96. Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Netw*, 18(3), 225-230. doi: S0893-6080(05)00006-7 [pii] 10.1016/j.neunet.2004.12.004
 97. Tobia, M. J., Iacovella, V., & Hasson, U. (2012). Multiple sensitivity profiles to diversity and transition structure in non-stationary input. *Neuroimage*, 60(2), 991-1005. doi: 10.1016/j.neuroimage.2012.01.041
 98. Tremblay, P., Baroni, M., & Hasson, U. (2013a). Processing of speech and non-speech sounds in the supratemporal plane: auditory input preference does not predict sensitivity to statistical structure. *Neuroimage*, 66, 318-332. doi: 10.1016/j.neuroimage.2012.10.055
 99. Tremblay, P., Deschamps, I., & Gracco, V. L. (2013b). Regional heterogeneity in the processing and the production of speech in the human planum temporale. *Cortex; a journal devoted to the study of the nervous system and behavior*, 49(1), 143-157. doi: 10.1016/j.cortex.2011.09.004
 100. Tremblay, P., Dick, A. S., & Small, S. L. (2013c). Functional and structural aging of the speech sensorimotor neural system: functional magnetic resonance imaging evidence. *Neurobiol Aging*, 34(8), 1935-1951. doi: 10.1016/j.neurobiolaging.2013.02.004
 101. Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *Neuroimage*, 57(4), 1561-1571. doi: 10.1016/j.neuroimage.2011.05.067
 102. Trude, A., & Brown-Schmidt, S. (2005). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27(7-8), 979-1001.
 103. Vaden, K. I., Jr., Kuchinsky, S. E., Keren, N. I., Harris, K. C., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2011a). Inferior frontal sensitivity to common speech sounds is amplified by increasing word intelligibility. *Neuropsychologia*, 49(13), 3563-3572. doi: 10.1016/j.neuropsychologia.2011.09.008
 104. Vaden, K. I., Jr., Piquado, T., & Hickok, G. (2011b). Sublexical properties of spoken words modulate activity in Broca's area but not superior temporal cortex: implications for models of speech recognition. *J Cogn Neurosci*, 23(10), 2665-2674. doi: 10.1162/jocn.2011.21620
 105. Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., . . . Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage*, 30(4), 1414-1432.
 106. Vitevitch, M. S. (2003). The influence of sublexical and lexical representations on the processing of spoken words in English. *Clin Linguist Phon*, 17(6), 487-499.
 107. Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Lang Speech*, 40 (Pt 1), 47-62.
 108. Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain Lang*, 68(1-2), 306-311. doi: 10.1006/brln.1999.2116
 109. Vitevitch, M.S., & Luce, P.A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325-329. doi: 10.1111/1467-9280.00064
 110. Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2015). Prediction During Natural Language Comprehension. *Cereb Cortex*. doi: 10.1093/cercor/bhv075
 111. Zaitsev, M., Hennig, J., & Speck, O. (2004). Point spread function mapping with parallel imaging techniques and high acceleration factors: fast, robust, and flexible method for echo-planar imaging distortion correction. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 52(5), 1156-1166. doi: 10.1002/mrm.20261
 112. Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256(5058), 846-849.
 113. Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb Cortex*, 6(1), 21-30.

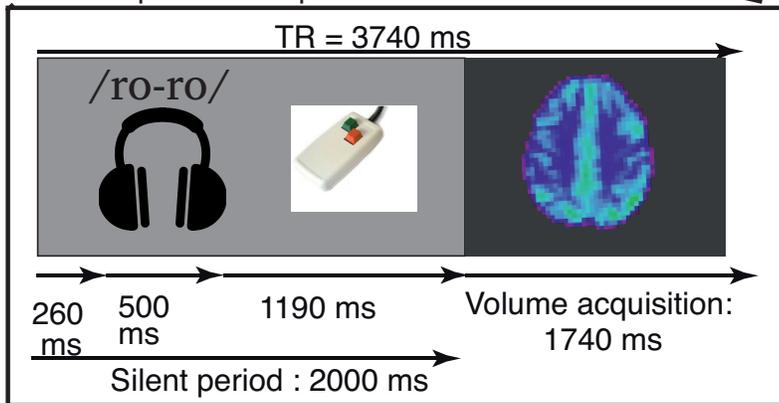
115. Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in

monitoring one's own voice during speech production. *J Cogn Neurosci*, 22(8), 1770-1781. doi: 10.1162/jocn.2009.21324.

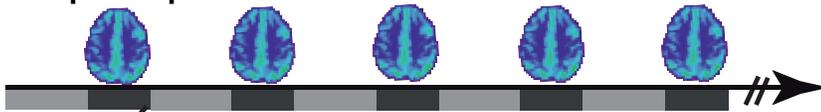
A. Speech perception



Example of an experimental trial



B. Speech production



Example of an experimental trial

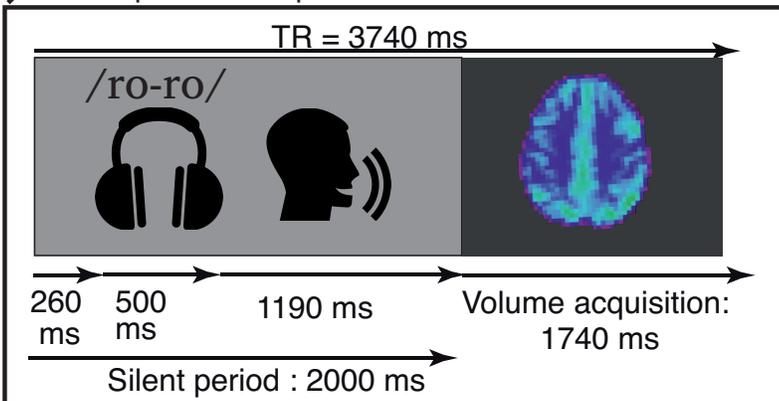


Figure 1. Schematic representation of trial structure for the perception task (A) and the production task (B). Note that trials were interspersed with silent “rest” intervals.

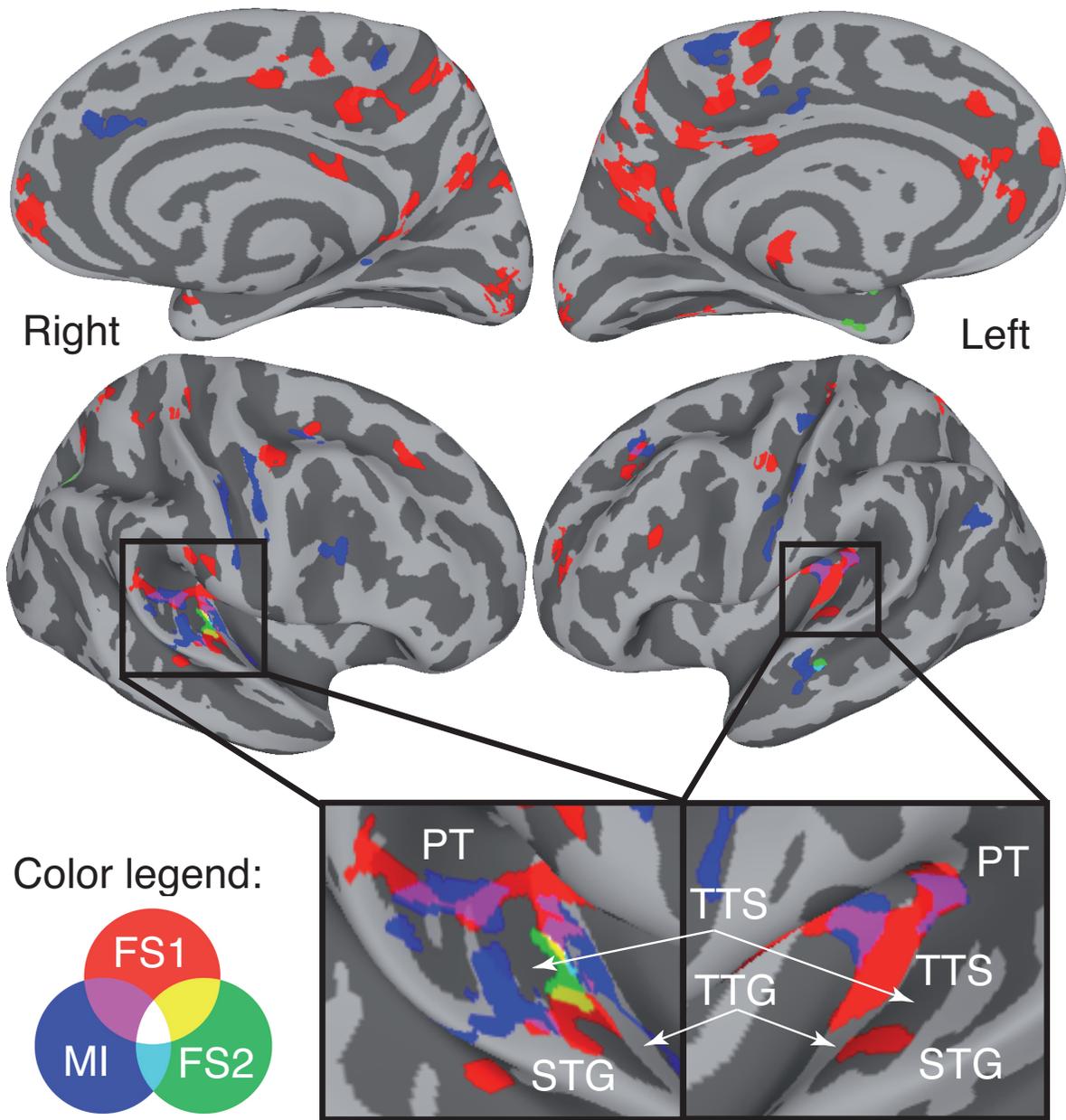


Figure 2. Group-level correlations between FS1, FS2 and MI and brain activation. Corrected group results for each statistic and relative overlap are shown on the group average smoothed white matter unfolded lateral and medial surfaces.

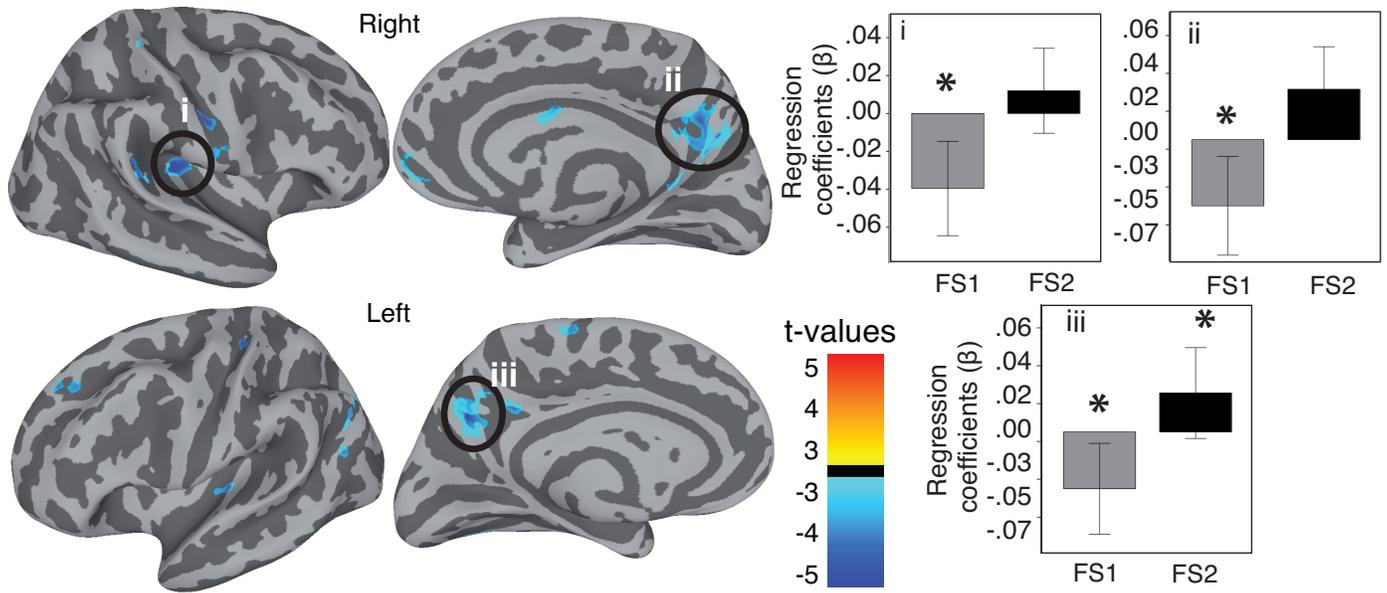
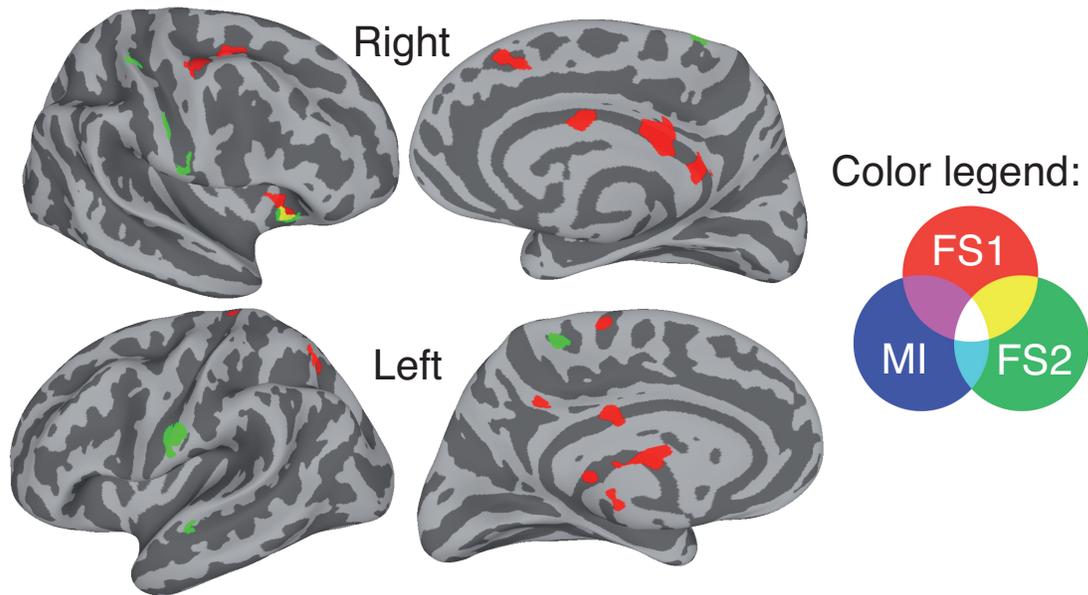


Figure 3. Group-level differences between FS1 and FS2 collapsed across tasks. Corrected group results are shown on the group average smoothed white matter unfolded lateral and medial surfaces.

A. Speech perception



B. Speech production

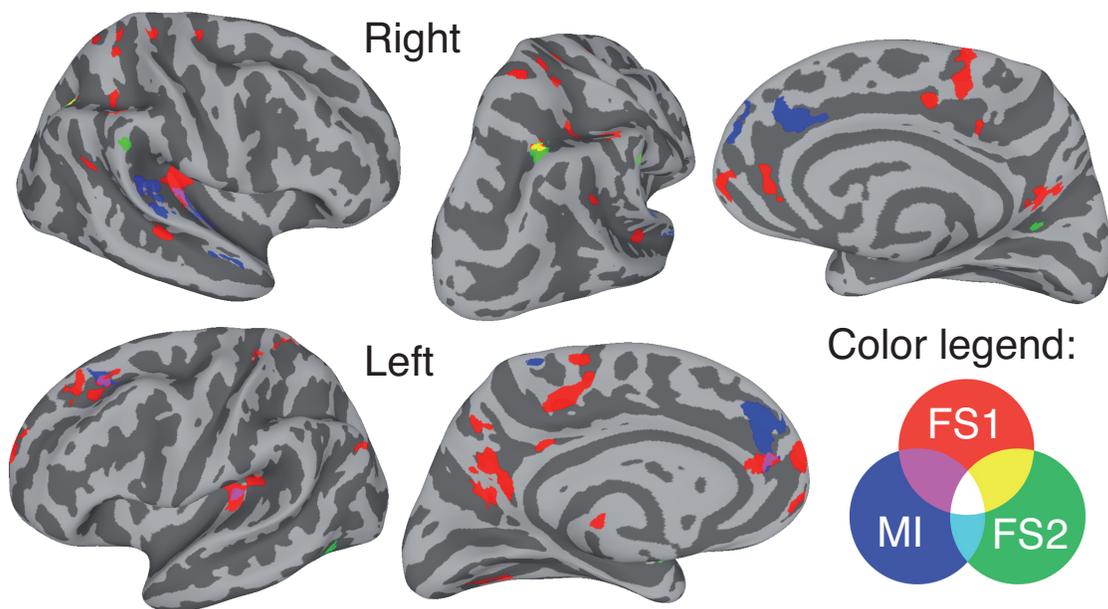
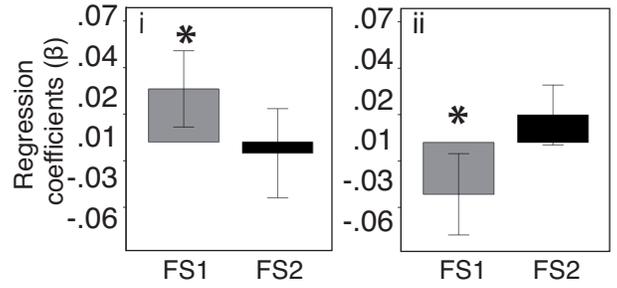


Figure 4. Group-level correlations between FS1, FS2 and MI and brain activation, shown separately for speech perception and speech production. Corrected group results for each statistic and relative overlap are shown on the group average smoothed white matter unfolded lateral and medial surfaces.

A. FS1 - FS2: Speech perception



B. FS1 - FS2: Speech production

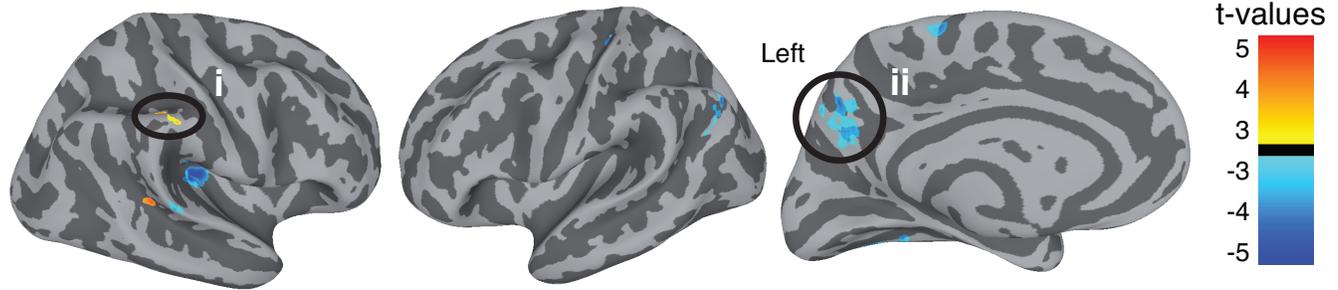


Figure 5. Group-level differences between FS1 and FS2. (A) for speech perception alone; (B) speech production alone. Cortical activations are shown on the group average smoothed white matter unfolded lateral and medial surfaces. The bar charts illustrate the different beta values for FS1 (black bars) and FS2 (gray bars) in two of the regions identified on the whole-brain maps.

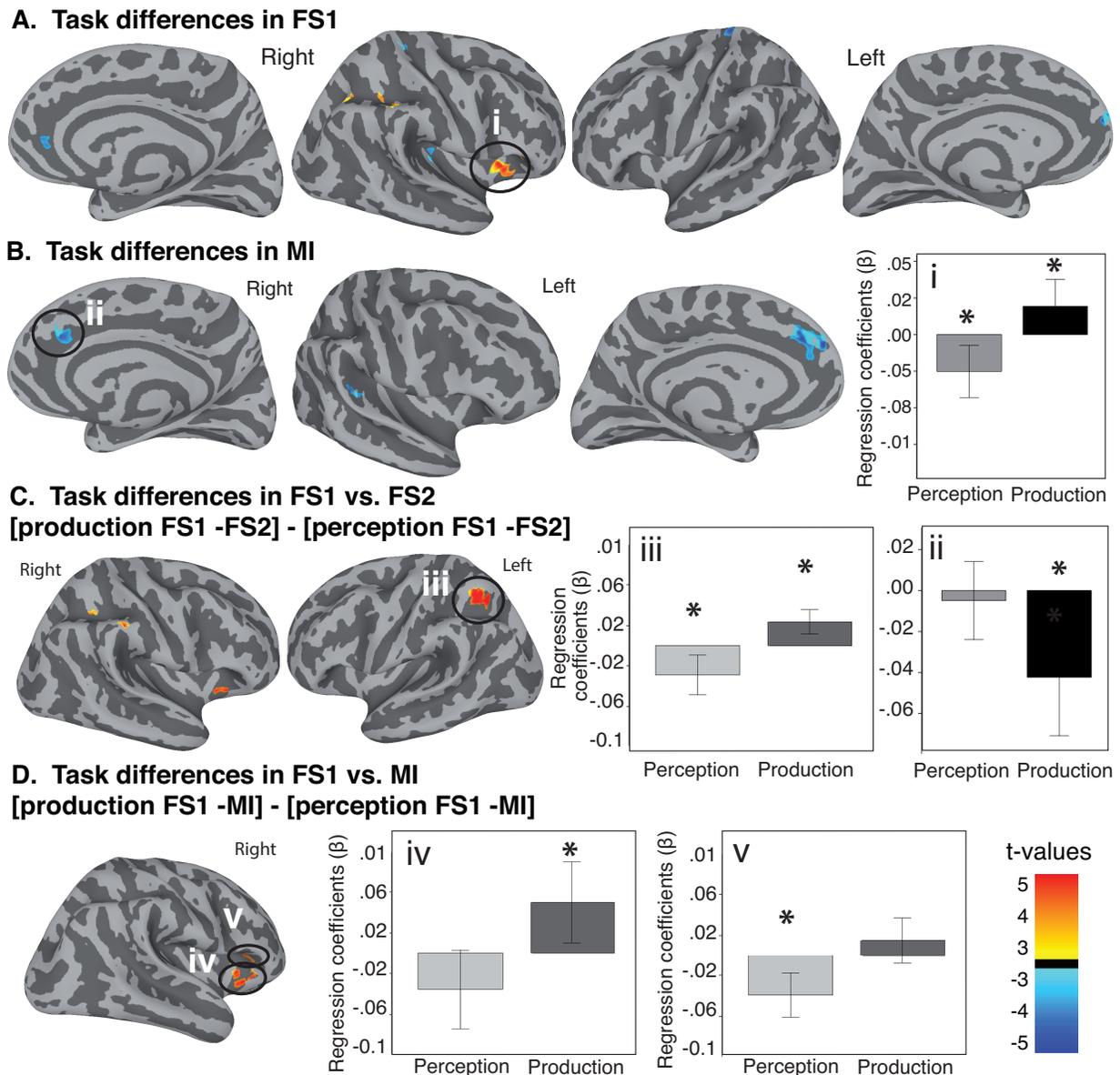


Figure 6. Group-level task differences within and across statistics. (A) Task differences in FS1, (B) task differences in MI, (C) Task differences in FS1 vs. FS2, and (D) Task differences in FS1 vs. MI. Cortical activations are shown on the group average smoothed white matter unfolded lateral and medial surfaces. The bar charts (i, ii) in panels A and B illustrate task differences for a subset of the regions identified on the whole-brain analyses. The bar charts in panels C and D (iii, iv, v) show the magnitude of difference between statistics, separately for each task. Asterisks indicate a significant difference against zero.