Noise differentially impacts phoneme representations in the auditory and speech motor systems

Yi Du[a], Bradley R. Buchsbaum[a,b], Cheryl L. Grady[a,b], and Claude Alain[a,b,1]

[a]Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, ON, Canada M6A 2E1; and [b]Department of Psychology, University of Toronto, Toronto, ON, Canada M8V 2S4

Edited by David Poeppel, New York University, New York, NY, and accepted by the Editorial Board April 2, 2014 (received for review October 3, 2013)

Although it is well accepted that the speech motor system (SMS) is activated during speech perception, the functional role of this activation remains unclear. Here we test the hypothesis that the redundant motor activation contributes to categorical speech perception under adverse listening conditions. In this functional magnetic resonance imaging study, participants identified one of four phoneme tokens (/ba/, /ma/, /da/, or /ta/) under one of six signal-to-noise ratio (SNR) levels (–12, –9, –6, –2, 8 dB, and no noise). Univariate and multivariate pattern analyses were used to determine the role of the SMS during perception of noise-impoverished phonemes. Results revealed a negative correlation between neural activity and perceptual accuracy in the left ventral premotor cortex and Broca's area. More importantly, multivoxel patterns of activity in the left ventral premotor cortex and Broca's area exhibited effective phoneme categorization when SNR ≥ –6 dB. This is in sharp contrast with phoneme discriminability in bilateral auditory cortices and sensorimotor interface areas (e.g., left posterior superior temporal gyrus), which was reliable only when the noise was extremely weak (SNR > 8 dB). Our findings provide strong neuroimaging evidence for a greater robustness of the SMS than auditory regions for categorical speech perception in noise. Under adverse listening conditions, better discriminative activity in the SMS may compensate for loss of specificity in the auditory system via sensorimotor integration.

forward sensorimotor mapping | speech categorization | fMRI | multivariate pattern analysis

The perception and identification of speech signals have traditionally been attributed to the superior temporal cortices (1–3). However, the speech motor system (SMS)—the premotor cortex (PMC) and the posterior inferior frontal gyrus (IFG), including Broca's area—that traditionally supports speech production is also implicated in speech perception tasks as revealed by functional magnetic resonance imaging (fMRI) (4–8), magnetoencephalography (9), electrocorticography in patients (10), and transcranial magnetic stimulation (TMS) (11, 12). Although there is little doubt about these redundant representations, contentious debate remains about the role of the SMS in speech perception. The idea of action-based (articulatory) representations of speech tokens was proposed long ago in the motor theory of speech perception (13) and has been revived recently with the discovery of "mirror neurons" (14). However, empirical evidence does not support a strong version of the motor theory (15). Instead, current theories of speech processing posit that the SMS may implement a sensorimotor integration function to facilitate speech perception (2, 16–18). Specifically, the SMS generates internal models that predict sensory consequences of articulatory gestures under consideration, and such forward predictions are matched with acoustic representations in sensorimotor interface areas located in the left posterior superior temporal gyrus (pSTG) and/or left inferior parietal lobule (IPL) to constrain perception (17, 18). Forward sensorimotor mapping may sharpen the perceptual acuity of the sensory system to the expected inputs via a top–down gain allocation mechanism (16), which, we assume, would be especially useful for disambiguating phonological information under adverse listening conditions. However, the assumption, that the SMS is more robust than the auditory cortex in phonological processing in noise so as to achieve successful forward mapping during speech perception, has not yet been substantiated.

In addition, there is a debate about whether the motor function is (11) or is not (16) essential for speech perception. Studies using TMS have found that stimulation of PMC resulted in declined phonetic discrimination in noise (11) but had no effect on phoneme identification under optimal listening conditions (16), suggesting a circumstantial recruitment of the SMS in speech perception. Moreover, neuroimaging studies have shown elevated activity in the SMS as speech intelligibility decreases (5, 17–21). For instance, there was greater activation in the PMC or Broca's area when participants listened to distorted relative to clear speech (19), or nonnative than native speech (17, 18). Activity in the left IFG increased as temporal compression of the speech signals increased until comprehension failed at the most compressed levels (20). For speech in noise perception, stronger activation in the left PMC and IFG was observed at lower signal-to-noise ratios (SNRs) (21), and bilateral IFG activity was positively correlated with SNR-modulated reaction time (RT) (5). Those findings have given rise to the hypothesis that the SMS contributes to speech in noise perception in an adaptive and task-specific manner. Presumably, under optimal listening conditions (i.e., no background noise), speech perception emerges primarily from acoustic representations within the auditory system with little or no support from the SMS. In contrast, the SMS would play a greater role in speech perception when the speech signal is impoverished under adverse listening conditions. However, there is likely a limit in the extent to which the SMS can compensate for poor SNR. That is, in some cases, information from articulatory commands fails to generate plausible

Significance

Contentious debate remains regarding the role of the redundant motor activation during speech perception. In this functional MRI study, multivariate pattern analysis revealed stronger multivoxel phoneme discrimination in speech motor regions than auditory cortices when the speech phonemes were moderately degraded by noise. Our findings provide neuroimaging evidence for the sensorimotor integration account. Preserved phoneme discrimination in speech motor areas may compensate for loss of specificity in noise-impoverished speech representations, which aids speech perception under adverse listening conditions.

Author contributions: Y.D., B.R.B., C.L.G., and C.A. designed research; Y.D. performed research; Y.D., B.R.B., and C.A. analyzed data; and Y.D., B.R.B., C.L.G., and C.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.P. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. E-mail: calain@research.baycrest.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1318738111/-/DCSupplemental.

predictions regarding the speech signals. Thus, the forward mapping may adaptively change with SNR in a linear or a convex (the forward mapping efficiency peaks at a certain SNR and decreases when the SNR increases or decreases) pattern. However, the SNR conditions under which the SMS can successfully compensate for perception of impoverished speech signals by such a forward mapping mechanism are unknown.

In the current fMRI study, 16 young participants identified English phoneme tokens (/ba/, /ma/, /da/, and /ta/) masked by broadband noise at multiple SNR levels (–12, –9, –6, –2, 8 dB, and no noise) via button press. A subvocal production task was also included at the end of scanning in which participants were instructed to repetitively and silently pronounce the four phonemes. Univariate General Linear Model (GLM) analysis and multivariate pattern analysis (MVPA) (22–25) were combined to investigate the recruitment [mean blood oxygenation level-dependent (BOLD) activation] and phoneme discriminability (spatial distribution of activity) of the SMS during speech in noise perception. MVPA compares the distributed activity patterns evoked by different stimuli/conditions across voxels and reveals the within-subject consistency of the activation patterns. It is robust to individual anatomical variability, is sensitive to small differences in activation, and provides a powerful tool for examining the processes underlying speech categorization (25). We predicted that (*i*) because the dorsal auditory stream (i.e., IFG, PMC, pSTG, and IPL) supporting sensorimotor integration is activated as a result of task-related speech perception (5, 17–21) and phonological working memory processes (26–28), the mean BOLD activity in those regions would negatively correlate with SNR-manipulated accuracy (increasing activity with increasing difficulty), supporting the compensatory recruitment of the SMS under adverse listening conditions; (*ii*) to implement effective forward sensorimotor mapping, the SMS would exhibit stronger multivoxel phoneme discrimination than auditory regions under noisy listening conditions; and (*iii*) when SNR decreases, the difference in phoneme discriminability between the SMS and auditory regions may increase linearly, or increase first and then decrease at a certain SNR level because of failed forward prediction processes under extensive noise interference. That is, the efficiency of the forward mapping would adaptively change with SNR in a linear or a convex pattern, respectively.

Results

Behavioral Performance. As anticipated, participants' accuracy and RT at identifying a phoneme embedded in noise strongly depended on SNR regardless of the phoneme type (Fig. 1*B*). Moreover, there was no significant difference in performance between the four phonemes. The group mean accuracy across phoneme types increased from 40% (chance = 25%) at –12 dB SNR to nearly 100% at 8 dB SNR and no noise conditions (Fig. 1*A*). Correspondingly, the group mean RT decreased linearly with increasing SNR. Both accuracy and RT followed a standard psychometric function, indicating successful manipulation of task difficulty by SNR. Because individual accuracy and RT were highly correlated ($R^2 = –0.871$, $P < 0.001$, Pearson correlation), we used accuracy as a predictor of the fMRI signal in further analyses.

Univariate GLM Results. ***Activation by phoneme perception and subvocal production.*** The GLM revealed minimal differences in BOLD activity across phonemes for both phoneme perception and subvocal production tasks. Therefore, the four phonemes were grouped and contrasted against the baseline. Compared with the silent intertrial baseline, the phoneme identification task under the no noise condition activated extensive bilateral regions in the auditory cortices (Heschl's gyrus and STG), the anterior insula, the ventral PMC (PMv) and adjacent Broca's area (pars opercularis), the frontal and parietal lobules, as well as the left dorsal primary motor cortex (M1), the dorsal PMC, and the somatosensory cortex (S1) [Fig. 1*C*;

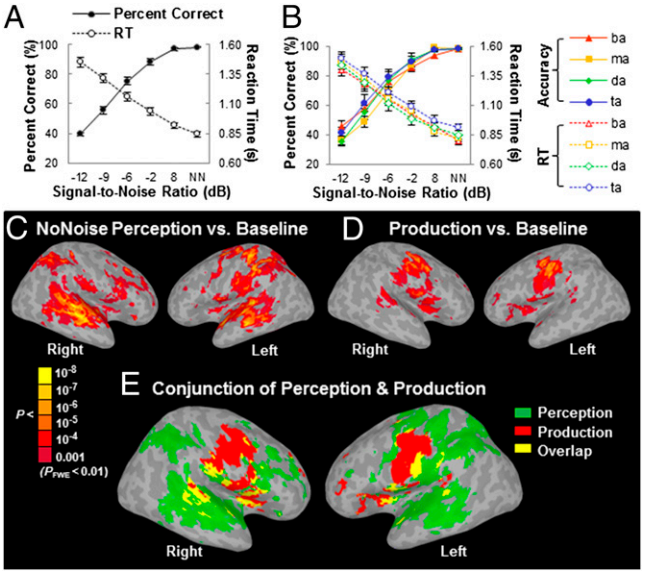

Fig. 1. Behavioral performance and activation elicited by phoneme perception and production. (*A*) Group mean accuracy and RT across phonemes as a function of SNR. (*B*) Group mean accuracy and RT for identifying individual phoneme as a function of SNR. NN, no noise condition. The error bars indicate SEM. (*C*) Activation elicited by phoneme identification without masking noise relative to silent baseline. (*D*) Activation elicited by subvocal phoneme production relative to rest. Maps are thresholded at FWE-corrected $P < 0.01$ with a cluster size ≥ 732 mm^3 for both perception and production. (*E*) Conjunction analysis of phoneme perception and subvocal phoneme production.

$P < 0.01$, family-wise-error (FWE)-corrected]. In contrast, compared with the resting baseline, the subvocal production task activated the M1 and PMC, the anterior insula and adjacent PMv and Broca's area, and the pSTG bilaterally (Fig. 1*D*; FWE-corrected $P < 0.01$). Brain regions that showed common activity for perception without noise and subvocal production (Fig. 1*E* and Table S1) included the anterior insula and adjacent PMv and Broca's area (pars opercularis), the dorsal PMC, the anterior portion of supplementary motor cortex (pre-SMA), the pSTG and the lentiform nucleus bilaterally, as well as the left ventral area of S1 in postcentral gyrus (poG) and IPL. Notably, the subvocal production task activated bilateral ventral M1/PMC, likely in response to articulation-related lips and tongue movements as suggested by previous reports (6). In contrast, the speech perception task activated the left dorsal M1/PMC, consistent with button pressing by the right fingers. Such dissociation of M1/PMC activity suggests that participants did not use a subvocal rehearsal strategy during the perception task.

Regions where BOLD signal correlated with accuracy. There was stronger BOLD activation in inferior frontal and premotor regions as well as weaker activation in temporal regions when phonemes were presented with increasing noise (Fig. S1). To quantify the noise effect directly and reveal regions where BOLD activity was modulated by task difficulty at the within-subject level, each participant's BOLD activity at each SNR was subjected to a within-subject regression analysis using each individual's mean accuracy (across phonemes) at each SNR as predictor variables. Brain regions in which BOLD signal negatively correlated with accuracy were observed bilaterally in the anterior insula and adjacent Broca's area (Ins/Broca) including pars opercularis (BA44) and pars triangularis (BA45), the pre-SMA and thalamus, as well as the left PMv, the left pSTG, the left IPL, and the right middle frontal gyrus (MFG) (blue voxels in Fig. 2*A* and Table S2; FWE-corrected $P < 0.01$). In contrast,

NEUROSCIENCE

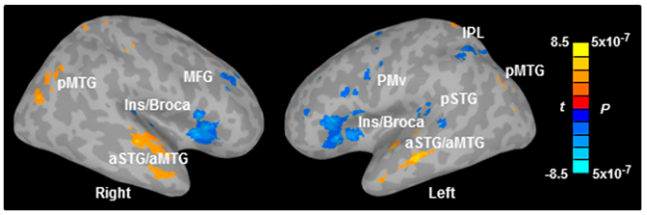

Fig. 2. Regions showing significant within-subject correlation between BOLD signal and behavioral accuracy. Maps are thresholded at FWE-corrected $P < 0.01$ with a cluster size ≥ 342 mm^3. Notably, BOLD activity in dorsal and ventral regions exhibited opposite correlations with accuracy. aSTG/aMTG, anterior superior temporal gyrus and anterior middle temporal gyrus; Ins/Broca, insula and Broca's area; IPL, inferior parietal lobule; MFG, middle frontal gyrus; pMTG, posterior middle temporal gyrus; PMv, ventral premotor cortex; pSTG, posterior superior temporal gyrus.

areas where BOLD signal positively correlated with accuracy were found bilaterally in the anterior regions of superior and middle temporal gyrus (aSTG/aMTG), the posterior MTG (pMTG), the anterior cingulate cortex, and the superior parietal lobule (yellow voxels in Fig. 2*A* and Table S2; FWE-corrected $P < 0.01$). Scatterplots in 10 regions of interest (ROIs) that exhibited a significant within-subject correlation between BOLD signal and accuracy (FWE-corrected $P < 0.01$; Fig. S2) show that although there were individual differences in absolute BOLD signal changes, all participants yielded a consistent linear correlation between BOLD activity and performance. Notably, areas that showed a negative correlation between BOLD signal and accuracy partially overlapped with regions involved in both phoneme perception and production (e.g., Broca's area, PMv, and pSTG). This suggests that areas involved in sensorimotor integration were recruited with increasing difficulty in speech in noise perception. Moreover, the relationships between task difficulty and brain activation differed for dorsal (e.g., Broca areas and PMv) and ventral (e.g., aSTG and aMTG) brain regions, providing evidence for their different roles in speech identification. Specifically, dorsal areas appear to serve a compensatory role in speech perception (increasing activity with increasing difficulty), whereas ventral areas appear to represent the intelligibility/semantic features of the speech stimuli (decreasing activity with increasing difficulty).

MVPA Results. ***Regions revealing phoneme-specific encoding.*** Given the likelihood of high intersubject anatomical variability and fine spatial scale of phoneme representations, we used MVPA to test whether phonemes from the same category evoked a more similar activation pattern than phonemes from different categories. In addition, phoneme perception at different SNRs may be represented not only by changes in mean BOLD activity in a given voxel/ROI but also by differential neural patterns in a given region. Here, a whole-brain MVPA searchlight analysis (23, 24) with a sphere of 10-mm radius was applied to the perception data at each SNR level separately. Data from five runs were randomly split into two partitions, and a similarity matrix was computed between the split halves using a Pearson correlation coefficient. At each searchlight location, a multivariate index of similarity, referred to as the phoneme-specificity index (PSI), was calculated as the overall difference in within- versus between-category correlations (average of diagonal elements minus average of off-diagonal elements in the correlation matrix) (24). Highly positive correlations were revealed for both within- and between-category comparisons (*SI Text* and Fig. S3). Thus, a positive PSI indicates that phonemes from the same category evoke more similar activation patterns than phonemes from different categories in a given region. In contrast, a PSI near zero means that different phoneme categories elicit a similar activation pattern, indicating a lack of specificity.

When there was no masking noise, several regions in the left hemisphere encoded phoneme categories (Fig. 3*A*; FWE-corrected $P < 0.05$). These included the M1, the PMC, the dorsal poG, the IFG (BA45 and BA46) and MFG, the anterior insula, the posterior STG and MTG, the IPL, and the angular gyrus. Phoneme-specific regions were also found in the right hemisphere, including the PMC, the MFG, the posterior STG and MTG, and the precuneus. Fig. 3*B* shows the mean PSI map across five SNRs (–12, –9, –6, –2, and 8 dB; FWE-corrected $P < 0.05$), as phoneme-specific representations at those SNRs revealed similar patterns (Fig. S4). Obviously, adding noise weakened the power of phoneme discrimination in almost all of the above mentioned areas except the left dorsal M1/PMC, which may index noise-irrelevant classification of button presses via the right four fingers.

Regions where phoneme discrimination correlated with accuracy. To reveal regions in which categorical speech representations were modulated by task difficulty, a within-subject regression analysis was applied to the PSI at each SNR using each individual's mean accuracy (across phonemes) at each SNR as multiple covariates. Areas that showed a positive correlation between the PSI and accuracy were observed in the left IFG including Broca's area (BA45 and BA46), the left PMv, the dorsal region of left poG, the left precuneus, the right IPL, and the right insula (Fig. 3*C* and Table S3; FWE-corrected $P < 0.05$). Among those regions, the PSI increased linearly from about 0 at the lowest SNR to significantly positive values at –6 dB SNR in the left IFG and –2 dB SNR in the left PMv [Fig. 3*D*; all $t(15) > 2.19$, $P < 0.05$, one-sample two-tailed t tests]. Notably, compared with the left IFG and left PMv regions where BOLD signal correlated with accuracy, here the left IFG region was more anterior and dorsolateral, and the left PMv region was more dorsolateral.

Phoneme discrimination in auditory cortices. The left and right dorsolateral STGs exhibited effective phoneme discrimination when there was no masking noise (Fig. 3*A*). However, the phoneme-specific representations in STGs were not significant when the

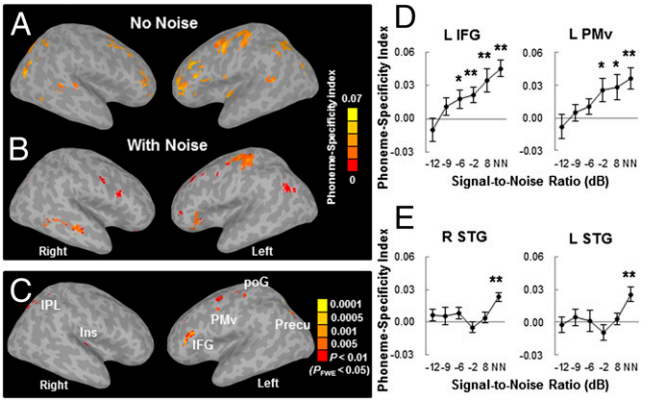

Fig. 3. PSI maps and within-subject correlation between the PSI and behavioral accuracy. PSI maps when the phonemes were presented (*A*) without noise and (*B*) with noise (average across five SNRs). Maps are thresholded at FWE-corrected $P < 0.05$ with a cluster size ≥ 293 mm^3 (without noise) or 439 mm^3 (with noise). (*C*) Regions that showed a significant within-subject correlation between the PSI and perceptual accuracy (FWE-corrected $P < 0.05$, clusters ≥ 293 mm^3). (*D*) PSI in two selected ROIs (left IFG and left PMv) that showed a significant correlation between the PSI and accuracy. (*E*) PSI in bilateral STG ROIs that exhibited significant (FWE-corrected $P < 0.05$) phoneme-specific encoding under the no noise condition. $^*P < 0.05$, $^{**}P < 0.01$, by one-sample t tests. NN, no noise condition. The error bars indicate SEM. IFG, inferior frontal gyrus; Ins, insula; IPL, inferior parietal lobule; PMv, ventral premotor cortex; poG, postcentral gyrus; Precu, precuneus; STG, superior temporal gyrus.

phonemes were embedded in noise (Fig. 3*B*). Moreover, STG areas did not reveal a significant correlation between the PSI and accuracy (Fig. 3*C*). Fig. 3*E* shows the PSI at each SNR in two ROIs from bilateral STGs that revealed effective phoneme categorization under the no noise condition (FWE-corrected $P < 0.05$). The PSI changed significantly with SNR in both ROIs [both $F(5,75) > 2.45$, $P < 0.05$, repeated-measures ANOVA]. Specifically, the PSI in both STG areas decreased markedly (no longer significant) with the addition of a small amount of noise (8 dB SNR), and this lack of phoneme discrimination was maintained with a further increase in noise level [all $t(15) < 1.45$, $P > 0.05$, one-sample two-tailed t tests]. This result suggests that phoneme-specific representations in auditory cortices are particularly sensitive to background noise, which clearly differed from the speech motor regions (i.e., Broca's area).

Phoneme discrimination in ROIs from univariate analysis. To reveal how the proposed forward mapping varies with SNR, the PSIs in four sensorimotor ROIs that showed a negative correlation between BOLD signal and accuracy (FWE-corrected $P < 0.01$) were compared. Four spherical ROIs (two motor ROIs, left PMv and Ins/Broca; two sensorimotor interface ROIs, left pSTG and IPL) with a 10-mm radius around the peak Talairach coordinates from the univariate regression analysis were defined (Fig. 4*A*). A two-way repeated-measures ANOVA on the PSI in four ROIs did not find a significant interaction between ROI and SNR [Fig. 4*B*; $F(15,225) < 1$, $P > 0.05$]. However, a significant ROI × SNR interaction was revealed between the left Ins/Broca and pSTG ROIs [$F(5,75) = 2.37$, $P < 0.05$], but not for other ROI pairs [all $F(5,75) < 1.17$, $P > 0.05$]. This was further confirmed by a SNR-dependent change in the PSI difference between the left Ins/Broca and pSTG (Fig. 4*C*; $F(5,75) = 2.37$, $P < 0.05$, one-way repeated-measures ANOVA] with a significant quadratic trend [$F(1,15) = 4.93$, $P < 0.05$]. Specifically, the left Ins/Broca exhibited stronger phoneme categorization than the left pSTG at –6, –2, and 8 dB SNRs [all $t(15) > 2.8$, $P < 0.05$] but not at lower or higher SNRs [all $t(15) < 1.4$, $P > 0.05$]. Moreover, one-sample t tests on the PSI at each ROI and SNR revealed more robust phoneme discrimination in motor ROIs than sensorimotor interface ROIs. That is, there was significant phoneme discrimination in the left PMv when SNR ≥ 8 dB and in the left Ins/Broca when SNR ≥ –6 dB [all $t(15) > 2.19$, $P < 0.05$]. However, discriminative activity was revealed only at the no noise condition in the left pSTG, and at 8 dB SNR and no noise conditions in the left IPL [all $t(15) > 2.64$, $P < 0.05$]. Thus, the difference in phoneme discrimination between motor regions (i.e., Broca's area) and auditory–motor interfaces (i.e., left pSTG) as a function of SNR suggests a convex instead of a linear pattern of forward mapping, which may peak at a medium SNR level (i.e., –2 to 8 dB) and then decrease as the noise level goes up or down. In other words, better discriminative activity in the SMS may compensate for loss of specificity in auditory and sensorimotor integration regions when the speech signal is moderately degraded by noise. However, such forward mapping may be ineffective or unnecessary under excessively adverse or optimal listening conditions when phoneme discriminability is equally poor or good in all regions of the speech processing system, respectively.

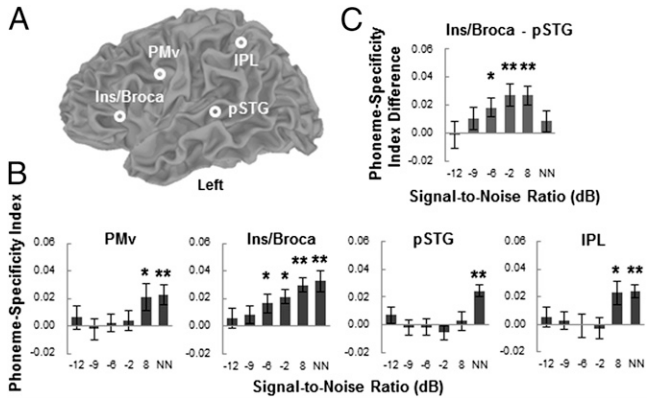

Fig. 4. PSI as a function of SNR in ROIs that showed a significant correlation between BOLD signal and accuracy. (*A*) Locations of ROIs are displayed on a template brain. ROIs were spheres with a 10-mm radius around the peak coordinates from univariate regression analysis. Peak Talairach coordinates are (–44, 0, 28) for the left PMv, (–40, 20, 5) for the left Ins/Broca, (–45, –32, 9) for the left pSTG, and (–36, –48, 38) for the left IPL. Data from the left insula and Broca's area (Ins/Broca) were pooled because a sphere with a 10-mm radius centering at the above coordinate occupied a part of both areas. (*B*) PSI in four ROIs as a function of SNR. (*C*) Difference in the PSI between the left Ins/Broca and pSTG ROIs as a function of SNR. *$P < 0.05$, **$P < 0.01$, by one-sample t tests. See Fig. 2 legend for abbreviations.

Discussion

In the present study, the overall BOLD responses and multivoxel phoneme discriminability in the left PMv and Broca's area exhibited opposite relationships with SNR-manipulated accuracy (BOLD, negative correlation; discriminability, positive correlation). The positive association between task difficulty and BOLD activity in those speech motor regions suggests a compensatory recruitment of the SMS in speech in noise perception. More importantly, those speech motor areas exhibited significant phonetic discrimination above –6 dB SNR, whereas bilateral auditory cortices encoded phoneme-specific information only when the noise was absent or extremely weak (SNR > 8 dB). Our findings provide direct neuroimaging evidence showing a greater robustness of the SMS than the auditory system for phoneme categorization in noise. In addition, Broca's area exhibited stronger phoneme categorization than the sensorimotor interface area (left pSTG) at medium levels of SNR (−2 to 8 dB) but not at lower or higher SNRs. Our results suggest a convex pattern of forward mapping during speech in noise perception, peaking when the speech sound is moderately distorted by noise but turning off under extremely adverse or optimal listening conditions.

Consistent with previous findings (4, 6, 7, 29), areas including bilateral anterior insula and adjacent PMv/Broca's area, bilateral pSTG, and the left IPL were activated by both identification of unmasked speech phonemes and subvocal phoneme production. Among those regions, the PMv, Broca's area, and the anterior insula belong to the prefrontal articulatory network, whereas the left pSTG and left IPL may function as an auditory–motor interface, which transforms acoustic representations of speech into their articulatory counterparts (1, 2). Unlike Pulvermüller et al. (6), who show an articulatory-feature-related somatotopic activation in the left PMC between [p] versus [t] phonemes during both subvocal production and passive listening (weak effect), here we did not observe such differential activation between bilabial/lip-related ([b] and [m]) versus alveolar/tongue-related ([d] and [t]) phonemes for either production or perception tasks. This may be due to different stimuli, different imaging paradigms (sparse sampling vs. continuous sampling), and different tasks (passive listening vs. active identification) between those two studies.

The task difficulty was effectively manipulated as participants' accuracy and RT negatively correlated with each other as a function of SNR. Brain–behavior correlation revealed a positive relationship between accuracy and BOLD signal in bilateral anterior STG and MTG regions, which may serve as neural substrates for speech intelligibility and semantic processing (1, 3). In comparison, BOLD activity in the dorsal auditory stream was negatively correlated with perceptual accuracy, consistent with previous findings showing task-relevant activations in the left PMC (17–19, 21), the left IFG (5, 17, 20, 21), the left pSTG (30), and the left IPL (31) during speech perception. Although

NEUROSCIENCE

not our primary purpose, those results confirm a functional dissociation between the dorsal and ventral auditory streams that mediate sensorimotor integration and speech comprehension, respectively (1, 2). The strong BOLD activity at difficult (low SNR) conditions in sensorimotor regions may represent enhanced activation of internal models by articulatory prediction against which sensory representations could be matched. Alternatively, it may reflect increasing demand on attention, working memory, or decision making. However, the univariate analysis could not dissociate the two possibilities, making MVPA crucial for the current study.

By applying MVPA searchlight analysis, we revealed phoneme-specific representations in multiple regions lateralized in the left hemisphere (e.g., PMC, IFG, dorsolateral STG, and IPL) when the phonemes were presented without noise, partially overlapping with the SMS. The results are in line with categorical speech representation in human Broca's area and the left STG as revealed by a recent fMRI study using MVPA (25), scalp-recorded event-related potentials (31), as well as intracranial local field potentials in patients (32). A positive correlation between the PSI and perceptual accuracy was revealed in several regions such as the left PMv and the left IFG. This suggests that the multivoxel phoneme categorization efficiency in specific brain regions could predict listeners' ability to correctly identify the phoneme embedded in noise to some extent. More importantly, the strength of phonetic discrimination in those motor regions decreased gradually with increasing external noise with effective phoneme categorization above –6 dB SNR. In contrast, bilateral auditory cortices were vulnerable to noise in encoding phoneme-specific information, which was readily disrupted by a small amount of noise (8 dB SNR). The greater robustness (~14 dB SNR difference) in phoneme discrimination in noise in the SMS than auditory cortices supports the sensorimotor integration account (2, 16–18) and provides evidence that the SMS indeed contributes to phonological processing by compensating for loss of discrimination in speech representations in the auditory system under adverse listening conditions.

Our results further suggest a convex pattern of forward mapping as a function of SNR. When the noise is too strong (SNR < –6 dB), the internal models (motor predictions) are likely inaccurate, resulting in ineffective forward mapping. Conversely, when the noise is weak (SNR > 8 dB), the forward mapping is likely unnecessary, as the speech representations in the auditory system are highly accurate. However, when the noise level is intermediate (–6 dB ≤ SNR ≤ 8 dB), the degraded speech representations in auditory cortices could benefit from motor predictions that constrain the sensory processing of the expected inputs in a top–down manner. Such a convex pattern of forward mapping as a function of task difficulty is consistent with previous reports showing strongest PMC (19) or IFG (20) activity only when the speech signal is moderately distorted but identifiable. Because functional connectivity between the PMC and posterior temporal plane has been demonstrated (8, 19), and increasing functional connectivity between prefrontal and parietal cortices facilitates speech comprehension under adverse listening conditions (30), we propose that the changes in sensorimotor mapping strength may be represented by alterations in functional connectivity between the prefrontal articulatory regions, sensorimotor interfaces, and auditory cortices.

In addition, strong BOLD responses do not always imply successful multivoxel phoneme discrimination. Indeed, phoneme discrimination failed in two prefrontal articulatory regions (left PMv and Broca's area) below –6 dB SNR and in two auditory–motor interfaces (left pSTG and IPL) below 8 dB SNR, despite strong BOLD activation. We speculate that the increased BOLD activity at low SNRs in articulatory regions may partially relate to inaccurate internal models as well as iterative "predict-correct loops" because of increased variance in prediction. Also, the strong BOLD activation under noisy conditions in the left pSTG and IPL may be partially caused by unmatched sensorimotor mapping and repeated error correction processes. Because categorical speech perception requires the listener to maintain sublexical representations in an active state as a metalinguistic judgment is made, it involves some degree of executive control and working memory functions (1). Thus, the linearly prolonged RT and incremental BOLD activity in the PMv, IFG, pSTG, and IPL with increasing task difficulty likely reflect an accumulation of effort-related changes in selective attention (33), phonological working memory (20, 26–28), or phoneme-category judgment and response selection (5, 8, 34).

In sum, analyses of regional-average activation and multivoxel pattern information told complementary stories. Specifically, we demonstrated distinct patterns between the recruitment (mean BOLD activity) and multivoxel phoneme discriminability in the SMS as a function of SNR in a phoneme categorization task. More importantly, MVPA revealed stronger phoneme discrimination at moderate noisy listening conditions in the left PMv and Broca's area than in bilateral auditory cortices, which provides neuroimaging evidence for the assertion that the superior phoneme categorization in the SMS may compensate for degraded bottom–up speech representations in a top–down fashion. In addition, the convex pattern of forward sensorimotor mapping here significantly advances and refines the sensorimotor integration theory by characterizing its range and limits. Further research is needed to investigate how the internal models in the SMS as well as the functional connectivity between the SMS and the auditory cortex are affected by task difficulty (e.g., SNR), cognitive aging, hearing loss, and diseases like aphasia. Our findings also emphasize the importance of the spatial distribution as well as the mean activation of cortical representations along the speech-processing pathways.

Materials and Methods

Participants. Sixteen right-handed adults (21–34 y old, M = 26.2; eight females) providing written informed consent according to the University of Toronto and Baycrest Hospital Human Subject Review Committee guidelines participated in the study. All participants were native English speakers and had normal pure-tone thresholds at both ears (<25 dB HL for 250–8,000 Hz).

Stimuli and Task. Four tokens from the standardized Nonsense Syllable Test (35) were used in this study. These tokens were naturally produced English consonant–vowel phonemes (/ba/, /ma/, /da/, and /ta/), spoken by a female talker. Each phoneme token was 500 ms in duration and matched in terms of average root mean square sound pressure level (SPL). The vowel was always [a] because its formant structure provides a superior SNR relative to the MRI scanner spectrum. The four consonants were stop consonants, chosen for their balanced articulatory features (bilabial/lip-related [b] and [m] vs. alveolar/tongue-related [d] and [t]). A 500-ms online-generated white noise segment (4 kHz low-pass cutoff, 10 ms rise–decay envelope) started and ended simultaneously with the phonemes. Sounds were presented by circumaural MRI-compatible headphones (HP SI01, MR Confon), acoustically padded to suppress scanner noise by 25 dB. The intensity level of the phonemes was fixed at 85 dB SPL, and the noise level at five SNRs was 97, 94, 91, 87, and 77 dB SPL (before attenuation by headphone), leading to five levels of SNR: –12, –9, –6, –2, and 8 dB. Besides the five SNR levels, phonemes were presented alone at 85 dB SPL.

Before scanning, phonemes were presented individually without noise (four trials per phoneme), and participants identified the phonemes by pressing corresponding keys on a four-button pad with an accuracy of 94% or better. During scanning, there were five perception runs followed by one subvocal production run. For each perception run, 80 noise-masked phonemes (four trials per phoneme per SNR) and 20 phonemes without noise (5 trials per phoneme) were randomly presented at an averaged interstimuli interval of 4 s (jittered from 2 to 6 s). Participants were asked to listen carefully and identify the phonemes by pressing corresponding keys as fast as possible.

For the subvocal production run, the letter string "say BAH," "say MAH," "say DAH," or "say TAH" occurred every 2 s on the screen for 16 s, and participants performed the respective articulation movement repeatedly (following the rate of flash) during the entire period. Participants were

instructed to "pronounce silently, without making any sound" and move the jaw as little as possible, so as to avoid movement artifacts. After that, the letter string was replaced by a fixation cross for another 16 s, and participants were told to close their mouth and have a rest. Four on–off blocks occurred for each phoneme in a pseudorandom order.

Data Acquisition and Analysis. Participants were scanned using a 3-T MRI system (Siemens Trio 3 T magnet) with a standard 12-channel "matrix" head coil. Functional imaging was performed to measure brain activation by means of the BOLD signal. T2* functional images were obtained using continuous echo planar imaging acquisition [30 slices, matrix size, 64 × 64, 5-mm thick; repetition time (TR), 2,000 ms; echo time (TE), 30 ms; flip angle, 70°; field of view (FOV), 200 mm; voxel size, 3.125 × 3.125 × 5 mm]. Structural T1-weighted anatomical volumes were obtained after three fMRI runs using spoiled gradient recalled echo (axial orientation, 160 slices, 1-mm thick; TR, 2,000 ms; TE, 2.6 ms; FOV, 256 mm). The preprocessed imaging data (including physiological motion correction, slice-timing correction, and realignment to a reference image, by Analysis of Functional Neuroimages software, AFNI version 2.56a; *SI Text*) were then analyzed by two complementary methods: (*i*) voxel-wise GLM analysis and (*ii*) whole-brain MVPA.

Univariate GLM Analysis. The concatenated imaging data during perception were fit with a GLM with different regressors for the four phonemes and six SNRs in AFNI. A GLM was also fit to the production data with the four articulation movements being modeled by different regressors. The predicted activation time course was modeled as a "gamma" function convolved with the canonical hemodynamic response function for event-related perception data and as a box-car function for block-designed production data. The four phonemes were grouped and contrasted against the baseline (silent intertrial intervals for perception and interblock rest period for production), as the GLM revealed similar activity across phonemes. Individual contrast maps and within-subject regression maps were normalized to Talairach stereotaxic space, resampled with a voxel size of 3 × 3 × 3 mm, and spatially smoothed using a Gaussian filter with a FWHM value of 6.0 mm. Maps were tested for random effect by one-sample *t* tests and corrected for multiple comparisons by AlphaSim procedure. Using an uncorrected $P = 0.001$, this procedure yielded a map-wise false-positive $P = 0.01$ by removing clusters smaller than certain spatial extents (see the legends of Figs. 1 and 2 and *SI Text* for details).

MVPA. The preprocessed data from the perception runs were fit with a GLM for each run separately with the four phonemes and six SNRs being modeled by individual regressors. Data were resampled (3 × 3 × 3 mm) but not smoothed to ensure maximal sensitivity to high spatial frequencies (36). To avoid introducing dependencies between conditions, data without subtraction of the mean pattern (normalization) were entered into MVPA (37). Specifically, pattern similarity between phonemes was assessed using a multivariate searchlight analysis (23, 24) in which a sphere of 10-mm radius was swept through the entire brain volume. Data from five runs were randomly split into two partitions for each phoneme category and each SNR separately, and a similarity matrix was computed between the split halves using a Pearson correlation coefficient. At each searchlight location, a multivariate index of reliability, referred to as the PSI, was computed as the average within-category correlations minus the average between-category correlations (24) (*SI Text* and Fig. S3). This was then averaged over the 10 random split halves for each searchlight center. The process was conducted independently for each part of the brain covered by the sphere, and information maps were created by mapping the multivariate index back to the corresponding voxels. The single-subject MVPA maps were normalized to Talairach space, subjected to random effect analysis by *t* tests, and corrected for multiple comparisons by AlphaSim procedure (uncorrected $P < 0.01$, FWE-corrected $P < 0.05$; see Fig. 3 legend and *SI Text* for cluster extent threshold).

ACKNOWLEDGMENTS. We thank Gregory Hickok, Gavin Bidelman, Stephen Arnott, and Jeffrey Wong for their insightful comments on an earlier version of this manuscript. This research was supported by grants from the Canadian Institutes of Health Research (MOP106619).

1. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8(5):393–402.
2. Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat Neurosci* 12(6):718–724.
3. Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26(2):100–107.
4. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7(7):701–702.
5. Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7(3):295–301.
6. Pulvermüller F, et al. (2006) Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci USA* 103(20):7865–7870.
7. Callan D, Callan A, Gamez M, Sato MA, Kawato M (2010) Premotor cortex mediates perceptual performance. *Neuroimage* 51(2):844–858.
8. Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013) Automatic phoneme category selectivity in the dorsal auditory stream. *J Neurosci* 33(12):5208–5215.
9. Herman AB, Houde JF, Vinogradov S, Nagarajan SS (2013) Parsing the phonological loop: Activation timing in the dorsal speech stream determines accuracy in speech reproduction. *J Neurosci* 33(13):5439–5453.
10. Edwards E, et al. (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. *Neuroimage* 50(1):291–301.
11. Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. *Curr Biol* 17(19):1692–1696.
12. Sato M, Tremblay P, Gracco VL (2009) A mediating role of the premotor cortex in phoneme segmentation. *Brain Lang* 111(1):1–7.
13. Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21(1):1–36.
14. Rizzolatti G, Arbib MA (1998) Language within our grasp. *Trends Neurosci* 21(5): 188–194.
15. Lotto AJ, Hickok GS, Holt LL (2009) Reflections on mirror neurons and speech perception. *Trends Cogn Sci* 13(3):110–114.
16. Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron* 69(3):407–422.
17. Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *Neuroimage* 22(3):1182–1194.
18. Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *Neuroimage* 33(1):316–325.
19. Osnes B, Hugdahl K, Specht K (2011) Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage* 54(3): 2437–2445.
20. Poldrack RA, et al. (2001) Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *J Cogn Neurosci* 13(5):687–697.
21. Scott SK, Rosen S, Wickham L, Wise RJ (2004) A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J Acoust Soc Am* 115(2):813–821.
22. Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425–2430.
23. Kriegeskorte N, Bandettini P (2007) Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage* 38(4):649–662.
24. Linke AC, Vicente-Grabovetsky A, Cusack R (2011) Stimulus-specific suppression preserves information in auditory short-term memory. *Proc Natl Acad Sci USA* 108(31): 12961–12966.
25. Lee YS, Turkeltaub P, Granger R, Raizada RD (2012) Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *J Neurosci* 32(11):3942–3948.
26. Buchsbaum BR, et al. (2011) Conduction aphasia, sensory-motor integration, and phonological short-term memory—An aggregate analysis of lesion and fMRI data. *Brain Lang* 119(3):119–128.
27. Poldrack RA, et al. (1999) Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10(1):15–35.
28. Liebenthal E, Sabri M, Beardsley SA, Mangalathu-Arumana J, Desai A (2013) Neural dynamics of phonological processing in the dorsal auditory stream. *J Neurosci* 33(39): 15414–15424.
29. Buchsbaum BR, et al. (2005) Reading, hearing, and the planum temporale. *Neuroimage* 24(2):444–454.
30. Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. *J Neurosci* 27(9):2283–2289.
31. Bidelman GM, Moreno S, Alain C (2013) Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage* 79:201–212.
32. Chang EF, et al. (2010) Categorical speech representation in human superior temporal gyrus. *Nat Neurosci* 13(11):1428–1432.
33. Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3(3):201–215.
34. Venezia JH, Saberi K, Chubb C, Hickok G (2012) Response bias modulates the SMS during syllable discrimination. *Front Psychol* 3:157.
35. Dubno JR, Schaefer AB (1992) Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. *J Acoust Soc Am* 91(4 Pt 1):2110–2121.
36. Thompson R, Correia M, Cusack R (2011) Vascular contributions to pattern analysis: Comparing gradient and spin echo fMRI at 3T. *Neuroimage* 56(2):643–650.
37. Garrido L, Vaziri-Pashkam M, Nakayama K, Wilmer J (2013) The consequences of subtracting the mean pattern in fMRI multivariate correlation analyses. *Front Neurosci* 7:174.