



Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients

Gangyi Feng^{a,b}, Erin M. Ingvalson^{c,d}, Tina M. Grieco-Calub^{e,f}, Megan Y. Roberts^e, Maura E. Ryan^{g,h}, Patrick Birmingham^{i,j}, Delilah Burrowes^{g,h}, Nancy M. Young^{d,f,k}, and Patrick C. M. Wong^{a,b,l,1}

^aDepartment of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong SAR, China; ^bBrain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR, China; ^cSchool of Communication Science and Disorders, Florida State University, Tallahassee, FL 32301; ^dDepartment of Otolaryngology - Head and Neck Surgery, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611; ^eRoxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL 60208; ^fKnowles Hearing Center, School of Communication, Northwestern University, Evanston, IL 60208; ^gDepartment of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611; ^hDepartment of Medical Imaging, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611; ⁱDepartment of Anesthesiology, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611; ^jDepartment of Anesthesiology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611; ^kDivision of Otolaryngology - Head and Neck Surgery, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611; and ^lDepartment of Otorhinolaryngology, Head and Neck Surgery, The Chinese University of Hong Kong, Hong Kong SAR, China

Edited by Robert J. Zatorre, Montreal Neurological Institute, McGill University, Montreal, QC, Canada, and accepted by Editorial Board Member Thomas D. Albright December 6, 2017 (received for review October 17, 2017)

Although cochlear implantation enables some children to attain age-appropriate speech and language development, communicative delays persist in others, and outcomes are quite variable and difficult to predict, even for children implanted early in life. To understand the neurobiological basis of this variability, we used presurgical neural morphological data obtained from MRI of individual pediatric cochlear implant (CI) candidates implanted younger than 3.5 years to predict variability of their speech-perception improvement after surgery. We first compared neuroanatomical density and spatial pattern similarity of CI candidates to that of age-matched children with normal hearing, which allowed us to detail neuroanatomical networks that were either affected or unaffected by auditory deprivation. This information enables us to build machine-learning models to predict the individual children's speech development following CI. We found that regions of the brain that were unaffected by auditory deprivation, in particular the auditory association and cognitive brain regions, produced the highest accuracy, specificity, and sensitivity in patient classification and the most precise prediction results. These findings suggest that brain areas unaffected by auditory deprivation are critical to developing closer to typical speech outcomes. Moreover, the findings suggest that determination of the type of neural reorganization caused by auditory deprivation before implantation is valuable for predicting post-CI language outcomes for young children.

neural preservation | cochlear implant | prediction | auditory deprivation | machine learning

Auditory neuroscience research has detailed mechanisms of neural plasticity resulting from restoration of hearing after a period of auditory deprivation (1, 2). Importantly, studies have found considerably greater plasticity in both humans (3) and animal models (4, 5) when restoration occurs at relatively younger (e.g., before age 4 y in humans) versus relatively older ages (e.g., after age 7 y). This is not surprising from a basic neuroscience standpoint. Neuronal processes, such as synaptogenesis, especially those associated with connections of the deeper cortical layers and convergence of bottom-up and top-down inputs (6), likely have a much better chance of approaching normal when sensory restoration occurs early (1, 5, 7). However, it is not clear how auditory deprivation, and the resulting neural reorganization, is associated with better or poorer communication outcomes in those affected from birth or very early in life. When applied to humans, these questions are not only theoretically interesting, as they allow for testing competing hypotheses

about neural plasticity within a young age, but also clinically relevant. Children with early-onset auditory deprivation (and their caregivers) seek cochlear implantation for hearing restoration, often times within infancy. However, given the large variability in postcochlear implant (post-CI) language outcomes, even among children implanted at a young age (8), there is currently no viable method to predict which children will achieve age-appropriate language skills or who will experience persistent language delays. This gap in clinical practice calls for new clinical tools to permit development and implementation of more effective rehabilitation strategies.

In a recent review, Gabrieli et al. (9) made a strong argument for outcome prediction as a key contribution of human cognitive neuroscience. In the present study, early-deafened children younger than 3.5 y underwent structural MRI as part of their presurgical clinical evaluation to rule out gross brain abnormalities

Significance

The ability to accurately predict speech improvement for young children who use cochlear implants (CIs) would be a first step in the development of a personalized therapy to enhance language development. Despite decades of outcome research, no useful clinical prediction tool exists. An accurate predictive model that relies on routinely obtained presurgical neuroanatomic data has the potential to transform clinical practice while enhancing our understanding of neural organization resulting from auditory deprivation. Using presurgical MRI neuroanatomical data and multivariate pattern analysis techniques, we found that neural systems that were unaffected by auditory deprivation best predicted young CI candidates' postsurgical speech-perception outcomes. Our study provides an example of how research in cognitive neuroscience can inform basic science and lead to clinical application.

Author contributions: G.F., E.M.I., N.M.Y., and P.C.M.W. designed research; E.M.I., T.M.G.-C., M.Y.R., M.E.R., P.B., D.B., and N.M.Y. performed research; G.F. and E.M.I. analyzed data; and G.F., N.M.Y., and P.C.M.W. wrote the paper.

Conflict of interest statement: G.F., N.M.Y., P.C.M.W., The Chinese University of Hong Kong, and the Ann & Robert H. Lurie Children's Hospital of Chicago disclose potential financial conflict of interest related to US patent application filed on December 21, 2017.

This article is a PNAS Direct Submission. R.J.Z. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

¹To whom correspondence should be addressed. Email: p.wong@cuhk.edu.hk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717603115/-DCSupplemental.

and define cochlear and eighth nerve anatomy. We compared neuro-morphological differences between these children and age-matched children with normal hearing to identify neural reorganization due to auditory deprivation at early stages of development. This comparison yields two brain templates for testing competing hypotheses about how neuro-morphological changes relate to postsurgical language outcomes. The first hypothesis, the “neural recycling hypothesis,” predicts that brain regions significantly affected by auditory deprivation would be most predictive of postsurgical outcomes. Both transient and permanent neural injuries can result from auditory deprivation due to prelingual deafness. Morphological reorganization in the affected regions would indicate the extent of the potential injury, which may be associated with the likelihood of functional restoration in the future. Candidate brain areas associated with this hypothesis include regions within the auditory system, especially the middle section of the superior temporal gyrus (STG), which is essential for basic acoustic analysis (10, 11). This area also shows consistent differences across studies of neuroanatomical differences in the hearing impaired, although these studies tend to focus on adults with prolonged hearing impairment (12–14).

The second hypothesis, the “neural preservation hypothesis,” postulates that brain areas that are unaffected by auditory deprivation would most likely contribute to post-CI language development. For relatively young children, these unaffected brain areas likely include higher-level brain areas, such as auditory association cortex (including the dorsal portion involving in speech perception) (10, 11) and higher-order cognitive and language-related brain regions, which are developed later in life (15, 16). When implantation is done at a young age, these later developing areas may not yet be significantly altered by auditory deprivation in comparison with lower-level auditory regions. These higher-level brain areas have been shown to have large individual variability in both neuroanatomy (17) and functions (18), which are associated with individual differences of cognitive functions or behaviors (19, 20). The preservation of these brain areas is also likely related to children’s ability to resolve the degraded input generated by the CI. To overcome poor signal quality, higher-level brain areas might be particularly engaged to modulate lower-level acoustic analysis to develop speech.

In the present study, 37 children with bilateral sensorineural hearing loss who underwent cochlear implantation [18 females; age at implant range 8–38 mo, mean = 17.9 mo (SD = 7.81)] and 40 children with normal hearing (NH) [14 females; age range 8–38 mo, mean = 18.0 mo (SD = 10.2)] underwent T1-weighted neuroanatomical MRI. These neuroanatomical scans were obtained before CI surgery. None of the children in the CI group showed acquired brain anomalies, severely malformed cochlea, or cochlear nerve deficiency. All children in the CI group underwent anesthesia during scanning to optimize patient comfort and to minimize movement artifacts. All children who used CI evaluated were below age 3.5 y at implant, as studies have found that children implanted before 3.5–4 y of age are more likely to have closer to age-typical clinical outcomes and electrophysiological responses than children implanted at older ages (3, 21, 22). Both before CI surgery and 6-mo postactivation, the children in the CI group underwent speech-perception assessments that included a battery of standardized measures typically used in CI clinical care and research (8, 23). To identify morphological brain patterns affected and unaffected by deafness, we used voxel-based morphometry and multivoxel pattern similarity analyses to compare neuroanatomical images of children in the CI and NH groups. From these analyses, brain regions that were potentially affected and unaffected by deafness were revealed. We then generated brain templates for the affected and unaffected regions and constructed machine-learning models to predict children’s future speech improvement in the CI group based on gains on speech perception observed 6-mo postactivation. We

employed machine-learning algorithms with different cross-validation procedures, which have been found to overcome an overestimation of predictive power in correlational approaches (24, 25) that were used in previous clinical “postdiction” studies (9, 26, 27; see ref. 9 for a review of postdiction studies and discussion of their limitations). Furthermore, using neuroanatomical measures to predict future speech improvement of individual pediatric CI recipients provides an avenue to objectively examine the brain that is task-free, not biased by residual hearing, and not influenced by sedation or anesthesia, which changes the quality of the functional MRI (fMRI) blood-oxygen level-dependent (BOLD) signal (28).

Results

Neural Reorganization Due to Auditory Deprivation in Young Children.

We used two types of neuro-morphological analyses to evaluate neural reorganization resulting from early deafness: voxel-based morphometry (VBM) and multivoxel pattern similarity (MVPS) analyses (see *SI Appendix*, Fig. S1 for MVPS analysis procedure). While VBM reveals local tissue (gray and white matter, GM and WM, respectively) density, MVPS measures similarity in local spatial morphological pattern that is independent of voxel-wise density. Thus, the two types of analyses are relatively complementary. Taken as a whole, the most obvious neuro-morphological differences between the children in the CI and NH groups lie in the bilateral auditory cortex, especially the middle portion of STG in both VBM density and MVPS (see Fig. 1 and *SI Appendix*, Table S1 for detailed regions). These effects in both GM and WM remained after demographic variables [i.e., age, sex, and estimated household income (socioeconomic status, SES)] were controlled (see *SI Appendix*, Fig. S2 for results when these variables were not controlled). Less consistently across tissue and analysis type, neuro-morphological differences can be observed in the inferior frontal gyrus, cingulate gyrus, the occipital lobe, hippocampus, and the precuneus, suggesting that auditory deprivation may have begun to affect cross-modal and general cognitive brain regions, but these effects are less consistent across measures compared with the effects found in the auditory cortex (see *SI Appendix*, Fig. S3 for results using a less-conservative threshold).

Predicting Auditory and Speech-Perception Outcomes. To evaluate whether brain regions affected (neural recycling hypothesis) or unaffected (neural preservation hypothesis) by auditory deprivation are the best predictors of speech-perception outcomes, we conducted a series of machine-learning prediction analyses.

Auditory and speech-perception outcomes. Speech perception was assessed once before surgery and again at 6 mo postactivation using the Speech Recognition Index in Quiet (SRI-Q) (29), following previous longitudinal CI studies (8, 23, 30). The SRI-Q consisted of a hierarchy of measures of speech-recognition abilities (see *Materials and Methods* for details). We performed a one-way ANOVA with pre- and post-CI SRI-Q scores as the main variable of interest and all other nonneural measures, including age, sex, SES, lateralization of CI (CIL, unilateral vs. bilateral), and pre-CI speech awareness threshold (SAT, an index of residual hearing) as covariates. We found a significant improvement in SRI-Q at 6 mo postactivation compared with pre-CI scores [$F_{(1, 66)} = 62.52, P < 0.001$]. To calculate improvement in speech perception, we subtracted the pre-CI scores from the 6-mo outcome scores for each individual. Large individual differences in speech improvement can be seen (Fig. 2A), as in previous studies (8). For our first set of machine-learning prediction analyses, children in the CI group were divided into high-vs. low-improvement groups by a median split (see patient classification results in the paragraph immediately below). In the second set of prediction analyses, more precise ranking predictions were also made (see *Neural Machine Learning and Precise Ranking Prediction*, below).

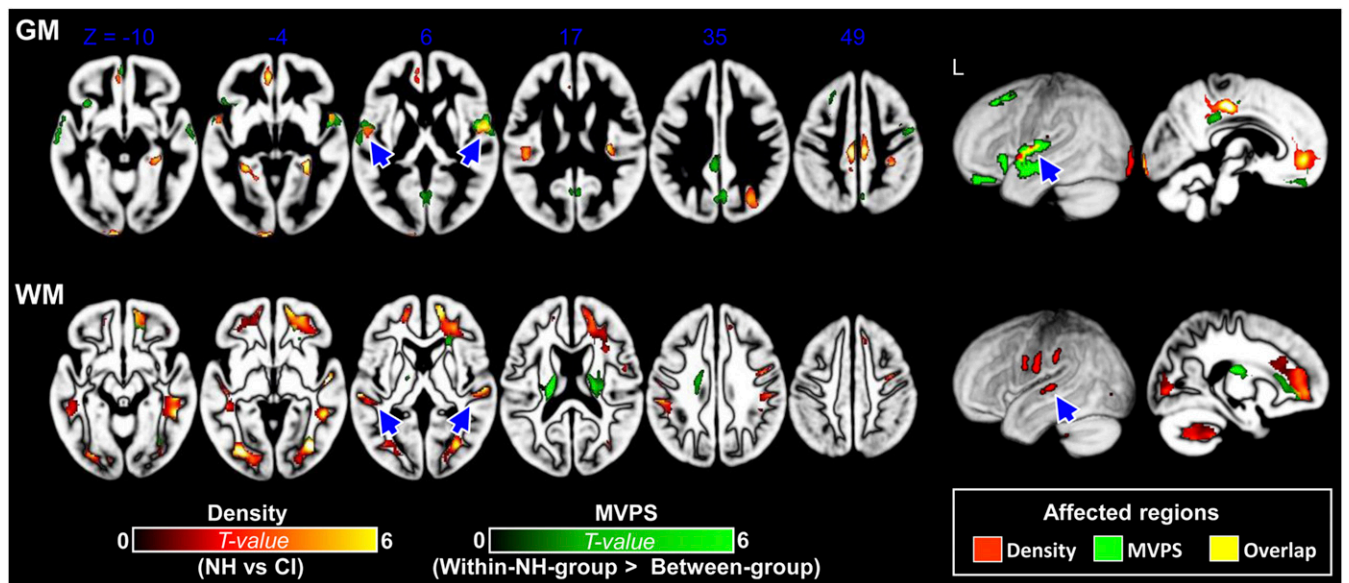


Fig. 1. GM and WM differences in VBM density (i.e., density; regions in red-yellow) and neuro-morphological pattern (i.e., MVPS; regions in green) between the NH and CI groups. (Upper) Children in the CI group showed significant differences in GM density compared with children in the NH group across auditory temporal, medial frontal, and posterior cingulate regions. Additionally, the between-group MVPS was significantly lower relative to within-NH group MVPS across frontotemporal regions, suggesting that there was significant brain reorganization in morphological pattern for children in the CI group compared with age-matched children in the NH group. The arrows highlight the affected auditory regions, many of which were in the middle portion of STG/STS (including core and belt regions). (Lower) WM differences in density and MVPS between children in the NH and CI groups. Children in the CI group showed a significant decrease in WM density compared with children in the NH group. The between-group MVPS was also significantly lower relative to within-NH group MVPS for frontal and temporal WM areas, suggesting there was a significant neuroanatomical reorganization in morphological pattern for children in the CI group compared with age-matched children with NH. The group-comparison brain maps were obtained by using a two-sample *t* tests with the variance of demographic variables controlled. Voxel-level threshold $P < 0.001$, cluster-level FWE-corrected $P < 0.05$. L, left hemisphere.

Within the narrow age range of our patients (8–38 mo), we found that age was only marginally correlated with pre-CI SRI-Q ($\rho = 0.28$, $P = 0.09$) and not at 6 mo postactivation ($\rho = -0.09$, $P = 0.61$). Fig. 2*B* shows the correlation matrix indicating cor-

relation strength between nonneural measures. Similar to previous studies (e.g., ref. 8), residual hearing was significantly correlated with pre-CI SRI-Q ($\rho = -0.65$, $P < 0.001$). However, residual hearing was not significantly correlated with SRI-Q at

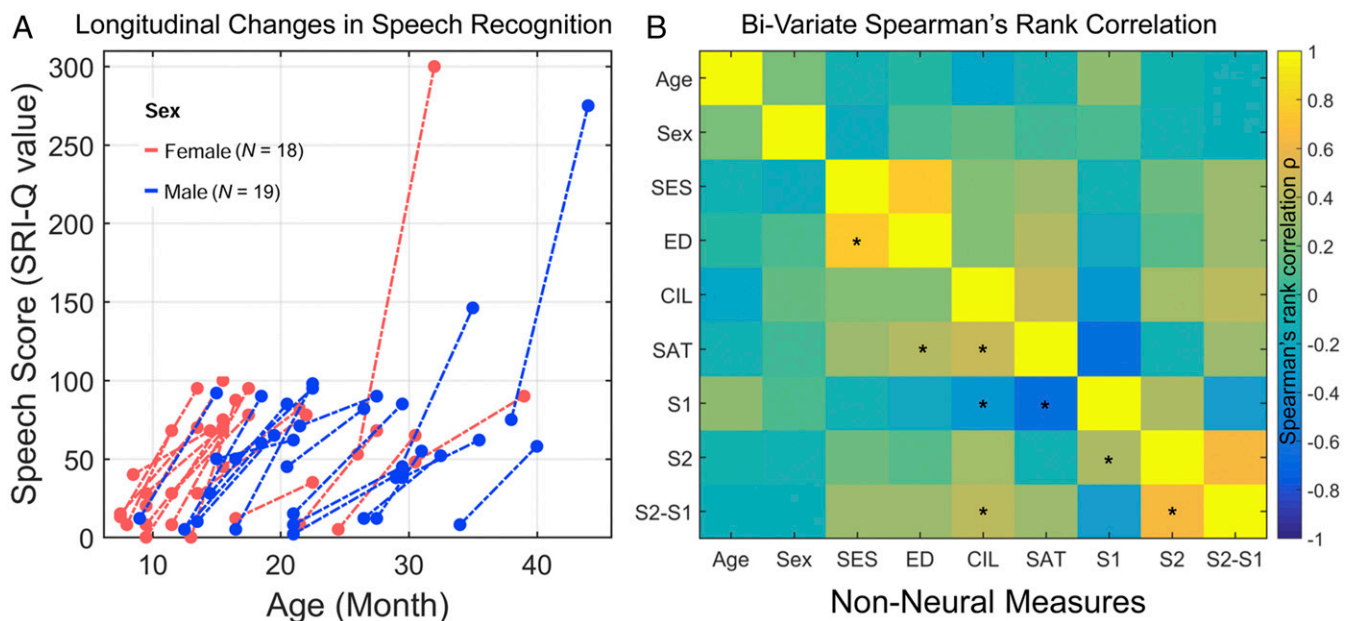


Fig. 2. Longitudinal changes in speech perception score (SRI-Q) for children in the CI group and Spearman's rank correlations between pairs of nonneural measures. (A) Scatterplot shows the SRI-Q scores before CI and 6-mo postactivation. The two scores are linked by a dashed line for each CI participant. (B) Colored correlation matrix indicating correlation strength between nonneural measures. Age, age of CI; CIL, lateralization of CI (0, unilateral; 1, bilateral implantation); ED, parents' education level; SAT, pre-CI speech awareness threshold (residual hearing); S1, pre-CI SRI-Q score; S2, 6-mo SRI-Q score; S2-S1, SRI-Q improvement. * $P < 0.05$.

6 mo postactivation ($\rho = -0.11$, $P = 0.51$) in our sample. Residual hearing was marginally correlated with SRI-Q improvement ($\rho = 0.28$, $P = 0.09$). SRI-Q improvement was significantly correlated with lateralization (unilateral vs. bilateral) of CI ($\rho = 0.43$, $P = 0.008$). SRI-Q improvement was not significantly correlated with sex ($\rho = -0.15$, $P = 0.39$), nor with duration of hearing-aid use (the time between hearing fitting and MRI scan) ($\rho = 0.22$, $P = 0.20$), and was marginally correlated with SES ($\rho = 0.30$, $P = 0.07$) and parents' education level ($\rho = 0.29$, $P = 0.07$).

Neural machine learning and patient classification. As discussed above, we conducted two types of analyses to compare neural morphological changes resulting from auditory deprivation. These two types of analyses (VBM and MVPS) provided us with complementary characterizations of the brain affected and unaffected by hearing deprivation. For each type of analysis, we constructed machine-learning classification models [support vector machine (SVM) with a nested leave-one-out cross-validation (LOOCV) procedure] (see *Materials and Methods* and *SI Appendix, Fig. S4* for details) to use presurgical neural morphological data to predict speech-perception improvement following CI activation. Here, we report results of these model performances in classifying children who use CIs into high vs. low SRI-Q improvement groups (Fig. 3A). To evaluate our two hypotheses, we separately incorporated brain templates of regions affected and unaffected (see *Materials and Methods* for details on the template construction approach) by auditory deprivation in the SVM models. The classification performances of these models were first compared against chance by using permutation tests with 10,000 iterations. Bootstrapping procedures (*Materials and Methods*) were then used to quantitatively compare model performances that were built using these brain templates and measures. Separate classification models were built for GM and WM to quantify the predictive effect of each tissue type. Whole-brain models (without applying a predefined template) were also constructed to evaluate how our hypothesis-driven

approach differs from a purely data-driven approach. Classification models were built with and without nonneural measures (i.e., demographic variables, pre-CI SAT, and pre-CI SRI-Q) and with only these measures for comparison purposes.

In general, our results support the neural preservation hypothesis. Convergence of evidence indicates that brain regions unaffected by auditory deprivation have the highest accuracy, sensitivity, and specificity in classifying children in the CI group into those with high or low speech-perception improvement (Fig. 3B). Bootstrapping procedures with 10,000 iterations confirm that models of unaffected brain regions did not differ from the whole-brain models and that they are reliably better than models of affected brain regions (see *SI Appendix, Table S2* for model comparisons). For example, although classification accuracy and specificity for the unaffected GM model were at around 76% and 82%, respectively, they were only 59% and 59% for the affected GM model, and 49% and 35% for nonneural measures (*SI Appendix, Table S3*). Additional results, which converge with the MVPS models, can be found in *SI Appendix Tables S4 and S5*. Those results also contain additional findings using different sets of model construction procedures [e.g., nested template definition procedure; that is, defining the affected/unaffected brain templates based on each training set, (*SI Appendix, Table S3*)], different machine-learning classifiers [e.g., linear and nonlinear SVM (*SI Appendix, Table S4*)], and feature selection procedures [e.g., using principle component analysis and different feature selection approaches to reduce the number of features and ensure each trained model (affected, unaffected, or whole-brain) contained the same number of features]. Moreover, it is worth pointing out that classification performances derived from models with nonneural measures were not significantly better than chance in classifying children in the CI group (Permutation test, $P = 0.53$). In general, the classification performance for each brain model did not improve even after the

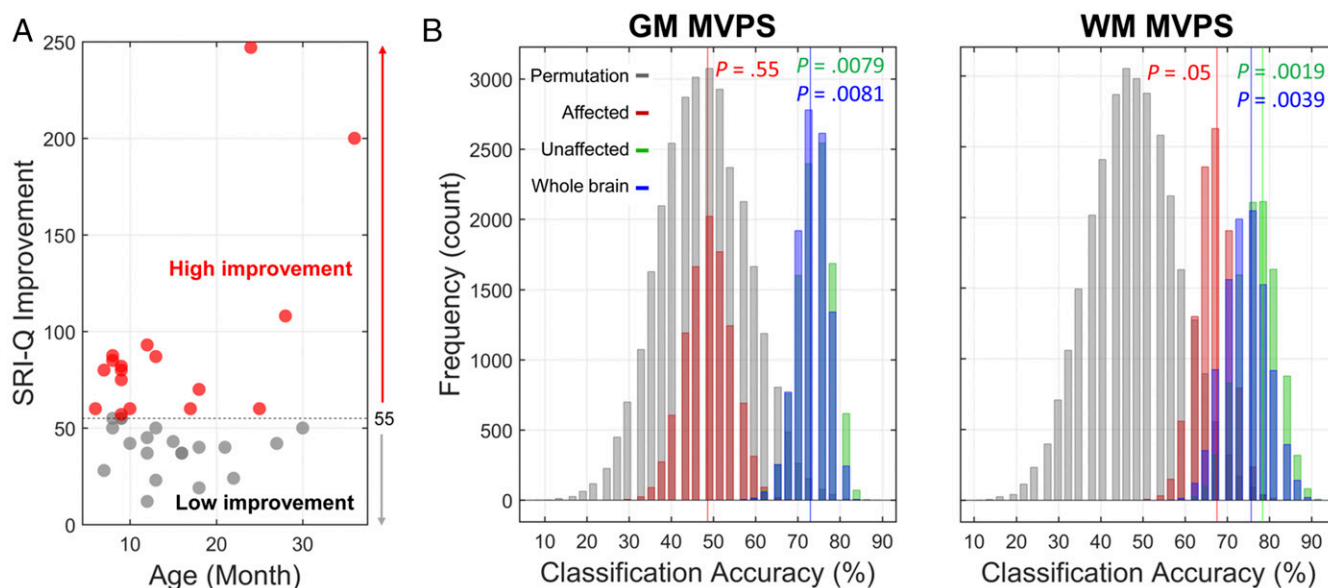


Fig. 3. High vs. low speech improvement CI subgroups and distributions of binary classification performance based on affected, unaffected, and whole-brain templates. (A) Children in the CI group were divided into two subgroups (high vs. low speech improvement) based on the median SRI-Q improvement score. (B) Histograms of distributions of classification accuracy for each template and permutation null distribution. SVM models using different brain templates were built. Three brain templates in each tissue type were used to restrict brain voxels and to define brain space for searching for the most informative voxels for optimizing and testing classification performance. Statistical significance of each template comparing with its corresponding null distribution are listed. In each fold of LOOCV, 90% of the data were used for model training, and 10% were reserved for model testing. This procedure was repeated 10,000 times for accurately estimating distribution of classification accuracy. This analysis was conducted for both tissue types (GM/WM) and neural morphological measures (density/MVPS). The unaffected models were as good as the models using all voxels from the whole brain. Only results from models using MVPS measure are reported here.

inclusion of nonneural measures as features (e.g., see SVM classification results reported in *SI Appendix, Tables S3 vs. S4*). Nevertheless, the classification performances of unaffected models still outperformed that of affected models (*SI Appendix, Table S5*). The predictive regions identified by an additional searchlight approach are widely distributed across frontal, temporal, and parietal cortices, and include the left STG, which is important for speech perception and learning (11, 18), and numerous areas in the putative cognitive network, such as the frontoparietal cortices (see Fig. 4A and *SI Appendix, Table S6* for all of the GM clusters that contributed to patient classification). As a quantitative analysis, Fig. 4 further revealed that most of the affected voxels had low predictive effect for both GM and WM (Fig. 4B), while most of the voxels with high predictive effect were in the unaffected areas (Fig. 4C) outside of the primary auditory cortex.

Neural machine learning and precise ranking prediction. In addition to making binary, high vs. low improvement patient classification to obtain clinically useful metrics, such as sensitivity and specificity, we further used machine learning for ranking approach to make a more precise prediction about postsurgical speech-perception outcomes. Because our primary outcome measure—the SRI-Q—is represented on an ordinal rather than a continuous scale, conventional regression techniques, such as support vector regression (31), are not suitable. Instead, a ranking SVM algorithm was used to predict each child's post-CI ranking in SRI-Q improvement relative to other patients in our sample. We found model predictions of both the unaffected and whole-brain ranking SVM models to be significantly better than chance for MVPS in GM [unaffected: $\rho_{(\text{predicted}, \text{observed})} = 0.44$, $P = 0.006$; whole-brain: $\rho_{(\text{predicted}, \text{observed})} = 0.44$, $P = 0.007$], whereas machine-learning models based on the affected brain regions failed to reach statistical significance [$\rho_{(\text{predicted}, \text{observed})} = -0.07$,

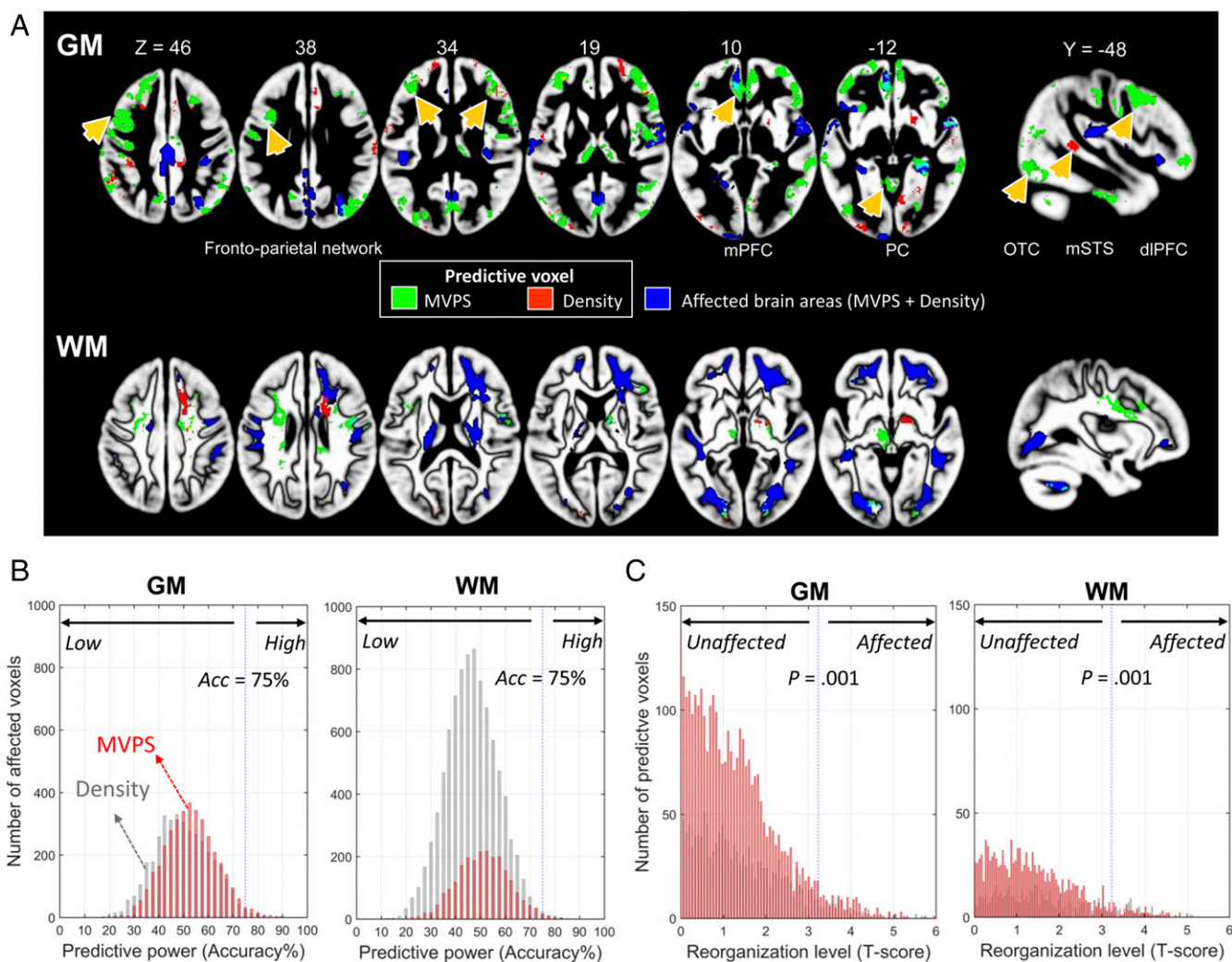


Fig. 4. Whole-brain searchlight patient classification results based on different types of neural measures (density and MVPS measures for both GM and WM). (A) GM and WM regions that best-classified children with CI into high vs. low improvement groups from the MVPS (green) and density (red) measures. Overlapping regions are shown in yellow. Affected regions are shown in blue for comparison purposes. Arrows point to regions that showed the greatest classification accuracy (see *SI Appendix, Table S6* for details). Region abbreviation: dlPFC, dorsal lateral prefrontal cortex; mPFC, medial prefrontal cortex; mSTS, middle superior temporal sulcus; OTC, occipital-temporal cortex; PC, posterior cortex. (B) Most of the affected voxels had low predictive power (classification accuracy) for both GM and WM. Voxels from the affected template were selected for quantitatively visualizing their predictive power (accuracy = 75% was highlighted). (C) Most of the voxels with high predictive power were located in the unaffected areas. The voxels with high predictive power (significant better than chance) were selected for quantitatively visualizing their reorganization level ($P = 0.001$ were highlighted). Reorganization level refers t-score derived from group comparison between CI and NH groups.

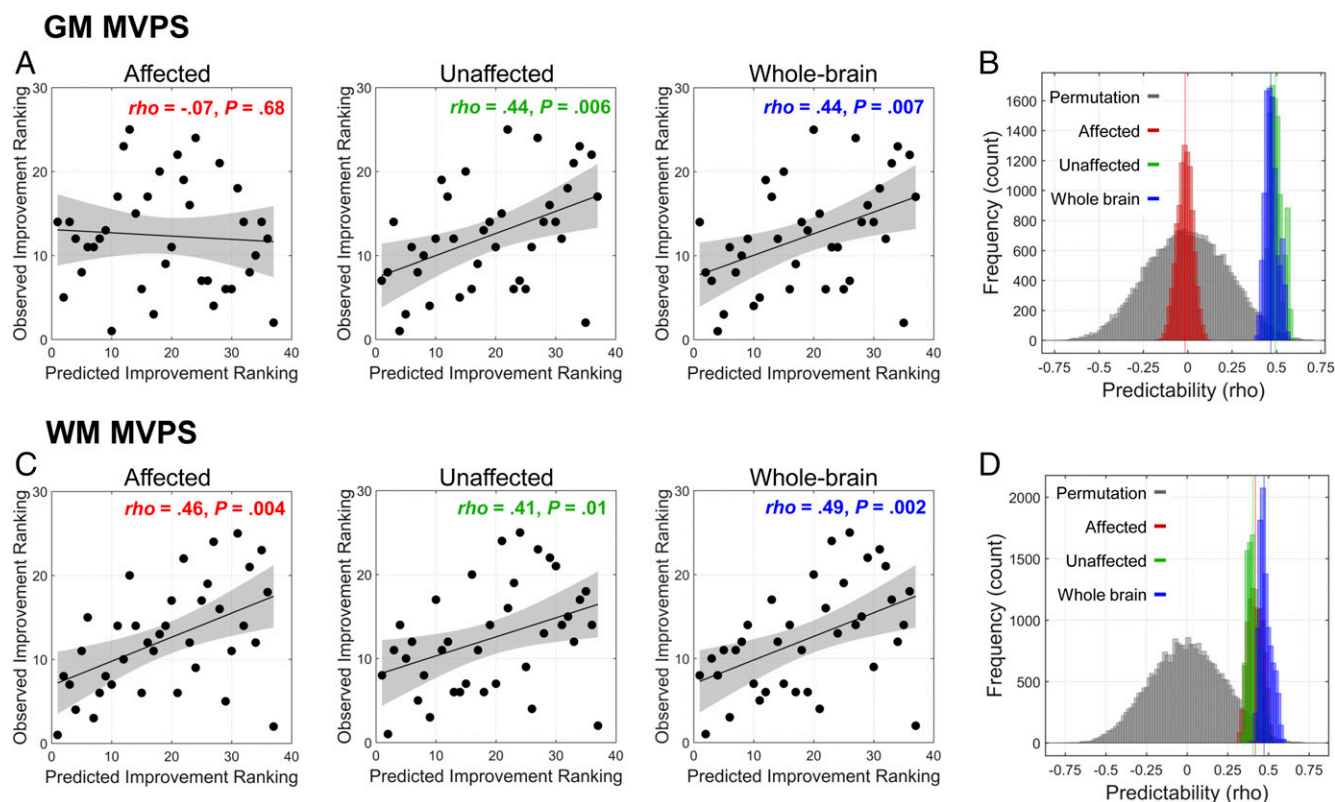


Fig. 5. Performance of RankSVM models. (A and C) The observed and predicted ranking scores of each patient's speech improvement from models built on GM and WM MVPS measures are shown. Spearman's rank correlations between the observed and predicted scores were calculated. (B and D) Permutation test (10,000 iterations) and bootstrapping procedures were used to determine the statistical significance and the reliability of the predictive models, respectively. In GM, the predictability of the unaffected model significantly outperformed the affected model. In WM, the predictability of the unaffected model was not significantly different from that of the affected model; all models performed significantly better than chance. See *SI Appendix Table S3* for the detailed statistical results of B and D.

$P = 0.68$] (Fig. 5A). Model comparisons further revealed that model performance of the unaffected template significantly outperformed that of the affected model for MVPS in GM (Fig. 5B) ($P_s < 0.001$). Model performance for all templates were significantly better than chance for WM (Fig. 5C) [affected: $\rho_{(\text{predicted}, \text{observed})} = 0.46, P = 0.004$; unaffected: $\rho_{(\text{predicted}, \text{observed})} = 0.41, P = 0.01$; whole-brain: $\rho_{(\text{predicted}, \text{observed})} = 0.49, P = 0.002$] and did not differ from each other (Fig. 5D) ($P_s > 0.1$). Predictions by none of the models based on VBM density were better than chance ($P_s > 0.1$). These results were confirmed by the 10,000-iteration bootstrapping procedure and permutation test (see Fig. 5B and D and *SI Appendix, Table S3* for details).

Discussion

The present study investigated how patterns of neural reorganization resulting from auditory deprivation might result in differences in spoken-language perception ability when the hearing is restored following CI activation in very young children. Most certainly, cochlear implantation has been tremendously effective in assisting spoken language development in children with bilateral, severe-to-profound, sensorineural hearing loss (1, 8, 32). CIs undoubtedly provide significant speech perception, spoken language acquisition (33), and academic gains (34) to children with significant sensorineural hearing loss. Despite these advances, we continue to see extensive variability in language outcomes among pediatric CI recipients. For example, Niparko et al. (8) found that even children who were implanted before 18 mo of age had language outcomes that lagged behind those of children with NH. In addition, the range of language outcomes among children who use CIs was much

more variable and fewer children attained age-typical performance in comparison with children with NH (8). Because successful spoken language development requires not only an intact peripheral auditory system but also the central nervous system to encode the bottom-up input, we hypothesized that variability in the brain would be associated with variability in language development, even after peripheral hearing was established. One obvious source of variability in the brain of children with hearing loss is the pattern of morphological changes resulting from auditory deprivation. To investigate this further, we first identified patterns of morphological changes following early-onset hearing loss and subsequently used patterns of neural changes to predict spoken language development in early-deafened children implanted when younger than 3.5 y. Supporting the neural preservation hypothesis, we found that mostly higher-level auditory and cognitive regions that are not affected by hearing deprivation best predicted pediatric CI candidates' future speech improvement. These unaffected regions included the dorsal auditory network, which is engaged in speech perception (11), and its functional variability is associated with individual differences in speech learning in adult learners (18).

The CI research to forecast outcomes has revealed two important predictors: age of implantation (earlier is better than later) and residual hearing (less-severe hearing loss leads to better post-CI outcomes than profound) (see refs. 8 and 29 for some of the best examples). These two predictors fare particularly well in conventional regression analyses when the sample size is large and when the variability in the age of implantation is also large. However, sizable variability in outcomes can be observed even when implantation takes place before the age of 5 y, and even

when implantation is as early as 18 mo (see ref. 8). Residual hearing, although statistically significant, may not have a strong impact on the individual patient because of its small effect size. For example, Niparko et al. (8) found an improvement in only about two points on the Reynell Developmental Language Scales over a 3-y period per 20-dB decrease (improvement) in the hearing threshold. As we have found, although residual hearing does correlate with pre-CI speech-perception scores, its accuracy and sensitivity in classifying individual children is at chance and is also substantially lower than neural measures (*SI Appendix, Tables S2 and S3*). Thus, when focusing on young CI candidates and clinical utility at the individual patient level, we must find more robust predictors than the age of implantation and residual hearing.

Over the past decades, investigators in the fields of auditory neuroscience and clinical audiology have extensively studied alterations of the nervous system caused by auditory deprivation. Knowledge gained from this research has influenced best practices in clinical decision making. For example, much is now understood about the neurological sequelae of auditory deprivation, focusing on differentiating innate versus experience-dependent effects, timing of deafness or auditory restoration (2), and cross-modal encroachment (35). An important revelation from this body of work is the advantage of auditory stimulation before 4 y of age to allow synaptogenesis to develop in the auditory system. Congenitally deaf children implanted after 7 y of age lack an N1 electrophysiological response, likely an indicator of deficits in higher-order auditory areas and decoupling of primary and association areas (3). These findings in humans corroborate with research in animals (especially the cat model), in which deafness is found to lead to abnormal activities in deep cortical layers where top-down and bottom-up inputs are likely to converge and interact (1, 35). The present effort builds on these comparisons of relatively younger and older children by determining how individual variability within younger children could provide meaningful information about later development.

It is worth noting that in the adult hearing-impairment literature, prolonged hearing restoration via hearing aid use has been associated with suppression of cross-modal reorganization in the auditory cortices (36). Our population differs in age, and had an abbreviated period of hearing aid use. CI intervention to limit the period of auditory deprivation is known to be advantageous and is a goal of standard clinical practice. Therefore, an abbreviated hearing aid experience is inherent in our study population of perilingually deaf children for whom candidacy evaluation had confirmed that amplification provided inadequate access to spoken language. Because hearing restoration via CI is more effective than hearing aid use for our population, we hypothesize, based on the previous findings (e.g., ref. 36), that CI would also prevent cross-modal encroachment. Although we found those unaffected regions mainly outside of the auditory cortices to be predictive of spoken language outcomes in the current study, it is possible that the auditory cortex becomes more predictive post-CI than pre-CI, as our participants gain experience with sound. This possibility requires additional studies that employ functional imaging techniques that are compatible with CI devices, such as functional near-infrared spectroscopy, to longitudinally measure postsurgical neuroplasticity.

Regarding clinical audiology, research on CI use in children has not only focused on when surgery should ideally take place but also what predictors besides the age of implantation and residual hearing levels can best predict postsurgical outcome. Several investigators have also studied the use of presurgical neural measures for predicting postsurgical language outcomes, although not all of these studies focused exclusively on early-deafened children. Lee et al. (26) conducted the first study attempting to link presurgical neural metabolism, measured by PET, with postactivation speech-perception ability. Across 15 patients ranging from 2- to

20-y old, they found a significant correlation between the extent of hypometabolism in the auditory cortex and speech-perception outcomes. Similarly, Giraud and Lee examined patients from 1- to 11-y old and found a significant correlation between neural metabolism and speech outcomes postactivation (27). As discussed earlier, drastic differences in neurophysiological outcomes have been observed among children who were implanted early (before 4 y) versus late (after 7 y). Thus, it is difficult to interpret and translate findings from studies using CI recipients with such a broad age range. Moreover, correlational measures tend to overestimate predictive power (9, 24, 25). In the case of the aforementioned PET studies (26), correlative neural hypometabolism could overpredict language outcomes.

It has been argued that if neural measures are to become useful in clinical practice, they must predict outcomes for unseen individuals based on models developed previously with other individuals (9). Using LOOCV procedures in combination with machine-learning algorithms could potentially overcome these issues and estimate the true predictive power of the neural measures (24, 25). The present study recruited only children younger than 3.5 y and used SVM learning algorithms in conjunction with a leave-one-out validation method for classification and prediction. This approach avoids having a broad age range of patients and minimizes the possibility of prediction overestimation.

Another important methodological feature of our study is that we used objective measures of neural morphology as the primary predictive measure. These neuroanatomical measures are not affected by sedation, which can influence blood pressure and flow and consequently the BOLD signal as measured by fMRI (28, 37, 38). Because our patients have hearing loss, requiring them to listen to auditory stimuli can potentially introduce confounds to the degree that residual hearing is predictive of outcomes (8). It is worth noting that a recent study also used machine learning to predict language-learning outcomes in CI candidates (39). Those children were again of a broader age range (8–67 mo), and they conducted a listening task under sedation. The BOLD signal was used for prediction. The high variability in patients' age, and relying on BOLD responses measured under sedation may explain why only a restricted region of the brain showed statistically significant predictive results in this data-driven study.

It is important to note that not all neuroanatomical measures were equally predictive of postsurgical speech-perception improvement outcomes for CI candidates. We found models relying on MVPS to outperform models relying on VBM density measures. While MVPS captures each child's morphological pattern against a normative group, VBM density measures tissue characteristics without making reference to a typically developing group. It is perhaps this norm-referencing nature of MVPS that enabled it to produce more accurate predictive models.

Advances in noninvasive neuroimaging techniques in recent decades have revolutionized human cognitive neuroscience research and clinical management of a multitude of medical conditions. More recently, prediction has been the dominant focus of basic and clinical research to unlock the potential of noninvasive neuroimaging. These investigations have explored the extent to which individual variation in neuroanatomical and neurophysiological processes may relate to variation in disease state and treatment outcomes. Measures from advanced neuroimaging are often more accurate predictors than nonneural measures alone (see ref. 9 for a review). For example, T1-weighted neuroanatomical data (local GM densities derived from VBM) and resting-state fMRI activity obtained at the beginning of an 8-wk math tutoring program significantly predicted performance improvement (40); no behavioral measures (IQ, working memory, and mathematical abilities) predicted performance improvement. López-Barroso et al. (41) found a significant correlation between word learning and WM integrity and

connectivity using diffusion tensor imaging. If outcome prediction is to be an essential contribution of cognitive neuroscience (9) for young, early-deafened children with a severe-to-profound hearing loss, the most pressing need is to accurately predict the high variability of language development outcomes, preferably with a theory-driven approach that provides a mechanistic explanation of the findings. The present study, focusing on objective, presurgical neural measures and their predictive power of postimplantation outcomes in early-deafened children, is one step toward this goal. The present study also lays the foundation for personalized therapy for children with CI (42).

Conclusions

By using presurgical, objective neuroanatomical data, the present study was able to construct machine-learning models that classified CI candidates into different speech-perception improvement groups with a high level of accuracy, specificity, and sensitivity, as well as making more precise predictions about each patient's ranking in speech outcome. These predictive models support the neural preservation hypothesis, which postulates that regions in the auditory association and cognitive networks to be most predictive of future speech-perception development. Despite our preliminary success, one limitation of the present study is that only CI candidates from one medical center were enrolled. Future research should include CI candidates from different medical centers to test the generalizability of our hypothesis and machine-learning models.

Materials and Methods

Participants. Children in the CI group were all implanted at the Ann & Robert H. Lurie Children's Hospital of Chicago. They received a T1-weighted structural whole-brain MRI as part of their clinical preoperative evaluation to identify cochlear labyrinthine, eighth cranial nerve, and gross brain abnormalities. Consent from the parent(s) or guardian(s) was obtained to use these scans and children's clinical outcome data for our research, as approved by the Institutional Review Boards of Lurie Children's Hospital of Chicago and Northwestern University. Children in the CI group had a bilateral severe to profound ($n = 27$) or moderate to severe ($n = 10$) sensorineural hearing loss. Exclusion criteria included: (i) English not the dominant language spoken at home; (ii) developmental delay, complicating conditions, or additional disabilities expected to impact CI outcomes; (iii) gross brain malformation, severely malformed cochlea, cochlear nerve deficiency, and postmeningitic deafness; and (iv) partial electrode array insertion likely to limit the development of speech perception. The same scale was used to obtain the speech perception score before CI in our standardized test batteries. All CI candidates were fitted with hearing aids before surgery, which is one of the CI candidacy requirements. The average time between hearing aid fitting and MRI scan was 7.49 mo (SD = 4.93). The time was significantly correlated with age of implant ($r = 0.41$, $P = 0.013$). Children in the NH group ($n = 40$) were selected from the NIH MRI Study of Normal Brain Development (PedsMRI; <https://pediatricmri.nih.gov/nihpd/info/index.html>). They had normal hearing at the time of image acquisition and were matched with the children in the CI group in age, sex, and SES (group differences, $P_s > 0.5$).

Auditory and Speech-Perception Outcome Measures. The primary outcome measure was the SRI-Q (29), as used in the Childhood Development after Cochlear Implantation (CDaCI) Study (8, 23). The SRI-Q consists of a hierarchy of measures of speech recognition abilities from parental reports for children with the lowest auditory abilities to direct measures of speech perception for children with better auditory abilities all on the same scale. Specifically, for very young children whose speech perception cannot be directly obtained, the SRI-Q includes the Infant-Toddler Meaningful Auditory Integration Scale/Meaningful Auditory Integration Scale (IT-MAIS/MAIS), which is a parental report and criterion-referenced rating scale of the child's basic auditory skills. For children with higher auditory abilities and who are suitable for direct speech measures, the SRI-Q includes the Early Speech Perception Test (ESP), Multisyllabic Lexical Neighborhood Test (MLNT), the Lexical Neighborhood Test (LNT), the Phonetically Balanced Word Lists-Kindergarten (PBK), and the Children's Hearing in Noise Test (HINT-C). The choice of these tests was based on the status of the patient (e.g., age, developmental ability, and hearing aptitude), and whether a criterion level was met. To reflect their position in this speech/auditory ability

hierarchy, scores from these measures were rescaled to 0–100 for the IT-MAIS/MAIS, 101–200 for the ESP, 201–300 for the MLNT, 301–400 for the LNT, 401–500 for the PBK, and 501–600 for the HINT-C in quiet resulting in a range of scores from 0 to 600 (29). It is important to note that this battery allows for a comparison of performance across the entire age range of our research. Because of the young age of our subjects, all were given the lowest level of the SRI-Q before surgery (IT-MAIS/MAIS). At 6-mo postactivation, most children's speech perception ability was still captured by the IT-MAIS or MAIS, except for three whose auditory abilities have improved enough to receive direct measures of speech perception, including the ESP and MLNT. In addition, residual hearing, obtained via live voice while using hearing aids in the to-be implanted ear before surgery, served as a proxy for residual hearing.

MRI Acquisition and Analyses. The T1-weighted images of the CI candidates were acquired either on a 3T General Electric MR scanner (DISCOVERY MR750) using a 3D brain volume sequence (BRAVO) ($n = 15$) or on a 3T Siemens scanner (MAGNETOM Skyra) using a MPRAGE sequence ($n = 22$) at Lurie Children's Hospital of Chicago. Scanning parameters were optimized for improving signal-to-noise ratio and increased GM and WM contrast (GE scanner: TE = 3.8 ms, TR = 9.4 ms, flip angle = 12°, matrix = 512 × 512, 188 slices of 1.4 mm thickness, voxel size = 0.47 × 0.47 × 1.4 mm; Siemens scanner: TE = 2.60–3.25 ms, TR = 1,900 ms, flip angle = 9°, matrix = 256 × 256, 176 slices of 0.9-mm thickness, voxel size = 0.9 × 0.9 × 0.9 mm). T1-weighted MR scans of the children in the NH group were selected and downloaded from the NIH open-access database (TE = 10 ms, TR = 500 ms, flip angle = 90°, matrix = 256 × 192, 30–60 slices of 3-mm thickness, voxel size = 1 × 1 × 3 mm) (43).

MRI images were processed using the FreeSurfer software package (freesurfer.net/) v6.0 and SPM 12 (Wellcome Department of Imaging Neuroscience, London, United Kingdom; www.fil.ion.ucl.ac.uk/spm/). To increase image quality, we used a nonlocal mean filter algorithm (44) to remove noise. The de-noised images were then subjected to intensity inhomogeneity correction and resampling into 1 × 1 × 1-mm voxel size and were further segmented into three different tissue types (GM, WM, and cerebrospinal fluid). Age-appropriate tissue probability templates (45, 46) were used to guide and improve tissue segmentation. Study-specific GM and WM templates were constructed using the Diffeomorphic Anatomical Registration through Exponentiated Lie Algebra (DARTEL) algorithm (47). The segmented brain tissues for each participant were then spatially normalized by using the DARTEL templates. Finally, all normalized images were modulated and smoothed with a 4-mm full width at half-maximum Gaussian kernel. These smoothed images were used for further analyses.

The VBM analysis consisted of conducting two-sample t tests of the GM and WM density between CI and NH groups, with or without demographic variables (i.e., age, sex, and SES) as covariates. We restricted this statistical comparison with voxels with a minimum probability value of 0.5 (absolute threshold masking) to avoid possible edge effects around tissue borders. The group-level statistical maps were thresholded at voxel-level $P < 0.001$ initially, and all reported brain areas have been corrected $P < 0.05$ at the cluster-level using the family-wise error (FWE) correction as implemented in the SPM package. The peak coordinates of clusters were reported in the Montreal Neurological Institute space.

The MVPS analysis involved first z -transforming the GM/WM density maps to ensure that pattern similarity effects were not related to the absolute magnitude of tissue density of individual voxels but spatial pattern. A searchlight approach (48) with a four-voxel-radius sphere centered on each voxel (~200 voxels in average across regions) was used to define the local spatial extent (*SI Appendix, Fig. S1*). Because the choice of spatial sphere extent is somewhat arbitrary, different searchlight extents (four- and six-voxel-radius sphere) were tested to ensure consistency. The local tissue pattern similarity (here we used Pearson correlation r value) was then calculated for each of the local spheres for each pair of individuals between CI and NH groups (i.e., between-group MVPS) and within the NH group (within-NH group MVPS), respectively. We used Pearson correlation because it captures the local neuroanatomical pattern similarity between each pair of individuals while not being influenced by the absolute magnitude of tissue density for each voxel. The similarity r values were subsequently normalized by Fisher's r -to- z transformation. Each child in the CI group had 40 between-group similarity brain maps for each tissue type, while each individual in the NH group had 39 within-NH group similarity brain maps (excluding the pair by itself) for each tissue type. These maps were then averaged for each child and each tissue separately. Finally, 37 between-group and 40 within-NH group MVPS maps were generated. To quantitatively measure the extent of neuro-morphological reorganization by the

multivoxel pattern for the children in the CI group, we compared the between-group and within-NH group MVPS maps, in which within-NH group MVPS maps were considered as a normal baseline. Therefore, any brain area with a similarity value that is significantly lower in the between-group compared with that in the within-NH group would be considered a reorganized region. Two-sample t tests were used to assess significant differences between between-group and within-NH group similarities for each voxel. The group-level statistical maps were thresholded at cluster-level FWE-corrected $P < 0.05$.

The affected templates were built by combining the voxels that showed significant group difference from both (VBM + MVPS/similarity) types of morphological analysis. The unaffected templates were built by including the voxel that did not show significant group difference from both types of analysis. It is worth to note that the templates were not built by using a conjunction (overlap) of VBM and MVPS/similarity data. By using templates of combination rather than conjunction, we can ensure that when SVM models for either morphological data (VBM or MVPS/similarity) are built, all potentially affected/unaffected voxels would be included.

Multivariate Machine-Learning Classification. Neural data from each child, masked by either the affected or unaffected brain templates, were converted into an S -by- V matrix, where S is the number of participants and V is the number of features (voxels), and the matrix was normalized so that mean = 0 and SD = 1. No spatial masking was used for the whole-brain models. To avoid overfitting of a large number of predictive features and to optimize model parameters, we used a nested LOOCV procedure with SVM classifier. We implemented a k -fold LOOCV procedure with three levels (inner, middle, and outer) of nesting for dimension reduction/feature selection, parameters optimization, and model validation (SI Appendix, Fig. S4). At the inner level, we employed two different feature dimension reduction or feature selection procedures for validation. The first approach is to use principal component analysis to reduce the number of feature (i.e., voxel) dimensions for model training. Principal component analysis can preserve the maximal amount of individual variance while keeping the same number of features (i.e., principal components) across brain templates. The second approach is to adopt the SVM recursive feature elimination procedure (49) to select the most relevant voxels. This feature selection procedure was independent of the outer-level model testing. Thus, we used $n - 1$ participants in the training dataset for training models, and the remaining participants was used for testing at each LOOCV fold. Features were ranked based on model weights and recursively removed until the optimal pattern that gave maximal model performance was found. To ensure that the difference in the number of voxels between brain templates would not be a confounding factor, we selected the same number of voxels (i.e., 5,000 voxels, ~5% of a total number of voxels) across different brain templates. Features with the best model performance were selected based on the inner level of LOOCV and fed into the other levels of LOOCV. Moreover, besides using predefined templates (affected vs. unaffected), we also used a nested template-definition procedure to select affected/unaffected voxels that were determined based on the training set in each fold of the cross-validation; therefore, different affected/unaffected templates would be generated during the classification. At the middle level, we used two different support vector classifiers (i.e., linear and nonlinear SVM) to validate our results. For linear SVM classifier, we used the default parameters (i.e., $C = 1$, $\gamma = 1/\text{number of features}$). For nonlinear SVM classifier, we generated 100 models with wide range of parameters on the selected features (or dimensions). We chose the model parameters (C and γ) in a nonlinear SVM classifier with radical base function with highest generalization accuracy based on the training dataset. Finally, at the outer level of LOOCV, we used the selected features and optimized parameters (for nonlinear SVM) from the inner and middle level to build SVM models based on all training datasets, then applied the trained model to classify an unseen sample (the remaining participant from the $n - 1$ participant for the LOOCV or 10% of the samples for the 10-fold cross-validation). We used functions from MATLAB package LIBSVM (50) combined with our in-house scripts to conduct our analyses. Classification accuracy, sensitivity and specificity, and area under the curve of the receiver operating characteristic curve (compares sensitivity vs. specificity) were calculated and subjected to various comparisons (e.g., affected vs. unaffected brain regions) using a bootstrapping and permutation test.

Whole-Brain Searchlight Classification Analysis. We used the searchlight algorithm (48) to further identify brain regions that show high discrimination between high and low SRI-Q improvement for the children who use CIs by using functions from the LIBSVM toolbox (50) and CoSMoMVA toolbox (51). The 10-fold cross-validation procedure was employed. At each voxel, local neuro-morphological values (VBM density or MVPS) within a spherical searchlight (four-voxel-radius sphere) were extracted for each child. Therefore, in each spherical searchlight,

a $V \times C$ value matrix was constructed, where V referred to voxel, C referred to the number of children. This matrix was input to a linear SVM classifier for training and testing. Then a whole-brain classification accuracy map was generated (SI Appendix, Fig. S5). We thresholded the classification map by utilizing a permutation procedure, in which we randomly assigned the group labels into two new groups and repeated the LOOCV procedure 1,000 times. Statistical significance was determined by comparing the actual classification accuracy and the permutation-based null distribution for each spherical searchlight.

Multivariate Ranking SVM. To further examine how precisely we can predict the speech improvement of individual CI-using children, we used linear multivariate ranking SVM (RankSVM) for building and validating predictive models. Due to the ordinal nature of the IT-MAIS/MAIS score, conventional support vector regression that relied on continuous variables could not be applied. The machine-learned ranking technique is suitable in this case. The machine-learned ranking technique is a type of supervised machine-learning technique that has been widely used in web search and information retrieval (52–54) and whose goal is to construct ordered models. The trained models can be used to sort unseen data according to their degrees of relevance or importance. RankSVM is a pairwise machine-learned ranking technique that can be used to form ranking models by minimizing a regularized margin-based pairwise loss. That is, SVM computes the weight vector that maximizes the differences of data pairs in ranking. Thus, the model training requires investigating every data pair as potential candidates of support vectors, and the number of data pairs is quadratic to the size of the training set. Therefore, the computational efficiency is low when the number of pairs and feature space is large. Implementing RankSVM with a primal truncated Newton method that is known to be fast for optimization problem (55, 56) could significantly increase the computation efficiency. Here, we employed linear RankSVM (57) with a primal Newton method (56) to construct ranking models and generate predictions with a 10-fold cross-validation procedure. For each fold in the cross-validation, speech improvement scores in the training set were first converted into an ordinal array and then fed into the linear RankSVM to build models for each template. We accessed the prediction power of the SRI-Q improvement ranking score from a given brain measure by calculating the Spearman's rank correlation between the predicted and observed values [$\rho(\text{predicted}, \text{observed})$]. In each prediction attempt, the statistical significance of the correlation was evaluated using a non-parametric permutation ($n = 10,000$) approach.

Nonparametric Permutation Test. To determine whether the classification accuracy or ranking prediction of each model was significantly better than chance, we employed a nonparametric permutation procedure to generate a chance distribution (the null hypothesis) and further used it to test whether the observed classification performance for each model occurred by chance. To achieve this, we first randomly assigned the group labels into two new groups or randomized the ranking scores and repeated the LOOCV procedure. This randomization procedure was repeated 10,000 times, and the 95th percentile points of each distribution were used as the critical values for a one-tailed t test of the null hypothesis with $P = 0.05$.

Bootstrapping Procedure for Estimating Distributions of Model Performance. To statistically test whether classification performance or ranking prediction of the unaffected model significantly outperformed that of the affected model, we additionally employed a bootstrapping procedure by randomly splitting the data into 10-fold, in which 90% of the data were used for training and 10% for testing, repeated 10 times. The same nested LOOCV procedure was used for each model. This random sampling procedure (bootstrapping) was repeated 10,000 times for generating a performance distribution for each model (e.g., Fig. 3B). Statistical comparisons between models and 95% confidence intervals for each model were then calculated.

ACKNOWLEDGMENTS. We thank Haejung Shin, Beth Tournis, and members of the cochlear implant team at the Ann & Robert H. Lurie Children's Hospital of Chicago for their support of this research. This research is supported by grants from Knowles Hearing Center at Northwestern University and Ann & Robert H. Lurie Children's Hospital of Chicago, and donations from the Global Parent Child Resource Centre Limited and the Dr. Stanley Ho Medical Development Foundation awarded to The Chinese University of Hong Kong.

1. Kral A, Sharma A (2012) Developmental neuroplasticity after cochlear implantation. *Trends Neurosci* 35:111–122.
2. Kral A (2013) Auditory critical periods: A review from system's perspective. *Neuroscience* 247:117–133.
3. Sharma A, Dorman MF, Spahr AJ (2002) A sensitive period for the development of the central auditory system in children with cochlear implants: Implications for age of implantation. *Ear Hear* 23:532–539.
4. Kral A, Hartmann R, Tillein J, Heid S, Klinke R (2002) Hearing after congenital deafness: Central auditory plasticity and sensory deprivation. *Cereb Cortex* 12:797–807.
5. Klinke R, Kral A, Heid S, Tillein J, Hartmann R (1999) Recruitment of the auditory cortex in congenitally deaf cats by long-term cochlear electrostimulation. *Science* 285:1729–1733.
6. Hackett TA (2011) Information flow in the auditory cortical network. *Hear Res* 271:133–146.
7. Kral A, Tillein J, Heid S, Klinke R, Hartmann R (2006) Cochlear implants: Cortical plasticity in congenital deprivation. *Prog Brain Res* 157:283–313.
8. Niparko JK, et al.; CDaCI Investigative Team (2010) Spoken language development in children following cochlear implantation. *JAMA* 303:1498–1506.
9. Gabrieli JD, Ghosh SS, Whitfield-Gabrieli S (2015) Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85:11–26.
10. Hickok G, Poeppel D (2004) Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67–99.
11. Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
12. Emmorey K, Allen JS, Bruss J, Schenker N, Damasio H (2003) A morphometric analysis of auditory brain regions in congenitally deaf adults. *Proc Natl Acad Sci USA* 100:10049–10054.
13. Smith KM, et al. (2011) Morphometric differences in the Heschl's gyrus of hearing impaired and normal hearing infants. *Cereb Cortex* 21:991–998.
14. Shibata DK (2007) Differences in brain structure in deaf persons on MR imaging studied with voxel-based morphometry. *AJNR Am J Neuroradiol* 28:243–249.
15. Hill J, et al. (2010) Similar patterns of cortical expansion during human development and evolution. *Proc Natl Acad Sci USA* 107:13135–13140.
16. Gilmore JH, et al. (2012) Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cereb Cortex* 22:2478–2485.
17. Wong PC, et al. (2007) Volume of left Heschl's gyrus and linguistic pitch learning. *Cereb Cortex* 18:828–836.
18. Wong PC, Perrachione TK, Parrish TB (2007) Neural characteristics of successful and less successful speech and word learning in adults. *Hum Brain Mapp* 28:995–1006.
19. Mueller S, et al. (2013) Individual variability in functional connectivity architecture of the human brain. *Neuron* 77:586–595.
20. Sowell ER, et al. (2004) Longitudinal mapping of cortical thickness and brain growth in normal children. *J Neurosci* 24:8223–8231.
21. Sharma A, Gilley PM, Dorman MF, Baldwin R (2007) Deprivation-induced cortical reorganization in children with cochlear implants. *Int J Audiol* 46:494–499.
22. Sharma A, Nash AA, Dorman M (2009) Cortical development, plasticity and reorganization in children with cochlear implants. *J Commun Disord* 42:272–279.
23. Fink NE, et al.; CDaCI Investigative Team (2007) Childhood development after cochlear implantation (CDaCI) study: Design and baseline characteristics. *Cochlear Implants Int* 8:92–116.
24. Whelan R, Garavan H (2014) When optimism hurts: Inflated predictions in psychiatric neuroimaging. *Biol Psychiatry* 75:746–748.
25. Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage* 45(1 Suppl):S199–S209.
26. Lee DS, et al. (2001) Cross-modal plasticity and cochlear implants. *Nature* 409:149–150.
27. Giraud AL, Lee HJ (2007) Predicting cochlear implant outcome from brain organization in the deaf. *Restor Neurol Neurosci* 25:381–390.
28. Peltier SJ, et al. (2005) Functional connectivity changes with concentration of sevoflurane anesthesia. *Neuroreport* 16:285–288.
29. Wang NY, et al.; CDaCI Investigative Team (2008) Tracking development of speech recognition: Longitudinal data from hierarchical assessments in the Childhood Development after Cochlear Implantation Study. *Otol Neurotol* 29:240–245.
30. Barnard JM, et al.; CDaCI Investigative Team (2015) A prospective longitudinal study of U.S. children unable to achieve open-set speech recognition 5 years after cochlear implantation. *Otol Neurotol* 36:985–992.
31. Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V (1997) Support vector regression machines. *Advances in Neural Information Processing Systems*, eds Mozer MC, Jordan MI, Petsche T (MIT Press, Cambridge, MA), Vol 9, pp 155–161.
32. Nicholas JG, Geers AE (2007) Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. *J Speech Lang Hear Res* 50:1048–1062.
33. Svirsky MA, Robbins AM, Kirk KI, Pisoni DB, Miyamoto RT (2000) Language development in profoundly deaf children with cochlear implants. *Psychol Sci* 11:153–158.
34. Marschark M, Rhoten C, Fabich M (2007) Effects of cochlear implants on children's reading and academic achievement. *J Deaf Stud Deaf Educ* 12:269–282.
35. Kral A (2007) Unimodal and cross-modal plasticity in the 'deaf' auditory cortex. *Int J Audiol* 46:479–493.
36. Shiell MM, Champoux F, Zatorre RJ (2015) Reorganization of auditory cortex in early-deaf people: Functional connectivity and relationship to hearing aid use. *J Cogn Neurosci* 27:150–163.
37. Kiviniemi VJ, et al. (2005) Midazolam sedation increases fluctuation and synchrony of the resting brain BOLD signal. *Magn Reson Imaging* 23:531–537.
38. Mhuircheartaigh RN, et al. (2010) Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: A functional magnetic resonance imaging study using propofol. *J Neurosci* 30:9095–9102.
39. Tan L, et al. (2015) A semi-supervised support vector machine model for predicting the language outcomes following cochlear implantation based on pre-implant brain fMRI imaging. *Brain Behav* 5:e00391.
40. Supekar K, et al. (2013) Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proc Natl Acad Sci USA* 110:8230–8235.
41. López-Barroso D, et al. (2013) Word learning is mediated by the left arcuate fasciculus. *Proc Natl Acad Sci USA* 110:13168–13173.
42. Wong PCM, Vuong LC, Liu K (2017) Personalized learning: From neurogenetics of behaviors to designing optimal language training. *Neuropsychologia* 98:192–200.
43. Almlí CR, Rivkin MJ, McKinstry RC; Brain Development Cooperative Group (2007) The NIH MRI study of normal brain development (Objective-2): Newborns, infants, toddlers, and preschoolers. *Neuroimage* 35:308–325.
44. Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M (2012) New methods for MRI denoising based on sparseness and self-similarity. *Med Image Anal* 16:18–27.
45. Shi F, et al. (2011) Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS One* 6:e18746.
46. Richards JE, Sanchez C, Phillips-Meek M, Xie W (2016) A database of age-appropriate average MRI templates. *Neuroimage* 124:1254–1259.
47. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38:95–113.
48. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103:3863–3868.
49. De Martino F, et al. (2008) Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage* 43:44–58.
50. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM Trans Intel Syst Technol* 2:27:1–27:27.
51. Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front Neuroinform* 10:27.
52. Richardson M, Prakash A, Brill E (2006) Beyond PageRank: Machine learning for static ranking. *Proceedings of the 15th International Conference on World Wide Web*, (ACM, New York), pp 707–715.
53. Burges CJ (2010) From RankNet to LambdaRank to LambdaMART: An overview (Microsoft Research, Redmond, WA), Microsoft Research Technical Report MSR-TR-2010-82.
54. Liu T-Y (2009) Learning to rank for information retrieval. *Found Trends Inf Ret* 3:225–331.
55. Chapelle O (2007) Training a support vector machine in the primal. *Neural Comput* 19:1155–1178.
56. Chapelle O, Keerthi SS (2010) Efficient algorithms for ranking with SVMs. *Inf Retrieval* 13:201–215.
57. Joachims T (2002) Optimizing search engines using clickthrough data. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York), pp 133–142.