

# Supporting Information

Abrams et al. 10.1073/pnas.1602948113

## SI Methods

**Participants.** The Stanford University Institutional Review Board approved the study protocol. Parental consent and children's assent were obtained for all evaluation procedures, and children were paid for their participation in the study.

A total of 32 children were recruited from around the San Francisco Bay Area for this study. Six participants were excluded because of excessive movement, one was excluded because of infrequent contact with their biological mother, who also was unavailable for a vocal recording, and another participant was excluded because of scores in the "severe" range on standardized measures of social function. Parent reports from the final sample of 24 participants showed that these children were raised in families with a wide range of socioeconomic backgrounds, with 25% of participants coming from households earning  $\leq$ \$100K/y. Socioeconomic status was not correlated with children's social communication skills as assessed using the SRS-2 ( $P > 0.50$ ), the key behavioral measure described in the analysis. All children were required to have a full-scale IQ  $>80$ , as measured by the WASI (41). All children were right-handed and had no history of: neurological, psychiatric, or learning disorders; no personal or family (first degree) history of developmental cognitive disorders or heritable neuropsychiatric disorders; no evidence of significant difficulty during pregnancy, labor, delivery, or the immediate neonatal period; and no abnormal developmental milestones as determined by neurologic history and examination. Participants were the biological offspring of the mothers whose voices were used in this study (i.e., none of our participants were adopted, and therefore none of the mothers' voices were from an adoptive mother), and all participants were raised in homes that included their mother. Participants' neuropsychological and language characteristics are provided in Tables S1 and S2, respectively.

**Data Acquisition Parameters.** All fMRI data were acquired in a single session at the Richard M. Lucas Center for Imaging at Stanford University. Functional images were acquired on a 3-T Signa scanner (General Electric) using a custom-built head coil. Participants were instructed to stay as still as possible during scanning, and head movement was minimized further by placing memory-foam pillows around the participant's head. A total of 29 axial slices (4.0-mm thickness, 0.5-mm skip) parallel to the anterior/posterior commissure line and covering the whole brain were imaged by using a T2\*-weighted gradient-echo spiral in-out pulse sequence (43) with the following parameters: repetition time (TR) = 3,576 ms; echo time = 30 ms; flip angle = 80°; one interleaf. The 3,576-ms TR is the sum of (i) the stimulus duration of 956 ms; (ii) a 300-ms silent interval buffering the beginning and end of each stimulus presentation (600 ms total of silent buffers) to avoid backward and forward masking effects; (iii) the 2,000-ms volume acquisition time; and (iv) an additional 20-ms silent interval which helped the stimulus computer maintain precise and accurate timing during stimulus presentation. The field of view was 20 cm, and the matrix size was 64 × 64, providing an in-plane spatial resolution of 3.125 mm. Reduction of blurring and reduction of signal loss arising from field inhomogeneities was accomplished by the use of an automated high-order shimming method before data acquisition.

**fMRI Task.** Auditory stimuli were presented in 10 separate runs, each lasting 4 min. One run consisted of 56 trials of mother's voice, female control voices, environmental sounds, and catch trials, which were pseudorandomly ordered within each run. Stimulus presentation order was the same for each subject. Each

stimulus lasted 956 ms. Before each run, child participants were instructed to play the "kitty cat game" during the fMRI scan. While lying down in the scanner, children were first shown a brief video of a cat and were told that the goal of the cat game was to listen to a variety of sounds, including "voices that may be familiar," and to push a button on a button box only when they heard kitty cat meows (catch trials). The function of the catch trials was to keep the children alert and engaged during stimulus presentation. During each run, four or five exemplars of each stimulus type (i.e., nonsense words produced by that child's mother, female control voices, and environmental sounds) and three catch trials were presented. At the end of each run, the children were shown another engaging video of a cat. Across the 10 runs, a total of 48 exemplars of each stimulus condition were presented to each subject. Speech stimuli were presented to participants in the scanner using E-Prime v1.0 (Psychological Software Tools, 2002). Participants wore custom-built headphones designed to reduce the background scanner noise to  $\sim 70$  adjusted dB (dBA) (44, 45). Headphone sound levels were calibrated before each data-collection session, and all stimuli were presented at a sound level of 75 dBA. Participants were scanned using an event-related design. Auditory stimuli were presented during silent intervals between volume acquisitions to eliminate the effects of scanner noise on auditory discrimination. One stimulus was presented every 3,576 ms, and the duration of the silent period was not jittered. The total silent period between stimulus presentations was 2,620 ms and consisted of a 300-ms silent period, 2,000 ms for a volume acquisition, another 300 ms of silence, and a 20-ms silent interval that helped the stimulus computer maintain precise and accurate timing during stimulus presentation.

**fMRI Preprocessing.** fMRI data collected in each of the 10 functional runs were subject to the following preprocessing procedures. The first five volumes were not analyzed to allow for signal equilibration. A linear shim correction was applied separately for each slice during reconstruction by using a magnetic field map acquired automatically by the pulse sequence at the beginning of the scan. Functional images were first realigned to their first volume using SPM8 analysis software ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) and then corrected for deviant volumes resulting from spikes in movement. Translational movement in millimeters ( $x$ ,  $y$ ,  $z$ ) was calculated based on the SPM8 parameters from the realignment procedure for each subject. We used a despiking procedure (46) similar to those implemented in the Analysis of Functional NeuroImages (AFNI) toolkit maintained by the National Institute of Mental Health (Bethesda, MD) (47). Volumes with movement exceeding 0.5 voxels (1.562 mm) or spikes in global signal exceeding 5% were interpolated using adjacent scans. The majority of volumes repaired occurred in isolation. After the interpolation procedure, images were further corrected for slice-timing errors. To normalize the functional images to standard Montreal Neurological Institute (MNI) space, each individual's functional images were first coregistered to their structural T1 images, and their T1 images were then transformed to MNI space. The transformation parameters were subsequently applied to the functional images, which were then resampled to 2-mm isotropic voxels. Finally, functional images were smoothed with a 6-mm full-width half-maximum Gaussian kernel to decrease spatial noise prior to statistical analysis.

**Movement Criteria for Inclusion in fMRI Analysis.** For inclusion in the fMRI analysis, we required that each functional run have a maximum scan-to-scan movement of  $<6$  mm and that no more

than 15% of volumes were corrected in the despiking procedure. Moreover, we required that all individual subject data included in the analysis consist of at least seven functional runs that met our criteria for scan-to-scan movement and percentage of volumes corrected; subjects who had fewer than seven functional runs that met our movement criteria were not included in the data analysis. All 24 participants included in the analysis had at least seven functional runs that met our movement criteria. Fifteen of the participants had 10 runs of data that met these movement criteria; two subjects had nine runs of data that met movement criteria; five subjects had eight runs of data; and two subjects had seven runs that met criteria.

**Voxelwise Analysis of fMRI Activation.** The goal of the voxelwise analysis of fMRI activation was to identify brain regions that showed differential activity levels in response to mother's voice, female control voices, and environmental sounds. Brain activation related to each speech task condition was first modeled at the individual subject level using boxcar functions with a canonical hemodynamic response function and a temporal derivative to account for voxelwise latency differences in hemodynamic response. Environmental sounds were not modeled to avoid collinearity, and this stimulus served as the baseline condition. Low-frequency drifts at each voxel were removed using a high-pass filter (0.5 cycles/min), and serial correlations were accounted for by modeling the fMRI time series as a first-degree autoregressive process (48). Voxelwise  $t$  statistics maps for each condition were generated for each participant using the general linear model (GLM) along with the respective contrast images. Group-level activation was determined using individual-subject contrast images and second-level one-sampled  $t$  tests. The main contrasts of interest were [mother's voice vs. female control voices], [female control voices vs. mother's voice], [female control voices vs. environmental sounds], [female control voice 1 vs. female control voice 2], and [female control voice 2 vs. female control voice 1]. Significant clusters of activation were determined using a voxelwise statistical height threshold of  $P < 0.01$ , with familywise error corrections for multiple spatial comparisons ( $P < 0.01$ ; 128 voxels) determined using Monte Carlo simulations (49, 50) using a custom Matlab script. To examine GLM results in the IC, a small subcortical brain structure, we used a small-volume correction at  $P < 0.01$ . To define specific cortical regions, we used the Harvard-Oxford probabilistic structural atlas (51) with a probability threshold of 25%.

**Signal-Level Analysis.** Group mean activation differences for key brain regions identified in the whole-brain univariate analysis were calculated to examine the basis for the [mother's voice > female control voices] group differences (Fig. 2). This analysis was performed because stimulus differences can result from a number of different factors. For example, both mother's voice and female control voices could elicit reduced activity relative to baseline, and significant stimulus differences could be driven by greater negative activation in response to female control voices. Significant stimulus differences were inherent to this ROI analysis, because they are based on results from the whole-brain GLM analysis (52); however, the results provide important information regarding the magnitude and sign of results in response to both stimulus conditions. The baseline for this analysis was calculated as the brain response to environmental sounds. A number of ROIs were constructed using coordinates reported in previous studies: IC ROIs were 5-mm spheres centered at  $\pm 6$ ,  $-33$ ,  $-11$  (53, 54); primary auditory cortical ROIs (Te1.0, Te1.1, and Te1.2) were identified a priori from cytoarchitectonic maps (55), and bilateral pSTS coordinates were identified from previous investigations of voice-selective cortex (11, 37). All other ROI coordinates used in the signal-level analysis were based on peaks identified in the [mother's voice > female control voices] group map. This analysis included 13 ROIs in bilateral STC, nine ROIs

in bilateral frontal cortex, seven ROIs in bilateral parietal cortex, three ROIs in bilateral occipital cortex, one ROI in the anterior cingulate, and five subcortical structures. Cortical ROIs were defined as 5-mm spheres, and subcortical ROIs were 2-mm spheres, centered at the peaks in the [mother's voice > female control voices] group map. Signal level was calculated by extracting the  $\beta$ -value from individual subjects' contrast maps for the [mother's voice > environmental sounds] and [female control voices > environmental sounds] comparisons. The mean  $\beta$ -value within each ROI was computed for both contrasts in all subjects. The group mean  $\beta$  and its SE for each ROI are plotted in Figs. S1–S4.

**Effective Connectivity Analysis.** Effective connectivity analysis was performed using gPPI (42), a method more sensitive than PPI to context-dependent differences in connectivity. At the individual subject level, the time series from the seed region is first deconvolved to uncover neuronal activity and then multiplied with the task design waveforms to form an interaction term. This interaction term is then convolved with the hemodynamic response function (HRF) to form the gPPI regressor, and the resulting time series is regressed against all other voxels in the brain. The goal of this analysis was to examine connectivity patterns of the voice-selective network, with a focus on voice-selective temporal cortex, which is hypothesized to be a hub of the network linking auditory regions with heteromodal regions important for reward and affective processing of these sounds (1, 21). Therefore, we constructed four STS/STG ROIs that were identified from the univariate analysis [mother's voice > female control voices] group  $t$  map (Fig. 2; mSTS and aSTS/STG) and two ROIs that were identified from previous investigations of voice-selective cortex (pSTS) (11, 37). For the STS regions identified from the univariate analysis, we identified peaks in this  $t$  map for mSTS and aSTS/STG regions bilaterally and constructed nonoverlapping 5-mm spherical ROIs centered at these peaks. These six ROIs then were used as seeds in six separate whole-brain gPPI models. Significant clusters of activation were determined using a voxelwise statistical height threshold of  $P < 0.01$ , with familywise error corrections for multiple spatial comparisons ( $P < 0.01$ ; 128 voxels) determined using Monte Carlo simulations (49, 50).

**Brain-Behavior Analysis.** Regression analysis was used to examine the relationship between brain signatures of mother's voice perception and social and language skills. Social function was assessed using the Social Communication subscale of the SRS-2 (22). For our measure of language function, we used the CELF-4 (23), a standard instrument for measuring language function in neurotypical children. Regression analyses were conducted using the Core Language score of the CELF, a measure of general language ability. Brain-behavior relationships were examined using analysis of both activation levels and effective connectivity. We first performed a voxelwise regression analysis in which the relation between fMRI activity and social and language measures was examined using images contrasting mother's voice vs. female control voices. We then performed a voxelwise regression analysis between STC connectivity and standardized social and language measures using gPPI images generated for each participant by contrasting responses to mother's voice vs. responses to female control voices. Significant clusters were determined using a voxelwise statistical height threshold of  $P < 0.01$ , with familywise error corrections for multiple spatial comparisons ( $P < 0.01$ ; 128 voxels) determined using Monte Carlo simulations (49, 50).

**Functional Brain Connectivity and Prediction of Social Function.** To examine the robustness and reliability of brain connectivity between STS and reward, affective, salience detection, and face-processing brain regions for predicting social communication scores, we performed a confirmatory cross-validation (CV)

analysis that employs a machine-learning approach with balanced fourfold CV combined with linear regression (25). In this analysis, we extracted individual subject connectivity beta values, taken from the [mother's voice > female control voices] gPPI contrast, in left-hemisphere NAc, right-hemisphere amygdala, fusiform cortex, and hippocampus (i.e., the left-hemisphere aSTS gPPI seed) and in right-hemisphere OFC, AI, and dACC (the right-hemisphere pSTS gPPI seed). Mean gPPI beta values for each brain connection (e.g., left-hemisphere aSTS seed to left-hemisphere NAc) were separately entered as the independent variable in a linear regression analysis with SRS-2 social communication standard scores as the dependent variable. First,  $r_{(\text{predicted, observed})}$ , a measure of how well the independent variable predicts the dependent variable, was estimated using a balanced fourfold CV procedure. Data were divided into four folds so that the distributions of dependent and independent variables were balanced across folds. Data were randomly assigned to four folds, and the independent and dependent variables were tested in one-way ANOVAs, repeating as necessary until both ANOVAs were insignificant to guarantee balance across the folds. A linear regression model was built using three folds, leaving out the fourth, and this model was used to predict the data in the omitted fold. This procedure was repeated four times to compute a final  $r_{(\text{predicted, observed})}$  representing the correlation between the data predicted by the regression model and the observed data. Finally, the statistical significance of the model was assessed using a nonparametric testing approach. The empirical null distribution of  $r_{(\text{predicted, observed})}$  was estimated by generating 1,000 surrogate datasets under the null hypothesis that there was no association between social communication subscores and brain connectivity.

**Stimulus Design Considerations.** Previous studies investigating the perception (2, 5) and neural bases (8, 9) of mother's voice processing have used a design in which one mother's voice serves as a control voice for another participant. However, for a number of reasons, in this study we used a design in which all participants heard the same two control voices. First, we wanted to be able to perform analyses comparing brain responses between the two control voices (see *Results, Analysis of Control Voices* in the main text), which would not have been possible had the participants heard different control voices. There also was an important practical limitation with using mother's voices as control voices for other children: Although we make every effort to recruit children from a variety of communities in the San Francisco Bay Area, some level of recruitment occurs through contact with specific schools, and in other instances our participants refer their friends to our laboratory for inclusion in our studies. In these cases, it is a reasonable possibility that our participants may know other mothers involved in the study and therefore may be familiar with these mothers' voices; that familiarity would limit the control we were seeking in our control voices. Importantly, the Health Insurance Portability and Accountability Act (HIPAA) guidelines are explicit that participant information is confidential, and therefore there would be no way to probe whether a child knew any of the other families involved in the study. Given these analytic and practical considerations, we concluded that it would be best to use the same two control voices, which we knew were unfamiliar to the participants, for all participants' data collection.

**Stimulus Recording.** Recordings of each mother were made individually while her child was undergoing neuropsychological testing. Mother's voice stimuli and control voices were recorded in a quiet conference room using a Shure PG27-USB condenser microphone connected to a MacBook Air laptop computer. The audio signal was digitized at a sampling rate of 44.1 kHz and was A/D converted with 16-bit resolution. Mothers were positioned in the conference room to avoid early sound wave reflections

from contaminating the recordings. To provide a natural speech context for the recording of each nonsense word, mothers were instructed to repeat three sentences, each of which contained one of the nonsense words, during the recording. The first word of each of these sentences was their child's name, which was followed by the words "that is a," followed by one of the three nonsense words. A hypothetical example of a sentence spoken by a mother for the recording was "Johnny, that is a keebudishawlt." Before beginning the recording, mothers were instructed on how to produce these nonsense words by repeating them to the experimenter until the mothers had reached proficiency. Importantly, mothers were instructed to say these sentences using the tone of voice they would use when speaking with their child during an engaging and enjoyable shared learning experience (e.g., if their child asked them to identify an item at a museum). The vocal recording session resulted in digitized recordings of the mothers repeating each of the three sentences ~30 times to ensure multiple high-quality samples of each nonsense word for each mother. A second class of stimuli included in the study was nonspeech environmental sounds. These sounds, which included brief recordings of laundry machines, dishwashers, and other household sounds, were taken from a professional sound effects library.

**Stimulus Postprocessing.** The goal of stimulus postprocessing was to isolate the three nonsense words from the sentences that each mother spoke during the recording session and to normalize them for duration and rms amplitude for inclusion in the fMRI stimulus presentation protocol, the pleasantness and excitement ratings experiment, and the mother's voice identification task. First, a digital sound editor (Audacity: <https://sourceforge.net/projects/audacity/>) was used to isolate each utterance of the three nonsense words from the sentences spoken by each mother. The three best versions of each nonsense word were selected based on the audio and vocal quality of the utterances (i.e., eliminating versions that were mispronounced, included vocal creak, or were otherwise not ideal exemplars of the nonsense words). These nine nonsense words then were normalized to 956 ms in duration, the mean duration of the nonsense words produced by the female control voices, using Praat software similar to previous studies (56). A 10-ms fade (ramp and damp) was performed on each stimulus to prevent click-like sounds at the beginning and end of the stimulus, and then stimuli were equated for rms amplitude. These final stimuli were evaluated for audibility and clarity to ensure that postprocessing manipulations had not introduced any artifacts into the samples. The same process was performed on the control voices and environmental sounds to ensure that all stimuli presented in the fMRI experiment were the same duration and rms amplitude.

**Pleasantness and Excitement Ratings for Vocal Stimuli.** To examine the relative pleasantness and engagingness of the vocal stimuli used in the fMRI experiment, we performed two behavioral experiments in an independent cohort of 27 children (mean age  $\pm$  SD:  $11.1 \pm 1.2$  y; sex: 10 female, 17 male). Participants were seated in a quiet room in front of a laptop computer, and headphones were placed over their ears. In one experiment, participants were presented with trials of either a mother's voice sample or a control voice sample of the nonsense word "teebudishawlt." After each stimulus presentation, the participant rated the vocal sample for pleasantness on a four-point scale as "very unpleasant," "unpleasant," "pleasant," or "very pleasant." In the second experiment the same procedures were used, but participants rated each vocal sample on a four-point scale for engagingness. Because there was concern that 8- to 10-y-old children might not understand the meaning of the word "engaging" in the context of this experiment, consistent with a previous study (57), we used the following four-point scale for this experiment: "totally boring," "a little boring," "a little exciting," or "totally exciting." Each vocal stimulus (i.e., 24

mother's voice samples plus the two control voice samples) was presented once to each child in both the pleasantness and engagingness ratings tasks. The order of stimulus presentation was randomized for each participant and experiment; half of the participants performed the pleasantness ratings task first, and the other half of the participants performed the engagingness ratings task first. The vocal samples used in these behavioral experiments are the same as those used in the fMRI experiment. To examine statistical differences between ratings for mother's voice and female control voices, we performed independent samples *t* tests comparing the mean ratings for mother's voice samples and participant ratings for both control voices.

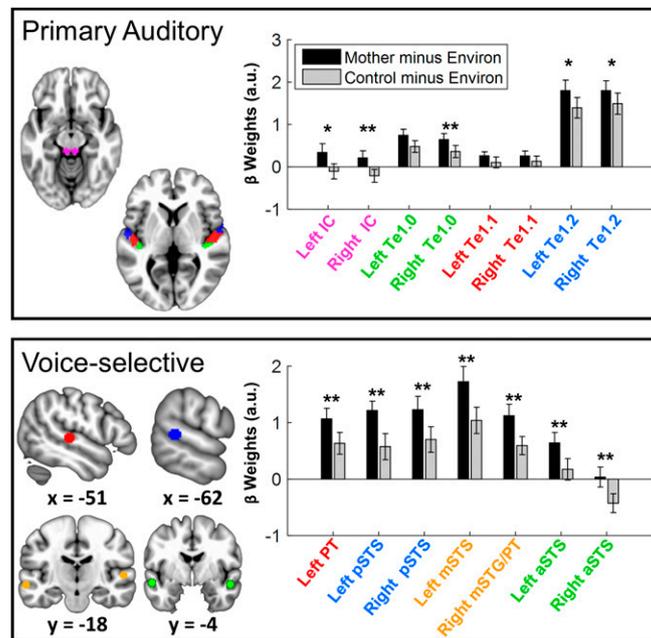
**Postscan Speaker Identity Recognition Task.** All participants who participated in the fMRI experiment completed an auditory behavioral test following the fMRI scan. The goal of the Speaker Identity Recognition Task was to determine if the participants could reliably discriminate their mother's voice from female control voices. Participants were seated in a quiet room in front of a laptop computer, and headphones were placed over their ears. In each trial, participants were presented with a recording of a multisyllabic nonsense word spoken by either the participant's mother or by a control mother, and the task was to indicate whether or not the participant's mother spoke the word. The multisyllabic nonsense words used in this behavioral task were the same samples used in the fMRI task. Each participant was presented with 54 randomly ordered nonsense words: 18 produced by the participant's mother and the remaining 36 produced by female control voices.

## SI Results

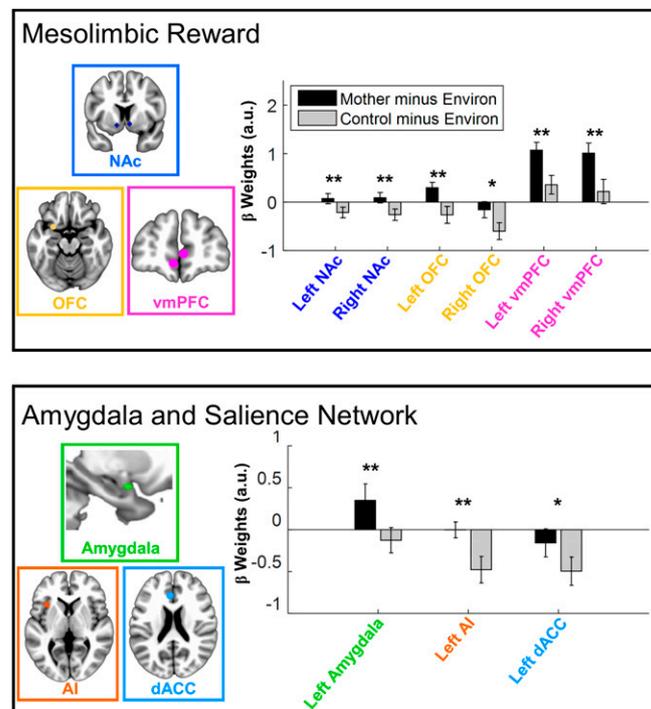
**fMRI Sex Difference Analysis.** We examined whether male and female participants showed different univariate activity for the [mother's voice > female control voices] contrast. We performed an ROI analysis within all 38 brain regions identified for the [mother's voice > female control voices] contrast in the group results (Fig. 2). This analysis included 13 ROIs in bilateral STC,

nine ROIs in bilateral frontal cortex, seven ROIs in bilateral parietal cortex, three ROIs in bilateral occipital cortex, one ROI in the anterior cingulate, and five subcortical structures. Signal level was calculated by extracting the  $\beta$ -value from individual subjects' contrast maps for the [mother's voice > female control voice] comparison. The mean  $\beta$ -value within each ROI was computed for this contrast in all subjects. Because of the small number of participants in each group (7 females; 17 males), we performed Wilcoxon rank sum tests on mean  $\beta$ -values within each ROI between female and male groups. Results from this analysis showed that none of the 38 regions identified in the combined group analysis showed a significant sex difference ( $P < 0.01$ ) for the [mother's voice > female control voices] contrast.

**Control Voice Analysis.** Behavioral ratings acquired from an independent cohort of children (*Pleasantness and Excitement Ratings for Vocal Stimuli*, above) showed that one of the female control voices had significantly greater pleasantness ratings compared to the other female control voice ( $P < 0.001$ ). The mean ( $\pm$  SD) rating for control voice 1 was 3.29 ( $\pm$  0.78) and for control voice 2 was 2.26 ( $\pm$  0.90). We therefore performed whole-brain and ROI analyses to examine whether more pleasant vocal features alone were sufficient to elicit increased activity in the brain systems highlighted in the [mother's voice > female control voices] contrast (Fig. 2). A whole-brain analysis was performed on the [female control 1 > female control 2] contrast and the [female control 2 > female control 1] contrast, and results were thresholded at  $P < 0.01$  and 128 voxels. An ROI analysis also was performed for these same two contrasts using the same 38 ROIs identified for the signal-level analysis. Signal level was calculated by extracting the  $\beta$ -value from individual subject's contrast maps for the [female control 1 > female control 2] and [female control 2 > female control 1] comparisons. The mean  $\beta$ -value within each ROI was computed for this contrast in all subjects, and one-sample *t* tests were performed on mean  $\beta$ -values within each ROI for both contrasts.

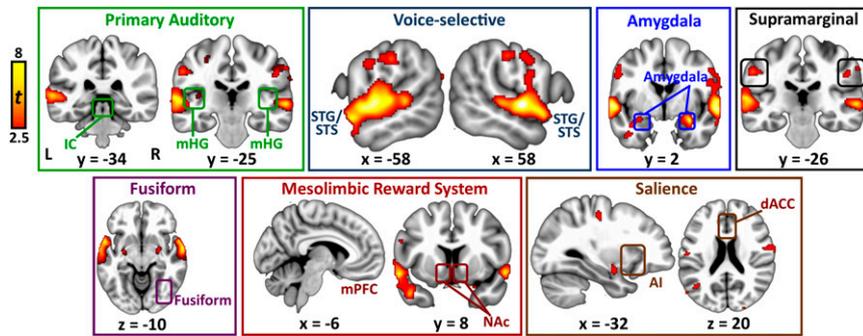


**Fig. 51.** Signal levels in primary auditory regions (*Upper*) and voice-selective cortex (*Lower*) in response to mother's voice and female control voices. Primary auditory regions were identified a priori from previous auditory studies (IC ROIs) (56) and from cytoarchitectonic maps (Te ROIs) (55), and voice-selective regions were selected for signal-level analysis based on previous investigations of voice-selective cortex (bilateral pSTS; refs. 11, 37) or their identification in the [mother's voice > female control voices] contrast (bilateral mSTS and aSTS; see Fig. 2 in the main text). Values plotted for mother's voice and female control voices are referenced to duration- and energy-matched environmental sounds, e.g., [mother's voice > environmental sounds]. The signal-level analysis was performed because stimulus-based differences in fMRI activity can result from a number of different factors. Significant differences were inherent to this ROI analysis, because they are based on results from the whole-brain GLM analysis (52); however, results provide important information regarding the magnitude and sign of fMRI activity. \*\* $P < 0.01$ ; \* $P < 0.05$ .

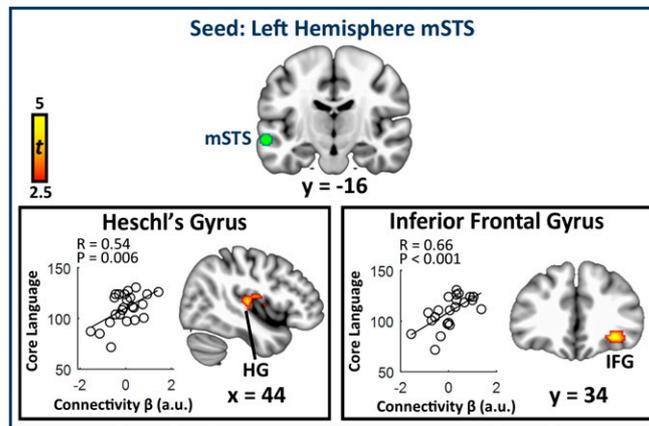


**Fig. 52.** Signal levels in mesolimbic reward regions (*Upper*) and the amygdala and salience network (*Lower*) in response to mother's voice and female control voices. Regions were selected for signal-level analysis based on their identification in the [mother's voice > female control voices] contrast (Fig. 2 in the main text). Values plotted for mother's voice and female control voices are referenced to duration- and energy-matched environmental sounds, e.g., [mother's voice > environmental sounds]. NAc and amygdala ROIs are 2-mm spheres centered at the peak for these regions in the [mother's voice > female control voices] contrast; all other ROIs in these bar graphs are 5-mm spheres centered at the peak for these regions in the [mother's voice > female control voices] contrast (Fig. 2 in the main text). \*\* $P < 0.01$ ; \* $P < 0.05$ .





**Fig. 55.** Brain activity in response to female control voices compared to environmental sounds. Compared to environmental sounds, female control voices elicit greater activity throughout a wide extent of voice-selective STG and STS (*Upper Center Left*), bilateral amygdala (*Upper Center Right*) and supramarginal gyrus (*Upper Right*), and a small extent of left hemisphere mHG (*Upper Left*). In contrast to brain activity in response to mother's voice, univariate results comparing female control voices to environmental sounds failed to identify brain regions in reward, salience, and face-processing regions or in the IC.



**Fig. 56.** Connectivity of left-hemisphere voice-selective cortex and language abilities. The whole-brain connectivity map shows that children's CELF core language covaried with the strength of functional coupling between the left-hemisphere mSTS (*Upper*) and right-hemisphere HG (*Lower Left*) and BA 47 of inferior frontal gyrus (IFG) (*Lower Right*). Scatterplots show the distributions and covariation of mSTS connectivity strength in response to mother's voice and standardized scores of language abilities in these children. Greater language scores are associated with greater brain connectivity between the STS and these brain regions.

**Table S1. Demographic and IQ measures**

| Measure               | TD, <i>n</i> = 24 | Population mean (SD) |
|-----------------------|-------------------|----------------------|
| Gender ratio, M:F     | 17:7              |                      |
| Age, y                | 10.22 ± 1.44      |                      |
| IQ, WASI scale        |                   |                      |
| Full-scale IQ         | 119.08 ± 12.00    | 100 (15)             |
| Verbal IQ             | 120.42 ± 13.87    | 100 (15)             |
| Performance IQ        | 113.54 ± 13.80    | 100 (15)             |
| Achievement, WIAT     |                   |                      |
| Word reading          | 114.04 ± 9.18     | 100 (15)             |
| Reading comprehension | 114.79 ± 9.02     | 100 (15)             |

Demographic and mean IQ scores are shown for the full sample. F, female; M, male; TD, typically developing children; WIAT, Wechsler Individual Achievement Test.

**Table S2. Neuropsychological measures of social and language abilities**

| Measure                               | TD, <i>n</i> = 24           | Population mean (SD) |
|---------------------------------------|-----------------------------|----------------------|
| <b>SRS-2</b>                          |                             |                      |
| Total Standard Score                  | 43.13 ± 5.24                | 30.9 (25.3)          |
| Social Communication Standard Score*  | 42.29 ± 4.63                |                      |
| Social Motivation Standard Score*     | 44.75 ± 6.22                |                      |
| <b>CELF-4</b>                         |                             |                      |
| Core Language Standard Score          | 109.92 ± 15.24 <sup>†</sup> | 100 (15)             |
| Receptive Language Standard Score     | 104.29 ± 17.06 <sup>†</sup> | 100 (15)             |
| Expressive Language Standard Score    | 113.67 ± 15.91 <sup>†</sup> | 100 (15)             |
| Concepts and following directions     | 9.71 ± 3.37                 | 10 (3)               |
| Recalling sentences                   | 11.54 ± 3.48                | 10 (3)               |
| Formulated sentences                  | 12.21 ± 3.13                | 10 (3)               |
| Word classes – total                  | 13.04 ± 2.94                | 10 (3)               |
| Receptive                             | 13.00 ± 2.77                | 10 (3)               |
| Expressive                            | 13.08 ± 3.16                | 10 (3)               |
| Spoken paragraphs                     | 10.17 ± 2.37                | 10 (3)               |
| Familiar sequences                    | 10.75 ± 3.40                | 10 (3)               |
| <b>CTOPP</b>                          |                             |                      |
| Phonological Awareness Standard Score | 107.50 ± 10.73              | 100 (15)             |
| Alternate Rapid Naming Standard Score | 99.25 ± 17.27               | 100 (15)             |
| Elision                               | 11.58 ± 2.02                | 10 (3)               |
| Blending words                        | 10.92 ± 2.43                | 10 (3)               |
| Nonword repetition                    | 9.46 ± 2.02                 | 10 (3)               |
| Rapid color naming                    | 9.92 ± 3.67                 | 10 (3)               |
| Rapid object naming                   | 9.83 ± 3.12                 | 10 (3)               |

Neuropsychological measures of Language and Phonetic abilities are shown for the full sample. CELF-4, *Clinical Evaluation of Language Fundamentals*, 4th edition (23); CTOPP, Comprehensive Test of Phonological Processing; SRS-2, *Social Responsiveness Scale, Second Edition* (22); TD, typically developing children.

\*Population means and SDs are not reported for Social Communication and Social Motivation Standard Scores (22).

<sup>†</sup>Word Classes 1 was used instead of Word Classes 2 to compute standardized scores for two subjects that were below the age of 7 y.

**Table S3. Superior temporal cortex seed coordinates for gPPI analysis**

| Brain region          | Coordinates  |
|-----------------------|--------------|
| Left-hemisphere pSTS  | –63, –42, 9  |
| Left-hemisphere mSTS  | –60, –16, –8 |
| Left-hemisphere aSTS  | –54, –6, –12 |
| Right-hemisphere pSTS | 57, –31, 5   |
| Right-hemisphere mSTS | 56, –20, 4   |
| Right-hemisphere aSTS | 52, –2, –14  |

**Audio File S1.** Exemplar of nonsense word “keebudishawlt” used in fMRI task produced by control female 1.

[Audio File S1](#)

**Audio File S2.** Exemplar of nonsense word “peebudishawlt” used in fMRI task produced by control female 1.

[Audio File S2](#)

**Audio File S3.** Exemplar of nonsense word “teebudishawlt” used in fMRI task produced by control female 1.

[Audio File S3](#)

**Audio File S4.** Exemplar of nonsense word “keebudishawlt” used in fMRI task produced by control female 2.

[Audio File S4](#)

**Audio File S5.** Exemplar of nonsense word “peebudishawlt” used in fMRI task produced by control female 2.

[Audio File S5](#)

**Audio File S6.** Exemplar of nonsense word “teebudishawlt” used in fMRI task produced by control female 2.

[Audio File S6](#)