Supporting Information

Experimental design: fMRI Voice Localizer. The Voice Localizer task was administered to 15 Hearing controls (age 30.73±5.46). A modified version of a classical fMRI voice localizer(Belin et al., 2000) was implemented to exclude any lexical vocalization. Three categories of stimuli were used: human neutral vocal (NV; from the Montreal Affective Voices dataset), scrambled human vocal (SCRB) and object (OB) sounds. The human NV belonged to 20 adult speakers and consisted of single articulations of the vowel /a/. The SCRB stimuli were obtained from the NV by randomly mixing their magnitude and the phase of each Fourier component while keeping global energy (root mean square) and envelope similar with the original sound; this condition was introduced to remove some low-level feature and isolate higher-level voice selective regions. OB stimuli consisted of sounds from man-made artefacts (e.g. train, cars, trumpets) that had been normalized for loudness using a root mean square function.

In the MRI scanner, a block-designed one-back identity task was implemented for this experiment in a single run that lasted approximately 12 minutes and consisted of 30 blocks, ten for each of the three experimental conditions. In each block, a single audio-file was delivered containing a sequence of 16 stimuli, which belonged to the same condition (i.e. NV, SCRB, OB) and lasted for about 1000 ms each with a 500ms ISI; in one to three occasions per block, the exact same stimulus was consecutively repeated that the participant had to detect. The presentation of sound blocks was alternated with that of resting-state silent inter-blocks lasting 7 to 9 seconds (duration jitter = 1000 ms).

Experimental design: fMRI Face-adaptation. In the present study we used a modified version of a fMRI adaptation paradigm validated and fully described in a recent study (Gentile and Rossion, 2014). The stimuli consisted of 18 different faces (males in the first and third run; female in the second run; see the original article for dataset information). Face stimuli were presented in blocks and were repeated with five variable stimulation rates: 4, 6, 6.6, 7.5 and 8.57 Hz (ranging from one face every 250ms to one face every 125ms). These rate were selected to cover a fast range of stimulation frequencies and compromise with the refresh rate constrain of the stimulation monitor (i.e. 60 Hz/frequency rate as integer) and scanning time constrains. In each block, the faces could be either identical (SF) or different (DF) from each other. Therefore, the complete experimental design consisted of a total of 10 conditions: 5 frequencies × same/different faces; two blocks for both the SF and DF condition were presented for each frequency in a run, which in total consisted of 20 blocks. A single block lasted for 27 s and was followed by a resting period of 9s in which a fixation cross was presented. Participants were instructed to attend to a black cross that was positioned at the level of the nose of each depicted face and to press a response key whenever it would turn red (between 2 and 3 times during a block and with random interval between each other). The entire testing session lasted approximately 35 minutes. For a schematic depiction of the experimental design see Supplementary Fig. 5.

BFRT and DFRT composite measure calculation. For the BFRT, individual raw total (i.e. on 54 items) scores of correct face recognition were computed for each individual across the three groups and converted to z-scores based on the mean and the standard deviation of the score distribution in the hearing group. For the DFRT, the number of correct hits (recognition

of previously seen faces) and false alarms (recognition of previously unseen faces) for each participant were used to compute the statistic *d-prime* as a measure of the sensitivity to known faces. After individual *d-prime* values were computed, they were also converted to corresponding *z*-scores based on the mean and the standard deviation of the score distribution in the hearing group. Finally, *z*-scores for the two tests were summed up to obtain the composite face recognition measure. Group-specific performance was analyzed using a one-way ANOVA with the composite face recognition measure as the dependent variable and the three groups as the between-subjects factor.

Beta Weights Extraction in right TVA/dTFA for face and house conditions. We first created two bilateral TVA masks by intersecting the (i) cluster of activation image generated by the conjunction analysis [Voice > Scrambled Voice ∩ Voice > Object Sound] at the group level and (ii) a sphere volume (15 mm radius = 14cm^3). The center of the sphere volume was defined by searching, within each left and right temporal cluster, the group peak-coordinates showing a geometrical distance lower than 5 mm from the peak-coordinates for the middle TVA reported in the STS/STG by Belin(10) and colleagues [62;-14;1 and -58;-18;-4]. This approach was chosen to ensure consistency in functional localization of voice-sensitive regions between studies and that inferences could be drawn within portions of the STS that functionally interact with FFA during speaker's voice recognition(13) and seems to be structurally connected with it(14).

Subsequently, we used the bilateral TVA ROIs as masks within which we searched the local activation maximum closest (sphere search = 10mm radius) to the peak of the group-maxima in the right and left mid-STG/STS (see Table S4) showing voice-selective response in hearing

controls during our independent voice-localizer experiment. The masks were used to avoid selecting peak-coordinates outside of our region of interest (i.e. mid-STG/STS) and potentially extending to the posterior STG/STS, which is known to also process face information in hearing individuals. The beta estimates were then extracted from the selected individual peak coordinate within a sphere volume of 5mm radius for both the face and house conditions of the face localizer separately and in each study participants.

Exploration of cross-modal regional response in left mid-TVA: Statistical inferences performed at 0.05 FWE voxel-corrected over a small spherical volume on the peak-coordinate for left mid-TVA [-60 -16 1] did not reveal cross-modal face selectivity in this region. For exploratory purposes we further extracted individual activity estimates from this region (see section above) and enter the individual measures in a repeated measure ANOVA with the two visual conditions as within-subject factor and the three groups as between-group factor, as well as in three within sample paired t-tests. These analyses revealed face selective responses only in the deaf group ($t = 6.206$, $p < 0.001$), which activated the left mid-STG/STS more for faces than for houses compared to both the hearing ($F = 51.96$; $p < 0.001$) and the hearing-LIS ($F = 33.62$, $p<0.001$) groups - as can be seen in supplemental Figure S4.

DCMs definition. In the right hemisphere, each region of interest was first defined as a sphere (5mm radius) centered individually on the local activation maximum closest to (i) the peak of the group-maxima in the regions showing face-selective response (i.e. FFA, pSTS and dTFA) and (ii) the peak of the group-maxima in the occipital region showing stronger functional connectivity to dTFA (i.e. V2/V3; for details on peak-coordinates see Supplementary Tab. 4). Then, correspondent time series were obtained by extracting the first

principal component from all raw voxel time series within each specific region, mean-corrected and high-pass filtered to remove low-frequency signal drifts. In all dynamic causal models (DCMs), inputs corresponded to the visual stimulation, regardless of the specific visual condition (i.e. face + house), and entered the system in V2/V3. In addition, in all DCMs visual information was allowed to flow within the dynamic system through 'all-to-all' endogenous connections running between all the four regions (e.g. between V2/V3-FFA, V2/V3-pSTS, V2/V3-dTFA, FFA-pSTS and so on). Instead, the three models differed on the specification of the modulatory term describing the effect driven by face information processing on endogenous connections. More specifically, face-selective responses in dTFA was hypothesized to be supported by: face-driven modulation of V2/V3 to dTFA connectivity in Model 1, face-driven modulation of FFA to dTFA connectivity in Model 2 and face-driven modulation of pSTS to dTFA connectivity in Model 3. See figure 4.B in the main text for a detailed depiction of the models.

Figure S1

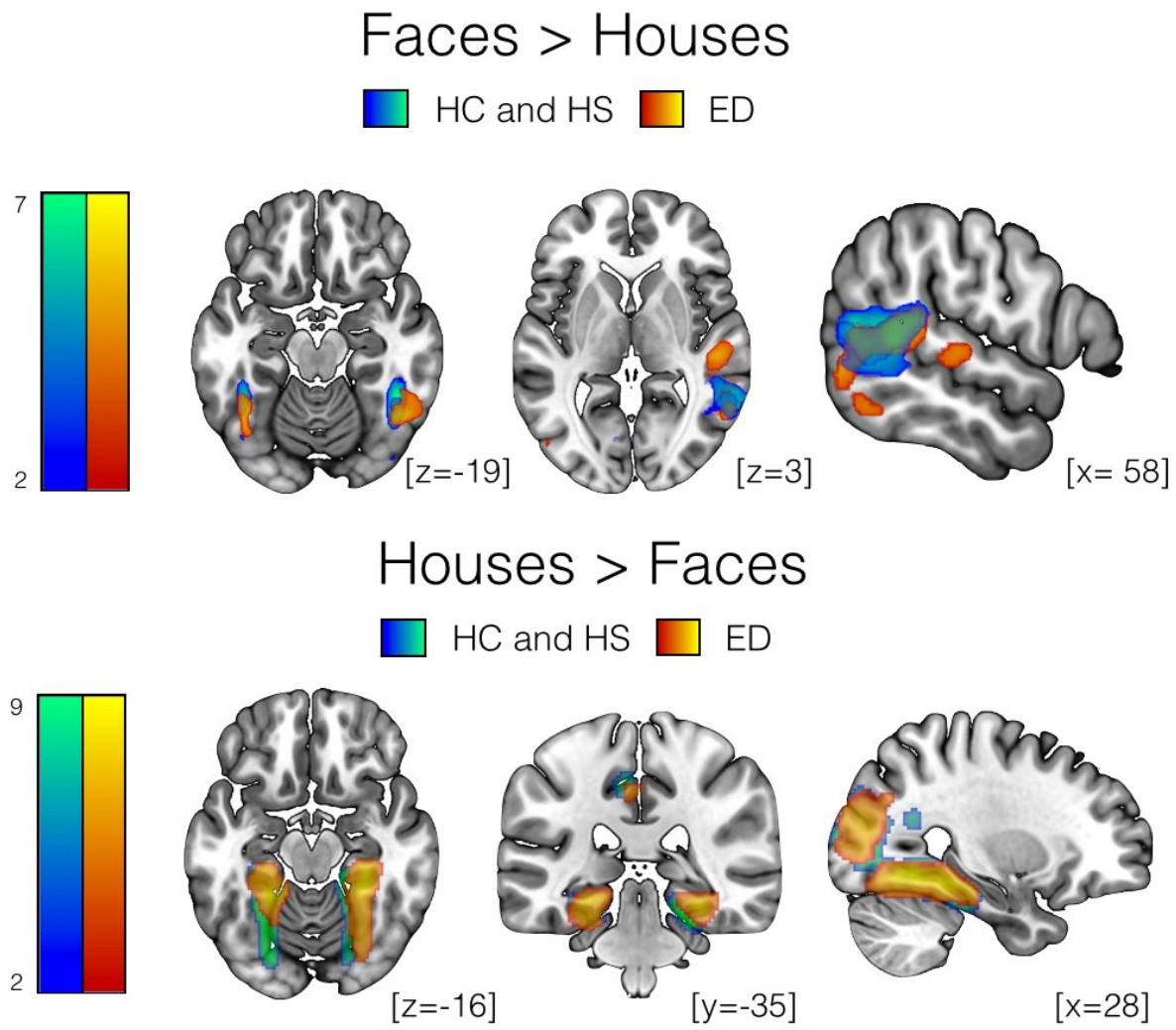

Figure S1 (Related to figure 1). Regional face- and house-selective responses in the three groups. Since no differences were observed between hearing and hearing-LIS individuals, the two groups are merged for visualization purposes. Supra-threshold ($P < 0.05$ FWE cluster-corrected; cluster size > 50) effects for hearing (blue/green) and deaf (red/yellow) individuals are superimposed on multiplanar slices of the MNI-ICBM152 template. Z-values are scaled accordingly to the color map.

Figure S2

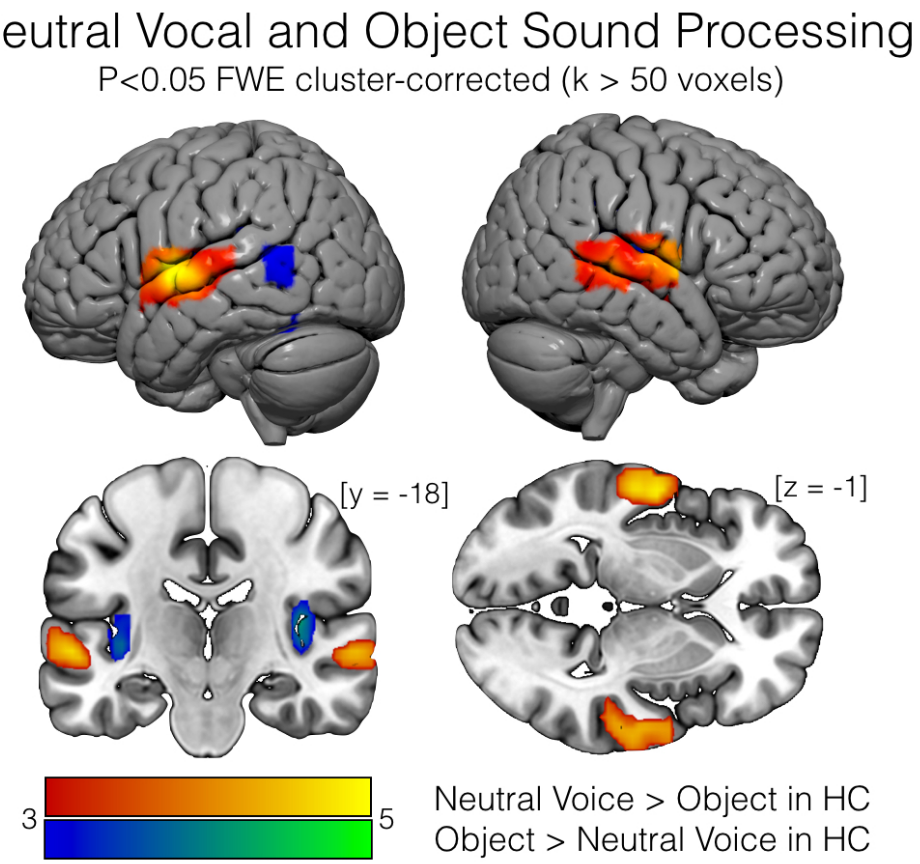

Figure S2 (Related to figure 1). Voice selective activations in the hearing group. Supra-threshold (*P* <0.05 FWE cluster-corrected, cluster size > 50) selective responses to neutral voices (red/yellow) and object sounds (blue/green) are shown in color scale (*z-values*) on a render (top panel) and axial/coronal slices of the MNI-ICBM152 template brain. The activations shown for Neutral Voice here refer to the conjunction contrast [(Neutral Voice > Scrambled Voice) ∩ (Neutral Voice > Object Sound)]; the activations shown for Object Sound here refer to the conjunction contrast [(Neutral Voice > Scrambled Voice) ∩ (Object Sound > Neutral Voice)]. Abbreviation: HC, Hearing Controls; FWE, Family-Wise Error; k, cluster size.

Figure S3

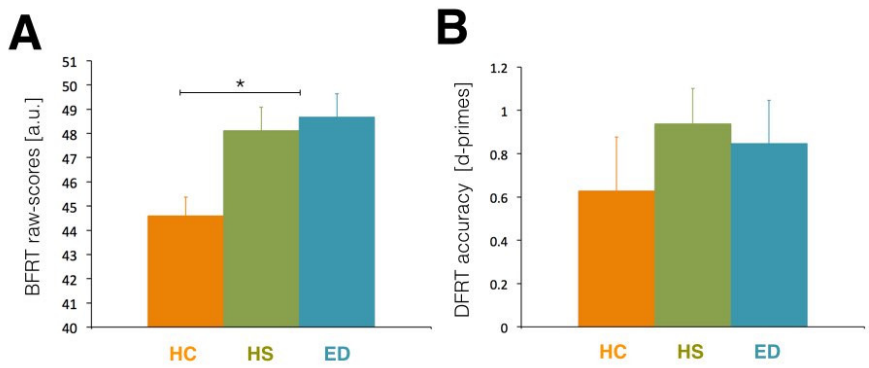

Figure S3 (Related to Figure 1). Face processing abilities. Behavioral performance on the Benton Face Recognition Test (BFRT) and Delayed Face Recognition Test (DFRT) separately. Bar graphs display: **(A)** the BFRT mean accuracies (*a.u.* ± SEM) and the significant difference between groups (**P* = 0.004) and **(B)** the DFRT mean accuracies (*d*-prime values . ± SEM), which do not differ between groups. Abbreviations: HC, Hearing Controls; HS, Hearing sign language users; ED, Early Deaf individuals

Figure S4

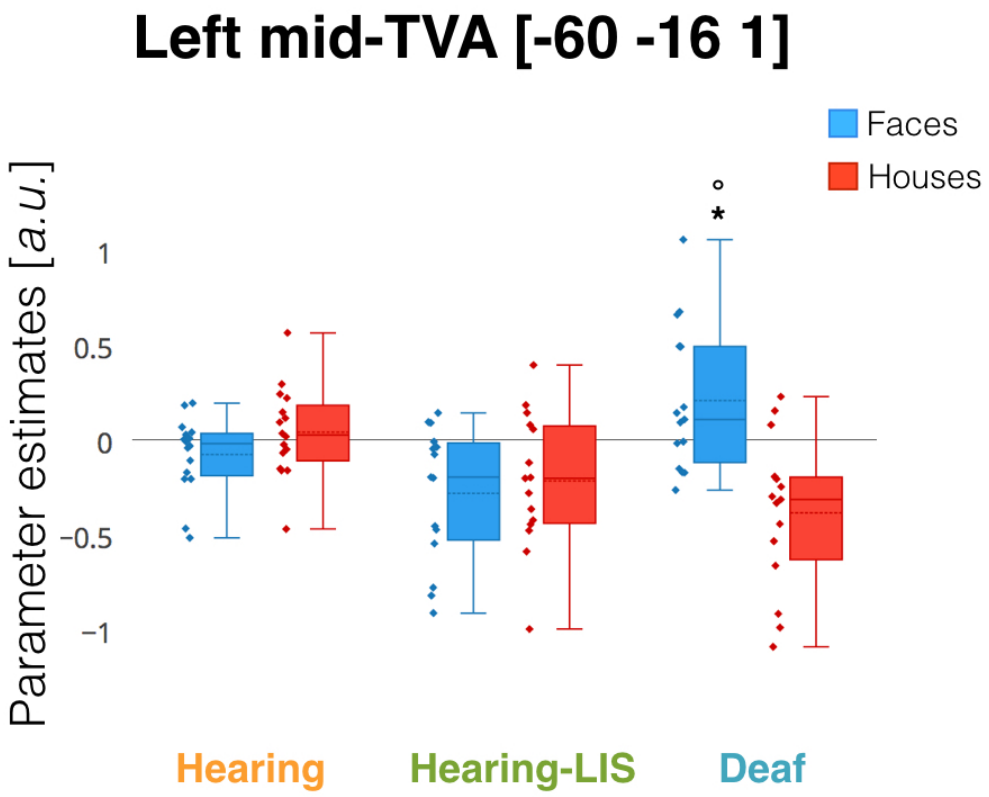

Figure S4 (Related to Figure 1). Face selectivity in the left mid-TVA in the deaf. Box-plots showing the central tendency (*a.u.;* median = solid line; mean = dashed line) of activity estimates for face (blue) and house (red) processing computed over individual parameters (diamonds) extracted at group-maxima for left-TVA in each group; * P<0.001 between groups; ° P<0.001 for Faces > Houses in deaf subjects.

Figure S5

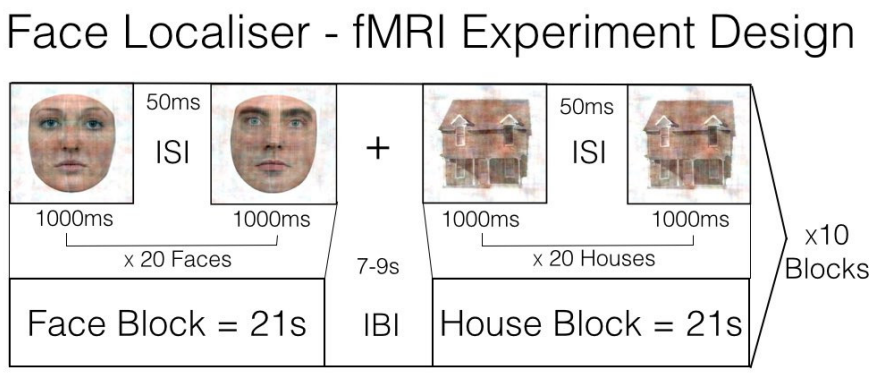

Figure S5 (Related to Figure 1). Face localizer paradigm. Schematic representation of the experimental design (one-back identity task) used for the fMRI Face Localizer acquisition. A run consisted of 20 blocks, 10 for condition (i.e. faces or houses); each block lasted for 21s and consisted of 20 stimuli; two stimuli were separated by a inter-stimulus-interval (ISI) of 50ms and two blocks by a resting inter-block interval (IBI) of 7 to 9s. Face and house stimuli were matched for low-level image properties and two stimuli were separated by an inter-stimulus interval of 50ms. Two exemplar blocks, one for each condition, are depicted.

Figure S6

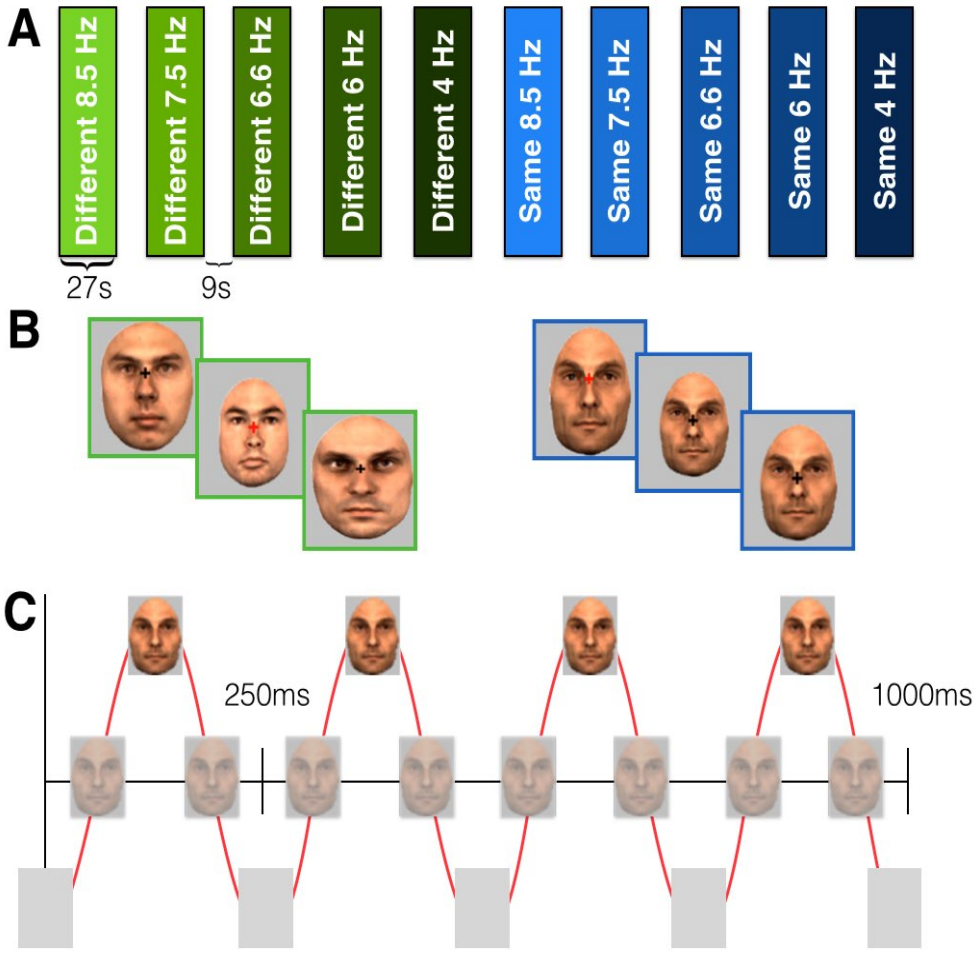

Figure S6 (Related to Figure 2). Face-adaptation paradigm. Schematic representation of the experimental design (one-back identity task) used for the fMRI Face-adaptation acquisition. (A) A run consisted of 20 blocks of trials and 10 different conditions (2 blocs for condition). Each block lasted 27s and two blocks were separated by a resting period (cross-fixation) of 9s. The order of block presentation was pseudo-randomized. (B) Example of stimuli presented in the different (left) and same (right) face condition. The size of the face image changed at every trial while a black cross was presented above the face nose; participants were asked to press the response button whenever the cross color

would turn to red. (C) An example of face-trial presentation within 1s: 4 cycles of the same face condition at 4Hz.

SUPPLEMENTAL TABLES

Table S1 (Related to table 2). Characteristics of the early deaf participants.

Code	Deafness Onset	Deafness Severity	Deafness Duration	Preferred Language	Hearing Aid Use	Experiment
ED1	Birth	Profound	25	LIS	no	fMRI
ED2	Birth	Profound	21	LIS	no	fMRI-MEG
ED3	Birth	Profound	45	LIS	Partial	fMRI-MEG
ED4	Age 0-4*	Profound	32	LIS/Italian	Full	fMRI-MEG
ED5	Birth	Severe/ Profound	39	Italian	Full	fMRI-MEG
ED6	Birth	Profound	31	Italian/LIS	Full	fMRI-MEG
ED7	Birth	Profound	34	LIS	No	fMRI-MEG
ED8	Birth	Profound	41	LIS/Italian	Partial	fMRI-MEG
ED9	Birth	Profound	31	LIS	No	fMRI
ED10	Birth	Severe	24	Italian/LIS	Full	fMRI
ED11	Birth	Profound	33	Italian	Full	fMRI-MEG
ED12	Birth	Severe	25	Italian	Full	fMRI-MEG
ED13	Birth	Profound	24	LIS/Italian	Full	fMRI-MEG
ED14	Birth	Profound	39	LIS	Full	fMRI-MEG
ED15	Birth	Profound	36	LIS/Italian	No	fMRI-MEG
ED16	Birth	Profound	49	Italian/LIS	Full	MEG
ED17	Birth	Profound	37	Italian/LIS	No	MEG
ED18	Birth	Profound	53	LIS/Italian	No	MEG
ED19	Birth	Severe	38	LIS/Italian	Full	MEG
ED20	Birth	Severe	26	Italian/LIS	Full	MEG

*Hearing Aid use: Partial = only during school or work hours; Full = on most of the day to support environmental sound detection (alarms, door bells, foot steps). Only ED11 and ED12 reported support during speech reading. Abbreviations: LIS, Italian Sign Language; ED, Early Deaf. *ED4 reported measles before age 4.*

Table S2 (Related to table 2). Italian Sign Language in the early deaf and hearing participants

Code	LIS Acquisition Age (Years)	LIS Exposure Duration (Years)	LIS Use Frequency (% Year-time)
	Early Deaf Participants		
ED1	0.5	25	100
ED2	19	2	100
ED3	16	29	100
ED4	23	13	45
ED5	18	21	3
ED6	21	10	14
ED7	11	23	100
ED8	2	39	100
ED9	16	15	100
ED10	0.5	24	100
ED11	--	0	0
ED12	--	0	0
ED13	0.5	24	100
ED14	19	20	100
ED15	2	34	100
ED16*	6	43	14
ED17*	20	18	45
ED18*	6	47	100
ED19*	0.5	38	100
ED20*	10	16	14
	Hearing Sign Language Users		
HS1	22	18	45
HS2	0.5	36	100
HS3	0.5	29	100
HS4	0.5	41	100
HS5	25	5	45
HS6	0.5	31	45
HS7	19	5	45
HS8	0.5	36	100
HS9	27	22	100
HS10	0.5	26	100
HS11	0.5	46	100
HS12	19	36	100
HS13	0.5	33	45
HS14	16	20	100
HS15	0.5	39	100

(*) Participated in the MEG experiment only; ED, Early Deaf; HS, Hearing LIS-users.

Table S3. fMRI Acquisition Parameters

Experiment	Volumes	Slices	TR	TE	Flip Angle	Matrix Size	Slice Gap	Slice Thickness
Voice Localizer	335	37	2200ms	33ms	76°	64x64	0.6mm	3mm
Face Localizer	274	37	2200ms	33ms	76°	64x64	0.6mm	3mm
Face Adaptation	329	38	2250ms	33ms	76°	64x64	0.4mm	3mm

TR = Repetition Time; TE= Echo Time

Table S4 (Related to figure 4). Group-specific peak-coordinates used for extraction of activity estimates (beta weights/time-series) and regions of interest definition.

Area	$X_{(mm)}$	$Y_{(mm)}$	$Z_{(mm)}$
fMRI Face Localizer: Beta Weights Extraction			
Right TVA in each group	63	-22	4
Left TVA in each group	-60	-16	1
fMRI Face-adaptation: Beta Weights Extraction			
Right dTFA in ED	62	-18	2
Right TVA in HC and HS	63	-22	4
Right FFA in ED	48	-56	-18
Right FFA in HC	44	-50	-16
Right FFA in HS	44	-52	-18
PPI on Face Localizer: Seed Region Definition			
Right dTFA in ED	62	-18	2
Right TVA in HC and HS	63	-22	4
DCM on Face Localizer: ROIs Definition			
Right dTFA in ED	62	-18	2
Right TVA in HC and HS	63	-22	4
Right FFA in ED	48	-56	-18
Right FFA in HC	44	-50	-16
Right FFA in HS	44	-52	-18
Right pSTS in ED	50	-44	14
Right pSTS in HC	52	-42	-16
Right pSTS in HS	52	-44	10
Right V2/V3 in ED	26	-94	4
Right V2/V3 in HC	28	-86	4
Right V2/V3 in ED	27	-92	-1

Search radius = 10mm; ROI radius= 5mm; Abbreviations: HC, Hearing Controls; HS, Hearing LIS-users; ED, Early Deaf; TVA, Temporal Voice Area; TFA, Temporal Face Area; FFA, Fusiform Face Area; pSTS, posterior Superior Temporal Sulcus.

Table S5. Increased functional connectivity from the right dTFA/TVA for the main effect of face condition in each group and differences between the three groups

Area	Cluster size	$X_{(mm)}$	$Y_{(mm)}$	$Z_{(mm)}$	*Z*	*D.F.*	P_{FWE}
HC Faces > Houses						15	
No significant effects							
HS Faces > Houses						14	
No significant effects							
ED Faces > Houses						14	
R lateral occipital cortex	2824	42	-86	8	7.16		< 0.001
R inferior occipital cortex	s.c.	42	-70	-8	5.26		0.004
R fusiform gyrus	s.c.	34	-48	-16	4.05		< 0.001*
L lateral occipital cortex	3596	-22	-90	-4	5.77		< 0.001
L inferior temporal gyrus		-34	-60	-6	4.47		< 0.001*
ED > HC ∩ HS - Faces > Houses						3,43	
R lateral occipital cortex	1561	42	-86	8	5.91		< 0.001
L lateral occipital cortex	1044	-20	-92	-4	4.83		0.027
ED > HC - Faces > Houses						30	
R lateral occipital cortex	2071	40	-88	8	5.93		< 0.001
R middle occipital gyrus	s.c.	32	-96	10	5.84		< 0.001
L lateral occipital cortex	2794	-24	-90	-4	5.37		0.002
ED > HS - Faces > Houses						29	
R lateral occipital cortex	2015	42	-86	8	5.91		< 0.001
R middle occipital gyrus	s.c.	36	-90	0	5.30		0.003
L lateral occipital cortex	1150	-20	-92	-4	4.83		0.027

Significance corrections are reported at the voxel level; cluster size threshold = 50; (*) brain activations significant after FWE cluster-correction over the whole brain. Abbreviations: HC, Hearing Controls; HS, Hearing LIS-users; ED, Early Deaf; D.F. = degrees of freedom; FWE, Family-Wise Error; s.c., same cluster.