Supporting Information

Bouton et al. 10.1073/pnas.1714279115

SI Results

fMRI Study. In the first fMRI study subjects performed a phonemic categorization task on a /ba/ /da/ continuum in which the onset value of the second formant (F2) and the F2 slope linearly covaried in six steps. Given the tonotopic organization of auditory cortical areas (1, 2), we expect F2 parameters to be spatially encoded, i.e., the larger the frequency span in the stimulus, the stronger is the BOLD response. Note that the choice of stimuli places us in a rather unfavorable situation with respect to our hypothesis that phonemes might be efficiently encoded on the basis of local spectro-temporal cues. Steps 1 and 6 corresponded to clear ba and da prototypes, whereas steps 2–5 were ambiguous stimuli identified as either ba or da (Fig. 1*A*). Subjects' behavior confirmed that ambiguous stimuli were less reliably assigned to a given syllable than prototype stimuli. As shown in Fig. 1*B*, *Middle*, d' values parametrically varied with category ambiguity, that is, with the distance to category boundary. Thus, the d' parameter can be taken as a graded index of the perceptual decision process, i.e., the effort associated with recognizing and classifying a stimulus among competitors (3). Subjects' behavior also confirmed that F2 and d' were naturally decorrelated ($r = 0.006$, $P = 0.59$). Using a generalized linear model, we separately probed brain regions where neural activity increased with F2 parameters and with decisional effort, respectively. Only the right pSTG including Heschl's gyrus was activated by F2 parameters (Fig. 1*C*). Decisional aspects of the task ($-d'$) implicated the left IFG, the left pSTG, and the right anterior temporal-ventral prefrontal region. In the IFG, the neural-response profile associated with decisional-related activity was an inverted U-shaped curve, reflecting increased difficulty in the perceptual categorization of ambiguous syllables and reduced responses for prototypical stimuli (4).

To address whether there were preferential representations for /ba/ and /da/, we probed neural activity that correlated more with one category than with the other independent of the acoustic structure of the stimuli. We compared trials in which the same stimulus gave rise to a different percept within each relevant level of the ba–da continuum, i.e., stimuli 2, 3, and 4 (Fig. S3). The comparison /ba/ minus /da/ within each level of the continuum revealed a significant cluster in the right frontal gyrus (x, y, z = 45, 35, 1, $T = 2.8$) and in the left supramarginal gyrus extending to the postcentral gyrus (x, y, z = −63, −16, 22, $T = 3.12$). The reverse contrast, /da/ minus /ba/, elicited a neighboring cluster in the supramarginal gyrus but at a slightly more dorsal location (x, y, z = −51, −37, 43, $T = 3.43$). The supramarginal gyrus is known to be involved in phonological and articulatory processing of words (5–7). These data confirm (8–10) that the articulatory network is mobilized to help distinguish phonemes when stimuli are ambiguous and categorization is challenging, as is the case for levels 2, 3, and 4 of the ba–da continuum.

We then carried out one last analysis to better characterize the ability to decide between categories irrespective of the specific phonemic category being presented. To do so, we compared brain activity for correct and incorrect responses within each level of the continuum (stimuli 1–6) and then calculated an averaged map across all levels (Fig. S4). Correct and incorrect responses, and hence subjective boundary separations, were defined according to the subject's ratings along the continuum. This analysis, in essence, corresponds to the correct-minus-incorrect analysis that we performed in i-EEG (see main text and Fig. 4). Here, comparing brain activity for correct and incorrect responses (correct minus incorrect) revealed a network of significant clusters in the left middle temporal cortex (peak MNI coordinates, x, y, z = −60, −22, −17, $T = 4.56$), in the angular gyrus (x, y, z = −48, −70, 37, $T = 7.17$), and in the inferior frontal area (x, y, z = −48, 32, −8, $T = 4.28$). The reverse contrast (incorrect minus correct) also revealed a strongly lateralized response but in the opposite hemisphere, with significant clusters in the pSTG (x, y, z = 57, −25, 46, $T = 8.49$) and in a network of regions known to be part of the executive control system, such as the anterior cingulate cortex (x, y, z = 30, −64, 49, $T = 7.24$), the precuneus (x, y, z = 12, −52, 67, $T = 6.28$), and the dorsolateral prefrontal cortex (x, y, z = 33, 35, 28, $T = 4.55$). Interestingly, these results suggest that language-related areas were involved when phoneme categorization was easy and effortless (correct identification), whereas the right pSTG and regions involved in executive control and conflict monitoring were recruited only when categorization was difficult.

To establish the spatial and functional specificity of F2-parameter tracking, we repeated the whole experiment using a /da/ /ta/ continuum in which the morphed acoustic parameter, the VOT, was a temporal rather than a spectral cue (Fig. S2). Critically, the consonant part is longer for /ta/ than for /da/, and therefore we expected a stronger BOLD response for the former. We thus predicted that increased activity in left auditory cortical areas would reflect the increased integration time needed to categorize. Accordingly, VOT-related activity was found in the left temporal lobe. Decision on VOT was easier than on F2 (*t* test, $P = 0.002$), and consequently the decisional effort was less marked in the left IFG. No correlate of VOT tracking and related decisional effort was present in the right hemisphere (Fig. S2*C*).

Independence of Decoding Time Course and Correctness Time Course. Note that the coincidence of decoding and correctness effects is genuine because they originate in independent data series. Decoding was computed from event-related potentials from all syllables (prototypes or ambiguous) (*Methods*), whether correctly or incorrectly identified, whereas correctness effects result from a contrast between correctly and incorrectly recognized prototype syllables.

Neural Activity Related to Perceptual Decision. When we broadened the i-EEG analyses to temporal and frontal locations outside the auditory cortex, two additional regions showing decision-related activity. Consistent with both fMRI and MEG findings, these regions were the right inferior prefrontal cortex and the temporal pole region. Correlates of correct perceptual decision were also found at this exact location (shaft 4, contact 4 and shaft 6, contact 1), indicating that this is another important site for categorical syllable perception. That the electrodes posterior to these (shaft 4, contact 3 and shaft 6, contact 2) also show categorical effect based on F2 confirms previous observations that spectro-temporal phonemic information is recoded in posterior temporal regions. This region presumably corresponds to an intermediate, noncritical processing level between the sensory and decision levels from which choice correlated activity is likely to emerge (11).

Poststroke Patient. Finally, since classifying was possible from both local and distributed activity, we addressed whether focal readout of the pSTG was necessary for human perceptual decisions or whether perceptual decisions could also be made from subthreshold noise-level activity distributed across the temporal lobe. Within a database of 310 poststroke patients we selected one right-handed patient who had a focal lesion of the right pSTG confined to those regions that fMRI and MEG indicated as primarily involved in F2 tracking and perceptual decision (Fig. S10*A*). The patient scored normally on all clinical psychological tests involving language, memory, attention, executive function, and spatial orientation. While she was able to perform the syllable categorization task (Fig. S10*C*), she could not discriminate between ba and da and identified all syllables as da (Fig. S10*B*). Her d' profile indicates that she had greater difficulty on the ba end. Critically, she performed well on the control da–ta task in which the VOT rather than F2 was varied. This behavioral dissociation effect confirms the double dissociation observed with fMRI (Fig. S2) and indicates that the right pSTG is necessary for phonemic identification based on spectral cues. It would be interesting to know whether machine-learning (as opposed to neuronal) decoding could still correctly categorize the phonemes from this subject's neural activity. However, to answer this question, we will have to wait until invasive recordings from a patient with a similarly selective lesion become available.

1. Nourski KV (2017) Auditory processing in the human cortex: An intracranial electrophysiology perspective. *Laryngoscope Investig Otolaryngol* 2:147–156.
2. Riecke L, et al. (2016) Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. *Cereb Cortex* 27:3002–3014.
3. Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: Neural and computational mechanisms. *Nat Rev Neurosci* 15:745–756.
4. Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7: 295–301.
5. Hickok G, Poeppel D (2015) Neural basis of speech perception. *Handb Clin Neurol* 129: 149–160.
6. Oberhuber M, et al. (2016) Four functionally distinct regions in the left supramarginal gyrus support word processing. *Cereb Cortex* 26:4212–4226.
7. Raizada RDS, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. *Neuron* 56:726–740.
8. Klein M, et al. (2015) Early activity in Broca's area during reading reflects fast access to articulatory codes from print. *Cereb Cortex* 25:1715–1723.
9. Lyu B, Ge J, Niu Z, Tan LH, Gao J-H (2016) Predictive brain mechanisms in sound-to-meaning mapping during speech processing. *J Neurosci* 36:10813–10822.
10. Yoo S, Lee K-M (2013) Articulation-based sound perception in verbal repetition: A functional NIRS study. *Front Hum Neurosci* 7:540.
11. Engel TA, Chaisangmongkon W, Freedman DJ, Wang X-J (2015) Choice-correlated activity fluctuations underlie learning of neuronal category representation. *Nat Commun* 6:6454.

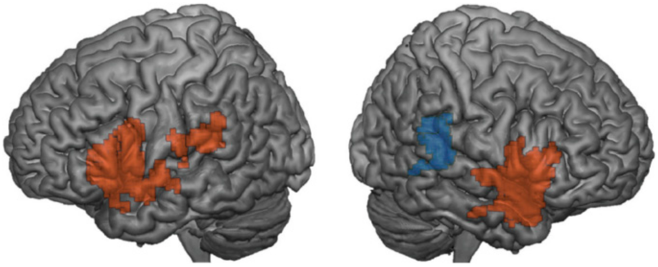

Fig. S1. The fMRI ba/da experiment. Bilateral inferior prefrontal cortex (red), left (red) and right (blue) pSTG displayed at a $P = 0.05$ uncorrected threshold, k = 20 (any-depth search mode).

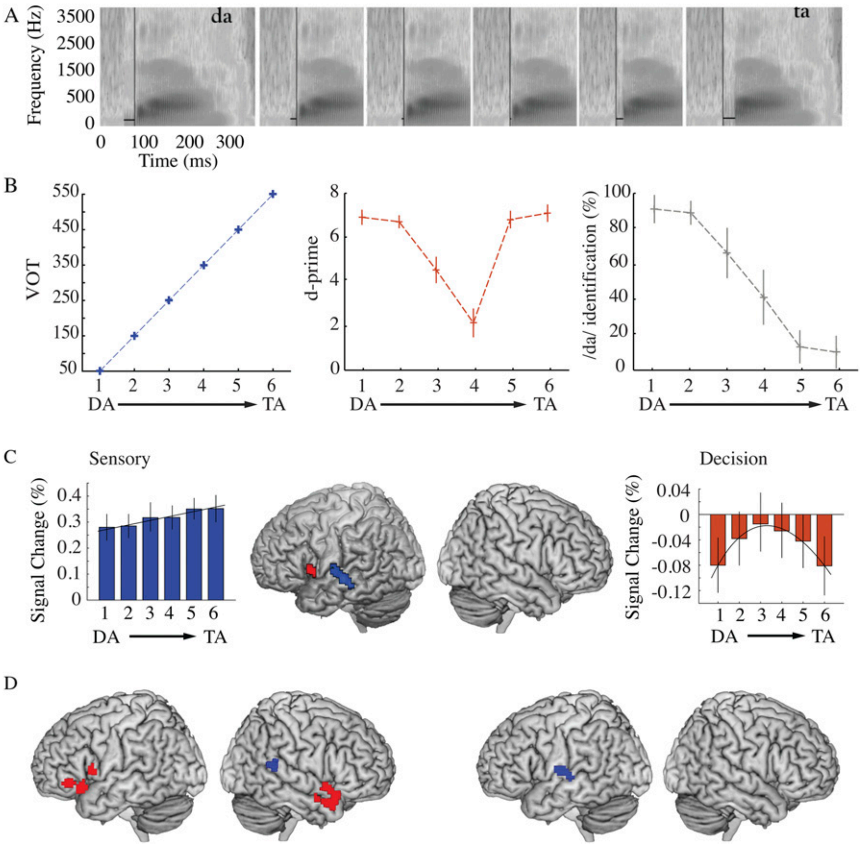

Fig. S2. Control da/ta fMRI experiment. (*A*) Spectrograms of the stimulus continuum between syllables /da/ and /ta/ synthesized with linear variation in the VOT (in 5-ms steps from −15 to +15 ms). The full spectrograms at the extremes of the continuum represent /da/ (*Left*) and /ta/ (*Right*) prototype syllables. The spectrograms in the middle are centered on the VOT. (*B*) Values for the VOT (blue; *Left*), average d' (red; *Center*), and percent of syllables identified as /ta/ (gray; *Right*). Data are shown as mean ± SEM. (*C*, *Middle*) Spatial localization of VOT neural encoding (blue) and d' (red) in fMRI BOLD signal, expressed as beta coefficients. Significant clusters were found in the left STG (peak MNI coordinates, x, y, z = −66, −22, 1, T = 4.29) for VOT tracking and in the left inferior prefrontal cortices (x, y, z = −48, 14, 7, T = 3.75) for perceptual decision (d'). Images are presented at a whole-brain threshold of $P < 0.001$. (*C*, *Right*) Percent signal change in the left inferior prefrontal cortex (red bars). (*Left*) Percent signal change in the in the left STG (blue bars). The BOLD signal increases with the VOT in the left STG and with the perceptual decision load in the left inferior prefrontal region (red). (*D*) Spatial localization of acoustic encoding (blue) and d' (red) in fMRI BOLD signal when comparing ba/da vs. da/ta (*Left*) or da/ta vs. ba/da (*Right*) conditions. (*Left*) The ba/da vs. da/ta contrast elicited significant clusters in the right pSTG (peak MNI coordinates, x, y, z = 36, −43, 10, T = 3.81) for the F2 tracking and in the left inferior prefrontal cortex (peak MNI coordinates, x, y, z = −39, 32, −5 extending to −45, 20, −8, T = 3.66) and the right temporal cortex extending to the right prefrontal cortex (peak MNI coordinates, x, y, z = 36, −43, 10 extending to 45, 11, −11, T = 4.28) for perceptual decision. The da/ta vs. ba/da contrast elicited significant clusters in the left superior temporal gyrus (peak MNI coordinates, x, y, z = −66, −22, 1, T = 3.82) for the VOT tracking only. All reported clusters survived a cluster-level uncorrected $P < 0.05$ combined with a voxel-level uncorrected $P < 0.001$.

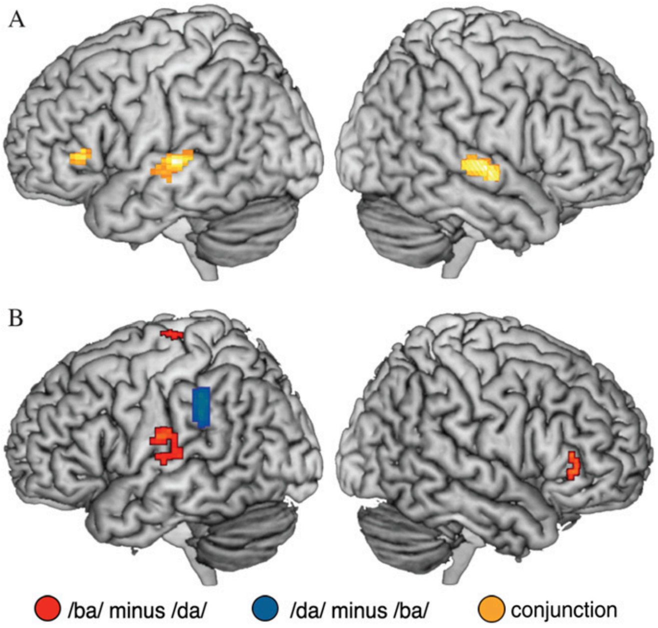

Fig. S3. Phonemic representations in the fMRI experiment. The analysis was performed on trials in which the same stimulus gave rise to a different percept within each relevant level of the ba–da continuum (i.e., stimuli 2, 3, and 4). (*A*) Conjunction results: Activations that are shared across /ba/ and /da/ responses are shown in yellow. (*B*) Contrast results: Activations that are specific to /ba/ are shown in red, and activations that are specific to /da/ are shown in blue. Images are presented at a whole-brain threshold of $P < 0.005$.

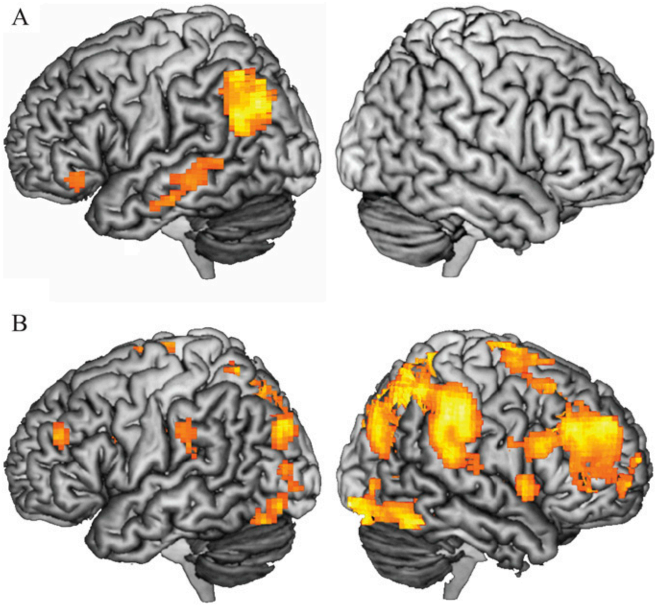

Fig. S4. fMRI experiment: correct versus incorrect perceptual decision. The analysis was performed on all levels of the continuum (from stimuli 1–6). (*A*) The correct-minus-incorrect contrast reveals a network of language-related regions. (*B*) The incorrect-minus-correct contrast reveals a network of regions for executive control. Images are presented at a whole-brain threshold of $P < 0.001$.

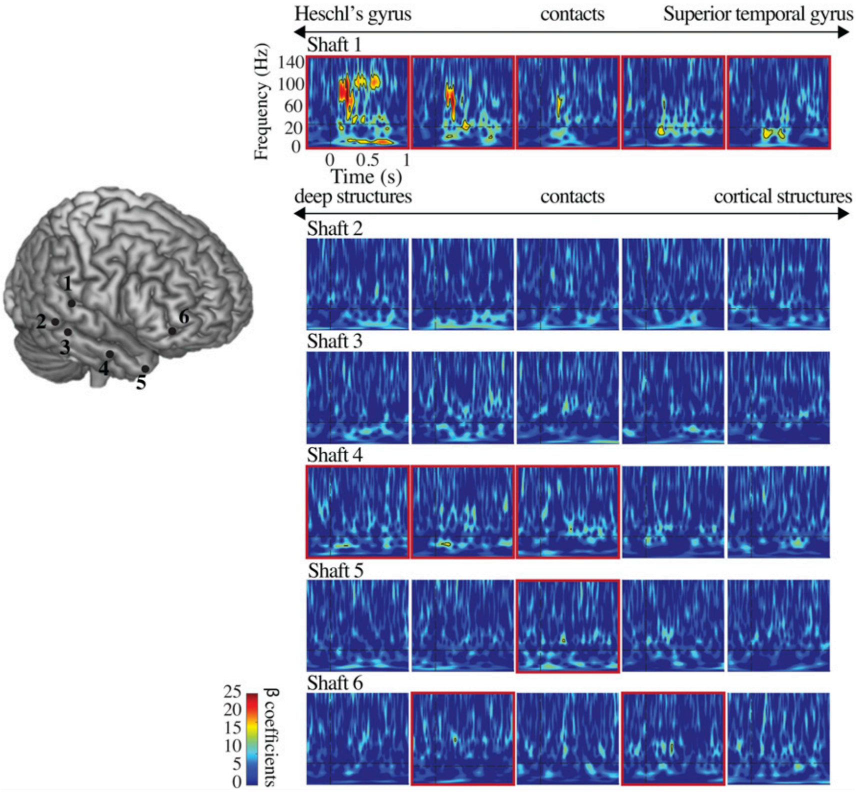

Fig. S5. i-EEG results in patient 1. Time–frequency representations of beta coefficients from the regression of F2 values against evoked activity on each contact of each shaft. Red frames indicate shaft contacts where significant F2 tracking was found (comparison with zero-mean normal distribution thresholded at $q < 0.05$).

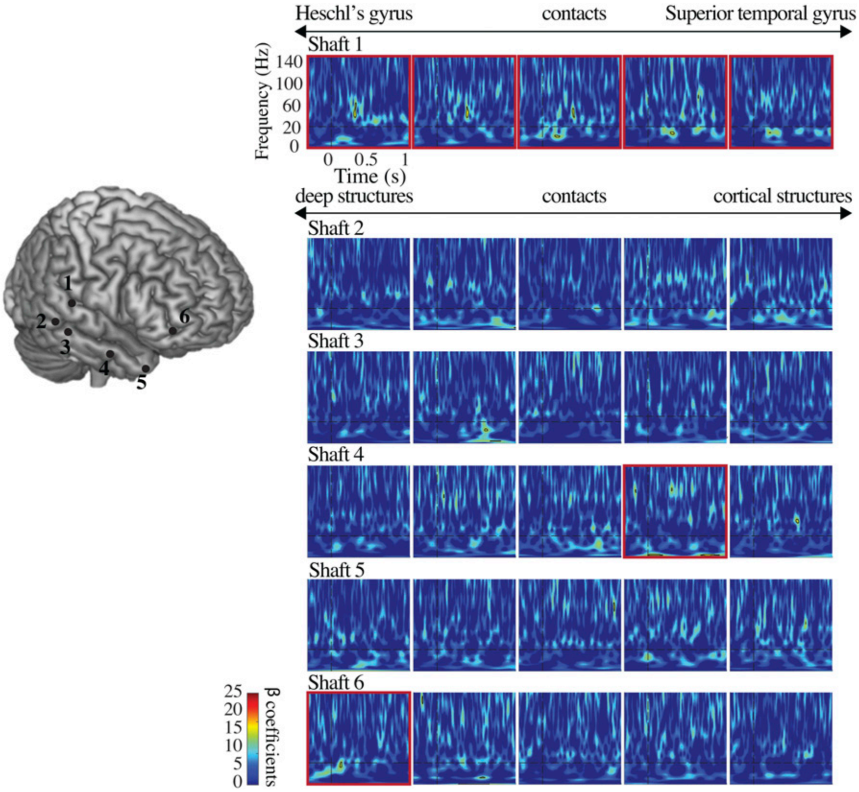

Fig. S6. i-EEG results in patient 1. Time–frequency representations of beta coefficients from the regression of d' values against evoked activity on each contact of each shaft. Red frames indicate contacts where decisional effects were significant (comparison with zero-mean normal distribution, thresholded at $q < 0.05$).

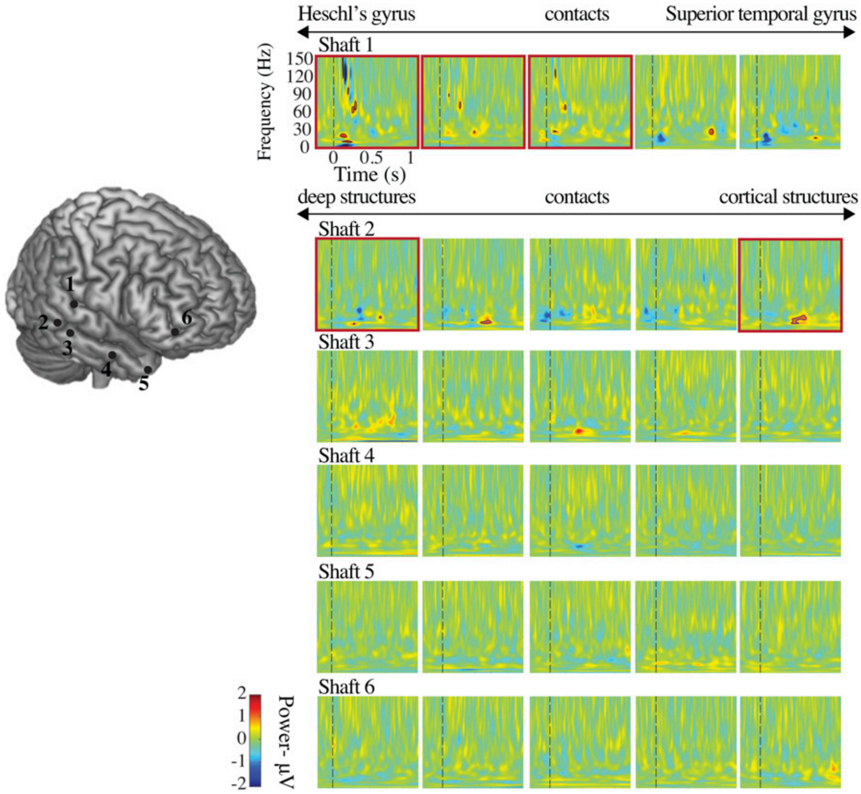

Fig. S7. i-EEG results in patient 1. Difference in the time–frequency response between correctly and incorrectly recognized syllables for each contact of each shaft. Red frames indicate contacts where significant correctness effects were found (comparison with zero-mean normal distribution, thresholded at $q < 0.05$).

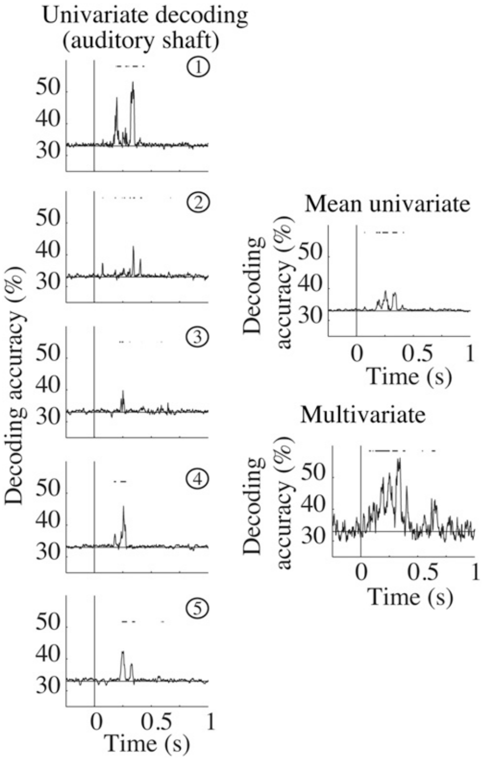

Fig. S8. Decoding on the auditory shaft (shaft 1) in patient 1. Significant univariate decoding was strongest on the first contact but was even stronger using the multivariate approach.

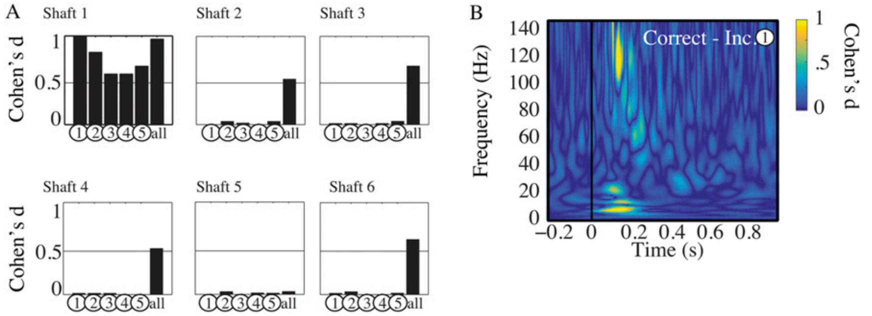

Fig. S9. Effect size (Cohen's *d*). (*A*) Effect size (Cohen's *d*) of machine decoding. For univariate decoding, contacts 1 and 2 of shaft 1 showed a large effect size (Cohen's $d > 0.8$), whereas contacts 3, 4, and 5 of shaft 1 showed a medium effect size ($0.5 <$ Cohen's $d < 0.8$). A small effect size (Cohen's $d < 0.2$) was found for all contacts of shafts 2, 3, 4, 5, and 6. For multivariate decoding, shaft 1 showed a large effect size, whereas shafts 2, 3, 4, and 6, showed a medium effect size. A small effect size was found for shaft 5. (*B*) Effect size (Cohen's *d*) of correct-minus-incorrect classifications. A large effect size (Cohen's $d > 0.8$) was found at 100–160 ms after stimulus onset in the alpha band (10 Hz) and from 130–150 ms poststimulus onset in the high-gamma band (100–140 Hz). A medium effect size (Cohen's $d = 0.55$) was found at 200–220 ms after stimulus onset between 60 and 80 Hz, which corresponds to significant cross-correlation between decoding accuracy and correct-minus-incorrect differences in the time–frequency domain (Fig. 4*B*).

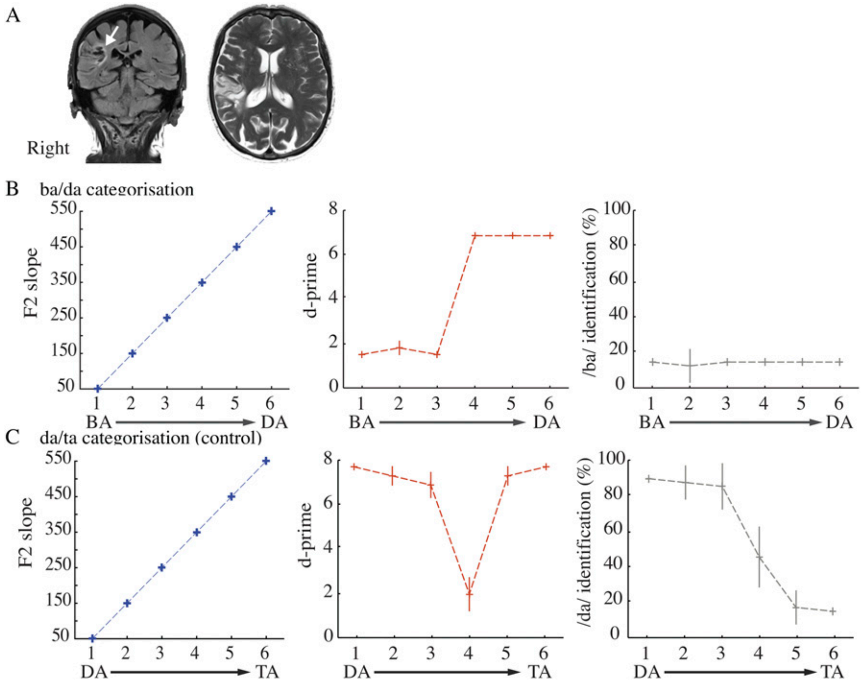

Fig. S10. Lesion data. (*A*) MRI scan of a patient with a focal lesion of the right pSTG. (*B* and *C*) Psychometric functions for ba–da (*B*) and da–ta (*C*) continua. The blue line represents the linear variations of F2 (*B*) and linear variations of the VOT (*C*), across stimuli. Gray curves show that the patient could not discriminate between /ba/ and /da/ and identified all syllables in the ba–da continuum as /da/ while performing normally on the da–ta continuum. Red curves indicate greater difficulty in performing the identification on the /ba/ end (*B*) but normal performance on the control da–ta task (*C*).