Supplementary Material for:
Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences

Revised in response to second set of Cerebral Cortex reviews

Rajeev D. S. Raizada,[1] Feng Ming Tsao,[2] Huei-Mei Liu,[3] and Patricia K. Kuhl[4]

[1] Neukom Institute for Computational Science, HB 6255, Dartmouth College, Hanover NH 03755.

[2] Dept. of Psychology, National Taiwan University, Taipei, Taiwan

[3] Dept. of Special Education, National Taiwan Normal University, Taipei, Taiwan

[4] Inst. for Learning & Brain Sciences, Univ. of Washington, Box 357988, Seattle WA 98195

S.1 Can standard fMRI reveal significant differences between the activations elicited by the phonemes?

A recent study by Obleser et al. (2006) put forward the interesting proposal that different phonemes may activate close-by but distinct regions of cortex, in sort of phonological topographic map. Specifically, the thresholded peak activation for back vowels (/u/ and /o/) and front vowels (/i/ and /e/) were found to be a few millimeters apart from each other.

One difficulty with this finding is that the fact that the locations of thresholded activation peaks are different does not necessarily imply that the underlying activation distributions are significantly different from each other. We carried out an analysis of this kind on our own fMRI data, using standard univariate fMRI analysis with smoothed BOLD data. Supplementary Figure S18 shows the activation profile along a line running in the y-direction through the center of the BOLD activation in left auditory cortex (x-coord = 60, y-coord range = -52.5 to 18.75, z-coord = -5.5). It can be seen the the activations for /ra/ (F3-low) and /la/ (F3-high) have different peaks and intensities. However, as Supp.Figs S18C and S18F illustrate, the differences between these two activation distributions are not significant. Thus, in our data at least, standard fMRI does not reveal a difference between the two phonemes.

S.2 On the interpretation of spatially distributed activation

One difficulty faced by pattern-based analyses is that their results are less straightforward to interpret than are those of standard fMRI analyses. Because the patterns of fMRI activation are spatially distributed over many voxels, one is no longer able to specify where exactly the useful information resides within a given cluster. If the information is truly distributed over multiple voxels, then there is no individual voxel where the information resides, there is only the region across which it is spread. One can state where the center of this region is, and that is what we do, but that does not alter the fact that the information is distributed over the region itself, rather than being located at the center, or at any other single voxel within the sphere-of-information. By analogy, although there is a single GPS coordinate uniquely specifying the location of the front door of the Disney Concert Hall in Los Angeles, there is no unique coordinate specifying the location of LA itself. Los Angeles is distributed over many square miles. An atlas may give a unique set of coordinates, perhaps corresponding to a point at the "center of gravity" of LA, but that does not make the metropolis any less distributed. Just as there is no single spot "where exactly the city of Los Angeles resides within the LA metropolitan region," there is no single voxel "where exactly the useful information resides within the cluster".

S.3 Discussion of using a 1.5T versus a 3T scanner

One aspect of the present study which is less than optimal is the fact that the scans were performed on a 1.5T scanner rather than a more powerful 3T machine. A possible concern might be that the unusual results that we observed (information-bearing activation in right auditory cortex, rather than the left) could perhaps have arisen artifactually due to using a lower field strength. In this scenario, our results would be false positive effects seeming to show activation where in fact there is none (Type I error).

However, several studies comparing fMRI results obtained from 1.5T vs. 3T scanners have concluded that the principal drawback of 1.5T machines is not the occurrence of false positives, but instead is that they may fail to see real effects. In statistical terms, 1.5T may lead to a lack of power, generating more false negatives (Type II error). This is precisely the opposite of the possible concern described above.

The recent study by Tieleman et al. (2007) is especially relevant to this point: measuring language laterality at 1.5T and 3T, they found that "The number of activated voxels and mean t-values were significantly higher at 3T for all paradigms. Using the same statistical threshold, language activation was significantly less lateralized, and more activation zones were depicted at 3T compared with 1.5T."

A similar result was found by Hoenig et al. (2005) in a motor task. They wrote: "Moreover, additional functional activation was detected in medial (supplementary motor area) and dorsal premotor regions ($P < .05$, corrected) at 3.0-T functional MR imaging, which was not detectable with corresponding 1.5-T imaging."

It should also be borne in mind that our results are group-level random effects analyses, which reflect the degree of consistency across subjects of the signal, rather than the magnitude of the signal. Studies of statistical power in multi-subject fMRI experiments have shown that the most effective way to increase power is by increasing the number of subjects (Desmond and Glover 2002; Thirion et al. 2007). Specifically, Scouten et al. (2006) drew the following conclusion: "Given that inter-subject noise dominates across a range of tasks, improvements in within-subject noise, through changes in acquisition strategy or even moving to higher field strength, may do little to improve group statistics." Although our scanner was only 1.5T, our number of subjects was reasonably sized by fMRI standard (n = 20).

Given the above, we consider it unlikely that our results are false-positive artifacts of using a 1.5T machine.

S.4 Subject gender balance

As is stated in the Methods section of the main text, twenty subjects participated in the experiment: 10 were native American English speakers, 10 were native Japanese speakers. All of the English speakers were female, and six out of the ten Japanese speakers were female.

Certainly our subject pool was skewed towards having more female than male participants, and it would be desirable for future follow-up studies to sample more evenly across gender.

Although there have been some reports of gender differences in language lateralisation, a recent meta-analysis suggests that the evidence for this effect is weak (Sommer et al. 2004). Another recent study argues that whether gender effects are observed or not may depend upon the specific details of task-design, as opposed to being a robustly reproducible phenomenon (Harrington and Farias 2008). Thus, while we acknowledge that a more gender-balanced subject pool would have been desirable, we consider it unlikely that this significantly impacts the validity of the work.

Supplementary References

Desmond JE, Glover GH, 2002. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. J Neurosci Methods 118:115–128.

Harrington GS, Farias ST, 2008. Sex differences in language processing: functional MRI methodological considerations. J Magn Reson Imaging 27:1221–1228.

Hoenig K, Kuhl CK, Scheef L, 2005. Functional 3.0-T MR assessment of higher cognitive function: are there advantages over 1.5-T imaging? Radiology 234:860–868.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP, 2006. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp 27:562–571.

Scouten A, Papademetris X, Constable RT, 2006. Spatial resolution, signal-to-noise ratio, and smoothing in multi-subject functional MRI studies. Neuroimage 30:787–793.

Sommer IEC, Aleman A, Bouma A, Kahn RS, 2004. Do women really have more bilateral language representation than men? A meta-analysis of functional imaging studies. Brain 127:1845–1852.

Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB, 2007. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. Neuroimage 35:105–120.

Tieleman A, Vandemaele P, Seurinck R, Deblaere K, Achten E, 2007. Comparison between functional magnetic resonance imaging at 1.5 and 3 Tesla: effect of increased field strength on 4 paradigms used during presurgical work-up. Invest Radiol 42:130–138.

Figures for Supplementary Material

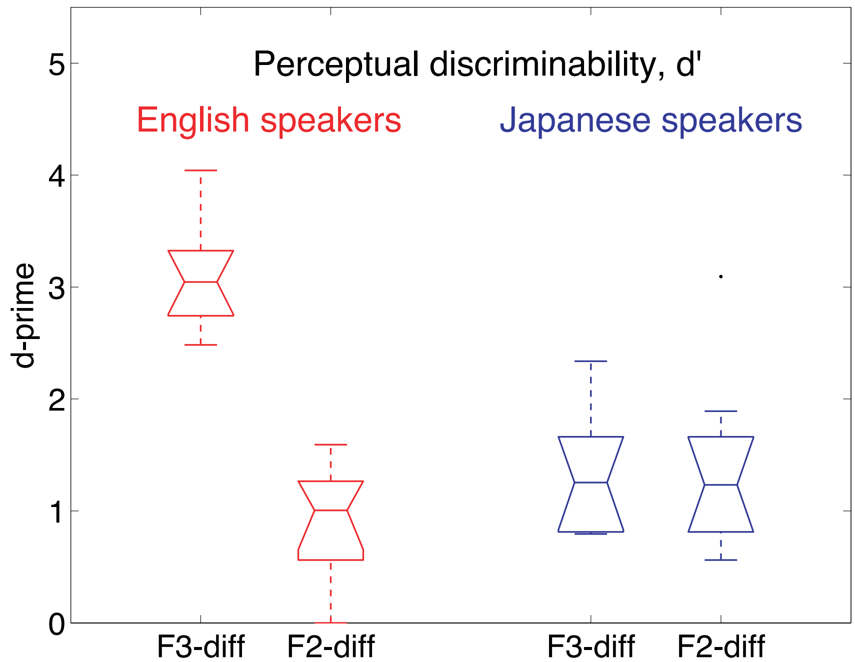


Figure S1: Behavioural responses to the stimuli, obtained by testing the subjects outside of the scanner. For the English speakers, a change in F3 corresponds to a change in category from /ra/ to /la/, and is therefore highly discriminable. However, F2 changes do not induce any category change for the English speakers, but instead are just a form of allophonic variation. For the Japanese speakers, neither F2 nor F3 changes induce category changes in the range covered by the stimuli presented in the fMRI experiment. They are correspondingly insensitive to such differences. However, Japanese speakers are slightly more able to discriminate F2 differences than are English speakers. A much more detailed illustration showing all of the stimuli, their formant frequencies, and how they were perceived, is shown in Supplementary Figure S2.

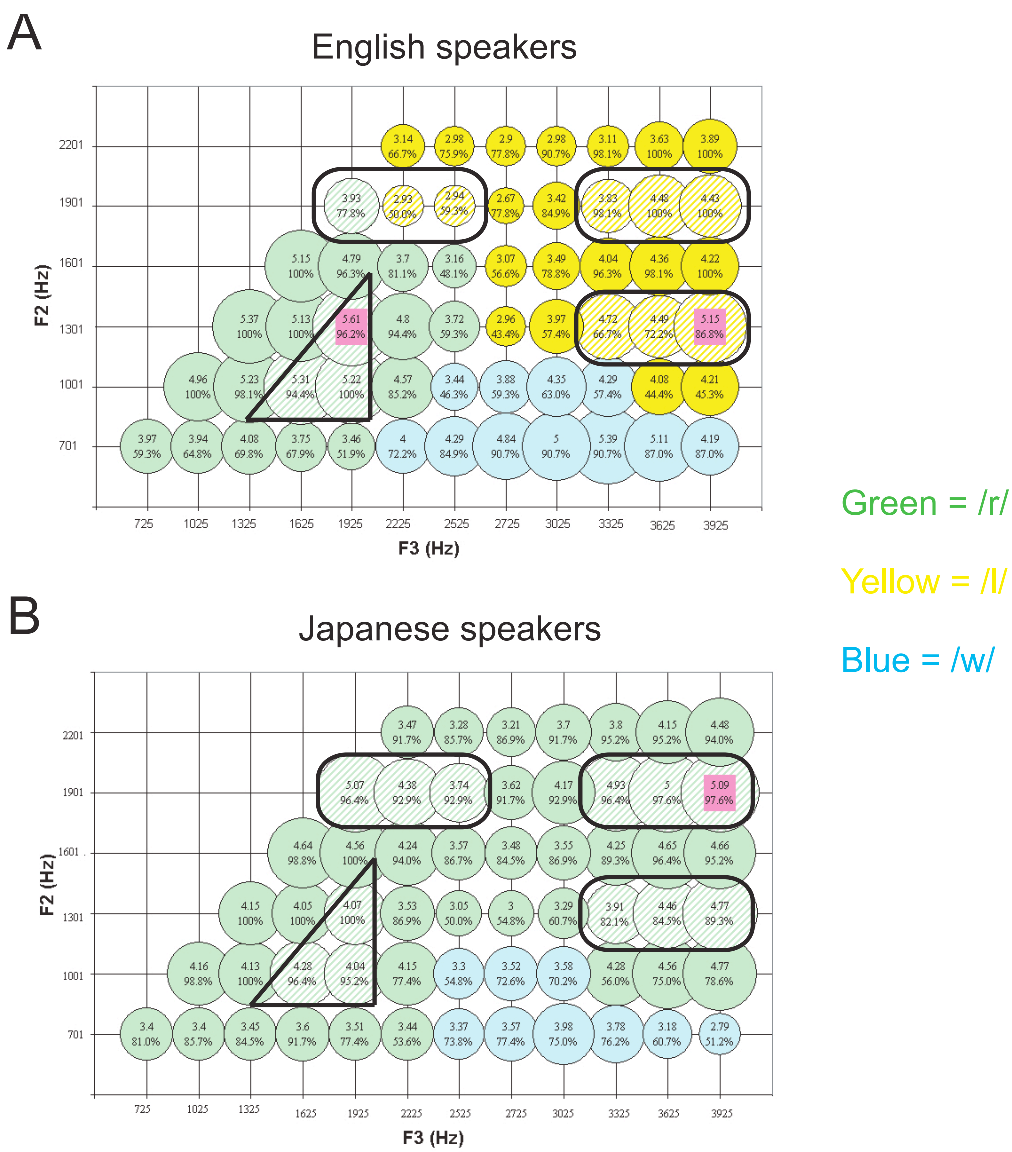


Figure S2: The stimuli presented in an initial behavioural test, in order to determine which should be used for the different group of subjects who would participate in the fMRI experiment. Subjects identified the syllable stimuli, and rated them for category goodness. The English speakers identified the syllables as English /ra/, /la/, and /wa/. The Japanese speakers identified the syllables as Japanese /ra/ and /wa/. The upper number within each circle is the average category-goodness rating given by the subjects to each stimulus, ranging from 1 (worst) to 7 (best). The colour indicates how each stimulus was most often identified (green = /ra/, yellow = /la/, blue = /wa/), and the lower number in each circle is the percentage of the time that identification was made. The stimuli that received the highest rating scores are marked with red squares. The twelve stimuli outlined in black are the ones that were selected to be used for the fMRI experiment. These stimuli were of four sorts: High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3 and Low-F2/Low-F3. The complete set of speech-synthesis input parameters and output results are provided in Table S1 on p. 24, Table S2 on p. 25, and Table S3 on p. 26.

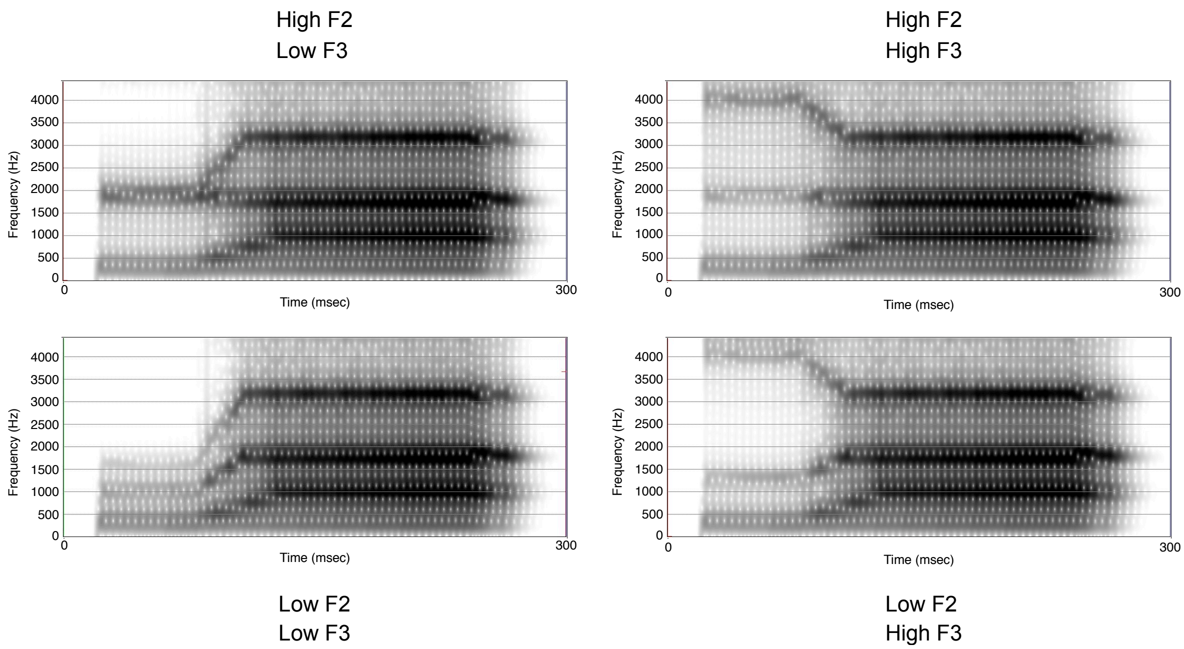


Figure S3: Spectrograms illustrating the formant properties of representative members of the four stimulus types: High-F2/High-F3, High-F2/Low-F3, Low-F2/High-F3 and Low-F2/Low-F3. The complete set of speech-synthesis input parameters and output results are provided in Table S1 on p. 24, Table S2 on p. 25, and Table S3 on p. 26.

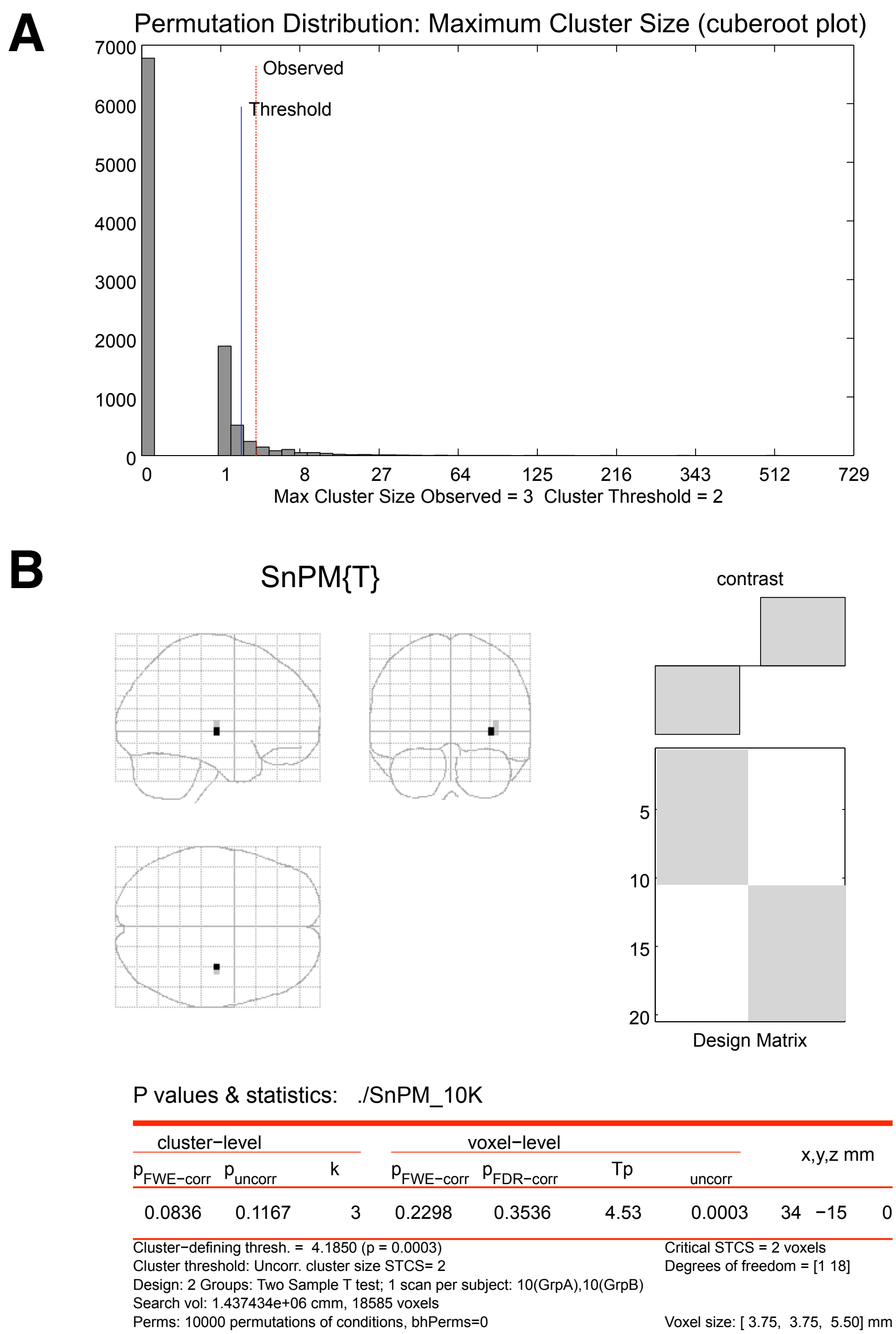


Figure S4: The right Heschl's cluster is significant at the Family-Wise Error (FWE) corrected p-value of p=0.083, as calculated using the SnPM non-parametric analysis package Nichols and Holmes (2002). **(A):** The distribution of maximum cluster size values for the 10,000 random permutations. **(B):** The "glass brain" maximum intensity projection of the right auditory cortex cluster, the simple structure of the English vs. Japanese two-sample t-test, and the corrected and uncorrected p-values and MNI-coordinates of the cluster. Ten thousand random labeling-permutations were performed, and no variance smoothing was applied.

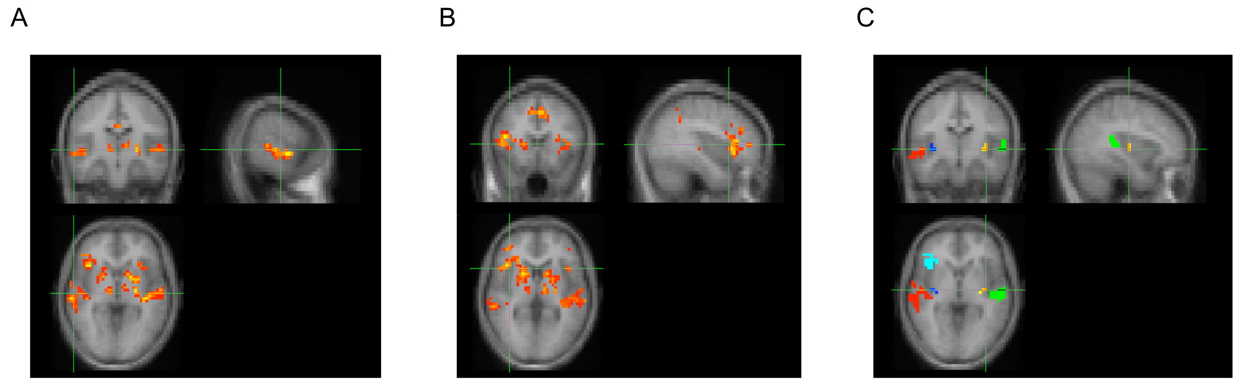


Figure S5: Areas activated by the speech stimuli, as assessed by conventional General Linear Model fMRI analysis. The typical set of language areas were found to be active: **(A):** the superior temporal gyrus (STG) bilaterally, and **(B):** Broca's area on the left. Other activated areas include the caudate and the anterior cingulate. **(C):** Regions of interest (ROIs) of the speech-related areas were calculated, for additional analysis. Cyan: Broca's area. Red: left STG (Wernicke's area). Green: right STG. For comparison purposes, the right primary auditory cortex ROI that revealed statistical separability differences is also shown, in yellow, and its mirror-image ROI on the left side, in dark blue.

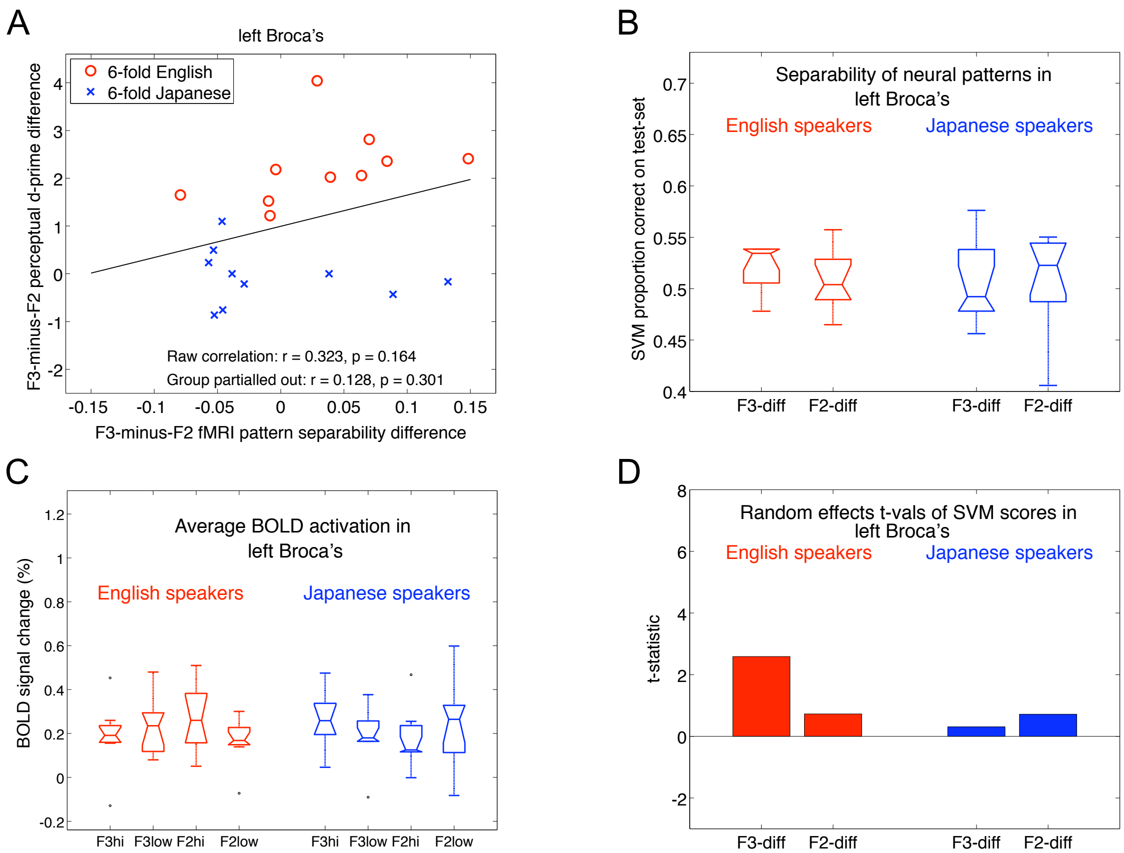


Figure S6: Pattern-based and also standard fMRI analysis results for the Broca's ROI in left frontal cortex shown in Supp.Fig. S5. This ROI was functionally defined using standard fMRI analysis. In panel **D**, the pattern separability of the English speakers' F3 high-vs.-low difference reaches statistical significance at $p < 0.05$, but the other comparisons do not.

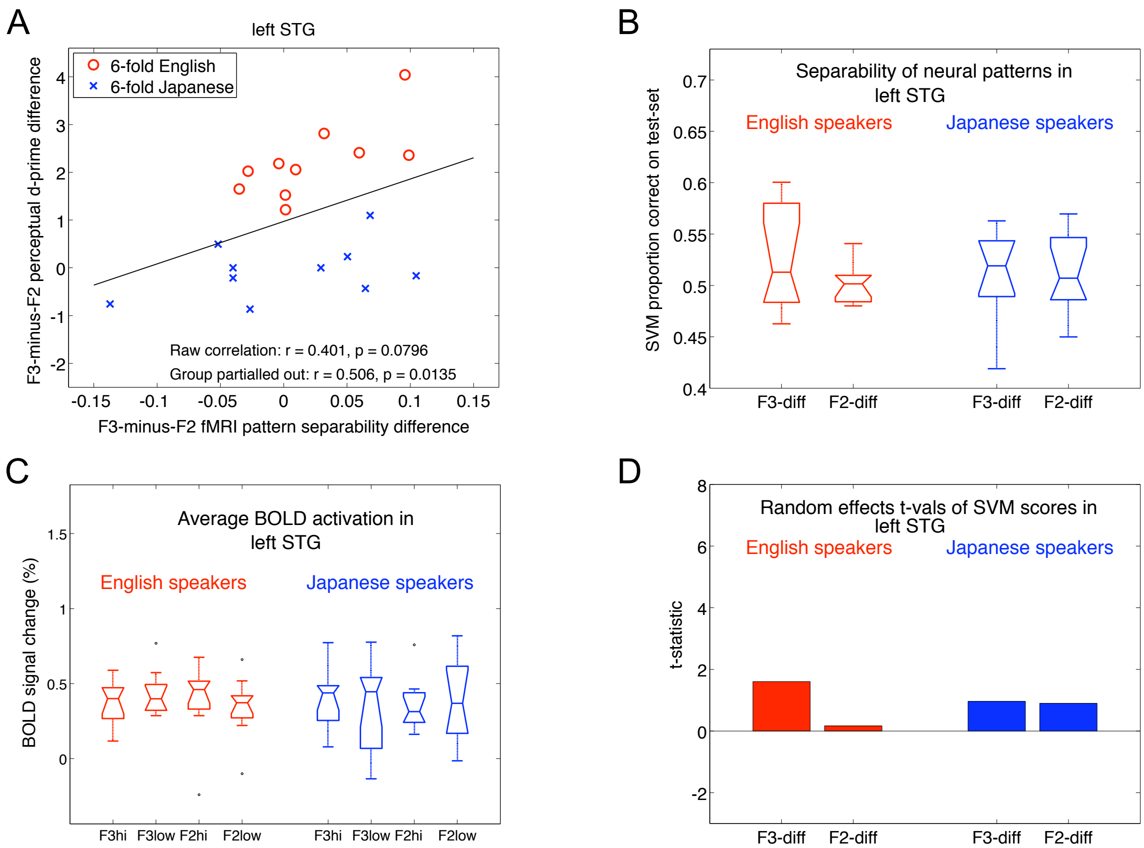


Figure S7: Pattern-based and also standard fMRI analysis results for the left superior temporal gyrus ROI shown in Supp.Fig. S5. This ROI was functionally defined using standard fMRI analysis. In panel **D**, none of the comparisons reaches statistical significance at $p < 0.05$.

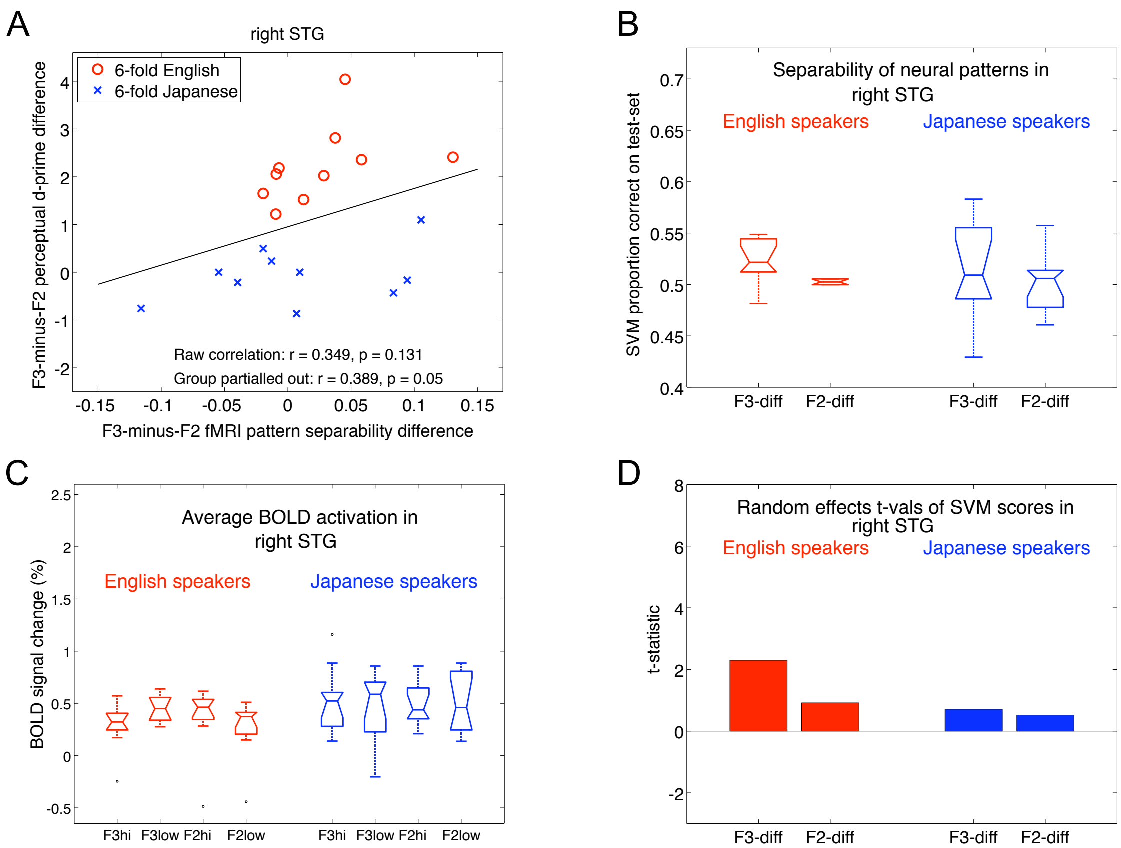


Figure S8: Pattern-based and also standard fMRI analysis results for the right superior temporal gyrus ROI shown in Supp.Fig. S5. This ROI was functionally defined using standard fMRI analysis. In panel **D**, the pattern separability of the English speakers' F3 high-vs.-low difference reaches statistical significance at $p < 0.05$, but the other comparisons do not.

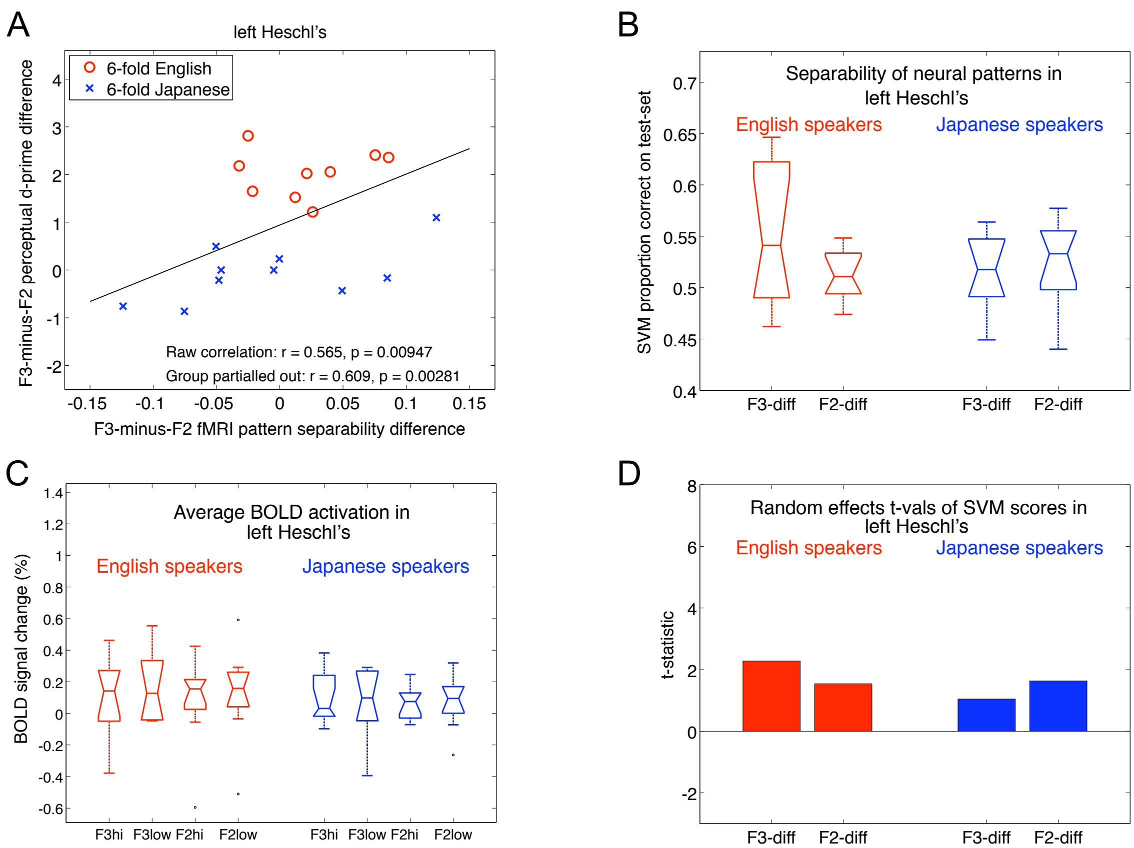


Figure S9: Pattern-based and also standard fMRI analysis results for the mirror-image left-hemisphere equivalent of the right Heschl's ROI found using the English vs. Japanese pattern-based analysis. This ROI is shown in dark blue in Supp.Fig. S5C. In panel **D**, the pattern separability of the English speakers' F3 high-vs.-low difference reaches statistical significance at $p < 0.05$, but the other comparisons do not.

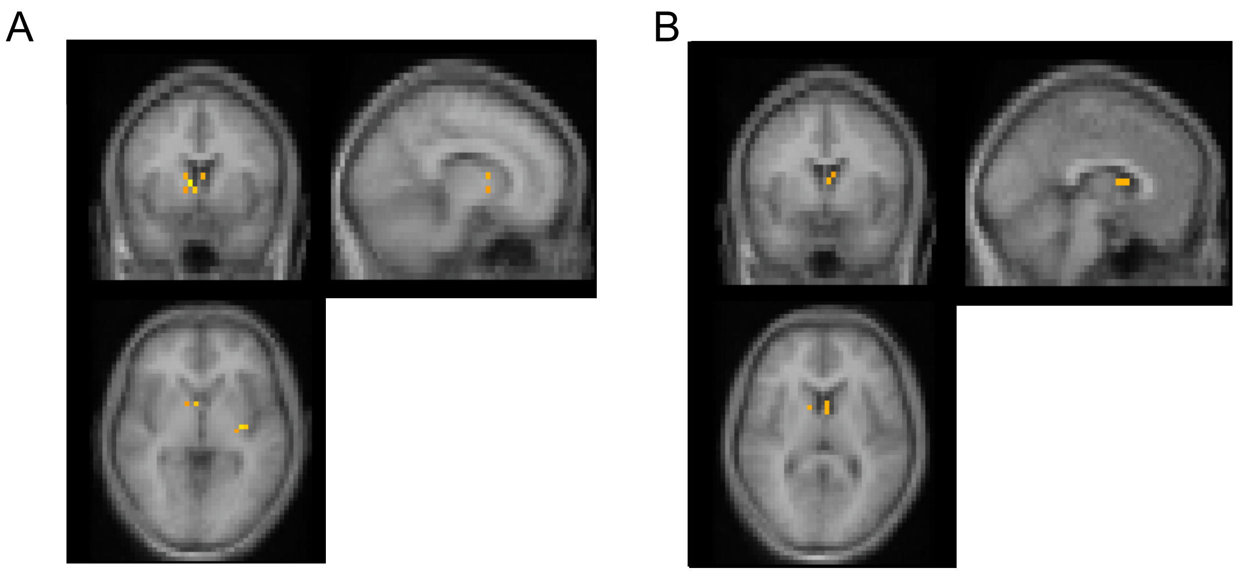


Figure S10: The two other ROIs found as a result of the English-vs.-Japanese F3-minus-F2 pattern separability contrast, in addition to the right Heschl's ROI. **A:** Left head of caudate **B:** Right head of caudate.

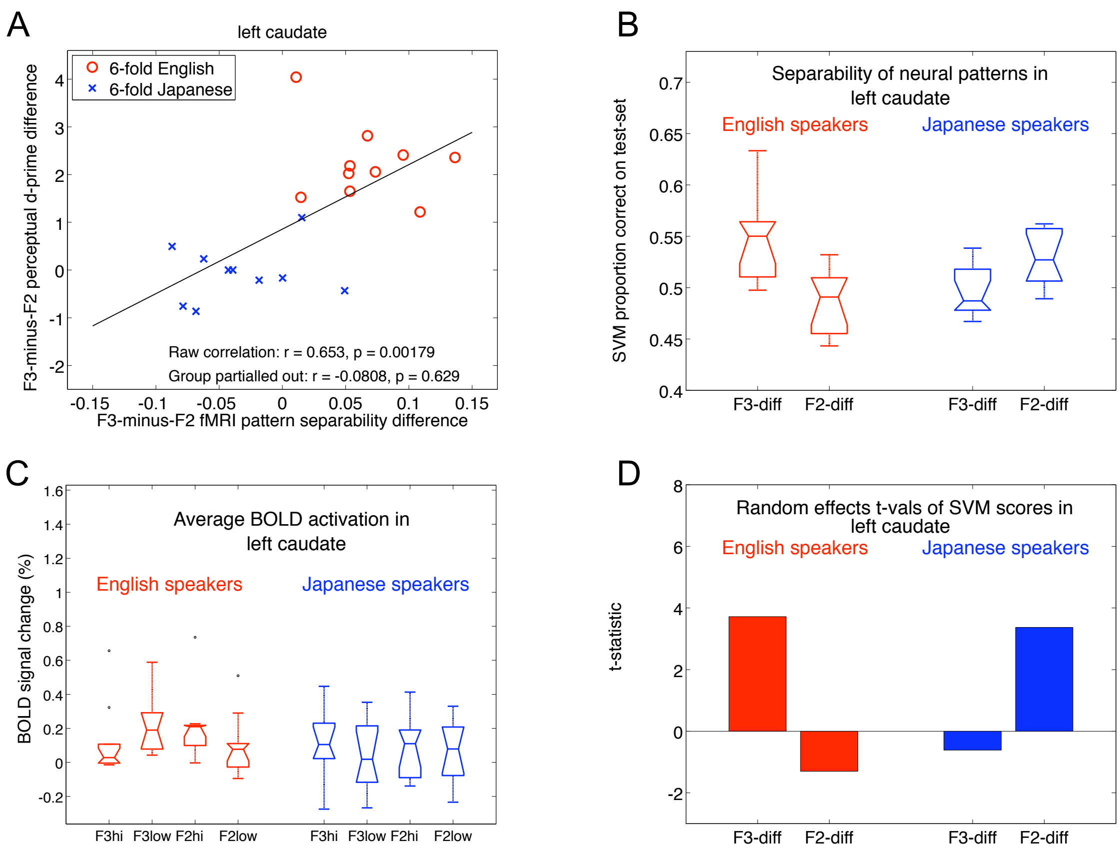


Figure S11: Pattern-based and also standard fMRI analysis results for the left head-of-caudate ROI which was found, along with right caudate and right Heschl's gyrus, using the English vs. Japanese pattern-based analysis. In panel **D**, the pattern separability of the English speakers' F3 high-vs.-low difference and of the Japanese speakers' F2 high-vs.-low difference reach statistical significance at $p < 0.05$, but the other comparisons do not.

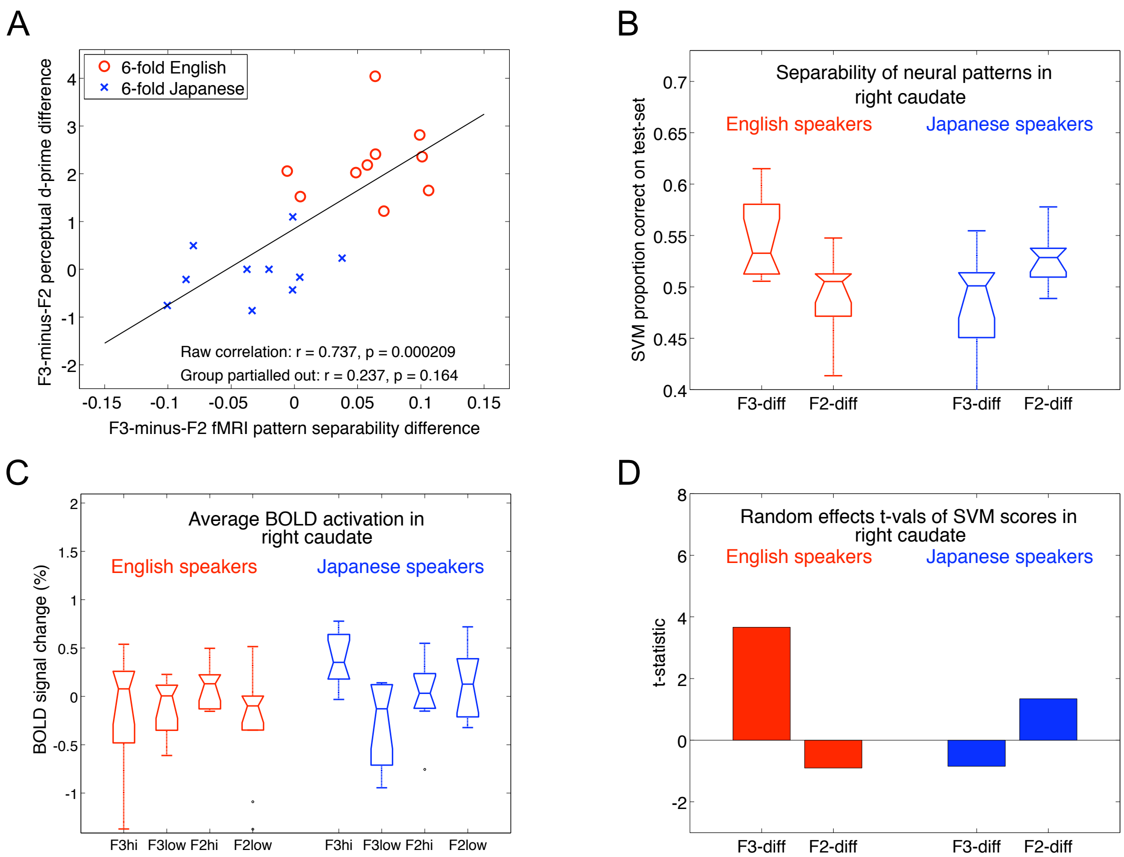


Figure S12: Pattern-based and also standard fMRI analysis results for the right head-of-caudate ROI which was found, along with left caudate and right Heschl's gyrus, using the English vs. Japanese pattern-based analysis. In panel **D**, the pattern separability of the English speakers' F3 high-vs.-low difference reaches statistical significance at $p < 0.05$, but the other comparisons do not.

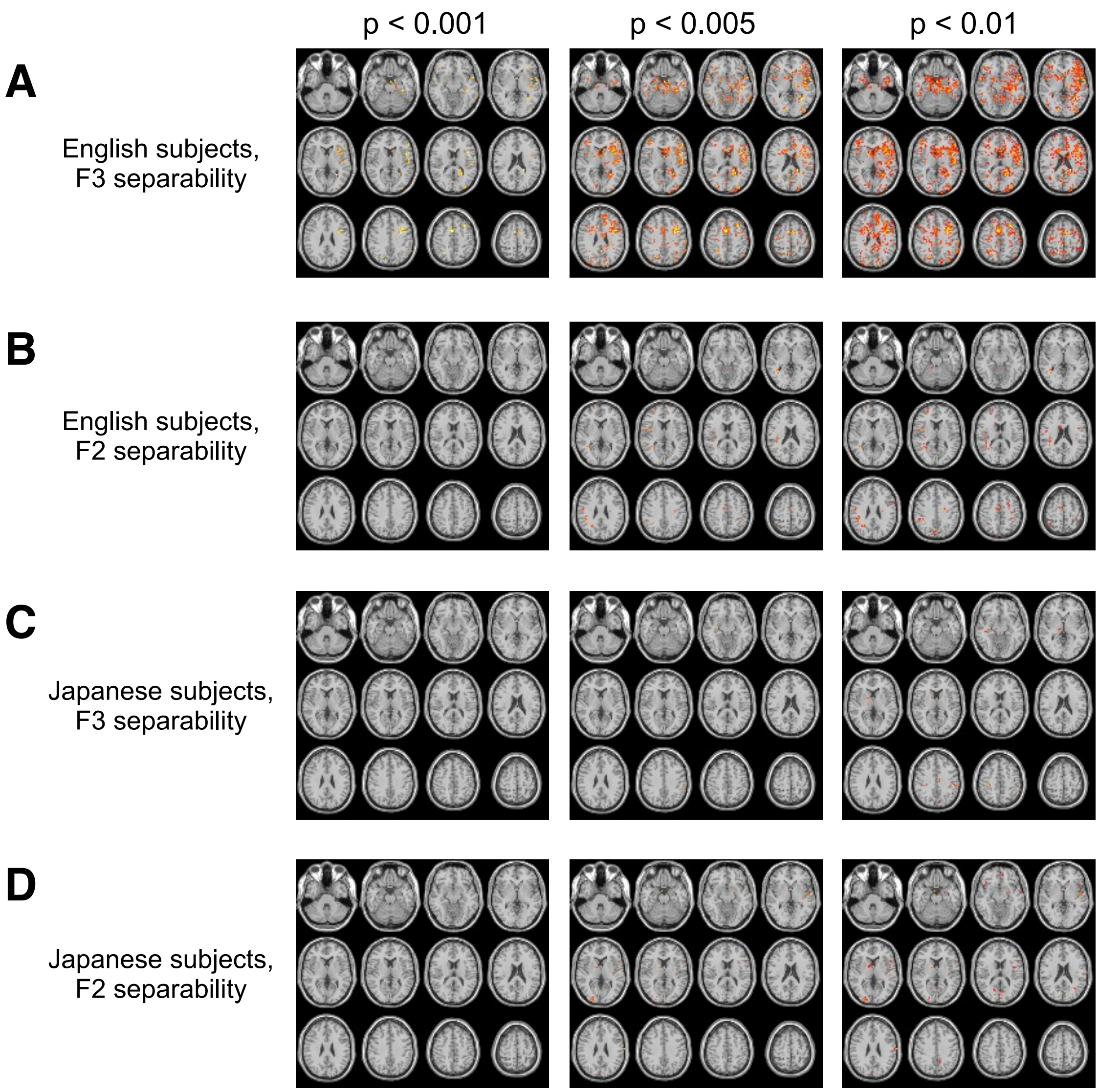


Figure S13: F3 high-vs.-low and F2 high-vs.-low pattern separability for the English and Japanese speakers, viewed across multiple slices of the brain at multiple statistical thresholds.

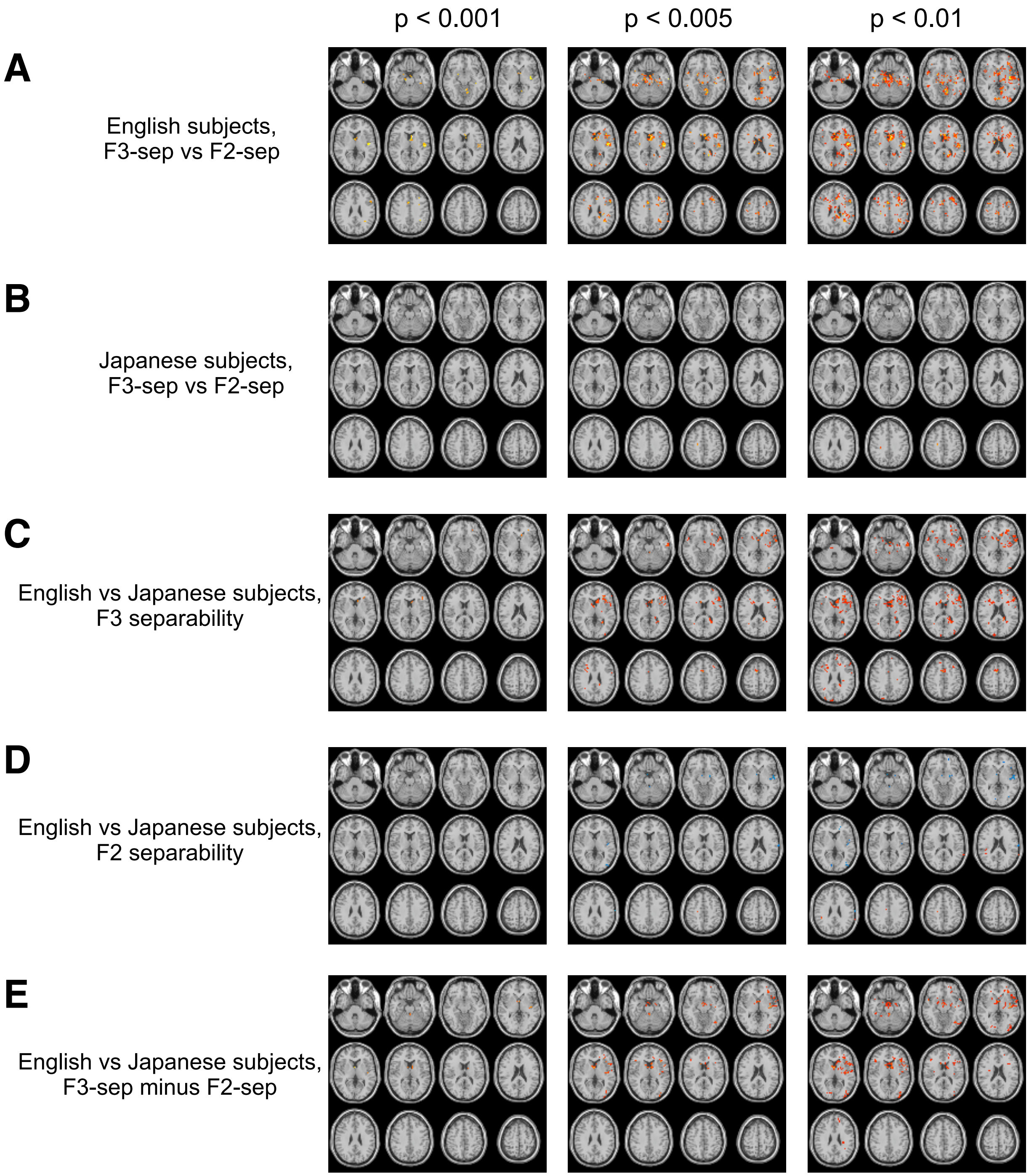
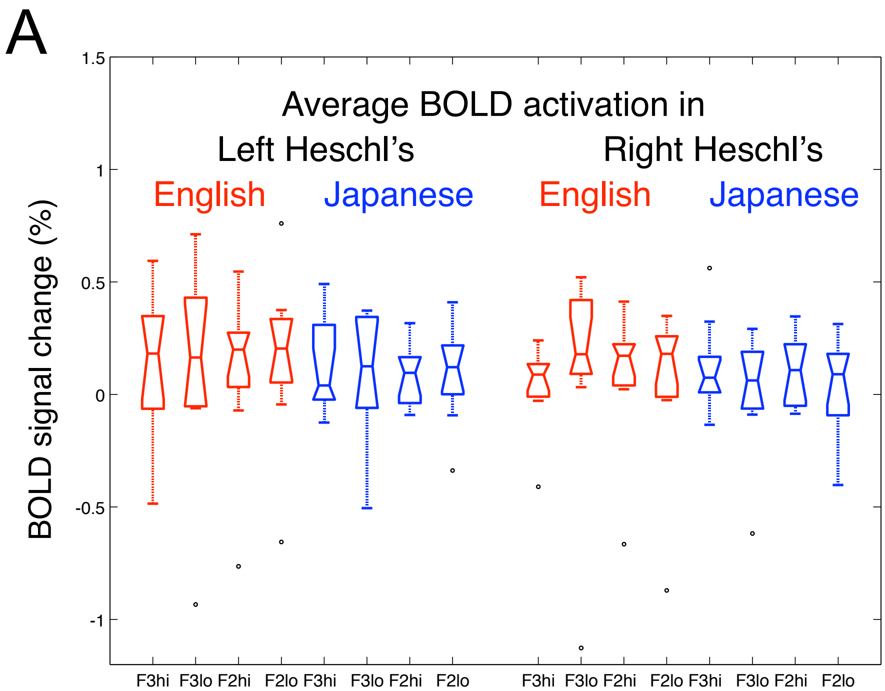


Figure S14: The full range of F3-vs.-F2 and English-vs.-Japanese pattern separability analyses for the English and Japanese speakers, viewed across multiple slices of the brain at multiple statistical thresholds.



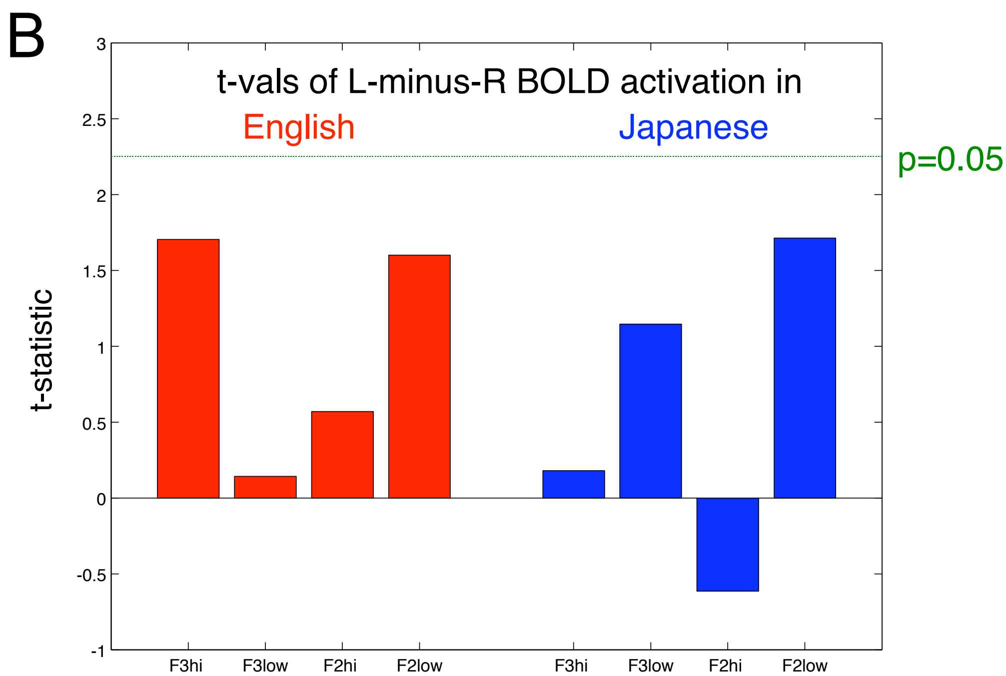


Figure S15: Direct left-vs.-right comparisons of standard fMRI BOLD activation in the right Heschl's gyrus ROI and its mirror-image ROI on the left. The left-vs.-right differences in panel **B** do not reach statistical significance at $p < 0.05$. Green dotted lines indicate the critical value for p=0.05 of t=2.26, two-tailed, df = 9.

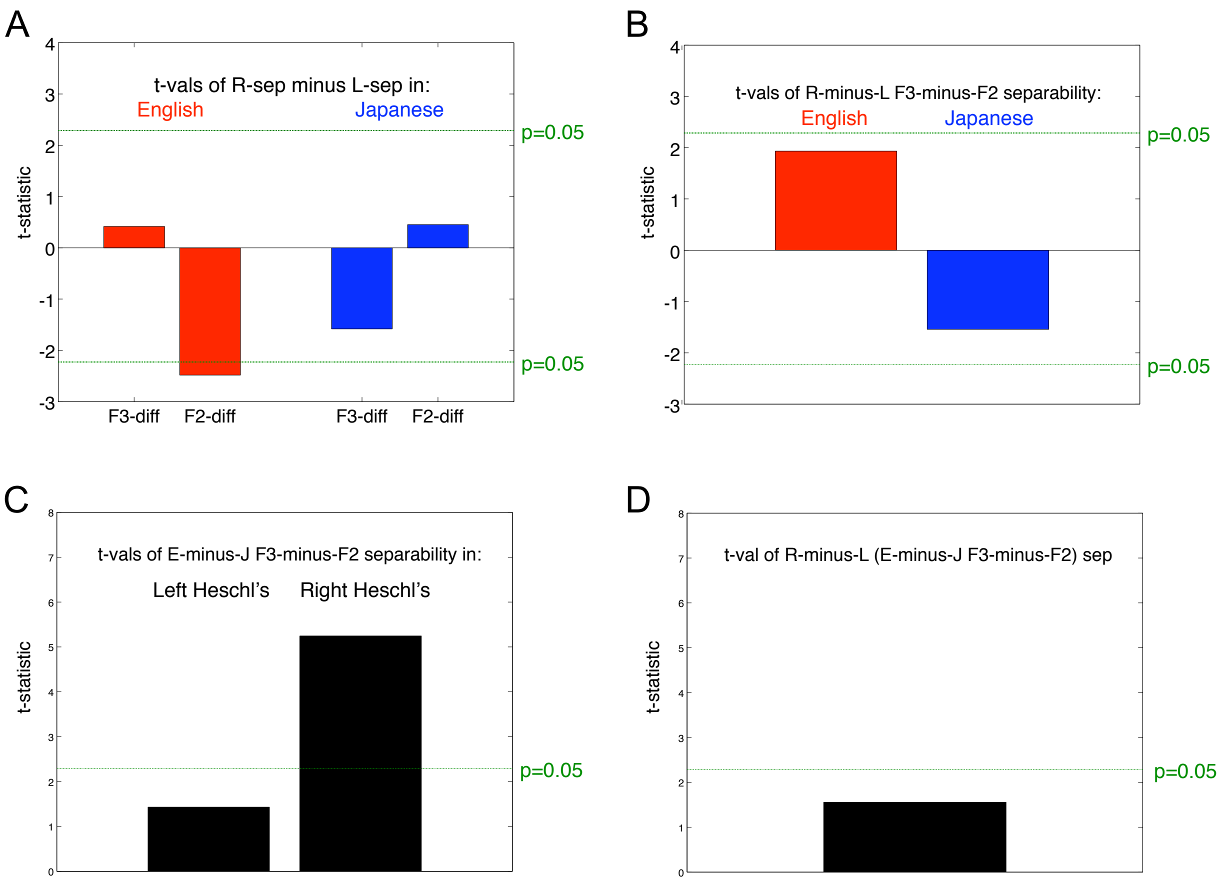


Figure S16: Direct left-vs.-right comparisons of fMRI pattern-separability differences in the right Heschl's gyrus ROI and its mirror-image ROI on the left. Green dotted lines indicate the critical value for p=0.05 of t=2.26, two-tailed, df = 9. **(A):** English speakers' F2 high-vs.-low pattern separability is greater in left than right at $p < 0.05$ (t = -2.48, p = 0.035, df = 9). The other comparisons do not reach significance. **(B):** The right-versus-left F3-minus-F2 separability comparison fail to reach significance for either the English speakers or Japanese speakers. **(C):** This comparison is significant in the right Heschl's ROI ($t = 5.24$, $p = 5 \times 10^{-5}$, df = 9), but trivially so, as this was how that ROI was defined. The comparison is not significant in the left-hemisphere mirror-image ROI. **(D):** This right-vs.-left comparison does not reach significance.

t

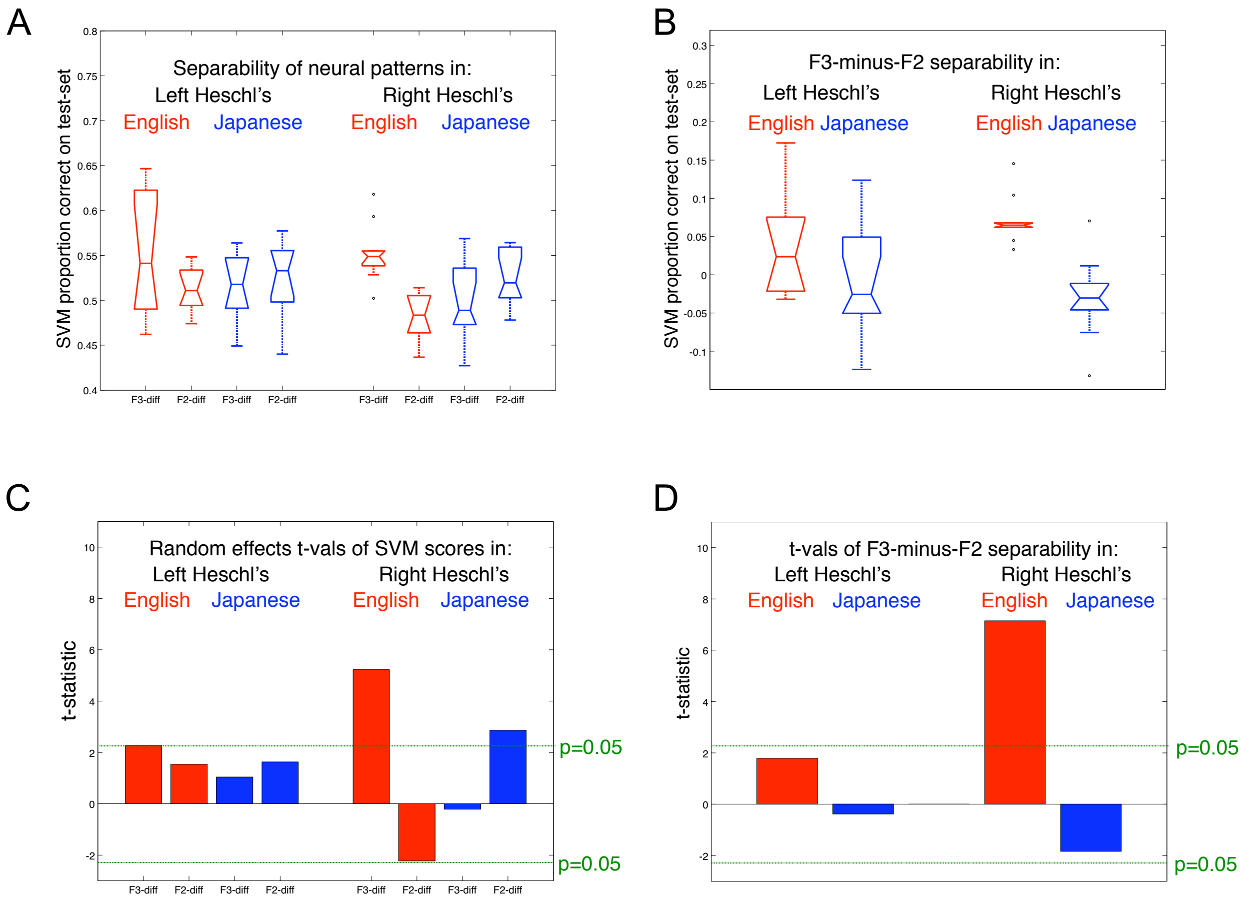


Figure S17: F3-versus-F2 comparisons of fMRI pattern-separability differences in the right Heschl's gyrus ROI and its mirror-image ROI on the left. Green dotted lines indicate the critical value for p=0.05 of t=2.26, two-tailed, df = 9. **(A):** The fMRI pattern separability values for F3-changes and F2-changes considered individually. **(B):** The differences between F3-change fMRI pattern separability and F2-change fMRI pattern separability. This subtraction will be greater than zero when F3-changes produce more separable fMRI patterns than do F2-changes, in a manner analogous to the behavioural result that for English speakers F3-changes are more perceptually discriminable than are F2-changes. **(C):** The following comparisons reach significance at $p < 0.05$: left Heschl's English speakers' F3-difference (t = 2.28, p = 0.49, df = 9), right Heschl's English speakers' F3-difference (t = 5.23, p = 5 $\times 10^{-4}$, df = 9), right Heschl's Japanese speakers' F2-difference (t = 2.86, p = 0.02, df = 9). **(D):** The F3-minus-F2 pattern separability difference for English speakers in right Heschl's is significant (t = 7.15, p = 5 $\times 10^{-5}$, df = 9), the other comparisons are not.

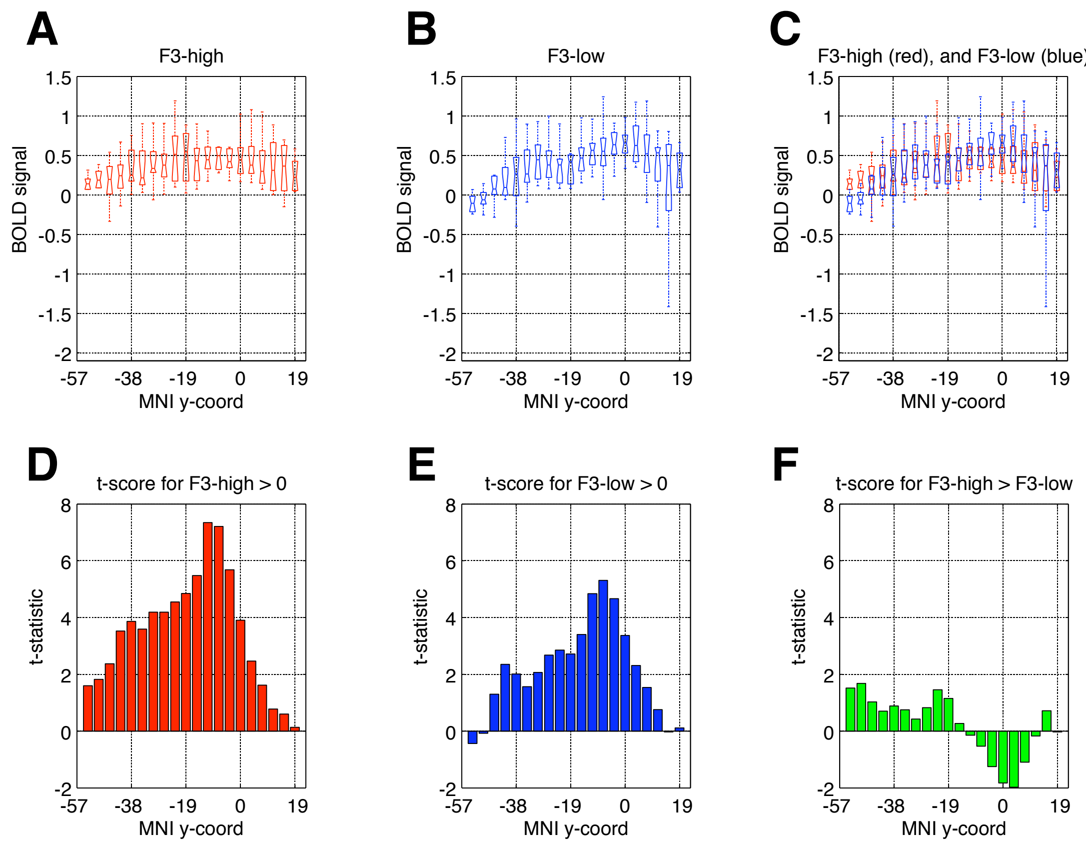


Figure S18: The activation profile along a line running in the y-direction through the center of the BOLD activation in left auditory cortex (x-coord = 60, y-coord range = -52.5 to 18.75 z-coord = -5.5). It can be seen the the activations for /ra/ (F3-low) and /la/ (F3-high) have different peaks and intensities. However, panels C and F illustrate, these two activation distributions are not significantly different. Thus, in our data at least, standard fMRI does not reveal a difference between the two phonemes.

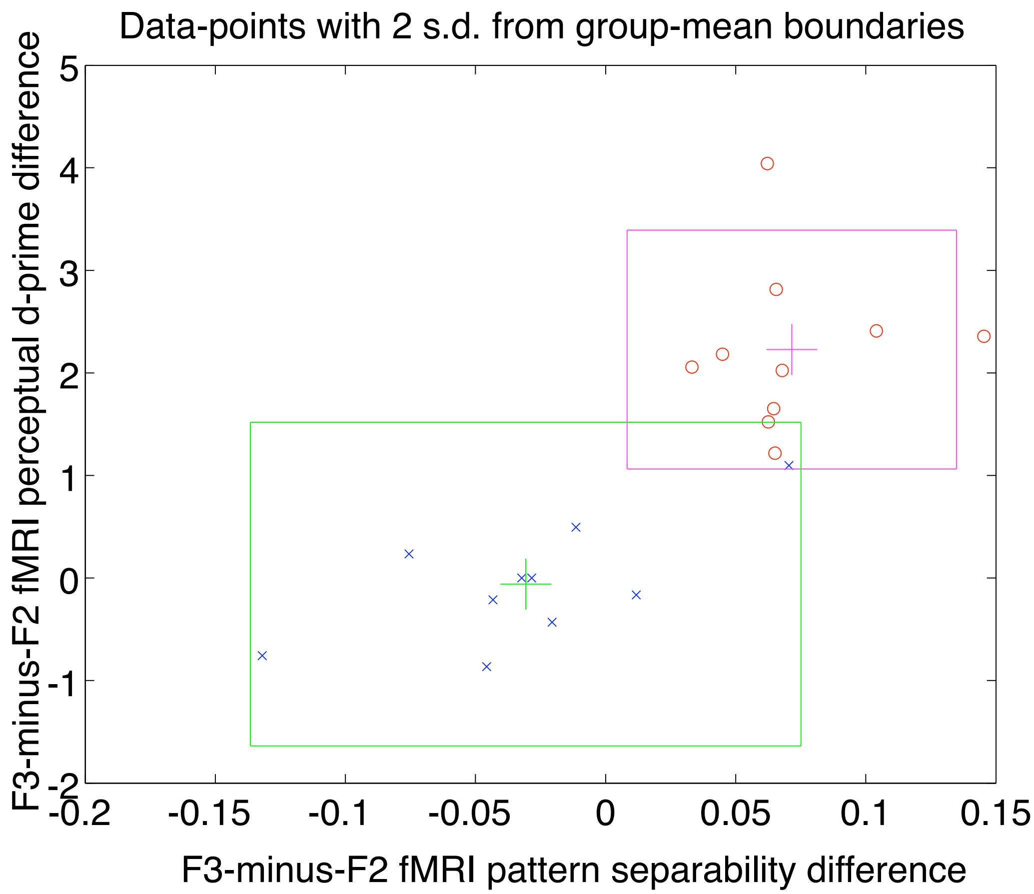


Figure S19: The F3-minus-F2 scatterplot from Fig. 4 of the main text, but now with $\pm$ 2 s.d. boundaries for the English and Japanese groups drawn on. This reveals that none of the Japanese speakers lie beyond this 2 s.d. range, but two of the English speakers do. The raw correlation, before excluding any outliers, is r = 0.796, $p < 3 \times 10^{-5}$. After partialling out the effect of group-membership, but again before excluding any outliers, this becomes r = 0.389, $p < 0.05$. With each of these two outlier points removed in turn, we then recalculated the correlations with and without group-membership partialled out. In both cases, the correlations remained significant. Specifically, one outlier (as determined by the 2 s.d. criterion) was the English speaker with the largest d-prime score. With this subject excluded, the partial correlation was r = 0.558 (p = 0.008). The other 2-s.d. outlier was the English speaker with the largest F3-minus-F2 SVM-separability score. With this subject excluded, the partial correlation was r = 0.408 (p = 0.047).

Tables for Supplementary Material

Parameter	Description	Values
DU	Duration of the utterance	400 ms (silence added after synthesis of 300ms long tokens)
UI	Update interval for parameter reset, in msec	55
SR	Output sampling rate	11025 Hz
NF	Number of formants	4
SS	Source switch	natural
RS	Random seed	8
SB	Same burst	1
GV	Overall gain scale factor for amplitude of voicing	60 dB
GH	Overall gain scale factor for amplitude of aspiration	60 dB
GF	Overall gain scale factor for amplitude of frication	60 dB
OQ	Open quotient	65%
SQ	Speech quotient	200%
AF	Amplitude of frication	0 dB
FNP	Frequency of nasal pole	500 Hz
BNP	Bandwidth of nasal pole	90 Hz
FNZ	Frequency of nasal zero	500 Hz
BNZ	Bandwidth of nasal zero	90 Hz

Table S1: KL parameters entered into the HLsyn program, for the stimuli with the lowest F2 and F3 frequencies: Part 1.

Parameter	Description	Values
AV	Amplitude of voicing	0 dB (0-10 ms) Transition from 0 to 53 dB (10-20 ms) 53 dB (20-60 ms) Transition from 53 to 55 dB (60-65 ms) Transition from 55 to 60 dB (65-110 ms) Transition from 60 to 65 dB (110-215 ms) 65 dB (215-240 ms) Transition from 65 to 60 dB (240-255 ms) Transition from 60 to 50 dB (255-265 ms) Transition from 50 to 0 dB (265-300 ms) 0 dB (300-400 ms)
TL	Extra tilt of voicing spectrum	0 dB (0-80 ms) Transition from 0 to 8 dB (80- 130 ms) 8 dB (130-400 ms)
AH	Amplitude of aspiration	0 dB (0-10 ms) Transition from 0 to 23 dB (10-20 ms) Transition from 23 to 28 dB (20-60 ms) 28 dB (60-65 ms) Transition from 28 to 40 dB (65-110 ms) 40 dB (110-240 ms) Transition from 40 to 30 dB (240-255 ms) Transition from 30 to 20 dB (255-265 ms) Transition from 20 to 0 dB (265-300 ms) 0 dB (300-400 ms)
F0	Fundamental frequency	0 Hz (0-20 ms) Transition from 0 to 231 Hz (20-25 ms) Transition from 231 to 221 Hz (25- 45 ms) 221 Hz (45 - 67 ms) Transition from 221 to 247 Hz (67-100 ms) 247 Hz (100 - 240 ms) Transition from 247 to 220 Hz (240- 265 ms) 220 Hz (265 - 400 ms)
F1	Frequency of 1st formant	365 Hz (0-80 ms) Transition from 365 to 965 Hz (80-130 ms) 965 Hz (130-400 ms)
B1	Bandwidth of 1st formant	200 Hz
F2	Frequency of 2nd formant	1001 Hz (0-80 ms) Transition from 1001 to 1807 Hz (80-130 ms) 1807 Hz (130- 400 ms)
B2	Bandwidth of 2nd formant	100 Hz
F3	Frequency of 3rd formant	1625 Hz (0-80 ms) Transition from 1625 to 3164 Hz (80-130 ms) 3164 Hz (130-400 ms)
B3	Bandwidth of 3rd formant	150 Hz
F4	Frequency of 4th formant	4512 Hz
B4	Bandwidth of 4th formant	100 Hz (0-80 ms) Transition from 100 to 400 Hz (80-130 ms) 400 Hz (130-400 ms)

Table S2: KL parameters entered into the HLsyn program, for the stimuli with the lowest F2 and F3 frequencies: Part 2.

Input values of KL parameters		Actual output values of KL parameters		Phonetic identification (%) (inside scanner)		Phonetic identification (%) (outside scanner)	
F2	F3	F2	F3	English speakers	Japanese speakers	English speakers	Japanese speakers
1001	1625	1069	1670	/ra/ (95%)	/ra/ (50%)	/ra/ (94.4%)	/ra/ (96.4%)
1001	1925	1062	1991	/ra/ (100%)	/ra/ (55%)	/ra/ (100%)	/ra/ (95.2%)
1301	1925	1350	2000	/ra/ (100%)	/ra/ (75%)	/ra/ (100%)	/ra/ (100%)
1301	3425	1349	3356	/la/ (70%)	/ra/ (55%)	/la/ (66.7%)	/ra/ (82.1%)
1301	3725	1350	3630	/la/ (75%)	/ra/ (60%)	/la/ (72.2%)	/ra/ (84.5%)
1301	4025	1351	3985	/la/ (90%)	/ra/ (80%)	/la/ (86.8%)	/ra/ (89.3%)
1901	1925	1850	2119	/ra/ (75%)	/ra/ (95%)	/ra/ (77.8%)	/ra/ (96.4%)
1901	2225	1930	2333	/la/ (55%)	/ra/ (100%)	/la/ (50%)	/ra/ (92.9%)
1901	2525	1948	2548	/la/ (55%)	/ra/ (90%)	/la/ (59.3%)	/ra/ (92.9%)
1901	3425	1952	3358	/la/ (100%)	/ra/ (100%)	/la/ (98.1%)	/ra/ (96.4%)
1901	3725	1955	3633	/la/ (100%)	/ra/ (95%)	/la/ (100%)	/ra/ (97.6%)
1901	4025	1943	4100	/la/ (100%)	/ra/ (90%)	/la/ (100%)	/ra/ (97.6%)

Table S3: Input and output values of the synthesised speech stimuli, and the subjects' behavioural responses to them.