Supplementary materials

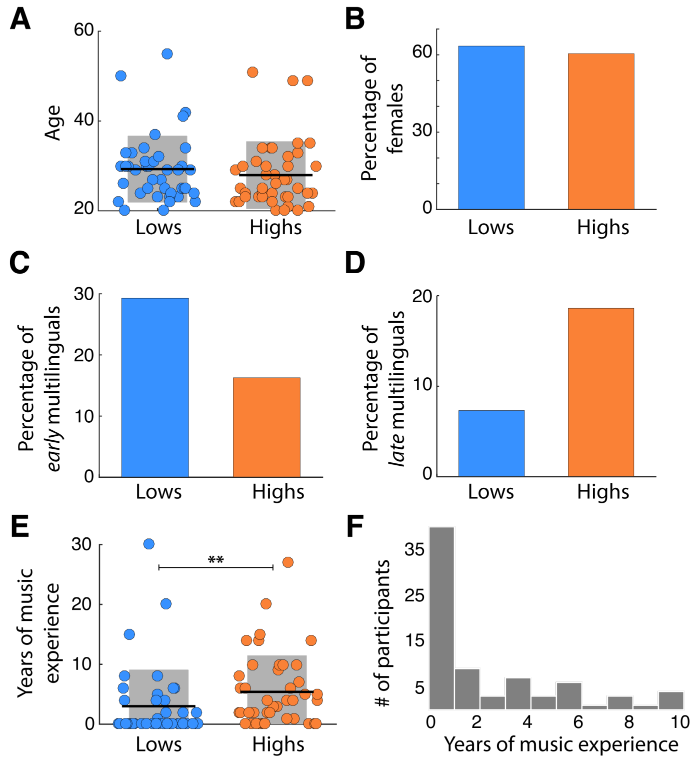

Fig. S1. Demographic information. (A) Age distribution within groups, no significant difference between groups' age (Mann-Whitney-Wilcoxon test, two-sided $p = 0.22$,). **(B)** Percentage of female participants within each group, no significant difference in the gender distribution between groups (Fisher's exact test, $p = 0.82$). **(C)** Percentage of early multilingual participants; they learned at least one other language in addition to English before 4 years of age. No significant difference between groups (Fisher's exact test, $p = 0.19$). **(D)** Percentage of late multilingual participants: participants who learned at least one other language in addition to English after 4 years of age. No significant difference between groups (Fisher's exact test, $p = 0.19$). **(E)** Years of music experience: number of years that the participant played at least one instrument, more than 2 hours per week (Mann-Whitney-Wilcoxon test, two-sided $p = 0.0032$). Double asterisks stands for significant difference between conditions. In all previous panels: Orange/light blue correspond to high/low synchronizers ($N_{high} = 43$ and $N_{low} = 41$), dots to individual participants, black lines to mean across participants, and shadowed region the SD. **(F)** Histogram of the years of music experience ($N = 75$). Participants above one SD were excluded.

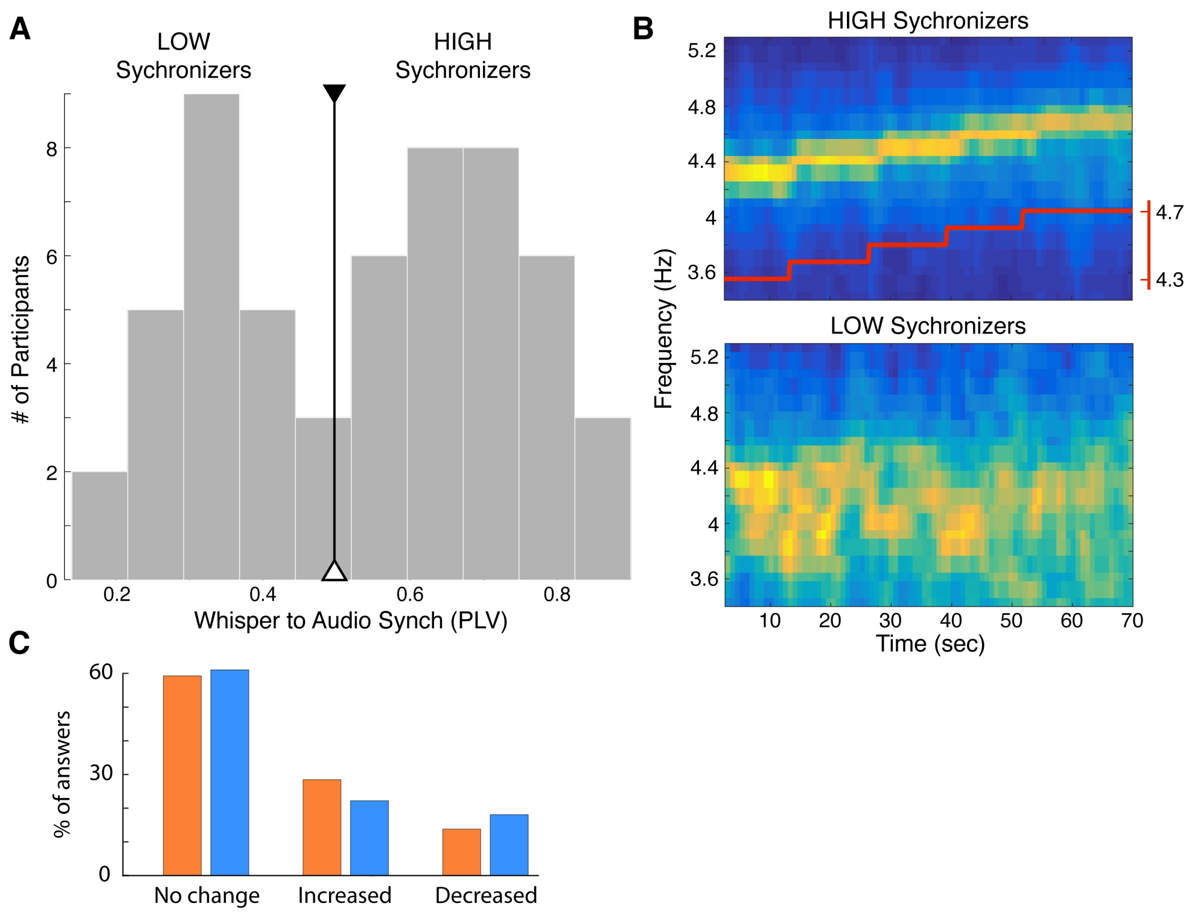

Fig. S2. SSS-test, accelerated version. (A) Histogram of the PLVs between the envelope of the perceived and produced speech signals, bandpass filtered at 3.5-5.5 Hz. Two clusters were obtained by a *kmeans* algorithm (black line represents the threshold; individuals above/below this line are labeled as high/low). **(B)** Produced speech envelopes' spectrograms. Upper panel: average across high synchronizers ($N_{high} = 33$); the red trace represents the time evolution of the perceived syllable rate. Lower panel: average across low synchronizers ($N_{low} = 22$). **(C)** Reported perception. Percent of blocks reported as: "The rate of the presented syllables did not change/increased/decreased". Orange/blue correspond to high/low synchronizers' answers.

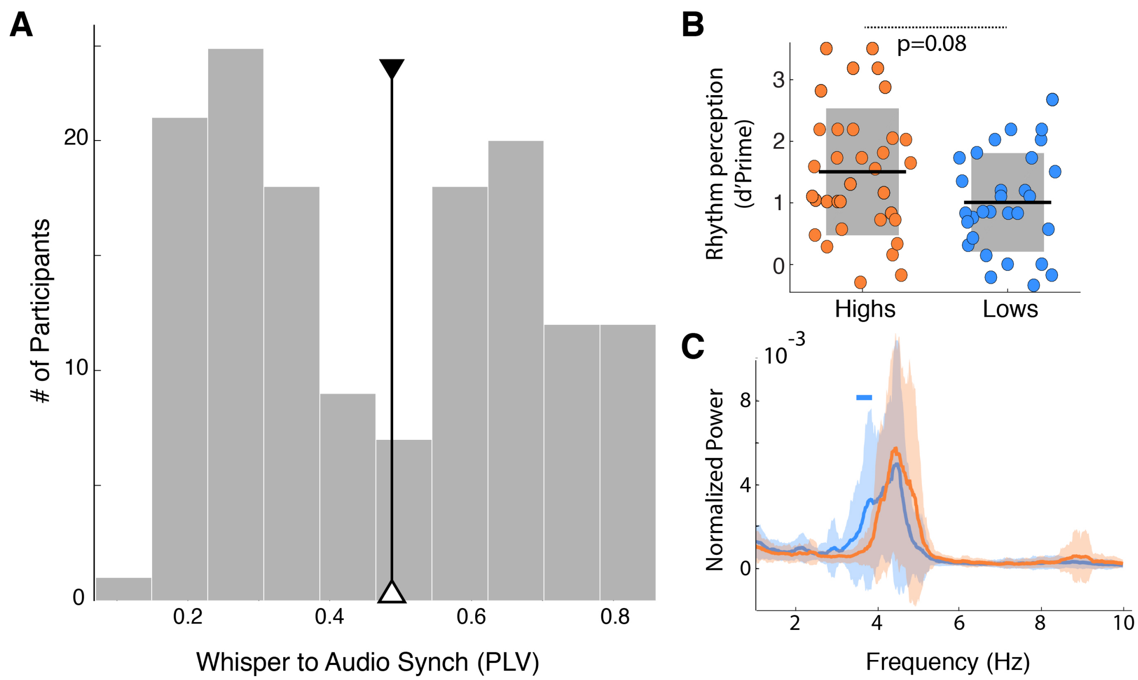

Fig. S3. SSS-test, online version. (A) Histogram of the PLVs between the envelope of the perceived and produced speech signals, bandpass filtered at 3.5-5.5 Hz. Two clusters were obtained by a *kmeans* algorithm (black line represents the threshold; individuals above/below this line are labeled as high/low). **(B)** Rhythm perception task: participants' d prime scores. High synchronizers are marginally better than the lows (N_{high} = 35 and N_{low} = 34; Mann-Whitney-Wilcoxon test, two-sided p = 0.08). Dots: individual subjects. Black lines: mean across participants. Shadowed region: SD. **(C)** Rhythm production task: Average spectra of the utterances' envelopes. Low synchronizers showed more power, although only marginally, than highs for frequencies from 3.6 to 3.7 Hz. Straight lines on top: marginal difference between groups (N_{high} = 31 and N_{low} = 44; Wilcoxon signed rank test, two-sided $p_{uncorrected} < 0.001$). No frequency survived a FDR correction under two-sided p = 0.05. Shadowed region: SD.

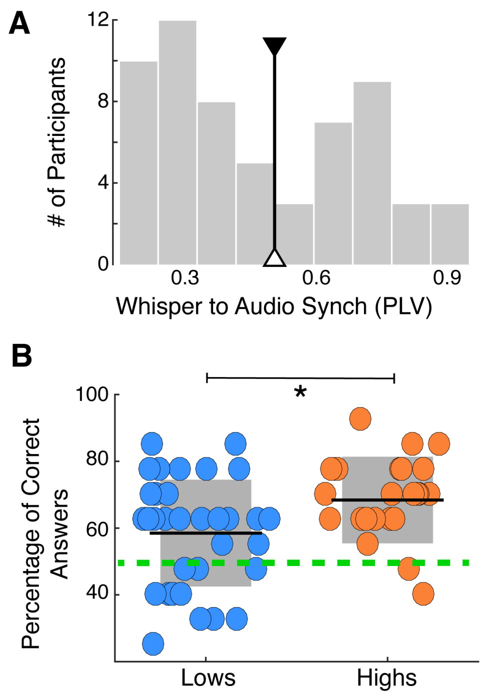

Fig. S4. Word learning task and accelerated SSS-test, online version. (A) Histogram of the PLVs between the envelope of the perceived and produced speech signals, bandpass filtered at 3.5-5.5 Hz. Two clusters were obtained by a *kmeans* algorithm (black line represents the threshold; individuals above/below this line are labeled as high/low). (**B**) Percent of correct answers for the statistical word-learning task (N_{high} =25, N_{low}=35; r = 0.37, Rank-Biserial Correlation; Mann-Whitney-Wilcoxon test, two-sided p = 0.015). Orange/light blue correspond to high/low synchronizers. * p<0.05. Dots: individual participants. Black lines: mean across participants. Shadowed region: SD. Green dashed line: chance level in a two alternative forced-choice post-learning task

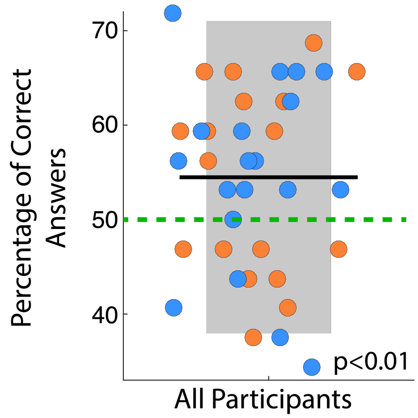

Fig. S5. Behavioral result, neurophysiological study. Syllable detection task, percent of correctly identified syllables (N = 37, Wilcoxon Signed-Rank test, two-sided p = 0.011). Orange/blue correspond to high/low synchronizers. Dots: individual participants. Black lines: mean across participants. Shadowed region: SD. Green dashed line: chance level.

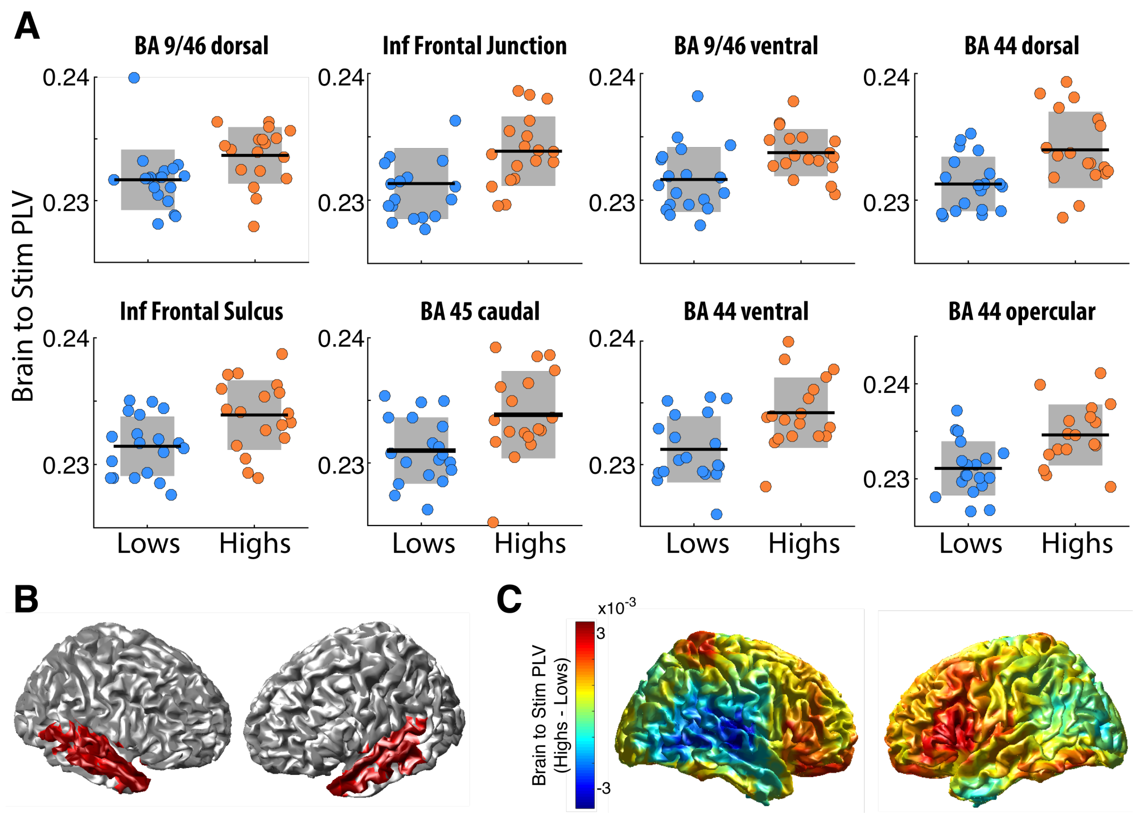

Fig. S6. Brain to stimulus synchronization, neurophysiological study. (A) PLV between brain activity and the cochlear envelope of the perceived syllables within each region where high and low synchronizers were significantly different. The aim of the scatter plots is to visualize the magnitude of the effect. Accordingly, the Rank-Biserial correlation for each region is: $r_{BA9/46d} = 0.55$, $r_{IFJ} = 0.52$, $r_{BA9/46v} = 0.51$, $r_{BA44d} = 0.52$, $r_{IFS} = 0.51$, $r_{BA45c} = 0.52$, $r_{BA44v} = 0.59$, and $r_{BA44op} = 0.53$. Orange/blue correspond to high/low synchronizers ($N_{high} = 18$ and $N_{low} = 19$). Dots: individual participants. Black lines: mean across participants. Shadowed region: SD. Green dashed line: chance level. **(B)** Temporal ROI comprising bilateral superior, middle and posterior temporal gyri. **(C)** Whole brain surface map showing the PLV differences between groups (PLV_{highs} - PLV_{lows}).

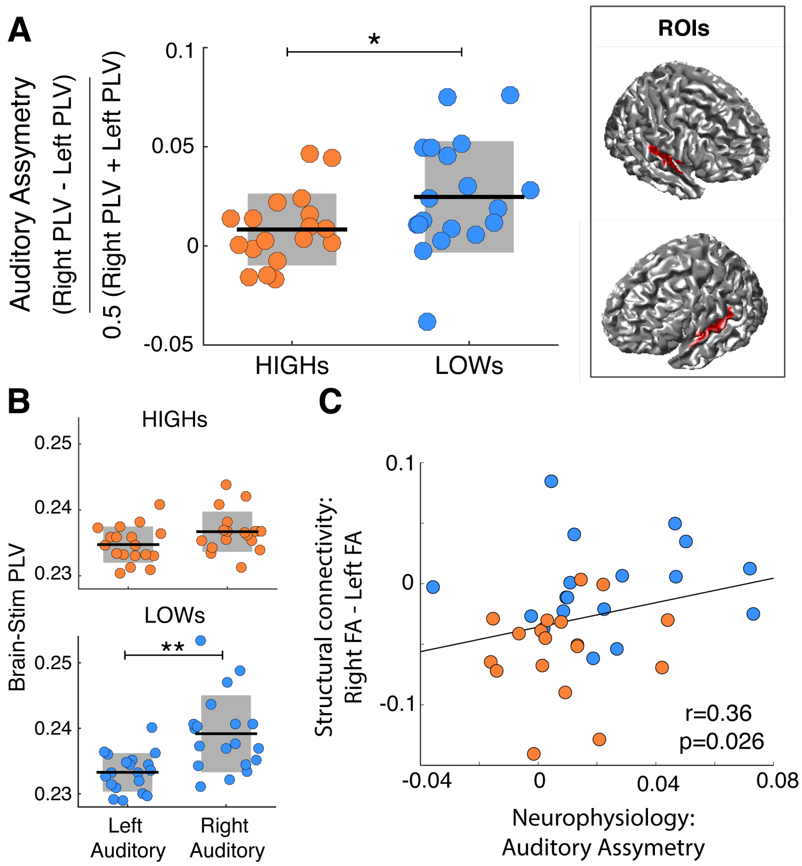

Fig. S7. **Asymmetry of auditory entrainment, neurophysiological study.** (**A**) Auditory asymmetry: comparison between groups. Asymmetry was computed as: $(PLV_{right} - PLV_{left})/0.5(PLV_{right} + PLV_{left})$. The asymmetry of auditory entrainment was significantly different between groups (r = 0.42, Rank-Biserial Correlation; Mann-Whitney-Wilcoxon test, two-sided p = 0.029). Right inset: ROIs, left and right early auditory regions. (**B**) Brain-to-stimulus synchrony in each hemisphere averaged within temporal ROIs. The data show that, while the typical rightward lateralization in tracking the speech envelope was present in low synchronizers, this was reduced in the high synchrony group (Wilcoxon signed-rank test, two-sided $p_{low} = 0.0013$ and $p_{high} = 0.089$). (**C**) Scatter plot of the correlation between structural and neurophysiological values. Mean FA laterality as a function of the auditory entrainment's asymmetry. There was a significant relationship (N = 36, Spearman r = 0.36, p = 0.026; Skipped Spearman r = 0.38, t = 2.40, CI = 0.04, 0.66) between the neurophysiological auditory asymmetry and the structural laterality of the white matter cluster (see Fig. 3 for the cluster) that differentiates between groups. While the structural leftwards laterality and the reduced rightward asymmetry shown by the high synchronizers might seem counterintuitive, in both cases high synchronizers show a more leftwards pattern of results *as compared to low synchronizers* (the correlation between structural laterality and auditory asymmetry is positive). Orange/light blue correspond to high/low synchronizers respectively. ** p < 0.005 (Wilcoxon signed-rank test), * p < 0.05 (Mann-Whitney-Wilcoxon test). Dots: individual participants. Black lines: mean across participants. Shadowed region: SD.

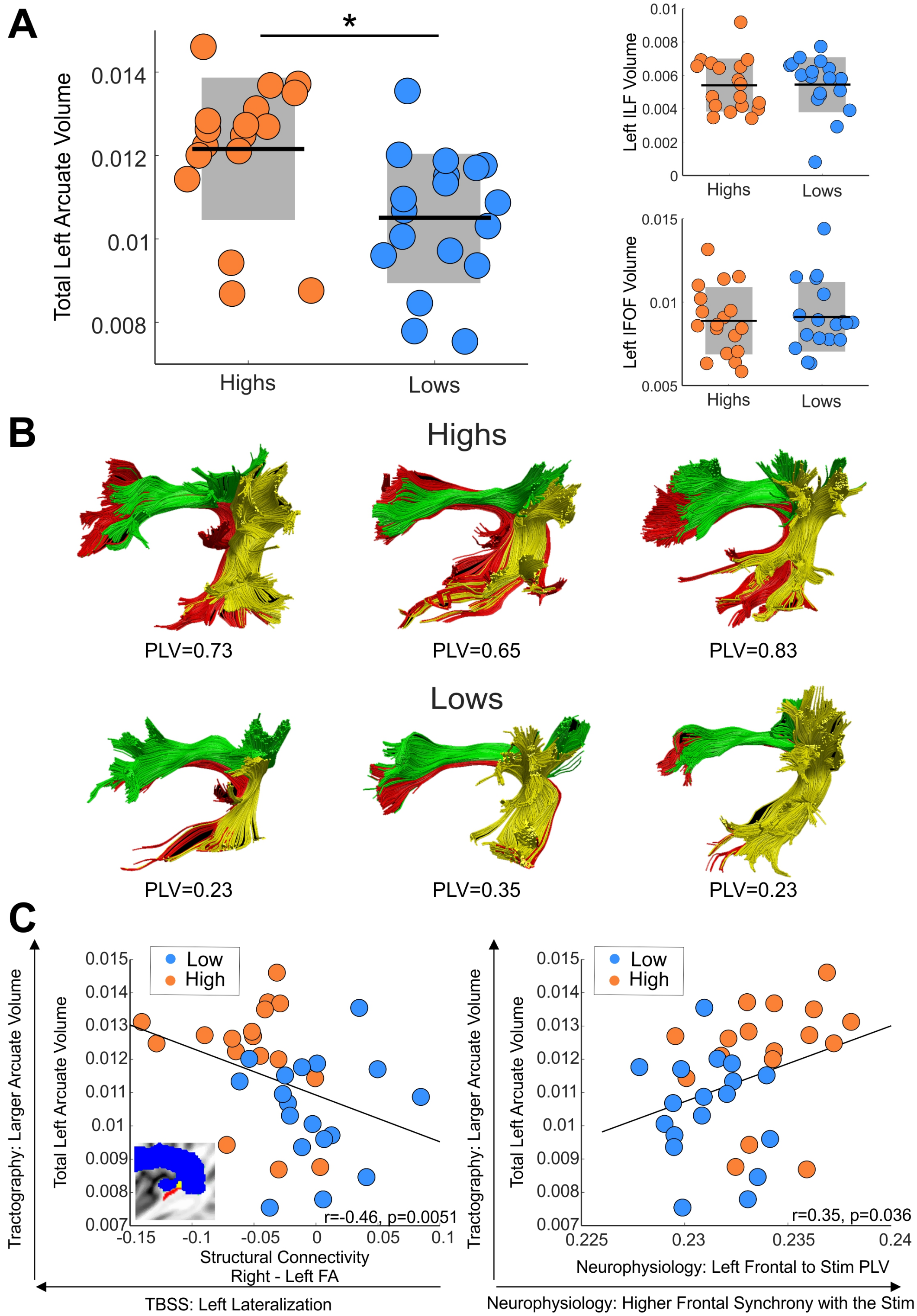

Fig S8. Tractography results. (A) Box-plots showing the mean (center line) and SD (grey areas) volumes (corrected for TIV, see tractography methods) for the total left arcuate (sum of the anterior, posterior and long segment volumes; left panel; N = 36, Wilcoxon signed-rank test, two-sided p = 0.0025; r = 0.60, Rank-Biserial correlation; FDR-corrected for multiple comparisons) and the two control tracts (right panel: left IFOF, two-sided p = 0.76, r = 0.06, Rank-Biserial correlation, and left ILF, two-sided p = 0.57, r = 0.11, Rank-Biserial correlation). Even though we cannot specify which exact segment of the arcuate is responsible for these differences (the volume of each of the three segments of the arcuate on their own did not differentiate between high and low synchronizers: long segment, $p = 0.21$, $r = 0.24$; anterior segment, $p = 1$, $r = 0$; posterior segment, $p = 0.079$, $r = 0.34$, Rank-Biserial correlation), the results do show that the dorsal pathway for language processing—connecting temporo-parietal regions with premotor areas and the inferior frontal cortex—is structurally enhanced in high synchronizers as compared to low. **(B)** Arcuate dissections (long segment in red, anterior in green and posterior in yellow) for representative high (top) and low (bottom) synchronizers. The arcuate dissections for the remaining 30 individuals showed a similar result. The behavioral PLV obtained using the SSS-Test is also shown. **(C)**. Scatter plots display the correlation (N = 36) between the total volume of the left arcuate and the FA laterality values of the TBSS cluster (see Fig 3 in the main manuscript; negative values imply a leftwards structural lateralization; left panel, an overlap-in yellow-between the TBSS cluster-in red- and a probabilistic atlas of the left arcuate fasciculus-in blue[63] is also shown) and also the synchrony of the left inferior/middle frontal gyri with the speech syllable rate (right panel). In other words, the larger the volume of the virtually dissected left arcuate, the more the TBSS FA cluster was lateralized to the left (see Fig. 3 in the main manuscript; Spearman's $r = -0.46$, $p = 0.0051$; Skipped Spearman $r = -0.43$, $t = -2.80$, CI = -0.10, -0.72); and the higher the brain-to-stimuli synchrony in frontal regions was (i.e., neurophysiology; Spearman´s $r = 0.35$, $p = 0.036$; Skipped Spearman $r = 0.35$, $t = 2.18$, CI = 0.02, 0.63). In contrast, the volumes of the left IFOF and left ILF did not differentiate between groups and were not correlated with the FA TBSS cluster (IFOF: Spearman´s $r = -0.12$, $p = 0.45$; Skipped Spearman $r = -0.09$, $t = -0.52$, CI = -0.42, 0.28; ILF: Spearman´s $r = -0.18$, $p = 0.28$; Skipped Spearman $r = -0.19$, $t = -1.17$, CI = -0.51, 0.15) and the frontal neurophysiological results (IFOF: Spearman´s $r = -0.006$, $p = 0.97$; Skipped Spearman $r = 0.07$, $t = 0.46$, CI = -0.29, 0.42; ILF: Spearman´s $r = -0.10$, $p = 0.52$; Skipped Spearman $r = -0.22$, $t = -1.35$, CI = -0.55, 0.13). No significant differences were found for FA or RD measures (all ps > 0.11). Orange/blue, high/low synchronizers. Dots: individual participants. Black lines: mean across participants. Shadowed region: SD.

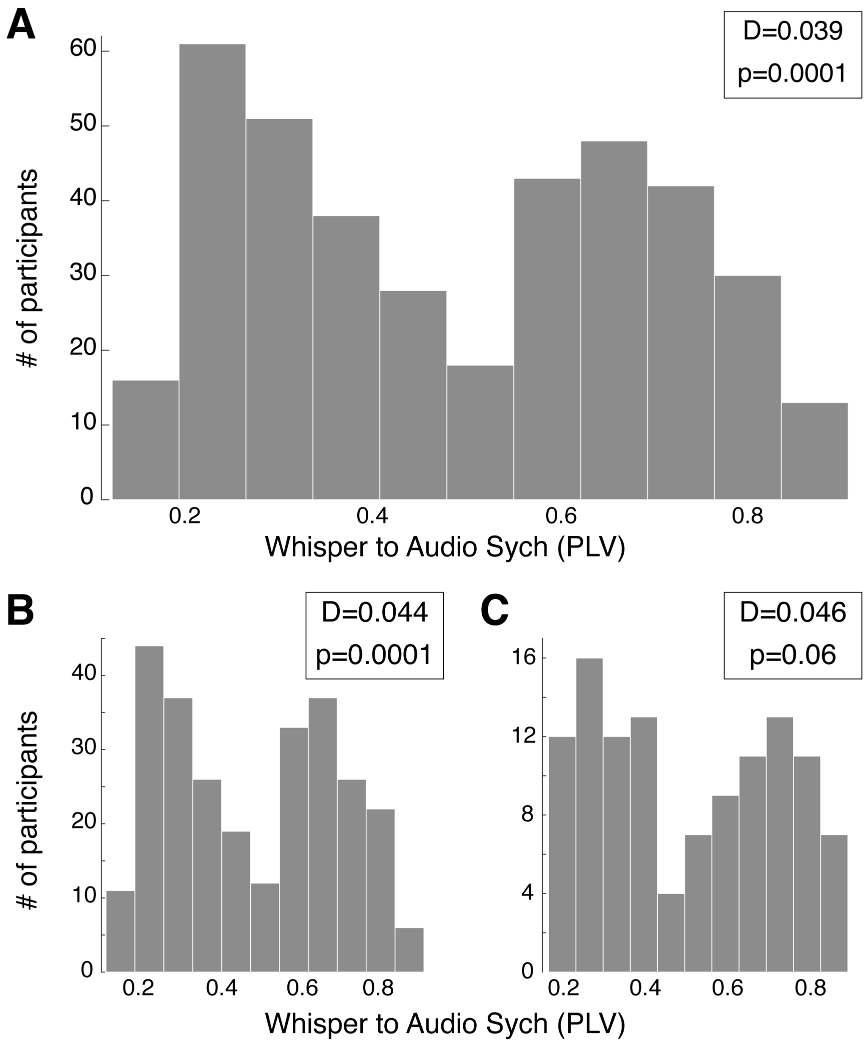

Fig. S9. SSS-test, joint bimodal distribution. For each joint distribution we calculated Hartigans' dip test statistic for testing unimodality (Hartigan, J.A., Hartigan, P.M. The Dip test of Unimodality. *Annals of Statistics*, 13, 70-84, 1985) using R (R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. 2013) and the *diptest* package (https://cran.r-project.org/web/packages/diptest/index.html). Although Hartigans´s dip test does not directly test for bimodality, it allows to reject the null hypothesis that the tested distribution is unimodal. The figure shows the histogram of the PLVs between the envelope of the perceived and produced speech signals, bandpass filtered at 3.5-5.5 Hz for **(A)** all the SSS-test experiments pooled together (N=388; stable rate in-lab, stable rate in-lab replication, stable rate online version, accelerated rate in-lab and accelerated online version); **(B)** all the SSS-test experiments with a stable rate pooled together (N=273; stable rate in-lab, stable rate in-lab replication and stable rate online version) and **(C)** all the SSS-test experiments with an accelerated syllable rate pooled together (N=155 accelerated rate in-lab and accelerated online version). D, dip tests statistic.