nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-019-0353-z

Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning

M. Florencia Assaneo[1,7*], Pablo Ripollés[1,7], Joan Orpella[2,3,4,7], Wy Ming Lin[1], Ruth de Diego-Balaguer[2,3,4,5,8] and David Poeppel[1,6,8]

We introduce a deceptively simple behavioral task that robustly identifies two qualitatively different groups within the general population. When presented with an isochronous train of random syllables, some listeners are compelled to align their own concurrent syllable production with the perceived rate, whereas others remain impervious to the external rhythm. Using both neurophysiological and structural imaging approaches, we show group differences with clear consequences for speech processing and language learning. When listening passively to speech, high synchronizers show increased brain-to-stimulus synchronization over frontal areas, and this localized pattern correlates with precise microstructural differences in the white matter pathways connecting frontal to auditory regions. Finally, the data expose a mechanism that underpins performance on an ecologically relevant word-learning task. We suggest that this task will help to better understand and characterize individual performance in speech processing and language learning.

The ability to synchronize a motor output to an auditory input—for example, tapping or dancing to music with a groove—is a basic trait present in humans from birth[1], with important cognitive implications[2,3]. From a phylogenetic perspective, spontaneous synchronization (without explicit training) to an external rhythm is argued to be a unique characteristic of vocal-learning species, including humans[4]. Study of this distinctive attribute has typically focused on how body movements are entrained by non-speech signals—for example, music[5] or a beat[6]. However, in this context, one foundational question has not been systematically investigated: do humans spontaneously align their speech motor output to auditory speech input? Resolving the role of audio–motor synchronization[7–9] in the context of speech processing is a critical step for characterization of the complex functional and structural neural architecture of language.

To address these questions, we designed a simple behavioral protocol to explore the spontaneous synchronization of speech (SSS test). The results demonstrate, in contrast to the previous literature, that a substantial part of the population does not show speech-to-speech synchronization. Thus, we further explored the functional and structural brain correlates associated with absence of synchrony. Finally, we turned to the key issue of whether the behavioral findings on audio–motor synchronization and their neural substrate extend to tasks related to more typical questions regarding speech processing and language learning.

Results

Spontaneous speech synchronization reveals a bimodal distribution. Participants ($N=84$) completed two behavioral blocks, each lasting 1 min, of listening to a rhythmic train of syllables at 4.5 syllables/s—the characteristic speech rate across languages[10,11]—while concurrently whispering the syllable 'tah' (Fig. 1a). At the end of each block, participants indicated whether a given target syllable was presented. Crucially, participants were instructed to correctly recall the syllables; there was no explicit instruction to synchronize utterances to the external rate. We first examined the degree of synchronization between the produced utterances and the input signal by computing the phase-locking value (PLV) for their envelopes around the given syllable rate (4.5 Hz; Fig. 1a). Surprisingly, participants' PLVs yielded a bimodal distribution (Fig. 1b), suggesting segregation of our cohort into distinct populations of high and low synchronizers ($N_{high}=43$, $N_{low}=41$). No difference was found between the groups in terms of language background, age or gender. However, high synchronizers showed, overall, more years of musical training than low synchronizers (Mann–Whitney–Wilcoxon test, two-sided $P=0.0033$; Supplementary Fig. 1). Still, musical experience by itself did not segregate the population into two groups (Supplementary Fig. 1f).

Next, we analyzed the distribution of phase differences between perceived and produced syllables for the high synchronizers. We found a nonuniform distribution ($N_{high}=43$; Rayleigh test, two-sided $P<0.001$; Fig. 1c) with the phase lags concentrated around 0 (95% confidence interval (CI) = −0.21, 0.26): high synchronizers adjusted their production to be in phase with the perceived syllables. The average spectra of the produced speech envelopes also exhibited striking differences between the groups. High synchronizers showed a pronounced peak at the given syllable rate, indicating high stability in maintaining the rhythm. The low-synchrony group was less stable, exhibiting a broader peak (Fig. 1d). Two additional experiments were conducted to further assess empirically whether synchrony of the produced utterances was indeed driven by an interaction between the perceived and produced speech rhythms (rather than being related to ability to maintain a tempo). First, a subset of the participants ($N_{high}=13$, $N_{low}=12$) completed an additional block of whispering 'tah' while listening to white noise (no-rhythm condition). Speech rhythmicity was strongly reduced

[1]Department of Psychology, New York University, New York, NY, USA. [2]Cognition and Brain Plasticity Unit, IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain. [3]Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain. [4]Institute of Neuroscience, University of Barcelona, Barcelona, Spain. [5]ICREA, Barcelona, Spain. [6]Neuroscience Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany. [7]These authors contributed equally: M. Florencia Assaneo, Pablo Ripolles, Joan Orpella. [8]These authors jointly supervised this work: Ruth de Diego-Balaguer, David Poeppel. *e-mail: fassaneo@gmail.com

in high synchronizers in the no-rhythm condition, relative to the rhythmic one, and remained unchanged in low synchronizers (Fig. 1e). Second, a new cohort of participants completed a modified version of the SSS test in which the rate of the perceived syllables was gradually increased (see Methods). Notably, the resulting distribution also displayed two peaks under this condition ($N=55$; Supplementary Fig. 2). High synchronizers adjusted their speech output to (multiple) changes in the tempo of the perceived speech (Supplementary Fig. 2b). This result provides additional compelling evidence that a subgroup of the population can adapt a produced rhythm to a perceived one. Furthermore, in most cases, participants were not aware of the shift in rate (Supplementary Fig. 2c), highlighting the unconscious and automatic nature of the phenomenon. These combined findings support the conjecture that participants exhibited two qualitatively different behaviors: whereas the speech output of high synchronizers was entrained by the external auditory speech rate, low synchronizers showed no interaction between the produced and perceived speech rhythms.

We next determined that subjects' degree of audio–motor synchrony was stable over time, suggesting that synchronization type is a consistent individual trait. This was demonstrated by the highly correlated PLVs, both across blocks within a session ($N=84$; Spearman correlation, $r=0.86$, $P<0.001$; Fig. 1f) and across sessions distant in time (34 participants repeated the test 1 month later; Spearman correlation, $r=0.78$, $P<0.001$; Fig. 1g). This stability over time proves the SSS test's reliability in clustering participants into high- and low-synchrony groups via a straightforward measure of phase locking to a regular pacing signal.

To further verify that the phenomenon is robust and replicable, we developed and conducted an online version of the SSS test for both the stable and accelerated syllable rates using the Amazon Mechanical Turk (AMT) platform. These two additional replications ($N=144$ for the stable version and $N=60$ for the accelerated version) underscore the reliability of the bimodal distribution under less controlled conditions (see Supplementary Fig. 3a for results with the stable rate and Supplementary Fig. 4a for those with the accelerated rate) and also allowed us to explore differences between groups in perception and production abilities. From a perceptual standpoint, the high synchronizers who completed the online version of the stable SSS test ($N=144$) were marginally better than the low synchronizers (Mann–Whitney–Wilcoxon test, two-sided

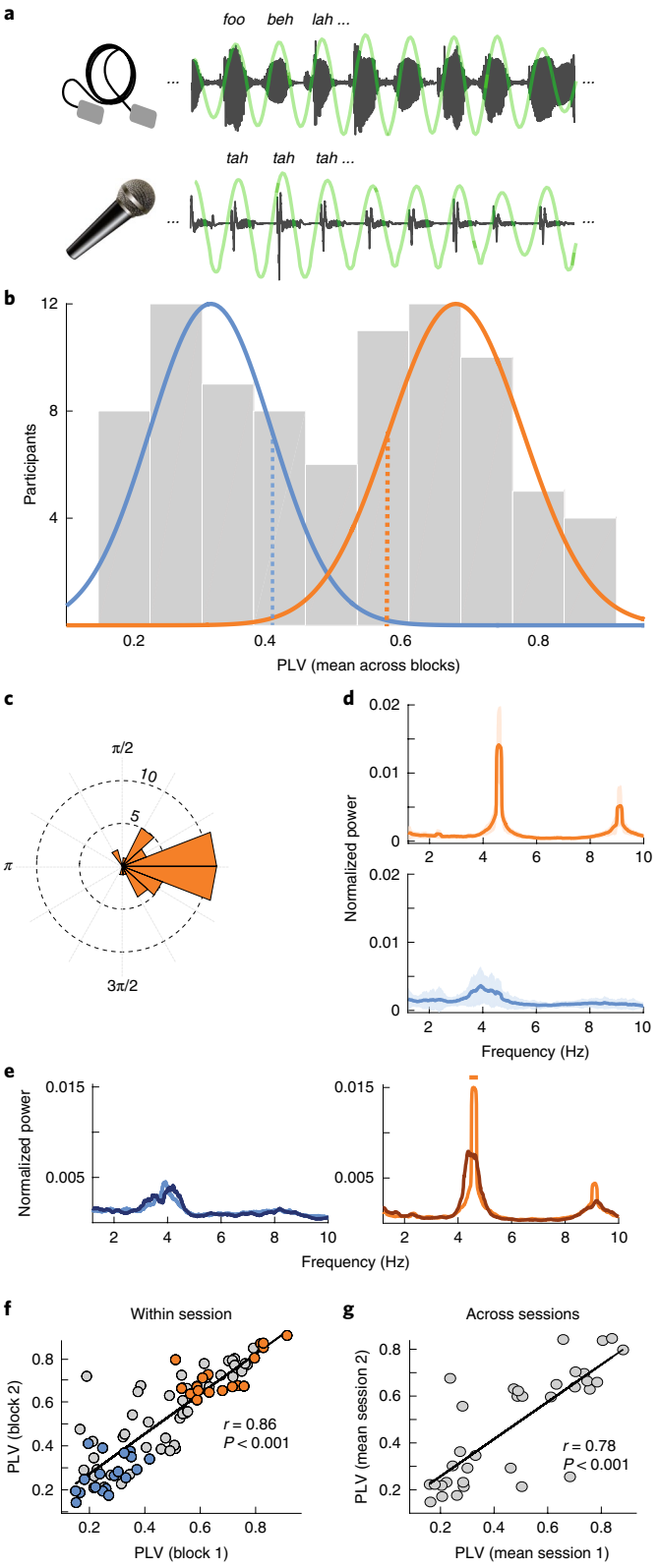

Fig. 1 | Spontaneous speech synchronization reveals a bimodal distribution. a, SSS test: example of the perceived (top) and produced (bottom) signals. Produced signals were independently recorded for each participant ($N=84$). Green lines represent the envelope, bandpass filtered from 3.5–5.5 Hz. To eliminate auditory interference resulting from listeners' own speech output, participants wore foam earplugs and whispered softly. **b**, PLV histogram (average across blocks). Colored lines represent normal distributions fitted to each of the two clusters obtained by a *k*-means algorithm (the number of participants in each cluster was $N_{high}=43$ (orange) and $N_{low}=41$ (blue)). Participants subsequently completing neurophysiology and neuroimaging sessions were randomly selected from 1σ below or above the mean (indicated by the blue and orange dashed lines). **c**, Phase histogram for the lag between the perceived and produced syllables. The histogram was computed only for the high-synchrony group; low-synchrony participants were not synchronized, and it was thus not possible to define a phase lag. **d**, Average spectra of the envelopes of utterances ($N_{high}=43$, $N_{low}=41$). Shaded regions represent s.d. **e**, Average spectra for a subgroup of participants ($N_{high}=13$, $N_{low}=12$). Dark and light lines correspond to the no-rhythm and rhythm conditions, respectively. A bar on top indicates a significant difference between the conditions (Wilcoxon signed-rank test, two-sided $P<0.05$, FDR corrected). **f**, PLV scatterplot of correlation between the first and second blocks (Spearman correlation, $r=0.86$, $P<0.001$). Dots represent individual subjects ($N=84$), and colored dots represent participants selected to complete subsequent neurophysiology and neuroimaging sessions ($N_{high}=18$, $N_{low}=19$). **g**, Scatterplot of mean PLVs, showing the correlation between the first and second sessions (conducted 1 month apart; Spearman correlation, $r=0.78$, $P<0.001$). In all panels, orange and blue correspond to high and low synchronizers, respectively.

$P=0.093$; Supplementary Fig. 3b). From a production standpoint, the low synchronizers were marginally less accurate (with no correction for multiple comparisons) in keeping a precise syllable rate (4.5 syllables/s) without feedback (on white noise): they tended to produce lower syllable rates (Supplementary Fig. 3c). Although these results imply that the two populations—identified by our 1-min test relying on speech audio–motor interactions—also differ, however modestly, at other levels of perception or production ability, further work is needed to delineate the extent of the predictive power of our test with regard to other capabilities.

In light of the reliability of our findings, we next pursued the following question: does the clear grouping based on the straightforward behavioral paradigm reflect neural properties and behavioral consequences that have broader relevance? These behavioral data invite the hypothesis that functional and/or structural brain differences underlie the segregation into high versus low synchronizers. To address this question, we next acquired both neurophysiological and structural data from a subgroup from the original cohort ($N_{high}=18$, $N_{low}=19$; Fig. 1f).

Neural distinction between groups: neurophysiological data. In a magnetoencephalography (MEG) experiment, participants listened to rhythmic trains of syllables (4.5 syllables/s), now passively (without whispering). They were instructed to listen attentively to the syllables and to indicate after each stream whether a given set of syllables had been presented. Task performace was above chance level, veryfing subject's attention to the syllables ($N=37$; Wilcoxon signed-rank test, two-sided $P=0.011$; Supplementary Fig. 5). However, there was not a significant difference in performance between the groups ($N_{high}=18$, $N_{low}=19$; Mann–Whitney–Wilcoxon test, two-sided $P=0.96$). Caution is required in interpreting this behavioral result, however, because syllable recognition was rather poor for the entire participant cohort. This derives from the fact that we designed an extremely difficult task (12 synthesized syllables, co-articulated for 2 min) to maximize participants' attention during the 2-min syllable perception task. We then computed the PLV between elicited brain activity and the envelope of the auditory stimuli (brain-to-stimulus synchrony; Fig. 2a) in the frequency band corresponding to the perceived syllable rate (4.5 ± 0.5 Hz). Given that group segregation relied on a speech audio–motor task, we centered our analyses in bilateral frontal and temporal regions implicated in speech production and perception, respectively. A first comparison between the groups restricted to frontal regions of interest (ROIs) revealed that high synchronizers showed enhanced brain-to-stimulus synchrony in left inferior and middle frontal gyri, more precisely, in left Brodmann areas 44, 45, 46 and 9 ($N_{high}=18$, $N_{low}=19$; Mann–Whitney–Wilcoxon test, two-sided $P<0.05$, false-discovery rate (FDR) corrected; Fig. 2b and Supplementary Fig. 6a). Interestingly, previous data showed that, during overt speech production, control of temporal speech patterns is likely governed by these regions[12,13]. Thus, our results suggest that areas related to speech timing during production are also implicated during speech perception to track the perceived syllable rate (note that no motor production occurred during the MEG session).

In contrast, the same analysis performed on temporal ROIs yielded no significant differences between the groups (Supplementary Fig. 6). However, the asymmetry of the entrainment in early auditory regions was significantly different between the groups (see Methods and Supplementary Fig. 7).

Structural distinction between groups: anatomical connectivity data. Having observed neurophysiological differences between the groups, we then acquired diffusion-weighted MRI (DW-MRI) data from the same cohort to quantify potential differences in the white matter pathways connecting the frontal and auditory regions that might distinguish the groups in terms of brain-to-stimulus synchrony. Excitingly, high synchronizers showed enhanced microstructural properties in the white matter neighboring the auditory cortex (Fig. 3a). Specifically, we found a distinct lateralization pattern in a white matter cluster ($N_{high}=18$, $N_{low}=18$; family-wise error (FWE)-corrected at the peak voxel, two-sided $P=0.024$; Fig. 3b), likely part of the arcuate fasciculus[14–16], that differentiated the groups, with high synchronizers showing significantly greater left lateralization. No significant clusters were obtained in ventral white matter pathways connecting frontal and temporo-occipital regions (see Methods). Notably, this structural difference relates to both the auditory (Supplementary Fig. 7b) and frontal (Fig. 2b) neurophysiological results: increased leftward lateralization in the white matter was related to higher brain-to-stimulus synchrony in left frontal regions (Fig. 3c) and to more symmetrical auditory entrainment (Supplementary Fig. 7c). Virtual dissections (tractography) further showed that the volume of the left arcuate (but not of the left inferior longitudinal or the inferior fronto-occipital fasciculi, which were also dissected as a control) not only differentiated between the groups but was also related to left frontal neurophysiological brain-to-stimulus synchrony (see Methods and Supplementary Fig. 8).

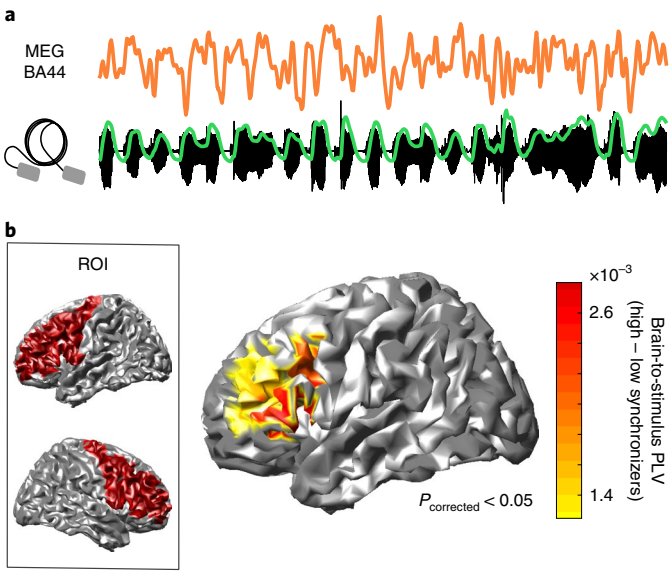

Fig. 2 | Neural distinction between groups: neurophysiological data. a, Activity from a high synchronizer, generated in Brodmann area 44 (BA44) (top) during passive listening to the stimulus (bottom; the stimulus envelope is shown in green). Similar signals were obtained for the other high synchronizers ($N=17$). **b**, Brain-to-stimulus synchronization. Left: ROI comprising bilateral precentral, middle frontal and inferior frontal gyri. Right: brain surface map showing PLV differences between the groups ($N_{high}=18$, $N_{low}=19$; Mann–Whitney–Wilcoxon test, two-sided $P<0.05$, FDR corrected).

Spontaneous speech synchronization test predicts word learning. Previous research documents that the early stages of word learning capitalize on interaction between auditory and frontal regions and the white matter pathways connecting them[15]. To test for a principled link between these observations and our simple behavioral test, a new sample of participants ($N=44$) was recruited to complete the SSS test as well as a word-form-learning task. More precisely, because we expected that speech synchronization should most clearly benefit segmentation abilities, a classical statistical learning paradigm was chosen[17,18]. In this paradigm, participants listened for 2 min to a continuous repetition of four trisyllabic pseudo-words, which were randomly concatenated without silence gaps between them. Next, they completed a testing phase that assessed whether they correctly

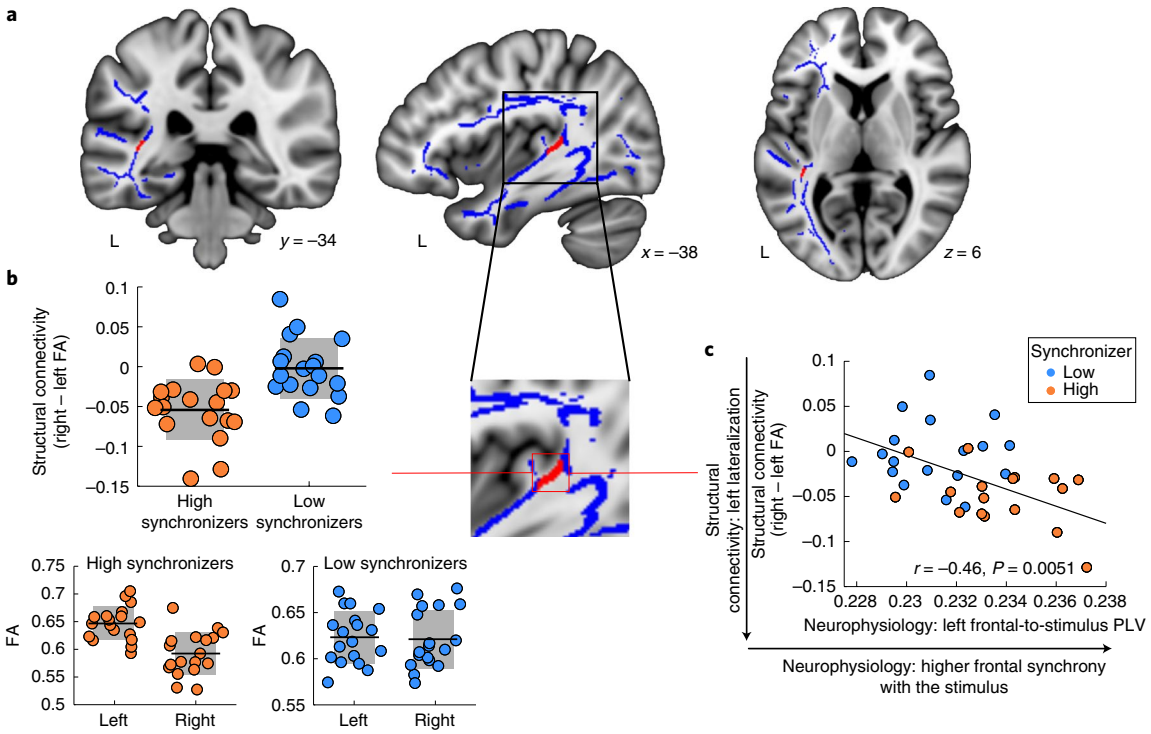

Fig. 3 | Structural distinction between groups: anatomical connectivity data. a, Laterality maps (generated using tract-based spatial statistics, TBSS) of fractional anisotropy (FA), showing right minus left values. White matter cluster (red) differentiating the groups ($N_{high}=18$, $N_{low}=18$; family-wise error (FWE) corrected, two-sided $P<0.05$, threshold-free cluster enhancement) is shown over the mean group skeleton (blue). Neurological convention is used, with MNI coordinates at the bottom of each slice. **b**, To facilitate visualization of the pattern of results, box plots ($N_{high}=18$, $N_{low}=18$) with the mean (center line) ± s.d. (gray area) FA values of the significant cluster are shown for laterality (right - left; top) and for each group and hemisphere separately (bottom). **c**, Scatterplot ($N=36$ participants) showing the correlation between mean FA laterality values (negative values imply leftward structural lateralization) and synchrony of the left inferior/middle frontal gyri with the speech syllable rate (Spearman correlation, $r=-0.46$, $P=0.0051$; skipped Spearman correlation, $r=-0.44$, $t=-2.90$, CI = −0.14, −0.68). In **b** and **c**, dots represent individual participants.

segmented the pseudo-words. The histogram of the PLVs obtained with the SSS test for this smaller group also displayed two peaks, replicating the bimodal distribution of the original cohort (Fig. 4a). When splitting this new population into high and low synchronizers (using the median PLV of the first cohort; Fig. 4a), we found that high synchronizers had a significant learning advantage in the phonological word-learning task ($N_{high}=24$, $N_{low}=20$; $r=0.4$, rank-biserial correlation; Mann–Whitney–Wilcoxon test, two-sided $P=0.024$; Fig. 4b). This learning benefit was also replicated in the additional cohort of participants who completed an online version of the word-learning task in addition to the online and accelerated version of the SSS test ($N_{high}=25$, $N_{low}=35$; $r=0.37$, rank-biserial correlation; Mann–Whitney–Wilcoxon test, two-sided $P=0.015$; Supplementary Fig. 4b). We hypothesize that (i) in line with previous work[8,19] the increment of synchronization in the frontal region, enhanced in high synchronizers, facilitates parsing of syllables by aligning attention to their onset; (ii) better parsing improves extraction of the statistical relationship between syllables; and, (iii) prediction of one syllable following another likewise helps to create a better phonological trace of whole words. In conclusion, enhanced audio–motor interaction as measured with our approach (the SSS test) not only is reflected in the functional and structural properties of frontal and temporal areas but also has compelling consequences for language learning.

Discussion

The combined behavioral, neurophysiological and neuroanatomical results reveal a fundamental phenomenon: whereas some individuals are compelled to spontaneously align their speech output to speech input, others remain impervious to external rhythm (see Supplementary Fig. 9 for a depiction of the joint bimodal distribution of all our experiments using the SSS test). We speculate that such distinct populations of high and low synchronizers emerge from the spontaneous nature of the synchrony induced by the SSS test (in which the goal of the task is orthogonal to synchronization). This contrasts with previous research showing more homogenous entrainment patterns when synchronization to an external auditory signal is explicitly requested[2,20]. The behavioral pattern we have discovered correlates with neurophysiological and structural features within key elements of the speech brain network[21–23], including production areas (inferior frontal gyrus), perception areas (early auditory cortex) and the white matter connecting them (see the tractography analysis in Supplementary Fig. 8)[24]. Excitingly, the fact that our results scale up to an ecologically relevant task[18], word learning in the context of speech segmentation, has theoretical and practical implications for how individual differences in cognition and learning are understood and studied[25,26].

Our ability to speak relies on a widely distributed and highly interconnected audio–motor network[27,28]. We hypothesize that an interplay between structural and physiological predispositions (roughly, nature) and experience-dependent tuning (roughly, nurture) can generate moderate modifications to the components of the speech audio–motor network that, owing to the coarseness of its connections[29], result in large consequences at the functional and behavioral levels. Specifically, a subtle enhancement in the structure of the white matter connecting auditory and motor regions could improve the synchronization (flow of information[30]) between temporal and frontal areas, in turn eliciting the effects

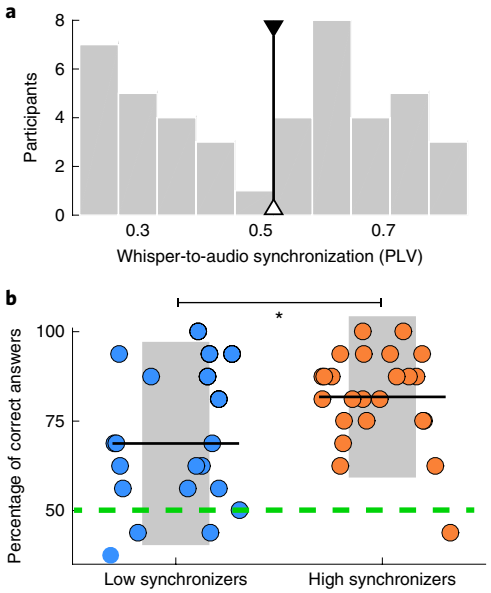

Fig. 4 | Spontaneous speech synchronization test predicts word learning. **a**, SSS test outcome. Histogram of PLVs between the envelopes of the perceived and produced speech signals, bandpass filtered at 3.5–5.5 Hz. The median of the first cohort's distribution is indicated (black line; individuals above and below this line were labeled as high and low synchronizers, respectively). **b**, Percentage of correct answers for the statistical word-learning task ($N_{high}=24$, $N_{low}=20$; Mann–Whitney–Wilcoxon test, two-sided $P=0.024$). Black lines represent the mean across participants and shaded regions represent s.d.; dots correspond to individual participants. The green dashed line indicates chance level. *$P<0.05$.

observed in these experiments. Previous research has shown that white matter located in the same region as the cluster highlighted by our study undergoes microstructural changes through musical training[9,31]. In line with these studies, we found that high synchronizers had, overall, more years of musical training than low synchronizers. However, in our work, musical training on its own did not follow a bimodal distribution, suggesting that musical experience is one of many factors defining group membership as a high or low synchronizer.

In summary, we introduce a deceptively simple behavioral task (the SSS test) capitalizing on individual differences that turn out to be predictive of audio–motor synchronization, neurophysiological function, brain anatomy and performance on an ecologically relevant word-learning task. Use of such a test can help to better characterize individual performance, leading to new discoveries related to speech processing and language learning that could previously have been masked by pooling populations with substantially different neural and behavioral attributes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-019-0353-z.

Received: 27 April 2018; Accepted: 1 February 2019;
Published online: 4 March 2019

References

1. Condon, W. S. & Sander, L. W. Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science* **183**, 99–101 (1974).
2. Repp, B. H. Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* **12**, 969–992 (2005).
3. Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L. & Kraus, N. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proc. Natl Acad. Sci. USA* **111**, 14559–14564 (2014).
4. Patel, A. D. The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol.* **12**, e1001821 (2014).
5. Janata, P. & Grafton, S. T. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* **6**, 682–687 (2003).
6. Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M. & Fitch, W. T. Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140093 (2015).
7. Assaneo, M. F. & Poeppel, D. The coupling between auditory and motor cortices is rate-restricted: evidence for an intrinsic speech–motor rhythm. *Sci. Adv.* **4**, o3842 (2018).
8. Park, H., Ince, R. A. A., Schyns, P. G., Thut, G. & Gross, J. Frontal top–down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **25**, 1649–1653 (2015).
9. Steele, C. J., Bailey, J. A., Zatorre, R. J. & Penhune, V. B. Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. *J. Neurosci.* **33**, 1282–1290 (2013).
10. Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J. & Lorenzi, C. A cross-linguistic study of speech modulation spectra. *J. Acoust. Soc. Am.* **142**, 1976–1989 (2017).
11. Ding, N. et al. Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* **81**(Pt. B), 181–187 (2017).
12. Long, M. A. et al. Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
13. Magrassi, L., Aromataris, G., Cabrini, A., Annovazzi-Lodi, V. & Moro, A. Sound representation in higher language areas during language generation. *Proc. Natl Acad. Sci. USA* **112**, 1868–1873 (2015).
14. Ripollés, P. et al. Strength of temporal white matter pathways predicts semantic learning. *J. Neurosci.* **37**, 11101–11113 (2017).
15. López-Barroso, D. et al. Word learning is mediated by the left arcuate fasciculus. *Proc. Natl Acad. Sci. USA* **110**, 13168–13173 (2013).
16. Thiebaut de Schotten, M. et al. Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *Neuroimage* **54**, 49–59 (2011).
17. Lopez-Barroso, D. et al. Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cereb. Cortex* **21**, 2742–2750 (2011).
18. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
19. Morillon, B. & Baillet, S. Motor origin of temporal predictions in auditory attention. *Proc. Natl Acad. Sci. USA* **114**, E8913–E8921 (2017).
20. Cummins, F. Rhythm as entrainment: the case of synchronous speech. *J. Phonetics* **37**, 16–28 (2009).
21. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
22. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
23. Pulvermüller, F. & Fadiga, L. Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* **11**, 351–360 (2010).
24. Catani, M., Jones, D. K. & Ffytche, D. H. Perisylvian language networks of the human brain. *Ann. Neurol.* **57**, 8–16 (2005).
25. Zatorre, R. J., Fields, R. D. & Johansen-Berg, H. Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat. Neurosci.* **15**, 528–536 (2012).
26. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D. Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490 (2017).
27. Hage, S. R. & Nieder, A. Dual neural network model for the evolution of speech and language. *Trends Neurosci.* **39**, 813–829 (2016).
28. Guenther, F. H. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychol. Rev.* **102**, 594–621 (1995).
29. Turken, A. U. & Dronkers, N. F. The neural architecture of the language comprehension network: converging evidence from lesion and connectivity analyses. *Front. Syst. Neurosci.* **5**, 1 (2011).
30. Fries, P. Rhythms for cognition: communication through coherence. *Neuron* **88**, 220–235 (2015).
31. Bengtsson, S. L. et al. Extensive piano practicing has regionally specific effects on white matter development. *Nat. Neurosci.* **8**, 1148–1150 (2005).

Acknowledgements
We thank J. Rowland and J.-R. King for comments and advice. This work was supported by NIH grant 2R01DC05660 (D.P.) and FP7 Ideas: European Research Council grant ERC-StG-313841 (R.d.D.-B.).

Author contributions
M.F.A., P.R., J.O., W.M.L., R.d.D.-B. and D.P. designed the research and wrote the manuscript; M.F.A., P.R., J.O. and W.M.L. acquired and analyzed the data.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0353-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.F.A.

Journal peer review information *Nature Neuroscience* thanks Sylvain Baillet, Narly Golestani and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Participants. A first cohort of 84 participants initially completed the SSS test (32 males; mean age, 28 years; age range, 19 to 55 years). From this, a subgroup of 37 subjects (right handed; 18 males; mean age, 30 years; age range, 21 to 55 years) also underwent the MEG and DW-MRI protocols. The original subgroup comprised four additional participants, but these had to be removed owing to artifactual MEG (three participants) or DW-MRI (one participant) data. The MEG session took place at least 4 d after the DW-MRI session. Both protocols were completed within 1 month of the SSS test.

A second cohort of 44 individuals (11 males; mean age, 21 years; age range, 19 to 31 years) completed the SSS test and the word-learning task.

A third cohort of 62 participants completed the accelerated version of the SSS test. Seven participants were removed because they spoke loudly instead of whispering or they stopped whispering for time periods longer than 4 s. The data from 55 participants (19 males; mean age, 23 years; age range, 18 to 36 years) were analyzed.

Two additional cohorts, one of 200 participants and one of 100 participants, completed the online version of the regular and accelerated SSS tests, respectively. 56 participants from the regular group and 40 participants from the accelerated group were removed for non-optimal conditions in their recordings (noisy recording, participant did not use headphones, participant spoke loudly instead of whispering or stopped whispering for time periods longer than 4 s). The final number of participants who submitted to the analyses was 144 (80 males; mean age, 34 years; age range, 19 to 55 years) for the regular SSS test and 60 (37 males; mean age, 35 years; age range, 19 to 51 years) for the accelerated SSS test.

All participants were native English speakers with self-reported normal hearing and no neurological deficits. They were paid for taking part in the study and provided written informed consent. All protocols were approved by the local institutional review board (New York University's Committee on Activities Involving Human Subjects).

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications[7,15,17].

Statistical analyses. Data distribution was not formally tested. Instead, nonparametric Mann–Whitney–Wilcoxon and Wilcoxon signed-rank tests were used for between- and within-subject comparisons, respectively. We controlled for multiple comparisons by using FDR correction (the only exception was TBSS white matter analyses, which used an FWE correction based on threshold-free cluster enhancement and a nonparametric permutation test). Nonparametric Spearman's rank correlations were used to assess the relationship between variables. In addition, we used the Robust Correlation Toolbox[32] to ensure the robustness of the relationship between the structural and neurophysiological data. In particular, we used Spearman skipped correlations[33,34] with percentile bootstrap 95% confidence intervals (calculated by resampling pairs of observations) for each correlation. Skipped correlations involve multivariate outlier detection and provide a more robust measure of correlation[35]. Bootstrap confidence intervals provide an additional way to test whether two variables are truly correlated; if the confidence intervals include 0, the null hypothesis cannot be rejected[32].

Effect sizes were calculated using rank-biserial correlations, which can be read as a Spearman correlation coefficient[36] or as the difference between the proportion of favorable and unfavorable evidence[37]. In our study, all effect sizes are above $r = 0.36$. An r value of 0.36 means that the favorable evidence outweighs the unfavorable evidence by 68% to 32%.

Data collection and analysis were performed with blinding to the conditions of the experiments for the behavioral tests but not for the MEG and structural analyses, as subjects had already been divided into high and low synchronizers.

Stimulus presentation order was randomized for all experiments with more than one stimulus.

Phase-locking value. Throughout the study, the synchronization between two signals was measured by the PLV between them. PLV was computed using the formula

$$\mathrm{PLV} = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i(\theta_1(t)-\theta_2(t))}\right|$$

where t is the discretized time, T is the total number of time points, and θ_1 and θ_2 are the phase of the first and second signals, respectively.

Stimuli. Five sets of syllables (G1 to G5) were created to be used in the MEG task. Each set consisted of 12 distinct syllables (unique consonant–vowel combinations) handpicked to maximize variability within and between sets. G5 was used in the SSS test, and the remaining four sets (G1 to G4) were used to create the stimuli (word streams, words and part-words) for the word-learning task.

Random syllable streams were created by randomly combining the syllables of a set, with no gap between them and with the sole constraint that the same syllable not be repeated consecutively. The total duration of the random syllable streams was 60 s for the SSS test (G5) and 120 s for the MEG task (G1 to G5).

For the word-learning task, we created four trisyllabic pseudo-words (henceforth, words) per set (G1 to G4) by unique combination of all the component syllables. This resulted in four 'languages' (henceforth, languages L1 to L4), each comprising four words. To improve learnability, we consulted the Irvine Phonotactic Online Dictionary (IPhOD version 2.0; http://www.IPhOD.com/) for minimum word-average biphoneme and positional probabilities. For the exposure phase, the four words of a language were randomly combined to form a 2-min-long stream for each language with no gaps, ensuring an equal number of non-consecutive repetitions per word. For the test phase, in addition to the words, we also created for each language all possible part-words by the combination of the final syllable of a word with the first two syllables of the remaining words (12 part-words per language). Written renderings and cross-ratings of all words and part-words were provided independently by five native speakers of American English. The written forms with the highest convergence were selected for visual presentation concurrent with the audio in the test phase of the word-learning task.

Random syllable streams, word streams, words and part-words were all converted to .wav files for auditory playback by using the MBROLA text-to-speech synthesizer with the American Male Voice diphone database (US2) at 16 kHz[38]. All phonemes were equal in pitch (200 Hz), pitch rise and fall (with the maximum at 50% of the phoneme), and duration, which was set to 111 ms to satisfy a presentation rate of 4.5 syllables/s throughout the entire study.

All auditory stimuli were presented binaurally at a mean sound pressure of 75 dB, via tubephones (E-A-RTONE 3A 50 Ω, Etymotic Research) attached to E-A-RLINK foam earplugs inserted into the ear canal.

SSS test. Participants (for the in-lab experiment in a sound isolation booth, seated in front of a PC with a microphone placed close to their mouth) completed three experimental steps:

1. Volume adjustment: subjects listened to the train of random syllables played backward while whispering 'tah' and increased the audio volume until they could not perceive their own voice;
2. Steady repetition example: an audio with a continuous repetition of the syllable 'tah' (recorded by a female speaker, manipulated to last 222 ms and concatenated to produce a rate of 4.5 syllables/s) was delivered through the earplug tubephones for 10 s. Subsequently, participants were instructed to whisper 'tah' at the same pace for 10 s. We primed the participants at the desired frequency, as previous research showed that synchronization to an external stimulation occurs when there is a close match between internal and external frequencies[39,40];
3. Syllable perception task: participants attended to the rhythmic syllable stream while steadily whispering 'tah'. After the presentation, they had to indicate whether a given set of target syllables were presented. For each run, four target syllables were randomly selected from a pool of eight (half of them were part of the stream). Importantly, participants were not explicitly instructed to synchronize to the external audio. According to the instructions, the assignment was to recall correctly the syllables and the 'tah' articulation was intended just to increase the difficulty of the task. By this, we encouraged attention to the audio while the goal of the task remained orthogonal to the synchronization (implicit synchronization). A subset of participants, randomly selected from the pool, completed an additional step at the end of the syllable perception task ($N_{high} = 13$, $N_{low} = 12$). During this step, they steadily whispered 'tah' for 1 min while listening to white noise (no-rhythm condition).

After the last step, participants filled out a questionnaire indicating age, handedness, gender, musical experience and spoken languages. Subsequently, they repeated steps 2 and 3. In this way, each participant completed two runs of the syllable perception task, which we named blocks 1 and 2.

A subgroup of participants, randomly selected from the original cohort, completed the whole experiment again 1 month after the first session ($N = 34$).

SSS test, accelerated version. The protocol for the accelerated version was the same as for the regular SSS test, but we modified the auditory stimulus as follows: we progressively increased the syllable rate from 4.3 to 4.7 syllables/s, using steps of 0.1 syllables/s; each rate was kept constant for 60 syllables with the exception of the last one, which remained constant until the end of the audio, which in this case was set to 70 s (Supplementary Fig. 2b). As in the regular version of the test, each participant completed two blocks. Once they finished, they indicated whether they perceived an increment, a decrement or no change in the rate of the presented syllables for each block. As for the stable-rate version of the SSS test, we computed the PLV between the envelopes of the produced utterances and input signal.

Online version. An online version of the SSS test (for both normal and accelerated versions of the paradigm) was developed using oTree, a Python-based framework for the development of controlled experiments on online platforms[41]. The online version mainly followed the same structure as the in-lab one (volume adjustment, steady repetition, syllable perception), but with some changes: (i) a microphone test phase was included before the volume adjustment phase and (ii) several restrictions were placed to ensure that the participants actually did the task

(for example, during steady repetition, a participant could not continue to the next page until they had heard the whole 10-s-long recording and recorded themself for another 10 s). All recordings were manually checked for errors (for example, not using headphones, speaking loudly instead of whispering, etc.). Instructions were exactly the same as those for the in-lab version.

Half of participants who completed the online version of the stable-rate SSS test also completed a rhythm perception test, while the other half undertook a rhythm production assessment. The perception task was the one described by Huss and colleagues[42]. Participants were presented with pairs of tunes (36 pairs, 18 the same and 18 different) and had to make a same/different judgment. In the different pairs, there was a mismatch in the duration of the accented note. For the production task, participants repeated the steady repetition example step and were instructed to keep whispering 'tah' at the same pace for 1 min while listening to white noise. The difference between this task and the no-rhythm condition wa s that all participants were primed again, before the 1 min of whispering under white noise, and explicitly instructed to keep the tempo. These tasks were always performed after the SSS test and were also programmed by using oTree.

Tasks developed with oTree can be deployed into AMT, a crowdsourcing platform that allows for acquisition of large datasets in a rapid, affordable, anonymous and easily replicable manner. Note that recent research has replicated a number of tasks from experimental psychology (for example, Stroop, Flanker, subliminal priming and category learning, among others) by using AMT[43]. The online SSS test was presented to AMT participants as an HTML webpage that ran in each participant's web browser. AMT participants were first presented with a summary of the task and then with an informed consent page. Upon acceptance, instructions for the task were presented.

Synchrony measurement. The degree of synchronization was measured by the PLV between the envelope of the produced speech and the cochlear envelope of the rhythmic syllable stream. The envelope was estimated as the absolute value of the Hilbert transform of the signal. Spectrograms of the auditory stimuli were computed using the NSL (Neural Systems Laboratory) Auditory Model MATLAB toolbox[44]. This toolbox filters the signal in a specific set of frequency bands, emulating the auditory filters applied by the cochlea (yielding what we call auditory channels), and computes their envelopes. The stimulus cochlear envelopes were calculated by adding the auditory channels between 180 and 7,246 Hz. Envelopes were resampled at 100 Hz, filtered between 3.5 and 5.5 Hz, and their phases were extracted by means of the Hilbert transform. The PLV was computed for windows of 5 s in length with an overlap of 2 s. The results for all time windows were averaged within each stimulus presentation, providing one PLV per block.

We computed the within- and across-session PLV Spearman rank correlation. For within-session blocks, correlation was computed between the PLV of block 1 and the PLV of block 2. For across-session computation, the mean PLV across blocks for the first session was correlated with the one obtained for the session completed 1 month later.

Definition of high and low synchronizers. We applied a k-means clustering algorithm[45] to the PLVs averaged across blocks, by using a squared Euclidean distance metric and two clusters. For each cluster (lower and higher), we fitted a normal distribution (mean μ and s.d. σ). Next, we defined low and high thresholds as $T_{low} = \mu_{lower} + \sigma_{lower}$ and $T_{high} = \mu_{higher} - \sigma_{higher}$. The groups of low and high synchronizers that underwent the MEG and DW-MRI sessions were randomly selected from the set of participants whose mean PLV across blocks was below T_{low} and above T_{high}, respectively.

Spectral analysis. We computed the discrete Fourier transform (DFT) for the envelope of the produced speech for each block without any windowing. We focused our analysis between 1 and 10 Hz. Thus, we kept the power values within this frequency window and they were normalized to sum 1. Spectra across blocks belonging to the same condition (no rhythm/rhythm) were averaged.

To assess differences between conditions (no rhythm/rhythm), within each group (high and low synchronizers), nonparametric paired Wilcoxon signed-rank tests were calculated for the power values at each frequency. Significant results are reported at FDR-corrected $P < 0.05$.

Neurophysiological study. *Task.* Once in the MEG system, participants completed five runs of the syllable perception task. Importantly, however, and in contrast to the SSS test, participants passively listened to the syllable streams in silence (without whispering). Each run involved a different random syllable stream corresponding to syllable sets G1 to G5. The order of presentation was counterbalanced across participants. Random syllable streams for the MEG experiment were 120 s in duration. Experimental runs were always preceded by 20 s of silence. Participants were instructed to listen attentively to the syllables and to indicate in the test phase after each stream was heard whether a given set of syllables had been presented. As in the SSS test, each test phase consisted of eight trials. In each, a syllable from a pool of eight (four presented and four foils) was randomly selected and presented visually. Participants indicated their decision by pressing a button with their right hand (index finger: yes, the syllable was present; middle finger: no, the syllable was not present). The following trial started between 900 and 1,100 ms after the response, for which there were no time constraints.

Data acquisition and processing. Neuromagnetic responses were recorded with a 1,000-Hz sampling rate using a 157-channel whole-head axial gradiometer system (KIT, Kanazawa Institute of Technology, Japan) in a magnetically shielded room. Five electromagnetic coils were attached to the subject's head to monitor head position during MEG recordings. The coils were localized to the MEG sensors at the beginning of the experiment and before the last two blocks of the main experiment. The position of the coils with respect to three anatomical landmarks—the nasion, and left and right tragus—was determined using 3D digitizer software (Source Signal Imaging) and digitizing hardware (Polhemus). This measurement allowed co-registration of subjects' anatomical MRI with the MEG data. An online bandpass filter between 1 and 200 Hz and a notch filter at 60 Hz were applied to the MEG recordings.

Data processing and analyses were conducted using custom MATLAB code and the FieldTrip toolbox[46]. For each participant's dataset, noisy channels were visually rejected. Two procedures were applied to the continuous MEG recordings. First, a least-squares projection was fitted to the data from the 2 min of empty room recorded at the end of each session. The corresponding component was removed from the recordings[47]. Second, the environmental magnetic field, measured with three reference sensors located away from the participant's head, was regressed out from the MEG signals using time-shifted PCA[48]. The MEG signals were then detrended and artifacts related to eyeblinks and heartbeats were removed by using independent-component analysis.

Source reconstruction. To reconstruct the brain activity generating the magnetic fields recorded by the MEG sensors, we used a linearly constrained minimum-variance beamforming approach. Using the subject's anatomical MRI, we first reconstructed the brain surface. Then, the brain volume was filled with a 1-cm grid, which was normalized to the MNI template (Montreal Neurological Institute brain) using SPM8 (http://www.fil.ion.ucl.ac.uk/spm). The lead fields were calculated for each grid point[49], and spatial filters were computed using the covariance matrix between all sensor pairs for all trials. Finally, the spatial filters were applied to the signals from sensors to reconstruct the time series for each source inside the brain (point on the grid).

Brain-to-stimulus synchronization. The degree of synchrony was measured by the PLV between brain activity and the cochlear envelope of the perceived train of syllables. Signals from sources were resampled at 100 Hz, filtered between 3.5 and 5.5 Hz, and their phases were extracted by means of the Hilbert transform. The PLV was computed for windows of 1 s in duration with an overlap of 0.5 s. The results for all time windows were averaged across the total presentation of the stimuli, obtaining one PLV per source and per subject.

PLVs were averaged for sources within the same region according to the Brainnetome Atlas[50] (38 mean PLV values were computed for the frontal ROI and 24 were computed for the temporal ROI; see below). ROI selection was theoretically driven, based on our initial hypothesis related to the audio–motor nature of the behavioral task. In other words, because segregation of participants relied on an audio–motor task, we focused our analyses on two broad regions comprising the cortical areas related to speech perception and production:

1. Bilateral frontal ROI: this was composed of 19 regions in each hemisphere, 38 in total. It comprised the middle, precentral and inferior frontal gyri from both hemispheres. Specifically, the selected Brainnetome Atlas regions were as follows: dorsal BA 9/46, inferior frontal junction, BA 46, ventral BA 9/46, ventro-lateral BA 8, ventro-lateral BA 6, lateral BA 10, dorsal BA 44, inferior frontal sulcus, caudal BA 45, rostral BA 45, opercular BA 44, ventral BA 44, head and face region BA 4, caudal dorso-lateral BA 6, upper limb region BA 4, trunk region BA 4, tongue and larynx region BA 4, and caudal ventro-lateral BA 6 (Fig. 2b);
2. Bilateral temporal ROI: this consisted of 12 regions in each hemisphere, 24 in total. This ROI covered bilaterally the superior, middle and posterior temporal lobe (medial BA 38, BA 41/42, TE 1.0/1.2, caudal BA 22, lateral BA 38, rostral BA 22, caudal BA 21, rostral BA 21, dorsolateral BA 37, anterior superior temporal sulcus (STS), rostro-posterior STS and caudo-posterior STS; Supplementary Fig. 6b).

Because a preference of right auditory areas for frequencies matching the syllable rate has been theoretically proposed[51,52] and experimentally reported[53–55], we also explored the degree of asymmetry in auditory entrainment to speech. On the basis of the previous literature[53,54], more restricted ROIs were chosen for this analysis. We defined early auditory regions as follows: BA 41/42, TE 1.0 and TE 1.2 (Supplementary Fig. 7a). Next, we calculated brain-to-stimulus synchrony within right and left early auditory areas for each group and calculated their neurophysiological asymmetry: $(PLV_{right} - PLV_{left})/0.5(PLV_{right} + PLV_{left})$.

To assess differences between groups (high and low synchronizers), nonparametric independent-samples Mann–Whitney–Wilcoxon tests were calculated for the PLVs of all regions within the corresponding ROI. Significant results are reported at FDR-corrected $P < 0.05$ within the ROI. To explore auditory

brain-to-stimulus synchrony between hemispheres within groups, a nonparametric paired Wilcoxon signed-rank test was computed.

For the frontal ROI, 8 of the 38 regions showed a significant difference between the groups, FDR corrected for multiple comparisons. All significant regions were located in the left hemisphere, specifically left: dorsal BA 9/46, ventral BA 9/46, inferior frontal junction, dorsal BA 44, ventral BA 44, opercular BA 44, inferior frontal sulcus and caudal BA 45.

Anatomical connectivity study. *Scanning parameters and diffusion measures.* DW-MRI data were acquired on a 3T scanner (Siemens Prisma 3T MRI scanner) using a 64-channel phased-array head coil, at the Center for Brain Imaging (New York University). Diffusion images were acquired with an EPI sequence optimized for DTI-MRI of white matter (81 axial slices; TR, 4,150 ms; TE, 85.2 ms; flip angle, 90°; slice thickness, 1.5 mm; acquisition matrix, 150×152; voxel size, 1.5×1.5×1.5 mm^3). One run with 10 interleaved non-diffusion-weighted volumes and 128 diffusion-weighted volumes (128 directions; b values of 1,500 s/mm^2) was acquired. To allow precise source reconstruction of the neurophysiological data, a high-resolution T1 MPRAGE image was also acquired during this MRI session (TR, 2,400 ms; TE, 2.24 ms; flip angle, 8°; voxel size, 0.80×0.80×0.80 mm^3; 256 sagittal slices; acquisition matrix, 320×300).

DTI-MRI analysis. Diffusion data processing started by correcting for eddy current distortions and head motion by using FMRIB's Diffusion Toolbox (FDT), which is part of the FMRIB Software Library (FSL 5.0.1; http://www.fmrib.ox.ac.uk/fsl/; ref. [56]). Subsequently, the gradient matrix was rotated to correct for head movement, to provide a more accurate estimate of diffusion tensor orientations, by using the fdt_rotate_bvecs program included in FSL[57]. Brain extraction was performed by using the Brain Extraction Tool[58], which is also part of the FSL distribution. Analysis continued with reconstruction of the diffusion tensors by using the linear least-squares algorithm included in Diffusion Toolkit 0.6.2.2[59]. Finally, FA and radial diffusivity (RD) maps for each participant were calculated using the eigenvalues extracted from the diffusion tensors.

Voxel-based analyses of FA and RD maps were performed with TBSS[60]. FA maps from all participants were registered to the FMRIB58_FA template (MNI152 space and 1×1×1 mm^3) using the nonlinear registration tool[61]. These registered FA maps were first averaged to create a mean FA volume. Then, a mean FA skeleton was derived, which represents the centers of all tracts common to all participants in the study. Each participant's aligned FA data were then projected onto this skeleton by searching for the highest FA value within a search space perpendicular to each voxel of the mean skeleton. This process was repeated for the RD maps by applying the transformations previously calculated with the FA maps. This resulted in individual FA and RD skeletons for each participant. In addition, given that laterality[62]—especially of the white matter paths connecting auditory and motor regions—is also related to cognitive function, laterality maps were also created. First, a symmetrical skeleton was created by using the script tbss_sym. Then, FA and RD data were projected onto this symmetrical skeleton, with left-hemispheric values subtracted from right-hemispheric ones. Thus, laterality FA and RD maps were also obtained (note that these maps reflect right minus left values; for the sake of clarity, results are shown on the left hemisphere; Fig. 3). A total of four analyses were performed (FA and RD, and FA and RD lateralization).

Finally, to assess white matter differences between high and low synchronizers, independent-samples t-tests were calculated for the FA and RD skeletons and laterality maps. For a more theoretically driven analysis, we focused on regions that are part of the dorsal and ventral pathways for language processing. In particular, we used an ROI approach to focus on the white matter pathways connecting auditory with frontal and motor regions—the arcuate fasciculus[21]. We also included in the analyses the ventral pathways connecting temporal, occipital and frontal areas as control regions. Our ROI was created by using well-known probabilistic atlases of white matter pathways in MNI space[16,63]. We included any voxel within the skeleton that had at least a 50% probability of being part of the long, anterior or posterior segments of the arcuate fasciculus, the inferior fronto-occipital fasciculus, the inferior longitudinal fasciculus or the uncinate fasciculus. Significant results are reported at FWE-corrected $P<0.05$ using threshold-free cluster enhancement[64] and a nonparametric permutation test with 5,000 permutations[65]. Significant voxels within the skeleton were filled to make the presentation of results easier to follow. Significant clusters (results) were averaged and a mean value per participant, reflecting individual microstructural differences, was obtained. These diffusion values were then correlated (using Spearman rank correlation) to MEG-derived measures of brain synchrony in both frontal and auditory regions (FDR corrected for the two correlations computed; see MEG results).

Tractography analyses. Given the TBSS results showing that a cluster consistent with the left arcuate fasciculus differentiated high and low synchronizers (Fig. 3), confirmatory ad hoc virtual dissections (deterministic tractography) were also performed to further locate the white matter pathways underlying the pattern of results. Specifically, for each participant, we manually dissected the three segments (long, anterior and posterior) of the left arcuate fasciculus[24]. As a control, the left inferior fronto-occipital fasciculus (IFOF) and the left inferior longitudinal fasciculus (ILF) were also dissected. The IFOF and ILF were selected on the basis of (i) the fact that their anatomy could partially overlap with the TBSS cluster and (ii) research suggesting that these pathways are part of the ventral pathway for language processing[14,29,66].

By using the previously processed diffusion data, whole-brain tractography was performed with Diffusion Toolkit 0.6.2.2[59] and the interpolated streamlines algorithm. Tractography was started only in voxels with an FA value greater than 0.2 and was stopped when the angle between two consecutive steps was larger than 35°. Manual dissection of the tracks was performed with Trackvis[59]. ROIs were defined by using the T1 high-resolution image and the FA and FA color-coded maps as a reference for individual anatomical landmarks. The three segments of the left arcuate were dissected by using established guidelines with a two-sphere approach. For the two control ventral tracts, three spherical ROIs at the level of the anterior temporal lobe (temporal ROI), the posterior region located between the occipital and temporal lobes (occipital ROI) and the anterior floor of the external/extreme capsule (frontal ROI) were created. To define each of the ventral tracts of interest, we applied a two-ROI approach: the ILF was obtained by connecting the temporal and occipital ROIs, while the streamlines passing through the occipital lobe and frontal ROIs were considered as part of the IFOF. All these ROIs were applied according to a well-defined anatomical atlas[16]. Exclusion of single-fiber structures that did not represent part of the dissected tract was achieved by using subject-specific no-ROIs. After dissection was completed, the volume and mean FA and RD values for each tract were extracted for further analysis. To take into account individual differences in head volume, volumes from all tracts were corrected by dividing the original value by the total intracranial volume (TIV) for each subject. TIV was calculated by submitting each participant's T1 high-resolution image to the standard Freesurfer pipeline (http://surfer.nmr.mgh.harvard.edu/). By using the extracted values (FA, RD and volume), we computed between-group comparisons for the arcuate as a whole (sum of the anterior, long and posterior segments), its three segments separately, and the IFOF and ILF as control tracts. Thus, we computed a total of 18 comparisons. We used an FDR-corrected significance threshold of $P<0.05$ to correct for these multiple calculations.

Phonological word-form-learning task. The task consisted of a volume adjustment step as above and two runs each with the following experimental steps:

1. Exposure phase: participants were exposed to a 2-min-long speech stream of words corresponding to one of the created languages (L1 to L4; Supplementary Table 1) and were asked to remain silent (no whispering) during the auditory presentation;
2. Test phase: each word stream was immediately followed by a test phase. Test trials consisted of a two-alternative forced choice between a word and a part-word, both randomly selected from the pool corresponding to a particular language. Each word and part-word appeared only twice, each time paired with a different item. The total number of test trials was thus eight. Test items were presented in both their auditory and written forms and were assigned a number (1 or 2) according to their auditory presentation and left–right presentation on the screen. Participants were asked to make their choice by pressing the corresponding number.

The language presentation order was counterbalanced across participants. All participants thus completed two runs, each testing a different language. The proportion of correct responses in the two runs was averaged before proceeding with group analyses.

The group of participants ($N=100$) who completed the online and accelerated versions of the SSS test also completed an online version of this word-form-learning task with the above specifications. The online word-learning task was also created using oTree.

To assess learning differences between the groups (high and low synchronizers), a nonparametric independent-samples Mann–Whitney–Wilcoxon test was calculated for the averaged proportion of correct responses.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

All computer code used for this study is available upon request.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Information. Additional data related to this paper may be requested from the authors.

References

32. Pernet, C. R., Wilcox, R. & Rousselet, G. A. Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. *Front. Psychol.* **3**, 606 (2013).
33. Rousseeuw, P. J. & Van Driessen, K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212–223 (1999).

34. Verboven, S. & Hubert, M. LIBRA: a MATLAB library for robust analysis. *Chemom. Intell. Lab. Syst.* **75**, 127–136 (2005).
35. Rousselet, G. A. & Pernet, C. R. Improving standards in brain–behavior correlation analyses. *Front. Hum. Neurosci.* **6**, 119 (2012).
36. Cureton, E. E. Rank-biserial correlation. *Psychometrika* **21**, 287–290 (1956).
37. Kerby, D. S. The simple difference formula: an approach to teaching nonparametric correlation. *Compr. Psychol.* https://doi.org/10.2466/11.IT.3.1 (2014).
38. Dutoit, T. & Pagel, V. Le projet MBROLA: vers un ensemble de synthétiseurs vocaux disponibles gratuitement pour utilisation non-commerciale. *Actes des Journées d'Études sur la Parole, Avignon* 441–444 (1996).
39. Ermentrout, G. B. & Rinzel, J. Beyond a pacemaker's entrainment limit: phase walk-through. *Am. J. Physiol.* **246**, R102–R106 (1984).
40. Strogatz, S. H. *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering* (Perseus Books, 1994).
41. Chen, D. L., Schonger, M. & Wickens, C. oTree—an open-source platform for laboratory, online, and fieldexperiments. *J. Behav. Exp. Financ.* **9**, 88–97 (2016).
42. Huss, M., Verney, J. P., Fosker, T., Mead, N. & Goswami, U. Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* **47**, 674–689 (2011).
43. Crump, M. J. C., McDonnell, J. V. & Gureckis, T. M. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS One* **8**, e57410 (2013).
44. Chi, T. & Shamma, S. NSL Matlab Toolbox https://isr.umd.edu/Labs/NSL/Software.htm (2003).
45. Arthur, D. & Vassilvitskii, S. k-means++: the advantages of careful seeding. in *Proc. 18th Ann. ACM-SIAM Symp. Discrete Algorithms* 1027–1025 (Society for Industrial and Applied Mathematics, 2007).
46. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
47. Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y. & Ochiai, M. Reduction of non-periodic environmental magnetic noise in MEG measurement by Continuously Adjusted Least squares Method. *IEEE Trans. Appl. Supercond.* **11**, 669–672 (2001).
48. de Cheveigné, A. & Simon, J. Z. Denoising based on time-shift PCA. *J. Neurosci. Methods* **165**, 297–305 (2007).
49. Nolte, G. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Phys. Med. Biol.* **48**, 3637–3652 (2003).
50. Fan, L. et al. The human Brainnetome Atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
51. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255 (2003).
52. Zatorre, R. J. & Belin, P. Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953 (2001).
53. Boemio, A., Fromm, S., Braun, A. & Poeppel, D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* **8**, 389–395 (2005).
54. Giraud, A. L. et al. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**, 1127–1134 (2007).
55. Telkemeyer, S. et al. Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* **29**, 14726–14733 (2009).
56. Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
57. Leemans, A. & Jones, D. K. The *B*-matrix must be rotated when correcting for subject motion in DTI data. *Magn. Reson. Med.* **61**, 1336–1349 (2009).
58. Smith, S. M. Fast robust automated brain extraction. *Hum. Brain Mapp.* **17**, 143–155 (2002).
59. Wang, R., Benner, T., Sorensen, A. G. & Wedeen, V. J. Diffusion Toolkit: a software package for diffusion imaging data processing and tractography. *Proc. Int. Soc. Magn. Reson. Med.* **15**, 3720 (2007).
60. Smith, S. M. et al. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487–1505 (2006).
61. Andersson, J. L. R., Jenkinson, M. & Smith, S. *Non-linear registration, aka spatial normalisation*. FMRIB Technial Report TR07JA2. (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University, 2007).
62. Catani, M. et al. Symmetries in human brain language pathways correlate with verbal recall.*Proc. Natl Acad. Sci. USA* **104**, 17163–17168 (2007).
63. Rojkova, K. et al. Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study. *Brain Struct. Funct.* **221**, 1751–1766 (2016).
64. Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
65. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* **15**, 1–25 (2002).
66. Duffau, H. et al. New insights into the anatomo-functional connectivity of the semantic system: a study using cortico-subcortical electrostimulations. *Brain* **128**, 797–810 (2005).

nature research

Corresponding author(s): M. Florencia Assaneo

Last updated by author(s): Jan 22, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used: Matlab_R2015a and PsychoToolbox-3 to collect behavioral data and subjects' responses during the MEG session; MEG160 (Yokogawa Electric Corp., Eagle Technology Corp., Kanazawa Institute of Technology) and Syngo Software 11C (Siemens Prisma 3T MRI scanner). Amazon Mechanical Turk Platform and Otree was used to collect online behavioral data. The computer code used for the data collection is available upon request
Data analysis	Matlab_R2015a, FieldTrip toolbox, FMRIB Software Library 5.0.1, Diffusion Toolkit 0.6.2.2, FreeSurfer 6.0, TrackVis 0.6.1, JASP 0.9.0.1, Auditory Model Matlab toolbox, 3D digitizer software Source Signal Imaging, SPM8, diptest package under R software, Robust Correlation Toolbox for Matlab and custom algorithms. The computer code used for the analyses is available upon request

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Additional data related to this paper may be requested from the authors. Correspondence and request for materials should be addressed to MFA (fassaneo@gmail.com)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications
Data exclusions	Exclusion criteria was pre-established and the following data points were excluded for the detailed reasons. 3 participants were excluded due to artifactual MEG noise and 1 extra participant due to excessive movement during the DW-MRI. 56 subject were removed form the stable rate SSS-test Mechanical Turk protocol for non-optimal conditions in their recording. 40 subject were removed form the accelerated rate SSS-test Mechanical Turk protocol for non-optimal conditions in their recording. 7 participants were excluded from the accelerated version of the SSStest because they spoke instead of whispering and/or stopped whispering for more than 4 seconds
Replication	The outcome from our behavioral test was replicated on four different cohort of participants. The learning differences between groups was replicated in an online version of the same task.
Randomization	Stimulus presentation order was randomized for all experiments with more than one stimulus. When comparing between groups we controlled for age, language background, gender and years of musical training.
Blinding	Data collection and analysis were performed blind to the conditions of the experiments for the behavioral tests but not for the MEG and structural analyses, as subjects had already been divided into high and low synchronizers.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	84 participants completed the first behavioral test (SSStest, 32 males; mean age, 28; age range, 19 to 55). From this, a subgroup of 37 subjects (right handed; 18 males; mean age, 30; age range, 21 to 55) also underwent MEG and DW-MRI protocols. A second cohort of 44 individuals (11 males; mean age, 21; age range, 19 to 31) completed a replication of the behavioral test and a word learning task. 55 participants completed a control behavioral test (19 males; mean age, 23; age range, 18 to 36). All participants the previously mentioned cohorts were native English speakers and self-reported normal hearing and no neurological deficits. 144 native english speakers completed a first Amazon Mechanical Turk experiment (80 males; mean age,34; age range, 19 to 5) and 60 native english speakers completed a second Amazon Mechanical Turk experiment (37 males; mean age, 35; age range, 19 to 51)
Recruitment	Participants were recruited from the grater New York University community and from the Amazon Mechanical Turk pool of participants. Participants recruitment was performed blind to their demographics.
Ethics oversight	Institutional Review Board (New York University's Committee on Activities Involving Human Subjects)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	N/A (diffusion weighted MRI)
Design specifications	N/A (diffusion weighted MRI)
Behavioral performance measures	N/A (diffusion weighted MRI)

Acquisition

Imaging type(s)	diffusion
Field strength	3 tesla
Sequence & imaging parameters	EPI sequence (81 axial slices, TR: 4150 ms, TE: 85.2 ms, flip angle: 90o, slice thickness: 1.5 mm, acquisition matrix: 150 × 152, voxel size: 1.5 × 1.5 × 1.5 mm3). T1 MPRAGE image (TR = 2400 ms, TE = 2.24 ms, flip angle = 8o, voxel size = 0.80 × 0.80 × 0.80 mm3, 256 sagittal slices, acquisition matrix = 320 × 300).
Area of acquisition	whole brain

Diffusion MRI ☒ Used ☐ Not used

Parameters: 128 directions, single shell, no cardiac gating, b=1500

Preprocessing

Preprocessing software	FMRIB Software Library 5.0.1, Diffusion Toolkit 0.6.2.2, FreeSurfer 6.0, TrackVis 0.6.1
Normalization	FNIRT (non-linear registration) for DW-MRI
Normalization template	FMRIB58_FA template (MNI152 space and 1×1×1 mm3)
Noise and artifact removal	N/A
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	Univariate independent samples
Effect(s) tested	Rank Biserial Correlation

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s): Long, anterior, and posterior segments of the arcuate fasciculus, the inferior-fronto occipital fasciculus, the inferior longitudinal fasciculus and the uncinate fasciculus.

Statistic type for inference (See Eklund et al. 2016)	Threshold-free cluster enhancement
Correction	FWE-corrected $p < 0.05$ value using a nonparametric permutation test with 5000 permutations; FDR

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis