Population-level differences in the neural substrates supporting Statistical Learning

Joan Orpella[1,†], M. Florencia Assaneo[2,*,†], Pablo Ripollés[1, 3, 4], Laura Noejovich[1], Diana López-Barroso[5,6], Ruth de Diego-Balaguer[7,8,9,10,‡], David Poeppel[1,4,11,‡]

[1]Department of Psychology, New York University, New York, NY, USA.

[2]Institute of Neurobiology, National Autonomous University of Mexico, Juriquilla, Querétaro, Mexico.

[3]Music and Audio Research Lab (MARL), New York University, New York, NY, USA.

[4]Center for Language, Music and Emotion (CLaME), New York University, New York, NY, USA and Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany.

[5]Cognitive Neurology and Aphasia Unit, Centro de Investigaciones Médico-Sanitarias, Instituto de Investigación Biomédica de Málaga – IBIMA and University of Málaga, Málaga, Spain.

[6]Department of Psychobiology and Methodology of Behavioral Sciences, Faculty of Psychology and Speech Therapy, University of Málaga, Málaga, Spain.

[7]Cognition and Brain Plasticity Unit, IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain.

[8]Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain.

[9]Institute of Neuroscience, University of Barcelona, Barcelona, Spain.

[10]ICREA, Barcelona, Spain.

[11]ESI, Frankfurt, Germany.

* Corresponding author (fassaneo@gmail.com)

† These authors contributed equally

‡ These authors jointly supervised this work

The ability to extract regularities from the environment is arguably an adaptive characteristic of intelligent systems. In the context of speech, statistical learning is thought to be an important mechanism for language acquisition. By considering individual differences in speech auditory-motor synchronization, an independent component analysis of fMRI data revealed that the neural substrates of statistical word form learning are not fully shared across individuals. While a network of auditory and superior pre/motor regions is universally activated in the process of learning, a fronto-parietal network is instead additionally and selectively engaged by some individuals, boosting their performance. Furthermore, interfering with the use of this network via articulatory suppression (producing irrelevant speech during learning) normalizes performance across the entire sample. Our work provides novel insights on language-related statistical learning and reconciles previous contrasting findings, while highlighting the need to factor in fundamental individual differences for a precise characterization of cognitive phenomena.

INTRODUCTION

Statistical learning (SL) is the capacity to use distributional information present in the environment to extract meaningful regularities. SL has been demonstrated across age groups from birth (1,2), sensory modalities (e.g., audition (3,4), vision (5), touch (6)), representational domains (5) (temporal, spatial), and even species (7,8). In the domain of speech and language processing, statistical *word form* learning (SWFL) is considered critical in the early stages of language acquisition as the ability to segment phonological word forms from continuous speech (3,9). Segmented word forms are readily associated with meanings (10) and can also be used in subsequent stages to discover grammatical relationships (11). While regarded fundamental to

language learning by most contemporary theories, the precise neural substrates of this ubiquitous phenomenon are not well understood and remain controversial.

There have been several experimental attempts to pinpoint the neural basis of SWFL, but the existing literature shows inconsistent results. Some studies report a correlation between SWFL performance and the activity of the superior temporal gyrus (STG) and dorsal pre/motor regions (12–15). Other experiments instead implicate the left inferior frontal gyrus (IFG) (16) and its interaction with superior temporal areas (17,18). We hypothesize that the inconsistency of results in the literature is a consequence of very specific individual differences in the neural resources allocated for SWFL.

In a recent study, we provided a first glance of how one might capitalize on individual differences to gain deeper mechanistic insights into SWFL: individual listeners grouped by their *spontaneous speech auditory-motor synchronization* abilities turn out to differ in their SWFL performance (19). Specifically, we showed there a behavioral task (the Spontaneous Speech Synchronization test, henceforth 'SSS-test') that robustly classifies participants into high and low speech auditory-motor synchronizers. At the brain level, high synchronizers showed a greater brain-to-stimulus synchrony in the left IFG during passive speech listening as well as more volume in the white matter pathways underlying the dorsal language stream (i.e., the arcuate fasciculus) (20). Critically, the high/low synchronizer distinction was predictive of SWFL performance (Fig 1A), but the relationship between auditory-motor synchrony, population-level brain differences, and SWFL remained elusive. Here we hypothesized that the dorsal language stream, including the IFG, is not only linked to auditory-motor synchrony as previously reported but also gives high synchronizers the advantage in SWFL over low synchronizers.

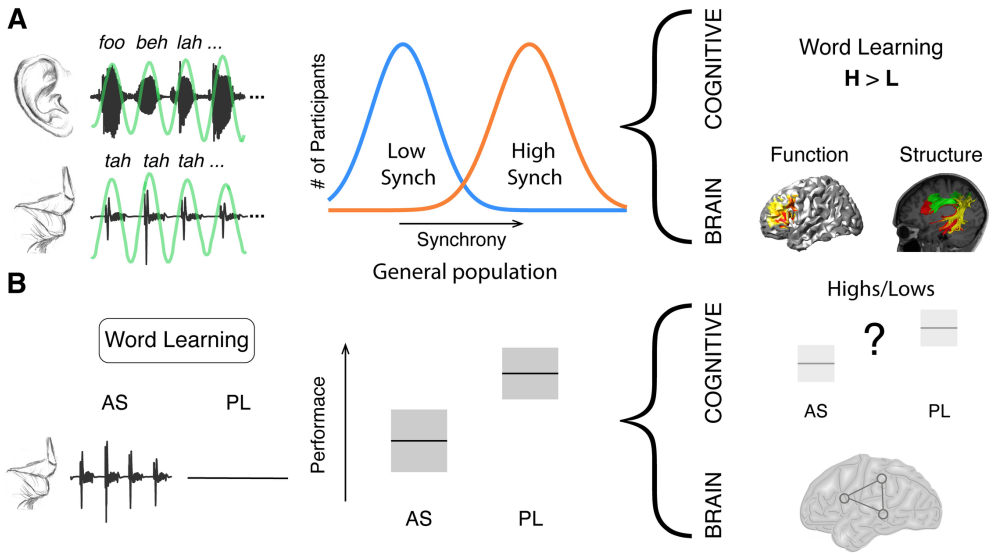

Fig 1. Previous work motivating the hypothesis. (A) During the SSS-test, participants listen to an isochronous stream of random syllables (rate 4.5 syllables/sec) while concurrently whispering the syllable "*tah*". *Left panel:* example of the perceived (upper panel) and produced (lower panel) signals. Green line, bandpass filtered envelope used to compute input-output synchrony. *Middle panel:* Synchrony between perceived and produced syllables yields a bimodal distribution, allowing the classification of participants into low (blue) and high (orange) synchronizers. While some participants spontaneously align the produced syllabic rate to the perceived one (high synchronizers), others show no modification of the produced rate due to the presence of the external rhythm (low synchronizers). *Right panel:* High synchronizers outperformed lows in a statistical word-learning task. They also showed enhanced brain-to-speech synchronization over left frontal regions and a greater volume in the white-matter pathways connecting temporal and frontal areas(19). **(B)** The word-learning task consists of a learning phase wherein four tri-syllabic pseudo-words are presented in a continuous stream. Learning is assessed post exposure. *Left panel:* Participants are instructed to repeat a nonsense syllable (articulatory suppression condition; AS) or to passively listen (passive listening condition; PL) during the entire learning phase. *Middle panel:* predicted performance decreases due to articulatory suppression (9). *Right panel:* differences between high and low synchronizers are hypothesized at the cognitive and brain levels.

To test this hypothesis, we used Independent Component Analysis (ICA) of fMRI data in combination with a classic behavioral paradigm designed to interfere with the availability of the dorsal language stream for SWFL (21). Specifically, the behavioral paradigm employed involves the contrast between passive listening (PL) and articulatory suppression (AS) conditions (Fig 1B). AS requires participants to repeat a nonsense syllable during (word) learning, which hampers SWFL performance (9). ICA, on the other hand, is a data-driven neuroimaging approach well-suited to identify spatially independent and temporally coherent brain networks that support specific cognitive processes (22). Previous work using this approach has related SWFL to a network comprising auditory and superior pre/motor areas (14). This earlier work, however, did not consider the high/low synchronizer distinction. In all, in the current experiment we investigate, in both high and low synchronizers, the brain networks engaged during SWFL under PL and AS conditions as well as the behavioral consequences of AS for performance (Fig 1). We hypothesize that, if high synchronizers show better learning – putatively due to a greater reliance on the dorsal language stream - they should show a greater recruitment of this functional anatomic stream than low synchronizers during learning as well as a greater AS effect.

RESULTS

Behavioral results: Articulatory Suppression modulates high but not low synchronizers' SWFL performance

An initial cohort (N = 55, 34 females; mean age, 22; age range, 18 to 37) underwent behavioral testing. Participants completed four blocks of statistical word-learning in two different experimental conditions, PL and AS, followed by the SSS-test (Method and Fig 1). In both tasks, the auditory stimuli were presented at a rate of 4.5 syllables per second, corresponding to the mean

syllable rate across languages (23–25) and the natural frequency of speech-motor regions(26). The outcome of the SSS-test showed the expected (19,27) bimodal distribution, allowing the classification of participants into high and low synchronizers (Fig 2A). Moreover, the synchrony between perceived and produced syllables in the SSS-test was highly correlated with that in the AS blocks (Fig 2B; N = 55, Spearman correlation coefficient r = 0.75, $p < 0.001$). This demonstrates that speech-to-speech synchrony is not only reliable across *time*, as was previously demonstrated (19), but also across *tasks*, confirming that auditory-motor synchrony is a stable feature of each individual.

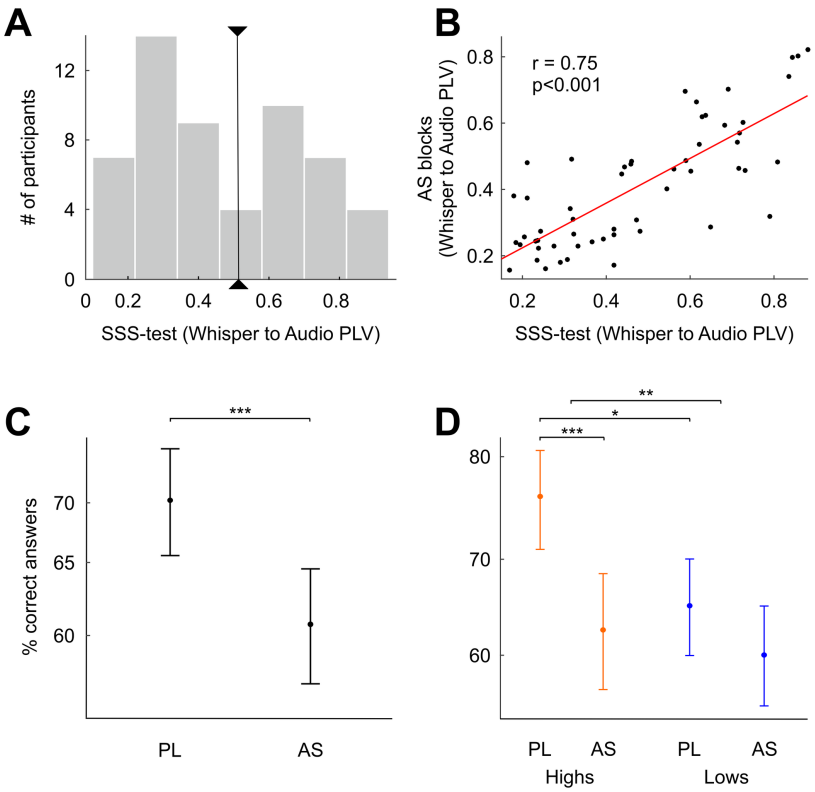

Fig 2. Articulatory Suppression modulates only high synchronizers' performance. (A) SSS-test outcome. Histogram of the phase locking values (PLV; measure of speech-to-speech synchrony, see Methods) between the envelope of the perceived and produced speech signals, bandpass filtered at 3.5-5.5 Hz. Black line, critical value

separating high and low synchronizers (N = 55, see Methods). (**B**) Participants' PLV during AS as a function of the PLV from the SSS-test. Red line represents the correlation of the data. (**C**) Percentage of correct responses during PL and AS on the entire sample. (**D**) Percentage of correct responses during PL and AS for the low (blue) and the high (orange) synchronizers. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ Linear mixed model results. Dots: model predicted group means. Bars: 95% confidence interval. AS: articulatory suppression. PL: passive listening.

A linear mixed-model analysis (N = 55; see Method) of the learning performance showed a significant decrement in AS relative to PL (Fig 2C; Main effect of Condition: $\chi2 = 15.4$, $p < 0.001$). This result thus replicates previously reported AS effects on SWFL(9). The analysis also showed a main effect of Group (Highs > Lows; $\chi2 = 9.11$, $p < 0.01$) in line with our previous work (19) and, importantly, a significant interaction between the two factors (Condition*Group; $\chi2 = 4.22$, $p < 0.05$). Critically, when the sample was next split into high and low synchronizers by estimating their corresponding marginal means (see Method), we observed the AS effect in the population of high synchronizers (Fig 2D; $N_{high} = 23$, zratio = 3.92, $p < 0.001$) but not in the population of low synchronizers (Fig 2D; $N_{low} = 32$, zratio = 1.63, $p = 0.1$); that is to say, the performance of low synchronizers was not modulated by the action of speaking during the learning phase. Additionally, in line with previously reported data(19), high synchronizers outperformed lows in the PL condition (zratio = 3.02, $p < 0.01$), but there was no difference in performance between groups in the AS condition (zratio = 0.64, $p = 0.52$). Importantly, for all groups and conditions, learning remained significantly above chance (signed rank tests against chance level, two-sided: $p_{high/AS} < 0.001$, $p_{high/PL} < 0.001$, $p_{low/AS} < 0.001$, $p_{low/PL} < 0.001$).

Neuroimaging results (I): High synchronizers activate an additional brain network during statistical word learning

Having established the expected behavioral differences between high and low synchronizers, we next acquired fMRI data from a new group of participants (N = 38) while they performed the same behavioral paradigm (see Method). The SWFL paradigm was optimized for fMRI testing. Specifically, we included both a *rest* block and a *speech motor* block as control conditions. During the speech motor block, participants were required to repeatedly whisper the syllable *"tah"* with no concurrent auditory input. The behavioral performance in the scanner showed the same trend as the learning pattern obtained with the first sample (Fig S1; Condition (PL > AS): $\chi2 = 5.40$, $p < 0.05$; Group (Highs > Lows): $\chi2 = 3.67$, $p = 0.055$; Condition*Group: $\chi2 = 2.74$, $p = 0.098$; Highs (PL>AS): zratio = 2.32, $p < 0.05$; Lows (PL vs AS): zratio = 0.08, $p = 0.93$; Highs vs Lows in PL: zratio = 1.92, $p = 0.055$; Highs vs Lows in AS: zratio = 0.05, $p = 0.96$). Even under notably adverse listening/learning conditions (i.e., during fMRI scanning), the detrimental effect of the AS condition was restricted to the high synchronizers.

Using the Group ICA of fMRI Toolbox(22) (GIFT; see Methods) we identified 5 brain networks that were significantly recruited during SWFL in the PL and/or the AS condition (Fig S2 and Fig S3). Critically, a fronto-parietal network including bilateral inferior and middle frontal gyri, inferior parietal cortex, and the supplementary motor area distinguished between high and low synchronizers during the PL condition (Fig 3A; $N_{high} = 18$, $N_{low} = 20$, Mann-Whitney-Wilcoxon test, two-sided $p = 0.038$, FDR-corrected). Moreover, while the activity of this network during PL was statistically significant for high synchronizers, it was not for the lows (Mann-Whitney-Wilcoxon test, two-sided $p_{high} < 0.005$ and $p_{low} = 0.9$, respectively, FDR-corrected). Moreover, we found moderate evidence in favor of the null hypothesis that the network was not activated during PL for the lows (Bayes Factor $BF_{01} = 4$).

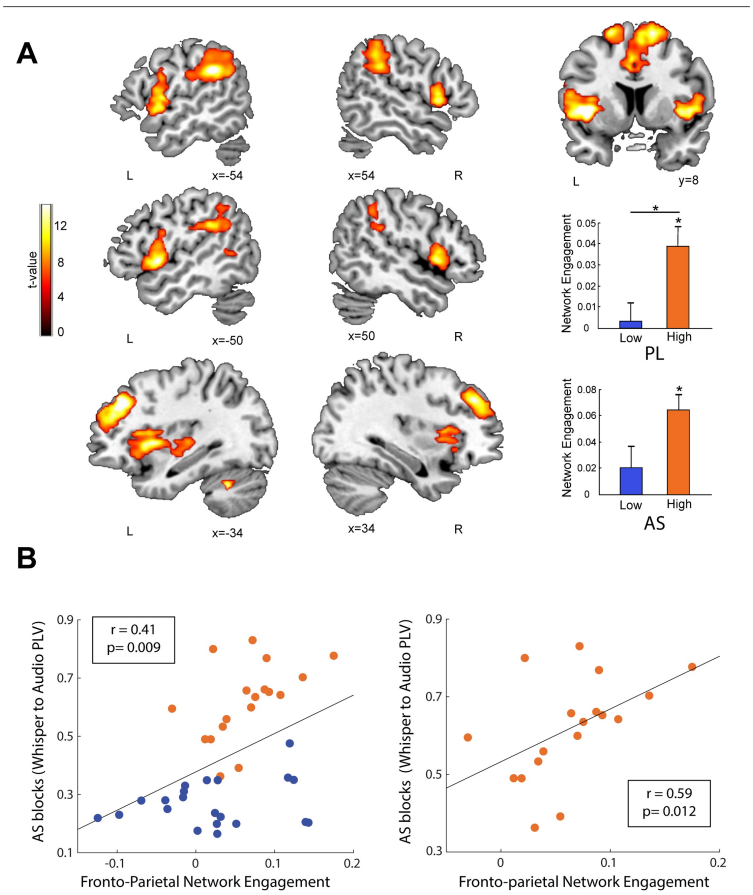

Fig 3. High synchronizers activate an additional brain network during statistical word learning. (A) In red/yellow, the fronto-parietal network is shown over a canonical template, with MNI coordinates on the lower portion of each slice. Neurological convention is used. A $p < 0.05$ FWE-corrected threshold at the cluster level with an auxiliary $p < 0.001$ threshold at the voxel is used. This is the only network showing significant differences in activity between high and low synchronizers during PL (see bar plots on the lower right; * $p < 0.05$, FDR-corrected). **(B)** Scatterplot displaying participants' PLV during AS as a function of the fronto-parietal network's engagement. Red line represents the correlation of the data. *Left panel:* all subjects. *Right panel:* high synchronizers.

Similarly, during AS, only high synchronizers significantly engaged the fronto-parietal network (Fig 3A; Mann-Whitney-Wilcoxon test, two-sided $p_{high} < 0.005$ and $p_{low} = 0.42$, FDR-corrected), again with moderate evidence in favor of the null hypothesis for the lows ($BF_{01} = 4.1$). In this condition, however, the network's activity did not differentiate between the groups. Given

that groups were defined by their speech auditory-motor synchrony, we then correlated the engagement of the fronto-parietal network with the synchronization (PLV) between the perceived and produced syllables during AS. Indeed, these measures were positively correlated in the entire sample as well as in the high synchronizers only (Fig 3B; Spearman correlation coefficient r_{all} = 0.41 and r_{high} = 0.56, p_{all} = 0.009 and p_{high} = 0.012), suggesting a link between spontaneous auditory-motor synchrony and fronto-parietal network engagement.

Neuroimaging results (II): The interplay between networks boosts learning

Next, we assessed whether the activity of any of the networks significantly engaged during the PL condition was predictive of SWFL. Replicating previous results(14,28), we found a network comprising mainly bilateral auditory regions and a small superior pre/motor cluster (henceforth, auditory network) whose activity positively correlated with learning performance in the whole sample (Fig 4A; Spearman correlation coefficient r = 0.42 and p = 0.032, FDR-corrected). We found no significant correlations between learning and network activity in the AS condition. Since during PL (*i)* highs behaviorally outperformed lows; (*ii)* the fronto-parietal network was only activated by high synchronizers; and (*iii)* the auditory network was related to learning performance, we next examined whether the learning benefit of high synchronizers over lows in SWFL was related to the interaction between the two networks (auditory and fronto-parietal). Specifically, we explored the relationship between the time courses of these two networks at the individual listener level and the learning benefit. As illustrated in Figure 4B, high synchronizers with a greater learning benefit (defined as PL minus AS) appeared to show a distinct pattern with interweaving time courses between the auditory and fronto-parietal networks. To quantify this observation, we employed an analysis typically used in electronics: XOR. Applied to our signals,

this logical operation assigns a single value per time point: *one* (1) when a single network is above baseline activity, or *zero* (0) otherwise; that is, a *one* is assigned when one or the other network (but not both) is active (Fig 4B, lower insets). For each high synchronizer, we averaged the XOR over time, and correlated this value with their learning benefit (PL-AS) (note that this analysis would be meaningless for low synchronizers, given the non-significant activation of their fronto-parietal network). In line with the observed pattern, a positive correlation was found (Fig 4C; Spearman correlation coefficient r = 0.65, $p < 0.005$). This suggests that the learning benefit shown by high synchronizers over lows is related to a specific pattern of activity highlighted by the XOR rather than a perfect correlation between time courses of the networks.

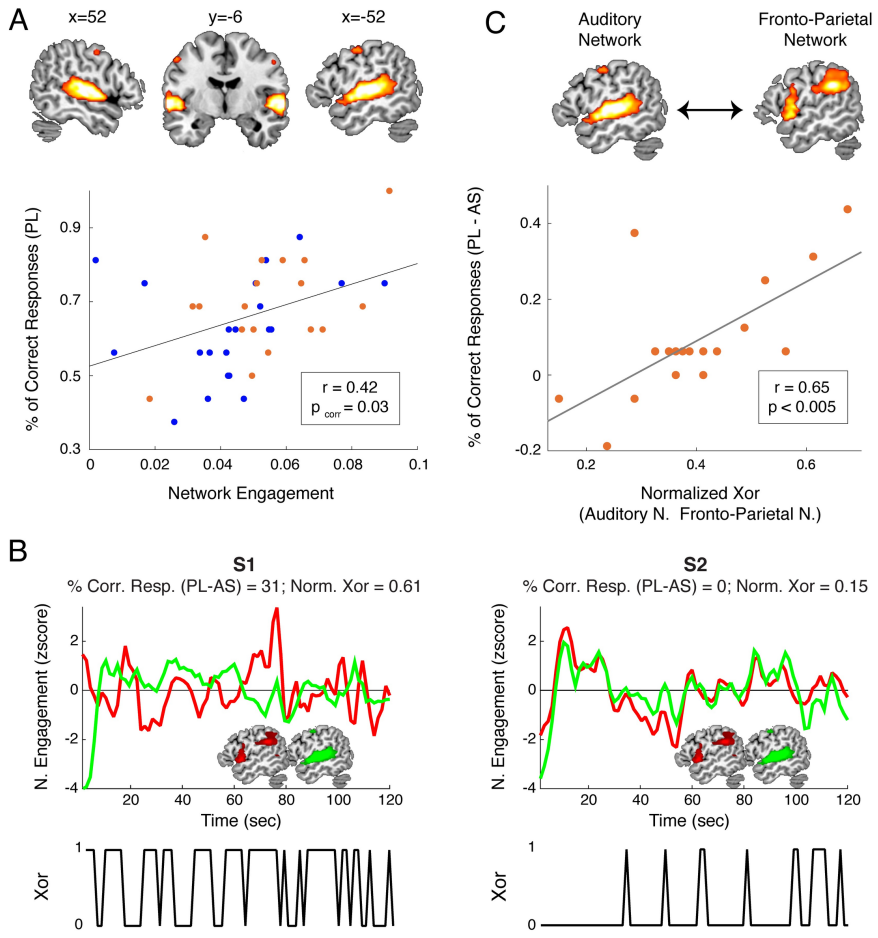

Fig 4. An interplay between networks boosts learning. (A) The auditory network supports learning during PL. *Upper panel:* In red/yellow, the auditory network is shown over a canonical template, with MNI coordinates on the upper portion of each slice. Neurological convention is used with a $p < 0.05$ FWE-corrected threshold at the cluster level, with an auxiliary $p < 0.001$ threshold at the voxel level. *Lower panel:* Scatterplot displaying participants' percentage of correct responses during PL as a function of the auditory network's engagement. (**B)** The learning benefit during PL (correct answers in PL − correct answers in AS) is related to the interplay between the time courses of the fronto-parietal (red) and the auditory (green) networks. *Left/right panel:* a representative high synchronizer with a greater/smaller learning benefit. *Lower panels:* Time evolution of the XOR analysis. (**C)** Scatterplot displaying high synchronizers' learning benefit as a function of the normalized XOR between the fronto-parietal (red) and the auditory (green) networks. Red line: correlation of the data.

DISCUSSION

The behavioral and neuroimaging data show that the neural substrates supporting SWFL vary across individuals in a systematic way. We arrived at this observation by splitting the population into two groups according to their spontaneous speech auditory-motor synchronization abilities (Fig 2), a classification that has now been shown to be robust in a number of experiments, both in-lab and online, in different languages, and with different experimental manipulations (19,27). Specifically, we found two distinct networks related to SWFL performance. One network encompasses mainly auditory regions and a small superior pre/motor cluster (auditory network), appears to be universally or generically recruited, and directly correlates with learning. Another network, including inferior frontal, inferior parietal and supplementary motor areas (fronto-parietal network), is neither necessary nor sufficient for learning, yet it boosts learning performance. This latter network, whose activity correlates with spontaneous auditory-motor synchrony, is exclusively recruited by high auditory-motor synchronizers during learning. These observations

parsimoniously account for the apparently disparate results in previous SWFL literature and provide a new way to discuss statistical learning in neural terms.

In terms of behavior, we find that the effects of AS are not universal. Typically, the execution of an articulatory suppression task leads to performance deficits. We demonstrate –in two independent cohorts– that only the performance of participants with a high degree of auditory-motor synchronization is affected by AS. Low synchronizers, in contrast, remain unaltered in their word learning performance. Our results thus indicate that articulatory rehearsal is not necessary for SWFL but its additional recruitment confers a learning benefit: high synchronizers, who show robust AS effects, performed better than lows during PL. Note that these results are not discordant with the previous literature (9) since averaging across high and low synchronizers yields the expected overall AS effects.

At the neural level, we found an important distinction between high and low synchronizers with respect to the engagement of a fronto-parietal network: only high synchronizers engage this network during PL. While SWFL performance correlates with the activity of the *auditory* network across the entire sample -in line with previous literature (14)- a synergistic relationship between *both* networks boosts learning performance in the high synchronizer group. Importantly, the engagement of the fronto-parietal network also predicted the degree of spontaneous synchronization of produced speech during the AS condition.

A relationship between auditory-motor synchronization and language skills has been previously reported in the literature (19,29,30). For example, precision in tapping to a metronome has been argued to correlate with reading and spelling abilities in children with developmental dyslexia (31). Similarly, synchronization to a beat correlates with phonological awareness and rapid naming in typically-developing preschoolers (32). Despite the cumulative evidence for a link

between auditory-motor synchronization and these various language skills, the existence of a neural substrate shared among these apparently unrelated cognitive abilities remains an empirical question. With this question in mind, our results suggest that the reported fronto-parietal cortical network subserves this shared role: on the one hand, the engagement of this network during PL confers a benefit in learning; on the other, during AS, the engagement of this network predicts the degree of speech auditory-motor synchronization.

Insofar as there are differences between high and low synchronizers at a structural level and differences in synchrony that are stable in time (19), we understand the high/low synchronizer differences reported in this and previous works as *trait* differences. We have also theorized on how structural connectivity differences can give rise to the synchrony differences between the groups under particular stimulation conditions (e.g., auditory stimulation within a specific frequency range; (33)). Because of this, our reported group differences could also be understood as *state* differences, the precise functional significance of which remains an open question.

From a mechanistic and more neurophysiologically motivated perspective, we propose that enhanced syllable-level segmentation or parsing -a key prerequisite for SWFL- results from the coordinated activity between auditory and fronto-parietal networks, ultimately leading to better SWFL performance. In line with this conjecture, we previously showed that the frontal-most component of the fronto-parietal network, the IFG, aligns with the onset of passively perceived syllables in high synchronizers. This frontal region has also been shown to send top-down signals to auditory cortex (the main component of the auditory network) to better align its activity to the speech input (34,35). A similar proposal has been advanced in the literature to account for the enhanced processing of phonology that results from improved auditory timing perception through auditory-motor training (36). Broadly speaking, therefore, our results line up with recent theories

of SL, which postulate the works of both learning systems (e.g., comprising auditory areas in the case of auditory input) and modulatory attentional/control systems (e.g., as supported by fronto-parietal networks) underlying learning performance (37,38). However, we add to these current views of SL by specifying the role of these modulatory systems in terms of timing operations critical to auditory-motor synchronization advantageous to learning.

On the other hand, there exist multiple and distinct fronto-parietal networks associated with attention that overlap with our reported fronto-parietal network in high synchronizers. Its ventral fronto-parietal component (i.e., inferior prefrontal to inferior parietal cortex), for example, has been related to stimulus-driven attention (39), which may in turn be related to the salience network (40). Note that the stimulus-driven attention network is mostly bilateral (e.g., (41)) but shows different patterns of *lateralization*, rightward for spatial attention and leftward for temporal attention (42). Given the relationship between our fronto-parietal network and auditory-motor synchronization (this paper and (19), a possibility therefore is that high synchronizers' fronto-parietal engagement relates to a temporal attention mechanism.

Another possibility is that fronto-parietal activity in high synchronizers relates to a control network (e.g., (43)) that flexibly interacts with other task-specific networks (e.g., (44,45)). This is possible given the activity in more dorsal frontal regions that also feature in our fronto-parietal network. Interestingly, recent articles (e.g., (46,47)) show that a supra-modal fronto-parietal network entrains to stimulation (sensory and via rhythmic TMS) in the theta band and that this entrainment (causally) enhances auditory working memory. This is very similar to our previous (19) and current findings, in which high synchronizers entrain to theta stimulation (higher behavioral PLV and brain-stimulus PLV during passive listening) and show a behavioral advantage over individuals that do not show this entrainment. The extent to which this and the

aforementioned fronto-parietal networks are one same network or different networks that interact for high synchronizers during the SL task cannot be answered by our current data and so remains an empirical question. However, our analysis (ICA for fMRI) indicates that, at the very least, these fronto-parietal regions' time-courses cohere in time.

There are also reasons to distinguish these fronto-parietal networks from the dorsal network for goal-directed attention (48), despite a similar involvement of dorsal prefrontal regions. In contrast to research showing SL benefits from interfering with this network (e.g., (49,50)), we show that articulatory suppression hinders learning. Moreover, fronto-parietal involvement, which correlates with auditory-motor synchronization, confers a learning benefit during passive listening. It is therefore likely that the dorsal prefrontal regions we report, which are shown to cohere in time with other fronto-parietal regions, perform a role different from goal-oriented attention within the context of our tasks. This is in line with the idea that the same region can undertake different roles depending on its interactions with other regions. It was not possible to determine the precise role of prefrontal regions alone from our data. On the other hand, we also show that the learning benefit relates to the way the fronto-parietal network interacts with the auditory network. Another possibility, therefore, is that different kinds of dorsal prefrontal involvement during learning incur in either learning benefits or hindrance.

A possible reason for the lateralization discrepancies with Assaneo et al. (2019) (19) (bilateral engagement vs left lateralization) is the use of radically different measures and analyses (ICA of the BOLD signal vs a phase-locking value between an auditory stimulus and the MEG signals). Thus, although bilateral frontal and parietal regions may work together for synchrony (and learning benefits) in high synchronizers, as reflected in the ICA for fMRI analysis, each region may perform different computations to achieve that goal that are not captured by the PLV

analysis, with entrainment in theta occurring only in left frontal regions. We similarly hypothesize that small structural differences (as those reported in (19), as captured by a particular method (diffusion-weighted MRI in that case), can give rise to large functional differences as appears to be the case in high synchronizers (33).

In sum, by considering individual differences in auditory-motor synchronization skills, our work sheds light onto the neural substrates of the statistical learning of phonological word forms and shows that what appeared to be disparate results in the existing literature stems from pooling together fundamentally distinct populations. More specifically, we reveal that, beyond a universally recruited network for SWFL, an additional fronto-parietal network that enables auditory-motor synchronization is selectively engaged by some individuals to produce a benefit in learning. The auditory-motor SSS-test we use thus emerges, once more, as a useful tool to achieve a more nuanced characterization of speech related phenomena. This work, therefore, not only highlights the importance of considering individual differences in SL (42) but also sounds a note of caution about assuming the existence of monolithic mechanisms underlying such cognitive tasks.

MATERIALS AND METHODS

Participants

A first cohort of 65 participants completed the behavioral protocol. Ten participants were removed because they spoke loudly instead of whispering or because they stopped whispering for longer than 4 sec (during the SSS-test and/or the AS blocks). The data from 55 participants (34 females; mean age: 22; age range: 18 to 37) were analyzed.

A second cohort comprising 22 low and 22 high synchronizers participated in the fMRI protocol. Participants were selected from a larger group of 388 individuals, which completed the

SSS-test in the context of a previous study (19) (Fig S1A). Three participants were initially removed due to recording measurement error or problems completing the learning task (e.g., participant fell asleep, responses not recorded) and were removed from the fMRI sample. Three additional participants were removed due to artifactual fMRI data. The final dataset thus comprised 38 individuals (23 females; mean age: 28; age range: 20 to 54; 19 High Synchronizers and 22 Low).

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

Overall experimental design

The behavioral protocol consisted of four blocks of statistical word-form learning performed under two different conditions (PL and AS), followed by the SSS-test (see Fig 5A). Four *pseudo-languages* were generated, and their order was randomized across participants. In the statistical learning blocks, participants were instructed to pay attention to the audio stream to be able to answer post-exposure questions about the perceived sounds. During the PL condition, participants passively listened to two of the *pseudo-languages*. During AS, participants repeatedly whispered the syllable *"tah"* while listening to the remaining two *pseudo-languages.* As in the SSS-test, participants were *not* instructed to synchronize their speech to the auditory stimulus. Instead, they were told that the point of the whispering was to make the listening task more challenging. In line with the previous literature (9,51), we assumed that the effects of AS on statistical learning would be due to an interference with the articulatory loop rather than to a higher executive load, which would be very unlikely given the highly automatized nature of the

articulation subtask. PL and AS conditions were interleaved, and the two possible orders (PL-AS-PL-AS or AS-PL-AS-PL) were randomized across participants. After listening to each *pseudo-language*, learning was tested on a two-alternative forced choice test.

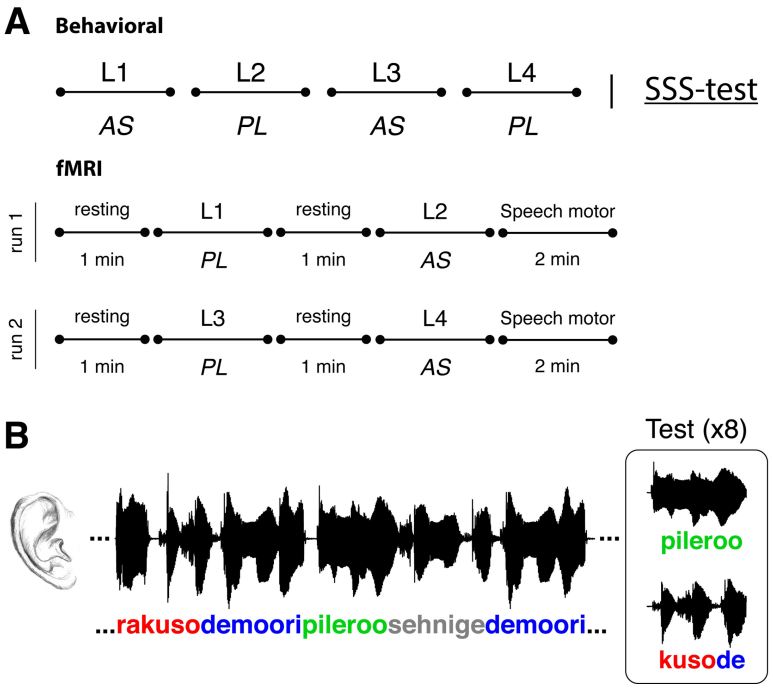

Fig 5. Overall experimental design. (A) The behavioral protocol consisted in 4 blocks of statistical word-form learning (each comprised a different *pseudo-language*) followed by the SSS-test. The statistical word-learning blocks were completed under 2 different conditions: PL, wherein participants passively listened to the *pseudo-languages* and AS, where participants concurrently, and repeatedly, whispered the syllable *"tah*". Conditions were interleaved and the order (AS-PL-AS-PL or PL-AS-PL-AS) was counterbalanced across participants. *Lower panel:* for the fMRI session, 2 speech-motor and 4 rest blocks were added to the behavioral protocol. (**B)** Schematic representation of a statistical word-learning block. *Left panel:* learning phase. Participants listened to the two-minute-long auditory stream containing the four words of the *pseudo-language*. *Right panel:* test phase. Learning was assessed after each pseudo-language exposure by an 8 trial two-alternative forced choice test contrasting a *word,* upper line, and a *part-word,* lower line.

The behavioral protocol was modified for fMRI acquisition. First, we divided the protocol into two experimental runs with one AS and one PL block each. In addition, one minute of rest was introduced before each statistical word-learning block (PL or AS) and two minutes of speech production without auditory input (speech motor condition) were introduced at the end of each run (see Fig 5A). Specifically, the speech motor condition consisted in repeatedly whispering the syllable *"tah"* with no auditory input. The SSS-test was not included in the fMRI session. Participants' speech synchronization abilities were assessed in a previous study (19).

Importantly, participants' whispered articulation was recorded during every AS block for both the behavioral and fMRI versions of the experiment (for the latter we used an MRI compatible noise cancelling microphone; OptoAcoustics FOMRI).

Stimuli

For the SSS-test and the word-learning task, we created four pseudo-languages (L1 to L4) each containing 12 distinct sets of syllables (unique consonant-vowel combinations) handpicked to maximize both between and within set variability. The syllables in each pseudo-language were combined to form four distinct tri-syllabic pseudo-words (henceforth, words). The words were relatively balanced on English word-average bi-phoneme and positional probabilities according to The Irvine Phonotactic Online Dictionary (IPhOD version 2.0; http://www.IPhOD.com/), to maximize their learnability. Words were concatenated in pseudorandom order to form auditory speech streams with no gaps between words, lasting 2 minutes each. An equal number of non-consecutive repetitions per word was ensured. For the learning test after each *pseudo-language* exposure, we created part-words by the concatenation of a word's final syllable and the first two syllables of another word of the same pseudo-language. A minute-long random syllable stream for the SSS-test was created by the random combination of a set of 12 syllables different from those

used for the pseudo-languages. The stream of syllables contained no pauses between them and no consecutive repetitions. Words, part-words and streams were converted to .wav files using the American Male Voice diphone database (US2) of the MBROLA text-to-speech synthesizer (52) at 16 kHz. All phonemes were equal in duration (111ms) - satisfying a constant syllable presentation rate of 4.5Hz, pitch (200Hz), and pitch rise and fall (with the maximum in the middle of the phoneme).

Statistical Word-Learning Task

The statistical word-learning task for each pseudo-language consisted of a learning phase, during which participants listened to the 2-minute-long streams containing the four words of the *pseudo-language* (L1 to L4); and a test phase, where each word of the *pseudo-language* was presented against four part-words (randomly selected from the pool of 12 possible part-words) in a two-alternative forced choice (see Fig 5B). During each test, words and selected part-words were presented twice each, within non-repeating pairs, making this a total of eight test trials. Test items were presented auditorily and in their written forms (left and right of the screen). Participants were required, for each test pair, to indicate their choice by pressing '1'/left or a '2'/right according to the order of auditory presentation and location on the screen. The presentation of the *pseudo-languages* was counterbalanced between participants. In order to select the best phonology to orthography matching for the visual presentation, the written renderings of all words and part-words with the highest convergence among five independent native speakers were selected.

SSS-test

Participants in a sound isolated booth and seated in front of a computer with a microphone placed close to their mouth, listened to a rhythmic stream of syllables while whispering the syllable "*tah"* concurrently and repeatedly for one minute. Next, they listened to isolated syllables and had

to indicate for each one whether they were present in the stream. Participants were not explicitly instructed to synchronize their speech to the external audio; instead, the instruction was to correctly recall the syllables. Before performing this task, participants were primed to repeatedly whisper *"tah"* at a rate of 4.5 syllables/sec; the same rate at which the external syllables were presented. For further details about the test, see (19,53).

Speech synchronization measurement

The degree of synchronization was measured by the phase locking value (PLV) between the envelope of the produced speech and the cochlear envelope (24,54) of the rhythmic syllable stream or the *pseudo-language*. The PLV was computed using the following formula:

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

where t is the discretized time, T is the total number of time points, and θ_1 and θ_2 the phase of the first and the second signals (i.e., of the envelopes of the auditory and produced speech signals). The PLV was computed for windows of 5 seconds length with an overlap of 2 seconds. The results for all time windows were averaged within each stimulus presentation, providing one PLV per block.

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.

Definition of high and low synchronizers

Each participant was classified as a low or a high synchronizer according to their corresponding speech synchronization value (Whisper to Audio PLV) obtained during the SSS-test. A threshold value was estimated from a previous dataset (19) comprising 388 PLVs obtained

with different versions of the SSS-test (Fig S1A). We applied a *k-means* clustering algorithm(55), using a squared Euclidean distance metric with 2 clusters, and computed the midpoint between the clusters' centers ($PLV_{threshold}$ = 0.49). Participants with a PLV below/above this value were classified as low/high synchronizers.

Scanning parameters

All fMRI scans took place at the NYU Center for Brain Imaging using a 3T scanner (Siemens Prisma 3T MRI scanner) and a 64-channel phased-array head coil. For the statistical word-learning fMRI task, two runs of at least 400 (note that the test is self-paced and total timings vary between participants) sequential whole-brain multi-echo echo-planar imaging (EPI) volumes were acquired (TR = 1500 ms, TE = 45 ms, flip angle = 77º, voxel size = 2.0 × 2.0 × 2.0 mm^3, 64 axial slices, acquisition size = 104×104). A high resolution T1 MPRAGE image was also acquired (TR = 2400 ms, TE = 2.24 ms, flip angle = 8º, voxel size = 0.80 × 0.80 × 0.80 mm^3, 256 sagittal slices, acquisition matrix = 320 × 300).

fMRI and ICA preprocessing

Data were preprocessed using MATLAB R2018a and the Statistical Parameter Mapping software (SPM12, Wellcome Trust Centre for Neuroimaging, University College, London, UK, www.fil.ion.ucl.ac.uk/spm/). For each subject, we first realigned the two word-learning runs to the mean image of all EPIs. The T1 was then co-registered to this mean functional image and segmented using Unified Segmentation (56). The deformation fields obtained during the segmentation step were used to spatially normalize all functional images from each run to the MNI template included in SPM12 (we maintained the original acquisition voxel size of 2.0 × 2.0 × 2.0 mm^3). Images were finally spatially smoothed with a 6 mm FWHM kernel.

We used the Group ICA of fMRI Toolbox (22) (GIFT v4.0b; http://mialab.mrn.org/software/gift/) to apply group spatial ICA to the previously preprocessed fMRI data. Based on previous research in clinical and healthy populations (14,57,58), the number of independent components to be extracted was set to 20. Note that this is a purely data-driven approach. Data were intensity normalized, concatenated and, using principal component analysis, reduced to 20 temporal dimensions. Then, this preprocessed data were fed to the *infomax* algorithm (59). The intensities of the spatial maps were in percentage of signal change after the intensity normalization, and thus no scaling was used.

To assess which of the 20 ICA networks retrieved were related to the different conditions of interest (PL and AS), both spatial and temporal classification methods were employed. First, for all subjects, the spatial map of each individual ICA network was submitted to a second-level analysis using a one sample *t*-test under SPM12 (60). We then obtained, for each network, a group map of activity that was thresholded using a $p < 0.05$ FWE-corrected threshold at the cluster level, with an auxiliary $p < 0.001$ threshold at the voxel level. Clusters with fewer than 50 voxels were not included in the analyses. We visually inspected these thresholded networks and 8 were discarded as they reflected artifacts related to movement or the presence of ventricles or blood vessels (14,58).

Using GIFT, for the remaining 12 networks, we calculated a multiple regression that fitted each participant's network time-course to a model. The model was created using SPM12 by convolving the timing of both the main (AS and PL) and control (Rest and Speech motor) conditions with a canonical hemodynamic response. To control for motion artifacts, the model included six movement regressors obtained from the realignment step. The test phase was also included in the GLM model as a separate nuisance condition. Therefore, the beta values for the

AS and PL conditions were computed from the part of the fMRI signal pertaining to the listening blocks (i.e., without the testing phase, which was modeled separately). By fitting a multiple regression between this model and each network's time-course, we obtained, for each condition, beta values that represented network engagement. For PL beta values, we used the Rest condition as a baseline. For AS we used both the rest and the speech motor control conditions as baseline to capture the activity related to the learning process during AS itself and not to the motor-activity related to the whispering. For any comparison using beta values, participants exceeding 2 SD were excluded from the analysis.

For the group spatial maps of each network, maxima and all coordinates are reported in MNI space. Anatomical and cytoarchitectonical areas were identified using the Automated Anatomical Labeling (61) and the Talairach Daemon(62) database atlases included in the xjView toolbox (http://www.alivelearn.net/xjview/).

Statistical Analyses

Group analyses were performed on the proportion of correct responses averaged across same-type conditions (PL and AS). To test for differences between learning conditions and groups, we performed generalized linear mixed modelling in *R* (version 4.0.2) and *RStudio* (version 1.3.959) using the *lme4* package (63). The dependent variable (responses to the learning tests) was assumed to have a binomial distribution and a logit link function was applied. An initial model included Condition (AS, PL), Group (High, Low) and their interaction as predictors. This model also included Subjects as a random effects factor to allow for varying intercepts between subjects. This was compared to alternative models with additional random effects factors Order, Language, and Trial number. These factors were removed, keeping the initial model, because the increase in variance explained by these more complex models was in all cases negligible. Akaike Information

Criterion (AIC) was used for this assessment, thus selecting the model with the best balance between goodness of fit and complexity. The effects of the different predictors and their interactions on learning performance were assessed by means of likelihood ratio tests using the *afex* package (64) in *R*. These tests were based on Type 3 sums of squares. Following a significant interaction between Group and Condition, we estimated marginal means, using the *emmeans* package in R, of participants' performance within each group (Highs, Lows) for the PL and AS conditions. Where specified, we additionally used non-parametric Mann-Whitney-Wilcoxon and Wilcoxon signed-rank tests for between and within subject comparisons, respectively. Multiple comparisons were controlled using a False-Discovery Rate correction. Non-parametric Spearman's rank correlations were used to assess the relationship between variables. Bayes factors (BF_{01}), which reflect how likely data are to arise from the null model (i.e., the probability of the data given H0 relative to H1), were also computed with the software JASP using default priors (65–67).

REFERENCES

1. Gervain J, Macagno F, Cogoi S, Pena M, Mehler J. The neonate brain detects speech structure. Proc Natl Acad Sci. 2008;105(37):14222–7.

2. Bulf H, Johnson SP, Valenza E. Visual statistical learning in the newborn infant. Cognition. 2011;121(1):127–32.

3. Saffran JR, Aslin RN, Newport EL. Statistical Learning by 8-Month-Old Infants. Science. 1996;274(5294):1926–8.

4. Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tone sequences by human infants and adults. Cognition. 1999;70:27–52.

5. Fiser J, Aslin RN. Statistical learning of higher-order temporal structure from visual shape sequences. J Exp Psychol Learn Mem Cogn. 2002;28(3):458–67.

6. Conway CM, Christiansen MH. Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. J Exp Psychol Learn Mem Cogn. 2005;31(1):24–39.

7. Toro JM, Trobalón JB. Statistical computations over a speech stream in a rodent. Percept Psychophys. 2005;67(5):867–75.

8. Meyer T, Olson CR. Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci. 2011;108(48):19401–6.

9. Lopez-Barroso D, De Diego-Balaguer R, Cunillera T, Camara E, Münte TF, Rodriguez-Fornells A. Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. Cereb Cortex. 2011;21(12):2742–50.

10. Estes KG, Evans JL, Alibali MW, Saffran JR. Can Infants Map Meaning to Newly Segmented Words? Statistical Segmentation and Word Learning. 2007.

11. Saffran JR, Wilson DP. From syllables to syntax: Multilevel statistical learning by 12-month-old infants. Infancy. 2003;4(2):273–84.

12. McNealy K, Mazziotta JC, Dapretto M. Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. J Neurosci. 2006;26(29):7629–39.

13. Cunillera T, Càmara E, Toro JM, Marco-Pallares J, Sebastián-Galles N, Ortiz H, et al. Time course and functional neuroanatomy of speech segmentation in adults. NeuroImage. 2009;48(3):541–53.

14. López-Barroso D, Ripolles P, Marco-Pallares J, Mohammadi B, Munte TF, Bachoud-Levi AC, et al. Multiple brain networks underpinning word learning from fluent speech revealed by independent component analysis. NeuroImage. 2015;110:182–93.

15. López-Barroso D, Catani M, Ripollés P, Dell'Acqua F, Rodríguez-Fornells A, de Diego-Balaguer R, et al. Word learning is mediated by the left arcuate fasciculus. Proc Natl Acad Sci U S A. 2013;110(32):13168–73.

16. Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D. The neural correlates of statistical learning in a word segmentation task: An fMRI study. BRAIN Lang. 2013;127(1):1–21.

17. Henin S, Turk-browne N, Friedman D, Dugan P, Doyle W, Devinsky O, et al. Statistical learning shapes neural sequence representations. 2019;

18. López-Barroso D, Catani M, Ripollés P, Dell'Acqua F, Rodríguez-Fornells A, de Diego-Balaguer R, et al. Word learning is mediated by the left arcuate fasciculus. Proc Natl Acad Sci U S A. 2013;110(32):13168–73.

19. Assaneo MF, Ripollés P, Orpella J, Lin WM, de Diego-Balaguer R, Poeppel D. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat Neurosci. 2019;22(4):627–32.

20. Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007;8(5):393–402.

21. Baddeley A, Lewis V, Vallar G. Exploring the Articulatory Loop. Q J Exp Psychol Sect A. 1984;

22. Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. Hum Brain Mapp. 2001;14:140–51.

23. Pellegrino F, Coupé C, Marsico E. Across-Language Perspective on Speech Information Rate. Language. 2011;87(3):539–58.

24. Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. Neurosci Biobehav Rev. 2017;81:181–7.

25. Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C. A cross-linguistic study of speech modulation spectra. J Acoust Soc Am. 2017;142(4):1976–89.

26. Assaneo MF, Poeppel D. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. Sci Adv. 2018;4(2):1–10.

27. Assaneo MF, Rimmele JM, Sanz Perl Y, Poeppel D. Speaking rhythmically can shape hearing. Nat Hum Behav. 2020;

28. Plante E, Patterson D, Gómez R, Almryde KR, White MG, Asbjørnsen AE. The nature of the language input affects brain activation during learning from a natural language. J Neurolinguistics. 2015;36:17–34.

29. Thomson JM, Fryer B, Maltby J, Goswami U. Auditory and motor rhythm awareness in adults with dyslexia. J Res Read. 2006;

30. Tierney AT, Kraus N. The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. Brain Lang. 2013;124(3):225–31.

31. Thomson JM, Goswami U. Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. J Physiol Paris. 2008;

32. Woodruff Carr K, White-Schwoch T, Tierney AT, Strait DL, Kraus N. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. Proc Natl Acad Sci. 2014;111(40):14559–64.

33. Poeppel D, Assaneo MF. Speech rhythms and their neural foundations. Nat Rev Neurosci. 2020 Jun;21(6):322–34.

34. Rimmele JM, Morillon B, Poeppel D, Arnal LH. Proactive Sensing of Periodic and Aperiodic Auditory Patterns. Trends Cogn Sci. 2018;22(10):870–82.

35. Park H, Ince RAA, Schyns PG, Thut G, Gross J. Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. Curr Biol. 2015;25(12):1649–53.

36. Tierney A, Kraus N. Auditory-motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). Front Hum Neurosci. 2014;8(NOV):1–9.

37. Conway CM. How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. Neurosci Biobehav Rev. 2020;112(January):279–99.

38. Orpella J, Ripollés P, Ruzzoli M, Amengual JL, Callejas A, Martinez-Alvarez A, et al. Integrating when and what information in the left parietal lobe allows language rule generalization. PLOS Biol. 2020 Nov 2;18(11):e3000895.

39. Corbetta M, Patel G, Shulman GL. The Reorienting System of the Human Brain: From Environment to Theory of Mind. Neuron. 2008 May 8;58(3):306–24.

40. Dosenbach NUF, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. Trends Cogn Sci. 2008 Mar 1;12(3):99–105.

41. Dorsal and Ventral Attention Systems: Distinct Neural Circuits but Collaborative Roles - Simone Vossel, Joy J. Geng, Gereon R. Fink, 2014 [Internet]. [cited 2022 May 12]. Available from: https://journals.sagepub.com/doi/full/10.1177/1073858413494269

42. Siegelman N, Bogaerts L, Christiansen MH, Frost R. Towards a theory of individual differences in statistical learning. Philos Trans R Soc B Biol Sci. 2017 Jan 5;372(1711):20160059.

43. Marek S, Dosenbach NUF. The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. Dialogues Clin Neurosci. 2018 Jun 30;20(2):133–40.

44. Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS. Multi-task connectivity reveals flexible hubs for adaptive task control. Nat Neurosci. 2013 Sep;16(9):1348–55.

45. Zanto TP, Gazzaley A. Fronto-parietal network: flexible hub of cognitive control. Trends Cogn Sci. 2013 Dec 1;17(12):602–3.

46. Albouy P, Weiss A, Baillet S, Zatorre RJ. Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. Neuron. 2017 Apr 5;94(1):193-206.e5.

47. Albouy P, Benjamin L, Morillon B, Zatorre RJ. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science. 2020 Feb 28;367(6481):1043–7.

48. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 2002 Mar;3(3):201–15.

49. Smalle EHM, Daikoku T, Szmalec A, Duyck W, Möttönen R. Unlocking adults' implicit statistical learning by cognitive depletion. Proc Natl Acad Sci. 2022 Jan 11;119(2):e2026011119.

50. Smalle EHM, Muylle M, Duyck W, Szmalec A. Less is more: Depleting cognitive resources enhances language learning abilities in adults. J Exp Psychol Gen. 2021;150(12):2423–34.

51. Baddeley A, Lewis V, Vallar G. Exploring the Articulatory Loop. Q J Exp Psychol Sect A. 1984 May 1;36(2):233–52.

52. Dutoit T, Pagel V. Le projet MBROLA : Vers un ensemble de synthétiseurs vocaux disponibles gratuitement pour utilisation non-commerciale. In: XXIèmes Journées d'Etude sur la Parole. 1996. p. 441–4.

53. Lizcano-Cortés F, Gómez-Varela I, Mares C, Wallisch P, Orpella J, Poeppel D, et al. Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers. STAR Protoc. 2022 Jun 17;3(2):101248.

54. Chi T. SS. NSL Matlab Toolbox. Md Neural Syst Lab Univ Md.

55. Arthur D, Vassilvitskii S. k-means++: The Advantages of Careful Seeding. Proc Eighteenth Annu ACM-SIAM Symp Discrete Algorithms. 2007;1–11.

56. Ashburner J, Friston KJ. Unified segmentation. NeuroImage. 2005;26(3):839–51.

57. Sihvonen AJ, Särkämö T, Ripollés P, Leo V, Saunavaara J, Parkkola R, et al. Functional neural changes associated with acquired amusia across different stages of recovery after stroke. Sci Rep. 2017;

58. Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, et al. Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci. 2009;

59. Bell AJ, Sejnowski TJ. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Comput. 1995;7(6):1129–59.

60. Calhoun VD, Adali T, McGinty VB, Pekar JJ, Watson TD, Pearlson GD. fMRI activation in a visual-perception task: Network of areas detected using the general linear model and independent components analysis. NeuroImage. 2001;

61. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002;15(1):273–89.

62. Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, et al. Automated Talairach Atlas labels for functional brain mapping. Hum Brain Mapp. 2000;

63. Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. Journalo F Stat Softw. 2015;67(1):1–47.

64. Singmann H, Bolker B, Westfall J, Aust F. afex: Analysis of Factorial Experiments.

65. JASP Team. JASP (Version 0.9)[Computer software]. [Computer software]. 2018.

66. Morey RD, Rouder JN, Jamil T, Urbanek S, Forner K, Ly A. Package 'BayesFactor'. Compr R Arch Netw. 2018;

67. Rouder JN, Morey RD. Default Bayes Factors for Model Selection in Regression. Multivar Behav Res. 2012;

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health (https://www.nih.gov/) grant 2R01DC05660 (D.P.), National Science Foundation (https://www.nsf.gov/) Grant 2043717 (P.R., D.P., J.O. and M.F.A.), and by the European Research Council (https://erc.europa.eu/) grant ERC-StG-313841 (TuningLang, RdD-B) and BFU2017-87109-P Grant from the Spanish Ministerio de Ciencia e Innovacion (https://www.ciencia.gob.es/)(RdD-B), which is part of Agencia Estatal de Investigacion (AEI) (Co-funded by the European Regional Development Fund. ERDF, a way to build Europe). We also thank the CERCA Program / Generalitat de Catalunya for the institutional support. DL-B was supported by a travel grant from the University of Malaga and the Ramón y Cajal program (RYC2020-029495-I) funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/10.13039/50110001103) and FSE invierte en tu futuro. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

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

COMPETING INTERESTS

The authors declare no competing interests.

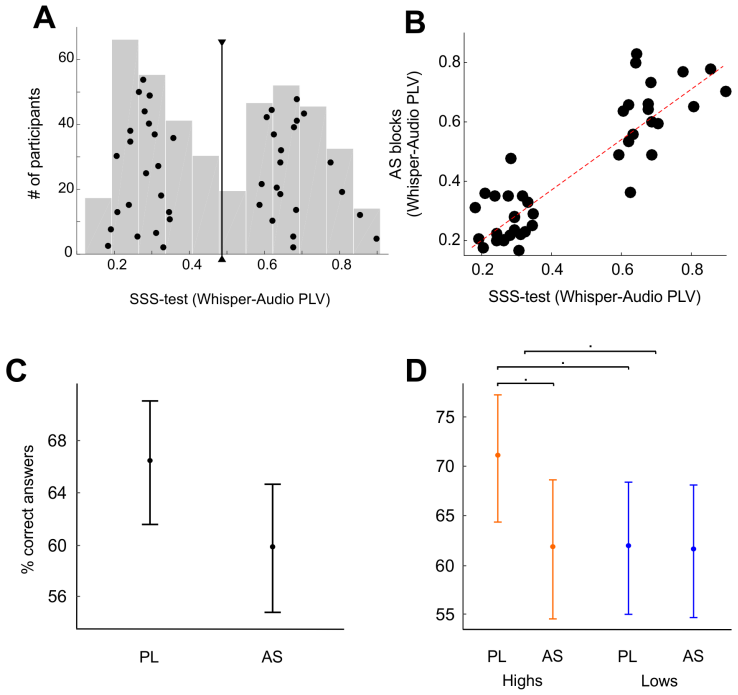

Fig S1. Behavioral performance in the scanner. (A) SSS-test outcome. Histogram of 388 PLVs obtained in a previous work[1] with two different versions of the SSS-test. Black dots represent the participants selected to complete the fMRI protocol. Black line represents the threshold value adopted in this work to separate high and low synchronizers: $PLV_{threshold}$ = 0.49. A *k-means* clustering algorithm using a squared Euclidean distance metric was applied over this distribution (N = 388). The threshold value is the midpoint between the two clusters' centers. **(B)** Scatterplot displaying participants' PLV during AS inside the scanner as a function of the PLV from the SSS-test. Red line represents the correlation of the data. The correlation is displayed for visualization purposes, to emphasize that the synchronization of low synchronizers is consistently worse than that of highs during the AS block. The correlation within groups remains significant only for high synchronizers (r_{HIGH}=0.45 p_{HIGH}=0.044; r_{LOW}=0.21 p_{LOW}=0.31). **(C)** Percentage of correct responses for the statistical word-learning task during PL and AS conditions inside the scanner on the entire sample. **(D)** Percentage of correct responses for the statistical word-learning task during PL and AS conditions inside the scanner for the low (blue color) and the high (orange color) synchronizers. The mixed-model analysis of this dataset yielded a significant difference between conditions (main effect of Condition (PL > AS), $\chi2$ = 5.40, $p < 0.05$)), and a main effect of group close to significance (Highs > Lows; $\chi2$ = 3.67, p = 0.055), and a trending

Condition*Group interaction ($\chi 2$ = 2.74, p = 0.098). Dots: model predicted group means. Bars: 95% confidence interval. AS: Articulatory Suppression; PL: Passive Listening.

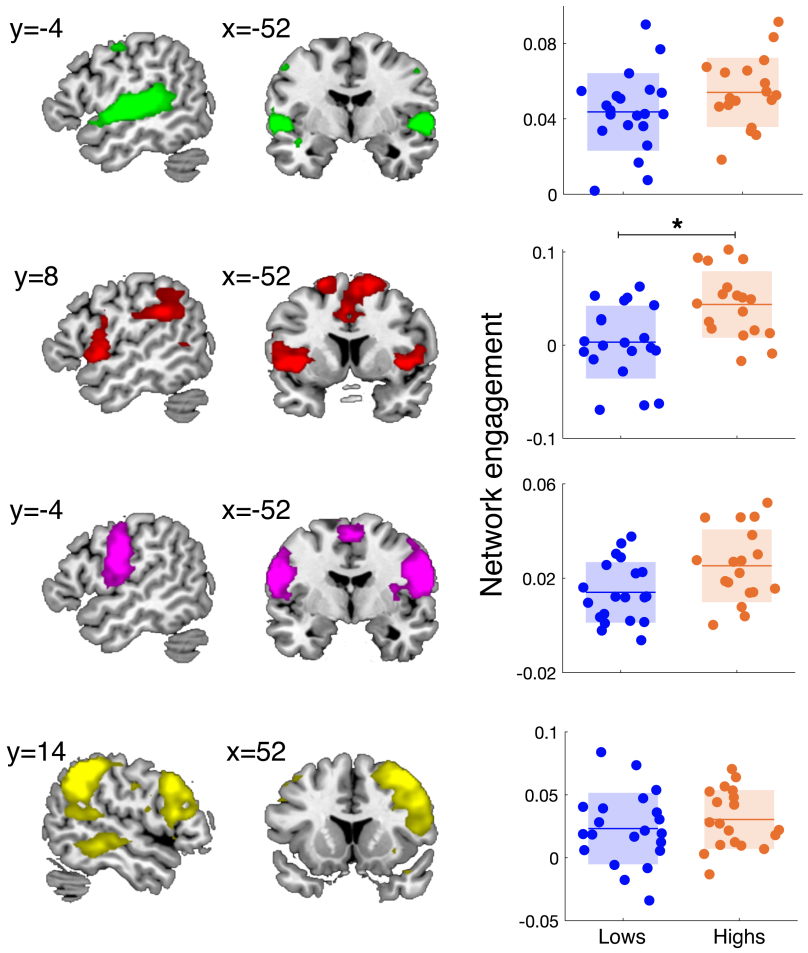

Fig S2. Brain networks significantly activated during PL. The different networks are shown over a canonical template with MNI coordinates on the upper portion of each slice. Neurological convention is used with a $p < 0.05$ FWE-corrected threshold at the cluster level and an auxiliary $p < 0.001$ threshold at the voxel level. In addition to the auditory (green) and fronto-parietal (red) networks described in the main manuscript, a sensorimotor (magenta) and a right lateralized fronto-temporo-parietal (yellow) networks were also activated during PL. All these networks were significantly activated during PL for both high and low synchronizers, except the fronto-parietal, which was only active for the high. *$p < 0.05$ Mann-Whitney-Wilcoxon between-group comparison, FDR corrected.

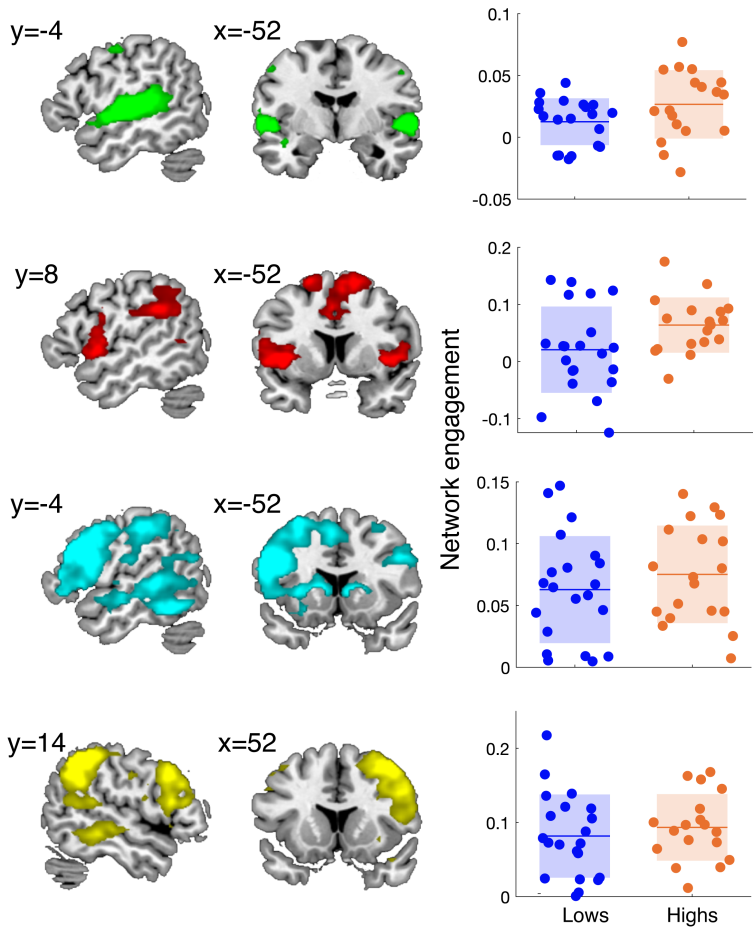

Fig S3. Brain networks significantly activated during AS. The different networks are shown over a canonical template with MNI coordinates on the upper portion of each slice. Neurological convention is used with a $p < 0.05$ FWE-corrected threshold at the cluster level and an auxiliary $p < 0.001$ threshold at the voxel level. In addition to the auditory (green) and fronto-parietal (red) networks described in the main manuscript, a left (light blue) and a right (yellow) lateralized fronto-temporo-parietal network were also activated during AS. All networks were significantly activated during AS for both high and low synchronizers, except for the fronto-parietal which was only active for the high and the auditory, which was marginally significant for the lows ($p_{unc} = 0.03$).