Statistical learning shapes neural sequence representations

Simon Henin[1,2*], Nicholas Turk-Browne[3], Daniel Friedman[1,2], Anli Liu[1,2], Patricia Dugan[1,2], Adeen Flinker[1,2], Werner Doyle[1,2], Orrin Devinsky[1,2], Lucia Melloni[1,2,4*]

1. New York University Comprehensive Epilepsy Center, 223 34th Street, New York, NY 10016, USA.
2. Department of Neurology, New York University School of Medicine, 240 East 38th Street, 20th Floor, New York, NY 10016, USA.
3. Department of Psychology, Yale University.
4. Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322, Frankfurt am Main, Germany.

*Address for correspondence
Simon Henin (simon.henin@nyumc.org) and Lucia Melloni (lucia.melloni@ae.mpg.de)

ABSTRACT

Although sensory input arrives continuously, we experience the world in discrete units, consisting of words, objects, scenes and events, which in turn form the basis of language, thought and memory. How does sensory input get parsed into meaningful units? A process known as statistical learning (SL) may underlie this ability. SL is ubiquitous, for example allowing babies to discover word boundaries in continuous speech by tracking transitional probabilities between syllables. Here we examine which cortical circuits extract such regularities, how these regularities are represented, and how this learning compares across sensory modalities. We exposed subjects to auditory and visual sequences containing temporal regularities while collecting direct, intracranial recordings (23 patients, 3689 electrodes). We used neural frequency tagging to first map the cortical circuits for SL and then representational similarity analysis to determine which aspect(s) of the regularities are learned. SL manifested into two distinct ways across electrodes, differing in terms of both anatomical location and hierarchical organization: one cluster of electrodes located in earlier processing stages (e.g., superior temporal gyrus, STG) represented both the constituent elements (e.g., syllables) and learned higher-order units (e.g., words); the other cluster was localized to later processing stages (e.g., inferior frontal gyrus, IFG) and represented only the higher-order units. Within these regions, SL shaped the similarity of neural representations at multiple levels, with a division of labor between earlier vs. later brain areas in terms of encoding of simple generic aspects of the sequences (i.e., transitional probability) vs. complex and specific information (i.e., ordinal position, identity). The anatomical and representational segregation of these circuits was observed for SL in both the auditory and visual modality, yet the anatomical areas (with the exception of IFG and anterior temporal pole), showed specificity to modality. These findings indicate the existence of multiple computational systems for sequence processing supporting learning across hierarchically segregated cortical circuits.

INTRODUCTION

We receive continuous input from the world and yet experience it in digestible chunks. In the domain of language, for example, acquisition and use require extracting meaningful sequences such as words, phrases and sentences out of a continuous stream of sounds, often without clear acoustic boundaries or pauses between linguistic elements[1]. This segmentation ability occurs incidentally and effortlessly, and is thought to be a core building block of development. Indeed, young infants can capitalize on transitional probabilities between syllables[2] or shapes[3] to extract embedded regularities after minimal exposure. Similar abilities have been demonstrated in adults[4, 5], who rely on transitional probabilities and other statistical properties[6-9] to detect repeating patterns in the temporal and spatial co-occurrences of stimuli. This behavior known as "statistical learning" (SL) occurs across many different stimuli, sensory modalities, ages, tasks, and even species[10]. SL represents a fundamental behavioral ability with cascading effects for cognition, and yet the brain mechanisms that give rise to this ability have only begun to be investigated.

Brain systems such as the hippocampus and the inferior frontal gyrus have been implicated in both visual[11-14] and auditory SL[12, 15-17]. However, most studies have compared brain changes before and after learning, focusing on the *outcome* of learning[18], principally because of challenges associated with analyzing continuous and dynamic data streams, especially with fMRI. Some studies have attempted to explore the *acquisition* of auditory regularities using neural measures[19-21], and yet left open several important questions: (1) Which brain areas are involved in detecting and representing regularities? (2) Which aspects of the regularities are represented — e.g., transitional probabilities, ordinal position and/or word identity [22]? (3) How are these brain areas similar or different across modalities[23]?

To answer these questions, we collected intracranial recordings (ECoG) from 23 human epilepsy patients with broad cortical coverage during an SL task. We used neural frequency tagging (NFT)[19, 24, 25] as an online measure of SL because it is designed for continuous stimuli and can separate neural processes operating over different timescales (e.g., syllables and words), allowing us to track the acquisition of regularities as SL evolves. Combining ECoG and NFT, we precisely localized *where* SL takes place in the brain and demonstrate *how* this learning emerges over time. We used representational similarity analysis (RSA) to determine *which* aspect(s) of temporal regularities are represented. Finally, we related the neural circuits, online dynamics and representational changes for SL across auditory and visual modalities.

We found that SL occurs quickly in both auditory and visual modalities and can be tracked precisely with NFT. In both modalities, partly overlapping neural circuits encoded statistical units (e.g., words) and their constituent sensory elements (e.g., syllables). This learning was supported by rapid changes in the similarity space of neural representations, with information about the regularities encoded at multiple levels: transitional probabilities, with elements grouped by probability strength; ordinal position, with elements grouped by their position in a regular sequence; and identity, with elements grouped by the specific regularity in which they are embedded. Auditory and visual

elements undergo similar SL-related representational changes, yet the areas affected overlap only partially (generally in supramodal areas, such as inferior frontal gyrus, anterior temporal pole and the hippocampus). These results provide mechanistic insight into a fundamental human learning ability, reveal the plasticity of cortical areas to be shaped by the structure of the world, and highlight NFT as a versatile tool for investigating incidental learning that could prove useful in preverbal infants and other non-verbal patient populations.

RESULTS

Neural tracking of auditory statistical learning

We exposed 17 participants to brief (2 min x 5 exposures) auditory streams of syllables while manipulating the structure of the sequence. In structured streams, each syllable was placed into the first, second, or third position of a three-syllable word or "triplet" (**Fig. 1a**). A continuous stream of syllables was generated by randomly inserting multiple repetitions of each word without breaks or prosodic cues between words. In random streams, syllables were inserted the same number of times but in a random order at the syllable level. As a result, the TPs were uniform and there were no words to segment. Participants were not informed about whether structure was present. Instead, they were asked to perform a 1-back cover task, in which they had to detect occasional repetitions of individual syllables that had been inserted into both stream types[26]. Performance on the cover task showed that participants attended to the stimuli across both conditions ($d'_{structured}$ = 1.34, $t(16)$ =10.72, $p<0.01$; d'_{random}=0.89, $t(16)$=9.88, $p<0.01$), with better detection of the syllable repetitions during the structured than the random condition ($t(16)$ = 5.38, $p<0.01$).

If participants learned the hidden regularities of the structured stream and used them to segment the stream into words, such learning should be reflected in the temporal dynamics of neural activity. Specifically, we used NFT to decompose representations of segmented units at two hierarchical levels of the stream[19, 20, 24, 25]: oscillations at the 4 Hz frequency — corresponding to individual syllables — should be present throughout exposure to both structured and random streams, given the different acoustic properties of each syllable; critically, oscillations at the 1.33 Hz frequency corresponding to triplet boundaries should emerge over exposure to the structured but not random streams, because they can only be extracted by picking up on higher within vs. across word TPs across repetitions in the structured stream.

We obtained direct neurophysiological signals from 1,898 intracranial electrodes in the 17 participants, comprehensively covering the frontal, parietal, occipital, and temporal lobes in both hemispheres (see **Supplementary Fig. S1** for reconstruction). To assess phase-locked neural activity across frequencies, we first evaluated within-electrode phase coherence [24] in the structured and random streams, respectively. Consistent with our hypothesis, there was a significant peak in the phase coherence spectrum at the syllable rate (i.e., 4 Hz) for both structured and random streams ($p<0.05$, FDR corrected). In addition, there was a significant peak at the word rate (i.e., 1.33 Hz), but only for the structured stream ($p<0.05$, FDR corrected) (**Fig. 1b**). Coherence at the word rate in the structured stream was highly replicable across participants with 16/17 patients exhibiting significant entrainment in at least one electrode; by contrast, no electrode showed

entrainment at the word rate for the random stream. Interestingly, there was also a significant phase coherence peak at 2.66 Hz for the structured stream ($p<0.05$, FDR corrected). This may reflect an oscillation at the rate of syllable *pairs*, consistent with evidence that participants can learn sequential pairs embedded in triplets, in addition to the triplets themselves[27]; alternatively, this may be a harmonic of the word rate. This neural tracking seems to have reflected implicit learning, as participants were unable to explicitly discriminate words vs. non-words in a recognition test administered after exposure (mean accuracy = 44.6%, $t(14) = -1.94$, $p = 0.08$). This finding supports NFT as a sensitive tool for assessing online statistical learning.

To examine the timecourse of SL, we calculated phase coherence at the word rate over smaller temporal windows of data (~60 s) concatenating the five exposures. In the structured stream, the word-rate response emerged rapidly within 260 seconds ($t(191) = 2.10$, $p=0.005$) and progressively increased over time (**Fig. 1c**). No increase in the strength of the word-rate response was observed for the random stream, ruling out effects of endogenous entrainment over time unrelated to learning.

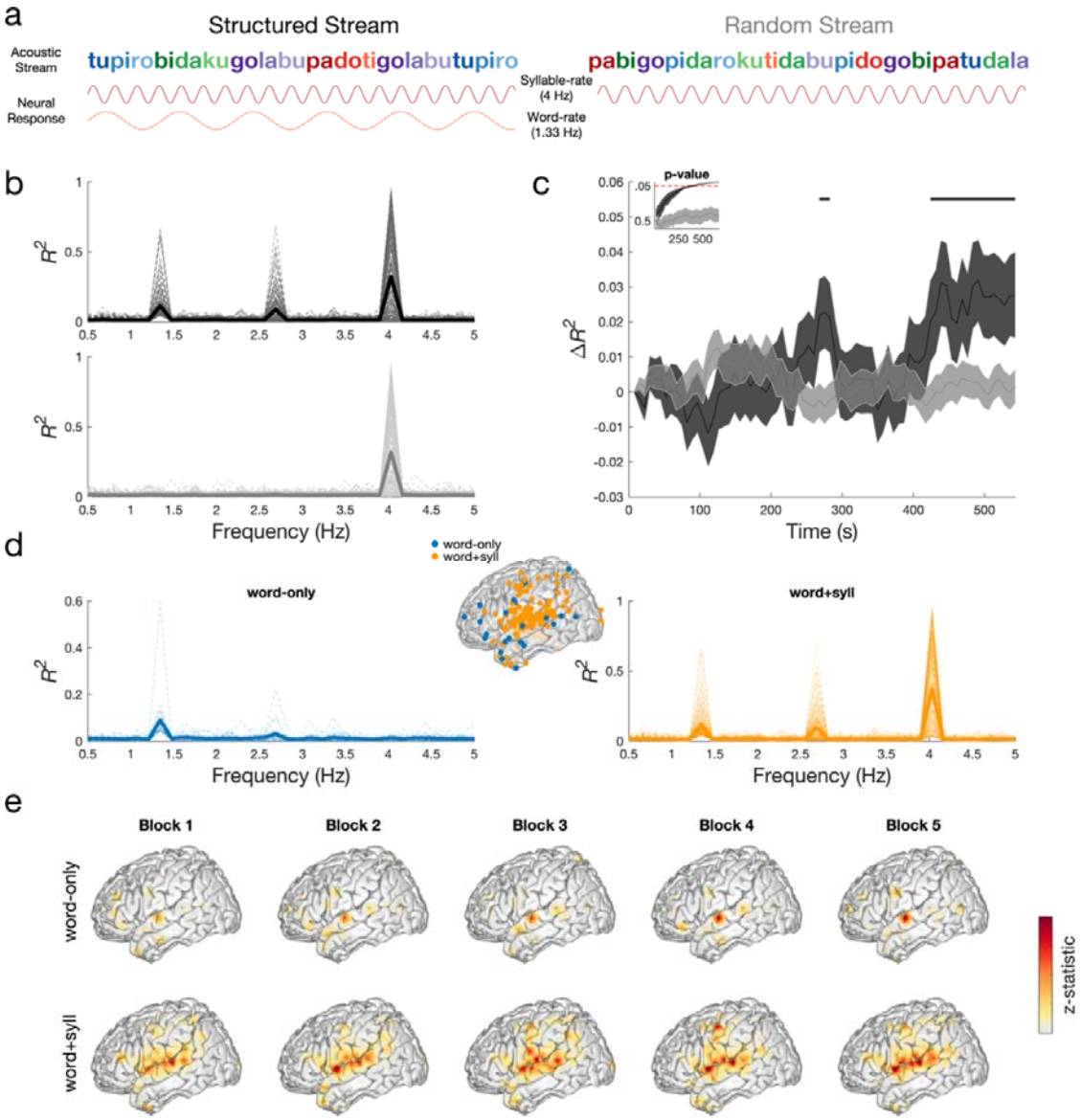

Figure 1 Neural tracking of auditory statistical learning **(a)** Schematic depiction of the auditory statistical learning task. The structured stream (left) contained 12 syllables (250 ms SOA, 4 Hz) in which the transitional probabilities formed four words (color coded for visualization, 750 ms SOA, 1.33 Hz). The random stream (right) contained the same 12 syllables in a random order. The predicted neural response is shown below each syllable stream: syllable tracking (top) was expected in both conditions, whereas word tracking (bottom) was expected only in the structured condition. **(b)** Phase coherence spectrum in neural data for the structured (top, black) and random (bottom, gray) conditions from 1,898 electrodes in 17 patients. Each electrode is depicted with a thin line and the average with a thick line. **(c)** Timecourse of coherence response for all electrodes showing word-rate tracking (1.33Hz) for the structured (black) and random (gray) conditions. Solid lines above the timecourse indicates when the change in coherence increased above zero ($p<0.025$, one-sided t-test). Inset shows accumulated p-value of the responses over time showing when temporal coherence reached significance at 0.05 level (red dashed line). **(d)** Phase coherence spectrum in the structured condition for electrodes showing word-tracking responses, in two groups: electrodes that tracked words only (left, blue) and electrodes that tracked both words and syllables (right, orange). Inset shows the localization of word-only (blue) and word+syll (orange) electrodes. **(e)** Localized word-rate coherence in the structured condition across each of the 5 exposure blocks for each electrode type (top: word-only; bottom: word+syll), showing the evolution of the coherence response across brain regions.

We exploited the excellent spatial resolution afforded by ECoG to examine the neural circuits responsible for SL. Specifically, we localized which cortical areas became synchronized to the word rate in the structured condition (**Fig. 1d**). Some electrodes that tracked words also tracked syllables (word+syll), and these were located primarily in the superior temporal gyrus (STG), with smaller clusters in motor cortex and pars opercularis. Other electrodes tracked the words only (word-only), and these were located in inferior frontal gyrus (IFG) and anterior temporal lobe (ATL). In other words, these areas encoded higher-order structure without responding to the syllabic-level acoustic features of the materials conveying the structure. This organization resembles the anatomical structure of the auditory processing hierarchy, with lower areas in STG and hierarchically higher areas in surrounding fronto- and temporo-parietal cortex[28]. Thus, word-only responses may arise from higher-level stages of processing than word+syll responses.

We investigated the timecourse of learning for both types of word response across cortical areas by dividing the structured stream into five blocks (**Fig. 1e**). The topographies were relatively stable over time (Intra-class correlation; $ICC_{word\text{-}only} = 0.51$, $ICC_{word+syll} = 0.45$) but increased in strength as a function of exposure, suggesting that learning of higher-order structure can lead to rapid changes in these two separate networks.

The same two types of responses were observed when restricting analysis to the high-gamma band (HGB), thought to mirror multiunit firing[29]. Topographies for the two types resemble those described above for raw LFP, though significant responses were sparser and less numerous (**Supplementary Fig. S2**). This replicates findings that the high-gamma response is more spatially confined compared to the LFP[30, 31].

Representational analysis in auditory statistical learning

The results so far provide compelling evidence of segmentation of the continuous auditory stream. But what is driving this segmentation? It could be based on at least three statistical cues in the stream: transitional probabilities (within word, 1.0; between word, 0.33), ordinal position (1st, 2nd, 3rd position), or word identity (sequential concatenation of specific syllables into four words). Although all three cues could be used to mark the start and end of the words, only the third gives access to individual words which can then be mapped onto meaning[32].

To test the role of these cues in SL, we employed a multivariate pattern similarity approach. In the case of word identity, for example, we reasoned that SL would change the representational space of stimuli such that syllables belonging to the same word would evoke more similar neural activity patterns across electrodes[11]; this clustering could in turn provide a basis for segmentation[7]. Alternatively, the neural representations of syllables may cluster by ordinal position or transitional probability, allowing us to test which of these cues was learned. We quantified the representational space of syllables from raw LFP, separately over electrodes showing word-only and word+syll coherence. Specifically, we calculated the cosine distance between the spatial patterns of neural activity across electrodes for every pair of syllables.

We found that the two types of electrodes encoded different cues: For word+syll electrodes, multidimensional scaling of the distances between syllables revealed a representation of transitional probabilities **(Fig. 2b)**, grouping syllables based on whether their probability given the preceding syllable was low (1st) vs. high (2nd, 3rd). In contrast, word-only electrodes represented ordinal position **(Fig. 2a),** grouping syllables based on which position they occupied in the words (1st vs. 2nd vs. 3rd), while not distinguishing words themselves. We performed the same analysis on the raw LFP but restricted to the subset of electrodes showing word-rate responses in the HGB (**Supplementary Fig. S2**). We found a similar pattern of clustering for word+syll electrodes, with grouping only by transitional probabilities (**Fig. 2d**) and for word-only electrodes, with grouping based on ordinal position (**Fig. 2c,** dimension 1). Remarkably, word-only electrodes also showed grouping by word identity (**Fig. 2c,** dimension 2). No clustering of responses in any of the electrode types was observed when performing the same analysis on the random stream (**Supplementary Fig. S3**), indicating that changes in representational space resulted from SL and do not reflect properties of the stimuli per se.

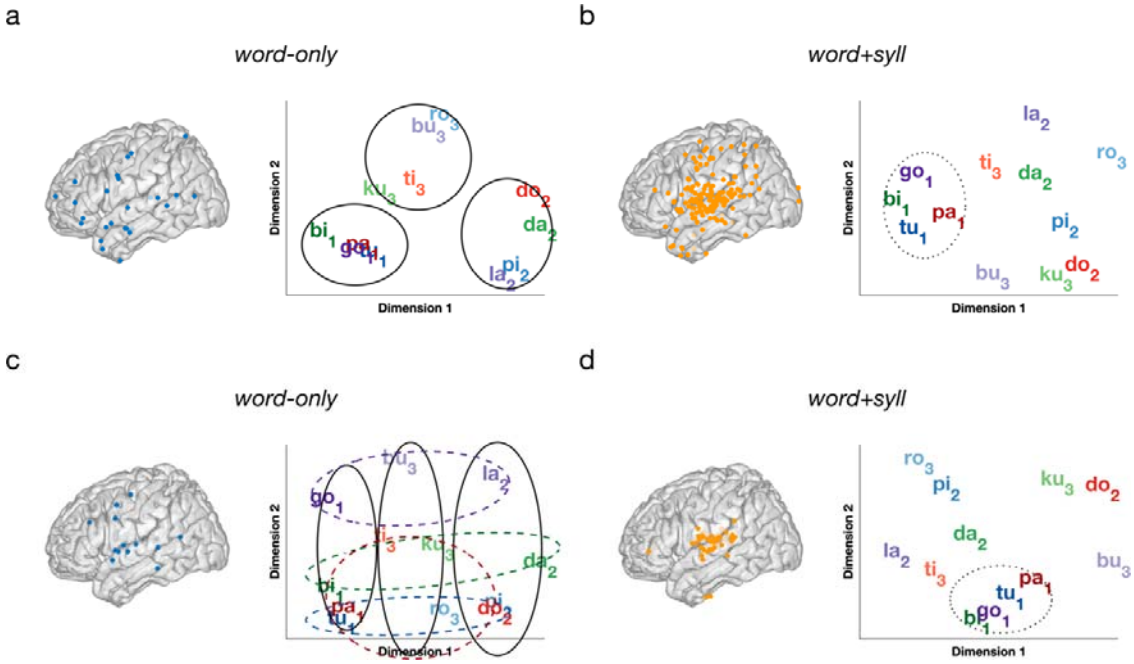

Figure 2 Pattern similarity results on raw LFP during auditory statistical learning. **Top row:** Multidimensional scaling (MDS) representation of distances between syllabic responses for electrodes showing significant LFP coherence for (a)

word-only and (b) word+syll. Individual words are color-coded, subscripts represent ordinal position (e.g., "$tu_1pi_2ro_3$") **Bottom row:** MDS representation on raw LFP restricted to electrodes showing significant HGB coherence at the (c) word-only and (d) word+syll rates. Dotted ellipses outline grouping by transitional probability in word+syll electrodes (Right panels). Solid ellipses outline grouping by ordinal position in word-only electrodes (Left panels). Dashed ellipses indicate grouping at the level of the individual words (color coded) for word-only electrodes showing segmental responses in the HGB.

To evaluate these different cues statistically, we collapsed pattern similarity across syllables of the same class, i.e., similarity of tokens with low vs. high transitional probability, same vs. different ordinal position, and same vs. different word identity, for both electrode types (word-only, word+syll) and both selection procedures (LFP, HGB). Consistent with the MDS, word+syll electrodes showed greater similarity for syllables with low transitional probability (**Fig. 3, top**). Word-only electrodes showed reliable coding for ordinal position **(Fig. 3, middle)** and for word identity only in electrodes showing segmental responses in the HGB **(Fig. 3, bottom)**. These results show that even brief exposure to auditory regularities can reshape the representational space of syllables, giving rise to clustered neural representations along several dimensions. That is, learning of sequences shapes representations at multiple levels concurrently, with a division of labor across lower- and higher-order brain areas in terms of simple and generic vs. complex and specific regularities.

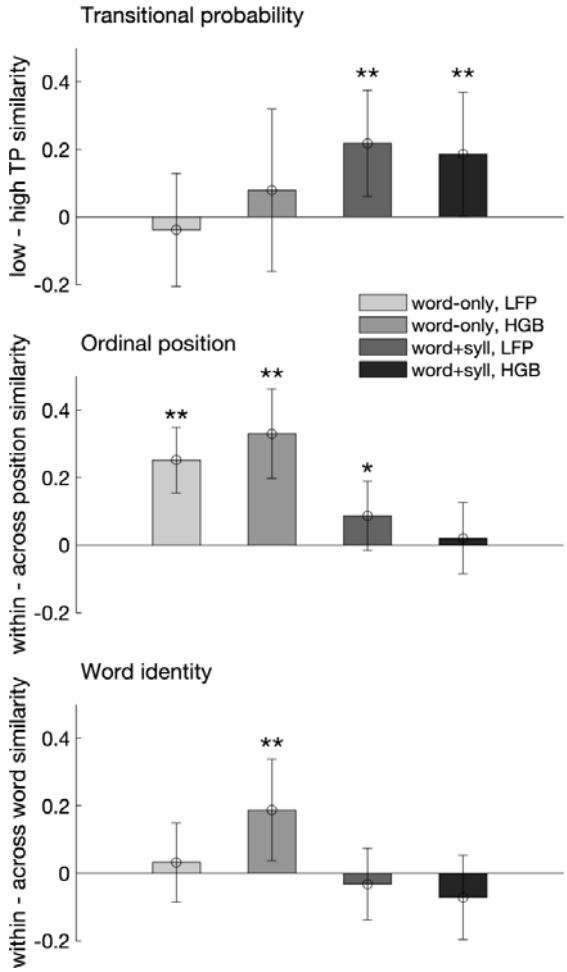

Figure 3. Comparison of multivariate similarity for syllables tracking transitional probabilities, ordinal position, and word identity in the auditory statistical learning task. **Top:** similarity by transitional probability. Greater within-class similarity indicates stronger grouping of syllables with a low transitional probability (0.33) than syllables with a high transitional probability (1.0). **Middle:** within vs. between similarity for ordinal position. Greater within-class similarity indicates greater grouping of syllables holding the same 1st, 2nd, or 3rd position in a word. **Bottom:** within vs. between similarity for word identity. Greater within-class similarity indicates grouping of syllables into individual words. (** p<0.01, * p<0.05, Wilcoxon rank sum test, error bars denote the population standard error of the mean).

Neural tracking of visual statistical learning

The need to segment continuous input into discrete units is not limited to auditory input but extends to stimuli in the visual domain, e.g., to build representations of scenes and events[33, 34]. Whether similar SL mechanisms are engaged in segmenting and acquiring structure across auditory and visual domains remains debated [23]. To investigate whether similar coding principles could be at work in the visual modality, we exposed 12 subjects to brief (2 min x 5 exposures) visual streams of fractal images (375 ms each) while manipulating the structure of the sequence [11]. In the structured streams, each fractal was assigned to the first or second position of a pair or "duplet" (**Fig. 4a**). We generated a continuous stream of fractals pairs by randomly inserting each pair without breaks or other cues between pairs. In the random streams, fractals were inserted the same number of times but in a random order at the fractal level. As a result, there were no pairs to segment. As in the auditory sequence, participants were not informed about the presence of structure in some of the sequences. Instead, they were asked to perform a 1-back cover task, in which they had to detect occasional repetitions of individual fractals that had been occasionally inserted into both stream types[26]. We observed significant behavioral d′ scores for both the structured and random streams ($d'_{structured}$ = 1.66, $t(11)$ =5.27, p<0.001; d'_{random}=1.78, $t(11)$=6.96, p<0.001), which did not differ ($t(11)$=-0.46, p=0.65). This indicates that participants were engaged with the task and attentive to the stimuli.

We collected neurophysiological recordings from 1606 intracranial electrodes in 12 patients, extensively covering frontal, parietal, temporal and occipital cortex (see **Supplementary Fig. S1** for electrode reconstruction). Neural oscillations at the 2.66 Hz frequency correspond to the individual fractals and should be present throughout exposure to both structured and random streams. If participants learned the regularities in the structured stream and used them to segment this stream into pairs, such learning should be reflected in oscillations at the 1.33 Hz frequency, corresponding to the pair boundaries, but should be absent in the random stream.

To test this hypothesis, we evaluated within-electrode phase coherence in the structured and random streams. We observed a significant peak in the phase coherence spectrum at the fractal rate (i.e., 2.66 Hz) for both structured and random streams (p<0.05, FDR corrected) **(Fig. 4b)**. Critically, there was also a significant peak at the pair rate (i.e., 1.33 Hz), but only for the structured stream (p<0.05, FDR corrected). Coherence at the pair rate was highly replicable across subjects with 12/12 patients exhibiting entrainment in at least one electrode; and appears to reflect implicit learning as subjects were unable to explicitly discriminate pairs vs. non-pairs in a follow up recognition test (mean performance = 53.4%, $t(11)$=1.41, p=0.18). Increases in pair-rate responses were observed within 160 s of exposure ($t(554)$= 3.8, p<0.01), subsequently plateauing despite further exposure to the

structured stream. Such an increase in the pair-rate response was absent in the random stream ruling out spurious entrainment as a function of time **(Fig. 4c)**.

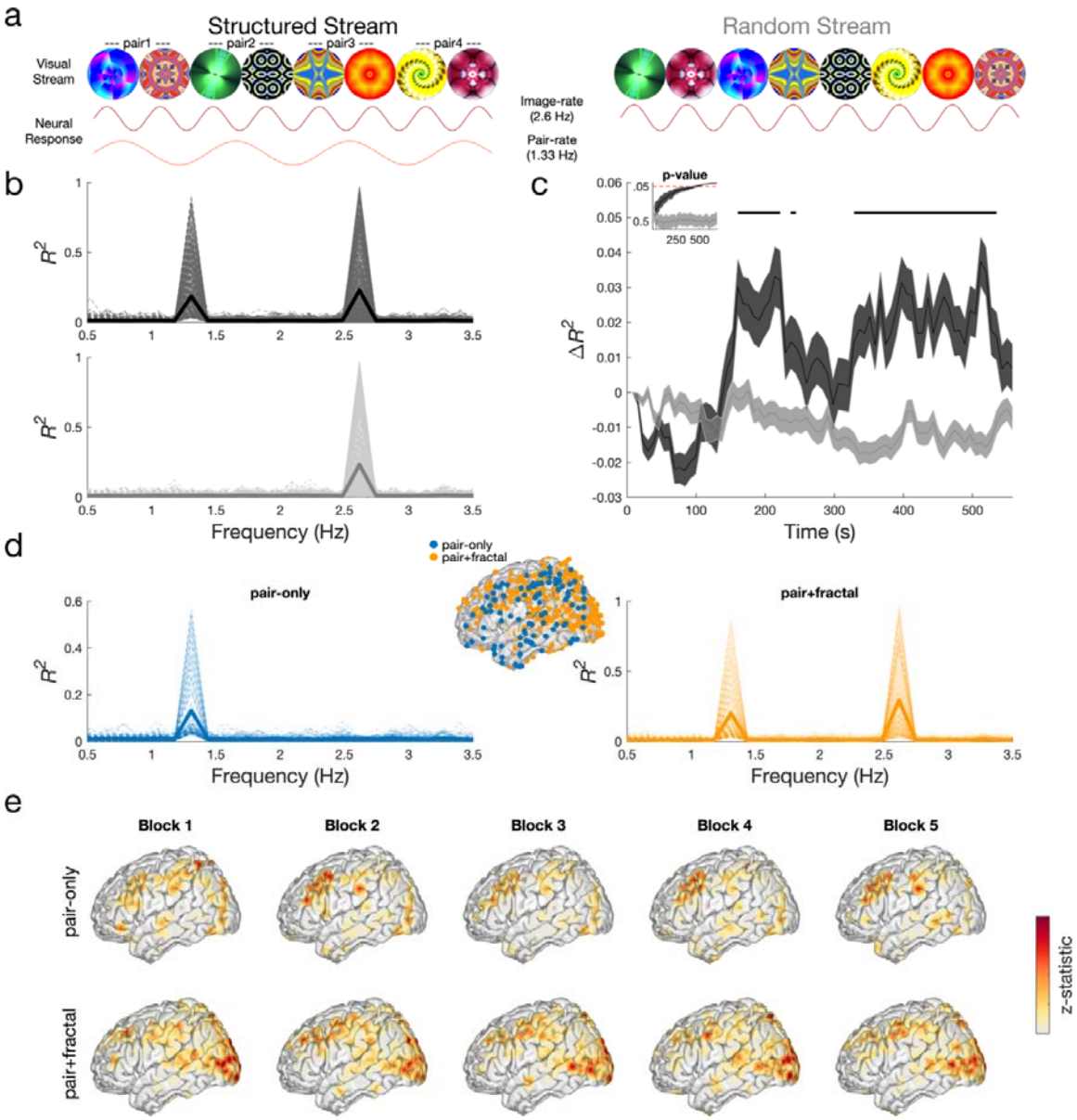

Figure 4 Neural tracking of visual statistical learning **(a)** Schematic depiction of the visual statistical learning task. The structured stream (left) consisted of a continuous visual stream of 8 fractals (375 ms SOA, 2.66 Hz). The transitional probabilities were adjusted to form four fractal pairs (750 ms SOA, 1.33 Hz). Note that the SOA of the fractals was elongated compared to the syllables to match the frequency of the learned units (pairs and words). The random stream (right) contained the same 8 fractals but in random order. The predicted neural responses are shown under each stream: fractal tracking is expected for both streams while pair tracking is expected for the structured stream only. **(b)** Phase coherence spectrum in neural data for the structured (top, black) and random (bottom, gray) conditions from 1606 electrodes in 12 patients. Each electrode is depicted with a thin line and the average with a thick line. **(c)** Timecourse of coherence response for all electrodes showing pair-rate tracking (1.33Hz) in the structured (black) and random (gray) conditions. Solid lines above the timecourse indicate when the change coherence increased above zero ($p<0.025$, one-sided t-test). Inset shows accumulated p-value of the responses over time showing when temporal coherence reached significance at 0.05 level (red dashed line). **(d)** Phase coherence spectrum in the structured condition for electrodes showing pair-tracking responses, in two groups: electrodes that tracked pairs only (left, blue) and electrodes that tracked pairs and fractals (right, orange). Inset shows the localization of pair-only (blue) and pair+fractal (orange) electrodes. **(e)**

Localized pair-rate coherence in the structured condition across each of the 5 exposure blocks for each electrode type (top: pair-only; bottom: pair+fractal), showing the evolution of the coherence response across brain regions

We next examined the neural circuits responsible for visual SL. Much like in the auditory SL, we observed an anatomical and hierarchical segregation between two types of electrodes: one clustered mostly within occipital (striate and extrastriate) and parietal cortex (IPS), showing entrainment at the fractal and pair rates (pair+fractal, **Fig. 4d, right**); the other localized more anteriorly with clusters in frontal (middle and superior), parietal and temporal cortex, showing entrainment at the pair-rate only (pair-only, **Fig. 4d, left**).

Figure 4e shows the temporal evolution for both types of pair response across cortical areas in each of the five blocks of the structured stream. As a function of exposure, responses increased in strength, but their spatial topographies remained stable ($ICC_{pair\text{-}only}$= 0.63, $ICC_{pair+fractal}$=0.75). Latency analysis on the different electrodes types showed that pair-only responses in superior frontal cortex and pair+fractal responses in occipital come online concurrently and earlier than responses in parietal or middle frontal cortex (**Supplementary Fig. S4**).

The same separation between pair+fractal and pair-only responses and spatial arrangement across electrodes was observed when restricting the analysis to the HGB (**Supplementary Fig. S2**).

Representational analysis in visual statistical learning

In the auditory modality we observed representational changes indicative of sequence learning based on transitional probabilities, ordinal position and word identity. How do visual regularities shape neural representations? Note that because the learned units in the visual modality contained only two elements (i.e., pairs), it is difficult to distinguish coding based on transitional probabilities and ordinal position — both cues predict grouping of the first fractal in each pair with the first fractals of other pairs and grouping of the second fractal with the other second fractals. Therefore we refer to this grouping as consistent with either transitional probability or ordinal position, in contrast with pair identity, which predicts that the first and second members of each pair will be grouped together and different from the other pairs[11].

We conducted multivariate pattern analysis on the raw LFP separately for electrodes showing pair-only and pair+fractal coherence and then calculated the cosine distance between the spatial patterns of neural activity across electrodes for every pair of fractals. Multidimensional scaling of the distances between fractals in the pair+fractal and the pair-only electrodes revealed a representation of first vs. second fractal (transitional probabilities and/or ordinal position) (**Fig. 5a-b**). When performing the same analysis on the raw LFP in electrodes showing HGB pair-rate responses (**Supplementary Fig. S2**), we found a similar pattern of clustering for pair-only electrodes, with grouping based on transitional probabilities and/or ordinal position (**Fig. 5c,** dimension 1). Importantly, and replicating the auditory findings, these electrodes also showed grouping by pair identity (**Fig. 5c,** dimension 2). Performing these analyses on the random stream did not reveal clustering of responses in any electrode type (**Supplementary Fig. S5**). These results indicate that rapid changes in representational space occur as a function of visual statistical learning.

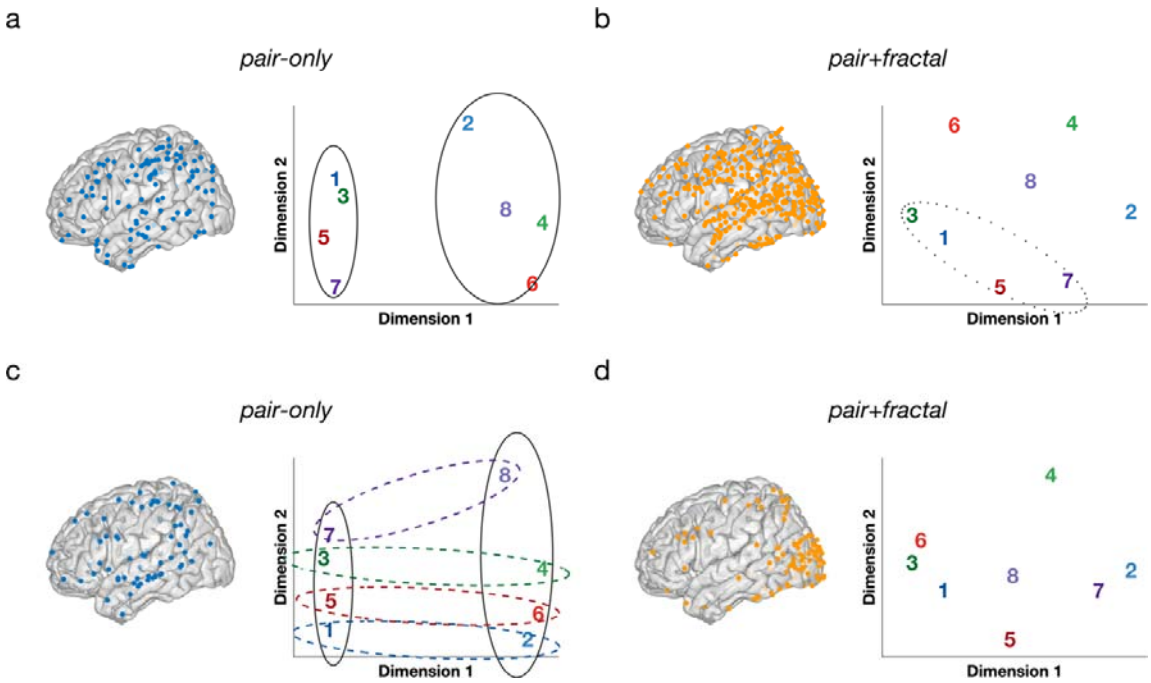

Figure 5 Pattern similarity results on raw LFP during visual statistical learning. **Top Row:** Multidimensional scaling (MDS) of distances between responses to individual fractals for electrodes showing significant LFP coherence in (a) pair-only and (b) pair+fractal electrodes. Pairs are color-coded, odd numbers refer to the first position, even to the second. **Bottom Row:** MDS representation on raw LFP in electrodes showing significant HGB coherence at the (c) pair-only and (d) pair+fractal rates. Dotted ellipses outline grouping by transitional probability/ordinal position in pair+fractal electrodes (Right panels). Solid ellipses outline grouping by transitional probability/ordinal position in pair-only electrodes (Left panels). Dashed ellipses indicate grouping (color coded) at the level of the individual pairs for pair-only electrodes showing segmental responses in the HGB.

To evaluate these different cues statistically, we collapsed pattern similarity across fractals belonging to different classes for both electrode types (pair+fractal, pair-only) and selection procedures (LFP, HGB). Note that although both transitional probability and ordinal position depend on grouping of first fractals together and second fractals together, the representational impact of these cues can be quantified in different, non-independent ways. Transitional probability was examined by comparing pattern similarity among first fractals (first-first) with low transitional probability (i.e., relatively unpredictable given preceding fractal) vs. among second fractals (second-second) with high transition probability (i.e., predictable given preceding fractal). Ordinal position was examined by comparing pattern similarity within the same position (first-first, second-second) vs. between different positions (first-second). Pair identity was examined by comparing pattern similarity within pair (e.g., first1-second1) vs. between pair (first1-second2).

In line with the MDS, pair-only electrodes showed greater similarity for fractals with low transitional probability (**Fig. 6, top**), reliable coding for ordinal position **(Fig. 6, middle),** and reliable coding for pair identity based on HGB selection **(Fig. 6, bottom)**. Pair+fractal

electrodes with LFP selection showed greater similarity for low transitional probability and reliable coding for ordinal position.

These results show that brief exposure to auditory but also visual statistical regularities molds representational spaces in comparable ways enabling learning of sequences along several dimensions. Though the circuits responsible for SL in the visual and the auditory modalities only partially overlap, several functional changes appear remarkably similar across modalities: from the speed at which learning effects are observed to the representational division of labor across lower- and higher-order brain areas.

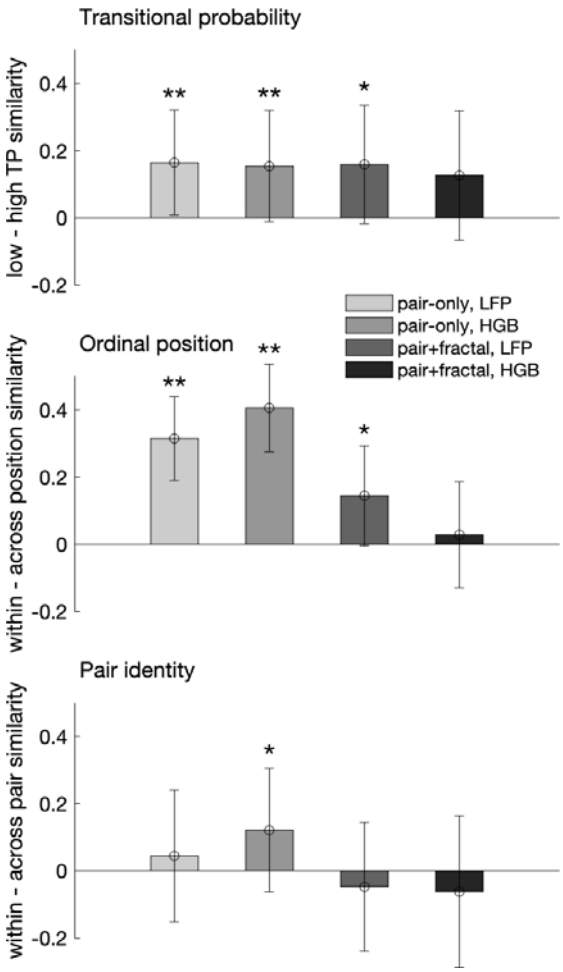

Figure 6. Comparison of multivariate pattern similarity for fractals tracking transitional probability, ordinal position, and pair identity in the visual statistical learning task. **Top:** Comparison of representational similarity for fractals with low vs. high transitional probability. Greater within-class similarity indicates stronger grouping of fractals with a low transitional probability (0.33) over fractals with a high transitional probability (1.0). **Middle:** within vs. between similarity for ordinal position. Greater within-class similarity indicates grouping of fractals holding the same 1st or 2nd position in a pair. **Bottom:** within vs. between similarity for pair identity. Greater within-class similarity indicates grouping of fractals into pairs. (** $p<0.01$, * $p<0.05$, Wilcoxon rank sum test, error bars denote the population standard error of the mean).

DISCUSSION

We have shown that statistical regularities in sensory inputs are extracted and represented in the brain in as little as ~2 min, in line with previous behavioral studies [2]. The underlying neurophysiology of this learning process can be reliably measured online and precisely localized using invasive electrophysiological recordings, even when subjects are not consciously aware of learning. SL is accompanied by rapid changes in neural representations, reflected in two functionally and anatomically segregated types of responses: those representing the base sensory elements plus learned higher-order units, and those exclusively representing the learned higher-order units. The localization of these responses appears hierarchical: the former map onto early, modality-specific processing stages (e.g., STS and occipital cortex), the latter onto late, amodal processing stages (e.g., IFG and anterior temporal pole). Although the precise areas exhibiting auditory and visual SL only partially overlap, this important organizing principle is evident across both modalities.

Our results are consistent with findings that areas across the cortical hierarchy integrate information over increasingly larger temporal windows[30, 35-37]. Yet, they go beyond topographical mapping of temporal receptive fields: First, our results do not only reflect stable response properties of different levels of processing but show how SL shapes the neural representational space within these areas profoundly and rapidly. This contrasts with the much more gradual representational change that occurs over development or with longer-term perceptual learning. Second, we discovered that qualitatively different aspects of sequence knowledge are encoded across different brain areas: sites representing the stimulus *and* higher-order units encode local and generic aspects of sequences, such as their transitional probabilities (or degree of uncertainty). In contrast, sites exclusively representing higher-order structure encode global and also more specific aspects of the sequences such as the ordinal position of the elements, but most importantly the specific identity of the sequential structure.

Studies on SL emphasize variations in predictive uncertainty as the primary cue for segmentation. Our results extend this body of work demonstrating that SL also involves the acquisition of higher-order sequence knowledge, i.e., ordinal position and identity. Although the former can aid word segmentation, the acquisition higher-order structures or "chunks" may serve as the mental units for mapping segmented word forms onto novel word referents [32]. Our latency analysis suggests that those two processes, while emerging quickly in the course of SL, may not be acquired simultaneously (see **Supplementary Fig. S4**). These results are in line with the hypothesis that the output of the word boundary discovery may provide cues to edges of constituents which in turn can serve as scaffolding for the subsequent discovery of internal structure, i.e., which elements are contained and in which positions[38].

Our finding that sequences are represented at multiple levels, from simple and generic to complex and specific regularities, may also help reconcile two theoretical models that have been proposed to explain sensitivity to statistical structure: the 'statistical model' posits that learners represent statistical relations between elements in the input and do not explicitly represent statistically coherent units in memory. In contrast, 'chunking

models' posit that learners represent statistically coherent units of information from the input in memory, such that the resulting representations are discrete chunks of information[39]. So far, these models have been only contrasted at the behavioral level, e.g., by studying sensitivities to illusory or embedded units[40, 41]. An intriguing possibility is that these two models actually coexist and map onto the different networks and sequence representations that we report here — i.e., simple representations encoding transitional probabilities and complex representations encoding positional information and identity — and that the discrepant behavioral results observed in the literature reflect the differential engagement of these two neural processes across different tasks.

The main organizational principles of neural changes underlying SL are shared across the auditory and visual domain. For example, we observed a similar functional clustering of responses, i.e., elemental + higher-order units and higher-order units-only, as well as a similar division of labor for the representation of the sequences for the auditory and the visual modalities (compare **Figs. 1** with **4**). In addition, no clear lateralization was observed either for the auditory or visual statistical learning in any of the electrodes following higher-order units (**Supplementary Fig. S6**). Nevertheless, there were also striking differences between domains: For instance, during the temporal evolution of SL, responses to higher-order units continuously increased in the auditory domain, whereas responses in the visual domain developed faster and plateaued. This may reflect the different chunk sizes (pair vs. triplet) in the two modalities. In addition, the cortical areas involved in auditory and visual SL only partially overlapped (**Fig. 7 and Supplementary Table 1; Supplementary Fig. S8**). This was to be expected at the level of the sensory responses: responses encoding the stimulus and higher-order units clustered around the STG for the auditory SL involving syllables and around occipital cortex for the visual SL involving fractals. Perhaps more interestingly, areas engaged exclusively in higher-order unit representation were also (partially) segregated: while areas such as IFG and ATL tracked higher-order units in both modalities, middle frontal and superior parietal cortices seemed to be differentially involved in auditory and visual SL. This result cannot be explained by the fact that different groups of subjects (and electrode coverage) contributed to the different tasks, as we confirmed this segregation in six subjects who completed both auditory and visual SL tasks (**Supplementary Fig. S7)**.

This suggests that while sequence operations performed across domains might build on similar representations (transitional probability, ordinal position and identity), the circuits performing these operations might be modality-specific to some extent, much like the nested tree structures involved in language, music, and mathematics that are each represented in distinct circuits[42]. Our results speak for a more modularized representation of sequences for the encoding of local and simple aspects such as transitional probability represented in sensory areas, but a less modularized representation as complexity increases, as IFG and ATL encoded positional and identity information for both visual and auditory SL, in line with a domain-general role in SL.

Intriguingly, we observed a dissociation between 'online' and 'offline' measures of SL. A clear NFT response at the rate of the higher-order units was observed for the structured stream in virtually all participants, and increased as a function of exposure to the regularities. Yet, the patients did not perform better than chance in the subsequent 'offline' behavioral recognition test. Hence, the online NFT responses seem to be more

sensitive and to reflect implicit learning. This suggests intact cortical circuits for automatic SL in our group of patients, but hints at episodic memory deficits, which are prevalent in temporal lobe epilepsy[43]. Relatedly, the fact that we observed SL in patients with epilepsy may argue against the involvement of the medial temporal lobe (MTL) and hippocampus. However, fMRI studies have demonstrated that visual SL leads to changes in representational similarity in the hippocampus[11, 14]; a result corroborated by lesion studies demonstrating the need for the MTL in both visual and auditory SL[10, 12]. Moreover, computational studies indicate a division of labor in the hippocampus between the monosynaptic pathway (connecting entorhinal cortex directly to CA1) and the trisynaptic pathway (connecting entorhinal cortex to CA1 through dentate gyrus and CA3), with the former supporting SL and the latter supporting episodic memory. Given that epilepsy can lead to selective deficits in hippocampal circuits[44], it remains possible that the observed dissociation between online and offline SL measures reflects partial damage to one pathway.

The dissociation between online and offline measures is also noteworthy given the large variability observed when SL is measured through offline recognition tasks[45]. NFT may represent an alternative and potentially more sensitive measure of learning, additionally enabling tracking of the precise temporal evolution of SL. Further studies could use this methodology to characterize the different learning trajectories across subjects as well as individual differences in SL. Because NFT does not require instructions or impose task demands, it is well-suited to tracking the acquisition of sequence knowledge across the lifespan from newborns to the elderly. Combining NFT with representational similarity analysis provides insight into the building blocks and computations involved in SL.

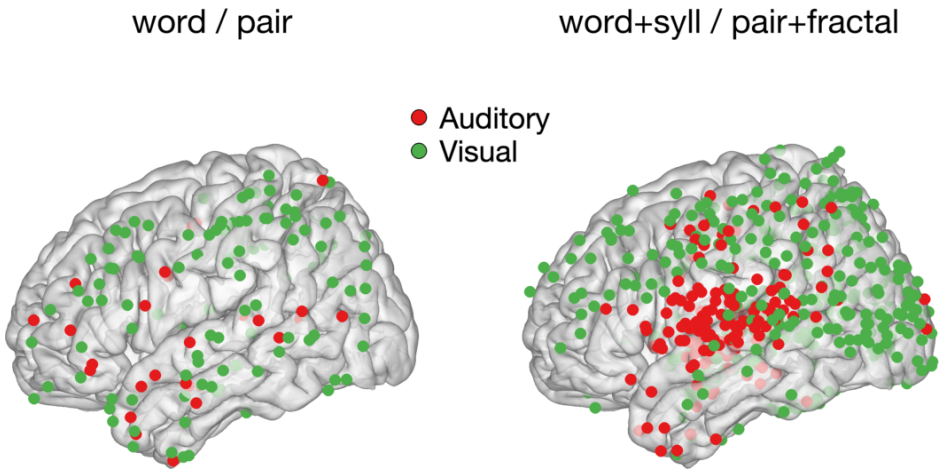

Figure 7 Electrode localization, projected onto left hemisphere, of higher-order structure across auditory (red) and visual (green) SL. Responses to higher-order structure that contain information about the constituent elements (right) segregate naturally by sensory modality, whereas sites with higher-order responses exclusively (left) show overlap between tasks in the IFG and ATL.

Acknowledgements

We thank P. Minhas, B. Mahmood, M. Hofstadter for help in intracranial data collection, H. Wang for electrode reconstruction, Sasha Devore, Bijan Pesaran, David Poeppel and Caspar M. Schwiedrzik for helpful comments on an earlier version of the manuscript. This work was supported by a grant from the German-Israeli Foundation for Scientific Research and Development (GIF I-99-105.4-2016) and NIH R01 MH069456. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no competing interests.

Author Contributions

Conceived and designed the experiments: LM; collected the data: SH and LM.; attending neurosurgeons: WD; supervised recordings and IRB approval process: AF, DF, OD, WD, PD and AL; analyzed the data: SH, NTB and LM; consulted on experimental design and interpretation: NTB; wrote the manuscript: SH, NTB and LM; revised the manuscript: all authors.

References

1. Kuhl, P.K. Early language acquisition: cracking the speech code. *Nat Rev Neurosci* **5**, 831-843 (2004).
2. Saffran, J.R., Aslin, R.N. & Newport, E.L. Statistical learning by 8-month-old infants. *Science* **274**, 1926-1928 (1996).
3. Kirkham, N.Z., Slemmer, J.A. & Johnson, S.P. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**, B35-42 (2002).
4. Turk-Browne, N.B. Statistical learning and its consequences. in *The influence of attention, learning, and motivation on visual search* (ed. M.D. Dodd & J.H. Flowers) 117-146 (Springer, 2012).
5. Aslin, R.N. Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdiscip Rev Cogn Sci* **8** (2017).
6. Orban, G., Fiser, J., Aslin, R.N. & Lengyel, M. Bayesian learning of visual chunks by human observers. *Proc Natl Acad Sci U S A* **105**, 2745-2750 (2008).
7. Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B. & Botvinick, M.M. Neural representations of events arise from temporal community structure. *Nat Neurosci* **16**, 486-492 (2013).
8. Meyniel, F., Maheu, M. & Dehaene, S. Human Inferences about Sequences: A Minimal Transition Probability Model. *PLoS Comput Biol* **12**, e1005260 (2016).
9. Maheu, M., Dehaene, S. & Meyniel, F. Brain signatures of a multiscale process of sequence learning in humans. *Elife* **8** (2019).
10. Schapiro, A.C. & Turk-Browne, N.B. Statistical learning. in *Brain mapping: an encyclopedic reference* (ed. A.W. Toga) 501–506 (Elsevier, 2015).
11. Schapiro, A.C., Kustner, L.V. & Turk-Browne, N.B. Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol* **22**, 1622-1627 (2012).
12. Schapiro, A.C., Gregory, E., Landau, B., McCloskey, M. & Turk-Browne, N.B. The necessity of the medial temporal lobe for statistical learning. *J Cogn Neurosci* **26**, 1736-1747 (2014).

13. Turk-Browne, N.B., Scholl, B.J., Chun, M.M. & Johnson, M.K. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci* **21**, 1934-1945 (2009).
14. Schapiro, A.C., Turk-Browne, N.B., Norman, K.A. & Botvinick, M.M. Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**, 3-8 (2016).
15. Karuza, E.A., *et al.* The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang* **127**, 46-54 (2013).
16. McNealy, K., Mazziotta, J.C. & Dapretto, M. Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci* **26**, 7629-7639 (2006).
17. McNealy, K., Mazziotta, J.C. & Dapretto, M. The neural basis of speech parsing in children and adults. *Dev Sci* **13**, 385-406 (2010).
18. Karuza, E.A., Emberson, L.L. & Aslin, R.N. Combining fMRI and behavioral measures to examine the process of human learning. *Neurobiol Learn Mem* **109**, 193-206 (2014).
19. Buiatti, M., Peña, M. & Dehaene-Lambertz, G. Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* **44**, 509-519 (2009).
20. Batterink, L.J. & Paller, K.A. Online neural monitoring of statistical learning. *Cortex* **90**, 31-45 (2017).
21. Getz, H.R., Ding, N., Newport, E.L. & Poeppel, D. Cortical entrainment to constituent structure in language acquisition. *Cognition* **2**, 37082-37082 (2018).
22. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* **88**, 2-19 (2015).
23. Frost, R., Armstrong, B.C., Siegelman, N. & Christiansen, M.H. Domain generality versus modality specificity: the paradox of statistical learning. *Trends Cogn Sci* **19**, 117-125 (2015).
24. Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* **19**, 158-164 (2016).
25. Zhou, H., Melloni, L., Poeppel, D. & Ding, N. Interpretations of frequency domain analyses of neural entrainment: periodicity, fundamental frequency, and harmonics. *Front Hum Neurosci* **10**, 274 (2016).
26. Turk-Browne, N.B., Junge, J. & Scholl, B.J. The automaticity of visual statistical learning. *J Exp Psychol Gen* **134**, 552-564 (2005).
27. Saffran, J.R., Newport, E.L. & Aslin, R.N. Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language* **35**, 606-621 (1996).
28. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat Rev Neurosci* **8**, 393-402 (2007).
29. Mukamel, R., *et al.* Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science* **309**, 951 (2005).
30. Honey, C.J., *et al.* Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* **76**, 423-434 (2012).
31. Manning, J.R., Jacobs, J., Fried, I. & Kahana, M.J. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J Neurosci* **29**, 13613-13620 (2009).
32. Graf Estes, K., Evans, J.L., Alibali, M.W. & Saffran, J.R. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychol Sci* **18**, 254-260 (2007).
33. Richmond, L.L. & Zacks, J.M. Constructing Experience: Event Models from Perception to Action. *Trends Cogn Sci* **21**, 962-980 (2017).

34. Zacks, J.M. & Tversky, B. Event structure in perception and conception. *Psychol Bull* **127**, 3-21 (2001).
35. Hasson, U., Chen, J. & Honey, C.J. Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn Sci* **19**, 304-313 (2015).
36. Lerner, Y., Honey, C.J., Silbert, L.J. & Hasson, U. Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* **31**, 2906-2915 (2011).
37. Chaudhuri, R., Knoblauch, K., Gariel, M.A., Kennedy, H. & Wang, X.J. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* **88**, 419-431 (2015).
38. Endress, A.D. & Bonatti, L.L. Words, rules, and mechanisms of language acquisition. *Wiley Interdiscip Rev Cogn Sci* **7**, 289 (2016).
39. Perruchet, P. & Vinter, A. PARSER: A model for word segmentation. *Lang Cogn Process* **39**, 246-263 (1998).
40. Perruchet, P. & Poulin-Charronnat, B. Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *J Mem Lang* **66**, 807-818 (2012).
41. Slone, L.K. & Johnson, S.P. When learning goes beyond statistics: Infants represent visual sequences in terms of chunks. *Cognition* **178**, 92-102 (2018).
42. Dehaene, S. The Error-Related Negativity, Self-Monitoring, and Consciousness. *Perspect Psychol Sci* **13**, 161-165 (2018).
43. Lemesle, B., Planton, M., Pages, B. & Pariente, J. Accelerated long-term forgetting and autobiographical memory disorders in temporal lobe epilepsy: One entity or two? *Rev Neurol (Paris)* **173**, 498-505 (2017).
44. Zhao, F., *et al.* Neuropsychological deficits in temporal lobe epilepsy: A comprehensive review. *Ann Indian Acad Neurol* **17**, 374-382 (2014).
45. Siegelman, N. & Frost, R. Statistical learning as an individual ability: theoretical perspectives and empirical evidence. *J Mem Lang* **81**, 105-120 (2015).
46. Yang, A.I., *et al.* Localization of dense intracranial electrode arrays using magnetic resonance imaging. *Neuroimage* **63**, 157-165 (2012).
47. Ding, N. & Simon, J.Z. Power and phase properties of oscillatory neural responses in the presence of background activity. *J Comput Neurosci* **34**, 337-343 (2013).
48. Prichard, D. & Theiler, J. Generating surrogate data for time series with several simultaneously measured variables. *Phys Rev Lett* **73**, 951-954 (1994).
49. Berens, P. CircStat: A MATLAB Toolbox for Circular Statistics. *J Stat Softw* **31**, 1-21 (2009).
50. McGraw, K.O. & Wong, S.P. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* **1**, 30-46 (1996).
51. Walther, A., *et al.* Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188-200 (2016).

METHODS

Participants and recordings

Electrocorticography (ECoG). ECoG recordings were obtained from a total of 23 patients (13 female, average age 35 yrs, range 16-59 yrs, 21 right-handed) with drug-resistant focal epilepsy undergoing clinically motivated invasive monitoring at the Comprehensive Epilepsy Center of the New York University Langone Medical Center. 11 subjects participated in the auditory statistical learning only, 6 in the visual statistical learning only, and 6 subjects participating both in the auditory and the visual statistical learning task. All subjects participating in the study provided oral and written informed consent prior to participation in the study, in accordance with the Institutional Review Board at the New York University Langone Medical Center. Patients were informed that participation in the study would not affect their clinical care and that they could withdraw from the study at any point without affecting medical treatment. Brain activity was recorded from a total of 3689 (average of 120+/-30 per subject) intracranially implanted subdural platinum-iridium electrodes embedded in silastic sheets (2.3 mm diameter contacts, Ad-Tech Medical Instrument). The decision to implant, electrode targeting, and the duration of invasive monitoring were determined solely on clinical grounds without reference to this or any other study. Macroelectrodes were arranged as grid arrays (8 × 8 contacts, 10 or 5 mm center-to-center spacing), linear strips (1 × 8/12 contacts), or depth electrodes (1 × 8/12 contacts), or some combination thereof. Subdural electrodes covered extensive portions of lateral and medial frontal, parietal, occipital, and temporal cortex of the left and/or right hemisphere (see **Supplementary Fig. 1** for electrode coverage across all subjects, and for the individual coverage of each subject). Recordings from grid, strip and depth electrode arrays were acquired using a NicoletOne C64 clinical amplifier (Natus Neurologics, Middleton, WI), bandpass filtered from 0.16-250 Hz and digitized at 512 Hz. Intracranial EEG signals were referenced to a two-contact subdural strip facing towards the skull near the craniotomy site. Data were subsequently downsampled to 250 Hz offline and a 60-Hz notch filter was applied to remove any line-noise artifacts. All electrodes were visually inspected, those with excessive noise artifacts were removed from subsequent analysis (185/3689 electrodes removed). In addition to analyzing the raw local field potential (LFP), high-gamma band (HGB) activity was extracted by applying an additional high-pass filter (fc=70 Hz) and the envelope of HGB activity was estimated by taking the square of the Hilbert transform of the filtered signal.

Stimulus materials and experimental procedure

Auditory statistical learning task. Twelve consonant-vowel (CV) syllables were synthetically generated using MacTalk. Syllable length was equated and prosody was flattened prosody using Praat (Boersma, Paul & Weenink, David, 2018). The individual syllables were concatenated in Matlab. Two sequences were created: a structured and a random sequence. In the structured sequence transitional probabilities between syllables was manipulated such that 4 hidden words (3 syllables each) were embedded in the sequence (see **Fig. 1**), resulting in a continuous artificial language stream with an underlying syllable presentation rate of 4Hz, and word-rate of 1.33Hz. In the random sequence transitional probabilities across syllables were the same (e.g., p=1/11 syllables). Each sequence lasted

approximately 2 minutes (540 syllable presentations) and was presented 5 times. To avoid potential cueing of the words at the start and end of the stream, the volume of the audio stream was ramped on and off, over the first and last 1.5s, respectively. Participants were not informed of the structure, and instead, to ensure task compliance, participants were asked to perform a cover task in which they indicated syllable repetitions that were randomly embedded in the auditory streams. Sixteen (16) syllable repetitions were randomly embedded into each presentation block of a sequence.

Once both streams (random and structured) had been played to the participants, they were then informed that one of the audio streams consisted of a hidden structured containing "words". Subjects then performed a two-alternative forced-choice (2AFC) task where they had to select from two audio segments, presented one after the other, the one containing a "word". One audio segment contained a "word" (e.g., "tupiro"), while the other was a lower probability "part-word" from the stream spanning word boundaries (e.g., "butupi"). Presentation order was counterbalanced across trials. Since exposure to the individual syllables is equated, a preference for the true words over the part-word is indicative of statistical learning. Two of the seventeen patients who participated in this experiment did not complete the 2AFC task, one for technical reasons and one because the participant was confused about the task.

Visual statistical learning task. The procedure for the visual statistical learning task was identical to the auditory statistical learning task, however, in this task, sequences were formed from 8 fractals (4-sets of two fractal pairs, duplets). Fractals were taken from the same set of images previously used in [11]. The stimulus onset asynchorony between fractals was set to 375, whereby each fractal was presented for 233 ms with an interstimulus interval of 150 ms. In the structured sequence transitional probabilities between fractals were manipulated such that 4 hidden fractal-pairs (2 fractal each) were embedded in the sequence (see **Fig. 4**), resulting in a continuous stream of fractals with a presentation rate of 2.6 Hz and a fractal-pair rate of 1.3 Hz. In the random sequence, transitional probabilities remained fixed between all possible fractals (e.g., p=1/7). Each sequence lasted approximately 2 minutes (360 fractals presentations). As in the auditory learning task, participants were not informed of the structure, however, to ensure task compliance, participants were asked to perform a cover task in which they indicated, using the keyboard, when a fractal had been repeated. Sixteen (16) fractal repetitions were randomly embedded within each sequence block.

Data analysis

Behavioral data. Performance during the online cover task was assessed by calculating a d′ score for every participant across all 5 exposure trials. As the cover task was embedded in the continuous stream of auditory or visual stimuli, detection of a syllable or image repetition was deemed accurate ("hit") if the participant made a keyboard response within 150-1500ms of the occurrence of the repetition (Results are robust to the selection of the response window, as comparable results were obtained using response windows up to 750ms, 1000ms, and 3000ms). All other keyboard responses outside the valid response window were deemed false alarms. Significance of d′ scores were assessed via a one-sample t-test and comparison of d′ scores across conditions (structured vs. random) was

assessed with a paired t-test. Performance on the offline 2AFC recognition test was assessed by determining the percent of correctly identified "words" or "fractal-pairs", and subjected to a one-sample t-test against chance performance (50%).

ECoG Surface reconstruction and electrode localization. Pre-surgical and post-surgical T1-weighted MRIs were acquired for each patient, and the location of the electrode relative to the cortical surface was determined from co-registered MRIs following the procedure described in Yang and colleagues[46]. Co-registered, skull-stripped T1 images were nonlinearly registered to an MNI-152 template and electrode locations were then extracted in Montreal Neurological Institute (MNI) space (projected to the surface) using the co-registered image. A three-dimensional reconstruction of each patient's brain was computed using FreeSurfer (http://surfer.nmr.mgh.harvard.edu). In all figures, electrode locations are projected onto the right hemisphere of the MNI-152 template brain, unless otherwise noted.

Phase coherence. For each experiment, signals from all five (5) blocks of a sequence (structured or random) were concatenated and then reshaped into 10-word segments (10 words x 90 trials x electrodes) and converted into the frequency domain via FFT (0.134 Hz resolution). Phase coherence was computed for each electrode, $R^2 = [\sum^N cos\, \emptyset]^2 + [\sum^N sin\, \emptyset]^2$ over the 90 trials [47]. Significance of the response at each frequency of interest (e.g., 1.33Hz & 4Hz) was determined by comparing the magnitude of the coherence response to 1000 phase-shuffled surrogate datasets [48] and then subjected to FDR correction across all electrodes.

Phase coherence latency. The response latency was computed for all electrodes with a significant phase coherence at the higher-order rate (i.e., word rate or pair rate, 1.33 Hz) across all blocks. To identify a time-point when each electrode exhibited a significant response (e.g., time to first significant response), phase coherence was computed on smaller segments of data (10 words x 9 trials), using a sliding window of 10 words. The significance of the response in each window was determined using a Rayleigh test (CircStat Toolbox [49]) on the phase angles computed across trials per FFT bin. The response latency for each electrode was then defined as the first time-point at which statistical significance was reached (p<0.05). Phase coherence was also estimated within each block (e.g., Fig. 2e), and the stability of coherence across blocks was assessed using Intraclass Correlation (ICC)[50].

Representational similarity analysis. To assess the similarity of neural responses to each token (syllables, fractals), a multivariate spatial pattern analysis was performed. First, all individual trials were 'whitened' by the noise covariance matrix computed across all tokens and trials[51] and the average response across all tokens (e.g., non-specific response) was subtracted from each trial. Next, the data were vectorized across all significant electrodes (e.g., samples x significant electrodes) within a cluster (e.g., word+syll or word-only) and the dissimilarity of the spatial patterns was computed between each pair of tokens. Dissimilarity was assessed using the correlation distance, $r = 1 - \frac{(A-\bar{A})\cdot(B-\bar{B})}{||A-\bar{A}||\,||B-\bar{B}||}$, where A and B represent the mean spatial vectors for two of the tokens being compared (e.g., syllables 'tu' vs. 'pi'), and using 5-fold cross-validation. The resultant representational dissimilarity matrix (RDM) was subjected to a principal component analysis and the first

two dimensions were plotted against each other to produce a 2-dimensional mapping of dissimilarity scores across all pairs (see **Fig. 2**).

Within vs. between category similarity. To quantify the degree to which responses are able to capture features of the learned streams i.e., transitional probability, ordinal position and/or identity, we calculated the difference in similarity (Pearson correlation) between items in the same category and items belonging to the other category in question. This similarity estimate was calculated by randomly sampling the significant electrodes by type (e.g., raw LFP, word-only) and computing a similarity matrix between all tokens (e.g., syllables or fractals), for each resampling. This resampling procedure was repeated 200 times (with replacement), and the average fisher-transformed correlations of all elements within a category was compared against all items that spanned the opposing category using a Wilcoxon's rank sum test (two-sided).

Code availability

The code used to generate the main findings of the current study are available from the corresponding author upon reasonable request.

Data availability.

The data sets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.