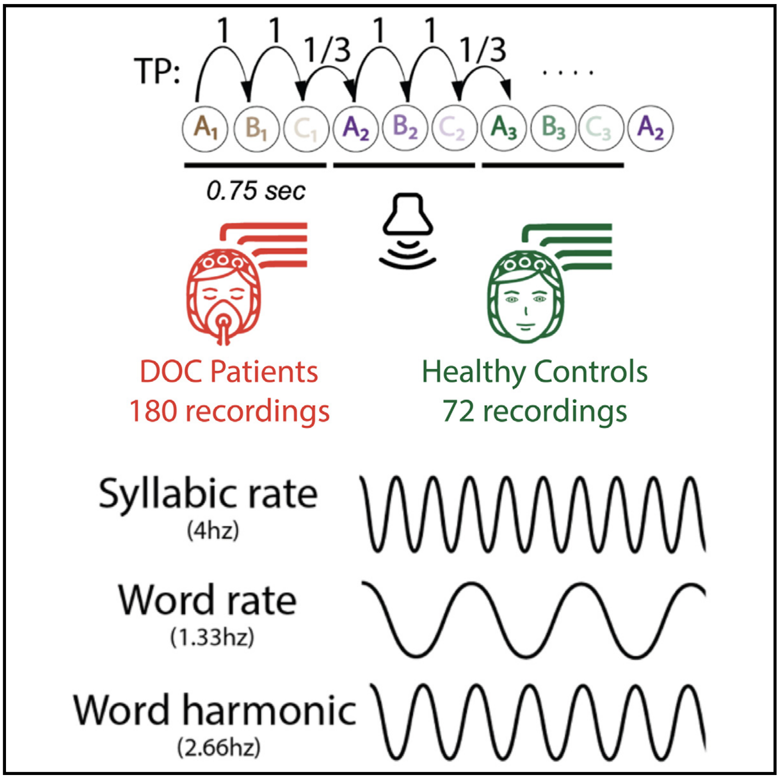
iScience

Article

The role of conscious attention in auditory statistical learning: Evidence from patients with impaired consciousness

Graphical abstract



Highlights

- Statistical learning is partially persevered in patients with DOC
- Statistical learning is a highly automatic process that does not need explicit attention
- Neural markers of statistical learning correlate with DOC clinical scores

Authors

Lucas Benjamin, Di Zang, Ana Fló, ..., Xuehai Wu, Peng Gui, Ghislaine Dehaene-Lambertz

Correspondence

pgui@ion.ac.cn (P.G.), ghislaine.dehaene@cea.fr (G.D.-L.)

In brief

Sensory neuroscience; Cognitive neuroscience

Benjamin et al., 2025, iScience *28*, 111591
January 17, 2025 © 2024 Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.111591

iScience

CellPress
OPEN ACCESS

Article

The role of conscious attention in auditory statistical learning: Evidence from patients with impaired consciousness

Lucas Benjamin,[1,11] Di Zang,[2,9,11] Ana Fló,[1,7] Zengxin Qi,[2,3,4,5,6] Pengpeng Su,[10] Wenya Zhou,[10] Liping Wang,[8] Xuehai Wu,[2,3,4,5,6] Peng Gui,[8,*] and Ghislaine Dehaene-Lambertz[1,12,*]

[1]Cognitive Neuroimaging Unit U992, CNRS, INSERM, CEA, DRF/Institut Joliot, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France
[2]Department of Neurosurgery, Huashan Hospital, Shanghai Medical College, Fudan University, Shanghai 200040, China
[3]National Center for Neurological Disorders, Shanghai 200040, China
[4]Shanghai Key Laboratory of Brain Function and Restoration and Neural Regeneration, Shanghai 200040, China
[5]Neurosurgical Institute of Fudan University, Shanghai, China
[6]Shanghai Clinical Medical Center of Neurosurgery, Shanghai, China
[7]Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy
[8]Institute of Neuroscience, Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China
[9]Department of Neurosurgery, China-Japan Friendship Hospital, Beijing 100029, China
[10]Shanghai Hebin Rehabilitation Hospital, Shanghai 201702, China
[11]These authors contributed equally
[12]Lead contact
*Correspondence: pgui@ion.ac.cn (P.G.), ghislaine.dehaene@cea.fr (G.D.-L.)
https://doi.org/10.1016/j.isci.2024.111591

SUMMARY

The need for attention to enable statistical learning is debated. Testing individuals with impaired consciousness offers valuable insight, but very few studies have been conducted due to the difficulties inherent in such studies. Here, we examined the ability of patients with varying levels of disorders of consciousness (DOC) to extract statistical regularities from an artificial language composed of randomly concatenated pseudowords by measuring frequency tagging in EEG. The objectives were firstly, to assess the automaticity of the segmentation process and the correlations between the level of covert consciousness and statistical learning capacities; secondly, to identify potential new diagnostic indicators. We observed that segmentation abilities were preserved in some minimally conscious patients, suggesting that auditory statistical learning is an inherently automatic low-level process. Due to significant inter-individual variability, word segmentation might not be robust enough for clinical use. In contrast, temporal accuracy of auditory syllable responses correlates strongly with coma severity.

INTRODUCTION

The structure of sensory sequences can be uncovered by analyzing statistical patterns present in the input, a mechanism known as statistical learning. It has been shown that such a mechanism operating between adjacent syllables in a speech stream can be exploited to detect words without any other cues.[1] In this behavioral experiment, four pseudo-words were randomly concatenated to form a continuous sequence with a high predictability of the next syllable within the pseudo-word but a drop of transition probability between pseudo-words. Despite the absence of other cues for word segmentation, such as those provided by prosodic indices, participants were able to distinguish words corresponding to triplets with high transition probabilities between syllables, from part-words comprising a low transition between syllables. This result was subsequently extended to non-linguistic streams[2,3] as well as to visual sequences.[4] This phenomenon is not exclusive to the human species,[5,6] nor limited to adjacent elements; it also extends to non-adjacent dependencies.[7] Furthermore, evidence indicates that humans are also capable of learning higher-order statistical structures in sequences.[8–12] Thus, it has been proposed that statistical learning represents a fundamental learning mechanism operating in different modalities and is sensitive to statistical regularities at different temporal scales.

The question of whether auditory statistical learning occurs automatically and whether this mechanism requires attention remains a matter of debate. While most studies have used passive exposure to sequences in awake and attentive subjects, the impact of attention on statistical learning has been explicitly questioned by some authors, with mixed results. While some studies have shown a drastic decline in performance under divided attention,[13,14] others have shown that auditory sequence segmentation is preserved when attention is diverted by asking subjects to focus on an independent visual sequence.[15,16] Another study even reports enhanced performance in participants experiencing cognitive fatigue following an effortful working memory task prior to sequence presentation.[17] Thus, studies that directly manipulate attentional focus do not provide any definitive answer regarding the interaction between auditory statistical learning and participants' attentional resources.

Another approach to answering this question would be to study statistical auditory learning in sleeping subjects. Although not directly testing attention, sleep studies address the automaticity of this mechanism. In a recent study, Batterink et al.[18] tested sleeping adults with pseudo-word segmentation tasks composed of either bi-syllabic or tri-syllabic pseudo-words. The results indicated that sleeping adults were only sensitive to bi-syllabic pairs but failed to extract tri-syllabic words. These results suggest that sleep might reduce the integration period of the statistical computation without preventing the computation between adjacent items. This interpretation appears congruent with the findings of Strauss and colleagues'[19] who observed during sleep, a preserved mismatch response to auditory violations of a sequence local regularity but no reaction to a violation concerning a longer timescale regularity. In contrast, one-to three-day-old neonates tested with EEG in a pseudo-word segmentation paradigm similar to that of Batterink et al. in adults[20–22] were able to segment tri-syllabic pseudowords during sleep (for quadrisyllabic pseudo-words see[22]).

Studying statistical learning in patients with coma could also reveal the automaticity of statistical learning. Among patients with Disorders of Consciousness (DOC), varying degrees of residual consciousness may be observed, as assessed by diagnostic tools such as CRS-R[23] or categorization into Unresponsive Wakefulness Syndrome (UWS) -no visible sign of awareness- or Minimally Conscious State (MCS) –visible partial awareness-[24,25]. One attempt to measure auditory statistical learning in patients with DOC was recently made by Xu et al.[26] using bi-syllabic words concatenated in a continuous, monotonic stream. They reported some learning in patients with emerging consciousness. Indeed, the power at the frequency of syllable pairs and its harmonics were significantly above zero, suggesting successful segmentation of sequences into word pairs. Although this study supports the possibility of auditory statistical learning in patients with DOC, it had certain limitations. Firstly, the bi-syllabic units employed in the study, which required only a pairing between two items, do not encompass all aspects of segmentation, as evidenced by the findings of the sleep studies discussed above. Furthermore, the pairs presented in the study could be either frequent and meaningful in natural language or reversed and thus meaningless (e.g., "go home" versus "home go"). However, even in the reversed condition, syllables within a pair are more related to each other in natural language than syllables between pairs. This feature makes it difficult to distinguish the current learning of statistical regularities in the stream from a reactivation in memory of previous exposure to natural language. This limitation was partially addressed by obtaining similar results with the learning of artificial tri-syllabic pseudowords, but the sample of minimally conscious patients was small (N=8).

In the present study, we used the experimental design and techniques previously employed in our research with sleeping neonates[20] (see STAR Methods): Our aim was to investigate the ability of patients with DOC to recognize statistical regularities in an artificial syllable sequence comprising four tri-syllabic pseudo-words which were concatenated in a pseudo-random manner (Figure 1). The syllables within the words followed each other predictively (transitional probabilities = 1), while between the words, the transitional probabilities dropped to 1/3. In addition, the subjects were presented with a random sequence composed of the same syllables with a flat transitional probability of 1/11 between syllables. A group of healthy awake adults was also included as a control to compare with the clinical population.

We utilized high-density EEG recordings (256 channels) to assess sequence segmentation using the frequency tagging method, which has been demonstrated to be a robust approach for evaluating non-responsive subjects in our previous studies.[16,20,22,27] This technique relies on detecting rhythmicity in brain activity driven by the rhythmicity of the input sequence. This method permits the monitoring of sequence segmentation by following the modulation of power and Phase Locking Value (PLV) at the syllabic and word frequencies. The power at the syllabic rate was analyzed in order to estimate basic auditory processing at the individual level and check which patients have intact auditory perception. Indeed, patients with severe auditory perception disorder would exhibit no frequency tagging at the syllabic rate (i.e., no stable auditory ERP). Auditory statistical learning and segmentation of the tri-syllabic words are revealed by an increase in the word-rate frequency and eventually its harmonics.

Not only does this paradigm address the question of conscious attention in auditory statistical learning, but also offers a promising way to enhance the clinical assessment of covert consciousness in patients with DOC. Previous studies have shown that EEG features are altered in relation to the level of consciousness. These include spectral power,[28] auditory ERP amplitude and latency,[19] and mismatch responses in EEG oddball paradigms.[29] The depth of language processing during time-locked natural language exposure has also been suggested as a useful metric for predicting patient outcomes.[30] Given the importance of language function[30–33] and correct metabolic functioning of the left middle temporal cortex[34] in the recovery of patients with DOC, investigating language-related paradigms is a promising avenue for the development of new clinical tools. Therefore, the present study also aims to investigate whether neural markers of auditory statistical learning can be used for diagnosis and outcome prediction in patients with DOC.

RESULTS

Evidence of word segmentation in the different experimental groups

To estimate whether patients were able to correctly segment the words concatenated in the structured stream, we measured the normalized PLV at the word rate and its first harmonic and compared the values to those measured in the random stream (Figure 2A).

Firstly, in healthy control subjects, numerous electrodes exhibited significant positive PLV at both the word rate and its

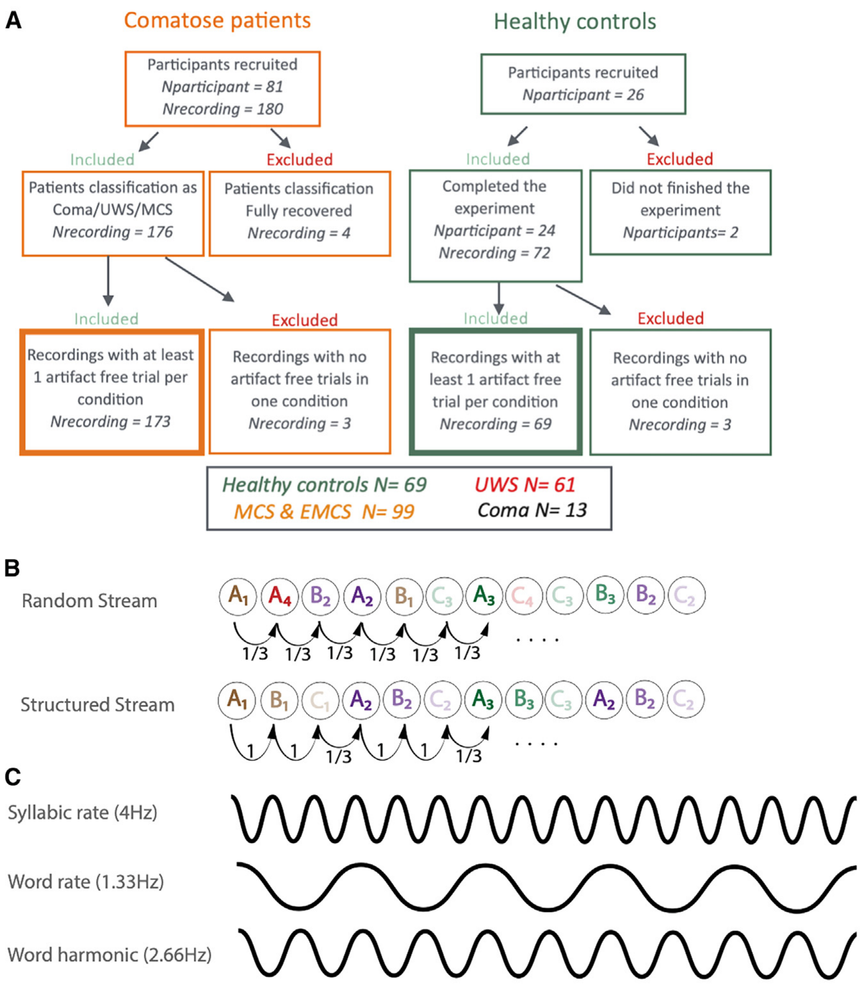


Figure 1. Presentation of the paradigm and dataset

(A) Flow chart of the inclusion of the recording. 180 recordings of comatose patients were performed. 4 were discarded because patients were classified as fully recovered and 3 more recordings were excluded because no artifact-free epochs could be found in the data, preventing further analysis. Similarly, 3 recordings were rejected out of 72 in the control group.

(B) Description of the two streams presented to participants. The random stream consists of syllables that can be followed by any of the other 11 syllables, resulting in a flat transitional probability of 1/11 throughout the sequence. The structured stream is composed of four tri-syllabic pseudo-words with transition probabilities between syllables equal to 1 inside the pseudo-words and 1/3 between the pseudo-words.

(C) Frequency tagging analyses: The syllables were presented at a rate of 4Hz, which is expected to elicit a 4Hz oscillation in the brains of normal hearing subjects. If, and only if, the structured sequence was segmented based on transition probabilities, the phase locking value (PLV) at the word rate (1.33Hz) and its harmonics (2.66Hz) should increase relative to the random stream.

harmonic when listening to the structured stream ($p < 0.05$ FDR corrected) but not to the random stream, with a significant difference between the two conditions in many electrodes (Figure 2). The same analysis in patients in the minimally conscious state (clinical assessment: MCS or EMCS) showed similar results, although with fewer significant electrodes. In the UWS group, a trend in the same directions as the other groups was visible (Structured stream at 2.66Hz: 23 channels with $p < 0.05$ uncorrected), but only one channel survived the FDR correction at the word rate harmonic. The comparison with the random stream showed a very modest effect (1 electrode with $p < 0.05$ FDR corrected). In the coma patients, no electrode showed a significant segmentation effect.

Correlation of the segmentation performance with coma recovery scale revised

We then estimated for each electrode the correlation of the word and harmonic PLV with the clinical Coma recovery scale revised (CRS-R) estimated just before the recording, excluding the healthy participants. We found a highly significant correlation spread on many channels of the frequency tagging (normalized PLV) and the CRS-R score during the structured stream only for the word first harmonic (24 channels with $p < 0.05$ FDR) with a significant difference with the random condition (Figure 2B).

Individual effect size

The effect size for each recording was estimated to assess inter-individual variability and to determine whether it was possible to separate subjects who segmented the structured sequence from those who did not (Figure 2C). This analysis replicated our previous results, with both healthy controls and patients with MCS showing an average effect size greater than zero for both the word rate and its harmonic. The distribution of the data revealed a high degree of inter-individual variability with no clear bi-modal distribution. This precludes the possibility to draw reliable conclusions about a patient's ability to perform the task on an individual level.

Controls taking into account low-level auditory perception as captured by the response at the syllabic rate

The previous correlation between PLV at the word harmonics and CRS-R score might indicate either a modulation in auditory statistical learning and segmentation abilities with DOC severity or a spurious correlation due to a higher number of patients with auditory perception impairment and a severe DOC. Fortunately, PLV at the syllabic rate effectively summarizes basic auditory perception and temporal synchronization of auditory ERP. We presented the distribution of the average PLV at a syllabic rate for each recording in Figure 3A. Therefore, we repeated the

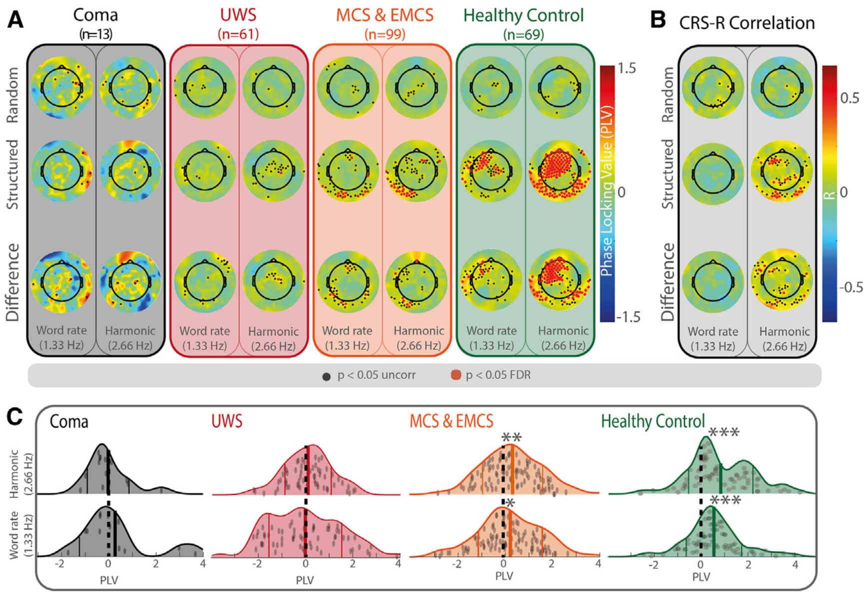


Figure 2. Frequency tagging analysis

(A) Normalized Phase Locking Value (PLV) for each electrode at the word rate (1.33Hz) and at its first harmonic (2.66Hz) during the random and structured streams in the recordings of Coma, UWS, MCS, and Healthy subjects. The bottom line presents the topography of the contrast [structured > random]. Dots represent electrodes with $p < 0.05$ before multiple comparison correction and red dots electrodes significant after FDR correction.

(B) Correlation of the PLV at word rate and its harmonic with Comatose Recovery Scale-Revised (CRS-R) in patients with DOC. The harmonic of the word rate (2.66Hz) significantly correlates with the clinical score only during the structured stream.

(C) Distribution of the effect size for each recording in each group and for the word rate (bottom row) and its harmonics (top row). Each dot represents the difference between structured and random PLV at the word and the harmonic rates for the 10 best electrodes in each condition of a particular recording. Bold vertical bars represent the average of the distributions. Significant difference in the distribution average with zero is represented by stars (* $<p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

previous analyses (Figure 2) while excluding recordings with negative syllabic rates to ensure that all recordings left are with patients with correct hearing and a level of signal/noise in the recording that enables correct measure of frequency tagging. The results remained similar, both considering the PLV measures in each group (Figure 3B) and the correlation with CRS-R (Figure 3C).

Finally, to mitigate the impact of the variation of the PLV at the syllabic rate of each electrode on the segmentation metrics, we regressed out the syllabic rate for each recording for each electrode before computing the correlation of the residuals with the CRS-R score. For this analysis, we included all recordings of patients with DOC, comprising those with negative average syllabic rates. We obtained results similar to the original analyses: 1) a significant correlation during the structured stream only with the first harmonic, but not at the fundamental; 2) a significant contrast between structured and random sequence streams (Figure 3D).

The analyses presented in Figures 2 and 3 have been replicated using power instead of PLV with similar results (Figures S1 and S2 in supplemental information).

Auditory ERP measured with syllabic rate frequency tagging

Independently from statistical learning, we found that the frequency tagging at the syllabic rate was highly correlated with the clinical assessment of the level of consciousness in patients with DOC. Indeed, we found many electrodes for which PLV and Power at the syllabic rate (4hz) robustly correlated with diagnostic scales, such as CRS-R (Figure 4). PLV at the syllabic rate significantly correlated with CRS-R on most electrodes. On Figure 4 B, we display the distribution of the average PLV value from the significant electrodes, for each subject by clinical assessment group. We clearly observe higher PLV at the syllabic rate associated with better clinical assessment. We also compared for each group the probability of improved outcome six months later depending on the PLV at the syllabic, word, and word harmonic rates and found no significant differences (for the syllabic rate analysis, recordings associated with an improved six-month outcome are displayed in green on Figure 4).

DISCUSSION

Frequency tagging provides a direct neural measure for monitoring word segmentation within a continuous speech stream obviating the need of an explicit behavioral response. This approach has proven its usefulness in assessing segmentation abilities in preverbal infants[7] and neonates.[20] Consequently, we proposed to apply this paradigm in patients with DOC as also did Xu et al.[26] Our first goal was to establish whether auditory statistical learning and auditory sequence segmentation might persist in patients with disorder of consciousness amidst conflicting literature on the role of conscious attention. Our second goal was to explore whether statistical learning metrics may help the clinical diagnosis and care of these patients.

Auditory statistical learning is partially preserved in patients with disorders of consciousness

The automaticity of statistical learning is still being debated. Indeed, while some studies showed a large decline in performance under divided attention[13,14] arguing for the need for focused attention on the task, others have reported that sequence segmentation persists even when outside the focus of attention[15,16,35] and even improves with cognitive fatigue that impairs focal attention.[17] Furthermore, sleeping neonates can automatically segment a stream based on its statistical

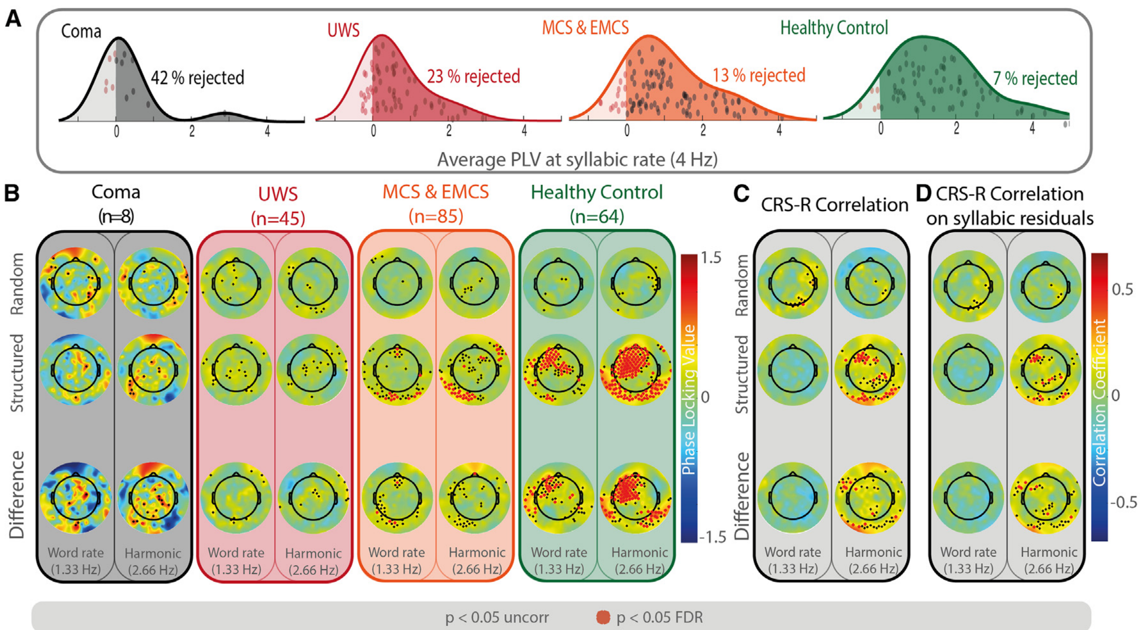


Figure 3. Frequency tagging analysis limited to subjects with a positive syllabic rate

(A) Average PLV at the syllabic rate (4Hz) for each recording (each dot represents one recording). Syllabic rate is a good metric of preserved auditory perception and correct signal/noise ratio in the recording. For the following analysis, we only kept the recordings with a positive average syllabic rate (black dots) as we cannot be sure that the other participants even heard the stimuli. Red dots represent recordings with negative syllabic rates, which were then excluded from the analyses presented in B and C.

(B) Topographies of the normalized PLV at each electrode at the word rate (1.33Hz) and its first harmonic (2.66Hz) during the random and structured streams in patients with Coma, UWS, MCS, and Healthy subjects, excluding recordings with negative average PLV at a syllabic rate (red dots on Panel A). The bottom line is the contrast structured > random. Black dots represent electrodes with $p < 0.05$ before multiple comparison correction and red dots electrodes significant after FDR correction.

(C) Topographies of the correlation score between the PLV at word rate (or harmonic) with CRS-R scores, excluding recordings with negative average PLV at syllabic rate. The harmonic of the word rate (2.66Hz) significantly correlates with this score during the structured stream only. (D) To account for the variation of frequency tagging at syllabic rate, we computed the correlation after having regressed out the effect of DOC on syllable entrainment, results remained similar.

properties.[20,22] Similarly, other studies investigating visual statistical learning with hypnosis,[36] or TMS disruption of DLPFC,[37] suggest that statistical learning abilities are enhanced when prefrontal activity is reduced.

Studies in patients with DOC provide valuable insight into this issue. In a recent study,[26] showed that some comatose patients were able to extract bi-syllabic real words (or mixed) arguing for a preserved minimal version of auditory statistical learning on pairs of real words. Here, we extend and strengthen this result by using a stream of tri-syllabic pseudo-words with flat intonation, i.e., without any additional linguistic cues to aid segmentation. Thus, any increase of PLV and power at the frequency of three syllables can only be based on the online calculation of the statistical relationship or transitional probability (TP), between the syllables, i.e., TP = 1 between syllables belonging to the same word and TP = 1/3 between syllables belonging to different words. We checked for any spurious effects by comparing this structured stream with a stream consisting of a concatenation of the same syllables with a flat probability (TP = 1/11). Significant differences between the two streams were observed in MCS and EMCS patient groups and possibly a very weak trend in patients with UWS. Since patients with MCS suffer from severe attentional dysfunction, this result provides evidence that the full focus of attention is not needed for this type of learning. Sleeping adults were only able to chunk bisyllabic pseudowords, unlike neonates who succeeded with trisyllabic pseudo-words as is the case here. It can be related to the linguistic expert adults' bias to segment the stream in shorter units (Franck et al., 2010) that might interfere more in a sleep stage than in a coma. In any case, chunking a stream in its tri-syllabic components reveals that even patients with MCS and EMC were able to integrate TP over several syllables to discover the TP drop that is used to chunk words.

Why are results more visible on the first harmonic compared to the word rate?

All our analyses pointed toward a greater frequency tagging at the first harmonic compared to the fundamental frequency at the word rate. In our design, the harmonic of the word rate (2.66Hz) is different from half of the frequency of the syllabic rate (2Hz); thus, the modulation seen here can only be due to the discovery of the word structure and not to the perception of the syllables. The evoked activity by each syllable and word superpose as there is no pause between syllables and words. The shape of this event-related activity is complex, and the Fourier transform decomposes it into a set of sinusoids with different

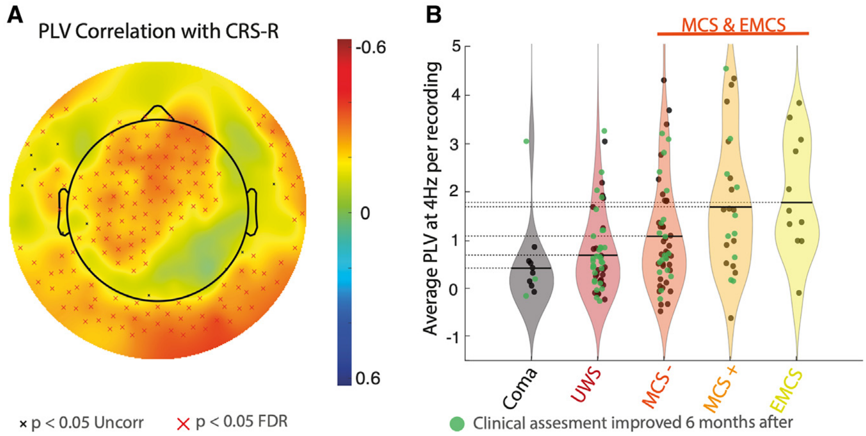


Figure 4. Syllabic rate correlation with CRS-R

(A) Topography of the correlation score between syllabic rate PLV and CRS-R. For each electrode, we correlated the PLV measure at the syllabic rate (average across random and structured conditions) with patients' CRS-R measured just before the recording.

(B) For better visualization of the effect, we report here the distribution of the average PLV at 4Hz across electrodes for each recording and patients' status. The average frequency tagging is modulated by the participant's depth of disorders of consciousness but the inter-recording variance within each diagnostic group stays higher than the variance explained by the depth of disorders of consciousness. Note that this figure is not independent from the analysis done in Panel A so we did not perform statistical analysis on this. This is just presented for a better visualization and estimation of the inter-recording variance. Recordings associated with an improved clinical assessment 6 months later are displayed in green. We found no systematic relation between PLV at the syllabic rate and the probability of improved clinical condition 6 months later.

powers depending on the ERP shape. A significant response at the first harmonic argues for a rhythmic response that vanishes faster than the word length, such as a larger response words' first syllable. In contrast, an activity drooling over the following syllables belonging to the word (e.g., integration of the three syllables) would be more visible at the fundamental frequency.[38] ERP shapes can explain the different sensitivity of the two measures observed in the above analyses. Further experiments are needed to investigate whether this difference reveals a different encoding of the word in memory. For instance, sleeping neonates segment a tri-syllabic non-word stream but only remember the first syllable of the words, contrary to adults who memorize the entire word.[20]

Correlation between the level of residual consciousness and statistical learning

We observed a significant correlation between CRS-R score and the first harmonic of the word rate in patients with DOC. This reveals that even though this auditory statistical learning task does not require attention, a deeper consciousness disorder penalizes learning. Therefore, we tried to separate whether impaired learning was related to deficient statistical computations or to a deficit in auditory perception due to degraded auditory/phonetic encoding or the result of the suboptimal synchronization of cortical activity with the stimulus. To do this, we used the 4Hz entrainment as a metric for the auditory low-level processing quality as well as a proxy of the recording quality which is sometimes impaired by the electrically noisy environment of the hospital wards. We linearly regressed it to CRS-R score. We then used the residuals of this regression in the correlation analysis with the word rate and its harmonic (Figure 3). Despite this stringent control, a significant correlation remained between CRS-R and the first harmonic only for the structured stream as well as a significant difference between the structured and random condition. This suggests that statistical learning abilities were affected by the degree of residual consciousness, even in cases where the brain exhibited the capacity to track the syllabic rhythm. It remains possible that in some cases, syllabic entrainment was based solely on the vocalic nucleus of the CV syllable, without the exact phonemes being encoded, thus ruling out the possibility of statistical learning. This could be the case for patients with lesions of the left perisylvian regions involved in phonetic processing. Fama et al.[39] reported that patients with stroke, notably those with anterior lesions, did not show evidence of statistical learning in a behavioral paradigm in which they had to rate their familiarity with words, part-words and non-words after 10 min of familiarization with a similar structured stream than here. Despite the high number of recordings studied here, we had not enough power to orthogonalize the DOC degree and the type, size, localization of the lesions. Further studies are needed, notably by testing musical tones to simplify the identification of the tokens in the stream.

Is there a clinical interest for this type of paradigm?

Despite the significant result described above, a word-segmentation task as implemented here might not be usable as a standalone clinical tool, although it could be relevant to include it in a battery of tests as this task targets a basic learning mechanism. The search for better indices of recovery, as well as indices quantifying the integrity of different brain functions beyond the anatomical lesions visible on MRI, is a necessity to guide care. However, here, the CRS-R effect size observed was smaller than the inter-individual variance and not highly significant, and language functions might be better quantified by sentences in the native language.[30,33,40] By contrast, the frequency tagging at the syllabic rate proved much more informative. Indeed, the correlation with CRS-R was greatly and highly significant and features of the auditory ERPs have been shown to be usable.[19] In our dataset, many electrodes showed a significant correlation between the auditory 4Hz frequency tagging and CRS-R measures. Steady-State measure is a more robust and time-economic way to elicit brain responses than isolated ERP, and frequency tagging is a more robust way, to look at the same neural response. This is confirmed by the typical auditory topography of the effect size of the correlation (Figure 4A). Previous studies suggested that the brain responses to the violation of auditory regularities, such as mismatch negativity or P300 waves,[32,41–44] can indicate the presence or absence of awareness in these patients. Thus, syllabic rate entrainment is a promising venue. Further research might be useful to better characterize which frequencies to be entrained are the most sensitive and which electrodes are the most informative for clinical application.

Conclusion

In this study, we discovered evidence of preserved auditory statistical learning of word boundaries in some patients with DOC using frequency tagging measurements. This result confirms our hypothesis that attention is not required for auditory statistical learning and extends the previous findings of efficient statistical learning abilities in sleeping neonates. Our study shows that statistical learning is an automatic process that scans the auditory environment even in conditions of disturbed conscious attention. In addition, we showed that these metrics of auditory statistical learning were significantly correlated with diagnostic metrics such as CRS-R, implying that they can be used as indicators of the level of consciousness and the prognosis of patients with DOC. Finally, we proposed that frequency tagging robustness could be of interest for better characterization of auditory ERP modification in patients with DOC, warranting further investigation of this measure in relation to the neural mechanisms and the clinical markers of consciousness disorders.

Limitations of the study

Running EEG in an ICU environment comes with limitations in terms of noise (electrical noise more pronounced than in Lab environment with Faraday cage, involuntary movements of patients ….). As described in the discussion, this high noise might prevent meaningful interpretation at the subject level, especially in a clinical context.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for data and the scripts should be directed to and will be fulfilled by the lead contact, Ghislaine Dehaene-Lambertz (ghislaine.dehaene@cea.fr).

Materials availability

Stimuli and MATLAB scripts used for this study have been deposited to https://osf.io/k5tsq/.

Data and code availability

- Data. All data reported in this article will be shared by the lead contact upon request.
- Code. The main scripts are provided at https://osf.io/k5tsq/.
- Other. Sample stimuli are accessible at https://osf.io/k5tsq/.
- Any additional information required to reanalyze the data reported in this article is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all patients and families who accepted to participate in the study. We also would like to thank the Huashan and Hebin hospitals, and the MEG Facility of CAS CEBSIT/ION in data collection. LB thanks the Treilles foundation for its support. LB, AF, and GDL were funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 695710) obtained by GDL. This work was also supported by the National Natural Science Foundation of China (grant No. 82201352) and the CAS Youth Innovation Promotion Association to PG, the CAS Project for Young Scientists in Basic Research YSBR-071 and the Lingang Laboratory Grant (LG202105-02-01) to LW, National High Level Hospital Clinical Research Funding (No. 2023-NHLHCRF-BQ-43) to DZ, Shanghai Zhou Liangfu Medical Development Foundation "Brain Science and Brain Diseases Youth Innovation Program" to ZQ, the National Natural Science Foundation of China (grant No. 82271224) to XW.

AUTHOR CONTRIBUTIONS

Design of the study: LB, AF, PG, and GDL. Data collection: DZ and PG. Data Analysis: LB, GDL, and PG. Results interpretation: LB, DZ, AF, DZ, and PG. Patients' coordination: PS and WZ. Original writing of the article: LB and GDL. Corrections and comments: All authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Comatose patients
  - Healthy control participants
- METHOD DETAILS
  - Stimuli
  - Procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data preprocessing
  - Frequency tagging
  - Statistical analyses
  - Individual effect size
  - Modulation of the neural responses by level of consciousness
  - Modulation of words segmentation by the syllabic response

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111591.

Received: January 13, 2024
Revised: November 10, 2024
Accepted: December 10, 2024
Published: December 12, 2024

REFERENCES

1. Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996). Statistical Learning by 8-Month-Old Infants. Science *274*, 1926–1928.
2. Saffran, J.R., Johnson, E.K., Aslin, R.N., and Newport, E.L. (1999). Statistical learning of tone sequences by human infants and adults. Cognition *70*, 27–52.
3. Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., and Kolinsky, R. (2008). Songs as an aid for language acquisition. Cognition *106*, 975–983.
4. Fiser, J., and Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. Proc. Natl. Acad. Sci. USA *99*, 15822–15826.
5. Hauser, M.D., Newport, E.L., and Aslin, R.N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. Cognition *78*, 53–64.

6. James, L.S., Sun, H., Wada, K., and Sakata, J.T. (2020). Statistical learning for vocal sequence acquisition in a songbird. Sci. Rep. *10*, 2248.
7. Kabdebon, C., Pena, M., Buiatti, M., and Dehaene-Lambertz, G. (2015). Electrophysiological evidence of statistical learning of long-distance dependencies in 8-month-old preterm and full-term infants. Brain Lang. *148*, 25–36.
8. Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., and Pallier, C. (2015). The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. Neuron *88*, 2–19.
9. Garvert, M.M., Dolan, R.J., and Behrens, T.E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. Elife *6*, e17086.
10. Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., and Botvinick, M.M. (2013). Neural representations of events arise from temporal community structure. Nat. Neurosci. *16*, 486–492.
11. Benjamin, L., Fló, A., Al Roumi, F., and Dehaene-Lambertz, G. (2023). Humans parsimoniously represent auditory sequences by pruning and completing the underlying network structure. Elife *12*, e86430.
12. Benjamin, L., Sablé-Meyer, M., Fló, A., Dehaene-Lambertz, G., and Al Roumi, F. (2024). Long-Horizon Associative Learning Explains Human Sensitivity to Statistical and Network Structures in Auditory Sequences. J. Neurosci. *44*, e1369232024.
13. Fernandes, T., Kolinsky, R., and Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. Atten. Percept. Psychophys. *72*, 1522–1532.
14. Toro, J.M., Sinnett, S., and Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. Cognition *97*, 25–34.
15. Batterink, L.J., and Paller, K.A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. Cortex *115*, 56–71.
16. Benjamin, L., Dehaene-Lambertz, G., and Fló, A. (2021). Remarks on the analysis of steady-state responses: spurious artifacts introduced by overlapping epochs. Cortex *142*, 370–378. https://doi.org/10.1016/j.cortex.2021.05.023.
17. Smalle, E.H.M., Daikoku, T., Szmalec, A., Duyck, W., and M€ Ott€ Onen, R. (2022). Unlocking adults' implicit statistical learning by cognitive depletion. Proc. Natl. Acad. Sci. USA *119*, e2026011119. https://doi.org/10.1073/pnas.2026011119/-/DCSupplemental.
18. Batterink, L.J., and Zhang, S. (2022). Neuropsychologia Simple statistical regularities presented during sleep are detected but not retained. Neuropsychologia *164*, 108106.
19. Strauss, M., Sitt, J.D., King, J.R., Elbaz, M., Azizi, L., Buiatti, M., Naccache, L., van Wassenhove, V., and Dehaene, S. (2015). Disruption of hierarchical predictive coding during sleep. Proc. Natl. Acad. Sci. USA *112*, E1353–E1362.
20. Fló, A., Benjamin, L., Palu, M., and Dehaene-Lambertz, G. (2022). Sleeping neonates track transitional probabilities in speech but only retain the first syllable of words. Sci. Rep. *12*, 4391.
21. Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., and Ferry, A.L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. Dev. Sci. *22*, e12802.
22. Benjamin, L., Fló, A., Palu, M., Naik, S., Melloni, L., and Dehaene-Lambertz, G. (2023). Tracking transitional probabilities and segmenting auditory sequences are dissociable processes in adults and neonates. Dev. Sci. *26*, e13300.
23. Giacino, J.T., Kalmar, K., and Whyte, J. (2004). The JFK Coma Recovery Scale-Revised: Measurement characteristics and diagnostic utility1. Arch. Phys. Med. Rehabil. *85*, 2020–2029.
24. Bruno, M.-A., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., and Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. J. Neurol. *258*, 1373–1384.
25. Giacino, J. T., Ashwal, S., Childs, N., Cranford, R., Jennett, B., Katz, D. I., Kelly, J. P., Rosenberg, J. H., Whyte, J., Zafonte, R. D., et al. (2002). The minimally conscious state: Definition and diagnostic criteria. Neurology *58*, 349–353.
26. Xu, C., Li, H., Gao, J., Li, L., He, F., Yu, J., Ling, Y., Gao, J., Li, J., Melloni, L., et al. (2023). Statistical learning in patients in the minimally conscious state. Cereb. Cortex *33*, 2507–2516. https://doi.org/10.1093/cercor/bhac222.
27. Fló, A., Benjamin, L., Palu, M., and Dehaene-Lambertz, G. (2024). Statistical learning beyond words in human neonates. eLife *13*. https://doi.org/10.7554/eLife.101802.1.
28. Goldfine, A.M., Victor, J.D., Conte, M.M., Bardin, J.C., and Schiff, N.D. (2011). Determination of awareness in patients with severe brain injury using EEG power spectral analysis. Clin. Neurophysiol. *122*, 2157–2168.
29. Laureys, S., and Schiff, N.D. (2012). Coma and consciousness: Paradigms (re)framed by neuroimaging. Neuroimage *61*, 478–491.
30. Gui, P., Jiang, Y., Zang, D., Qi, Z., Tan, J., Tanigawa, H., Jiang, J., Wen, Y., Xu, L., Zhao, J., et al. (2020). Assessing the depth of language processing in patients with disorders of consciousness. Nat. Neurosci. *23*, 761–770.
31. Claassen, J., Doyle, K., Matory, A., Couch, C., Burger, K.M., Velazquez, A., Okonkwo, J.U., King, J.R., Park, S., Agarwal, S., et al. (2019). Detection of Brain Activation in Unresponsive Patients with Acute Brain Injury. N. Engl. J. Med. *380*, 2497–2505.
32. Sanz, L.R.D., Thibaut, A., Edlow, B.L., Laureys, S., and Gosseries, O. (2021). Update on neuroimaging in disorders of consciousness. Curr. Opin. Neurol. *34*, 488–496.
33. Sokoliuk, R., Degano, G., Banellis, L., Melloni, L., Hayton, T., Sturman, S., Veenith, T., Yakoub, K.M., Belli, A., Noppeney, U., and Cruse, D. (2021). Covert Speech Comprehension Predicts Recovery From Acute Unresponsive States. Ann. Neurol. *89*, 646–656.
34. Aubinet, C., Cassol, H., Gosseries, O., Bahri, M.A., Larroque, S.K., Majerus, S., Martial, C., Martens, G., Carrière, M., Chatelle, C., et al. (2020). Brain Metabolism but Not Gray Matter Volume Underlies the Presence of Language Function in the Minimally Conscious State (MCS): MCS+ Versus MCS− Neuroimaging Differences. Neurorehabil. Neural Repair *34*, 172–184.
35. Batterink, L.J., and Choi, D. (2021). Optimizing steady-state responses to index statistical learning: Response to Benjamin and colleagues. Cortex *142*, 379–388. https://doi.org/10.1016/j.cortex.2021.06.008.
36. Nemeth, D., Janacsek, K., Polner, B., and Kovacs, Z.A. (2013). Boosting Human Learning by Hypnosis. Cereb. Cortex *23*, 801–805.
37. Ambrus, G.G., Vékony, T., Janacsek, K., Trimborn, A.B., Kovács, G., and Nemeth, D. (2020). When less is more: Enhanced statistical learning of non-adjacent dependencies after disruption of bilateral DLPFC. J. Mem. Lang. *114*, 104144.
38. Zhou, H., Melloni, L., Poeppel, D., and Ding, N. (2016). Interpretations of Frequency Domain Analyses of Neural Entrainment: Periodicity, Fundamental Frequency, and Harmonics. Front. Hum. Neurosci. *10*, 274.
39. Fama, M.E., Schuler, K.D., Newport, E.L., and Turkeltaub, P.E. (2022). Effects of healthy aging and left hemisphere stroke on statistical language learning. Lang. Cogn. Neurosci. *37*, 984–999.
40. Edlow, B.L., and Naccache, L. (2021). Unmasking Covert Language Processing in the Intensive Care Unit with Electroencephalography. Ann. Neurol. *89*, 643–645.
41. Chennu, S., and Bekinschtein, T.A. (2012). Arousal Modulates Auditory Attention and Awareness: Insights from Sleep, Sedation, and Disorders of Consciousness. Front. Psychol. *3*, 65.
42. Qin, P., Di, H., Yan, X., Yu, S., Yu, D., Laureys, S., and Weng, X. (2008). Mismatch negativity to the patient's own name in chronic disorders of consciousness. Neurosci. Lett. *448*, 24–28.
43. Rivera-Lillo, G., Rojas-Líbano, D., Burgos, P., Egaña, J.I., Chennu, S., and Maldonado, P.E. (2018). Reduced delta-band modulation underlies the loss of P300 responses in disorders of consciousness. Clin. Neurophysiol. *129*, 2613–2622.

44. Wang, X.-Y., Wu, H.Y., Lu, H.T., Huang, T.T., Zhang, H., and Zhang, T. (2017). Assessment of mismatch negativity and P300 response in patients with disorders of consciousness. Eur. Rev. Med. Pharmacol. Sci. *21*, 4896–4906.
45. Zhang, G., Garrett, D.R., Simmons, A.M., Kiat, J.E., and Luck, S.J. (2023). Evaluating the effectiveness of artifact correction and rejection in event-related potential research. Preprint at: bioRxiv. https://doi.org/10.1101/2023.09.16.558075
46. Jonathan, D. & Reece, D. eSpeaker, [Computer program] retrieved October 2020. (2020). https://espeak.sourceforge.net/
47. Boersma, P., and Weenink, D. (2001). Praat, a system for doing phonetics by computer. Glot International *5*, 341–345.
48. Fló, A., Gennari, G., Benjamin, L., and Dehaene-Lambertz, G. (2022). Automated Pipeline for Infants Continuous EEG (APICE): A flexible pipeline for developmental cognitive studies. Dev. Cogn. Neurosci. *54*, 101077.
49. Desjardins, J.A., van Noordt, S., Huberty, S., Segalowitz, S.J., and Elsabbagh, M. (2021). EEG Integrated Platform Lossless (EEG-IP-L) pre-processing pipeline for objective signal quality assessment incorporating data annotation and blind source separation. J. Neurosci. Methods *347*, 108961.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Stimuli	This paper	https://osf.io/k5tsq/
EEG Analysis	MATLAB/This paper	https://osf.io/k5tsq/
EEG preprocessing	APICE	https://github.com/neurokidslab/eeg_preprocessing
Stimuli presentation	MATLAB/This paper	https://osf.io/k5tsq/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Comatose patients

81 patients from Huashan hospital with disorders of consciousness were included in the experiment for a total of 180 recordings (including 53 in females) from March 2021 to September 2022. Clinical assessment of the patient state was made just before each recording. 4 recordings were discarded as participants were labeled as fully recovered. We also rejected recordings that had not at least one artifact free epoch for each condition (random and structured) to increase signal/noise ration and prevent false positive or false negative results.[45] Indeed, in the absence of at least one epoch per condition, we could not estimate the PLV difference between random and structured sequences. 3 recordings were then rejected based on this criterion leaving a total of 173 recordings. Out of those 173 recordings, 13 were classified as coma, 61 as UWS and 99 as minimally conscious or emerging of consciousness. The experiment consisted of three stimulus sequences (see details in the following), of which 78 patients completed 84 recordings of list 1, 50 patients completed list 2 recordings and 46 patients completed list 3 recordings. Therefore, a total of 40 patients completed three different lists at different time points.

Healthy control participants

26 healthy control adults (>18yo, average age:25.27; 18 females and 9 males) from local community were also tested along which 24 completed the experiments (72 recordings). Similar to the patient group, we rejected 3 recordings that had not at least one artifact-free epoch in each condition, and thus analyzed the 69 remaining recordings. All 24 healthy volunteers completed all three lists.

No male/female differences was expected in this experiment so we recruited patients regardless of their sex.

The study protocol was approved by the Ethical Committee of Huashan Hospital of Fudan University (approval no. HIRB-2014-281 and updated approval no. HIRB-2023-051), and informed consent was obtained from all healthy participants and caregivers of all patients. All documents were submitted and archived at the Ethical Committee of Huashan Hospital of Fudan University.

METHOD DETAILS

Stimuli

We synthesized the syllables using the open-source text-to-speech synthesizer eSpeaker[46] with Mandarin Chinese language (zh) and the female voice variant f2. We set the pitch parameter to 70 and adapted the speed to obtain syllables with a duration as close as possible to 250 ms. We further corrected the syllables using Praat.[47] Specifically, we removed silent periods in the beginning and the end to obtain syllables lasting exactly 250ms and removed pitch changes setting a constant pitch of 225 Hz. The syllables audio files were concatenated without pauses to obtain the streams, and the first and last 4.5 s were ramped up and down to avoid the start and end of the stream might serve as perceptual anchors.

The structured streams consisted of a semi-random concatenation of four tri-syllabic pseudowords. The only restriction on the concatenation process was that the same pseudoword could not appear twice in a row, and that the same two pseudowords could not repeatedly alternate more than two times (i.e., the sequence WkWjWkWj, where Wk and Wj are two words, was forbidden). In order to prevent the phonetic features of the pseudowords from serving as segmentation cues, we balanced these features across the three syllables of the pseudowords. In addition, we created three different structured streams by changing the arrangement of the 12 syllables (each syllable occupied a different position within the pseudoword in each stream). The random stream resulted from concatenating the 12 syllables semi-randomly (without syllables repetition), giving an average uniform TP of 1/11. We also created test words to be presented in isolation but we do not present the results here.

Procedure

Scalp electrophysiological activity was recorded using a 256-electrodes net (GTEN 200, Magstim EGI) referred to the vertex with a sampling frequency of 1000 Hz. The recording procedure was similar to the one used in neonates.[20] Participants first heard a random sequence of 4 minutes (960 syllables) followed by a structured stream of 4 minutes (960 syllables, 320 words). After, participants listen to 8 repetitions of 30s (120 syllables, 40 words) of structured sequences followed by a block of 16 pseudo-words test trials. Patients were told to pay attention to the auditory stimuli.

All Healthy subjects were tested with the three structured streams (lists) on different days. For the DOC patients, we tried to follow the same pattern by testing each subject on each list as much as the hospital constraints allowed us. On average DOC subjects were tested 2.27 times, with 40 subjects having been tested on the three lists. We obtained 88 recordings with list 1, 50 with list 2 and 46 with list 3.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data preprocessing

Data were resampled to 250 Hz, band-pass filter 0.3–30 Hz and pre-processed using APICE pipeline for MATLAB[48] with default parameters for rejection. The process, similar to other preprocessing pipelines like Desjardin et al.[49] consists of several key steps: 1) Identification of motion artifacts in continuous data using relative thresholds applied to individual electrodes through iterative loops. Specifically, any sample with a value exceeding a set threshold (e.g., further away than 2.5 standard deviation from the average of the distribution) was marked as bad; 2) Correction of artifacts in continuous data when they involve only a few channels or occur over a short period of time, taking advantage of EEG data redundancy to reconstruct the brief portions of the affected signal; 3) Definition of contaminated samples (bad times) and non-functional channels (bad channels) based on the rejected data. Indeed, long periods of artifacts, or simultaneous failure of several channels cannot be corrected with neighbors' data and the corresponding epochs are removed from further analysis. From this artifact detection and correction, epochs with more than 50% of its data interpolated or epochs with more than 15% of remaining bad data were discarded from further analyses.

To assess the robustness of the analysis, the replicated the main effect (CRS-R correlation with PLV) by keeping all epochs from all subjects despite noise and found nosier but still significant correlation (some electrodes with ps<0.05 FDR, see Figure S3).

In order to remove eye movements and blinks, we also performed ICA on the healthy control data.

Frequency tagging

The pre-processed data were segmented from the beginning of each sequence into segments comprising 13 words to approach 10s long epochs (13*0.750=9.75s). Segments were not overlapping to avoid an artefact in the frequency domain related to the length of the overlap.[16] Epochs with artifacts were rejected. Data were converted to the frequency domain using the Fast Fourier Transform (FFT) algorithm and the Phase Locking Value (PLV) and Power were estimated for each electrode in both random and structured conditions. The phase locking value ranges from 0 (completely asynchronized data) to 1 (completely timed-locked activity). The value at each frequency bin estimation was then normalized by subtracting the mean value of eight neighboring frequency bins on each side. All analysis have been replicated using power instead of PLV (see supplemental information).

Statistical analyses

Statistical analysis was conducted on the three frequencies of interest: 4Hz corresponding to the syllabic rate; 1.33 and 2.66 corresponding to the frequency of the trisyllabic words and its harmonic. The results of the analyses performed on the PLV are presented in the main text, while those for power are presented in the supplemental information. Results of the two metrics are largely similar.

For each electrode and for each frequency, it was first tested whether the PLV (and power) was above 0 with a one-sample t-test against zero, second whether these values for the word frequencies were larger during the structured stream relative to the random stream (structured > random one-way paired t-test). In all analyses, p-values were corrected for multiple comparisons (256 electrodes) using FDR (Figure 2).

Individual effect size

Subsequently, the effect size for each recording was estimated in order to determine whether it was possible to distinguish between patients who performed the task from those who did not segment the structured sequence. To achieve this, we calculated the difference at the frequency of interest between the 10 electrodes with the highest PLV in the structured condition and the 10 electrodes with the highest PLV in the random condition. The rationale behind selecting the 10 best electrodes per recording was to obtain a robust measurement of the effect by selecting the most effective electrodes without being affected by possible differences in topography at the individual level induced by different brain lesions.

This analysis was conducted at the word rate (1.33Hz) and its first harmonic (2.66 Hz) and used one-way tests to compare the mean of the distribution with zero (no PLV difference between structured and random sequences).

Modulation of the neural responses by level of consciousness

To examine how the responses were influenced by the level of consciousness in DOC patients, we computed correlations between the Phase Locking Value (PLV) of each electrode and the Coma Recovery Scale-Revised (CRS-R) score measured prior to the recording in DOC patients only. Importantly, healthy controls were not included in this analysis to ensure that any observed correlation was not solely driven by the differences between healthy and DOC subjects, but rather by the varying degrees of consciousness within the DOC group.

Modulation of words segmentation by the syllabic response

To be able to segment words in such a stream, patients should at least have a minimal auditory function and some patients may not meet this criterion. We considered the responses at the syllabic rate as a proxy of low-level auditory function. Therefore, the previous analyses were done again retaining only patients with a positive average syllabic rate PLV, that is patients for whom the mean PLV across all electrodes for both the random and structured conditions was > 0 (Figure 3). This rejection metric is quite stringent as it rejected participants with impaired auditory processing but also recording with too low signal/noise ratio to correctly detect this metric. We also re-calculated the correlation between the PLV at the word rates and level of consciousness after regressing out the PLV at the syllabic rate for each electrode: we first performed a regression between word rate PLV and the syllabic rate PLV for each electrode; then the residuals of these regressions were correlated with CRS-R (Figure 3D).