

Current Biology

Simultaneous mnemonic and predictive representations in the auditory cortex

Highlights

- Stimulus memory and prediction can be simultaneously decoded from neural responses
- Decoding memory and prediction relies on uncorrelated data features
- Predictive representations are dynamically updated over the course of stimulation
- Memory and prediction are decodable in passive sequence processing under anesthesia

Authors

Drew Cappotto, Kongyan Li,
Lucia Melloni, Jan Schnupp,
Ryszard Auksztulewicz

Correspondence

drew.cappotto@my.cityu.edu.hk

In brief

Cappotto et al. show that mnemonic and predictive representations of auditory stimuli can be simultaneously decoded from neural activity at overlapping latencies and based on largely uncorrelated data features. Predictive representations are dynamically updated over the course of stimulation, suggesting a gradual formation of prediction.

Report

Simultaneous mnemonic and predictive representations in the auditory cortex

Drew Cappotto,^{1,5,*} HiJee Kang,¹ Kongyan Li,¹ Lucia Melloni,² Jan Schnupp,^{1,4} and Ryszard Auksztulewicz^{1,2,3,4}

¹Department of Neuroscience, City University of Hong Kong, 31 To Yuen Street, Kowloon Tong, Hong Kong

²Neural Circuits, Consciousness and Cognition Research Group, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany

³European Neuroscience Institute Göttingen: A Joint Initiative of the University Medical Center Göttingen and the Max Planck Society, Grisebachstraße 5, 37077 Göttingen, Germany

⁴These authors contributed equally

⁵Lead contact

*Correspondence: drew.cappotto@my.cityu.edu.hk

<https://doi.org/10.1016/j.cub.2022.04.022>

SUMMARY

Recent studies have shown that stimulus history can be decoded via the use of broadband sensory impulses to reactivate mnemonic representations.^{1–4} However, memories of previous stimuli can also be used to form sensory predictions about upcoming stimuli.^{5,6} Predictive mechanisms allow the brain to create a probable model of the outside world, which can be updated when errors are detected between the model predictions and external inputs.^{7–10} Direct recordings in the auditory cortex of awake mice established neural mechanisms for how encoding mechanisms might handle working memory and predictive processes without “overwriting” recent sensory events in instances where predictive mechanisms are triggered by oddballs within a sequence.¹¹ However, it remains unclear whether mnemonic and predictive information can be decoded from cortical activity simultaneously during passive, implicit sequence processing, even in anesthetized models. Here, we recorded neural activity elicited by repeated stimulus sequences using electrocorticography (ECoG) in the auditory cortex of anesthetized rats, where events within the sequence (referred to henceforth as “vowels,” for simplicity) were occasionally replaced with a broadband noise burst or omitted entirely. We show that both stimulus history and predicted stimuli can be decoded from neural responses to broadband impulses, at overlapping latencies but based on independent and uncorrelated data features. We also demonstrate that predictive representations are dynamically updated over the course of stimulation.

RESULTS

In the present experiment, we adapt recent techniques for decoding auditory working memory traces^{1,3,4} to simultaneously probe both memory and predictive processes. Electrocorticography (ECoG) was recorded from the auditory cortex (AC; [Figure S1A](#)) of anesthetized rats ($n=8$) while repeated stimulus streams of vowels were presented, with vowels occasionally omitted or replaced with a broadband noise burst ([Figure 1A](#)). Two types of blocks were employed. In “predictable” blocks, vowels were grouped into one of six triplets (AAO, AOO, AAI, AII, OOI, or OII) with each triplet presented at least 25 times in a given block of identical triplets before being replaced with another triplet (see [STAR Methods](#)). In control blocks, we presented the vowels in a pseudo-randomized order while keeping the position of bursts and omissions fixed (relative to their corresponding predictable block) to tap into mnemonic processing without predictive components ([Figure 1B](#)). In both types of blocks, 5% of vowels were replaced with omissions and 5% with bursts.

Univariate analyses: Only vowel-evoked activity differentiates between vowels

To test whether vowel identity influences average neural activity, we tested for the effects of a vowel (A, I, or O) and a block (predictable vs. random) on vowel-evoked ECoG activity (event-related potentials). We observed that vowel-evoked activity differentiates between the three vowels, both in predictable blocks ([Figure S2A](#); 13–260 ms; $F_{\max} = 58.56$; $p_{\text{FWE}} < 0.001$) and in random blocks ([Figure S2B](#); 13–207 ms; $F_{\max} = 58.21$; $p_{\text{FWE}} < 0.001$). The main effect of block (predictable vs. random) on vowel-evoked activity was not significant (all $p_{\text{FWE}} > 0.05$).

We then tested whether burst-evoked and/or omission-evoked activity also differentiates between the (preceding) vowels at different “positions” in the sequence, relative to the burst/omission (N-1 position: the immediately preceding vowel; N-2 position: two stimuli before the burst/omission; or N-3 position: three stimuli before the burst/omission). This analysis revealed that, similarly to the vowel-evoked responses, burst-evoked responses did not significantly differentiate between predictable and random blocks ([Figure S2C](#); all $p_{\text{FWE}} > 0.05$).

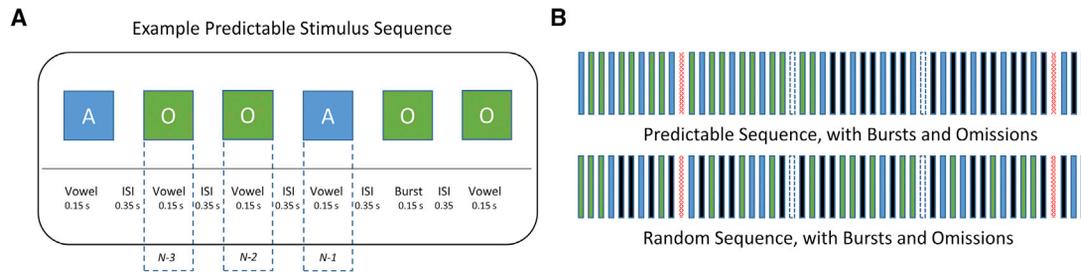


Figure 1. Stimulus sequences

(A) An example of an AOO predictable stimulus sequence, where one vowel of the triplet has been randomly substituted by a noise burst (or alternately omitted entirely [data not shown]) following a minimum of three triplet repetitions. In paired random blocks, the relative position of the burst/omission substitution remains unchanged, while the surrounding vowels are randomized. Vowel positions relative to the burst/omission are denoted as N-1, N-2, and N-3.

(B) Segment of an example predictable sequence in which vowel tokens are omitted or replaced with a noise burst after three repetitions (top) and the randomized version of that sequence where vowel tokens from the full sequence are presented pseudo-randomly while burst and omission tokens remained in the same relative positions (bottom).

However, unlike the vowel-evoked activity (which was modulated by vowel identity), noise-burst-evoked activity was not significantly modulated by (preceding) vowel identity when neural activity was analyzed in a mass-univariate manner. Specifically, neither the effect of the immediately preceding vowel on burst responses (N-1: all $p_{\text{FWE}} > 0.05$; [Figure S2D](#)) nor of the previous vowels (N-2 or N-3: all $p_{\text{FWE}} > 0.05$) were significant.

Omission-evoked responses peaked relatively early (83–93 ms) with a rising activity visible already prior to expected stimulus onset, possibly marking the offset response to the interrupted stimulus train rather than a true omission.¹² Nevertheless, just like burst-evoked activity, omission-related activity was also not significantly modulated by block type ([Figure S2E](#); all $p_{\text{FWE}} > 0.05$) or preceding vowel identity ([Figure S2F](#); N-1, N-2, or N-3: all $p_{\text{FWE}} > 0.05$).

Multivariate analysis: specific decoding boost for predictable vowels

Although in the univariate analysis burst-evoked activity did not differentiate between preceding vowels, based on our previous study,¹ we hypothesized that preceding stimuli can be decoded in a multivariate analysis. Specifically, by analyzing the spatio-temporal pattern of activity evoked by noise bursts, which did not carry overt information about the preceding vowels given that noise tokens were always identical and presented after vowel-evoked responses had returned to baseline (400 ms after stimulus offset), we sought to determine if activity evoked by noise bursts contained information about the preceding vowels (separately for N-1, N-2, and N-3 vowels). This analysis revealed significant decoding of vowels up to N-3 in predictable blocks and up to N-2 in random blocks ([Figure 2A](#); [Table S1](#)). Overall, immediately preceding stimuli could be decoded better than previous stimuli ([Table S2](#)) but not as well as currently processed stimuli ([Table S3](#)).

Crucially, if burst-evoked activity can reactivate not only mnemonic representations (irrespective of the currently processed stimulus) but also predictive representations (tokens that would have been predicted but are replaced by a noise burst), we would expect a specific decoding improvement for N-3 (but not N-2 or N-1) vowels presented in predictable blocks vs. random blocks. The decoding results were

consistent with this hypothesis. Specifically, decoding was significantly improved for the N-3 vowels presented in predictable blocks relative to the random blocks (paired *t* test, early cluster: 77–103 ms, $t_{\text{max}} = 3.45$, cluster-level $p_{\text{FWE}} = 0.010$; late cluster: 227–270 ms, $t_{\text{max}} = 3.79$, cluster-level $p_{\text{FWE}} < 0.001$; [Figure 2A](#)), suggesting that we could access a predictive representation of the vowel replaced by a noise burst. In a follow-up analysis using representational dissimilarity matrices, we found that this predictive representation contained information not only about the specific N-3 vowel replaced by the burst but also about the entire triplet preceding the burst ([Figure S3](#)).

While mnemonic and predictive representations could be decoded based on burst-evoked activity, decoding stimulus history based on omission-evoked activity did not yield any significant results ([Figure 2B](#); all $p_{\text{FWE}} > 0.05$). This suggests that, at least in this experimental protocol (vowel triplets) and in ECoG recorded under anesthesia, a stronger activation of the network (e.g., burst presentation) is necessary to make mnemonic and/or predictive representations observable.

Multivariate analysis: Decoding of predicted vowels gradually improves over time

Having established that decoding of the predicted vowel (N-3) shows a specific improvement in predictable vs. random blocks, we sought to determine whether this boost shows features of a predictive representation. We reasoned that, in predictable blocks, predictions should be learned over time and, consequently, the decoding of the N-3 vowel should gradually improve within and across blocks containing identical triplets ([Figures 3A](#) and [3E](#)). To test this, we performed a linear regression analysis on single-trial decoding estimates, using two “learning” regressors—one quantifying possible gradual improvements of decoding within each sequence containing identical triplets (within blocks) and one quantifying possible gradual improvements of decoding over the course of the entire recording session (across blocks). We treated the random blocks as a control for passage of time (including gradual suppression of activity due to habituation, short-term plasticity to repeated presentations of stimuli, and changes in stimulus-related and baseline activity due to prolonged anesthesia), since no learning was expected in this

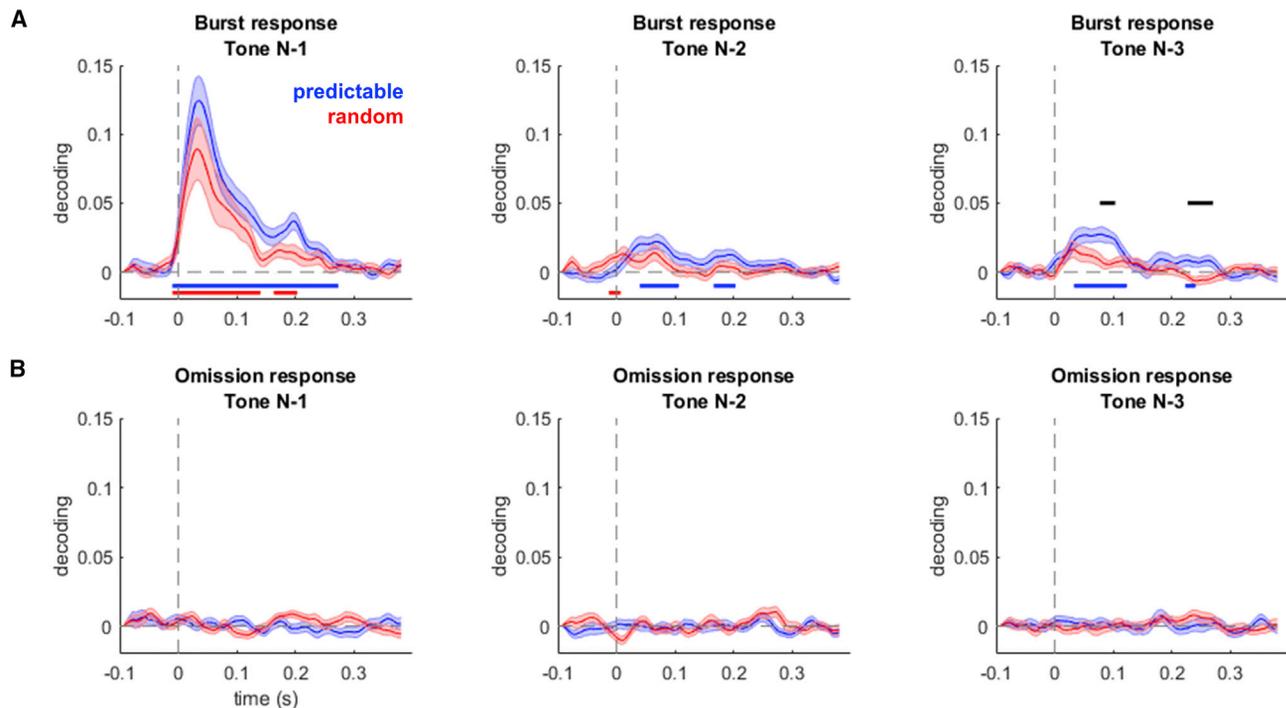


Figure 2. Multivariate analyses

(A) Time courses of decoding of preceding vowels based on burst-evoked activity. Left/middle/right panel: decoding N-1/N-2/N-3 vowel (blue, predictable blocks; red, random blocks; shaded area, SEM across recording sessions; blue/red horizontal line, decoding in predictable/random blocks significantly different from zero, $p_{FWE} < 0.05$; black horizontal line, decoding significantly different between predictable and random blocks, $p_{FWE} < 0.05$); shaded area, SEM across recording sessions. See also Tables S1–S3.

(B) Decoding based on omission responses. Legend as above.

condition. This analysis revealed that, for the early time window in which we observed a decoding boost in the predictable vs. random condition (77–103 ms), the “within-blocks” learning effect was significantly higher in the predictable than in random blocks (Wilcoxon sign rank test, $Z_{21} = 2.485$, $p = 0.013$; Figures 3B–3D), although significance testing of regression coefficients within conditions against zero did not yield significant effects (predictable: $Z_{21} = 1.477$, $p = 0.139$; random: $Z_{21} = -1.825$, $p = 0.068$). No significant learning effects across blocks were observed for the early time window (all $p > 0.5$). Conversely, for the later time window in which we observed a decoding boost (227–270 ms), the “across-blocks” learning effect (Figures 3F–3H) showed borderline significance in the predictable condition against zero ($Z_{21} = 2.033$, $p = 0.042$; uncorrected), but not in the random condition ($Z_{21} = 0.122$, $p = 0.903$), although a direct comparison of learning coefficients between conditions did not yield a significant effect ($Z_{21} = 1.303$, $p = 0.192$). The “within-blocks” learning effect did not yield any significant effects in the later time window (all $p > 0.5$). An additional analysis of N-2 and N-1 stimuli decoding revealed neither significant learning at either time scale nor a significant difference in learning coefficients between predictable and random blocks (all $p > 0.1$). Taken together, these results provide evidence that the early N-3 decoding in predictable blocks improves at faster time scales (within blocks) relative to random blocks, but the evidence for any decoding improvement at longer timescales (across blocks) is weak.

Multivariate analysis: Predictive and mnemonic representations rely on uncorrelated data features

While the decoding boost observed for the N-3 vowel in predictable blocks, and its gradual improvement over time, bears the hallmarks of a predictive representation, we have also accessed mnemonic representations by decoding previous vowels (N-1 and N-2) in random blocks. To test whether the decoding of predictive and mnemonic representations rely on the same data features, we performed three further analyses. First, we repeated decoding using a searchlight, where each decoding estimate was based on a subset of channels. While no significant N-3 decoding was found in random blocks based on all channels and correcting for multiple comparisons across time points, a searchlight could, in principle, uncover channels more sensitive to N-3 vowel identity. We then correlated the spatial maps of decoding estimates between the predictable and random blocks. We reasoned that if predictive and mnemonic representations rely on similar data features, the N-3 maps should be correlated across blocks. This analysis revealed significant correlations between spatial decoding maps in predictable and random blocks only for the N-1 vowel (Figure 4A; 33–70 ms; $t_{max} = 5.72$; cluster-level $p_{FWE} < 0.001$), but not for the earlier vowels (N-2 or N-3: all $p_{FWE} > 0.05$). Specifically, while for the N-1 vowel the spatial maps of decoding obtained in predictable and random blocks were similar (t test Bayes Factor: 865.33, indicating extremely strong evidence for a correlation) and showed the strongest contribution of the anterior/inferior channels, for the N-3 vowel,

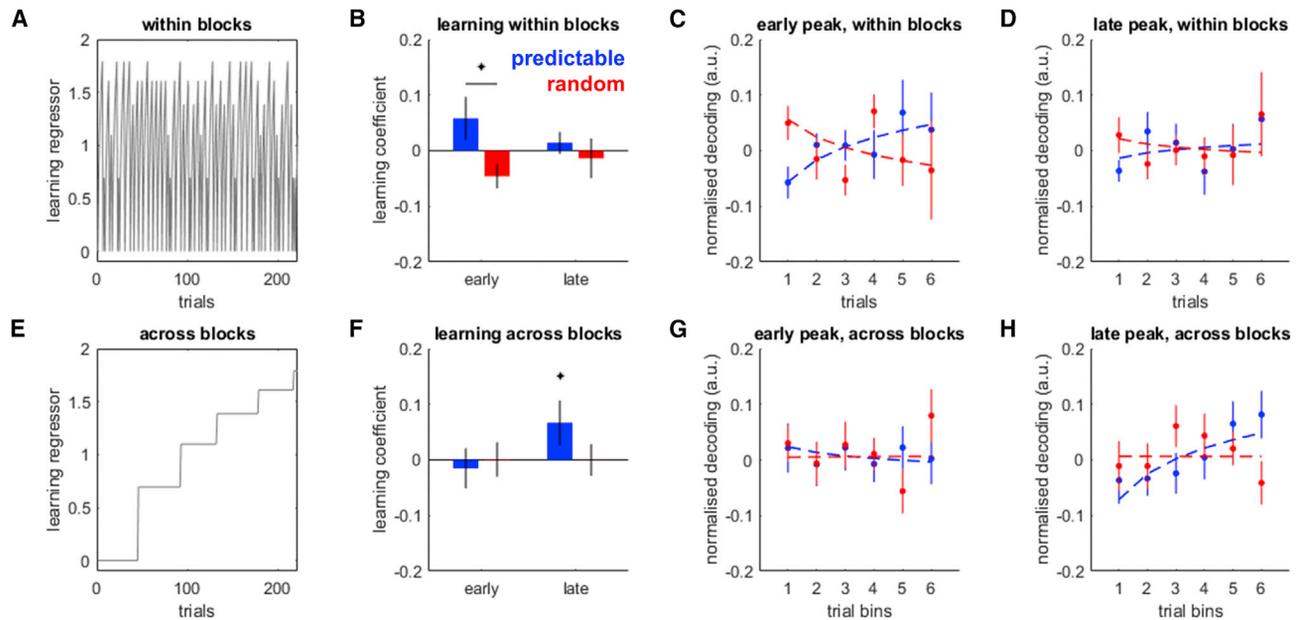


Figure 3. Learning effects

(A) A trial-by-trial regressor of learning within blocks (faster time scale) was quantified as the (log) burst number in a block of identical triplets. (B) Regression coefficients (“within-blocks” learning) for two time windows with significant N-3 decoding boost (see Figure 2A [right]). Error bars denote SEM across recording sessions. Asterisk denotes a significant Wilcoxon sign rank test. (C) Normalized decoding per trial within a block of identical triplets: early time window. Error bars denote SEM across recording sessions. (D) Normalized decoding per trial within a block of identical triplets: late time window. (E) A trial-by-trial regressor of learning across blocks (slower timescale) was quantified as the block number in a recording session, binned into six bins. (F–H) Learning effects across blocks, figure legend as in (B)–(D).

they were more orthogonal (t test Bayes Factor: 0.2569, indicating moderate evidence against correlation; cf. N-2: Bayes Factor 0.3558), showing an inferior-superior gradient in predictable blocks and an anterior-posterior gradient in random blocks (Figures 4B and 4C). This contrasted with correlations between decoding maps obtained for odd vs. even trials, which were significant for each vowel position (N-1: $r_{\max} = 0.27$; N-2: $r_{\max} = 0.13$; N-3: $r_{\max} = 0.11$; all significant at $p_{\text{FWE}} < 0.05$ correcting across time points). While the latter correlation coefficients were moderate to low, likely due to a decreased signal-to-noise ratio as a result of splitting the dataset in half, this finding suggests that N-3 decoding maps are relatively stable across trials (odd vs. even) but uncorrelated across conditions (predictable vs. random).

Second, we repeated the decoding of vowels in each position, this time training on trials drawn from one type of blocks (e.g., random) and testing on trials from the other type of blocks (e.g., predictable). This analysis (Figure S4A) revealed that only N-1 decoding generalized across block types (train on random, test on predictable: $T_{\max} = 12.92$, $p_{\text{FWE}} < 0.001$; train on predictable, test on random: $T_{\max} = 13.39$, $p_{\text{FWE}} < 0.001$), with no differences observed between blocks (all paired t test $p_{\text{FWE}} > 0.05$). Conversely, for N-2 and N-3 decoding, no significant cross-block decoding was observed in either direction (all $p_{\text{FWE}} > 0.05$).

Third, we performed a cross-temporal generalization analysis (Figures 4D and 4E), training on one vowel position (e.g., N-1) and testing on another (e.g., N-3). This analysis revealed that while decoding generalizes across time points within each vowel position (e.g., training on neural activity 100 ms and testing on

150 ms after vowel onset; cf. Cappotto et al., 2021¹), it does not generalize across vowel positions (e.g., training on N-1 and testing on N-3) except for a temporally limited interference effect between N-1 and N-2 vowels (Table S4).

These results suggest that the decoding boost observed for N-3 vowels in predictable blocks (reflecting a predictive representation) relies on data features that are specific to these blocks and are not generalizable to the random blocks or to other vowels.

DISCUSSION

In the present study, we demonstrated that stimulus history (sensory memory traces of token values up to N-3) can be decoded from neural responses to broadband noise bursts in both repeated triplet and randomized blocks, expanding on previous research.^{1,3,4} Crucially, we also provide evidence for the decoding of predictive mechanisms by linking increased N-3 decodability to predictable blocks, further established through the presence of learning effects as the number of triplet pattern repeats increases. This demonstrates that neural responses to noise bursts tap into predictive mechanisms, establishing a novel method for decoding both phenomena simultaneously and independent of attentional tasks. Our results suggest that mnemonic and predictive decoding rely on largely uncorrelated data features—specifically, decoding N-3 stimuli in predictable blocks cannot be generalized to decoding other stimuli in the same blocks or to the data features present in random blocks.

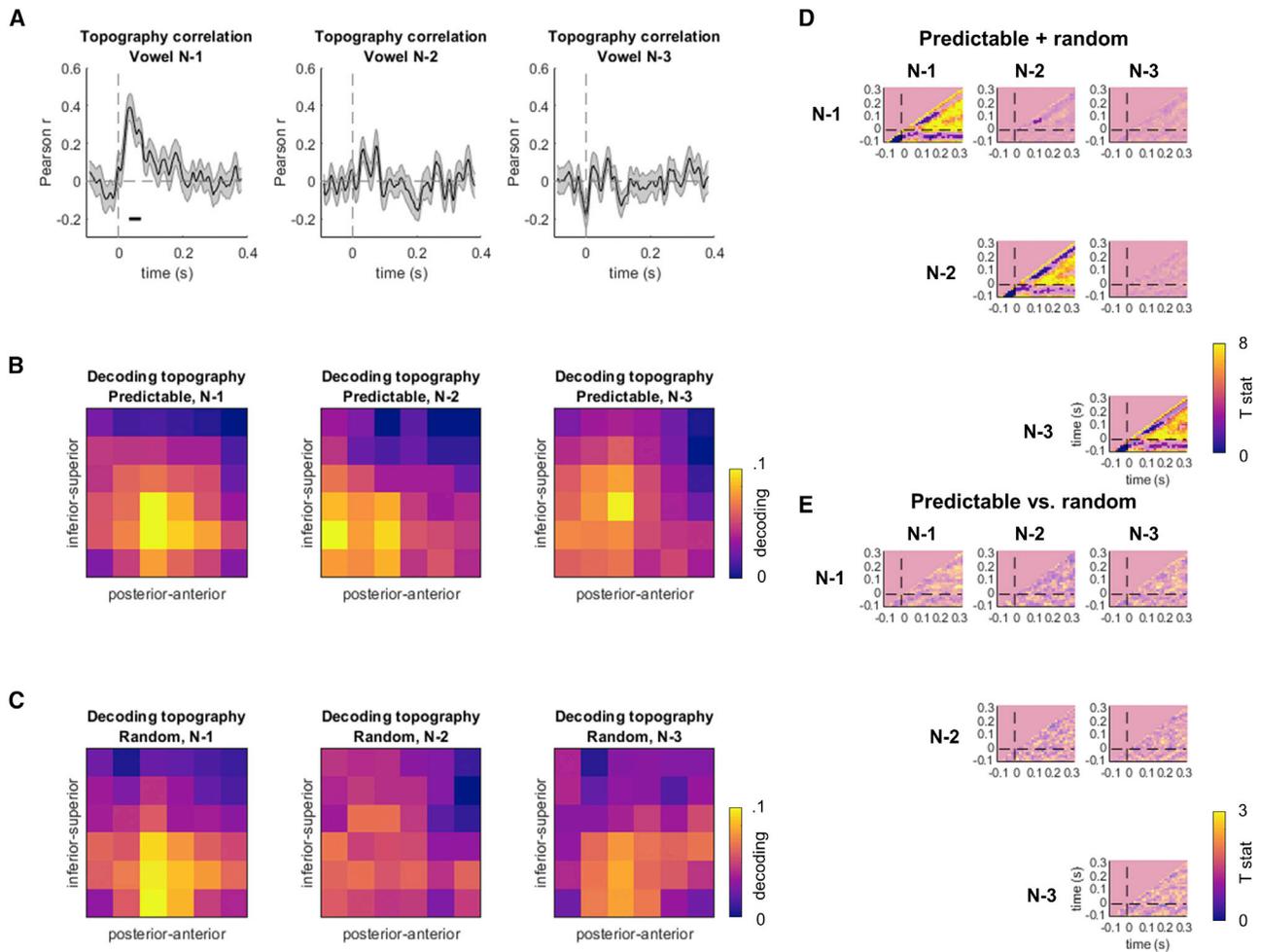


Figure 4. Spatial topography of predictive and mnemonic representations

(A) Time courses of correlation coefficients between decoding topographies in predictable vs. random blocks. Left/middle/right panel: decoding N-1/N-2/N-3 vowel (shaded area, SEM across recording sessions; black horizontal line, correlation coefficients significantly different from zero, $p_{FWE} < 0.05$).

(B) Decoding topographies based on the 0–100 ms decoding time window, predictable blocks. Left/middle/right panel: decoding N-1/N-2/N-3 vowel.

(C) Decoding topographies based on the 0–100 ms decoding time window, random blocks. Figure legend as in (B).

(D) Cross-temporal generalization averaged across conditions (predictable + random). Rows: test data; columns: remaining data used for estimating decoding matrices. Each panel shows a cross-temporal decoding matrix with each time point representing decoding based on the Mahalanobis distance between a particular vowel position (N-1, N-2, or N-3) and latency of neural activity and another vowel position and latency of neural activity. Unmasked areas represent significant cross-temporal decoding generalization at $p_{FWE} < 0.05$, cluster-level corrected. Only one (symmetric) side of the diagonal is plotted.

(E) Cross-temporal generalization: differences between conditions (predictable vs. random). Figure legend as in (D).

See also [Figure S4A](#) and [Table S4](#).

Previous work has established the use of broadband noise impulses in decoding sensory memory tokens^{3,4} mediated by mechanisms that function under anesthesia in animal models.¹ Here, we expand on these findings by decoding further stimulus history, showing that it is possible to decode memory representations of both sequences and individual tokens up to N-3. We also expand on another recent study¹³ showing that sequence contents can be preferentially decoded from auditory-cortical activity in rat models but that this decoding benefit is only observed for rats with prior training. Similarly, previous work in the visual system of awake mice found that prior training elicits predictive representations that can be decoded.¹⁴ Unlike these studies, which used several interleaved sequences in a continuous stream during prior

exposure blocks, we used a protocol in which a sequence (triplet) was repeated and then replaced, without prior training. This suggests that for such repetitive sequences, decoding can be achieved in naive and anesthetized rats. In contrast to the previous study,¹⁴ our results did not reveal any significant decoding on the omission responses. One possible explanation is that, in anesthetized brains that had not undergone prior training, predictive representations require a stronger activation (e.g., broadband noise bursts) to become observable than would be the case for awake brains of trained participants. In both the present and previous studies,¹ we have demonstrated that univariate analysis was not sufficient to decode memory tokens and multivariate methods provided significant decoding.

The present literature on animal models of predictive processing is largely within the context of stimulus-specific adaptation (SSA), making it difficult to separate predictive from adaptive mechanisms. Our findings in the AC are not likely to be explained by a simple SSA explanation, given that we observed the decodability of randomly substituted tokens within repeated sequences as well as within non-repeating triplets. If adaptation were responsible for decodability, this effect would be unlikely to increase with overall triplet repetition, as pattern sensitivity and resulting deviance detection has been shown to rely on hierarchical and contextual error detection.¹⁵ Our results, suggesting that decoding N-3 tone identity and triplet identity may occur at different latencies (Figure S3), are also consistent with the latter hypothesis, as they suggest that predictive processing of single elements might be more short-lived than the encoding of entire sequences.

Importantly, by contrasting responses to noise bursts in predictable vs. random sequences, we tapped into both predictive and mnemonic representations. This goes beyond recent findings in humans showing that predictive neural activity can be explained by memory of past stimuli but that could not access mnemonic representations independently of predictive processing.¹⁶ Interestingly, a recent study on auditory associative learning in awake mice showed that neural activity evoked by a predicted stimulus contains information both about its most likely predictor and its actual past but that this information relies on orthogonal neural codes, suggesting that mnemonic and predictive representations coexist within sensory cortices.¹¹ Although our paradigm did not test for this explicitly, our observation of uncorrelated data features enabling decoding in predictable vs. random blocks and a lack of decoding generalization across blocks and across vowels (N-1 vs. N-3) suggest that such mechanisms are not dependent on active processes, and they can also be observed indirectly over broad neural populations. It is important to note that the spatial resolution of ECoG makes it difficult to identify discrete neural populations due to changes in spatio-temporal representation, and finer recording techniques with single-cell resolution would be required to accurately discern if mnemonic and predictive representations decoded in our paradigm rely on unique neural populations or are multiplexed within the same population.

Importantly, we also establish that the decodability of predictable N-3 tokens gradually increases with repeated triplet presentations (relative to random blocks), implicating passive learning effects as a measure of predictive mechanisms. Recent studies have successfully paired concepts of statistical learning and predictive coding by investigating neural correlates of melodic expectation to naturalistic music, observing that neural responses to less statistically likely notes elicit markers consistent with their level of statistical predictability.¹⁷ Human fMRI studies in the visual domain have further established the role of temporal regularity in sequence learning and their resultant effects on the decodability of predictable stimuli.¹⁸ However, studies employing animal models and different attention states to investigate predictive mechanisms have been lacking.¹⁹ Although further investigations would be required to clearly verify the role of learning effects at multiple timescales, our results provide an indication of prediction formation at a relatively fast timescale (prediction updating following a presentation of a new triplet).

While it is intrinsically interesting that anesthesia did not abolish the emergence of predictive representations in our study, one must acknowledge that this raises questions about the extent to which our results are representative of neural functions in a normal, awake state. Different types of anesthetic agents (e.g., ketamine, equithesin, or pentobarbital) have been shown to affect various features of neuronal activity—such as spontaneous rate, response threshold, or oscillations—in the AC to a greater or lesser extent.^{20–22} However, experiments under anesthesia are still considered an efficient and useful tool in identifying neural mechanisms when carefully controlled. We selected urethane as our main agent for controlling the anesthesia, as it has been widely used for memory-related studies as an agent with minimal effect on spectral tuning, neural discriminability, and information processing.^{23–26} Importantly, a comparable hierarchical gradient across subcortical and cortical regions observed for prediction error signaling between urethane-anesthetized and awake rodents²⁷ supports the notion of preserved predictive processing even under anesthesia.

In summary, the present study observed concurrent mnemonic and predictable representations under anesthesia, indicating mechanisms at work in passive preparations and thus providing a new model for investigating simultaneous memory and predictive mechanisms independent of attentional state.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Subjects
 - Anesthesia and surgical procedures
 - Experimental apparatus
- METHODS DETAILS
 - Stimulus design
 - Experimental paradigm
 - Neural data acquisition and pre-processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Univariate analysis: Summarizing vowel-evoked, omission-evoked, and frozen noise burst-evoked activity
 - Univariate analysis: Oscillatory activity
 - Multivariate analysis: Decoding sensory, mnemonic, and predicted vowel information
 - Multivariate analysis: Learning effect on decoding
 - Multivariate analysis: Similarity between predictive and mnemonic representations
 - Multivariate analysis: Cross-temporal generalization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.04.022>.

ACKNOWLEDGMENTS

This work has been supported by the European Commission's Marie Skłodowska-Curie Global Fellowship (750459 to R.A.), the Hong Kong General Research Fund (11100518 to R.A. and J.S.) and a grant from the European Commission/Hong Kong Research Grants Council Joint Research Scheme (9051402 to R.A., D.P., and J.S.).

AUTHOR CONTRIBUTIONS

Conceptualization and methodology, D.C., R.A., and J.S.; investigation, D.C., H.K., and K.L.; writing – original draft, D.C.; writing – review & editing, D.C., R.A., L.M., and J.S.; funding acquisition, R.A. and J.S.; supervision, R.A., L.M., and J.S.

DECLARATION OF INTERESTS

The authors have no conflicts of interest to declare.

Received: December 2, 2021

Revised: March 3, 2022

Accepted: April 8, 2022

Published: April 28, 2022

REFERENCES

- Cappotto, D., Auksztulewicz, R., Kang, H., Poeppel, D., Melloni, L., and Schnupp, J. (2021). Decoding the Content of Auditory Sensory Memory Across Species. *Cereb. Cortex* 31, 3226–3236. <https://doi.org/10.1093/cercor/bhab002>.
- Stokes, M.G. (2015). Activity-silent" working memory in prefrontal cortex: A dynamic coding framework (Elsevier Ltd).
- Wolff, M.J., Ding, J., Myers, N.E., and Stokes, M.G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* 9, <https://doi.org/10.3389/fnsys.2015.00123>.
- Wolff, M.J., Kandemir, G., Stokes, M.G., and Akyürek, E.G. (2019). Unimodal and bimodal access to sensory working memories by auditory and visual impulses. *J. Neurosci.* 40, 671–681. <https://doi.org/10.1523/jneurosci.1194-19.2019>.
- Rust, N.C., and Palmer, S.E. (2021). Remembering the Past to See the Future. *Annu. Rev. Vis. Sci.* 7, 349–365. <https://doi.org/10.1146/annurev-vision-093019-112249>.
- Barron, H.C., Auksztulewicz, R., and Friston, K. (2020). Prediction and memory: a predictive coding account. *Prog. Neurobiol.* 192, 101821. <https://doi.org/10.1016/j.pneurobio.2020.101821>.
- Fairhall, A.L., Lewen, G.D., Bialek, W., and de Ruyter Van Steveninck, R.R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature* 412, 787–792. <https://doi.org/10.1038/35090500>.
- Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87. <https://doi.org/10.1016/j.jphysparis.2006.10.001>.
- Rubin, J., Ulanovsky, N., Nelken, I., and Tishby, N. (2016). The Representation of Prediction Error in Auditory Cortex. *PLoS Comput. Biol.* 12, e1005058. <https://doi.org/10.1371/journal.pcbi.1005058>.
- Schröger, E., Bendixen, A., Denham, S.L., Mill, R.W., Böhm, T.M., and Winkler, I. (2014). Predictive Regularity Representations in Violation Detection and Auditory Stream Segregation: From Conceptual to Computational Models. *Brain Topogr.* 27, 565–577. <https://doi.org/10.1007/s10548-013-0334-6>.
- Libby, A., and Buschman, T.J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* 24, 715–726. <https://doi.org/10.1038/s41593-021-00821-9>.
- Chien, V.S.C., Maess, B., and Knösche, T.R. (2019). A generic deviance detection principle for cortical On/Off responses, omission response, and mismatch negativity. *Biol. Cybern.* 113, 475–494. <https://doi.org/10.1007/s00422-019-00804-x>.
- Luo, D., Li, K., An, H., Schnupp, J.W., and Auksztulewicz, R. (2021). Learning boosts the decoding of sound sequences in rat auditory cortex. *Curr. Res. Neurobiol.* 2, 100019. <https://doi.org/10.1016/j.crneur.2021.100019>.
- Gavornik, J.P., and Bear, M.F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* 17, 732–737. <https://doi.org/10.1038/nn.3683>.
- Casado-Román, L., Carbajal, G.V., Pérez-González, D., and Malmierca, M.S. (2020). Prediction error signaling explains neuronal mismatch responses in the medial prefrontal cortex. *PLoS Biol.* 18, e3001019. <https://doi.org/10.1371/journal.pbio.3001019>.
- Baumgarten, T.J., Maniscalco, B., Lee, J.L., Flounders, M.W., Abry, P., and He, B.J. (2021). Neural integration underlying naturalistic prediction flexibly adapts to varying sensory input rate. *Nat. Commun.* 12, 2643. <https://doi.org/10.1038/s41467-021-22632-z>.
- Di Liberto, G.M., Pelofi, C., Bianco, R., Patel, P., Mehta, A.D., Herrero, J.L., de Cheveigné, A., Shamma, S., and Mesgarani, N. (2020). Cortical encoding of melodic expectations in human temporal cortex. *eLife* 9, e51784. <https://doi.org/10.7554/elife.51784>.
- Luft, C.D.B., Meeson, A., Welchman, A.E., and Kourtzi, Z. (2015). Decoding the future from past experience: learning shapes predictions in early visual cortex. *J. Neurophysiol.* 113, 3159–3171. <https://doi.org/10.1152/jn.00753.2014>.
- Heilbron, M., and Chait, M. (2018). Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience* 389, 54–73. <https://doi.org/10.1016/j.neuroscience.2017.07.061>.
- Cheung, S.W., Nagarajan, S.S., Bedenbaugh, P.H., Schreiner, C.E., Wang, X., and Wong, A. (2001). Auditory cortical neuron response differences under isoflurane versus pentobarbital anesthesia. *Hear. Res.* 156, 115–127. [https://doi.org/10.1016/s0378-5955\(01\)00272-6](https://doi.org/10.1016/s0378-5955(01)00272-6).
- Gaese, B.H., and Ostwald, J. (2001). Anesthesia Changes Frequency Tuning of Neurons in the Rat Primary Auditory Cortex. *J. Neurophysiol.* 86, 1062–1066. <https://doi.org/10.1152/jn.2001.86.2.1062>.
- Zurita, P., Villa, A.E.P., de Ribaupierre, Y., de Ribaupierre, F., and Rouiller, E.M. (1994). Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. *Neurosci. Res.* 19, 303–316. [https://doi.org/10.1016/0168-0102\(94\)90043-4](https://doi.org/10.1016/0168-0102(94)90043-4).
- Astikainen, P., Stefanics, G., Nokia, M., Lipponen, A., Cong, F., Penttonen, M., and Ruusuvirta, T. (2011). Memory-Based Mismatch Response to Frequency Changes in Rats. *PLoS One* 6, e24208. <https://doi.org/10.1371/journal.pone.0024208>.
- Capsius, B., and Leppelsack, H.-J. (1996). Influence of urethane anesthesia on neural processing in the auditory cortex analogue of a songbird. *Hear. Res.* 96, 59–70. [https://doi.org/10.1016/0378-5955\(96\)00038-x](https://doi.org/10.1016/0378-5955(96)00038-x).
- Ruusuvirta, T., Penttonen, M., and Korhonen, T. (1998). Auditory cortical event-related potentials to pitch deviances in rats. *Neurosci. Lett.* 248, 45–48. [https://doi.org/10.1016/s0304-3940\(98\)00330-9](https://doi.org/10.1016/s0304-3940(98)00330-9).
- Schumacher, J.W., Schneider, D.M., and Woolley, S.M.N. (2011). Anesthetic state modulates excitability but not spectral tuning or neural discrimination in single auditory midbrain neurons. *J. Neurophysiol.* 106, 500–514. <https://doi.org/10.1152/jn.01072.2010>.
- Parras, G.G., Nieto-Diego, J., Carbajal, G.V., Valdés-Baizabal, C., Escera, C., and Malmierca, M.S. (2017). Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nat. Commun.* 8, 2148. <https://doi.org/10.1038/s41467-017-02038-6>.
- Malmierca, M.S., Niño-Aguillón, B.E., Nieto-Diego, J., Porteros, Á., Pérez-González, D., and Escera, C. (2019). Pattern-sensitive neurons reveal encoding of complex auditory regularities in the rat inferior colliculus. *NeuroImage* 184, 889–900. <https://doi.org/10.1016/j.neuroimage.2018.10.012>.
- Woods, V., Trumpis, M., Bent, B., Palopoli-Trojani, K., Chiang, C.-H., Wang, C., Yu, C., Insanally, M.N., Froemke, R.C., and Venti, J. (2018). Long-term recording reliability of liquid crystal polymer μ CoG arrays. *J. Neural Eng.* 15, 066024. <https://doi.org/10.1088/1741-2552/aae39d>.

30. Ball, T., Kern, M., Mutschler, I., Aertsen, A., and Schulze-Bonhage, A. (2009). Signal quality of simultaneously recorded invasive and non-invasive EEG. *NeuroImage* 46, 708–716. <https://doi.org/10.1016/j.neuroimage.2009.02.028>.
31. Salisbury, D.F. (2012). Finding the missing stimulus mismatch negativity (MMN): emitted MMN to violations of an auditory gestalt. *Psychophysiology* 49, 544–548. <https://doi.org/10.1111/j.1469-8986.2011.01336.x>.
32. Kilner, J.M., Kiebel, S.J., and Friston, K.J. (2005). Applications of random field theory to electrophysiology. *Neurosci. Lett.* 374, 174–178. <https://doi.org/10.1016/j.neulet.2004.10.052>.
33. Henin, S., Turk-Browne, N.B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., and Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Sci. Adv.* 7, eabc4530. <https://doi.org/10.1126/sciadv.abc4530>.
34. Ding, N., and Simon, J.Z. (2013). Power and phase properties of oscillatory neural responses in the presence of background activity. *J. Comput. Neurosci.* 34, 337–343. <https://doi.org/10.1007/s10827-012-0424-6>.
35. Myers, N.E., Rohenkohl, G., Wyart, V., Woolrich, M.W., Nobre, A.C., and Stokes, M.G. (2015). Testing sensory evidence against mnemonic templates. *eLife* 4, e09000. <https://doi.org/10.7554/elife.09000>.
36. van Ede, F., Chekroud, S.R., Stokes, M.G., and Nobre, A.C. (2018). Decoding the influence of anticipatory states on visual perception in the presence of temporal distractors. *Nat. Commun.* 9, 1449. 12. <https://doi.org/10.1038/s41467-018-03960-z>.
37. Grootswagers, T., Wardle, S.G., and Carlson, T.A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *J. Cogn. Neurosci.* 29, 677–697. https://doi.org/10.1162/jocn_a_01068.
38. Nemrodov, D., Niemeier, M., Patel, A., and Nestor, A. (2018). The Neural Dynamics of Facial Identity Processing: Insights from EEG-Based Pattern Analysis and Image Reconstruction. *eNeuro* 5, ENEURO.0358, 17.2018. <https://doi.org/10.1523/eneuro.0358-17.2018>.
39. De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D.L. (2000). The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* 50, 1–18. [https://doi.org/10.1016/s0169-7439\(99\)00047-7](https://doi.org/10.1016/s0169-7439(99)00047-7).
40. Auksztulewicz, R., Myers, N.E., Schnupp, J.W., and Nobre, A.C. (2019). Rhythmic Temporal Expectation Boosts Neural Activity by Increasing Neural Gain. *J. Neurosci.* 39, 9806–9817. <https://doi.org/10.1523/jneurosci.0925-19.2019>.
41. Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411. [https://doi.org/10.1016/s0047-259x\(03\)00096-4](https://doi.org/10.1016/s0047-259x(03)00096-4).
42. Bobadilla-Suarez, S., Ahlheim, C., Mehrotra, A., Panos, A., and Love, B. (2019). Measures of Neural Similarity. *Comput. Brain Behav.*
43. Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage* 137, 188–200. <https://doi.org/10.1016/j.neuroimage.2015.12.012>.
44. HiJee, K., Ryszard, A., Hong, C.C., Drew, C., Gurusamy, R.V., and Hendrik, S.J.W. (2021). Memory Transfer of Random Time Patterns Across Modalities. Preprint at bioRxiv. <https://doi.org/10.1101/2020.11.24.395368>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Adult female Wistar rats	Chinese University of Hong Kong	RGD_13525002
Software and algorithms		
MATLAB	Mathworks	SCR_001622
Python	Python	SCR_008394
SPM12	University College London	SCR_007037
Deposited data		
Code and Processed Data	Zenodo	https://doi.org/10.5281/zenodo.6407267

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Drew Cappotto (drew.cappotto@my.cityu.edu.hk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Pre-processed data and code for generating results figures are publicly available via the following online repository: <https://doi.org/10.5281/zenodo.6407267>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

Eight young adult female Wistar rats, acquired from the Chinese University of Hong Kong, were used in the experiment. The rats were “naive”, i.e. had no experience or training with the stimulus sets prior to recording, were aged between 8 and 13 weeks (median age = 10.5 weeks), and weighed between 216 and 289 g (median weight = 238 g). Normal hearing was ascertained by measuring auditory brainstem response at thresholds < 20 dB sound pressure level (SPL) to broadband click trains.

Anesthesia and surgical procedures

Anesthesia was induced with an intraperitoneal (i.p.) injection of ketamine (80 mg/kg) and xylazine (12 mg/kg), and maintained throughout the experiment via 20% urethane injections. A first dose of 0.25 ml/kg of the urethane solution was administered one hour after the induction with ketamine and xylazine, and further 0.25 ml/kg doses were delivered as required, based on periodic assessments of anesthesia depth via the toe pinch withdrawal reflex. Dexamethasone (0.2 mg/kg, i.p.) was delivered before surgery as an anti-inflammatory. This protocol, based on previous rodent studies,^{1,28} allowed for fast induction of anesthesia via the initial administration of ketamine and xylazine, while avoiding later NMDA-specific inhibitory effects of ketamine through the use of urethane to maintain anesthesia for ECoG recordings. The anesthetized animal was placed in a stereotaxic frame, and the animal's head was fixed with hollow ear bars to allow sound delivery. An isothermal heating pad and a rectal thermometer were used to maintain body temperature at $36 \pm 1^\circ\text{C}$ throughout the experiment. The skin and muscle tissue over the temporal lobe of the skull were removed, and a craniotomy was performed to expose a 5×4 mm region over the right AC, leaving the dura intact. The anterior edge of the craniotomy was 2.5 mm posterior from Bregma, and the dorsal edge was 2 mm ventral from Bregma (Figure S1A, adapted from¹).

Experimental apparatus

The ECoG array was placed on the exposed cortex and a cotton roll was placed between the remaining skin and the array to hold the array securely in place and ensure a stable, low impedance contact between the recording sites and the dura. A hole was drilled through the skull anterior to the Bregma on the animal's left to place a small stainless steel screw which served as ground and reference electrode for the electrode array and headstage amplifier. Correct placement of the ECoG array was verified by recording a set

of Frequency Response Areas (FRAs; [Figure S1B](#)) from each site by collecting responses to 100 ms pure tones varying in sound level (30 - 80 dB SPL) and frequency (500 - 32,000 Hz, ¼ octave steps). Each tone was presented 10 times, in a randomly interleaved fashion, with an onset-to-onset ISI of 500 ms.

METHODS DETAILS

Stimulus design

The artificial vowels were generated using custom Python scripts. Consecutive vowels were separated by 350 ms of silence (500 ms onset to onset ISI). We deemed artificial vowels preferable to tones as they activate larger parts of the tonotopic array and they resemble many types of natural sounds, including many vertebrate vocalizations or insect sounds, making them arguably more ecologically valid than pure tones. These generated pulse trains which were subsequently passed through a cascade of two 2nd-order Butterworth bandpass filters with a bandwidth equal to 20% of the center (formant) frequency (`scipy.signal.functions.butter()` and `lfilter()`). The formant frequencies for these artificial vowels were chosen to lie between 900 and 9000 Hz to bring them well into the auditory range of rats, and the fundamental frequencies (F0s) of the vowels were relatively low, between 260 and 420 Hz, to generate a large number of closely stacked harmonics under each formant. Stimulus sequences consisted of combinations of three possible artificial vowels, one we refer to as “A” with formants and 3000 and 5400 Hz and an F0 of 420 Hz, an “O” with formants 900 and 2700 Hz and F0 260 Hz, and an “I” with formants 1050 and 9000 Hz and F0 300 Hz. On occasion, as described further below, one of the vowels in the sequence could be replaced by either a 150 ms frozen pink noise burst computed according to the algorithm described in <https://github.com/python-acoustics/python-acoustics/blob/master/acoustics/generator.py>, or by a silent pause. The artificial vowel and pink noise tokens were loaded onto a Tucker Davis Technologies (TDT) RZ6 digital sound processor which was programmed using custom written software to present the tokens in a predefined order at a sample rate of 48,828 Hz through headphone drivers connected to the hollow ear bars via 3D printed adapters.

Experimental paradigm

Two types of blocks were employed. In “predictable” blocks, vowels were grouped into triplets, which repeated at least 25 times (range 25-100, mean 30) before being replaced with another triplet (e.g., AOOAOOAOO...AAIAAIAAI...). In “random” blocks, vowels were presented in a random order, while keeping the base frequency of each vowel constant and comparable to the predictable block (e.g., AOIOAIOAOOAIIA...). Each session contained ~72 such blocks (amounting to a total of 2100 triplets per session), presented in a different order per session. Triplets were selected to prevent redundant combinations from occurring during presentation (e.g., AOO, OAO, and OOA would result in identical sequences with different starting points, and thus only AOO was used). The triplets were then concatenated to form the long stimulus sequences presented in the experimental sessions. In these sequences, 5% of stimulus events were replaced with omissions, and 5% were similarly replaced with a burst of pink noise. The vowels that were replaced with noise bursts or silent pauses were chosen pseudo-randomly, subject to the constraint that a minimum of three repetitions of a given triplet had to have occurred before a vowel could be replaced. In a control condition (“random” sessions), vowels were presented randomly, rather than in predefined triplets. The positions of omissions and noise bursts within the stimulus sequences were kept the same across the predictable and random blocks.

Neural data acquisition and pre-processing

An 8 x 8 Viventi ECoG electrode array with 400 µm electrode spacing²⁹ was used to acquire ECoG recordings, employing three ground channels located in the corners of the array, and a common reference. A (TDT) PZ5 neurodigitizer was used to record signals from the array via a RZ2 processor. FRA responses were recorded with BrainWare software at a sampling rate of 24,414 Hz, and responses to the vowel sequences were recorded using custom Python code at a sampling rate of 6104 Hz. The recorded electrode signals were first low-pass filtered at a cutoff frequency of 90 Hz using a 5th order Butterworth filter, and downsampled to 300 Hz to extract neural activity evoked by acoustic stimuli. The pre-processed signals were re-referenced to the average of all channels,³⁰ and segmented by extracting 500 ms long voltage traces from -100 ms to +400 ms relative to the onset of each token. Epoched traces were baseline-corrected by subtraction of the mean pre-stimulus voltage values, and linearly detrended.³¹

QUANTIFICATION AND STATISTICAL ANALYSIS

Univariate analysis: Summarizing vowel-evoked, omission-evoked, and frozen noise burst-evoked activity

Univariate analysis was performed to assess whether vowel types (A, I, O) modulated vowel-evoked, burst-evoked, and omission-evoked activity on a channel-by-channel basis ([Figure S2](#)). Additionally, in the analysis of burst-evoked and omission-evoked activity, we tested whether it is modulated by the preceding sounds at different “positions” relative to the burst/omission (N-1 position: the immediately preceding vowel, N-2 position: two stimuli before the burst/omission, N-3 position: three stimuli before the burst/omission). Epoched data were separated per vowel, position, and condition, and then averaged across trials. First, to visualize the evoked responses, trial-averaged ECoG responses were concatenated across sound types/positions/conditions/animals, resulting in 2 two-dimensional matrices per condition with single channels along one dimension and concatenated time points along the second dimension. A principal component analysis using singular value decomposition was performed on the resulting matrices. The output provided spatial principal components describing channel topographies, and temporal principal components describing voltage

time-series concatenated across vowels/positions and animals, sorted by the ratio of explained variance. A weighted average was calculated to summarize the top principal components explaining 95% of the original variance, weighted by the proportion of variance explained. These resulting voltage time-series were averaged per vowel across animals. Frozen noise burst-evoked and omission-evoked single-trial data were similarly averaged across trials, separately for each preceding vowel and position, and subject to the same principal component analysis described above.

The above principal component analysis was used only for the purposes of visualizing the data. In order to test if any time points and channels showed significant amplitude modulations by vowel (in case of vowel-evoked responses) or preceding vowel in each position (in case of burst-evoked and omission-evoked responses), single-subject trial-average ECoG data in the original electrode grid were converted into three-dimensional matrices containing two spatial dimensions and one temporal dimension. These matrices were then converted to 3D images and entered into a repeated-measures ANOVA with one within-subjects factor (vowel; three levels) and one repeated-measures factor (rat), implemented in SPM12 (University College London) as a general linear model (GLM). This was done separately for each stimulus type (vowel-evoked responses, burst-evoked responses, and omission-evoked responses). The effects of preceding vowels on burst-evoked and omission-evoked responses were analyzed in separate ANOVAs per position. To test for the effect of vowel on evoked activity amplitude, an omnibus F test across 3 vowels was used. The resulting statistical parametric maps were thresholded at $p < 0.005$ (two-tailed) and corrected for multiple comparisons across spatiotemporal voxels at a family-wise error (FWE)-corrected $p_{FWE} = 0.05$ (cluster-level).³²

Univariate analysis: Oscillatory activity

To test whether sequence processing is associated with spectral peaks in the neural response spectrum at the syllable and triplet rate³³, we analyzed phase coherence of neural activity (Figure S4B). Specifically, for each rat and recording session, we split the continuous single-channel ECoG data into 175 chunks of 12 triplets, and, for each chunk, calculated the Fourier spectrum of neural activity measured during that chunk. Inter-trial phase coherence (ITPC) was calculated according to the following equation³⁴:

$$ITPC_f = \left([\Sigma^N \cos \varphi_f]^2 + [\Sigma^N \sin \varphi_f]^2 \right) / N,$$

where φ_f denotes the Fourier phase at a given frequency f and $N = 175$ chunks. In the initial univariate analysis, phase coherence estimates were averaged across channels. To test for the presence of statistically significant phase coherence peaks, coherence values at the token rate (2 Hz) and triplet rate (0.667 Hz) were compared against the mean of coherence values at their respective neighboring frequencies (single token rate: 1.944 and 2.056 Hz; triplet rate: 0.611 and 0.722 Hz) using Wilcoxon's signed rank tests.

Multivariate analysis: Decoding sensory, mnemonic, and predicted vowel information

Data were subjected to multivariate analyses to test if information about vowel type could be decoded from the pattern of burst-evoked and omission-evoked activity observed across multiple channels and time points. To this end, we adapted methods established in previous multivariate decoding research, which has demonstrated decodability in similar data and experimental contexts.^{1,3,4,35,36}

Prior to decoding, single-trial omission or frozen noise burst-evoked responses were sorted by the preceding vowel, separately for each vowel position. While the randomized order of vowel presentation in relation to noise bursts (see [experimental paradigm](#) and [stimulus design](#)) effectively equalized the ratio of vowels presented at each position, we imposed an additional constraint on trial selection to ensure that decoding N-3 vowels is not confounded by the vowels presented immediately before the noise burst (N-1). Specifically, in decoding N-3 stimuli relative to noise burst X, we excluded trials for which N-3 and N-1 were identical (e.g., AAOAAX was included, since vowel N-1 corresponds to A and N-3 to O; however, AAOAXO was excluded, since both vowels N-1 and N-3 correspond to A). To equalize the number of trials across decoding conditions, the same constraint was imposed on N-2 stimuli (excluding trials for which N-2 and N-1 were identical) and on random blocks.

Decoding time-courses were estimated using a sliding window approach,^{1,4} pooling information over multiple time-points and channels to boost decoding accuracy.^{37,38} Specifically, for each channel, trial, and time point, we first pooled voltage values within a 50 ms window relative to a given time point. Then, a vector of 5 average voltage values was calculated per channel and trial by downsampling the voltage values over 10 ms bins. In other words, a single vector of multivariate data corresponding to the test trial (multiple channels \times 5 time points within a 50 ms window, concatenated into a long vector) is compared against three vectors (one per vowel), each of exactly the same length as for the test trial but based on the remaining trials. The data were then de-meant to remove the channel-specific average voltage over the entire 50 ms time window from each channel and time bin, ensuring that the multivariate analysis approach was optimized for decoding transient activation patterns.^{1,4} For the subsequent leave-one-out cross-validation decoding, the vectors of binned single-trial temporal data were then concatenated across channels. We used the Mahalanobis distance³⁹ as a multivariate decoding metric to take advantage of the potentially monotonic relation between vowel category and neural activity.^{1,4,40} Responses to dissimilar vowels are expected to yield large Mahalanobis distance metrics, while responses to similar vowels are expected to yield low Mahalanobis distance metrics. Having been shown to be optimal for decoding,³⁷ a leave-one-out cross-validation approach was used per trial, wherein we calculated 3 pairwise distances between ECoG amplitude fluctuations measured in a given test trial and mean vectors of ECoG amplitude fluctuations averaged for each of the 3 vowels/positions in the remaining trials. A shrinkage-estimator covariance obtained from all trials, excluding the test trial, was used to compute the Mahalanobis distances.⁴¹ Combining Mahalanobis distance with Ledoit-Wolf shrinkage has been shown to

have performance advantages over other correlation-based methods of measuring brain-state dissimilarity,⁴² while Mahalanobis distance-based decoding has known advantages over linear classifiers and simple correlation-based metrics.⁴³

Single-trial relative Mahalanobis distance estimates were averaged across trials, resulting in a 3 x 3 distance matrix for each rat, time point, relative vowel position (N-1, N-2, N-3), and substitution type (noise vs. omission). To obtain overall decoding quality traces, the 3 x 3 distance matrices were subject to a subtraction of the averaged off-diagonal elements (mean distance between vowels) from the averaged diagonal elements (mean distance within vowels). The resulting decoding time-series were entered into a 2x3 repeated-measures ANOVA with within-subjects factors Block (predictable vs. random) and Position (N-1, N-2, N-3), separately for the two substitution types (noise vs. omission). The resulting statistical parametric maps were thresholded at $p < 0.005$ (uncorrected). Across time points, p values were corrected using a FWE approach at a cluster-level $p_{FWE} = 0.05$ ³².

We reasoned that significant decoding of the N-3 vowel in the predictable blocks, but not in the random blocks, would reveal predictive representations of the expected vowel. However, such representations may be formed both on an element-by-element basis (e.g., when hearing AOOAOOAOO, an "A" may be predicted because one is heard every 3 tokens), and also for an entire triplet (e.g., when hearing AOOAOOAOOX, "X" might also reactivate a representation of the AOO context). In a follow-up analysis, we wanted to test whether bursts/omissions reactivate representations containing (1) information about the entire preceding triplet, or (2) specific information about the N-3 vowel, independent of the rest of the triplet. To this end, we ran an additional decoding analysis, this time using a 18 x 18 stimulus matrix (corresponding to 18 possible triplets, with 3 phase shifts for each of the 6 unique triplets; e.g., for a unique triplet AAO, the three phase shifts would correspond to AAO, AOA, and OAA), yielding 18 x 18 Mahalanobis distance matrices. This analysis focused on the predictable blocks only and zoomed into two time clusters in which we observed significant N-3 vowel decoding (see [results](#)). To quantify the decoding of the entire triplet, we subtracted the mean of all off-diagonal elements of the 18 x 18 stimulus matrix from the mean of all diagonal elements ([Figure S3A](#)). To quantify the decoding of information about the N-3 vowel independent of the entire triplet identity, we subtracted the mean of those elements of the 18 x 18 stimulus matrix which did not share the first vowel from the mean of those elements of the matrix which did share the first vowel (excluding the diagonal elements, corresponding to identical triplets). The decoding estimates based on these representational dissimilarity matrices were subject to one-sample t -tests (two-tailed) across recording sessions (see [Figures S3B](#) and [S3C](#) for results).

Since vowel decoding was relatively weaker for N-3 and N-2 vowels (see [results](#); [Figure 2A](#)), we have performed an additional analysis aiming at verifying whether spatial maps of decoding sensitivity can be reasonably established for these vowel positions. To this end, we performed an additional analysis in which we repeated the spatial correlation analysis, but rather than correlating predictable and random blocks, we correlated decoding based on odd vs. even trials within each block.

In an additional analysis, since we observed univariate differences in vowel-evoked responses (see [results](#)), we tested whether decoding primarily relies on those channels that are also associated with sensory encoding of vowels. To this end, we repeated the decoding analysis for two subsets of channels - those which strongly differentiated between vowels (with the corresponding F statistic of the main effect of vowel on the vowel-evoked responses higher than the median across channels) and those which differentiated weakly between vowels (F statistic below median across channels). The resulting decoding time-series were compared between the two groups of channels using a series of paired t -tests, correcting for multiple comparisons across time points at a FWE-corrected $p_{FWE} = 0.05$ (cluster-level).³²

While we did not observe univariate differences in spectral peaks at the single vowel rate between condition (and we did not observe peaks at the triplet level overall; see [results](#)), in a further analysis we also tested whether decoding might rely on those channels which show relatively higher triplet-rate peaks than other channels. Again, we repeated the decoding analysis for two subsets of channels, this time splitting them based on the single-channel phase coherence estimates for the single vowel rate (2 Hz; above/below median). The two resulting decoding time-series were compared using a series of paired t -tests, correcting for multiple comparisons as above.

For completeness, we also performed the decoding analysis on the vowel-evoked responses themselves (see [Table S3](#) for results). While, given that vowel-evoked responses showed univariate effects of vowel identity, multivariate decoding was expected to be significant, we could use this analysis to compare the magnitude of decoding mnemonic information (N-1) based on burst-evoked responses, relative to decoding of vowel identity (N) based on vowel-evoked responses.

Multivariate analysis: Learning effect on decoding

Another question we wanted to address is whether any decoding benefit we might observe in the predictable stimulus condition reflects predictive neural processing. In particular, we hypothesized that, if the decoding boost in predictable blocks is related to predictive processing, it should gradually build up, as the auditory system needs time to detect repeating patterns and learn to use them for predictions of which sound token is expected when. This can occur at two time scales: first, decoding can improve with each subsequent vowel token embedded in a block of identical triplets (reflecting learning within blocks); second, decoding can improve over subsequent blocks (reflecting learning across blocks). To test these hypotheses, we constructed two trial-by-trial learning regressors - a "within blocks" regressor quantifying the vowel position within a block of identical triplets, and an "across blocks" regressor quantifying which block of a particular triplet it is within the entire recording session. To facilitate comparisons between the two regressors, the "within blocks" regressor only included vowel position from 1 (first burst within a sequence) to 6 (sixth burst), while the "across blocks" regressor was binned into 6 bins of 2 blocks in each bin (e.g., bin 1 contained the first 2 blocks of a particular triplet, while bin 6 contained the last 2 blocks of the same triplet). Both regressors were log-transformed to increase the relative effect of the first bursts/sessions relative to the last bursts/sessions.⁴⁴ We then repeated the decoding analysis of the N-3 vowel and,

per recording session and condition, performed a multiple linear regression with a constant term and the two learning regressors on single-trial decoding estimates. Specifically, for both of the time clusters in which we identified significant differences between decoding in predictable vs. random blocks, we selected the single-trial peak decoding within a given time cluster, and then normalized (z-scored) the trial-by-trial peaks per rat, recording session, and condition. This resulted in 8 sets of learning coefficients: (1) for predictable vs. random conditions, (2) quantifying learning within vs. across blocks, (3) estimated for early vs. late time window. The resulting regression coefficients (betas) were tested for significant differences between predictable and random blocks (treated as a baseline condition) using Wilcoxon sign rank test. While we hypothesized that learning effects should be specific to N-3 stimuli, in an additional analysis we also tested for the same learning effects on the decoding of N-2 and N-1 stimuli.

Multivariate analysis: Similarity between predictive and mnemonic representations

To test whether the predictive and mnemonic representations are shared, we quantified the spatial correlation of decoding topographies between predictable and random blocks. Our reasoning was that, if predictive and mnemonic representations are shared, decoding topographies should be similar between predictable and random blocks. On the other hand, if predictive and mnemonic representations are independent, the decoding topographies should be different between the two types of blocks. To this end, we repeated the decoding analysis, this time using a searchlight approach. Specifically, rather than using all channels for decoding, we used subsets of channels, with each subset forming a 3x3 grid. Different subsets overlapped by 1 row or column, resulting in 36 (6x6) decoding estimates based on the 3x3 grids, separately for each recording session, condition, and time point. We then correlated the spatial maps obtained for predictable and random blocks, separately for each recording session and time point. The resulting Pearson correlation coefficients were entered into a series of one-sample t-tests, correcting for multiple comparisons across time points at $p_{FWE} = 0.05$.³²

Multivariate analysis: Cross-temporal generalization

In a further analysis, we tested whether decoding a particular vowel generalizes across time points (suggesting that the reinstated representations rely on a similar neural code, independent of the latency of measured neural activity) and/or across vowel positions (suggesting that decoding one triplet element relies on a similar neural code as decoding another triplet element). To this end, we performed a cross-temporal generalization analysis, in which we repeated our multivariate decoding analysis but with an important modification of the leave-one-out cross-validation approach. First, to quantify generalization across time points, in calculating the Mahalanobis distance we incrementally shifted the latency of the test data with respect to the remaining trials, in 16 ms time steps - such that decoding was trained on one latency but tested on another. As a result of this approach, rather than decoding time series, per recording session and condition (predictable vs. random) we obtained decoding matrices with each matrix element representing the Mahalanobis distance between data measured at two different latencies. Second, to quantify generalization across vowel positions, we allowed the test data labels to be replaced by labels corresponding to another vowel than the remaining trials. As a result of this approach, rather than obtaining 3 decoding matrices (one per vowel position), we obtained 6 decoding matrices with the 3 additional matrices representing the Mahalanobis distance between data measured at two different vowel positions. The resulting decoding matrices were entered into a series of 6 GLMs (one per vowel position pair), each implementing a paired t-test between decoding estimates obtained for the predictable and random conditions. The resulting statistical parametric maps were thresholded at $p < 0.005$ (two-tailed) and corrected for multiple comparisons across spatiotemporal voxels at a FWE-corrected $p_{FWE} = 0.05$ (cluster-level).³²