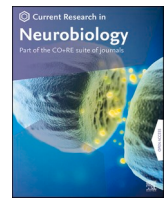




Contents lists available at ScienceDirect

Current Research in Neurobiology

journal homepage: www.sciencedirect.com/journal/current-research-in-neurobiology

Learning boosts the decoding of sound sequences in rat auditory cortex

Dan Luo^a, Kongyan Li^b, HyunJung An^a, Jan W. Schnupp^{a,*}, Ryszard Auksztulewicz^{a,c,**}^a Department of Neuroscience, City University of Hong Kong, Hong Kong Special Administrative Region^b Department of Biomedical Sciences, City University of Hong Kong, Hong Kong Special Administrative Region^c Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

ARTICLE INFO

Keywords:

Auditory cortex
ECoG
Sequence processing
Entrainment
Decoding

ABSTRACT

Continuous acoustic streams, such as speech signals, can be chunked into segments containing reoccurring patterns (e.g., words). Noninvasive recordings of neural activity in humans suggest that chunking is underpinned by low-frequency cortical entrainment to the segment presentation rate, and modulated by prior segment experience (e.g., words belonging to a familiar language). Interestingly, previous studies suggest that also primates and rodents may be able to chunk acoustic streams. Here, we test whether neural activity in the rat auditory cortex is modulated by previous segment experience. We recorded subdural responses using electrocorticography (ECoG) from the auditory cortex of 11 anesthetized rats. Prior to recording, four rats were trained to detect familiar triplets of acoustic stimuli (artificial syllables), three were passively exposed to the triplets, while another four rats had no training experience. While low-frequency neural activity peaks were observed at the syllable level, no triplet-rate peaks were observed. Notably, in trained rats (but not in passively exposed and naïve rats), familiar triplets could be decoded more accurately than unfamiliar triplets based on neural activity in the auditory cortex. These results suggest that rats process acoustic sequences, and that their cortical activity is modulated by the training experience even under subsequent anesthesia.

1. Introduction

Chunking of auditory input streams is an essential aspect of auditory perception for human and non-human listeners alike. It is thought that the mammalian auditory system discovers the appropriate rules for segmenting sound streams from observing statistical regularities in its auditory inputs (Dehaene et al., 2015). Developmental studies in humans have shown that infants could discriminate transitional probabilities in auditory sequences after brief passive exposure to statistically regular stimulus streams (Saffran et al., 1996), which suggests that chunking might be a relatively automatic process and a likely precursor of language acquisition (Romberg and Saffran, 2010; Wilson et al., 2017). Noninvasive recordings of neural activity in humans (Ding et al., 2015, 2017) have identified plausible correlates of auditory chunking, relying on the entrainment (phase-synchronization) of cortical activity to distinct time scales of speech streams (monosyllabic words, two-word syntactic phrases, and longer sentences). Crucially, the neural entrainment to speech streams was modulated by the participants' familiarity with the language in which speech was presented. While these results

provide a robust neural correlate of the human ability to segment speech streams, it is unknown whether animals could use similar chunking mechanisms to process familiar stimulus sequences.

In nonhuman primate models, sequence processing has typically been studied using artificial grammar paradigms and focused on cross-species comparisons. In the auditory domain, neural oscillatory signatures of sequence learning in the auditory cortex have been shown to be largely conserved between macaques and humans (Kikuchi et al., 2017), although unlike in the human studies (Batterink and Paller, 2019; Ding et al., 2015), the low-frequency entrainment effects were not frequency-specific to the stimulus presentation rates used in the artificial grammar sequences. Behavioral studies identified differences between species. For instance, Jiang et al. (2018) demonstrated that macaques could learn center-embedded relationships, such as ABC|CBA, in visuospatial sequences after extensive training, while human infants were sensitive to such artificial grammar structures after only brief exposure. In an earlier study (Wilson et al., 2013), macaques could differentiate more complex auditory artificial grammar structures relative to marmosets, which could detect only simple structure violations, e.g., based

* Corresponding author.

** Corresponding author. Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany.

E-mail addresses: wschnupp@cityu.edu.hk (J.W. Schnupp), ryszard.auksztulewicz@ae.mpg.de (R. Auksztulewicz).¹ These authors contribute equally.<https://doi.org/10.1016/j.crneur.2021.100019>

Received 12 March 2021; Received in revised form 11 June 2021; Accepted 7 July 2021

Available online 10 July 2021

2665-945X/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>.

on the first position in a sequence.

In rodents, the evidence for learning temporal regularities based on stimulus sequences is mixed. While one earlier study suggested that rats cannot extract abstract rules from stimulus sequences but are sensitive to the statistical co-occurrence of elements within the sequence (Toro and Trobalón, 2005), another study (Murphy et al., 2008) found that rodents can be trained to learn abstract rules from stimulus sequences and transfer them to new stimuli. Moreover, Bale et al. (2017) found that mice could be trained to discriminate tactile temporal sequences based on stimulus order only. Another study (Gavornik and Bear, 2014) provided electrophysiological evidence of sequence learning in awake mice, namely that the primary visual cortex of trained (but not naïve) animals showed larger evoked responses and higher peak firing rates for familiar vs. unfamiliar visual sequences. These results suggest that cortical neural activity of rodents could be shaped by previous experience of (visual) sequences. However, whether similar effects generalize to the auditory modality, and to what extent animals need to be trained to show sensitivity to sequences, remains unclear.

Here, we recorded neural activity in response to auditory sequences in anesthetized rats ($N = 11$). Prior to recording, the animals were split into three groups: a trained group was familiarized with syllable sequences using operant conditioning; a passive group received passive exposure to the same sequences; finally, a naïve group did not receive any exposure to these sequences. Using ECoG recordings from the auditory cortex, we quantified neural entrainment to segments of different lengths (syllables and triplets), tested whether entrained signals can be found in cortical responses of anesthetized rats, and whether they can be modulated by either active or passive auditory experience. Finally, using multivariate decoding of acoustic segments based on spatiotemporal patterns of neural activity, we tested whether stimulus decoding is modulated by prior experience. In this manner, we were able to identify a new neural signature of acoustic stream segmentation in anesthetized rodents.

2. Materials and methods

Eleven female Wistar rats were used in this study. The hearing threshold of all rats was normal, as confirmed by auditory brainstem response (ABR) recordings. Rats in the training group ($n = 4$) started behavioral training at 8 weeks of age, and once they finished training, they were subject to ECoG recordings. Another, passive group of rats ($n = 3$) was exposed to prolonged continuous acoustic stimulation prior to ECoG recordings, without any additional training. Finally, a group of naïve rats ($n = 4$) without any training experience were only subject to ECoG recordings. The mean age (\pm SD) at the time of ECoG recordings was 28.58 ± 7.83 weeks, and the weight 306.36 ± 43.36 g. All

experimental procedures were approved by the Animal Research Ethics Sub-Committee of the City University of Hong Kong under the animal license No. (17-76) in DH/SHS/8/2/5 Pt.1.

Artificial acoustic syllables selected from a database of Consonant-Vowel syllables (Ives et al., 2005) were used in behavioral training and subsequent recordings (see Fig. 1A). The speech syllables were analyzed and resynthesized by an open-source vocoder, STRAIGHT (Kawahara, 2006) for MATLAB R2018b (MathWorks), to match the stimulus onset and duration of all syllables (531 ms) and shift the fundamental frequency and formant scalar of each CV syllable upward 1 octave to match the optimal rat hearing range (Kelly and Masterton, 1977).

Animals in the trained group performed two-alternative-forced-choice behavior tasks using drinking water as a positive reinforcer. The equipment and task were adapted from (Li et al., 2019). Behavioral training consisted of three separate training stages: a syllable learning stage (training rats to discriminate entire syllables and scrambled syllables), a triplet learning stage (to discriminate entire triplets and scrambled triplets), and a triplet discrimination training stage (to discriminate familiar triplets and novel unfamiliar triplets). In the first two stages (Fig. 1B), we used four syllable triplets as “familiar” stimuli (/pisore/, /gusima/, /dazolu/, /pekina/) and scrambled stimuli as “unfamiliar” stimuli (generated by epoching the waveform of each syllable into 10 segments of equal length, and shuffling the segments; Fig. 1A). In the third and final training stage, we used the same four syllable triplets as “familiar” stimuli and other syllable triplets (e.g., /gukewo/) as “unfamiliar stimuli”. Once the performance of animals in a particular training stage approached 75% accuracy, they progressed to the next stage. Rats performed approximately 300 trials per session and one session per day. The total duration of training was 33.15 ± 0.52 weeks (mean \pm SD). To test whether this ability could transfer to novel stimuli, three animals were tested in a behavioral session using familiar stimuli and previously unheard unfamiliar stimuli.

Animals in the passive group were exposed to continuous presentation of “familiar” triplets (/pisore/, /gusima/, /dazolu/, /pekina/) for 24 h immediately prior to the electrophysiological recordings. Triplets were presented in a random order through loudspeakers at 80–85 dB in a sound-attenuated box. No further behavior or task was required.

In this manner we created three cohorts of rats. The trained group had experienced the target syllable triplets as reinforced stimuli, the passive exposure group had been exposed to the target triplets for 24 h, and the naïve group had no experience at all of the target triplets. All three groups were then subjected to identical electrophysiological recordings. There were two conditions in the ECoG experiment, delivered in separate trials. The order of the trials was randomized across rats. In the first (familiar) condition, each trial contained a 35 s long auditory

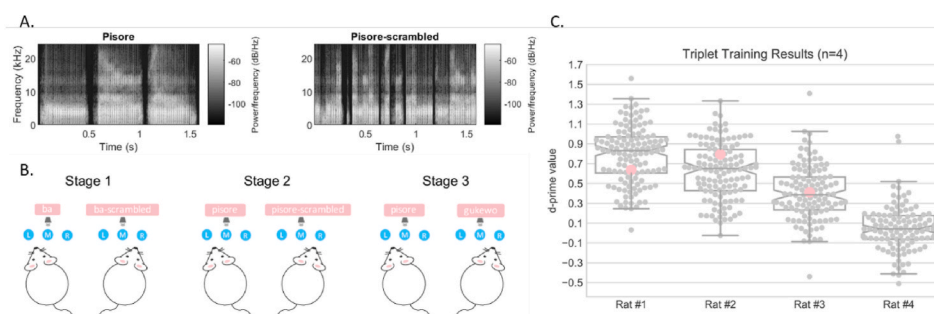


Fig. 1. Behavioral stimuli and training paradigm. (A) Spectrograms of two example stimuli (/pisore/ and its scrambled equivalent). (B) Two-alternative forced choice training paradigm. There were three training stages in total. In training stage 1, there was an additional spatial cue for helping rats to recognize the syllables. This spatial cue was gradually removed as rats reached a threshold of 75% accuracy in discriminating familiar (i.e./ba/ → lick left side) and unfamiliar stimuli (i.e./ba/-scrambled → lick right side). In training stages 2 and 3, rats were conditioned to discriminate familiar (i.e./pisore/ → lick left side) and unfamiliar (i.e./pisore/-scrambled or /gukewo/ → lick

right side) triplets rather than single syllables. (C) Behavioral performance of individual rats during training. The grey dots denote the training results of each training session (d' -prime values) and the pink dots represent the test results (d' -prime values) in the testing session. Performance of all rats was above chance level (bootstrapped, all $p < 0.05$), and three out of four rats which approached 70% accuracy in the triplet learning stage also performed significantly above-chance in the test session (chance level: $d' = 0$, bootstrapped, $p < 0.05$).

stimulus, composed of 20 familiar triplets (4 triplets, each with 5 repeats, presented in a random order) with no gap between triplets, padded with a 1.5 s silence at the onset and offset of each trial. In the second (unfamiliar) condition, triplets were modified to form unfamiliar sequences the rats had never heard before. Unfamiliar triplets were generated by (1) replacing the middle syllable with another familiar syllable (taken from another triplet), or (2) replacing the middle syllable with an unfamiliar syllable (never heard before), or (3) switching the order of the middle and final syllables, or combinations thereof. Each rat was exposed to 60 familiar stimulus trials and 60 unfamiliar stimulus trials.

After anesthetic induction and ABR recordings, a 4×5 mm craniotomy was performed, extending from 0 to 4 mm ventral from the temporal edge and 0–5 mm posterior from the Bregma, to expose the right temporal lobe, and the dura matter was carefully removed. Electrophysiological data were recorded by an 8×8 rodent ECoG electrode grid at a sampling rate of 24,414 Hz, acquired and amplified by Tucker-Davis Technologies (TDT) RZ2 Bioamp Processor and TDT PZ5 Neuro-Digitizer, and controlled by the BrainWare software. While in the present study we recorded responses from both primary and secondary auditory cortical fields, we did not acquire data allowing for channel mapping into separate regions. Auditory stimuli were delivered by a TDT RZ6 multiprocessor sampling at 48,828 Hz and presented by a custom-made earphone with a flat frequency response (calibrated by a G.R.A.S 46DP-1 microphone). Continuous ECoG data were first band-pass filtered between 0.1 and 48 Hz using a 6th order two-pass Butterworth filter, and then downsampled to 150 Hz. To emphasize signal components which increase neural response repeatability across stimuli, we epoched the data into segments corresponding to a single triplet presentation and denoised the epoched data using the Dynamic Separation of Sources toolbox (de Cheveigné and Simon, 2008).

3. Results and discussion

First, over multiple training sessions, the performance of all trained animals ($n = 4$), quantified as the sensitivity index d' , was significantly above chance level (see Fig. 1C and Table 1, one sample t -test, all $p < 0.05$). This result indicates that all trained animals could differentiate the familiar triplet stimuli from the unfamiliar control stimuli during the training sessions. The test result showed that the performance (d') of all tested animals ($n = 3$) was above chance (bootstrap, $p < 0.001$), implying that animals did learn the familiar triplets in one context, and could recognize them when presented among different, unfamiliar triplets in another context. One remaining trained rat was not tested as it did not achieve 70% accuracy in the triplet training sessions which we had set as a criterion for concluding the training, but the performance during training was still above chance (see Table 1). In the present study, to maintain the novelty of the unfamiliar triplet stimuli used in subsequent electrophysiological recordings, we did not expose rats to the same unfamiliar triplets during the training. While rats could only respond after sound sequence offset, we cannot exclude the possibility that they performed the task by memorizing parts of the triplets, instead of learning entire triplets. However, for the purpose of this study it was sufficient to ensure that the trained and the passive exposure group had extensive opportunity to become familiar with the sound sequences used

in subsequent recordings.

To test whether previous findings in humans – namely that familiarity with sound sequences yields spectral peaks in neural activity at rates specific to the duration of those sequences (Ding et al., 2015) – generalize to animal models, we analyzed the ECoG signals in the frequency domain (Fig. 2A). We calculated the Fourier power spectrum values of the continuous ECoG signals and normalized each power spectrum by dividing the power estimate for each frequency point within the 0.25–2.5 Hz range by the sum of all power estimates in this range. We observed that, across all rats, the spectral peaks for the syllable rate (1.88 Hz) were significantly higher than for the neighboring frequencies (Fig. 2AB; Wilcoxon sign rank tests; $Z_{\text{familiar}} = 2.934$, $P_{\text{familiar}} = 0.003$; $Z_{\text{unfamiliar}} = 2.934$, $P_{\text{unfamiliar}} = 0.003$). While triplet rate peaks were nominally higher relative to neighboring frequencies, reflecting minor peaks in the stimulus spectrum (Fig. 2C), this effect was weak and did not survive Bonferroni correction for multiple comparisons across the four tests ($Z_{\text{familiar}} = 2.045$, $P_{\text{familiar}} = 0.041$, $Z_{\text{unfamiliar}} = 2.401$, $P_{\text{unfamiliar}} = 0.016$). Furthermore, no differences in the triplet rate peak were observed between groups (pairwise Wilcoxon rank sum tests between trained, naïve, and passive groups: all $p > 0.2$), suggesting that there were no training-specific spectral signatures in our sample of anesthetized rats. It should be noted that while some earlier studies in humans (Batterink and Paller, 2019) suggested that low-frequency spectral peaks can be observed without attention, other studies demonstrated that diverted attention (Ding et al., 2018) and sleep (Makov et al., 2017) disrupt the neural correlates of acoustic chunking. Therefore, it cannot be ruled out that the animals' state might have influenced the amplitude of spectral peaks, and future studies should test whether performing neural recordings in awake animals could uncover stronger signatures of neural entrainment to the sequence presentation rate.

Next, we turned to the analysis of neural signals in the time domain. We applied a principal component analysis to the denoised data after concatenating data across all animals and trials, and retained the first component (explaining 52.96% of the variance). Then, we segmented the data into epochs, separately for each animal, trial, and triplet. Based on the segmented data, we calculated the average root-mean-square (RMS) amplitude over 20 triplets, separately for each trial and time point. Here we found that the response waveforms of ECoG responses were quite diverse across rats, and there was no consistent difference in responses to familiar and novel stimuli between trained, passive, and naïve animals. Specifically, a repeated-measures ANOVA was conducted to compare the single-trial RMS values of time courses between conditions (familiar vs. unfamiliar) and groups (trained vs. naïve). In the ANOVA, neither the main effect of condition (all $p > 0.05$, FDR-corrected across time points) nor a significant interaction between group and condition was observed (all $p > 0.05$, FDR-corrected). However, we observed a significant main group effect at all three syllables (all $F_{2,1298} > 10.826$, all $p < 0.05$, FDR-corrected), indicating that acoustic training had modulated the overall neural responses to sound stimuli. Furthermore, we compared the RMS amplitude curves separately for each animal between the two conditions (familiar vs. unfamiliar; Fig. 3) and found significant differences at the second and third syllable positions in all animals except one (t -test, all $p < 0.05$, FDR-corrected across time points). Still, the direction of these differences

Table 1
Behavioral training information and results of all animals.

Behavioral Results in Triplet Discrimination Training Stage							
	Total Session No.	Trial No./Session (Mean \pm SD)	Accuracy (%) (Mean \pm SD)	Significance (P-values)	T-values	df	Accuracy in Test Session (%) / Trial No.
Rat #1	127	287.42 \pm 97.02	64.14 \pm 5.39	$P < 0.0001$	29.54	126	62.54/289
Rat #2	123	301.08 \pm 93.77	61.88 \pm 5.26	$P < 0.0001$	25.02	122	64.88/197
Rat #3	118	235.34 \pm 91.54	57.03 \pm 6.05	$P < 0.0001$	12.62	117	56.58/239
Rat #4	112	260.02 \pm 89.33	50.86 \pm 4.13	$P = 0.029$	2.02	111	Nil

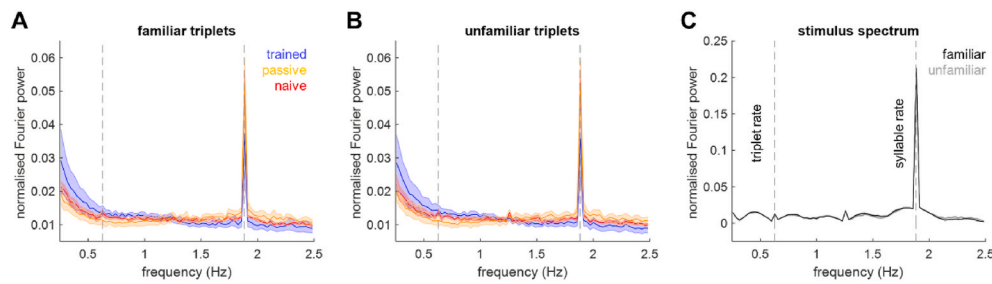


Fig. 2. Frequency-domain neural activity. (A) Normalized power spectrum values calculated for familiar triplets in the trained group ($n = 4$; blue), passive group ($n = 3$, orange) and naïve group ($n = 4$, red), showing a robust peak at the syllable rate but no robust peak at the triplet rate. No significant differences were observed between groups. Shaded areas denote SEM across rats. (B) Average power spectra for unfamiliar triplets. Legend as in (A). (C) Average power spectra based on stimulus envelope. Black: familiar stimuli; grey: un-

familiar stimuli; shaded areas denote SEM across trials.

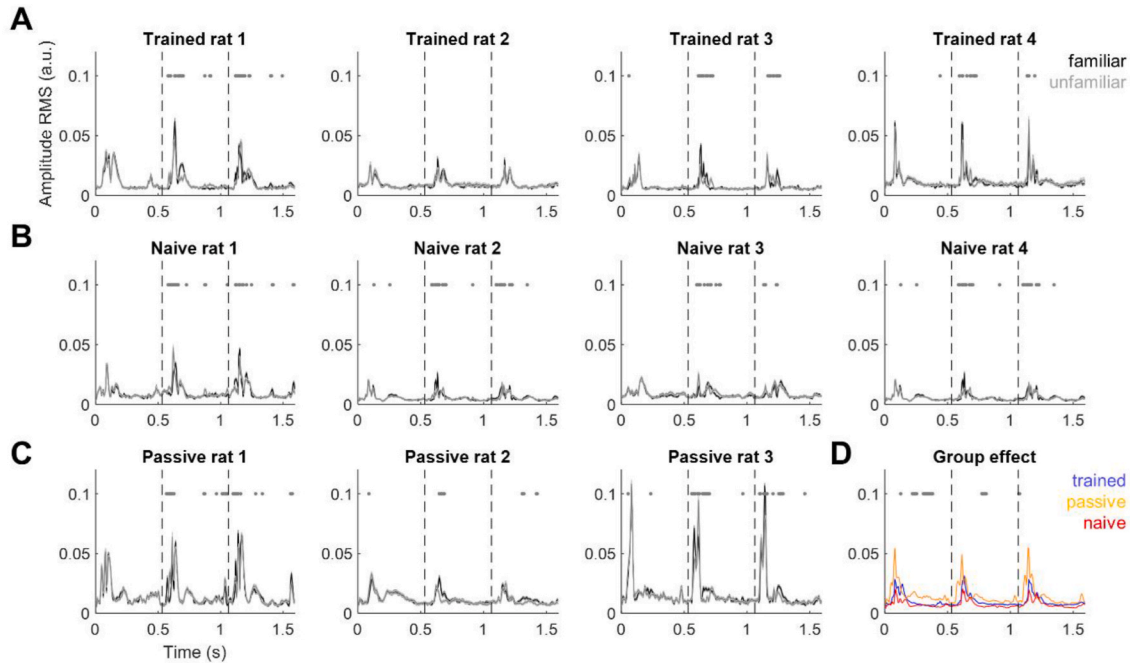


Fig. 3. Time-domain neural activity. (A–C) Blue/red lines denote the (RMS) amplitude over triplets of the ECoG response to familiar/unfamiliar sequences for each individual rat from the trained group (A), naïve group (B), and passive group (C). Grey dots show the significance of the main effect of stimulus familiarity per time point of the RMS amplitude (t -test, $P < 0.05$, FDR corrected). (D) RMS time courses averaged across individual rats, per group. Grey dots show a significant main effect of group (ANOVA, $P < 0.05$, FDR corrected).

was not consistent across syllables and rats (as evident from the lack of the main effect of condition in the repeated-measures ANOVA), indicating that average neural responses to familiar and unfamiliar stimuli were heterogeneous in our sample (An et al., 2021). Thus, unlike previous findings (Sanders et al., 2002, 2009), showing that a larger N1 amplitude of the EEG response at the first-syllable position in human subjects after familiarization may be a neural correlate of auditory chunking, in the present study we did not observe consistent amplitude differences or peaks in the time course analysis. While this suggests that the neural responses averaged over the entire AC region are not a sensitive measure of prior experience with familiar sequences, it does raise the possibility that sequence familiarity might instead affect more fine-grained population responses, accessible through multivariate analyses.

To determine whether stimuli can be decoded from neural activity, and whether stimulus familiarity affects decoding, we quantified the relative multivariate Mahalanobis distance (dissimilarity; (Ledoit and Wolf, 2004)) between spatial patterns of neural responses to different familiar and unfamiliar stimuli. In the first step, we calculated the RMS over time points of the ECoG amplitude for each channel, stimulus (i.e., each triplet), syllable, and rat. Then, per stimulus type (familiar vs.

unfamiliar), syllable, and rat, we used a leave-one-out cross-validation method to calculate the single-trial Mahalanobis distance values between the vector of RMS amplitudes concatenated across channels for a particular trial, and four vectors of average RMS amplitudes (averaged across the remaining trials, separately for each of the four triplets) concatenated across channels. The single-trial decoding estimate was calculated as a difference between (1) the dissimilarity between RMS amplitudes observed in this trial vs. other trials of the same stimulus, and (2) the average dissimilarity between RMS amplitudes observed in this trial vs. other stimuli. These single-trial decoding estimates were averaged across trials, separately for each stimulus type (familiar vs. unfamiliar), syllable, and rat. Finally, per syllable and rat, we calculated a ratio between the averaged decoding estimates for familiar vs. unfamiliar stimuli. These ratios were compared between trained, passive, and naïve rats, separately for each syllable.

As expected, no difference between trained, passive, and naïve animals was observed for the first syllable (permutation tests; trained vs. naïve: $p = 0.301$; trained vs. passive: $p = 0.192$; passive vs. naïve: $p = 0.337$; Fig. 4), since the first syllables were physically identical in both types of triplets. At the second syllable, while we did not find evidence for a consistent, significant difference between groups (trained vs. naïve:

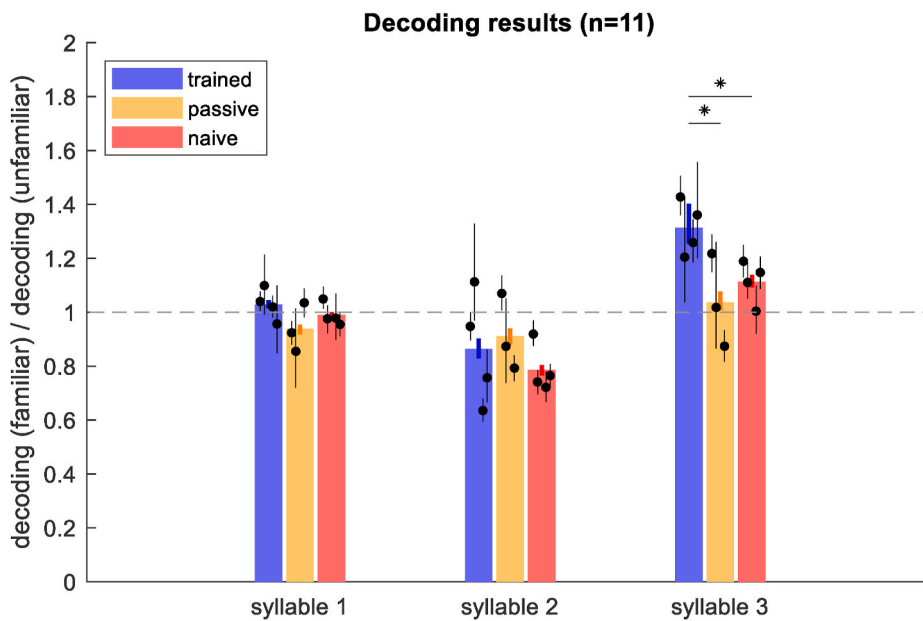


Fig. 4. Multivariate analysis. The blue bars show the decoding index (the ratio of decoding familiar stimuli to decoding unfamiliar stimuli) for trained rats whereas the orange/red bars indicate the decoding index for the passive/naïve groups respectively. Thick blue/orange/red lines indicate the 95% confidence interval (CI) range of decoding indices at the group level. Each black dot represents the respective decoding index per rat, and each black line stands for the 95% CI of individual rats' decoding indices. Asterisks denote significant differences between groups ($p < 0.05$).

$p = 0.190$; trained vs. passive: $p = 0.300$; passive v. naïve: $p = 0.080$; Fig. 4), one of the four trained rats showed robust improvement in decoding for familiar vs. unfamiliar stimuli. Crucially, we found a robust difference between groups at the third syllable position, where the decoding values in the trained group were significantly higher than in the remaining groups (trained vs. naïve: $p = 0.010$; trained vs. passive: $p = 0.002$; passive v. naïve: $p = 0.240$; Fig. 4).

Consistent with our hypothesis and with previous studies in the visual modality (Gavornik and Bear, 2014), all trained animals were more sensitive to the familiar stimuli than to the unfamiliar stimuli (decoding ratio, including 95% confidence intervals, above 1 for all trained animals; Fig. 4). Our results are also consistent with a previous study in humans (Batterink et al., 2015), in which familiarity effects were strongest on the late/final sequence elements, and were specifically improved by previous extensive training, but not by passive exposure. Importantly, in our study the decoding boost was observed in neural activity recorded during post-training anesthesia. While one previous study in cats has identified rapid effects of passive exposure to visual sequences (movies with natural scenes) on cortical activity under anesthesia (Yao et al., 2007), most studies in animal models used extensive behavioral training and recordings in awake animals to identify the behavioral and/or neural correlates of sequence processing (Bale et al., 2017; Gavornik and Bear, 2014; Homann et al., 2017; Murphy et al., 2008). However, given a recent study based on single-neuron recordings in the auditory cortex of awake mice (Libby and Buschman, 2021), in which neural signatures of passive sequence learning were shown to gradually increase over several days, it cannot be ruled out that decoding would also have been observed in our passive group after more extensive exposure. Nevertheless, the effect of prior training on shaping the neural responses to familiar stimuli under anesthesia is consistent with earlier findings in rats that training-induced receptive field plasticity shifts the sensitivity of neuronal populations in the auditory cortex toward the reinforced sound frequency (David et al., 2012; Fritz et al., 2005). Here, we show that learning experience boosts the ability to decode stimulus information from neural activity also in the context of learning temporally extended sequences.

In summary, we show that training experience can improve the neural sensitivity to sequences in rodents, although the neural correlates typically observed in humans (low-frequency entrainment to sequence presentation rate) are not detectable in the neural activity of

anesthetized rodents. Instead, we show that behavioral training leads to improvements in decoding stimulus-related information from the spatial pattern of neural activity in the auditory cortex, even under anesthesia. Future studies should test the behavioral relevance of these signals by relating the neural activity to behavioral responses in awake and behaving animals.

Author contributions

J.W.S. and R.A. provided the initial idea and designed the research. D.L. performed animal behavioral training and ECoG recording. K.L. and H.A. helped with the ECoG recording. R.A. and D.L. analyzed the data. D.L. drafted the manuscript. R.A. and J.W.S. revised the manuscript and supervised this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Dr. Ambika Prasad Mishra for the technical assistance during the ECoG experiment preparation. We also thank Drs Nai Ding, Jianfeng Zhang, Manuel S. Malmierca, and two anonymous reviewers for constructive comments on previous versions of the manuscript. This work was supported by European Commission's Marie Skłodowska-Curie Global Fellowship (750459 to R.A.), the Hong Kong General Research Fund (11100518 to R.A. and J.W.S.), and a grant from European Community/Hong Kong Research Grants Council Joint Research Scheme (9051402 to R.A. and J.W.S.).

Appendix A. Peer Review Overview and Supplementary data

A Peer Review Overview and (sometimes) Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.crneur.2021.100019>.

References

- An, H.J., Auksztulewicz, R., Kang, H.J., Schnupp, J.W.H., 2021. Cortical mapping of mismatch responses to independent acoustic features. *Hear. Res.* 399, 107894.
- Bale, M.R., Bitzidou, M., Pitas, A., Brebner, L.S., Khazim, L., Anagnou, S.T., Stevenson, C. D., Maravall, M., 2017. Learning and recognition of tactile temporal sequences by mice and humans. *Elife* 6, e27333.
- Batterink, L.J., Paller, K.A., 2019. Statistical learning of speech regularities can occur outside the focus of attention. *Cortex* 115, 56–71.
- Batterink, L.J., Reber, P.J., Paller, K.A., 2015. Functional differences between statistical learning with and without explicit training. *Learn. Mem.* 22 (11), 544–556.
- David, S.V., Fritz, J.B., Shamma, S.A., 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109 (6), 2144–2149.
- de Cheveigné, A., Simon, J.Z., 2008. Denoising based on spatial filtering. *J. Neurosci. Methods* 171 (2), 331–339.
- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., Pallier, C., 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88 (1), 2–19.
- Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., Poeppel, D., 2017. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Front. Hum. Neurosci.* 11, 481.
- Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2015. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19 (1), 158–164.
- Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., Zhang, J., 2018. Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. *J. Neurosci.* 38 (5), 1178–1188.
- Fritz, J., Elhilali, M., Shamma, S., 2005. Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear. Res.* 206 (1–2), 159–176.
- Gavornik, J.P., Bear, M.F., 2014. Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* 17 (5), 732–737.
- Homann, J., Koay, S.A., Glidden, A., Tank, D., Berry, M., 2017. Predictive coding of novel versus familiar stimuli in the primary visual cortex. *BioRxiv*. <https://doi.org/10.1101/197608>, 197608.
- Ives, D.T., Smith, D.R.R., Patterson, R.D., 2005. Discrimination of speaker size from syllable phrases. *J. Acoust. Soc. Am.* 118 (6), 3816–3822.
- Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., Wang, L., 2018. Production of supra-regular spatial sequences by macaque monkeys. *Curr. Biol.* 28 (12), 1851–1859.
- Kawahara, H., 2006. STRAIGHT, exploitation of the other aspect of VOCODER: perceptually isomorphic decomposition of speech sounds. *Acoust. Sci. Technol.* 27 (6), 349–353.
- Kelly, J.B., Masterton, B., 1977. Auditory sensitivity of the albino rat. *J. Comp. Physiol. Psychol.* 91 (4), 930.
- Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A.E., Nourski, K.V., Gander, P.E., Kovach, C. K., Kawasaki, H., Griffiths, T.D., Howard III, M.A., Petkov, C.I., 2017. Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol.* 15 (4), e2000219.
- Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. *J. Portfolio Manag.* 30 (4), 110–119.
- Li, K., Chan, C.H.K., Rajendran, V.G., Meng, Q., Roskoth-Kuhl, N., Schnupp, J.W.H., 2019. Microsecond sensitivity to envelope interaural time differences in rats. *J. Acoust. Soc. Am.* 145 (5), EL341–EL347.
- Libby, A., Buschman, T.J., 2021. Rotational dynamics reduce interference between sensory and memory representations. *Nat. Neurosci.* 24 (5), 715–726.
- Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., Golumbic, E.Z., 2017. Sleep disrupts high-level speech parsing despite significant basic auditory processing. *J. Neurosci.* 37 (32), 7772–7781.
- Murphy, R.A., Mondragón, E., Murphy, V.A., 2008. Rule learning by rats. *Science* 319 (5871), 1849–1851.
- Romberg, A.R., Saffran, J.R., 2010. Statistical learning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1 (6), 906–914.
- Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical learning by 8-month-old infants. *Science* 274 (5294), 1926–1928.
- Sanders, L.D., Ameal, V., Sayles, K., 2009. Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia* 47 (4), 1183–1186.
- Sanders, L.D., Newport, E.L., Neville, H.J., 2002. Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat. Neurosci.* 5 (7), 700–703.
- Toro, J.M., Trobalón, J.B., 2005. Statistical computations over a speech stream in a rodent. *Percept. Psychophysiology* 67 (5), 867–875.
- Wilson, B., Marslen-Wilson, W.D., Petkov, C.I., 2017. Conserved sequence processing in primate frontal cortex. *Trends Neurosci.* 40 (2), 72–82.
- Wilson, B., Slater, H., Kikuchi, Y., Milne, A.E., Marslen-Wilson, W.D., Smith, K., Petkov, C.I., 2013. Auditory artificial grammar learning in Macaque and Marmoset monkeys. *J. Neurosci.* 33 (48), 18825–18835.
- Yao, H., Shi, L., Han, F., Gao, H., Dan, Y., 2007. Rapid learning in cortical coding of visual scenes. *Nat. Neurosci.* 10 (6), 772–778.