Title: Time-independent sensory representations are essential for behavioral learning and perception

Authors: Sophie Bagur[1,†,*], Jacques Bourg[1,†], Alexandre Kempf[1], Thibault Tarpin[1], Khalil Bergaoui[1], Yin Guo[1], Sebastian Ceballo[1], Joanna Schwenkgrub[1], Jean-Luc Puel[2], Jérôme Bourien[2], Brice Bathellier[1,*]

Affiliations:

[1] Institut Pasteur, Université Paris-Cité, INSERM, Institut de l'Audition, 63 rue de Charenton, F-75012 Paris, France.
[2] Institut des Neurosciences de Montpellier, Université de Montpellier, INSERM, Montpellier, France.
[†] equal contribution
[*] Corresponding authors : brice.bathellier@pasteur.fr, sophie.bagur@pasteur.fr

Abstract:

The brain constantly associates time-varying sensory inputs with behavioral decisions. However, how temporal information in sensory neuronal circuits is linked to perception and behavioral output remains unclear. Here, by training mice to categorize patterned optogenetic stimulations in the auditory cortex, we show that they fail to learn a discrimination between two activity patterns that differ only by their temporal structure. Moreover, using large scale recordings throughout the auditory system, we show that the auditory cortex reformats upstream representations of sounds to make temporal information available as a time-independent rate code, coexisting with temporally structured activity. Thus, by resolving the inefficiency of downstream learning rules in the temporal domain, the auditory cortex enables specific associations of temporally-structured sounds to perceptual decisions.

One Sentence Summary: Auditory cortex reformats temporal information to make it available for downstream learning rules blind to time sequences.

Main text: Sensory inputs to the brain contain temporal motifs which generate neuronal activity sequences, commonly observed throughout the visual (*1*, *2*), auditory (*3–5*), tactile (*6*, *7*) and olfactory (*8*, *9*) systems, including sensory cortex. In single cortical neurons, these activity sequences carry information which is not available in the neuron's mean firing rate (*1*, *3*, *7*, *8*). However, it is unknown if this temporal information present in neuronal circuits is used to compute behavioral decisions. On the one hand, several computational models (*10–12*) and experimental findings (*13*, *14*) have proposed neuronal and synaptic plasticity mechanisms by which temporal sequences can be associated to a behavioral response, e.g. using spike-timing-dependent synaptic plasticity (*15*). On the other hand, numerous learning models assume that the key factor for associative learning is the firing rate of pre- and post-synaptic neurons (*16*, *17*) and spike-timing-dependent plasticity rules also support efficient learning from firing rate information alone, particularly with irregular *in vivo* firing patterns (*12*, *18*, *19*). Since temporal and firing rate information almost always coexist in neural representations (*20*), the concrete role played by temporal information in behaving animals cannot be resolved without causal manipulations of neuronal activity to test if temporal activity sequences alone can drive sensory-motor learning. The auditory system is uniquely suited to address this fundamental question. Natural sounds contain rich temporal variations of intensity and frequency (*21*) that are key features influencing sound recognition (*22*) and perceptual properties such as timber or loudness (*23*, *24*). Moreover, sounds bearing biological meaning, such as words or vocalizations often develop over several hundreds of milliseconds, a time scale that challenges models which learn spike sequences based on much shorter neuronal and synaptic plasticity time windows (STDP : < 50-100 ms) (*10*).

Mice cannot learn to discriminate artificial temporal patterns in auditory cortex

To address the role of temporal information in sensory-motor learning, we tested if mice can learn to discriminate between artificial patterns which can only be differentiated using temporal cues. To do this, we trained mice to discriminate between a neuronal activity sequence in the auditory cortex (AC) and its time-reversed rendition. We generated these activity patterns by optogenetically activating pyramidal neurons in Emx-Cre x flex-ChR2 mice using mesoscopic light spots reliably positioned in the tonotopic fields of AC (**Fig. 1A,B, S1A**). Mice were trained to lick within a 1.5s opportunity window after Go pattern onset in order to receive a rewarding medial forebrain bundle stimulation (*25*) and not lick for the NoGo pattern to avoid a timeout (**Fig. 1C,D, S1B,C**). As a reference task in which temporal information is irrelevant, we trained mice to discriminate the optogenetic activation of two spatially separate spots (rate-coded task : A vs B, **Fig. 1C**) activated at 20 Hz during 500 ms. Consistent with previous results (*26*), mice rapidly learned to discriminate these inputs reaching 70% accuracy within 69+/-385 trials (**Fig. 1E**). Counterbalancing task order, we trained the same mice to discriminate the successive activation of spot A then spot B against the reverse sequence (temporal-coded task : A-B vs B-A) (**Fig. 1C)**. Each individual spot was activated for 250ms with 20Hz light-pulse trains. The activation of A-driven and B-driven neurons was separated by 25ms to minimize synaptic interactions which could locally convert temporal information to rate information (**fig. S2**). In contrast to the rate-coded task, learning was extremely slow and inefficient in the temporal task. After 3000 training trials, none of the mice could discriminate the temporal

sequences, while most of them could discriminate the rate-coded patterns (**Fig. 1E, F**). Pushing training to even higher trial numbers, only two of seven mice reached slightly above chance levels for the temporal-coded task (**fig. S1D,E**). This demonstrates that temporal information present in the cortical activity driven by the two time-reversed optogenetic sequences is not easily accessible to downstream learning processes for sensory-motor associations.

Time-independent sound representations emerge in the auditory cortex

This striking observation seems to contradict the strong perceptual difference between a sound and its time-reversed rendition (*22–24*) and the ability of rodents, including mice, to learn to discriminate these stimuli, an ability which critically relies on AC (*27*, *28*). To solve this apparent contradiction, we investigated whether AC transforms temporal information about sounds into a new format adapted to the inability of downstream circuits to learn from temporal cues. We performed large-scale recordings in three successive regions of the auditory system: the inferior colliculus (IC), the auditory thalamus (TH) and the AC (**Fig. 2A, table S1**), combined with auditory nerve (AN) responses simulated with a detailed biophysical model. In each region, we measured the responses to a set of 140 sounds, mainly of 500 ms duration, which covered simple, spectral and temporal features (**Fig. 2B, table S2**). In the IC, we coupled electrophysiology to record 563 single units in the primary IC (central nucleus of IC) with 2-photon calcium imaging to record 13.132 ROIs from the more superficial secondary IC (dorsal cortex of IC). We also imaged 39.191 TH axonal boutons spread throughout AC and recorded 498 single units directly in TH. Finally, we imaged 60.822 ROIS throughout all subregions of the AC down to layer V. Calcium signals were linearly deconvolved (*29*), providing a temporal resolution of ~150ms. Full details of the dataset are provided in the **Supplementary text** and **fig. S3, S4**.

Contrasting neural responses to a frequency-modulated sound and its time-reversed rendition provides qualitative insight into the transformation of temporal information in AC. In the IC, the two sounds are represented by neural activity sequences that involve the same neurons and mirror each other in time (**Fig. 2C**). In the AC, the symmetry is lost and each sound is encoded by sequences involving different neurons (**Fig. 2C**). This observation suggests that temporal information is required in IC but dispensable in AC to discriminate between these sounds. To quantify this more generally, we trained linear decoders to identify each sound from neural population responses either based on the full temporal code, obtained by concatenating all time bins from all neurons, or based on a time-independent rate code, obtained by time-averaging each neuron's response over 750ms (**Fig. 2D**). In subcortical areas, classifier accuracy based on rate coding was much lower than based on temporal coding. By contrast, in the cortex, rate coding accuracy increased such that the temporal and rate codes carried almost the same information (**Fig. 2E-F**). This result holds when the numbers of neurons are matched across datasets (**Fig S5A**). In line with our hypothesis, this suggests that sound representations are transformed from the subcortical to the cortical stage such that temporal information is made available to temporally-blind learning mechanisms by reformatting into a time-independent rate code.

The change in decoding performance in AC could either result from a change of the representation as suggested in **Fig. 2C**, or from a change in the signal-to-noise ratio which varies across datasets (**fig. S5B**). To rule out the latter possibility, we quantified the similarity between population vectors evoked by a given sound pair using a numerically and analytically validated noise-corrected version of the Pearson correlation (*30*) (**fig. S5C,** see **Supplementary text** for mathematical derivations). Applying this method to all sound pairs, we constructed representational similarity analysis (RSA) matrices that summarize the relations between sound representations in the neural space based either on the rate or on the temporal code (**Fig. 2G, fig. S6A**). We first observed that, subcortically, the mean similarity between representations of different sounds (mean of RSA matrix) is higher for the temporal than for the rate code, but that this difference becomes minimal in AC (**Fig. 2H,I**). Second, the difference of structure between RSA matrices for rate and temporal codes is also reduced in AC (**Fig. 2J**). These results hold in all subfields of AC (**Fig. 2H**) and are robust to the number of neurons included in the analysis (**fig. S5D**). Hence, temporal and rate representations of sounds in AC become highly similar in their structure, explaining why decoders perform almost equally with both representations. Notably, this transformation is non-monotonic, with a strong increase of representation similarity in TH (**Fig. 2H**), indicative of a dense representation in this area (**fig. S5E,F**).

To test if the convergence between temporal and rate codes in AC is the consequence of the known decrease of temporal resolution in the cortex (*5*, *31*), we decomposed neural population activity using Fourier analysis and measured classifier accuracy at each specific timescale (**Fig. 2K, S5G**). This analysis showed that relevant temporal resolution is preserved on the time scales considered in our study. In all datasets, classifier accuracy is maximal at ~1.5 Hz resolution (**fig. S5G).** Thus all datasets contain temporally structured neural activity that is sufficient to identify the relevant temporal cues in our sounds. Moreover, accumulating information from low to high temporal resolutions shows a saturation of classifier accuracy at around 3 Hz for all datasets (**Fig. 2K**). Therefore all temporal information needed to discriminate our sounds is available below 3 Hz, which is much lower than the putative 30 Hz cutoff for temporal resolution in AC (*5*), with information at faster timescales being redundant. Only in AC did the mean firing rate information (0 Hz) reach a level similar to that of the full cumulative temporal information (**Fig. 2K**). Therefore, in AC, all information, including temporal information, is also accessible from the identity of active neurons. Accordingly, each sound is represented by small sets of highly active neurons in AC as shown by the high population sparseness (**fig. S5E**) (*32*). This can be explained by increased specificity of AC neurons' responses to particular sounds, i.e. to a particular set of spectro-temporal features, as measured by lifetime sparseness (**fig. S5F**). Interestingly, the level of tuning to simple, individual temporal features (e.g. frequency modulation direction) was stable from IC to AC (**fig. S7**), consistent with previous reports (*33*, *34*). Therefore, the cortical transformation of sound representations does not correspond to a sharpening of particular tuning properties but to the emergence of more complex tuning properties (*35*).

Time-independent representations determine learning speed under biologically-realistic learning rules

To quantify how this transformation affects sensory-motor learning, we used a feedforward neural network model which simulates discrimination learning in an auditory Go/NoGo task based on reinforcement learning principles (*36*). We upgraded this model by complementing its Hebbian synaptic learning rules with an eligibility trace (*37*, *38*). This mechanism flags active synapses with a signal that decays over ~1 second, allowing even delayed post-synaptic activity driven by the reward to gate plasticity based on presynaptic input. We parameterized the eligibility trace with synaptic plasticity measurements in the striatum (*39*), the key site of auditory reinforcement learning (*40*). (**Fig. 3A**). We trained this biologically realistic model to discriminate between the population responses to pairs of sounds taken from the AC, TH or IC datasets and measured the duration of discrimination learning for sound pairs covering a broad range of representation similarities for temporal and rate codes (**Fig. 3A**). We observed that similarity based on time-independent rate influences duration of discrimination learning in the model, but not temporal similarity (**Fig. 3B)**. Moreover, learning duration rises in a steep and non-linear manner for high rate-coded similarity (**Fig. 3C**). This observation recapitulates our optogenetic experiments (**Fig. 1**), which showed that purely temporal sequences with maximal rate representation similarity are hard to learn. The model also provides an explanation for the long standing observation that AC's involvement in sound discrimination depends on the particular pair of sounds, with subcortical (likely thalamo-striatal) pathways sufficing for certain sounds (*26–28*) (**Fig. 3D**). Our unbiased noise-corrected measurements of population representation similarities (**Fig. 2**) show that sounds differing only in frequency, including simple pure tones, have a correlation below 0.75 for their rate representations at all stages of the auditory system (**Fig. 3E, S6B-D**). For this range of correlation values, the model shows that learning occurs quickly and the impact of representation similarity on learning speed is marginal. Hence, our model predicts similar learning rates based on thalamic or cortical representations (**Fig. 3E**). Contrariwise, rate representations of time-symmetric frequency modulations are highly correlated subcortically (>0.9), and clearly less in the cortex (0.74) (**Fig. 3F**, similar results for amplitude modulations, **fig. S6E-G**). Based on these values, the non-linear relationship between rate-based correlation and learning speed in our model predicts a ~3-fold decrease in learning duration with cortical compared to subcortical representations (**Fig. 3F**). This is in line with the observation that pre-training AC ablation does not affect discrimination learning between distant pure tones but prolongs it for time-reversed frequency sweeps (*27*). Similarly, the strong impact of post-training AC inactivation on discrimination of temporal features, but not of pure tones (*26–28*, *41*) can be explained if one postulates that learning speed, determined by representation similarity, determines which auditory system stage is recruited for the task.

Time-independent representations emerge in artificial neuronal networks performing perceptual categorizations

These results suggest that auditory representations in cortex are well-suited to the downstream machinery for learning sensory-motor associations which is blind to temporal structure. We wondered whether this adaptation also reflects more general constraints on the cortical code. To explore this, we analyzed representations in convolutional neural networks (CNNs)

categorizing key features of the stimuli presented to our mice: the frequency and intensity range, and the type of frequency and amplitude modulations present in the sounds (**Fig. 4A, S8A**). This analysis showed a decorrelation of rate representations, converging with temporal representations in deep layers of the trained, but not untrained, networks (**Fig. 4A, S8B**). Typical CNNs implement pooling mechanisms which reduce the precision of sensory receptive fields in deeper layers (*42*). In our CNNs, the rate and temporal convergence could be due to this shrinkage for the temporal dimension and the associated loss of temporal resolution, not observed in cortex. However, we observed thattemporal and rate representations still converge in CNNs without temporal shrinking (**Fig. 4B**). The main effect of temporal shrinking in our simulations was to accelerate learning while preserving performance (**Fig. 4B, S8A**). In addition, representations of the categorization network qualitatively reproduced all aspects of the evolution of rate and temporal codes observed in the auditory system (compare **Fig. 4C,D** and **Fig. 2F,J**), including preserved temporal information in the deeper layers (compare **Fig. 4E, S8C** and **Fig. 2K, S5G**). These results extended to a previously published CNN trained to classify words and musical styles (*43*) (**fig. S8D**) and to networks trained to perform single sound identification in noise (**fig. S8E**). However, when we trained an auto-encoder network to perform sound compression and denoising without assigning sounds to specific labels, rate representations did not converge with temporal representations (**Fig. 4F, S8F,G**). Therefore, our CNN modeling strongly supports the view that the emergence of time-independent representations we observe in the cortex is driven by the computational constraints of classifying sounds into perceptual objects.

Discussion

The emergence of time-independent representations in AC is consistent with its proposed role in the encoding of auditory objects (*44*) such as words, vocalizations or musical notes with typical durations of ~100-1000ms, the timescale we considered in this study. Recording methods such as functional magnetic resonance or ultrasound imaging, which cannot access temporal representations, have identified representations of such complex sounds in the spatial domain (*43*, *45*). By showing that, at population scale, time-averaged activity below 1s recapitulates the temporal information present in AC, our results validate and extend these observations, but indicate that key temporal information encoded in subcortical auditory regions will be lost with these methods.

Information over longer time scales as in sentences or musical phrases is likely exploited by other processes such as persistent network dynamics. At shorter timescales, below a few tens of milliseconds, feedforward and recurrent motifs between neurons or over dendritic arborization form efficient computational architectures to detect temporal coincidence, extract local temporal features and eventually convert them in a firing rate information (*13*, *14*) (**fig. S2**). Accordingly, a previous study showed that rats can discriminate short temporal deviations in the synchrony of stimulation at two different cortical locations (*46*). This may be based on a conversion of fast temporal information into a firing rate code via synaptic interactions (**fig. S2**) or on spike-timing-dependent plasticity mechanisms efficient only on short timescales, such as those proposed to underlie learning of behavioral sequences after temporal compression in the hippocampus (*47*). By contrast, our results show that temporal information at longer

time-scales cannot be utilized by sensory-motor learning rules unless reformatted into a time-independent code. We expect that the knowledge of this constraint will be key to re-encode perception at the cortical level e.g. in the context of the development of cortical implants.

Interestingly, CNNs, like the ones used in our study, rely on local linear-non-linear computations along the time and frequency dimensions that iteratively detect local spectrotemporal features. The networks we trained demonstrate that such local computations, which could be implemented by local circuit motifs in the auditory system, are sufficient to make temporal information accessible through a mean firing rate code (*48*). However, we observed differences between transformations implemented in CNNs and in the auditory system. In particular, CNNs monotonically decorrelate sound representations, while, in the auditory system, the transformation from IC to AC is non-monotonic (**Fig. 2**). This dense, correlated, representations observed in TH may allow to increase representation robustness before dilution in the cortical circuit (*49*, *50*).

It is remarkable that neural temporal information is largely preserved in the AC despite the transfer of temporal information to firing rate representations (**Fig. 2**). This likely extends to non-auditory temporal features such as visual motion or vibrotactile textures, as suggested by the existence of visual cortex neurons whose firing rate is selective to motion direction (*1*) or barrel cortex neurons selective to particular textures (*6*) while temporal information still exists in these regions (*2*, *7*). The coexistence of temporal and time-independent coding schemes could serve to combine object-like representations, useful for categorical decisions, with an explicit representation of the temporal details that are also perceived together with the object.

References and Notes:

1. G. T. Buračas, A. M. Zador, M. R. DeWeese, T. D. Albright, Efficient Discrimination of Temporal Patterns by Motion-Sensitive Neurons in Primate Visual Cortex. *Neuron*. **20**, 959–969 (1998).
2. D. A. Butts, C. Weng, J. Jin, C.-I. Yeh, N. A. Lesica, J.-M. Alonso, G. B. Stanley, Temporal precision in the neural code and the timescales of natural vision. *Nature*. **449**, 92–95 (2007).
3. K. M. M. Walker, J. K. Bizley, A. J. King, J. W. H. Schnupp, Multiplexed and Robust Representations of Sound Features in Auditory Cortex. *J. Neurosci.* **31**, 14565–14576 (2011).
4. I. Nelken, G. Chechik, T. D. Mrsic-Flogel, A. J. King, J. W. H. Schnupp, Encoding Stimulus Information by Spike Numbers and Mean Response Time in Primary Auditory Cortex. *J. Comput. Neurosci.* **19**, 199–221 (2005).
5. X. Wang, T. Lu, D. Bendor, E. Bartlett, Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience*. **154**, 294–303 (2008).
6. J. L. Chen, S. Carta, J. Soldado-Magraner, B. L. Schneider, F. Helmchen, Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex. *Nature*. **499**, 336–340 (2013).
7. K. H. Long, J. D. Lieber, S. J. Bensmaia, Texture is encoded in precise temporal spiking patterns in primate somatosensory cortex. *Nat. Commun.* **13**, 1311 (2022).
8. R. Shusterman, M. C. Smear, A. A. Koulakov, D. Rinberg, Precise olfactory responses tile the sniff cycle. *Nat. Neurosci.* **14**, 1039–1044 (2011).
9. R. W. Friedrich, G. Laurent, Dynamic optimization of odor representations by slow temporal

patterning of mitral cell activity. *Science*. **291**, 889–94 (2001).
10. P. J. Drew, L. F. Abbott, Extending the effects of spike-timing-dependent plasticity to behavioral timescales. *Proc. Natl. Acad. Sci.* **103**, 8876–8881 (2006).
11. R. Gutig, H. Sompolinsky, The tempotron: a neuron that learns spike timing-based decisions. *Nat Neurosci*. **9**, 420–8 (2006).
12. J. Gjorgjieva, C. Clopath, J. Audet, J. P. Pfister, A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *Proc Natl Acad Sci U A*. **108**, 19383–8 (2011).
13. T. Branco, B. A. Clark, M. Häusser, Dendritic Discrimination of Temporal Input Sequences in Cortical Neurons. *Science*. **329**, 1671–1675 (2010).
14. B. Sivyer, S. R. Williams, Direction selectivity is computed by active dendritic integration in retinal ganglion cells. *Nat. Neurosci.* **16**, 1848–1856 (2013).
15. H. Markram, J. Lubke, M. Frotscher, B. Sakmann, Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*. **275**, 213–5 (1997).
16. D. O. Hebb, *The organization of behavior; a neuropsychological theory* (Wiley, New York, 1949), *A Wiley book in clinical psychology*.
17. E. L. Bienenstock, L. N. Cooper, P. W. Munro, Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci*. **2**, 32–48 (1982).
18. T. Toyoizumi, J. P. Pfister, K. Aihara, W. Gerstner, Generalized Bienenstock-Cooper-Munro rule for spiking neurons that maximizes information transmission. *Proc Natl Acad Sci U A*. **102**, 5239–44 (2005).
19. M. Graupner, P. Wallisch, S. Ostojic, Natural Firing Patterns Imply Low Sensitivity of Synaptic Plasticity to Spike Timing Compared with Firing Rate. *J. Neurosci.* **36**, 11238–11258 (2016).
20. E. Salinas, A. Hernández, A. Zainos, R. Romo, Periodicity and Firing Rate As Candidate Neural Codes for the Frequency of Vibrotactile Stimuli. *J. Neurosci.* **20**, 5503–5515 (2000).
21. I. Nelken, Y. Rotman, O. B. Yosef, Responses of auditory-cortex neurons to structural features of natural sounds. *Nature*. **397**, 154–157 (1999).
22. K. Saberi, D. R. Perrott, Cognitive restoration of reversed speech. *Nature*. **398**, 760–760 (1999).
23. K. W. Berger, Some Factors in the Recognition of Timbre. *J. Acoust. Soc. Am.* **36**, 1888–1891 (1964).
24. G. C. Stecker, E. R. Hafter, An effect of temporal asymmetry on loudness. *J. Acoust. Soc. Am.* **107**, 3358–3368 (2000).
25. A. Verdier, N. Dominique, D. Groussard, A. Aldanondo, B. Bathellier, S. Bagur, Enhanced perceptual task performance without deprivation in mice using medial forebrain bundle stimulation. *Cell Rep. Methods*, 100355 (2022).
26. S. Ceballo, Z. Piwkowska, J. Bourg, A. Daret, B. Bathellier, Targeted Cortical Manipulation of Auditory Perception. *Neuron*. **104**, 1168-1179.e5 (2019).
27. F. W. Ohl, W. Wetzel, T. Wagner, A. Rech, H. Scheich, Bilateral Ablation of Auditory Cortex in Mongolian Gerbil Affects Discrimination of Frequency Modulated Tones but not of Pure Tones. *Learn. Mem.* **6**, 347–362 (1999).
28. T. Dalmay, E. Abs, R. B. Poorthuis, J. Hartung, D.-L. Pu, S. Onasch, Y. R. Lozano, J. Signoret-Genest, P. Tovote, J. Gjorgjieva, J. J. Letzkus, A Critical Role for Neocortical Processing of Threat Memory. *Neuron*. **104**, 1180-1194.e7 (2019).
29. T. Deneux, A. Kempf, A. Daret, E. Ponsot, B. Bathellier, Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nat. Commun.* **7**, 12682 (2016).
30. C. Spearman, The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **15**, 72 (1904).
31. M. M. Asokan, R. S. Williamson, K. E. Hancock, D. B. Polley, Inverted central auditory hierarchies for encoding local intervals and global temporal patterns. *Curr. Biol.* **31**, 1762-1770.e4 (2021).
32. B. Willmore, D. J. Tolhurst, Characterizing the sparseness of neural codes. *Netw. Bristol Engl.* **12**, 255–270 (2001).
33. C. Kopp-Scheinpflug, J. L. Sinclair, J. F. Linden, When Sound Stops: Offset Responses in the Auditory System. *Trends Neurosci.* **41**, 712–728 (2018).

34. A. J. King, I. Nelken, Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat. Neurosci.* **12**, 698–701 (2009).
35. R. C. deCharms, D. T. Blake, M. M. Merzenich, Optimizing Sound Features for Cortical Neurons. *Science*. **280**, 1439–1444 (1998).
36. B. Bathellier, S. P. Tee, C. Hrovat, S. Rumpel, A multiplicative reinforcement learning model capturing learning dynamics and interindividual variability in mice. *Proc. Natl. Acad. Sci.* **110**, 19950–19955 (2013).
37. W.-X. Pan, R. Schmidt, J. R. Wickens, B. I. Hyland, Dopamine Cells Respond to Predicted Events during Classical Conditioning: Evidence for Eligibility Traces in the Reward-Learning Network. *J. Neurosci.* **25**, 6235–6242 (2005).
38. W. Gerstner, M. Lehmann, V. Liakoni, D. Corneil, J. Brea, Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules. *Front. Neural Circuits*. **12** (2018) (available at https://www.frontiersin.org/articles/10.3389/fncir.2018.00053).
39. S. Yagishita, A. Hayashi-Takagi, G. C. R. Ellis-Davies, H. Urakubo, S. Ishii, H. Kasai, A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*. **345**, 1616–1620 (2014).
40. P. Znamenskiy, A. M. Zador, Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature*. **497**, 482–485 (2013).
41. H. Li, J. Wang, G. Liu, J. Xu, W. Huang, C. Song, D. Wang, H. W. Tao, L. I. Zhang, F. Liang, Phasic Off responses of auditory cortex underlie perception of sound duration. *Cell Rep.* **35** (2021), doi:10.1016/j.celrep.2021.109003.
42. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature*. **521**, 436–44 (2015).
43. A. J. E. Kell, D. L. K. Yamins, E. N. Shook, S. V. Norman-Haignere, J. H. McDermott, A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron*. **98**, 630-644.e16 (2018).
44. Z. A. Slonina, K. C. Poole, J. K. Bizley, What can we learn from inactivation studies? Lessons from auditory cortex. *Trends Neurosci.* **45**, 64–77 (2022).
45. A. Landemard, C. Bimbard, C. Demené, S. Shamma, S. Norman-Haignere, Y. Boubenec, Distinct higher-order representations of natural sounds in human and ferret auditory cortex. *eLife*. **10**, e65566 (2021).
46. Y. Yang, M. R. DeWeese, G. H. Otazu, A. M. Zador, Millisecond-scale differences in neural activity in auditory cortex can drive decisions. *Nat. Neurosci.* **11**, 1262–1263 (2008).
47. Z. Nádasdy, H. Hirase, A. Czurkó, J. Csicsvari, G. Buzsáki, Replay and Time Compression of Recurring Spike Sequences in the Hippocampus. *J. Neurosci.* **19**, 9497–9507 (1999).
48. R. Haddad, A. Lanjuin, L. Madisen, H. Zeng, V. N. Murthy, N. Uchida, Olfactory cortical neurons read out a relative time code in the olfactory bulb. *Nat. Neurosci.* **16**, 949–957 (2013).
49. S. Muscinelli, M. Wagner, A. Litwin-Kumar, "Optimal routing to cerebellum-like structures" (preprint, Neuroscience, 2022), , doi:10.1101/2022.02.10.480014.
50. G. Laurent, Olfactory network dynamics and the coding of multidimensional signals. *Nat Rev Neurosci*. **3**, 884–95 (2002).
51. S. Tang, Y. Zhang, Z. Li, M. Li, F. Liu, H. Jiang, T. S. Lee, Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *eLife*. **7**, e33370 (2018).
52. E. T. Rolls, M. J. Tovee, Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **73**, 713–726 (1995).
53. S. Ceballo, J. Bourg, A. Kempf, Z. Piwkowska, A. Daret, P. Pinson, T. Deneux, S. Rumpel, B. Bathellier, Cortical recruitment determines learning dynamics and strategy. *Nat Commun*. **10**, 1479 (2019).
54. J. Bourien, Y. Tang, C. Batrel, A. Huet, M. Lenoir, S. Ladrech, G. Desmadryl, R. Nouvian, J.-L. Puel, J. Wang, Contribution of auditory nerve fibers to compound action potential of the auditory nerve. *J. Neurophysiol.* **112**, 1025–1039 (2014).
55. R. Meddis, Auditory-nerve first-spike latency and auditory absolute threshold: A computer model. *J. Acoust. Soc. Am.* **119**, 406–417 (2006).
56. A. M. Taberner, M. C. Liberman, Response Properties of Single Auditory Nerve Fibers in the

Mouse. *J. Neurophysiol.* **93**, 557–569 (2005).
57. J.-F. Léger, E. A. Stern, A. Aertsen, D. Heck, Synaptic Integration in Rat Frontal Cortex Shaped by Network Activity. *J. Neurophysiol.* **93**, 281–293 (2005).
58. S. Romero, A. E. Hight, K. K. Clayton, J. Resnik, R. S. Williamson, K. E. Hancock, D. B. Polley, Cellular and Widefield Imaging of Sound Frequency Organization in Primary and Higher Order Fields of the Mouse Auditory Cortex. *Cereb. Cortex*. **30**, 1603–1622 (2020).
59. M. Nahmani, A. Erisir, VGluT2 immunochemistry identifies thalamocortical terminals in layer 4 of adult and developing visual cortex. *J. Comp. Neurol.* **484**, 458–473 (2005).
60. E. Yaksi, R. W. Friedrich, Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging. *Nat. Methods*. **3**, 377–383 (2006).

Acknowledgments: We thank Maia Brunstein of the Hearing Institute Bioimaging Core Facility of C2RT/C2RA for help in acquiring Airyscan images of thalamocortical boutons in the auditory cortex and Alexander Kell for help implementing the word and music classification network. We also thank Yves Boubenec, Yves Frégnac, Andrew King, Srdjan Ostojic and Christine Petit for their feedback on the manuscript. We acknowledge the support of the Fondation pour l'Audition to the Institut de l'Audition.

Funding:

Fondation pour l'Audition, FPA IDA02 (BB) and APA 2016-03 (BB)
European Research Council, ERC CoG 770841 DEEPEN, (BB)
Fondation pour la Recherche Médicale SPF202005011970 (SB)
European Union's Horizon 2020 research and innovation programme under grant agreement No 964568, project Hearlight (BB)

Author contributions: S.B., Ja.B., A.K. and B.B. conceived experiments, designed the study and interpreted data. S.B., Ja.B., A.K., T.T., J.S. and B.B. collected data and S.B., Ja.B., A.K., S.C. and B.B. performed data analysis. Je.B. and J.L.P conceived and implemented the cochlear model. B.B. and S.B. implemented the reinforcement learning model. S.B., K.B. and Y.G. implemented the deep learning models. S.B., Ja.B. and B.B. prepared figures. S.B. and B.B. wrote the manuscript. S.B and B.B. managed the project.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All datasets are freely available at 10.12751/g-node.sz67di, hosted by G-Node Infrastructure. Custom codes used in this study are freely available at 10.12751/g-node.sz67di, hosted by G-Node Infrastructure.

Supplementary Materials:

Materials and methods

Supplementary Text

Figs. S1 to S8

Tables S1 to S3

References (51-60)

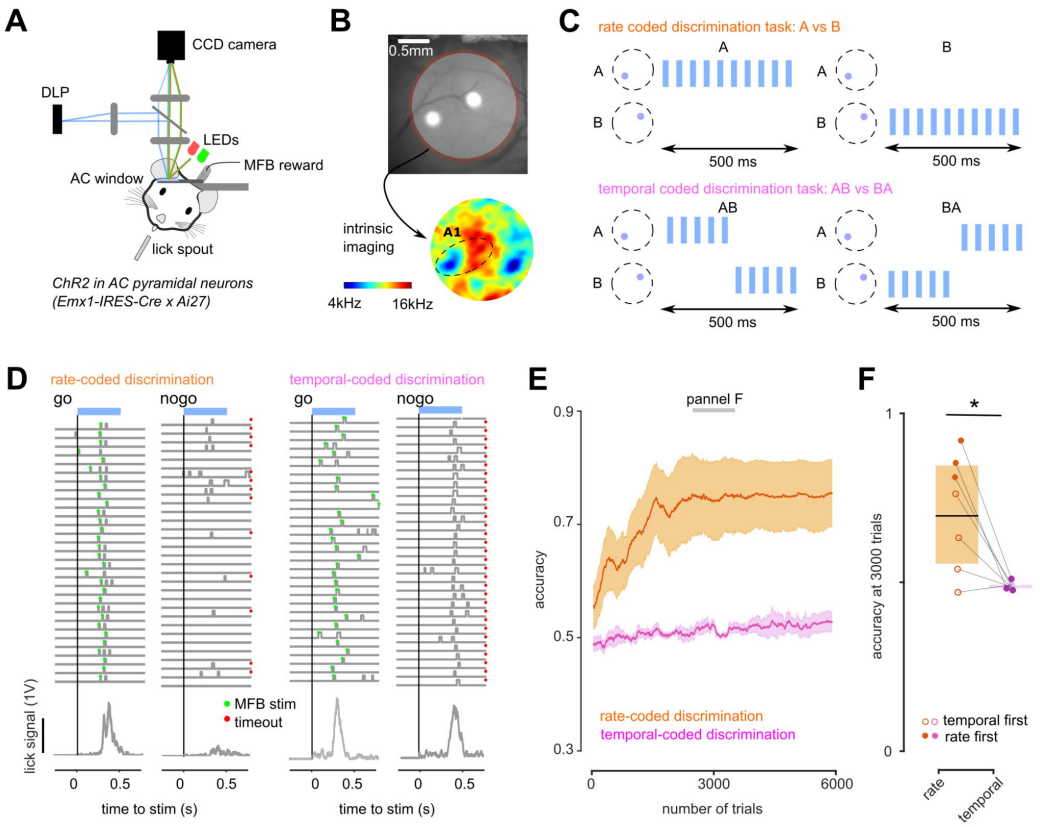

Fig. 1. Sensory-motor learning requires time-independent rate representations. A. Sketch of patterned optogenetic experiment in AC (MFB: medial forebrain bundle). **B.** Cortical window from an example mouse showing the location of the stimulation spots and the corresponding tonotopic map based on intrinsic imaging. **C.** Sketch of the optogenetic stimulation time courses for each discrimination task. **D.** Sample lick traces (top) and mean lick signal (bottom) for Go and NoGo trials in the rate-coded (left) and temporal-coded (right) discrimination tasks. **E.** Learning curves for all mice performing each task (n=7, error bars are sem). **F.** Accuracy at 3000 trials for all mice. (paired Wilcoxon test, p = 0.032, signed rank value = 27, n=7).

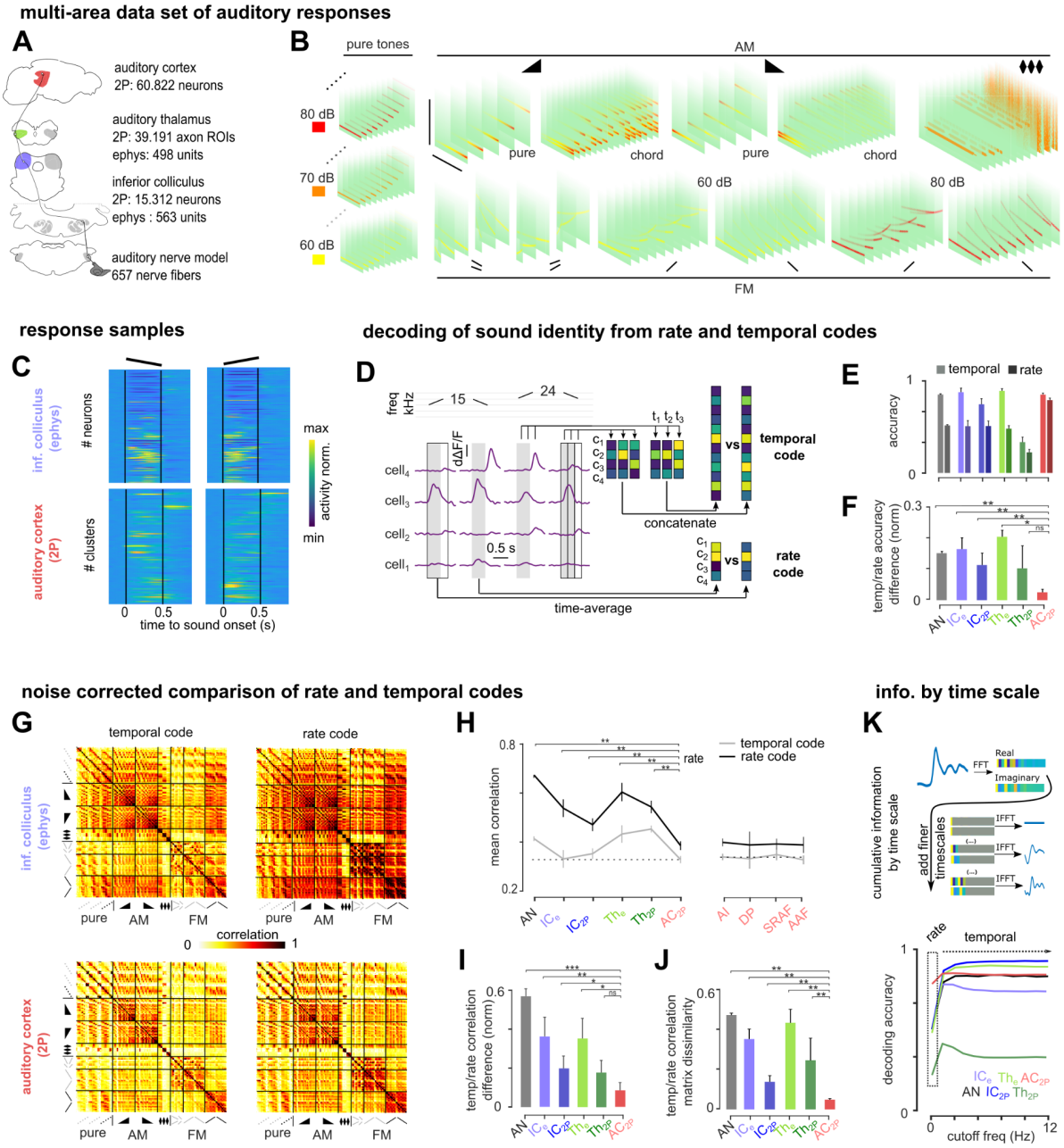

Fig. 2. Emergence of time-independent rate representations in the auditory cortex. A. Sketch of the auditory system and sample sizes at each level. **B.** Spectrograms of the sound set. **C.** Sample responses to up and down frequency sweeps from IC and AC neurons ordered by response amplitude. **D.** Responses of 4 AC neurons to different up and down frequency sweeps illustrating how temporal and rate codes are extracted. **E-F.** Mean sound decoding accuracy for temporal and rate codes in each area (**E**) and normalized difference between the two (**F**). (p-value for 100 bootstraps, error bars are S.D). **G.** Noise-corrected RSA matrices for all sound pairs for temporal (left) or rate (right) codes in IC and AC. **H.** Mean noise-corrected correlation by area. (p-value for 100 bootstraps comparing rate correlation of each region to AC, error bars are bootstrapped S.D). **I.** Normalized difference between mean noise-corrected correlation for temporal and rate codes. (p-value for 100 bootstraps, error bars are S.D). **J.** Noise-corrected dissimilarity between RSA matrix structure of temporal and rate codes. (p-value for 100 bootstraps, error bars are S.D). **K. Top:** Sketch illustrating the decomposition of

population responses by timescale and the concatenation of successive Fourier coefficients to accumulate increasingly fine timescales**. Bottom:** Mean decoding accuracy based on cumulative Fourier coefficients of neural responses. Full statistics are reported in **table S3.**

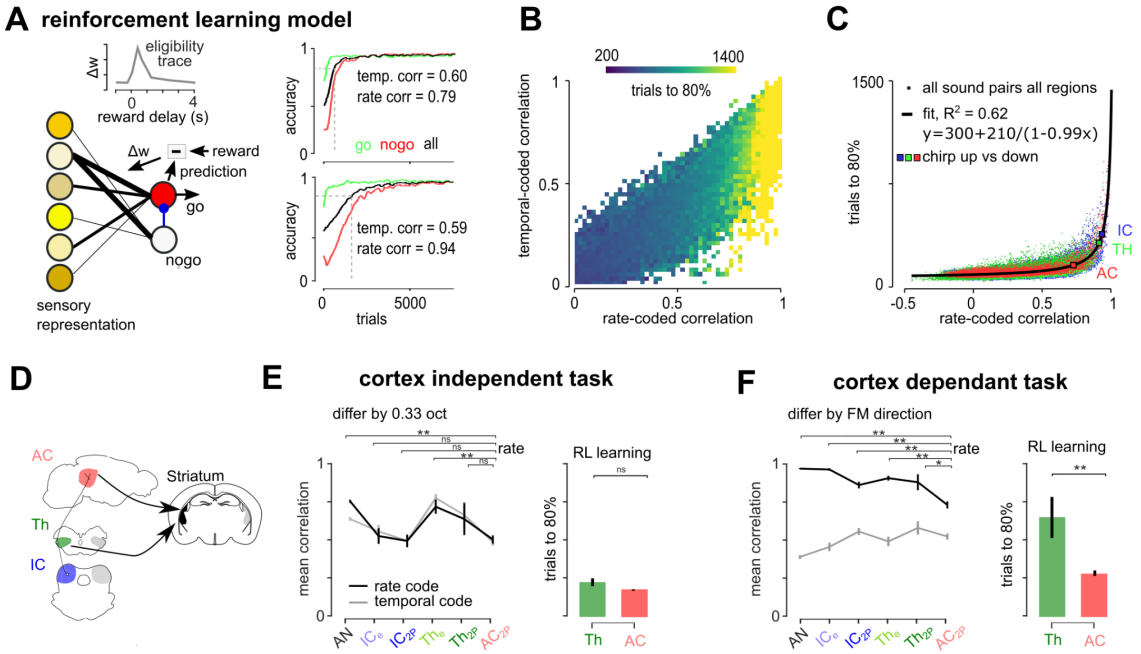

Fig. 3. A realistic learning model explains the role of time-independent representations in the acquisition of sound discriminations. A. Sketch of the reinforcement learning model (bottom left), eligibility trace dynamics (top left) and example learning curves for two recorded representations that have similar sequence code correlations but different rate code correlations. **B.** Heatmap of the number of trials needed to reach 80% accuracy at discriminating between a pair of sounds as a function of the temporal and rate code correlations between the representations of these sounds (averaged over all pairs of representations for all brain regions). **C.** Number of trials to 80% accuracy as a function of the correlations of rate representations. Large square dots show the mean correlation and learning time for time-symmetric frequency sweeps in IC, TH and AC and the black line shows the fit to data. **D.** Sketch showing the thalamic and cortical pathways for auditory learning. **E.** Mean noise-corrected correlation between sound pairs differing only by frequency (0.33 octave difference) and predicted learning rate for learning on a pure tone discrimination task based on thalamic (average of THe and TH2P) and cortical representations. **F.** Mean noise-corrected correlation between sound pairs differing only by the direction of the frequency sweep and predicted learning rate for learning on a pure tone discrimination task based on thalamic (average of THe and TH2P) and cortical representations. Full statistics are reported in **table S3.**

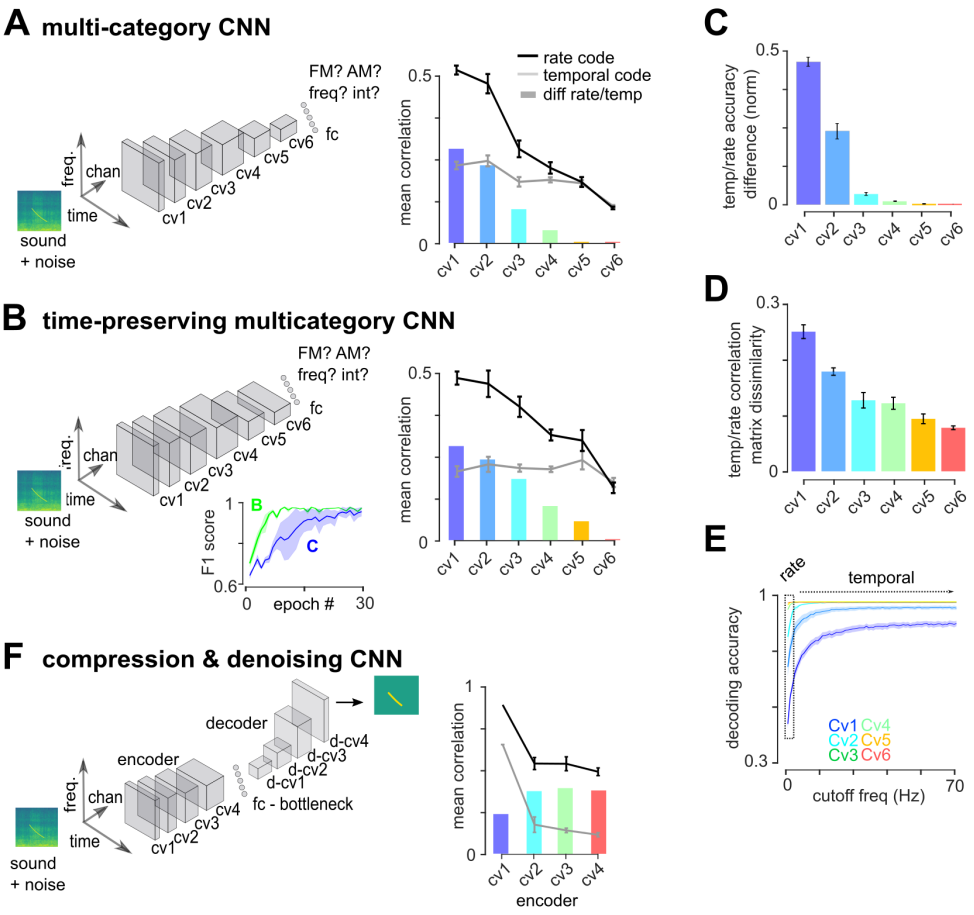

Fig. 4. Categorization deep networks implement a time-independent rate code in deeper layers. A-B. Left : Schematic of CNN architectures and target categories. **Right :** mean response correlations for the rate and temporal codes from RSA matrices constructed with the set of 140 sounds presented to mice (line) and difference between the two codes (bars). **A.** Multi-category CNN (n=8 networks). **B.** Multi-category CNN without shrinking of the temporal dimension (n=8 networks). Inset shows learning curves from training epochs for networks in A and B. **C-E.** All graphs refer to the time-preserving categorization CNN and reproduce analysis shown in **Fig. 2** for neural data : **C.** Normalized difference between mean sound decoding accuracy for temporal and rate codes. (error bars are sem over trained networks). **D.** Noise-corrected dissimilarity between RSA matrix structure of rate and temporal codes. **E.** Mean decoding accuracy based on cumulative Fourier coefficients of neural responses. **F.** Autoencoder CNN performing sound compression and denoising through a 20-unit bottleneck. (cv : convolution block, d-cv : deconvolution block - see methods for architecture details).