Cortical adaptation to sound reverberation

Aleksandar Z Ivanov[1], **Andrew J King**[1], **Ben DB Willmore**[1†], **Kerry MM Walker**[1†], **Nicol S Harper**[1†]

[1]Department of Physiology, Anatomy & Genetics, University of Oxford, United Kingdom

***For correspondence:**
aleksandar.ivanov@dpag.ox.ac.uk (FMS)

[†]These authors contributed equally to this work

Abstract In almost every natural environment, sounds are reflected by nearby objects, producing many delayed and distorted copies of the original sound, known as reverberation. Our brains usually cope well with reverberation, allowing us to recognize sound sources regardless of their environments. In contrast, reverberation can cause severe difficulties for speech recognition algorithms and hearing-impaired people. The present study examines how the auditory system copes with reverberation. We trained a linear model to recover a rich set of natural, anechoic sounds from their simulated reverberant counterparts. The model neurons achieved this by extending the inhibitory component of their receptive filters for more reverberant spaces, and did so in a frequency-dependent manner. These predicted effects were observed in the responses of auditory cortical neurons of ferrets in the same simulated reverberant environments. Together, these results suggest that auditory cortical neurons adapt to reverberation by adjusting their filtering properties in a manner consistent with dereverberation.

Introduction

Reverberations accompany almost all natural sounds that we encounter and are the reflections of sound off objects in the environment such as walls, furniture, trees, and the ground (***Huisman and Attenborough, 1991***; ***Sakai et al., 1998***). Compared to the original sound, these reflections are attenuated and distorted due to frequency-selective absorption and delayed due to increased path length (***Kuttruff, 2017***).

Reverberation can be useful, helping us judge room size, sound-source distance, and realism (***Shinn-Cunningham, 2000***; ***Trivedi et al., 2009***; ***Kolarik et al., 2021***). However, strong reverberation can impair sound-source localization (***Hartmann, 1982***; ***Shinn-Cunningham and Kawakyu, 2003***; ***Rakerd and Hartmann, 2005***; ***Shinn-Cunningham et al., 2005***) and segregation (***Culling et al., 1994***; ***Darwin and Hukin, 2000***), pitch discrimination (***Sayles and Winter, 2008***) and speech recognition (***Knudsen, 1929***; ***Nábělek et al., 1989***; ***Guediche et al., 2014***; ***Houtgast and Steeneken, 1985***). Notably, reverberation can be detrimental for people with hearing impairments, increasing tone detection thresholds and reducing intelligibility of consonants (***Humes et al., 1986***; ***Helfer and Wilber, 1990***). It can also impede the effectiveness of auditory prostheses such as hearing aids (***Schweitzer, 2003***; ***Qin and Oxenham, 2005***; ***Poissant et al., 2006***) and substantially reduces the performance of automatic speech recognition devices (***Yoshioka et al., 2012***; ***Kinoshita et al., 2016***).

The auditory system has mechanisms to help us cope with reverberation, to the extent that healthy listeners often only directly notice it when it is strong (in environments such as cathedrals). In the presence of mild-to-moderate reverberation, healthy listeners can continue to perform sound localization (***Hartmann, 1982***; ***Rakerd and Hartmann, 2005***) and speech and auditory object recognition tasks (***Houtgast and Steeneken, 1985***; ***Bradley, 1986***; ***Darwin and Hukin, 2000***; ***Culling et al., 2003***; ***Nielsen and Dau, 2010***). Since it is such a ubiquitous property of natural sounds, these findings highlight the importance, for both normal and impaired hearing, of understanding how the brain copes with reverberation (***Xia et al., 2018***).

What are the neurophysiological mechanisms that support listening in reverberant environments? Previous studies have examined subcortical processes that facilitate localization of reverberant sounds (***Yin, 1994***; ***Litovsky and Yin, 1998***; ***Fitzpatrick et al., 1999***; ***Spitzer et al., 2004***; ***Tollin et al., 2004***; ***Pecka et al., 2007***; ***Devore et al., 2009***; ***Kuwada et al., 2012***; ***Kim et al., 2015***; ***Brughera et al., 2020***), and how subcortical processing of synthetic periodic sounds is disrupted by reverberation (***Sayles and Winter, 2008***) and partially restored by compensatory mechanisms (***Slama and Delgutte, 2015***). Much less is known about the neural processing of speech and other complex natural sounds in the presence of reverberation. However, converging evidence from electrophysiological recordings in animals (***Rabinowitz et al., 2013***; ***Moore et al., 2013***; ***Mesgarani et al., 2014***) and from human EEG (***Khalighinejad et al., 2019***) and fMRI (***Kell and McDermott, 2019***) studies suggests that representations of sounds that are invariant to non-reverberant background noise emerge at the level of auditory cortex via neuronal adaptation to stimulus statistics (but see also ***Lohse et al. , 2020***). Auditory cortex may play a similar role in adaptation to reverberation. Indeed, speech and vocalization stimuli reconstructed from auditory cortical responses in awake ferrets more closely resemble their anechoic versions than the reverberant ones, even if the sounds were presented in reverberant environments (***Mesgarani et al., 2014***). Similar results have been found in humans using sound reconstructions from EEG measurements (***Fuglsang et al., 2017***). It remains unclear, however, whether the observed cortical invariance to reverberation can occur in the absence of top-down attention, and through what neural mechanisms this is achieved.

Here, we addressed these questions by using a model to predict what neural tuning properties would be useful for effective attenuation of reverberation (a normative “dereverberation model”). We then test these predictions using neural recordings in the auditory cortex of anesthetized ferrets. More specifically, we made reverberant versions of natural sounds in simulated rooms of different sizes. Next, we trained a linear model to retrieve the clean anechoic sounds from their reverberant versions. Our trained model provided specific predictions for how the brain may achieve this task: with increased reverberation, neurons should adapt so that they are inhibited by sound energy further into the past, and this should occur in a sound frequency-dependent manner. We observed these predicted effects in the responses of auditory cortical neurons to natural sounds presented in simulated reverberant rooms, and show that they arise from an adaptive process. These results suggest that auditory cortical neurons may support hearing performance in reverberant spaces by temporally extending the inhibitory component of their spectrotemporal receptive fields.

Results

Dereverberation model kernels show reverberation-dependent inhibitory fields

We trained a dereverberation model to estimate the spectrotemporal structure of anechoic sounds from reverberant versions of those sounds. The anechoic sounds comprised a rich 10-minute-long set of anechoic recordings of natural sound sources, including speech, textures (e.g. running water) and other environmental sounds (e.g. footsteps) (see Sound stimuli and virtual acoustic space). Reverberation in small (3.0 x 0.3 x 0.3m) and large (15 x 1.5 x 1.5m) tunnel-shaped rooms was simulated using the virtual acoustic space simulator Roomsim (***Campbell et al., 2005***) (Figure 1A). The simulation also modelled the acoustic properties of the head and outer ear by using a ferret head-related transfer function (HRTF, ***Schnupp et al. , 2001***). The dimensions of the smaller room made it less reverberant (reverberation time, RT_{10} = 130ms) than the larger room (RT_{10} = 430ms).

After the reverberant sounds were generated, they were converted to cochleagrams (Figure 1B). These spectrotemporal representations of the sound estimate the filtering and resulting representation of the sound by the auditory nerve (***Brown and Cooke, 1994***; ***Rahman et al., 2020***). Cochleagrams of an example sound clip presented in the anechoic, small and large room conditions are

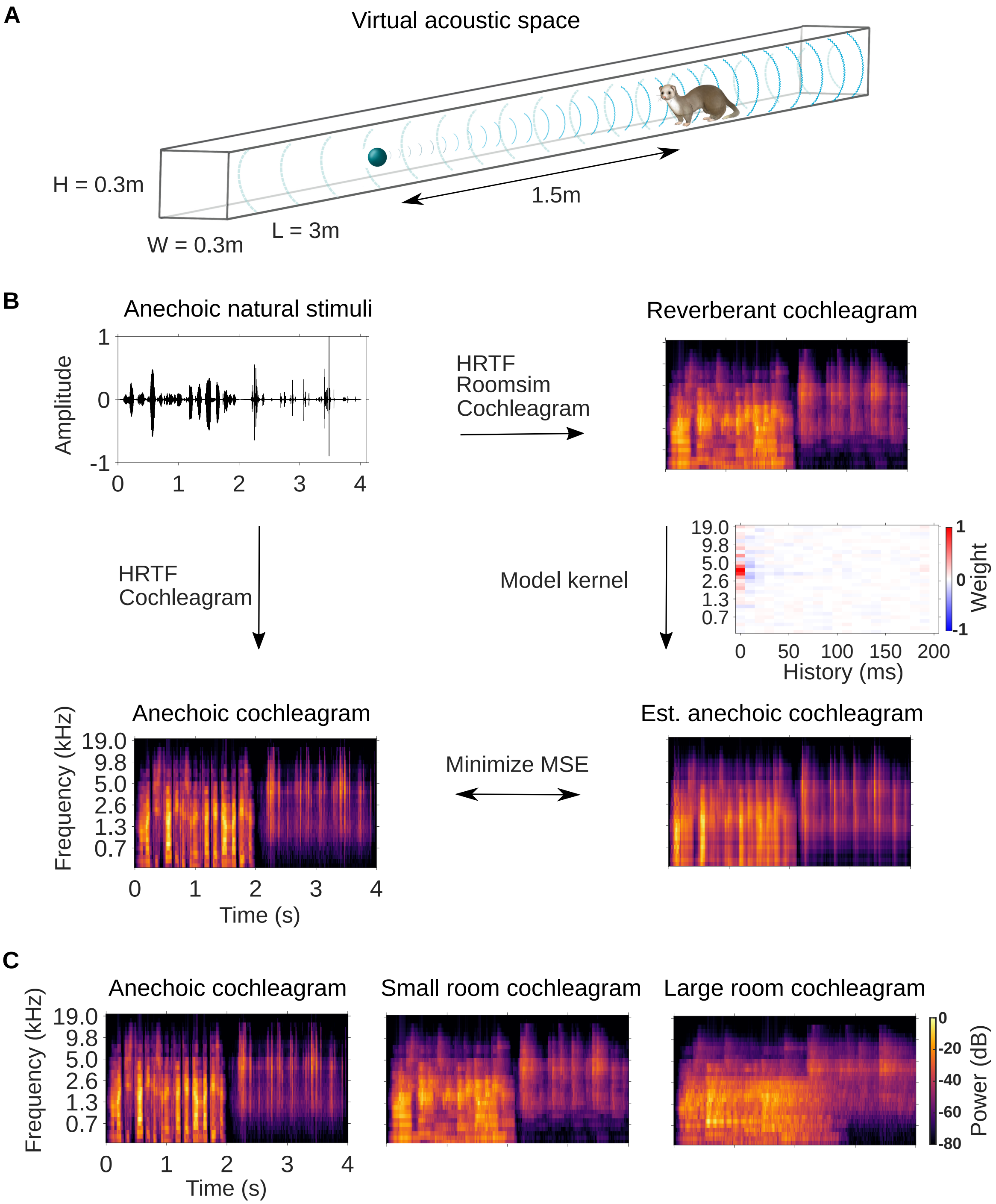

Figure 1. Dereverberation model.
A, Virtual acoustic space was used to simulate the sounds received by a ferret from a sound source in a reverberant room for diverse natural sounds. Schematic shows the simulated small room (length (L) = 3m, width (W) = 0.3m, height (H) = 0.3m) used in this study, and the position of the virtual ferret's head and the sound source (1.5m from the ferret head) within the room. We also used a medium (x2.5 size) and large room (x5). The acoustic filtering by a ferret's head and ears was simulated by a head-related transfer function (HRTF). **B**, Schematic of the dereverberation model. The waveform (top left panel) shows a 4s clip of our anechoic recordings of natural sounds. For a given room, simulated room reverberation and ferret HRTF filtering were applied to the anechoic sound using Roomsim (***Campbell et al., 2005***), and the resulting sound was then filtered using a model cochlea to produce a reverberant cochleagram (top right panel). A cochleagram of the anechoic sound was also produced (bottom left panel). For each room, a linear model was fitted to estimate the anechoic cochleagram from the reverberant cochleagram for diverse natural sounds. Each of the 30 kernels in the model was used to estimate one frequency band of the anechoic sound. One such model kernel is shown (middle right panel). Generating the estimated anechoic cochleagram (bottom right panel) involved convolving each model kernel with the reverberant cochleagram, and the mean squared error (MSE) between this estimate and the anechoic cochleagram was minimized with respect to the weights composing the kernels. **C**, Sample cochleagrams of a 4s sound clip for the anechoic (left panel), small room (middle panel), and large room (right panel) reverberant conditions.

shown in Figure 1C.

We trained a dereverberation model to recover the anechoic cochleagram, using either the small or large room cochleagrams as an input (Figure 1B). The dereverberation model was comprised of a set of "dereverberation" kernels, one for each frequency in the anechoic cochleagram (see Model kernels). Each model kernel used the full reverberant cochleagram (up to 200ms in the past) to estimate the current power in the anechoic cochleagram within a single frequency band. This resulted in a set of positive and negative weights in each model kernel. Obtaining the estimated anechoic sounds involved convolution over time between the model kernels and the reverberant cochleagrams, and the model was trained to minimize the difference between this estimate and the original anechoic sound (Figure 1B). The model was trained separately to dereverberate the small and large room cochleagrams. For each room, on a held-out test set, the dereverberation model reduced the difference between the incoming reverberant cochleagram and the anechoic cochleagram (small room mean squared error reduction 26%; large room reduction 20%).

Three examples of model kernels are shown in Figure 2A for the large room and the small room, with the anechoic frequency band they estimate indicated at the top. For each model kernel, the excitatory (red) and inhibitory (blue) weights represent spectrotemporal features in the reverberant cochleagrams that are associated with increased or decreased power in the specified frequency band of the anechoic cochleagram, respectively. The majority of the excitatory and inhibitory weights appear localized around a particular frequency, resembling the frequency tuning seen in auditory cortical neurons (***Bizley et al., 2005***). This is expected in our dereverberation model since each kernel aims to estimate the power in a given frequency band of the anechoic cochleagram.

The model kernels had temporally asymmetric structure, where strongest excitatory weights tended to occur first (Figure 2A), followed soon after by a longer inhibitory field. These excitatory and inhibitory timings are readily apparent when we plot the frequency-averaged positive and negative kernel weights (Figure 2B), and are a common feature across all kernels in the model (Figure 2-Figure supplement 1A, and Figure 2-Figure supplement 2A). This pattern has been commonly observed in the spectrotemporal receptive fields (STRFs) of auditory cortical neurons (***deCharms et al., 1998***; ***Linden et al., 2003***; ***Harper et al., 2016***; ***Rahman et al., 2019***), so our model qualitatively reproduces the basic frequency tuning and temporal characteristics of these auditory cortical neurons.

Importantly, we can compare the model kernels for the large room with those for small room. The inhibitory components of the large-room kernels tended to be delayed and longer in duration, relative to the small-room kernels (Figure 2B). In contrast, the temporal profile of the excitatory components was similar for the small and large rooms. We predicted that a comparable shift in inhibitory filtering could play a role in reverberation adaptation in auditory cortical neurons.

Auditory cortical neurons have reverberation-dependent inhibitory fields

To test the predictions of our dereverberation model in vivo, we presented to anesthetized ferrets an 80 sec subset of the natural sounds in the simulated small and large reverberant rooms (see Sound stimuli and virtual acoustic space). We did this while recording the spiking activity of neurons in the auditory cortex using Neuropixels high-density extracellular microelectrodes (***Jun et al., 2017***) (see Surgical procedure). Stimuli were presented as 40 sec blocks, in which all sounds were in the same reverberant room condition. This allowed neurons to adapt to the reverberation acoustics of the room. We recorded the responses of 2,244 auditory cortical units. Of these, the 696 units (160 single units, 23%) which were responsive to the stimuli were used for further analysis (see Spike sorting).

We estimated the filtering properties of each unit by fitting a separate STRFs to the neuronal responses for each reverberant condition. Neuronal STRFs are linear kernels mapping the cochleagram of the sound stimulus to the time-varying firing rate of the neuron (***Theunissen et al., 2001***). The positive regions of an STRF represent sound features whose level is positively correlated with the the neuron's spike rate, providing the "excitatory" part of the receptive field. Similarly, negative regions of the STRF indicate features whose level is negatively correlated with the neural unit's spike rate, providing the "inhibitory" receptive field.

Examples of typical neuronal STRFs are shown in Figure 2C, and these can be compared to the model kernel properties of our dereverberation model above (Figure 2A). As mentioned above, the model kernels show some similarity to the STRFs typically reported for auditory cortical neurons (***deCharms et al., 1998***; ***Linden et al., 2003***; ***Harper et al., 2016***; ***Rahman et al., 2019***). Likewise, the model kernels show similarity to the STRFs we present here, including having frequency tuning, early excitatory receptive fields and delayed inhibitory receptive fields (Figure 2D). These consistencies between the general features of our model and neurophysiological responses validated our use of this normative approach to capture neural response properties. We next examined if the model could predict neural adaptation to different reverberant conditions.

The important prediction we observed in the model was that the inhibitory fields tended to be more delayed and of longer duration in the large-room kernels versus the small-room kernels, whereas the excitatory field remained unchanged. Strikingly, we observed the same pattern in the neuronal STRFs in Figure 2D. This observation also held across different frequency channels in both the model and the data. (Figure 2-Figure supplement 1, Figure 2-Figure supplement 2).

Similar effects of reverberation on the inhibitory fields of model kernels and auditory cortical neurons

Since both the dereverberation model and the neuronal STRFs had structure which varied according to the reverberation condition, we sought to investigate these effects quantitatively. We used two metrics to estimate the temporal dynamics of the inhibitory (and excitatory) components of the model kernels and neuronal STRFs: Center of mass (COM) and peak time (PT) (see Quantification of the temporal effects in model kernels and neuronal STRFs). The COM measured the average temporal delay of the inhibitory (COM^-) or excitatory (COM^+) components of the model kernels/neuronal STRFs (Figure 2B,D). The PT is the time at which the maximal inhibition (PT^-) or excitation (PT^+) occurred.

For each anechoic frequency channel in the dereverberation model, we calculated the difference between the $COM^-$for the kernels in the large room and small room conditions, providing 30 COM^-differences (1 for each channel), and did the same for the COM^+. We plotted the distribution of these differences as histograms in Figure 3A. Similarly, a histogram of the COMdifference between the neuronal STRFs in the large and small room conditions is plotted for 696 cortical units in Figure 3B. We found that the COM^+did not differ significantly between the small and large rooms, either for model kernels (median COM^+difference = 0.97ms, Wilcoxon signed-rank test, p = 0.066) or neuronal STRFs (median COM^+difference = 0.32ms, p = 0.39). In contrast, the COM^-showed clear dependence on room size. The inhibitory centers of mass were higher in the larger room for both the model kernels (median COM^-difference = 7.9ms, p = 1.9x10^{-6}), and neuronal STRFs (median COM^-difference = 9.3ms, p = 1.5x10^{-66}).

The results of our analysis of PTwere largely consistent with our COMfindings (Figure 3C,D). The peak time of the excitatory component (PT^+) of model kernels did not differ between the small and large room (median PT^+difference = 0.0ms, p = 1.0), while PT^+in the neural data showed a small but statistically significant increase in the large room (median PT^+difference = 0.0ms, p = 0.014). The peak time of the inhibitory component, on the other hand, occurred much later in the large room, in both the model kernels (median PT^-difference = 10ms, p = 3.7x10^{-3}) and neuronal STRFs (median PT^-difference = 10ms, p = 1.5x10^{-39}). In general, there was more spread in the COM and PT in the neuronal data comparing to the dereverberation model. This is likely because, unlike in the model, which was focused purely on dereverberation, the auditory cortex subserves multiple functions and a diversity of STRF spans is useful for other purposes (e.g. prediction, ***Singer et al. , 2018***). Despite this, it is notable that the median $COM$and PTdifferences of the dereverberation model were of similar magnitude to those of the real data.

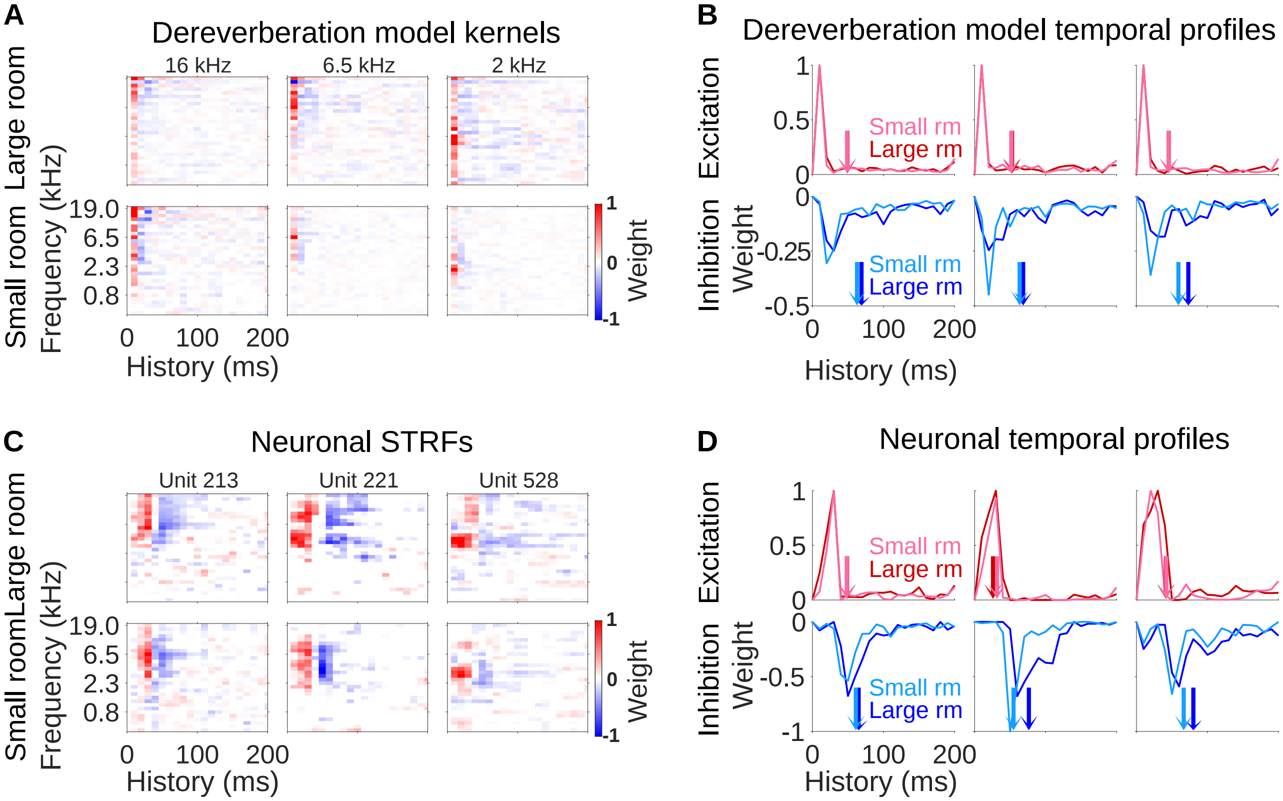

Figure 2. Comparison of dereverberation model kernels and neuronal receptive fields from auditory cortex.
A, Example model kernels resulting from the dereverberation model. Three example model kernels are shown, after training on the large (top row) or small (bottom row) room reverberation. The frequency channel which the model kernel is trained to estimate is indicated above each kernel. The color scale represents the weights for each frequency (y-axis) and time (x-axis). Red indicates positive weights (i.e. excitation), and blue indicates negative weights (i.e. inhibition; color bar right). **B**, Each plot in the top row shows the temporal profile of the excitatory kernel weights for the corresponding example model kernels shown in A. Excitatory temporal profiles were calculated by positively rectifying the kernel and averaging over frequency (the y-axis), and were calculated separately for the small (pink) and large (red) rooms. The center of mass of the excitation, COM^{+}, is indicated by the vertical arrows, which follow the same color scheme. The bottom row plots the inhibitory temporal profiles for the small (cyan) and large (blue) rooms. Inhibitory temporal profiles were calculated by negatively rectifying the kernel and averaging over frequency. The COM^{-} is indicated by the colored arrows. **C**, Spectrotemporal receptive fields (STRFs) of three example units recorded in ferret auditory cortex, measured for responses to natural sounds in the large room (top row) or small room (bottom row), plotted as for model kernels in A. **D** Temporal profiles of the STRFs for the three example units shown in C, plotted as for the model kernels in B.

Figure 2–Figure supplement 1. Model kernels and neuronal STRFs across frequency channels.

Figure 2–Figure supplement 2. Model and neuronal temporal profiles across frequency channels.

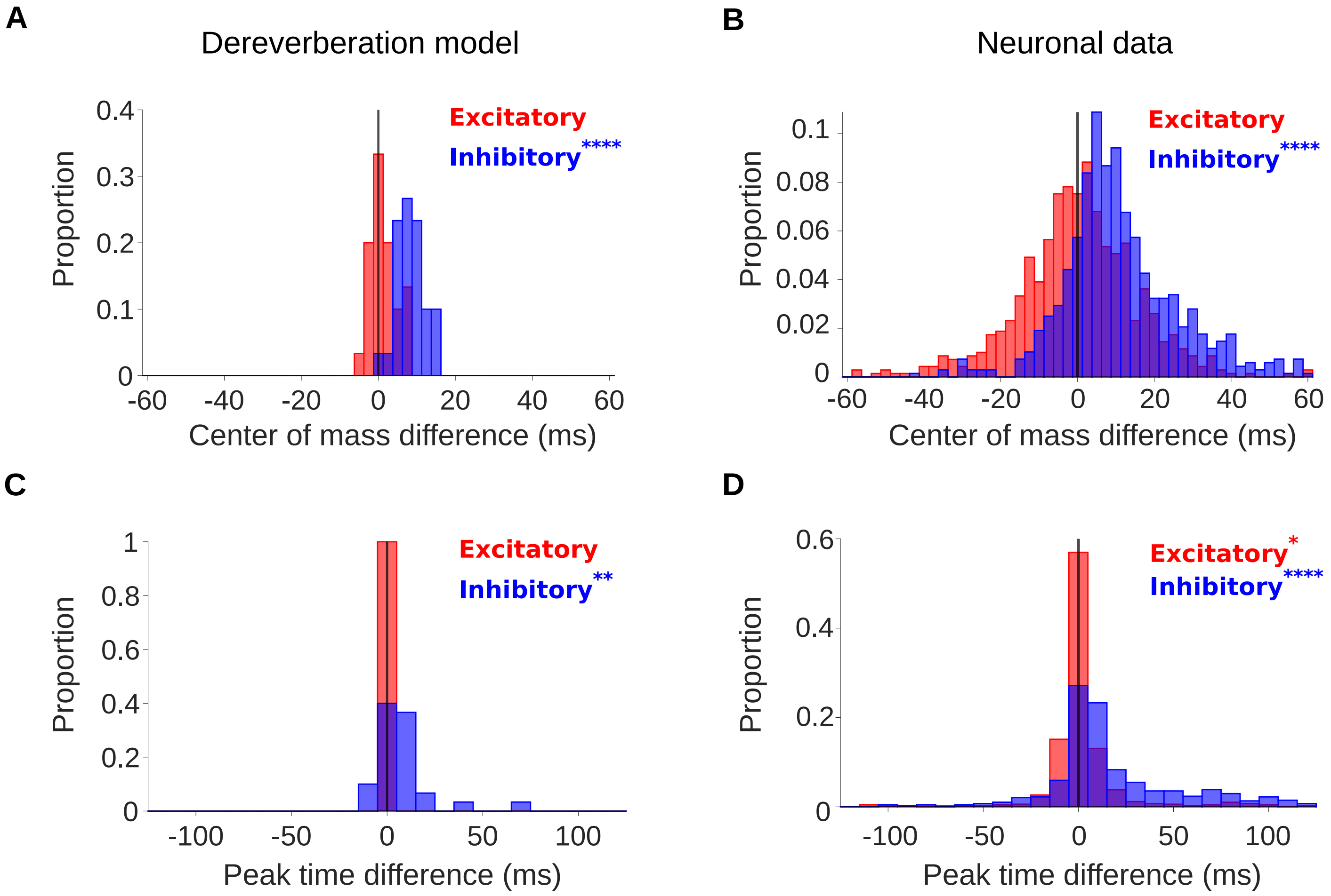

Figure 3. Differences in the temporal profiles for large and small rooms.
A, Histograms of the difference in center of mass of the temporal profiles (for the inhibitory field, COM^{-}, blue; excitatory field, COM^{+}, red) of dereverberation model kernels between the two different reverberant conditions (large - small room). The COM^{-} increased in the larger room with a median difference = 7.9ms; COM^{+} did not differ significantly between the rooms, median difference = 1.0ms. **B**, Center of mass differences, plotted as in A, but for the auditory cortical units. The COM^{-} increased in the larger room, median difference = 9.3ms; COM^{+} was not significantly different, median difference = 0.3ms. **C**, Histograms of the large - small room difference in peak time for the temporal profiles of the model kernels (inhibitory, PT^{-}, blue; excitatory, PT^{+}, red). The PT^{-} values were larger in the larger room, median difference = 10ms, whereas PT^{+} values were not significantly different, median difference = 0.0ms. **D**, Peak time differences for neuronal data, plotted as in C. The PT^{-} values increased in the larger room, median difference = 10ms, and PT^{+} showed a weakly significant change, but the median difference was 0ms. Asterisks indicate the significance of Wilcoxon signed-rank tests: $^{****}p < 0.0001, ^{**}p < 0.01, ^{*}p < 0.05$.

Figure 3–Figure supplement 1. A medium room condition shows intermediate center of mass and peak time values compared to the small and large room conditions.

Figure 3–Figure supplement 2. Simulated neurons suggest a role for adaptation in cortical dereverberation.

Figure 3–Figure supplement 2. Simulated neurons suggest a role for adaptation in cortical dereverberation (caption).

Figure 3–Figure supplement 3. Neural response to noise probe shows slower adaptation in the more reverberant condition.

Figure 3–Figure supplement 4. Adaptation to reverberation is confirmed using stimuli that switch between the small and large room.

As our stimulus set described above included only 2 reverberant rooms, it was not clear if the neurons treated these simulated rooms as two points along an ordered reverberation scale. To further examine whether the timing of the neuronal STRF inhibitory component scales with the amount of reverberation in our simulated room, we added a third “medium” sized room with the same relative proportions and absorption properties as the small and large rooms. We measured auditory cortical responses to this extended stimulus set in 2 ferrets (266 neural units).

The COM and PT measures of neuronal STRF dynamics were compared across the small, medium and large room conditions, and are shown in Figure 3-Figure supplement 1. As expected, there was little effect of room size on the timing of the excitatory STRF components (Figure 3-Figure supplement 1A,C). The COM^+ showed a weak but significant overall increase with room size (Kruskal-Wallis test; $\chi^2(2) = 6.4$, p = 0.042), but there was no effect of room size on the peak time of excitation, PT^+ ($\chi^2(2) = 1.4$, p = 0.50). In post-hoc pairwise comparisons, COM^+ only differed between the small and medium rooms (Fisher’s least significant differences; large-small: p = 0.21; large-medium: p = 0.21; medium-small: p = 0.012).

In contrast, as predicted, we found that the delay of the inhibitory STRF components increased with greater room reverberation. The COM^- was generally larger for larger rooms (Kruskal-Wallis test; $\chi^2(2) = 37$, p = 7.6x10^{-9}) (Figure 3-Figure supplement 1B). Post-hoc pairwise tests confirmed that COM^- differed between all three reverberant conditions (Fisher’s least significant differences; large-small: p = 1.3x10^{-9}; large-medium: p = 2.0x10^{-4}; medium-small: p = 0.019). The peak time of STRF inhibition, PT^-, also increased with room size across all 3 rooms ($\chi^2(2) = 27$, p = 1.6x10^{-6}; large-small: p = 2.7x10^{-7}; large-medium: p = 0.0024; medium-small: p = 0.036) (Figure 3-Figure supplement 1D).

Thus, as room size, and hence reverberation time, was increased, we observed an increase in the delay of inhibition in the tuning properties of auditory cortical neurons. This increase is consistent with a normative model of dereverberation, suggesting that the tuning properties of auditory cortical neurons may adapt in order to dereverberate incoming sounds.

Reverberation effects result from an adaptive neural process

In principle, there could be other reasons, unrelated to adaptation, why the temporal profile of the inhibitory field is delayed and broader in the more reverberant room. An important possibility is that differences in sound statistics between the reverberation conditions could result in different STRFs, even if the underlying neuronal tuning is unchanged. For example, the cochleagrams of more reverberant sounds are more temporally blurred (Figure 1C). This could lead to slower features in the neuronal STRFs for the larger room, purely due to systematic model fitting artefacts (***Christianson et al., 2008***). In combination with changing sound statistics, a non-adaptive static non-linearity in the neural system could produce apparent differences in neuronal tuning between the reverberation conditions (***Christianson et al., 2008***). Here, we perform several additional experiments and analyses to test whether the reverberation-dependent effects observed above are likely to result from a genuine adaptive process.

As a first test, for each recorded unit, we fitted a simulated linear-nonlinear-Poisson model neuron (***Schwartz et al., 2006***), composed of a single STRF (fitted to the combined small and large room stimuli) feeding into a non-linear output function (see subsection Simulated neuron), which, in turn, fed into a non-homogeneous Poisson process. Since this model did not have an adaptive component, we used it to assess whether our reverberation-dependent results could arise from fitting artefacts in a non-adaptive neuron. To do this, we presented the same stimuli to the simulated non-adaptive neurons as we did to the real neural responses and performed the same analyses. Hence, we fitted STRFs to the simulated neural responses separately for the large and small room conditions. We then extracted COM and PT parameters from the excitatory and inhibitory temporal profiles of these STRFs, and compared them to those of the measured cortical units. The simulated results are shown alongside the neural results in Figure 3-Figure supplement 2.

We asked whether the shift in inhibition observed in the dereverberation model and neural data was also present in this adaptation-free simulation. In the simulation, although the inhibitory COM^- was slightly larger for the more reverberant condition (Figure 3-Figure supplement 2B), the effect size for the simulated neurons (median COM^- difference = 0.90ms, p = 1.5x10^{-5}) was very small compared to that observed in the real neuronal data (median COM^- difference = 9.3ms, p = 1.5x10^{-66}, Figure 3-Figure supplement 2C). We directly compared the COM^- room differences between cortical units and their simulated counterparts (Figure 3-Figure supplement 2D), and found that the reverberation effects on COM^- were consistently larger in the neuronal data (median difference = 9.3ms, p = 3.9x10^{-35}). An analysis of the peak time of inhibitory STRF components for neural and simulated units was in agreement with the center of mass results (Figure 3-Figure supplement 2E-G). The simulation predicted a near zero shift in the peak inhibitory component between rooms (median PT^- difference = 0ms, p = 3.5x10^{-9}), and was unable to account for the 10ms large room delay observed in the neuronal responses (median difference = 10ms, p = 4.7x10^{-31}). For the simulation, differences in $COM^+$and PT^+between the two reverberation conditions were small (Figure 3-Figure supplement 2B,E, median COM^+ difference = 1.7ms, p = 3.4x10^{-4}; median PT^+ difference = 0.0ms, p = 2.3x10^{-30}), with a slight difference from the real responses for PT^+difference (median difference = 0.0ms, p = 7.7x10^{-6}) but not COM^+difference (median difference = 0.0ms, p = 0.72). In summary, differences in stimulus properties alone were not able to account for the ~10ms delay in inhibitory COM timing in the large reverberant room, and these are likely to arise instead from neural adaptation to room reverberation.

To further confirm that the shift in inhibitory receptive fields arises from neuronal adaptation to reverberation and not to differences in stimulus statistics between the room conditions, we compared how all neurons in our dataset respond to a probe stimulus (a non-reverberated noise burst) interspersed within the small and large room reverberation stimuli (see Noise burst analysis). If the neurons adapt to the current reverberation condition, we should expect them to respond differently to the noise probe when it occurs within the small room and large room stimuli, reflecting the different adaptation states of the neurons. The neuronal responses to the noise probe showed a similar initial onset excitation (0-20ms) in both conditions, but the return to baseline firing was slower in the large room condition (Figure 3-Figure supplement 3A). This is consistent with the previous STRF analysis, wherein the excitatory temporal profile was similar between the small and large rooms (Figure 3B,D) , while the inhibitory components were delayed in time in the large room (Figure 3B,D). For each cortical unit, we compared the center of mass of the noise burst response between the small and large rooms (Figure 3-Figure supplement 3B). The COM of the noise response increased slightly in the large room (median COM difference = 1.0ms, p = 0.0063). Therefore, responses to an anechoic probe noise show further evidence for reverberation adaptation in auditory cortical neurons, and are consistent with the predicted delayed inhibition in the presence of increased reverberation.

To further confirm and explore the adaptive basis of our results, we presented our reverberant sounds in blocks, which switched between the small and large room every 8s (see Figure 3-Figure supplement 4A and Switching stimuli analysis). This switching stimulus was tested in 310 neurons across 4 ferrets. If the room adaptation accumulates throughout the 8s following a room switch, we would expect the inhibitory component of neuronal STRFs to be increasingly delayed throughout this period. To test this prediction, we fitted STRFs to neuronal responses separately from the first and last half of each 8s room block, for the small (S1 early and S2 late halves) and large room (L1 early and L2 late halves). The switching stimulus was designed to ensure that the stimulus set of L1 and L2 (or S1 and S2) was the same, but the order of stimuli was shuffled differently for these two time periods. Specifically, we predicted that the neuronal STRFs would have a larger COM^- during the L2 than the L1 period, while COM^+ should remain unchanged. By the same reasoning, in a large-to-small room switch, we expected the COM^- to be smaller in S2 than in S1, while COM^+ should remain similar.

We observed these predicted trends in our data, as show in Figure 3-Figure supplement 4B,C. The COM^- decreased from S1 to S2 (median difference = -0.9ms, Wilcoxon signed-rank test, p =

0.019), while COM^+ did not change across these two periods (median difference = 0.52ms, p = 0.85). In the switch to a large room, COM^- increased from the first (L1) to second (L2) half of the block (median difference = 1.5ms, p = 0.0088), while COM^+ did not change (median difference = 0.8ms, p = 0.35). These results further suggest that auditory cortical receptive fields are genuinely adapting dynamically to the changing reverberant conditions.

Frequency dependence of the temporal profile of adaptation

Reverberation is a frequency-dependent effect, as higher frequencies are usually attenuated by air and surfaces faster than lower ones in natural conditions (***Traer and McDermott, 2016***; ***Kuttruff, 2017***). Therefore, we explored whether our dereverberation model and auditory cortical neurons also show frequency-dependent reverberation effects.

Figures 2-Figure supplement 1 and Figure 2-Figure supplement 2 plot the reverberation model kernels and neuronal STRFs as a function of their frequency tuning. A visual inspection of these plots reveals that in both the model and the neuronal data, while the temporal spread of the excitatory components stays relatively constant across the preferred frequency, the inhibitory components tend to extend less far back in time as the preferred frequency increases. This temporal narrowing of the inhibitory fields is observed for both the large and the small reverberant rooms. Therefore, the frequency-dependent effects predicted by our dereverberation model are confirmed in our cortical recordings.

To further examine these frequency-tuning effects, we plotted the excitatory and inhibitory center of mass values (COM^+, COM^-) as a function of the anechoic frequency estimated by the model kernels (Figure 4A) or the best frequency of the neuronal STRFs, i.e. the sound frequency of the highest weight (Figure 4B). The inhibitory components occurred systematically later in model kernels that were tuned to lower frequencies, in both the small (Pearson's correlation: r = -0.57, p = 0.0037) and large room (r = -0.80, p = $2.6\text{x}10^{-6}$) simulations. The same correlation between best frequency and COM^- was present in the neuronal STRFs (small room: r = -0.80, p = $3.0\text{x}10^{-6}$; large room: r = -0.85, p = $1.6\text{x}10^{-7}$). In contrast, the dereverberation model showed a smaller magnitude but significant increase of the excitatory COM^+ with best frequency (small room: r = 0.52, p = 0.0087; large room: r = 0.55, p = 0.0049), while there was no relationship between COM^+ and best frequency in the neuronal data (small room: r = -0.34, p = 0.1; large room: r = -0.25, p = 0.24).

Figure 4A,B also show that the inhibitory components were later in time in the large room than in the small room across the entire best frequency range, for both the dereverberation model and neuronal data. The COM^+ values, on the other hand, were largely overlapping between the two rooms across this frequency range. This is in agreement with our observations that the inhibitory components of the receptive fields shift reliably with room size, while the excitatory components do not.

The frequency dependence of the inhibitory shift may reflect a frequency dependence in the reverberation acoustics themselves. The decay rate of the power in the impulse response of a reverberant environment depends on sound frequency, and this dependence can change across different environments. However, many man-made and natural environments show a gradual decrease in decay rate above about ~0.5kHz (***Traer and McDermott, 2016***). The early decay rate can be measured as the reverberation time RT_{10}, which is the time necessary for the sound level to decay by 10dB relative to an initial sound impulse. The frequency-dependent RT_{10} values for our small and large rooms are plotted in Figure 4C. The impulse responses of both rooms exhibited a decrease in RT_{10} values as a function of frequency (Pearson's correlation; small room: r = -0.82, p = $1.1\text{x}10^{-10}$; large room: Pearson's correlation: r = -0.91, p = $8.0\text{x}10^{-10}$). Therefore, the frequency-dependent delay in the inhibitory components of our dereverberation model and cortical STRFs paralleled the RT_{10} frequency profile of the virtual rooms in which the sounds were presented.

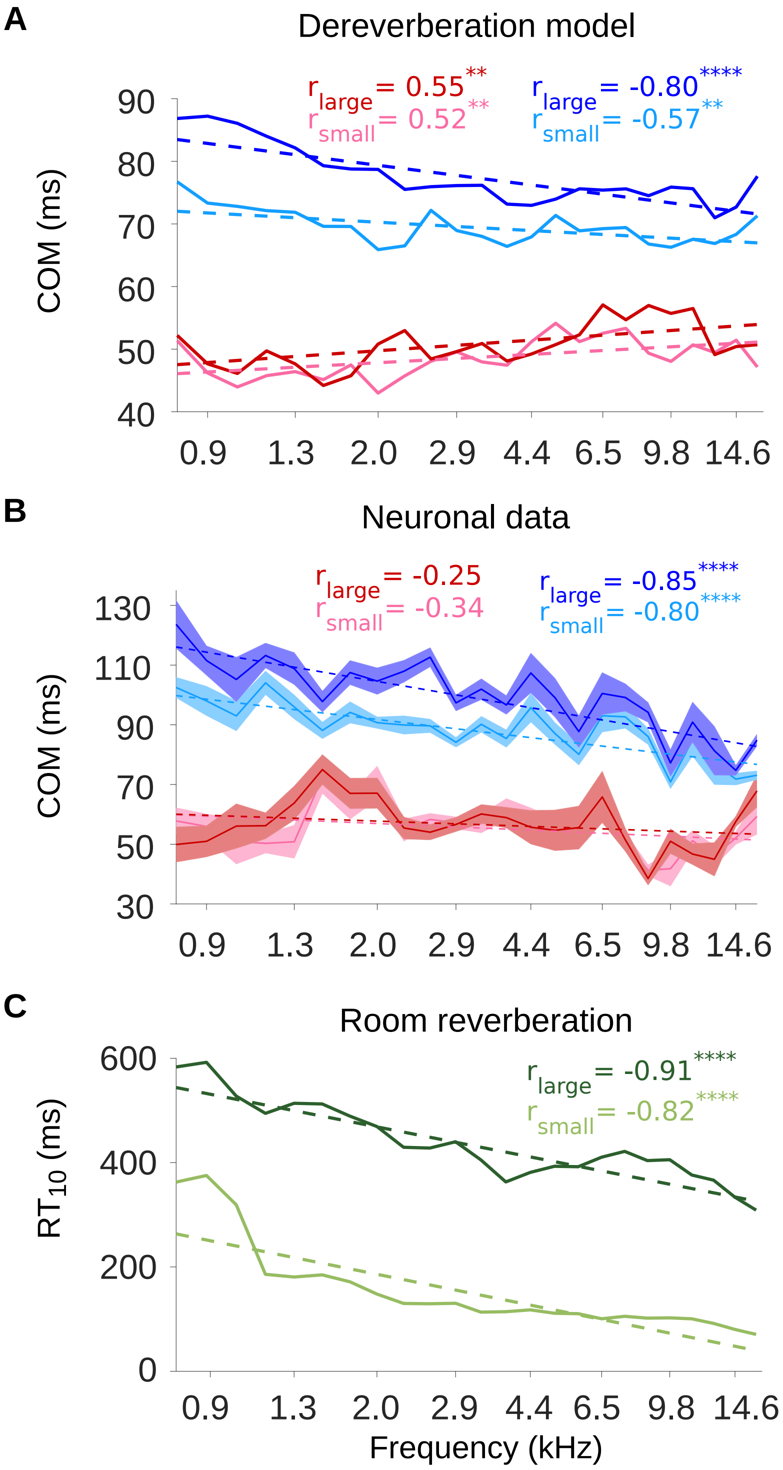

Figure 4. The inhibitory field latencies are frequency dependent, consistent with the reverberation.
A, Center of mass values (COM) are plotted against the anechoic frequency channel being estimated, for the excitatory and inhibitory fields of each model kernel for the large room and for the small room. These are color coded as follows: excitatory COM (large room, COM^{+}_{large}, red; small room, COM^{+}_{small}, pink) and their inhibitory counterparts (COM^{-}_{large}, blue; COM^{-}_{small}, cyan). The dashed lines show a linear regression fit for each room, and the Pearson's r value for each fit is given in the top right corner of the plot. **B**, COM values are plotted against the best frequency for the neuronal data (sound frequency of highest STRF weight). Each neuron was assigned a best frequency and the COM values measured. The solid lines represent the mean COM value for each best frequency, the shaded areas show ± SEM; color scheme and other aspects as in A. **C**, RT_{10} values are plotted as a function of cochlear frequency bands, for the large (dark green) and small (light green) rooms. Linear regression fit (dotted line) was used as in A and B to calculate r. Significance of Pearson's correlation: $^{****}p < 0.0001, ^{**}p < 0.01$.

Discussion

In this study, we applied a normative modelling approach to ask the question: If a function of the auditory system is to remove reverberation from natural sounds, how might the filtering properties of neurons adapt to achieve this goal? To answer this question we used a rich dataset of anechoic speech and natural environmental sounds, adding different amounts of reverberation to them. We then trained a linear dereverberation model to remove this reverberation. We constructed our model in such a way that the selectivity (kernels) of the model units after training can be compared to the filtering properties (STRFs) of real auditory cortex neurons in the ferret (Figure 1). We confirmed the validity of our dereverberation model by showing that it recapitulated known properties of auditory cortical neurons, such as frequency tuning and temporally asymmetric STRFs with excitation followed by inhibition (Figure 2). Interestingly, our dereverberation model also makes two novel predictions: (1) the inhibitory components of neuronal STRFs should be more delayed in more reverberant conditions (Figure 3); and (2) the inhibition should occur earlier for higher sound frequencies (Figure 2-Figure supplement 1, 2, Figure 4).

We verified both of these predictions using electrophysiological recordings from ferret auditory cortex neurons, fitting STRFs to neuronal responses to sounds from the same rich dataset, and comparing them to the model kernels. Finally, we used three additional methods - non-adaptive simulated neurons, probe stimuli and switching stimuli - to confirm that the observed changes in the neuronal STRFs are consistent with a truly adaptive dynamic process (Figure 3-Figure supplement 2, 3, 4). Thus, our results suggest that the population of auditory cortex neurons adapt to reverberation by extending their inhibitory field in time in a frequency-dependent manner. This proposed auditory cortical adaptation is summarized in Figure 5. In the following, we explore these findings in the broader context of previous studies and possible mechanisms for adaptation to reverberation.

Auditory cortical neurons adapt their responses to reverberation

Previous studies have shown that human hearing is remarkably robust to reverberation when listeners discriminate speech and naturalistic sounds (***Houtgast and Steeneken, 1985***; ***Bradley, 1986***; ***Darwin and Hukin, 2000***; ***Culling et al., 2003***; ***Nielsen and Dau, 2010***). Our neurophysiological results in the ferret auditory cortex are consistent with such robust representation. We find that neurons recorded in the auditory cortex tend to adapt their responses in a way that is consistent with the computational goal of removing reverberation from natural sounds (Figures 2, 3), even in anesthetized animals. Our results are also in good agreement with a previous study in awake passive listening ferrets, which showed that anechoic speech and vocalizations were more readily decodable from the responses of auditory cortex neurons to echoic sounds, than the echoic sounds themselves (***Mesgarani et al., 2014***). A similar study in humans using EEG corroborated these findings, showing better decoding accuracy of anechoic speech envelope compared to distorted reverberant speech, but only when listeners attended to the sound sources (***Fuglsang et al., 2017***).

Interestingly, a human MEG study suggests that auditory cortex may contain both reverberant and dereverberated representations of speech in reverberant conditions (***Puvvada et al., 2017***). In addition, ***Traer and McDermott*** (***2016***) found that humans were able to discriminate different reverberant conditions well with both familiar and unfamiliar sounds. In line with this, a minority of neurons in our study did not change the timing of their inhibitory responses in different reverberant conditions or showed the opposite effect from our model prediction (i.e. their COM^- and PT^- decreased in the more reverberant room) (Figure 3B,D). Thus, although most cortical neurons adapted to reverberation, it is possible that some of them might carry information about the reverberant environment or even represent it more explicitly.

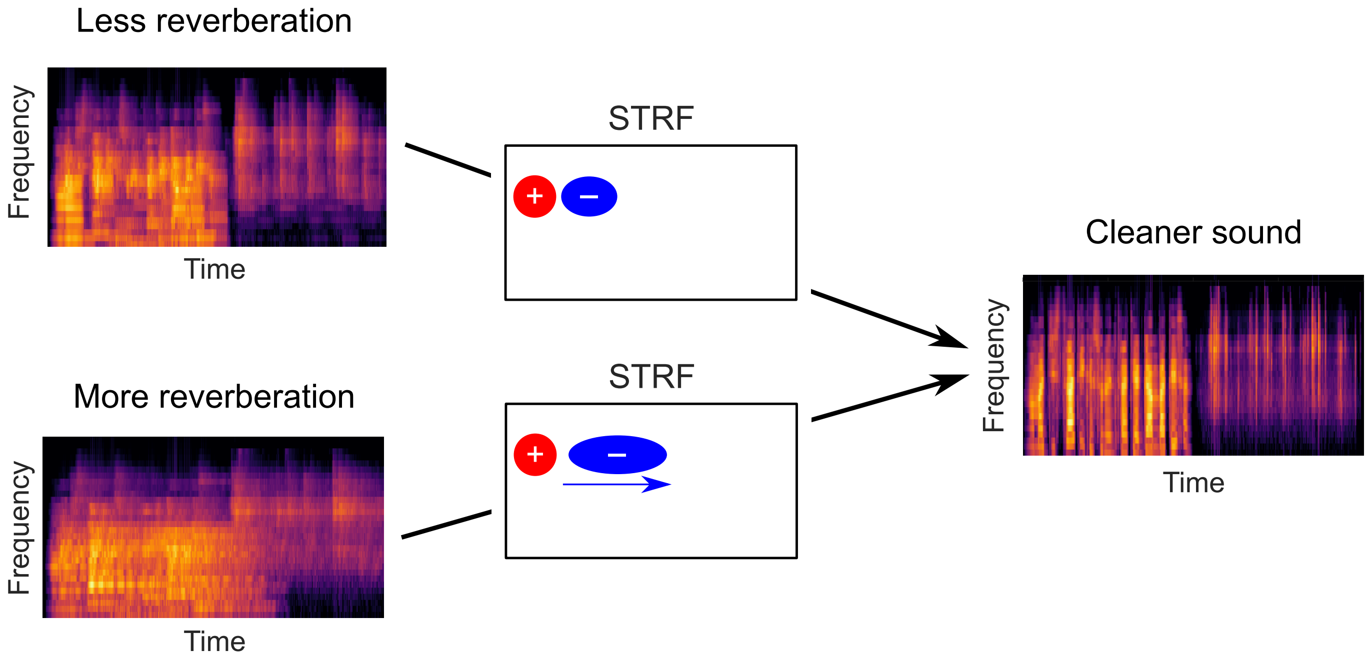

Figure 5. Schematic of dereverberation by auditory cortex.
Natural environments contain different levels of reverberation (illustrated by the left cochleagrams). Neurons in auditory cortex adjust their inhibitory receptive fields to ameliorate the effects of reverberation, with increased latency of inhibition for more reverberant environments (center). The consequence of this adaptive process is to arrive at a representation of the sound in which reverberation is reduced (right cochleagram).

Temporal shifts in inhibition underlie adaptation to reverberation

Our findings build on and provide an explanation for those of ***Mesgarani et al.*** (***2014***). These authors approximated a reverberant stimulus by convolving speech and vocalizations with decaying white noise. In contrast, we used a more diverse stimulus set, which included many environmental sounds that can have very different acoustical statistics (***Attias and Schreiner, 1996***; ***Turner, 2010***), and a model of reverberation that included early reflections and their frequency dependence, which are known to have important perceptual effects (***Traer and McDermott, 2016***). ***Mesgarani et al.*** (***2014***) proposed a combination of subtractive synaptic depression and multiplicative gain change as a potential mechanism for the observed adaptation in their study. However, they acknowledged that other functionally equivalent mechanisms might also be feasible. Notably, their study did not test different echoic conditions with varying amounts of reverberation. Therefore, the time constants of the synaptic depression and gain components in their model were fixed. ***Mesgarani et al.*** (***2014***) speculated that these time constants might have an important impact in conditions with different amounts of reverberation. This is indeed one of our main novel findings: more reverberant environments require more temporally delayed inhibitory responses within the STRFs of auditory cortical neurons.

Adaptation to reverberation is frequency dependent

Another novel finding of the present study was that the temporal lag of the inhibition was frequency dependent in both the model kernels and neuronal STRFs (Figure 2-Figure supplement 1, 2). For both the small and large rooms, the temporal lag of the inhibition, but not the excitation, approximately tracked the reverberant profile over sound frequency of the acoustic spaces (measured by the reverberation time (RT_{10}, Figure 4). Natural and man-made environments exhibit certain regularities, and the decline in reverberation over this frequency range is one of them (***Traer and McDermott, 2016***). Future studies could examine whether neurons adapt their responses accordingly to room impulse responses with more unusual RT_{10} frequency profiles.

The frequency-dependence of the delay in inhibition likely relates to some degree to the time constants of mean-sound-level adaptation (***Dean et al., 2008***), which also decrease with frequency in inferior colliculus neurons responding to non-reverberant noise stimuli (***Dean et al., 2008***). A study by ***Willmore et al.*** (***2016***) found that this frequency dependence of mean-sound-level adaptation may impact cortical responses and is consistent with removing a running average from natural sounds with undefined reverberation levels. Hence, the frequency dependence we observe in the present study may to some extent reflect general mechanisms for removing both reverberation and the mean sound level, and may be at least partially inherited from subcortical areas.

Possible biological implementations of the adaptation to reverberation

What might be the basis for the cortical adaptation to reverberation that we have observed? Some plausible mechanisms for altering the inhibitory field include synaptic depression (***David et al., 2009***), intrinsic dynamics of membrane channels (***Abolafia et al., 2011***), hyperpolarizing inputs from inhibitory neurons (***Li et al., 2015***; ***Natan et al., 2015***; ***Gwak and Kwag, 2020***), or adaptation inherited from subcortical regions such as the inferior colliculus or auditory thalamus (medial geniculate body) (***Dean et al., 2008***; ***Devore et al., 2009***; ***Willmore et al., 2016***; ***Lohse et al., 2020***). The physiological data obtained in this study do not allow us to discriminate among these mechanisms.

Hence, it would be important to investigate whether the adaptive phenomenon we have found occurs at subcortical levels too, namely the inferior colliculus and the medial geniculate body. Previous research in the inferior colliculus of rabbits has shown that neural responses to amplitude-modulated noise partially compensate for background noise and, for some neurons, particularly when that noise comes from reverberation (***Slama and Delgutte, 2015***). However, this study only examined one room size, so it did not investigate the temporal phenomenon we observed. ***Rabinowitz et al.*** (***2013***) found that neurons in the inferior colliculus in ferrets generally adapt less to the addition of non-reverberant background noise than those recorded in auditory cortex. This and other studies indicate that an increase in adaptation to sound statistics from auditory nerve to midbrain to cortex helps to construct noise-invariant sound representations in the higher auditory brain (***Dean et al., 2005***, ***2008***; ***Watkins and Barbour, 2008***; ***Wen et al., 2009***; ***Lohse et al., 2020***). However, subcortical adaptation phenomena may be influenced by cortical activity through descending connections (***Robinson et al., 2016***), making it challenging to dissect the neuroanatomical origin of these effects. Similarly, it is possible that reverberation adaptation also becomes more complete as we progress along the auditory pathway.

Considerations and future work

We undertook our electrophysiological recordings in the present study under general anesthesia in order to control for the effects of attention on reverberation adaptation and to facilitate stable recording of neural responses during our large stimulus set. Cortical adaptation to reverberation has been previously observed in awake listeners ((***Mesgarani et al., 2014***; ***Fuglsang et al., 2017***)), and we observed adaptive inhibitory plasticity in the anesthetized animal that is also consistent with dereverberation. This indicates that this form of adaptation is at least in part driven by stimulus statistics and can occur independently of activity and feedback from higher auditory areas (***Krom et al., 2020***).

Previous work has shown no effect of anaesthesia on another kind of adaptation, contrast gain control, in either the ferret auditory cortex (***Rabinowitz et al., 2011***) or the mouse inferior colliculus (***Lohse et al., 2020***). There is therefore no *a priori* reason to expect that cortical adaptation to reverberation would be substantially different in awake ferrets. Nevertheless, the effects of attention and behavior on auditory cortical STRFs in the ferret are well documented (***David, 2018***). These can manifest, for example, as gain changes and tuning shifts. Considering the importance of reverberation to perception, it would be interesting to explore the effects described here in behaving animals.

Another point for future research to consider is how our normative model could be further developed. For simplicity and interpretability, we used an elementary linear model. However, there are many more complex and powerful models for dereverberation in acoustical engineering, some of which may provide insight into the biology (***Naylor and Gaubitch, 2010***). Also, in our modelling we were focused on assessing what characteristics of dereverberation model kernels might change under different conditions, not on how the brain learns to make these changes. Hence, we gave our dereverberation model access to the true anechoic sound, something the brain would not have access to. However, there are blind dereverberation models that aim to dereverberate sounds from just one or two microphones, without access to the original anechoic sounds or room impulse response (***Li et al., 2018***; ***Jeub et al., 2010***). These blind dereverberation models will be particular useful to compare to biology if we want to explore how the brain learns to perform dereverberation with just two ears. It is also worth considering that the auditory system will be performing other functions in addition to dereverberation and these may be useful to add into a model.

Summary

We have observed in auditory cortical neurons a form of adaptation where the inhibitory component of the receptive fields is delayed in time as the room impulse response increases in a larger room. This is consistent with the cortex adapting to dereverberate its representation of incoming sounds in a given acoustic space. Dereverberated representations of sound sources would likely be more invariant under different acoustic conditions and thus easier to consistently identify and process, something valuable for any animal's survival. Reverberation is a ubiquitous phenomenon in the natural world and provides a substantial challenge to the hearing impaired and speech recognition technologies. Understanding the adaptive phenomena of the brain that allow us to effortlessly filter out reverberation may help us to overcome these challenges.

Methods and Materials

Animals

All animal procedures were approved by the local ethical review committee of the University of Oxford and performed under license from the UK Home Office. Three adult female and four adult male ferrets (*Mustela putorius furo*; Marshall BioResources, UK) were used in the electrophysiology experiments (mean age = 8.4 months; standard deviation = 4.2 months).

Surgical procedure

Terminal electrophysiological recordings were performed on each ferret under general anesthesia. Anesthesia was induced with an intramuscular injection of ketamine (Vetalar; 5 mg/kg) and medetomidine (Domitor; 0.02 mg/kg), and was maintained with a continuous intravenous infusion of these two drugs in Hartmann's solution with 3.5% glucose and dexamethasone (0.5 mg/ml/hr). The animal was intubated and artificially ventilated with medical O_2. Respiratory rate, end-tidal CO_2, electrocardiogram and blood oxygenation were continuously monitored throughout the recording session. Eye ointment (Maxitrol; Alcon, UK) was applied throughout and body temperature was maintained at 36-38°C. Atropine (Atrocare; 0.06 mg/kg i.m.) was administered every 6 hours, or when bradycardia or arrhythmia was observed.

Once anesthetized, each ferret was placed in a custom-built stereotaxic frame and secured with ear bars and a mouthpiece. After shaving the scalp and injecting bupivacaine (Marcain, <1mg/kg s.c.), the skin was incised and the left temporal muscle removed. A steel holding bar was secured to the skull using dental cement (SuperBond; C&B, UK) and a stainless steel bone screw (Veterinary Instrumentation, UK). A circular craniotomy (10 mm diameter) was drilled over the left auditory cortex, and the dura was removed in this region. The brain surface was covered with a solution of 1.25% agarose in 0.9% NaCl, and silicone oil was applied to the craniotomy regularly throughout recording.

With the ferret secured in the frame, the ear bars were removed, and the ferret and frame were placed in an electrically isolated anechoic chamber for recording. Recordings were then carried out in the left auditory cortex. An Ag/AgCl external reference wire was inserted between the dura and the skull on the edge of craniotomy. A Neuropixels Phase 3a microelectrode probe (***Jun et al., 2017***) was inserted orthogonally to the brain surface through the entire depth of auditory cortex. The cortical area of each penetration was determined based on its anatomical location in the ferret ectosylvian gyrus, the local field potential response latency, and the frequency response area (FRA) shapes of neurons. Based on these citeria, 95% of the recorded neurons were either within or on the ventral border of the primary auditory areas (primary auditory cortex, A1 and anterior auditory field, AAF), while the remaining neurons were located in secondary fields on the posterior ectosylvian gyrus. Following each presentation of the complete stimulus set, the probe was moved to a new location within auditory cortex. Data were acquired at a 30 kHz sampling rate using SpikeGLX software (https://github.com/billkarsh/SpikeGLX) and custom Matlab scripts (Mathworks).

Spike sorting

The recorded signal was processed offline by first digitally highpass filtering at 150Hz. Common average referencing was performed to remove noise across electrode channels (***Ludwig et al., 2009***). Spiking activity was then detected and clustered using Kilosort2 software (***Pachitariu et al., 2016***) (https://github.com/MouseLand/Kilosort2). Responses from single neurons were manually curated using Phy (https://github.com/cortex-lab/phy) if they had stereotypical spike shapes with low variance and their autocorrelation spike histogram showed a clear refractory period. Spikes from a given cluster were often measurable on 4-6 neighboring electrode channels, facilitating the isolation of single units. Only well isolated single units and multiunit clusters that were responsive to the stimuli (noise ratio <40 , (***Sahani and Linden, 2003***; ***Rabinowitz et al., 2011***)) were included in subsequent analyses.

Sound presentation

Stimuli were presented binaurally via Panasonic RP-HV094E-K earphone drivers, coupled to otoscope speculae inserted into each ear canal. The speculae were sealed in place with Otoform (Dreve Otoplastik). The earphones were driven by a System 3 RP2.1 multiprocessor and headphone amplifier (Tucker-Davis Technologies). Sounds were presented at a sampling rate of 48828Hz. The output response of the earphones was measured using a Brüel & Kjær calibration system with a GRAS 40DP microphone coupled to the end of the otoscope speculae with a silicone tube. An inverse filter was applied to the speaker output to produce a flat spectral response (±3dB) over the stimulus frequency range (200Hz-22kHz). Sound intensity was calibrated with an Iso-Tech TES-1356-G sound level calibrator.

Sound stimuli and virtual acoustic space

There are two stimulus sets, the set used to train the dereverberation model, and the set played to the ferrets, which was prepared from a subset the sounds used to make the first set. The stimuli used to train the dereverberation model were constructed from a dataset consisting of clips of anechoic sounds containing human speech and other natural sounds, such as cracking branches, footsteps, and running water. Most of the sound clips were recorded in an anechoic chamber using a Zoom H2 or Zoom H4 sound recorder, apart from some that came from the RWCP Sound Scene Database in Real Acoustic Environments (***Nakamura et al., 1999***). The clips varied in duration from 3s to 10s. A portion of the clips from the dataset was concatenated together to make a single stimulus of 600s duration. A 0.25s cosine ramp was applied to the onset and offset of each snippet to avoid clipping artifacts in concatenation. The 600s stimulus was then band-pass filtered from 200Hz–20kHz using an 8th-order Butterworth filter. We also constructed a held-out test set of 100s duration in the same manner using different examples of the same types of sounds from the dataset.

Finally, this stimulus was played in a virtual acoustic space (VAS), providing it with reverberation and head-related filtering. We used the "Roomsim" software (***Campbell et al., 2005***) to generate the virtual acoustic space. This software creates a cuboidal room of arbitrary x, y and z dimensions and simulates its acoustic properties for a listener at a particular position and orientation in space, for a sound source at a particular position. The simulations are based on the room-image method (***Allen and Berkley, 1979***; ***Heinz, 1993***; ***Shinn-Cunningham et al., 2001***). One difference between the standard room-image method and Roomsim is that the latter incorporates the absorption properties of different materials, which can be summarized by their frequency-dependent absorption coefficients. In principle, the amount of reverberation in a room will depend on its size, shape and the material from which the walls are made. For our room simuloations the walls, ceiling and floor use the frequency-dependent absorption coefficients of stone (***Álvarez Morales et al., 2014***). We decided to vary the amount of reverberation by changing the room size whilst keeping the other parameters fixed. Four different corridor-shaped rooms were created:

1. **Anechoic room**
2. **Small room** (length x width x height, 3mx0.3mx0.3m, RT_{10} = 130ms)
3. **Medium room** (7.5mx0.75mx0.75m, RT_{10} = 250ms)
4. **Large room** (15mx1.5mx1.5m, RT_{10} = 430ms)

Thus processing the 600s stimulus for each room provided four 600s stimuli. Note that the anechoic room does not have a clearly defined "shape", having no reflecting walls, ceiling or floor, with the acoustic filtering determined only by the relative orientation and distances of the sound source and receiver. Roomsim simulates the orientation-specific acoustic properties of the receiver's head and outer ear, represented by the head-related transfer function (HRTF). In all simulations, we used the same ferret HRTF provided from measurements previously made in the lab on a real ferret (from ***Schnupp et al. , 2001***). The joint filtering properties of the ferret's HRTF and the room were simulated together by Roomsim to produce a binaural room impulse response (BRIR). The ferret head position and orientation were simulated in the VAS, positioning it 0.15m from the floor, at the midpoint of the room's width (0.15m for the small, 0.375m for the medium and 0.75m for the large) and 1/4 of the room's length from one end (0.75m for the small, 1.875m for the medium and 3.75m for the large) and directly facing the opposite end. In all four room conditions, the sound source was positioned at the same height as the ferret's head (0.15m) and at a distance of 1.5m straight ahead in the direction faced by the ferret (0° azimuth and 0° elevation relative to the ferret's head). The reverberation time RT_{10} is the time necessary for the sound level to decay by 10dB relative to an initial sound impulse. We measured this using a cochlear model, as explained in the next section Cochlear model.

The stimuli presented to the ferrets were constructed from a representative subset of the anechoic natural stimuli used to train the dereverberation model. We cut 40 different snippets of natural sounds, each 2s in duration, from the clips in the datatset. These 2s snippets were concatenated together into two 40s long stimuli. A 0.25s cosine ramp was applied to the onset and offset of each snippet to avoid clipping artifacts in concatenation. The two 40s stimulus blocks were then processed in VAS in exactly the same way as with the modelling stimulus set, for the same small, medium and large rooms. This provided two 40s blocks for each reverberant condition (a small, medium or large room, see subsection Sound stimuli and virtual acoustic space). We played the small and large room conditions in 7 animals and the small, medium and large room conditions in 2 out of those 7. The 40s blocks were presented in pseudo random order, with ~5s of silence between blocks. This presentation was repeated ten times, with a different order each time.

Cochlear model

We used a power-spectrogram based model of cochlear processing as described in ***Rahman et al.*** (***2020***). Briefly, a spectrogram was produced from the sound waveform by taking the power spectrum through a short-time Fourier transform (STFT) using 10-ms Hanning windows, overlapping by 5 ms. The power of adjacent frequency channels was summed using overlapping triangular windows (using code adapted from melbank.m, http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html) to produce 30 log-spaced frequency channels ranging from 400Hz to 19kHz center frequencies. The resulting power in each channel at each time point was converted to log values and any value below a low threshold was set to that threshold.

We used the cochleagram to measure the frequency-band-specific reverberation times (RT_{10}) shown in Figure 4C. Our method is similar to that of ***Traer and McDermott*** (***2016***), but for consistency we used our cochlear model rather than theirs. First, we produce an impulse response, the sound induced at the right ear of the ferret in the virtual room, by a simple click at the standard source position. Then, we put this impulse response through our cochlear model to generate a cochleagram. Next, for each frequency band in this cochleagram, we fitted a straight line to the plot of the decaying log power output (dB) of the cochleagram over time. Using the slope of this line of best fit, we found the amount of time it took for this output to decay by 10dB. This provided the RT_{10} for each frequency band. We measured the overall RT_{10} of each room by taking the median RT_{10} over all 30 frequency bands.

Model kernels

The dereverberation model consisted of a set of linear kernels, one for each of the 30 frequency channels in the anechoic cochleagram. The kernels were fitted separately for each reverberant condition, thus providing 30 kernels for each room. The dereverberation model is summarized by the following equation:

$$\hat{x}^{\text{anech}}_{f't} = \sum_{f=1}^{f_{max}} \sum_{h=1}^{h_{max}} k_{f'fh} x^{\text{reverb}}_{f(t-h+1)} + b_{f'} \tag{1}$$

Here, $\hat{x}^{\text{anech}}_{f't}$ is the estimate of the anechoic cochleagram for frequency channel f' and time bin t. Obtaining $\hat{x}^{anech}_{f't}$ involved convolving the kernels $k_{f'fh}$ with the reverberant cochleagram x^{reverb}_{ft}. Here f is the frequency channel in the reverberant cochleagram and h indexes the time lag used in the convolutions. The model weights $k_{f'fh}$ are composed of 30 kernels, one for each frequency channel f' in the anechoic cochleagram. Finally, the bias term for frequency channel f' is $b_{f'}$.

For each anechoic frequency channel f', the associated model kernel was separately fitted to minimize the mean squared error between the kernel's estimate of that frequency channel of the anechoic cochleagram $\hat{x}^{anech}_{f't}$ and that actual channel of the anechoic cochleagram $x^{anech}_{f't}$, subject to L_2 regularization ("ridge" regression) on $k_{f'fh}$. The weights were fitted using the glmnet package (GLM, J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon, Stanford University, Stanford, CA; http://web.stanford.edu/~hastie/glmnet_matlab/index.html). To select the regularization strength (the hyperparameter λ), we performed 10-fold cross-validation, using 90% of the data for the training set and 10% (an unbroken 60s segment) for the validation set. Our validation sets over folds were non-overlapping. We found the λ that gave the lowest mean-squared error averaged over the 10 folds. Using this λ, we then re-fitted the model kernels using the whole cross-validation set (training + validation set). These resulting kernels are the ones shown and used in all analyses. These kernels were also used to estimate the dereverberation capacity of the model on the held-out test set. Note that here onward we typically refer to individual model kernels by k_{fh} for brevity, dropping the f' index used for the full set of kernels $k_{f'fh}$.

Neuronal STRFs

For each cortical unit, for each reverberation condition, we separately estimated its spectro-temporal receptive field (STRF) using its response to the natural stimuli under that condition (***Theunissen et al., 2001***). We used the STRF, a linear model, as this enabled comparison to our linear dereverberation model. The STRF model can be summarized by the following equation:

$$\hat{y}_{nt} = \sum_{f=1}^{f_{max}} \sum_{h=1}^{h_{max}} w_{nfh} x^{\text{reverb}}_{f(t-h+1)} + b_n \tag{2}$$

Here, $\hat{y}_{nt}$ is the estimated spike counts of neuron n at time bin t. Also, x^{reverb}_{ft} is the reverberant cochleagram in frequency channel f and at time t. For each neuron n, the weights in w_{nfh} over frequency channel f and history (time lag) index h provide its STRF. Finally, b_n is the bias term of unit n.

Notice the similarity of Equation 2 to Equation 1 of the dereverberation model. In both cases, we used the reverberant cochleagram as an input (from either the small, medium, or large room) and fitted the best linear mapping to the output. In the case of neuronal STRFs, the output is the neuronal spike count over time, whereas in the model kernel it is a frequency channel of the anechoic cochleagram. For each neuron and room, we seperately fitted an STRF by minimizing the mean squared error between the estimated spike counts $\hat{y}_{nt}$ and the observed spike counts y_{nt}. To do this, for a given room, we used the first 36s of neural response to the two 40s-stimuli associated with that room (as the last 4s contained a noise probe, see subsection Noise burst analysis). The weights were fitted using the glmnet package (GLM, J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon, Stanford University, Stanford, CA; http://web.stanford.edu/~hastie/glmnet_matlab/index.html). As for the model kernels (above), the fitting was subject to L_2 regularization. To select the regularization strength (the hyperparameter λ), we performed 10-fold cross-validation, using 90% of the data for the training set and 10% (an unbroken 7.2s segment) for the validation set. Our validation sets over folds were non-overlapping. We found the λ that gave the lowest mean-squared error averaged over the 10 folds. Using this λ, we then re-fitted the STRFs using the whole cross-validation set (training + validation set). The resulting STRFs are the ones shown and used in all analyses. As with the model kernels, from here onwards we typically refer to an individual STRF for a given neuron by the form w_{fh} for brevity, dropping the neuron index n used here in w_{nfh}.

Quantification of the temporal effects in model kernels and neuronal STRFs

To quantify the temporal profiles of the model kernels and neuronal STRFs, we chose two different measures:

1. Center of mass (COM)
2. Peak time (PT)

To compute them, we first obtained the averaged excitatory and inhibitory temporal profiles of the model kernels/neuronal STRFs as follows:

$$w_h^+ = \frac{1}{f_{max}} \sum_{f=1}^{f_{max}} [w_{fh}]_+ \tag{3}$$

$$w_h^- = \frac{1}{f_{max}} \sum_{f=1}^{f_{max}} [w_{fh}]_- \tag{4}$$

where w_{fh} is the neuronal STRF with f and h subscripts denoting frequency channel and history, respectively. Equations 3 and 4 are the same for the dereverberation model kernels but with k instead of w, as with all subsequent equations in this section. f_{max} is the number of frequencies (30) in the model kernel/neuronal STRF w_{fh}. The notation $[w_{fh}]_+$ and $[w_{fh}]_-$ stand for the element-wise operations max(w_{fh},0) and min(w_{fh},0), that is:

$$[w_{fh}]_+ = \begin{cases} w_{fh} \text{ if } w_{fh} \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

$$[w_{fh}]_- = \begin{cases} w_{fh} \text{ if } w_{fh} \leq 0 \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

Thus w_h^+ and w_h^- are the frequency-averaged positive-only, $[w_{fh}]_+$, and negative-only, $[w_{fh}]_-$, parts of the kernel w_{fh}.

From this, the COM was defined as follows:

$$COM^+ = \frac{\tau}{\sum_{h=1}^{h_{max}} w_h^+} \sum_{h=1}^{h_{max}} (h-1) w_h^+ \tag{7}$$

$$COM^- = \frac{\tau}{\sum_{h=1}^{h_{max}} w_h^-} \sum_{h=1}^{h_{max}} (h-1) w_h^- \tag{8}$$

The duration of a time bin is $\tau = 10$ms, hence time lag in the history of the neuronal STRF/model kernel ranges from $\tau(h-1) = 0$ms to $\tau(h_{max} - 1) = 190$ms. Thus COM^+ is the temporal center of mass for the positive (excitatory) components of the neuronal STRF/model kernel and COM^- the temporal center of mass for the negative (inhibitory) components.

The peak time (PT) was defined as the time at which the excitation and inhibition in the frequency averaged neuronal STRFs/model kernels peaked:

$$PT^+ = (\underset{h}{\text{argmax}}(w_h^+) - 1)\tau \tag{9}$$

$$PT^- = (\underset{h}{\text{argmin}}(w_h^-) - 1)\tau \tag{10}$$

Simulated neuron

In order to explore whether the changes that we observed are truly adaptive, we used simulated neurons that lacked adaptive receptive fields to generate responses. We then applied the same analyses to these simulated neuronal responses as we did to the actual responses. For each neuron n, we constructed a corresponding simulated neuron in the following way. First, we fitted a single STRF as described in section Neuronal STRFs. However, in this case we used the full dataset from the "small" and "large" conditions together, rather than fitting separate STRFs to the two conditions as we did previously.

Next, we fitted a sigmoid output non-linearity by first generating a spike count prediction $\hat{y}_{nt}$ for the full dataset according to equation 2 from section Neuronal STRFs, using this single STRF and then finding the sigmoid that best fits the actual spike count y_{nt} according to the following equation:

$$\hat{y}_{nt}^{\text{nonlin}} = \frac{\rho_1}{1 + \exp(-(\hat{y}_{nt} - \rho_3)/\rho_2)} + \rho_4 \tag{11}$$

Here $\hat{y}_{nt}^{\text{nonlin}}$ is the output of the point non-linearity at time bin t, providing a new estimate of the neuron's spike count. As mentioned, $\hat{y}_{nt}$ is the predicted spike count from the linear stage (see Equation 2) at time bin t, when fitted to the small and large room responses together. It is the four parameters ρ_1, ρ_2, ρ_3 and ρ_4 that are optimized in the fit.

We then used the fitted simulated model to produce an approximation of the real neuronal response to the reverberant stimulus sets for both the small and large conditions. In order to simulate realistic neuronal noise, we used the $\hat{y}_{nt}^{\text{nonlin}}$ output, at each time bin t, as the mean of a Poisson distribution from which we generated 10 "virtual" trials. Finally we performed the same analyses on these simuluated neural responses as we did for the real data; we fitted STRFs for the two reverberation conditions separately using these simulated model responses in place of the actual responses and then analyzed the resulting STRFs as outlined in the section above (Quantification of the temporal effects in model kernels and neuronal STRFs).

Noise burst analysis

To further confirm the adaptive change in properties of neurons across the two reverberant conditions, we presented a 500ms long unreverberated broadband noise burst embedded at a random time in the last 4s of each 40s sound block (i.e., from 36-40s) for each condition (small and large). Seven out of the ten repeats of any stimulus block contained a noise burst, with those seven randomly shuffled within the ten. The random timing distribution of the noise bursts was uniform and independent across repeats and conditions. For each neuron, responses to the noise bursts were assessed using peristimulus time histogram (PSTHs) with 10ms time bins. For the majority of neurons, the firing rate had returned to baseline by 100ms , so we decided to use the 0-100ms time window for further analysis (Figure 3-Figure supplement 3A). Different neurons had different response profiles, so in order to compare the adaptive properties in the two conditions we chose the center of mass (COM) of the firing rate profile within this window as a robust measure. This was defined similarly to the COM measure in subsection Quantification of the temporal effects in model kernels and neuronal STRFs (see also Equations 7 and 8). The COM for the noise bursts in the large and small conditions was calculated for each neuron individually and the difference between the two conditions computed (Figure 3-Figure supplement 3B).

Switching stimuli analysis

In order to confirm and explore the adaptive nature of the neuronal responses to reverberant sounds, we presented "switching stimuli" (Figure 3-Figure supplement 4A). These stimuli switched back and forth every 8 seconds between the large room and the small room and were created in the following way. First, we took our original reverberant stimuli for both the small room (80s duration) and large room (80s duration) conditions and divided them into consecutive 4s snippets, providing 20 snippets for each condition. We duplicated these two sets and shuffled each one independently, providing a total of four sets of 20 4s-long snippets. We then combined the snippets into eight 40s-long switching stimuli. These switching stimuli comprised 5 epochs of 8s duration each, with 4 "switches" between the small and large epochs. Half of the stimuli started from the large room condition and the other half from the small room condition. Within each 8s epoch, we defined two periods (period 1: 0-4s and period 2: 4-8s). The large-room periods were denoted by L1 (0-4s) and L2 (4-8s), and the small-room periods by S1 (0-4s) and S2 (4-8s) (Figure 3-Figure supplement 4A). The snippets from the first small-room set of 20 snippets populated the 20 S1 periods in order, while those from the second small-room set populated the S2 periods in a different order, due to the shuffling. Likewise, snippets from the first large-room set of 20 snippets populated the 20 L1 periods, and those from the second large-room set populated the L2 periods. Thus, the same set of stimuli were included in S1 and S2, and in L1 and L2, with the only differences being their ordering, and between the small and large room stimuli the amount of reverberation. When the 4s periods and 8s epochs were spliced together, they were cross-faded into each other with a 10ms cosine ramp with 5ms overlap, such that the transition from one period to the next was smooth with no detectable clicks between them. We played the eight 40s stimuli in random order to the ferrets; this was repeated 10 times with the order different each time.

The cortical responses recorded with these stimuli were analyzed using the procedure outlined in subsection Neuronal STRFs. For each neuron, we fitted four separate STRFs using the neural responses to the S1, S2, L1 and L2 periods. We did not use the first 8s of each of the eight 40s stimuli, since there was no prior sound (silence) and thus they would not be directly comparable to the other 4 epochs. We also did not use the first 500 ms of any of the periods, to avoid potential non-reverberation-related responses from the rapid transitions between them. From the resulting four STRFs, we extracted the COM^+ and COM^- values for each and compared S1 to S2 (Figure 3-Figure supplement 4B) and L1 to L2 (Figure 3-Figure supplement 4C).

Acknowledgments

We are grateful to Dr Quentin Gaucher for assistance with these electrophysiological experiments. We would also like to thank Zlatina Dimitrova for her artwork in Figure 1. This work was supported by a Wellcome Principal Research Fellowship to AJK (WT108369/Z/2015/Z), a BBSRC New Investigator Award (BB/M010929/1) to KMMW, and a Christopher Welch Scholarship (Oxford University Press) to AZI.

References

Abolafia JM, Vergara R, Arnold MM, Reig R, Sanchez-Vives MV. Cortical Auditory Adaptation in the Awake Rat and the Role of Potassium Currents. Cerebral Cortex. 2011 May; 21(5):977–990. https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhq163, doi: 10.1093/cercor/bhq163.

Allen JB, Berkley DA. Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America. 1979 Apr; 65(4):943–950. https://asa.scitation.org/doi/10.1121/1.382599, doi: 10.1121/1.382599, publisher: Acoustical Society of America.

Attias H, Schreiner CE. Temporal Low-Order Statistics of Natural Sounds. In: *NIPS*; 1996. .

Bizley JK, Nodal FR, Nelken I, King AJ. Functional organization of ferret auditory cortex. Cerebral Cortex. 2005; 15(10):1637–1653. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15703254.

Bradley JS. Speech intelligibility studies in classrooms. The Journal of the Acoustical Society of America. 1986 Sep; 80(3):846–854. http://asa.scitation.org/doi/10.1121/1.393908, doi: 10.1121/1.393908.

Brown GJ, Cooke M. Computational auditory scene analysis. Computer Speech & Language. 1994 Oct; 8(4):297–336. http://www.sciencedirect.com/science/article/pii/S0885230884710163, doi: 10.1006/csla.1994.1016.

Brughera A, Mikiel-Hunter J, Dietz M, McAlpine D. Auditory brainstem models: adapting cochlear nuclei improve spatial encoding by the medial superior olive in reverberation. bioRxiv. 2020 Jan; p. 694356. http://biorxiv.org/content/early/2020/03/27/694356.abstract, doi: 10.1101/694356.

Campbell DR, Palomaki KJ, Brown G. A MATLAB simulation of" shoebox" room acoustics for use in research and teaching. Computing and Information Systems J. 2005; 9:48–51.

Christianson GB, Sahani M, Linden JF. The Consequences of Response Nonlinearities for Interpretation of Spectrotemporal Receptive Fields. Journal of Neuroscience. 2008 Jan; 28(2):446–455. https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.1775-07.2007, doi: 10.1523/JNEUROSCI.1775-07.2007.

Culling JF, Hodder KI, Toh CY. Effects of reverberation on perceptual segregation of competing voices. The Journal of the Acoustical Society of America. 2003; 114(5):2871. http://scitation.aip.org/content/asa/journal/jasa/114/5/10.1121/1.1616922, doi: 10.1121/1.1616922.

Culling JF, Summerfield Q, Marshall DH. Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels. Speech Communication. 1994 Feb; 14(1):71–95. doi: 10.1016/0167-6393(94)90058-2.

Darwin CJ, Hukin RW. Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. The Journal of the Acoustical Society of America. 2000 Jul; 108(1):335–342. http://asa.scitation.org/doi/10.1121/1.429468, doi: 10.1121/1.429468.

David SV, Mesgarani N, Fritz JB, Shamma SA. Rapid Synaptic Depression Explains Nonlinear Modulation of Spectro-Temporal Tuning in Primary Auditory Cortex by Natural Stimuli. Journal of Neuroscience. 2009 Mar; 29(11):3374–3386. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5249-08.2009, doi: 10.1523/JNEUROSCI.5249-08.2009.

David SV. Incorporating behavioral and sensory context into spectro-temporal models of auditory encoding. Hearing Research. 2018 Mar; 360:107–123. https://linkinghub.elsevier.com/retrieve/pii/S0378595517303611, doi: 10.1016/j.heares.2017.12.021.

Dean I, Harper NS, McAlpine D. Neural population coding of sound level adapts to stimulus statistics. Nature neuroscience. 2005; 8(12):1684–1689. doi: 10.1038/nn1541, iSBN: 1097-6256 (Print).

Dean I, Robinson BL, Harper NS, McAlpine D. Rapid neural adaptation to sound level statistics. J Neurosci. 2008; 28(25):6430–6438. doi: 10.1523/JNEUROSCI.0470-08.2008, iSBN: 1529-2401 (Electronic)\r0270-6474 (Linking).

deCharms RC, Blake DT, Merzenich MM. Optimizing Sound Features for Cortical Neurons. Science. 1998 May; 280(5368):1439–1444. https://www.science.org/doi/10.1126/science.280.5368.1439, doi: 10.1126/science.280.5368.1439.

Devore S, Ihlefeld A, Hancock K, Shinn-Cunningham B, Delgutte B. Accurate Sound Localization in Reverberant Environments Is Mediated by Robust Encoding of Spatial Cues in the Auditory Midbrain. Neuron. 2009 Apr; 62(1):123–134. https://linkinghub.elsevier.com/retrieve/pii/S0896627309001639, doi: 10.1016/j.neuron.2009.02.018.

Fitzpatrick DC, Kuwada S, Kim DO, Parham K, Batra R. Responses of neurons to click-pairs as simulated echoes: Auditory nerve to auditory cortex. The Journal of the Acoustical Society of America. 1999 Dec; 106(6):3460–3472. http://asa.scitation.org/doi/10.1121/1.428199, doi: 10.1121/1.428199.

Fuglsang SA, Dau T, Hjortkjær J. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. NeuroImage. 2017 Aug; 156:435–444. https://linkinghub.elsevier.com/retrieve/pii/S105381191730318X, doi: 10.1016/j.neuroimage.2017.04.026.

Guediche S, Blumstein SE, Fiez JA, Holt LL. Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. Frontiers in Systems Neuroscience. 2014; 7(4):328–32. doi: 10.3389/fnsys.2013.00126.

Gwak J, Kwag J. Distinct subtypes of inhibitory interneurons differentially promote the propagation of rate and temporal codes in the feedforward neural network. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2020 May; 30(5):053102. http://aip.scitation.org/doi/10.1063/1.5134765, doi: 10.1063/1.5134765.

Harper NS, Schoppe O, Willmore BDB, Cui Z, Schnupp JWH, King AJ. Network Receptive Field Modeling Reveals Extensive Integration and Multi-feature Selectivity in Auditory Cortical Neurons. PLOS Computational Biology. 2016 Nov; 12(11):e1005113. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005113, doi: 10.1371/journal.pcbi.1005113, publisher: Public Library of Science.

Hartmann WM. Localization of sound in rooms. The Journal of the Acoustical Society of America. 1982 Nov; 72(S1):S93–S93. http://asa.scitation.org/doi/10.1121/1.2020159, doi: 10.1121/1.2020159.

Heinz R. Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. Applied Acoustics. 1993 Jan; 38(2):145–159. http://www.sciencedirect.com/science/article/pii/0003682X9390048B, doi: 10.1016/0003-682X(93)90048-B.

Helfer KS, Wilber LA. Hearing loss, aging, and speech perception in reverberation and noise. Journal of speech and hearing research. 1990 Mar; 33(1):149–55.

Houtgast T, Steeneken HJM. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. The Journal of the Acoustical Society of America. 1985 Mar; 77(3):1069–1077. http://asa.scitation.org/doi/10.1121/1.392224, doi: 10.1121/1.392224.

Huisman WHT, Attenborough K. Reverberation and attenuation in a pine forest. The Journal of the Acoustical Society of America. 1991 Nov; 90(5):2664–2677. http://asa.scitation.org/doi/10.1121/1.401861, doi: 10.1121/1.401861.

Humes LE, Dirks DD, Bell TS, Ahlstrom C, Kincaid GE. Application of the Articulation Index and the Speech Transmission Index to the recognition of speech by normal-hearing and hearing-impaired listeners. Journal of speech and hearing research. 1986 Dec; 29(4):447–62.

Jeub M, Löllmann HW, Vary P. Blind Dereverberation for Hearing Aids with Binaural Link. In: Bochum; 2010. p. 4. http://ikspub.iks.rwth-aachen.de/pdfs/jeub10c.pdf.

Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydın Barbic M, Blanche TJ, Bonin V, Couto J, Dutta B, Gratiy SL, Gutnisky DA, Häusser M, Karsh B, Ledochowitsch P, et al. Fully integrated silicon probes for high-density recording of neural activity. Nature. 2017 Nov; 551(7679):232–236. http://www.nature.com/articles/nature24636, doi: 10.1038/nature24636.

Kell AJE, McDermott JH. Invariance to background noise as a signature of non-primary auditory cortex. Nature Communications. 2019 Dec; 10(1):3958. http://www.nature.com/articles/s41467-019-11710-y, doi: 10.1038/s41467-019-11710-y.

Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N. Adaptation of the human auditory cortex to changing background noise. Nature Communications. 2019 Dec; 10(1):2509. http://www.nature.com/articles/s41467-019-10611-4, doi: 10.1038/s41467-019-10611-4.

Kim DO, Zahorik P, Carney LH, Bishop BB, Kuwada S. Auditory Distance Coding in Rabbit Midbrain Neurons and Human Perception: Monaural Amplitude Modulation Depth as a Cue. Journal of Neuroscience. 2015 Apr; 35(13):5360–5372. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3798-14.2015, doi: 10.1523/JNEUROSCI.3798-14.2015.

Kinoshita K, Delcroix M, Gannot S, P Habets EA, Haeb-Umbach R, Kellermann W, Leutnant V, Maas R, Nakatani T, Raj B, Sehr A, Yoshioka T. A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research. EURASIP Journal on Advances in Signal Processing. 2016 Dec; 2016(1):7. https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-016-0306-6, doi: 10.1186/s13634-016-0306-6.

Knudsen VO. The hearing of speech in auditoriums. The Journal of the Acoustical Society of America. 1929; 1(1):56. doi: 10.1121/1.1901470.

Kolarik AJ, Moore BCJ, Cirstea S, Aggius-Vella E, Gori M, Campus C, Pardhan S. Factors Affecting Auditory Estimates of Virtual Room Size: Effects of Stimulus, Level, and Reverberation. Perception. 2021 Jul; 50(7):646–663. http://journals.sagepub.com/doi/10.1177/03010066211020598, doi: 10.1177/03010066211020598.

Krom AJ, Marmelshtein A, Gelbard-Sagiv H, Tankus A, Hayat H, Hayat D, Matot I, Strauss I, Fahoum F, Soehle M, Boström J, Mormann F, Fried I, Nir Y. Anesthesia-induced loss of consciousness disrupts auditory responses beyond primary cortex. Proceedings of the National Academy of Sciences. 2020 May; 117(21):11770–11780. http://www.pnas.org/lookup/doi/10.1073/pnas.1917251117, doi: 10.1073/pnas.1917251117.

Kuttruff H. Room acoustics. Sixth edition ed. Boca Raton: CRC Press/Taylor & Francis Group; 2017.

Kuwada S, Bishop B, Kim DO. Approaches to the study of neural coding of sound source location and sound envelope in real environments. Frontiers in Neural Circuits. 2012; 6. http://journal.frontiersin.org/article/10.3389/fncir.2012.00042/abstract, doi: 10.3389/fncir.2012.00042.

Li C, Wang T, Xu S, Xu B. Single-channel Speech Dereverberation via Generative Adversarial Training. arXiv:180609325 [cs, eess]. 2018 Jun; http://arxiv.org/abs/1806.09325, arXiv: 1806.09325.

Li Ly, Xiong XR, Ibrahim LA, Yuan W, Tao HW, Zhang LI. Differential Receptive Field Properties of Parvalbumin and Somatostatin Inhibitory Neurons in Mouse Auditory Cortex. Cerebral Cortex. 2015 Jul; 25(7):1782–1791. https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bht417, doi: 10.1093/cercor/bht417.

Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM. Spectrotemporal Structure of Receptive Fields in Areas AI and AAF of Mouse Auditory Cortex. Journal of Neurophysiology. 2003 Oct; 90(4):2660–2675. https://journals.physiology.org/doi/full/10.1152/jn.00751.2002, doi: 10.1152/jn.00751.2002, publisher: American Physiological Society.

Litovsky RY, Yin TCT. Physiological Studies of the Precedence Effect in the Inferior Colliculus of the Cat. I. Correlates of Psychophysics. Journal of Neurophysiology. 1998 Sep; 80(3):1285–1301. https://www.physiology.org/doi/10.1152/jn.1998.80.3.1285, doi: 10.1152/jn.1998.80.3.1285.

Lohse M, Bajo VM, King AJ, Willmore BDB. Neural circuits underlying auditory contrast gain control and their perceptual implications. Nature Communications. 2020 Dec; 11(1):324. http://www.nature.com/articles/s41467-019-14163-5, doi: 10.1038/s41467-019-14163-5.

Ludwig KA, Miriani RM, Langhals NB, Joseph MD, Anderson DJ, Kipke DR. Using a Common Average Reference to Improve Cortical Neuron Recordings From Microelectrode Arrays. Journal of Neurophysiology. 2009 Mar; 101(3):1679–1689. https://www.physiology.org/doi/10.1152/jn.90989.2008, doi: 10.1152/jn.90989.2008.

Mesgarani N, David SV, Fritz JB, Shamma SA. Mechanisms of noise robust representation of speech in primary auditory cortex. Proceedings of the National Academy of Sciences. 2014 May; 111(18):6792–6797. http://www.pnas.org/cgi/doi/10.1073/pnas.1318017111, doi: 10.1073/pnas.1318017111.

Moore RC, Lee T, Theunissen FE. Noise-invariant Neurons in the Avian Auditory Cortex: Hearing the Song in Noise. PLoS Computational Biology. 2013 Mar; 9(3):e1002942. https://dx.plos.org/10.1371/journal.pcbi.1002942, doi: 10.1371/journal.pcbi.1002942.

Álvarez Morales L, Zamarreño T, Girón S, Galindo M. A methodology for the study of the acoustic environment of Catholic cathedrals: Application to the Cathedral of Malaga. Building and Environment. 2014 Feb; 72:102–115. https://linkinghub.elsevier.com/retrieve/pii/S0360132313003065, doi: 10.1016/j.buildenv.2013.10.015.

Nakamura S, Hiyane K, Asano F, Endo T. Sound scene data collection in real acoustical environments. Journal of the Acoustical Society of Japan (E). 1999; 20(3):225–231. http://www.jstage.jst.go.jp/article/ast1980/20/3/20_3_225/_article, doi: 10.1250/ast.20.225.

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN. Complementary control of sensory adaptation by two types of cortical interneurons. eLife. 2015 Oct; 4:e09868. https://elifesciences.org/articles/09868, doi: 10.7554/eLife.09868.

Naylor PA, Gaubitch ND. Speech Dereverberation. Springer Science & Business Media; 2010. https://www.springer.com/gp/book/9781849960557.

Nielsen JB, Dau T. Revisiting perceptual compensation for effects of reverberation in speech identification. The Journal of the Acoustical Society of America. 2010 Nov; 128(5):3088–3094. http://asa.scitation.org/doi/10.1121/1.3494508, doi: 10.1121/1.3494508.

Nábělek AK, Letowski TR, Tucker FM. Reverberant overlap- and self-masking in consonant identification. The Journal of the Acoustical Society of America. 1989 Oct; 86(4):1259–65.

Pachitariu M, Steinmetz N, Kadir S, Carandini M, Harris KD. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv. 2016; p. 061481. http://biorxiv.org/lookup/doi/10.1101/061481, doi: 10.1101/061481.

Pecka M, Zahn TP, Saunier-Rebori B, Siveke I, Felmy F, Wiegrebe L, Klug A, Pollak GD, Grothe B. Inhibiting the Inhibition: A Neuronal Network for Sound Localization in Reverberant Environments. Journal of Neuroscience. 2007 Feb; 27(7):1782–1790. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.5335-06.2007, doi: 10.1523/JNEUROSCI.5335-06.2007.

Poissant SF, Whitmal NA, Freyman RL. Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. The Journal of the Acoustical Society of America. 2006 Mar; 119(3):1606–15.

Puvvada KC, Villafañe-Delgado M, Brodbeck C, Simon JZ. Neural Coding of Noisy and Reverberant Speech in Human Auditory Cortex. bioRxiv. 2017 Jan; p. 229153. http://biorxiv.org/content/early/2017/12/04/229153.abstract, doi: 10.1101/229153.

Qin MK, Oxenham AJ. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. Ear and hearing. 2005 Oct; 26(5):451–60.

Rabinowitz NC, Willmore BDB, King AJ, Schnupp JWH. Constructing Noise-Invariant Representations of Sound in the Auditory Pathway. PLoS Biology. 2013; 11(11):e1001710. doi: 10.1371/journal.pbio.1001710.

Rabinowitz N, Willmore BB, Schnupp JH, King A. Contrast Gain Control in Auditory Cortex. Neuron. 2011; 70(6):1178–1191. doi: 10.1016/j.neuron.2011.04.030, publisher: Elsevier Inc.

Rahman M, Willmore BDB, King AJ, Harper NS. A dynamic network model of temporal receptive fields in primary auditory cortex. PLOS Computational Biology. 2019 May; 15(5):e1006618. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006618, doi: 10.1371/journal.pcbi.1006618, publisher: Public Library of Science.

Rahman M, Willmore BDB, King AJ, Harper NS. Simple transformations capture auditory input to cortex. Proceedings of the National Academy of Sciences. 2020 Nov; 117(45):28442–28451. http://www.pnas.org/lookup/doi/10.1073/pnas.1922033117, doi: 10.1073/pnas.1922033117.

Rakerd B, Hartmann WM. Localization of noise in a reverberant environment. In: Pressnitzer D, de Cheveigné A, McAdams S, Collet L, editors. *Auditory Signal Processing* New York, NY: Springer New York; 2005.p. 413–421. http://link.springer.com/10.1007/0-387-27045-0_51, doi: 10.1007/0-387-27045-0_51.

Robinson BL, Harper NS, McAlpine D. Meta-adaptation in the auditory midbrain under cortical influence. Nature Communications. 2016 Dec; 7(1):13442. http://www.nature.com/articles/ncomms13442, doi: 10.1038/ncomms13442.

Sahani M, Linden JF. How Linear are Auditory Cortical Responses? In: Becker S, Thrun S, Obermayer K, editors. *Advances in Neural Information Processing Systems 15* MIT Press; 2003.p. 125–132. http://papers.nips.cc/paper/2335-how-linear-are-auditory-cortical-responses.pdf.

Sakai H, Sato Si, Ando Y. Orthogonal acoustical factors of sound fields in a forest compared with those in a concert hall. The Journal of the Acoustical Society of America. 1998 Sep; 104(3):1491–1497. http://asa.scitation.org/doi/10.1121/1.424360, doi: 10.1121/1.424360.

Sayles M, Winter IM. Reverberation Challenges the Temporal Representation of the Pitch of Complex Sounds. Neuron. 2008 Jun; 58(5):789–801. https://linkinghub.elsevier.com/retrieve/pii/S0896627308003036, doi: 10.1016/j.neuron.2008.03.029.

Schnupp JW, Mrsic-Flogel TD, King AJ. Linear processing of spatial cues in primary auditory cortex. Nature. 2001 Nov; 414(6860):200–204. doi: 10.1038/35102568.

Schwartz O, Pillow JW, Rust NC, Simoncelli EP. Spike-triggered neural characterization. Journal of Vision. 2006 Jul; 6(4):13. http://jov.arvojournals.org/article.aspx?doi=10.1167/6.4.13, doi: 10.1167/6.4.13.

Schweitzer HC. Reducing the Negative Effects of Reverberation in Hearing Aid Processing. Hearing Review. 2003 Nov; p. 1–5. https://www.hearingreview.com/hearing-products/accessories/components/reducing-the-negative-effects-of-reverberation-in-hearing-aid-processing.

Shinn-Cunningham B, Kawakyu K. Neural representation of source direction in reverberant space. In: *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)* New Paltz, NY, USA: IEEE; 2003. p. 79–82. http://ieeexplore.ieee.org/document/1285824/, doi: 10.1109/ASPAA.2003.1285824.

Shinn-Cunningham B. Learning Reverberation: Considerations for Spatial Auditory Displays. Proceedings of the 2000 International Conference on Auditory Display. 2000; (April):126–134. doi: 10.1.1.22.5056.

Shinn-Cunningham BG, Lin IF, Streeter T. Trading Directional Accuracy for Realism in a Virtual Auditory Display. In: Las Vegas; 2005. p. 11. https://www.cmu.edu/dietrich/psychology/shinn/publications/pdfs/2005/2005vrinthcii_shinn.pdf.

Shinn-Cunningham BG, Desloge JG, Kopco N. Empirical and modeled acoustic transfer functions in a simple room: effects of distance and direction. In: *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)* New Platz, NY, USA: IEEE; 2001. p. 183–186. http://ieeexplore.ieee.org/document/969573/, doi: 10.1109/ASPAA.2001.969573.

Singer Y, Teramoto Y, Willmore BD, Schnupp JW, King AJ, Harper NS. Sensory cortex is optimized for prediction of future input. eLife. 2018; 7. doi: 10.7554/eLife.31557.

Slama MCC, Delgutte B. Neural Coding of Sound Envelope in Reverberant Environments. Journal of Neuroscience. 2015 Mar; 35(10):4452–4468. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3615-14.2015, doi: 10.1523/JNEUROSCI.3615-14.2015.

Spitzer MW, Bala ADS, Takahashi TT. A Neuronal Correlate of the Precedence Effect Is Associated With Spatial Selectivity in the Barn Owl's Auditory Midbrain. Journal of Neurophysiology. 2004 Oct; 92(4):2051–2070. https://www.physiology.org/doi/10.1152/jn.01235.2003, doi: 10.1152/jn.01235.2003.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network: Computation in Neural Systems. 2001 Jan; 12(3):289–316. https://doi.org/10.1080/net.12.3.289.316, doi: 10.1080/net.12.3.289.316, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/net.12.3.289.316.

Tollin DJ, Populin LC, Yin TCT. Neural Correlates of the Precedence Effect in the Inferior Colliculus of Behaving Cats. Journal of Neurophysiology. 2004 Dec; 92(6):3286–3297. https://www.physiology.org/doi/10.1152/jn.00606.2004, doi: 10.1152/jn.00606.2004.

Traer J, McDermott JH. Statistics of natural reverberation enable perceptual separation of sound and space. Proceedings of the National Academy of Sciences. 2016 Nov; 113(48):E7856–E7865. http://www.pnas.org/lookup/doi/10.1073/pnas.1612524113, doi: 10.1073/pnas.1612524113.

Trivedi U, Dieckman E, Xiang N. Reciprocal maximum-length and related sequences in the generation of natural, spatial sounding reverberation. The Journal of the Acoustical Society of America. 2009 Apr; 125(4):2735–2735. http://asa.scitation.org/doi/10.1121/1.4784525, doi: 10.1121/1.4784525.

Turner RE. Statistical Models for Natural Sounds. PhD Thesis, Gatsby Computational Neuroscience Unit, UCL; 2010.

Watkins PV, Barbour DL. Specialized neuronal adaptation for preserving input sensitivity. Nature Neuroscience. 2008 Nov; 11(11):1259–1261. http://www.nature.com/articles/nn.2201, doi: 10.1038/nn.2201.

Wen B, Wang GI, Dean I, Delgutte B. Dynamic Range Adaptation to Sound Level Statistics in the Auditory Nerve. Journal of Neuroscience. 2009 Nov; 29(44):13797–13808. https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.5610-08.2009, doi: 10.1523/JNEUROSCI.5610-08.2009.

Willmore BDB, Schoppe O, King AJ, Schnupp JWH, Harper NS. Incorporating Midbrain Adaptation to Mean Sound Level Improves Models of Auditory Cortical Processing. Journal of Neuroscience. 2016 Jan; 36(2):280–289. http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.2441-15.2016, doi: 10.1523/JNEUROSCI.2441-15.2016.

Xia J, Xu B, Pentony S, Xu J, Swaminathan J. Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners. The Journal of the Acoustical Society of America. 2018 Mar; 143(3):1523–1533. http://asa.scitation.org/doi/10.1121/1.5026788, doi: 10.1121/1.5026788.

Yin T. Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. The Journal of Neuroscience. 1994 Sep; 14(9):5170–5186. http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.14-09-05170.1994, doi: 10.1523/JNEUROSCI.14-09-05170.1994.

Yoshioka T, Sehr A, Delcroix M, Kinoshita K, Maas R, Nakatani T, Kellermann W. Making Machines Understand Us in Reverberant Rooms: Robustness Against Reverberation for Automatic Speech Recognition. IEEE Signal Processing Magazine. 2012 Nov; 29(6):114–126. http://ieeexplore.ieee.org/document/6296524/, doi: 10.1109/MSP.2012.2205029.

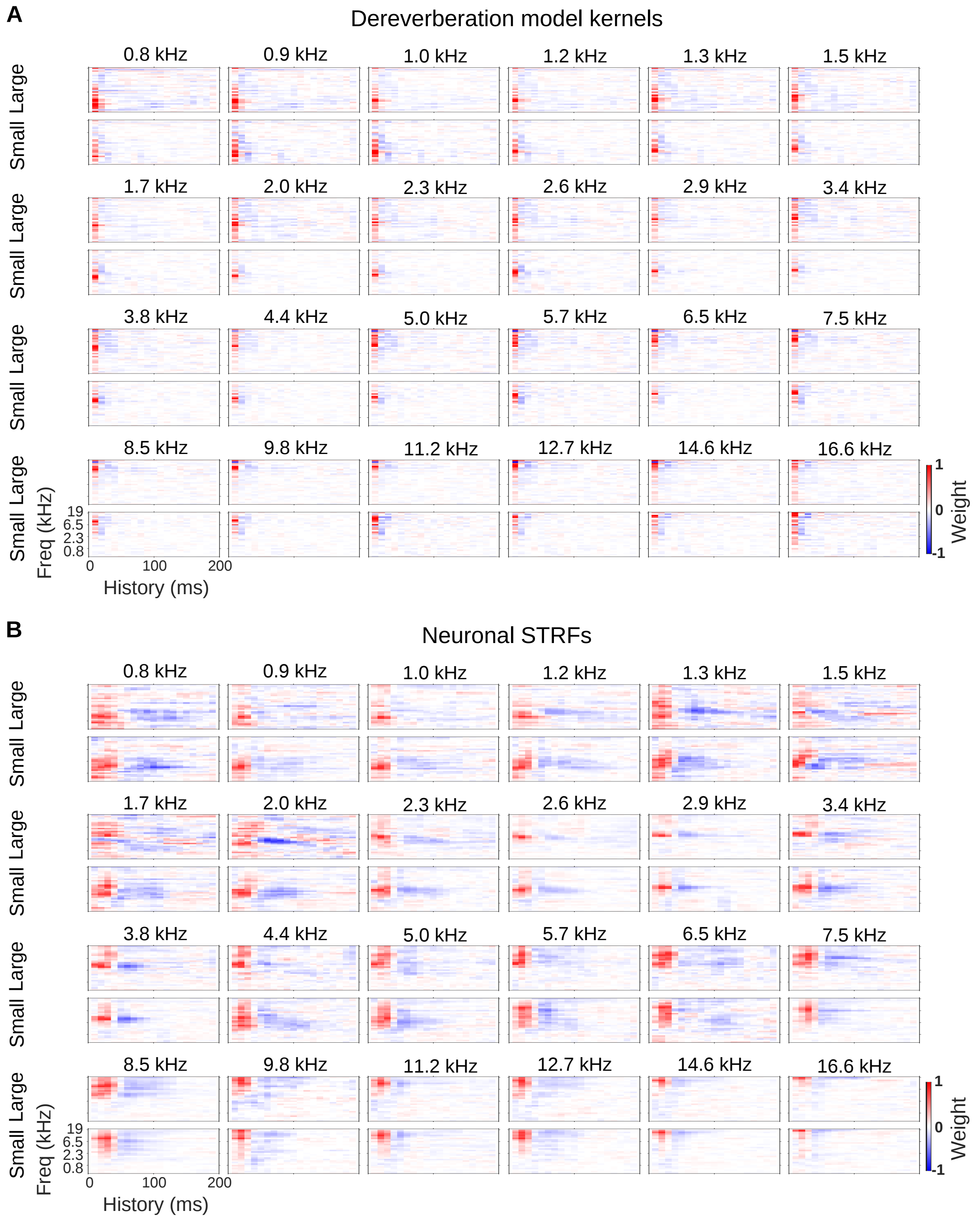

Figure 2–Figure supplement 1. Model kernels and neuronal STRFs across frequency channels.

A, Model kernels arranged by the anechoic frequency that they were trained to estimate. For each anechoic frequency, the top row shows the kernel for the large room condition, and the bottom row shows the kernel for the small room condition. In each plot, frequency is on the vertical axis and history on the horizontal. **B**, Neuronal STRFs arranged by best frequency, the frequency in the STRF with the largest weight. The STRFs of all neural units with the same best frequency were averaged to produce these plots. Plots are arranged as in A.

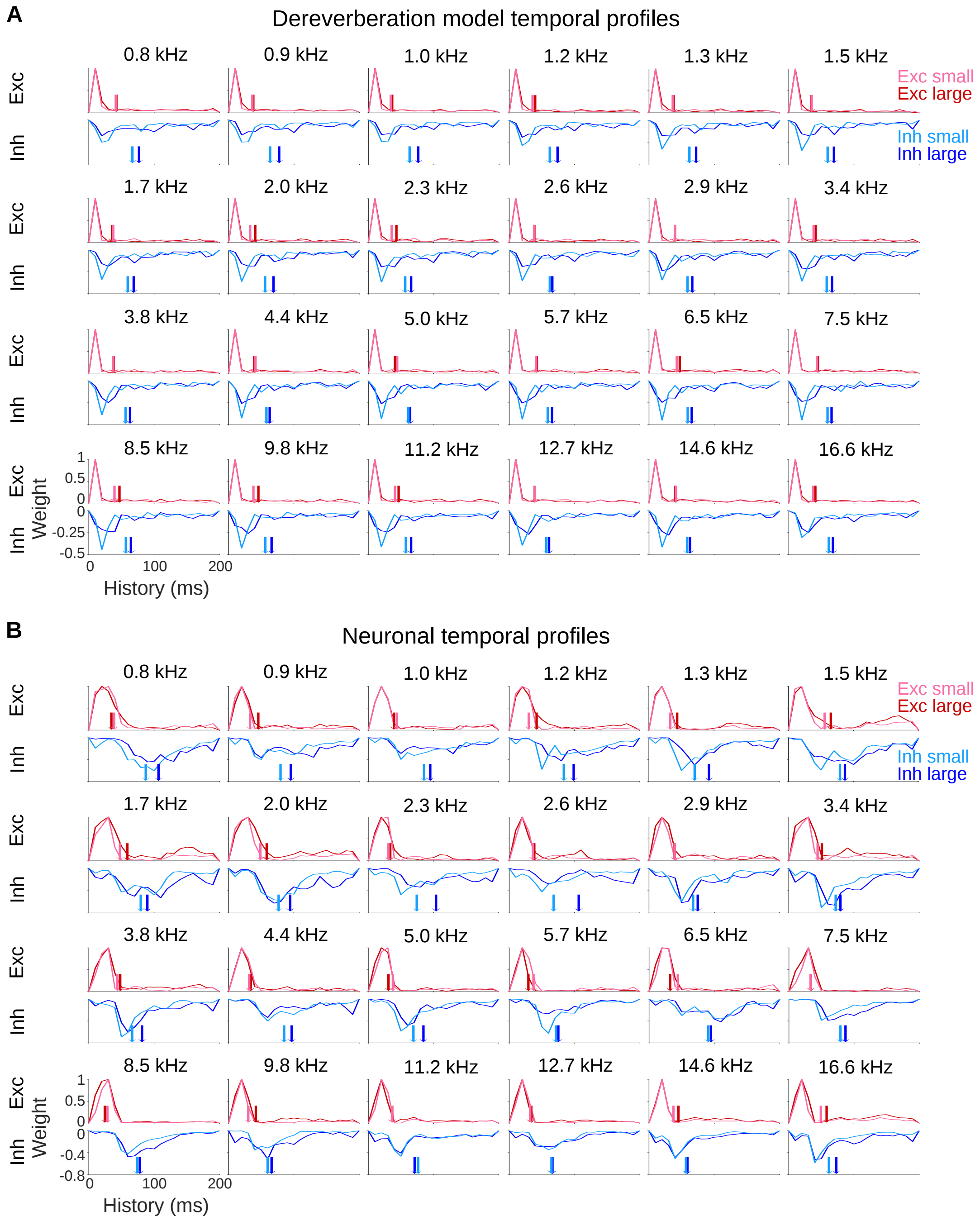

Figure 2–Figure supplement 2. Model and neuronal temporal profiles across frequency channels.

A, Temporal profiles of the excitatory (top rows) and inhibitory (bottom rows) weights of the model kernels, plotted as in Figure 2B. The estimated anechoic frequency channel is indicated above each pair of plots, as in Figure 2-Figure supplement 1A. The color code is as in Figure 2B: pink = small room excitation; red = large room excitation; cyan = small room inhibition; blue = large room inhibition. The center of mass (COM) values for the excitation and the inhibition in each room are indicated by the colored arrows. For each anechoic frequency, each temporal profile was normalized by dividing by the maximum value for the excitatory temporal profile of the same room. **B**, Temporal profiles of the excitatory and inhibitory components of the averaged neuronal STRFs shown in Figure 2-Figure supplement 1B, plotted and normalized as for the model kernels in A.

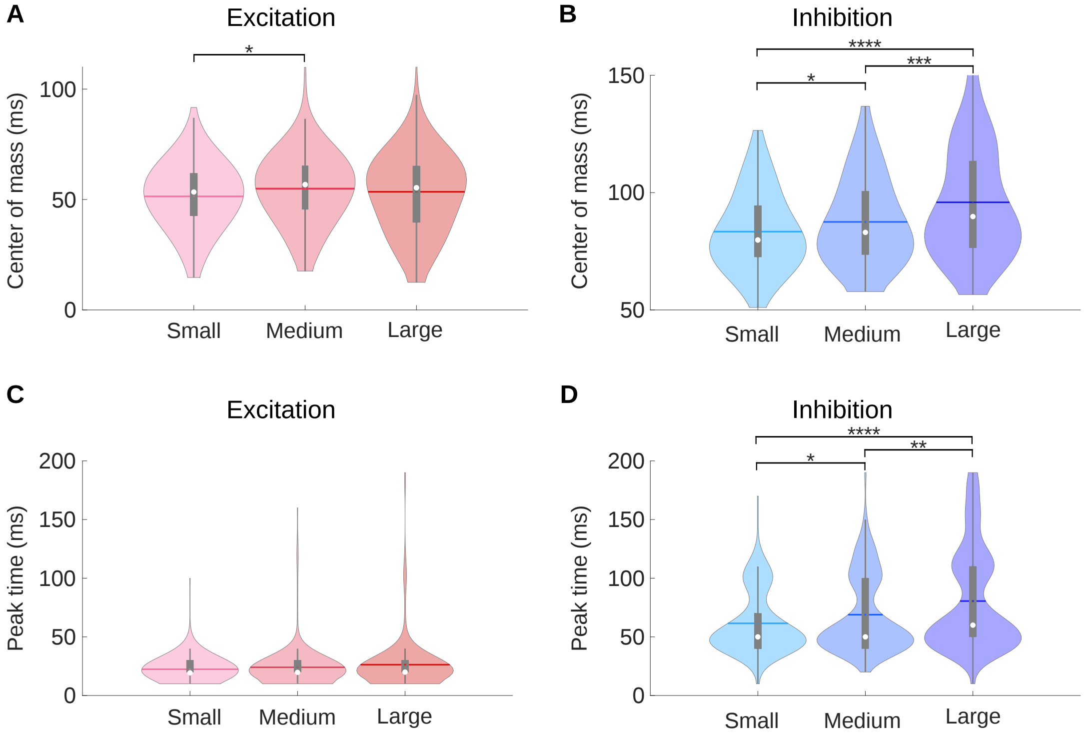

Figure 3–Figure supplement 1. A medium room condition shows intermediate center of mass and peak time values compared to the small and large room conditions.

A, Violin plots for the center of mass (COM^+) of the excitatory fields of the neuronal STRFs for the small, medium and large room conditions computed. **B**, Same as A, but here the violin plots show the center of mass (COM^-) of the inhibitory fields for the neuronal STRFs. **C**, Violin plots for the peak time of the excitatory fields (PT^+). **D**, The same data as C, but here the violin plots show the peak time (PT^-) of the inhibitory fields. In all violin plots, the white dot represents the median, the horizontal thick line the mean, the thick gray lines the interquartile range, the thin gray lines 1.5x interquartile range, and the colored shaded area represents the distribution. The results of Kruskal–Wallis tests followed by multiple comparisons using Fisher's least significant difference (LSD) procedure are indicated above the bars in A, B and D: $^*p < 0.05,^{**}\ p < 0.01,^{***}\ p < 0.001,^{****}\ p < 0.0001$.

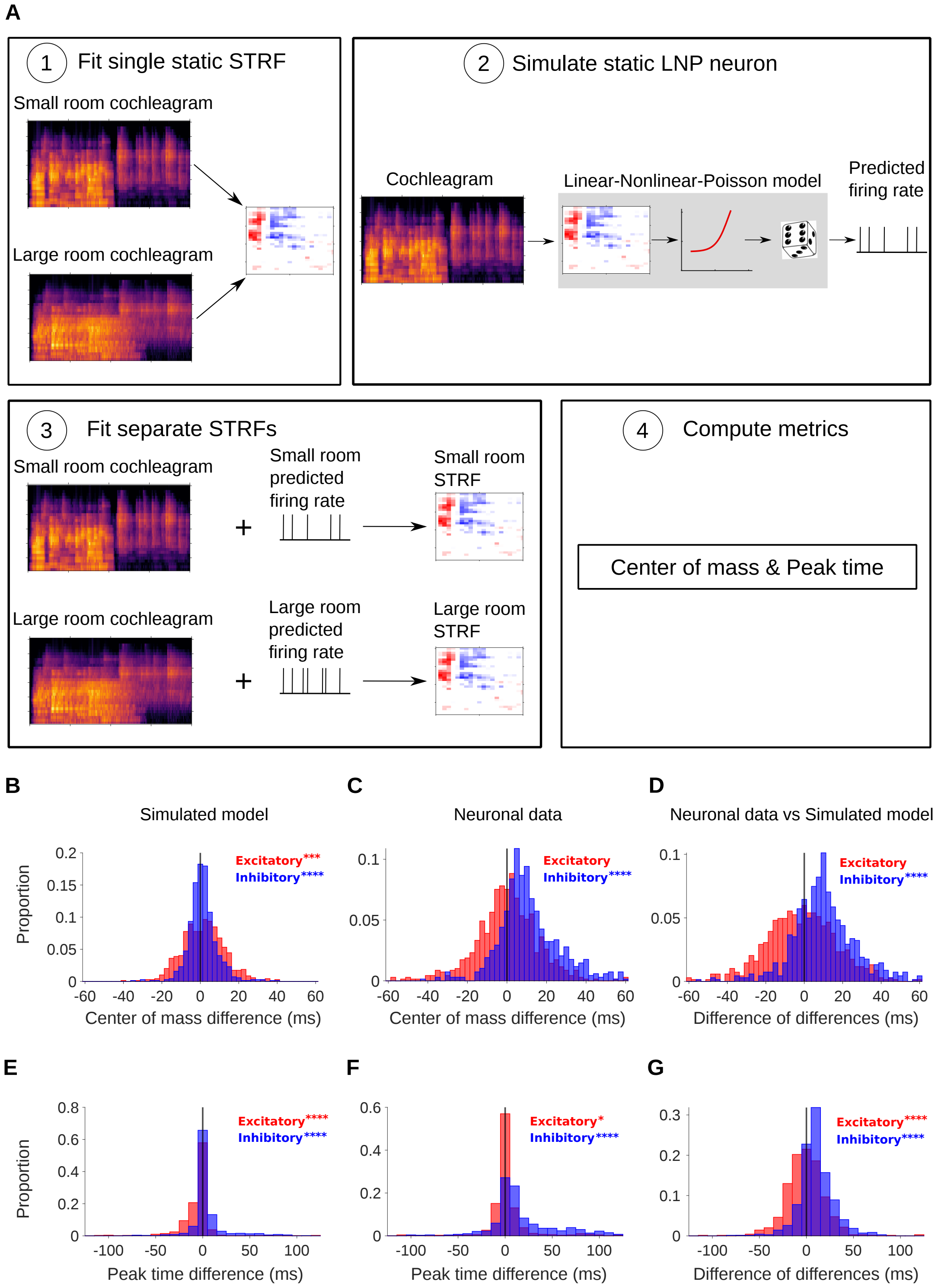

Figure 3–Figure supplement 2. Simulated neurons suggest a role for adaptation in cortical dereverberation. See next page for caption.

Figure 3–Figure supplement 2. Simulated neurons suggest a role for adaptation in cortical dereverberation.
To confirm that STRF differences between rooms were genuinely a result of adaptation, we simulated the recorded neurons using a non-adaptive linear-nonlinear-Poisson model and compared STRF measures of the simulated responses with those of the real neuronal STRFs in the different room conditions. **A**, The simulated neurons were made in the following way: 1) We fitted a single STRF for each neuron using the combined data from the small and large rooms; 2) We used this STRF along with a fitted non-linearity and a Poisson noise model to generate the simulated firing rate for the small and large rooms separately; 3) Using the small and large room cochleagrams and simulated firing rates, we fitted separate STRFs for the two conditions; 4) We computed the center of mass and peak time metrics as before. **B**, Difference in center of mass between the large and small room conditions (large - small room) for the simulated model neurons. The COM^{-} values (blue) were slightly larger in the large room, median difference = 0.90ms, and the COM^{+} values (red) were slightly elevated too, median difference = 1.7ms. **C**, Reproduction of Figure 3B showing the difference in center of mass of neuronal STRF components between the large and small room conditions (large - small room). The COM^{-} values increased in the larger room (median difference = 9.3ms), whereas COM^{+} did not differ significantly (median difference = 0.32ms). **D**, The center of mass differences shown in B and C were subtracted for each unit and plotted as the resulting difference of differences (real cortical unit - simulated model neuron). The COM^{-} differences between rooms were consistently larger in the neuronal data (median difference = 9.3ms), while the COM^{+} differences did not differ significantly (median difference = -1.1ms). **E**, Difference in peak time between the large and small rooms (large - small) for the simulated model neurons. The PT^{-} median difference = 0ms and the PT^{+} median difference = 0ms. **F**, Reproduction of Figure 3D showing the difference in peak time between the large and small rooms (large - small), calculated from neuronal STRFs. The PT^{-} values were larger in the large room (median difference = 10ms). PT^{+} did differ significantly between the rooms, but with a median difference = 0ms. **G**, Histogram of the difference in peak time room differences between the neural units and corresponding simulated model neurons (neural unit - simulated model neuron), plotted as in D above. The PT^{-} values were consistently larger in the large room for the neuronal data vs the simulated model neurons (median difference = 10ms). PT^{+} did significantly differ, but the median difference = 0ms. Asterisks indicate the significance of Wilcoxon signed-rank tests: $^{****}p < 0.0001$, $^{***}p < 0.001$, $^{*}p < 0.05$.

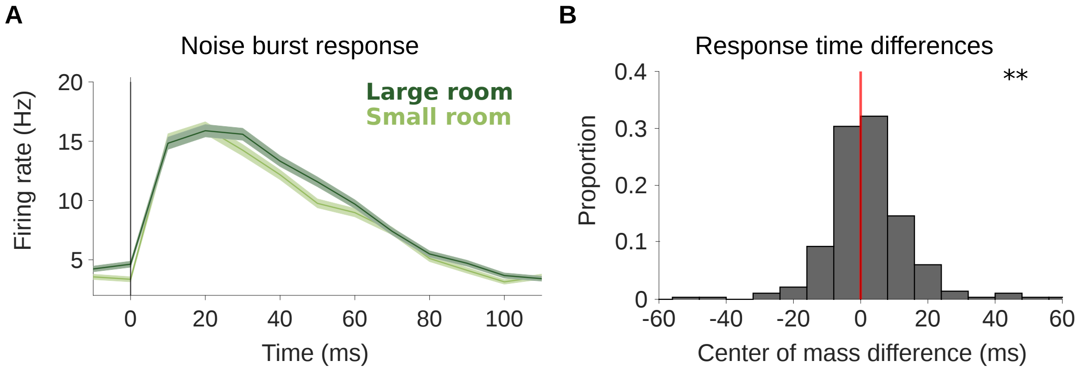

Figure 3–Figure supplement 3. Neural response to noise probe shows slower adaptation in the more reverberant condition.

A, Average firing rate across all cortical units in response to a noise burst that was embedded within the reverberant stimuli. Responses to the noise within the small (light green) and large (dark green) rooms are plotted separately. Shaded areas show ±SEM across units. The vertical line indicates the noise onset. **B**, Histogram of the difference in center of mass of the neuronal response to the noise probe (shown in A) between the two room conditions (large - small room). The center of mass shifted to a later time in the larger room (median difference = 1.0ms). Asterisks indicate significance of a Wilcoxon signed-rank test: $^{**}p < 0.01$.

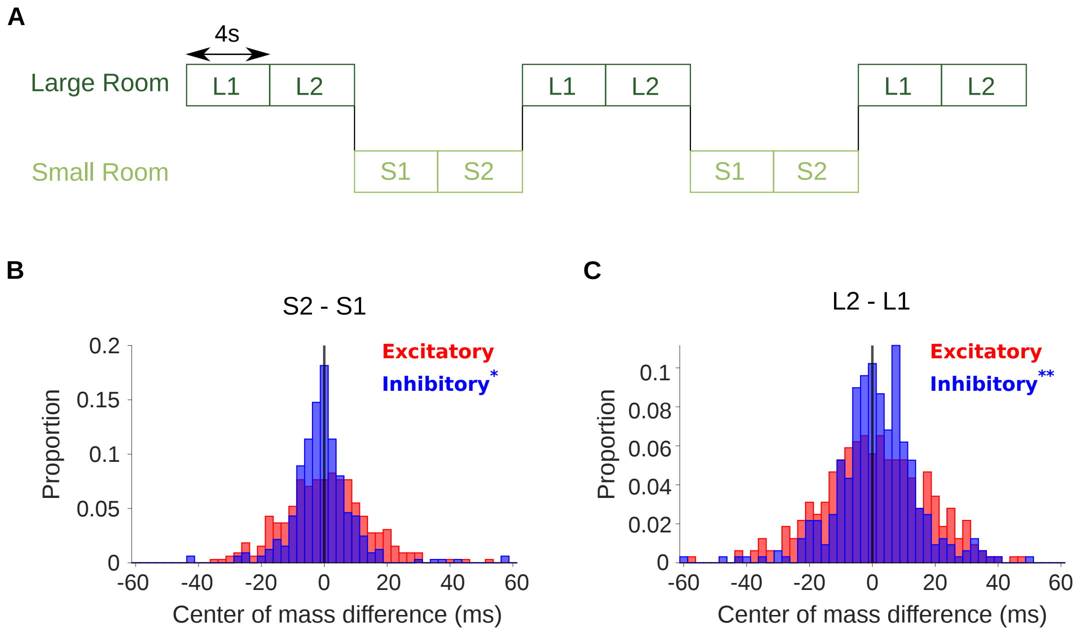

Figure 3–Figure supplement 4. Adaptation to reverberation is confirmed using stimuli that switch between the small and large room.

A, Schematic shows the structure of the stimulus, which switched between the large (dark green) and small room (light green) conditions. Letters indicate the reverberant condition in each stimulus block (S: small room, L: large room). Each 8s block within a given room condition was divided for analysis into an early (S1,L1) and late (S2,L2) period. STRFs were fitted to the data from each of the 4 periods independently (S1, S2, L1, L2). **B**, Difference in center of mass of inhibitory (COM^-, blue) and excitatory (COM^+, red) STRF components between the late and early time period of the small room stimuli (S2 - S1, see A). The COM^- decreased in S2 relative to S1 with a median difference = -0.9ms; COM^+ did not differ significantly, median difference = 0.52ms. **C**, Center of mass difference plotted as in B, but for the large room stimuli (L2 - L1). The COM^- values were larger in L2 relative to L1, median difference = 1.5ms, while the COM^+ values were not significantly different, median difference = 0.8ms. Asterisks indicate the significance of Wilcoxon signed-rank tests: $^{**}p < 0.01, ^{*}p < 0.05$.