

A cerebellum-like circuit in the auditory system cancels responses to self-generated sounds

Shobhit Singla¹, Conor Dempsey¹, Richard Warren¹, Armen G Enikolopov² & Nathaniel B Sawtell¹

The dorsal cochlear nucleus (DCN) integrates auditory nerve input with a diverse array of sensory and motor signals processed in circuitry similar to that of the cerebellum. Yet how the DCN contributes to early auditory processing has been a longstanding puzzle. Using electrophysiological recordings in mice during licking behavior, we show that DCN neurons are largely unaffected by self-generated sounds while remaining sensitive to external acoustic stimuli. Recordings in deafened mice, together with neural activity manipulations, indicate that self-generated sounds are cancelled by non-auditory signals conveyed by mossy fibers. In addition, DCN neurons exhibit gradual reductions in their responses to acoustic stimuli that are temporally correlated with licking. Together, these findings suggest that DCN may act as an adaptive filter for cancelling self-generated sounds. Adaptive filtering has been established previously for cerebellum-like sensory structures in fish, suggesting a conserved function for such structures across vertebrates.

The first central stage of mammalian auditory processing occurs in the dorsal and ventral divisions of the cochlear nucleus¹. Based on similarities in their evolution, development, gene expression patterns and anatomical arrangement, the DCN is considered to belong to a class of so-called cerebellum-like sensory structures^{2–6}. Other cerebellum-like structures include the first central stages of electrosensory and mechanosensory lateral line processing in several groups of fish. Many cell and fiber types are shared among these cerebellum-like structures and the cerebellum itself, including mossy fibers, granule cells, parallel fibers, Golgi cells, molecular layer interneurons and Purkinje or Purkinje-like cells. A hallmark of the circuitry of cerebellum-like sensory structures is the integration of direct input from peripheral sensory receptors (for example, electroreceptors in the case of cerebellum-like structures in fish and auditory nerve fibers in the case of DCN) with a diverse array of sensory and motor signals conveyed by a granule cell–parallel fiber system.

A primary site of this integration in DCN is the fusiform cell. Fusiform cells are also the main output cell of DCN and project to higher stages of auditory processing such as the inferior colliculus. The basilar dendrites of fusiform cells are contacted by auditory nerve fibers, which form a tonotopic map in the deep layer of DCN (Supplementary Fig. 1)^{1,6}. Their apical dendrites extend into a superficial molecular layer, where they are contacted by parallel fibers. Parallel fibers arise from granule cells located in so-called granule cell domains (GCDs) around the margins of the nucleus and cross through different tonotopic regions of DCN⁴. Granule cells receive a wide variety of signals, both auditory and non-auditory, from mossy fibers originating in a number of different brain regions⁶. Parallel fibers, but not auditory nerve fiber synapses, have been shown to exhibit forms of long-term associative synaptic plasticity *in vitro*^{7–9}. Though previous *in vivo* studies of DCN have extensively characterized auditory response properties in anesthetized or decerebrate

animals¹⁰, much less is known about the functional significance of its cerebellum-like circuitry^{11–13}.

Potential clues regarding DCN function come from studies of cerebellum-like structures associated with electrosensory processing in fish. Such studies have shown that anti-Hebbian synaptic plasticity acting on proprioceptive, electrosensory and motor corollary discharge signals conveyed by parallel fibers serve to cancel principal cell responses to self-generated electrosensory inputs—for example, those arising from the fish's own movements or electromotor behavior^{14,15}. Cancellation of self-generated electrosensory inputs allows externally generated, behaviorally relevant stimuli to be processed more effectively. Motivated by these past findings in fish, we set out to test the hypothesis that the cerebellum-like circuitry of the mammalian DCN functions to cancel responses to self-generated sounds.

To this end, we developed a preparation to study neural responses to self-generated sounds in the auditory brainstem of awake, behaving mice. We chose licking behavior because it is stereotyped and repetitive, can be elicited in head-fixed animals during electrophysiological recordings and, as we demonstrate, generates sounds that are a potential source of interference for the mouse auditory system.

RESULTS

DCN neurons respond preferentially to external versus self-generated sounds

We found that rhythmic licking generated sounds within the hearing range of the mouse and that such sounds exhibited stereotyped spectral and temporal profiles that were similar across mice (Fig. 1a, Supplementary Fig. 2 and Supplementary Video 1). The temporal profile of the licking sound is shown by the root mean squared (r.m.s.) amplitude of the microphone recording aligned to tongue contact with the lick spout (Fig. 1a). As can be seen in both the spectrogram and r.m.s. amplitude trace, licking sounds typically consist of

¹Department of Neuroscience and Kavli Institute for Brain Science, Columbia University Medical Center, New York, New York, USA. ²Department of Biological Sciences, Columbia University, New York, New York, USA. Correspondence should be addressed to N.B.S. (ns2635@cumc.columbia.edu).

Received 15 August 2016; accepted 19 April 2017; published online 22 May 2017; doi:10.1038/nn.4567

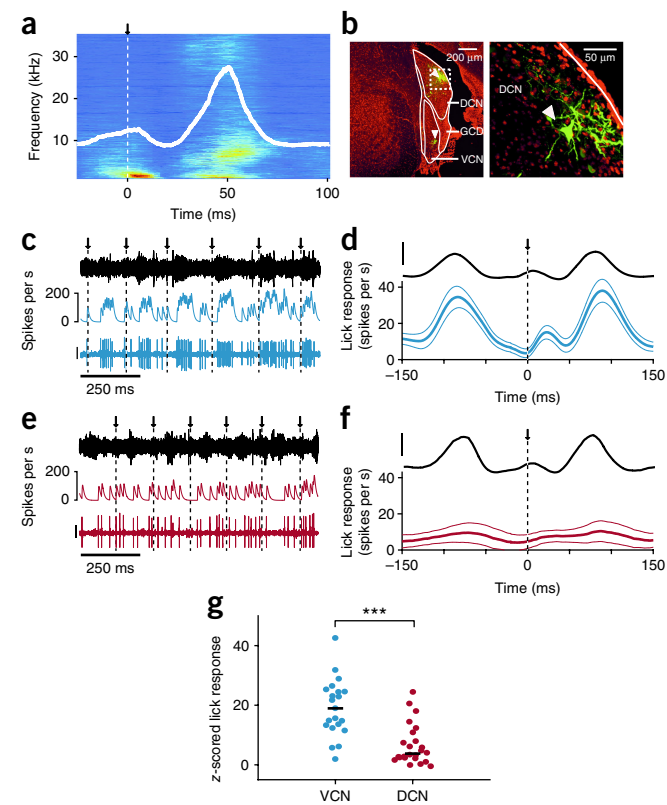


Figure 1 Self-generated sounds strongly affect VCN but not DCN neurons. (a) Average spectrogram of self-generated sounds during licking for a representative mouse (of $n = 20$). Arrow and dotted line indicate time of tongue contact with the lick spout. Solid white line indicates the r.m.s. amplitude of the microphone recording. (b) Left, dextran-conjugated Alexa 594 labeling (green) at recording sites in DCN and VCN (arrowheads). DAPI, red. Right, higher magnification of dashed white box on left showing a labeled fusiform cell (arrowhead). (c) Example VCN unit response during licking. Arrows and dotted lines indicate times of tongue contact with the lick spout. Traces represent the microphone recording (top), smoothed firing rate (middle) and VCN unit recording (bottom; scale: 30 μ V). (d) Top, average r.m.s. amplitude of the licking sound during VCN unit recordings (scale bar: 1 a.u.). Bottom, average VCN lick-triggered firing rate ($n = 21$). Thin lines are s.e.m. (e) Example DCN unit response during licking. (f) Top, average r.m.s. amplitude of the licking sound during DCN unit recordings. Bottom, average lick-triggered responses of all DCN units ($n = 25$), excluding those exhibiting complex spikes. Compared to VCN units, DCN units exhibited smaller temporal modulations related to licking (peak-to-trough firing rate for VCN: 43.8 ± 26.9 Hz, $n = 21$; for DCN: 19.7 ± 19.9 Hz, $n = 25$, mean and s.d., $P = 0.0005$, Wilcoxon rank sum test). Display same as in d. (g) z-scored lick responses (see Online Methods) were significantly smaller in DCN than in VCN units (*** $P = 0.00002$, Wilcoxon rank sum test). Median responses are indicated by solid lines.

an early component that begins before contact as well as a larger late component that peaks ~ 50 ms after contact, during tongue retraction (Fig. 1a and Supplementary Fig. 2).

To determine whether licking sounds evoke neural responses that could interfere with auditory processing and, if so, whether such responses are cancelled out in the DCN, we compared neural activity during licking in well-isolated single units in the ventral cochlear nucleus (VCN) and DCN. Since VCN receives direct auditory nerve input but lacks cerebellum-like circuitry, we hypothesized that VCN units would respond to acoustic stimuli regardless of whether they are self- or externally generated. Recording locations were judged based

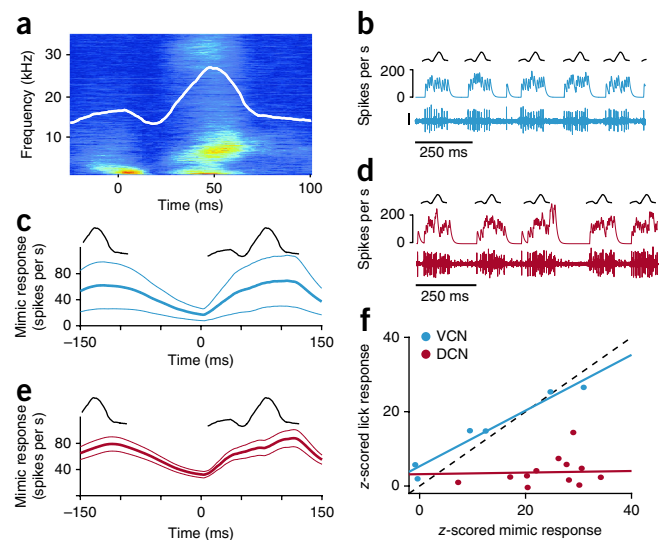


Figure 2 Responses to self-generated versus external sounds in VCN and DCN. (a) Spectrogram of the lick mimic generated from microphone recordings from 5 mice (Online Methods). Overlaid white line represents the r.m.s. amplitude. (b) Example VCN unit response to the lick mimic. Same unit as in Figure 1c. Traces represent a schematic of the r.m.s. of the mimic (top), smoothed firing rate (middle) and the VCN unit recording (bottom; scale bar: 30 μ V). (c) Top, schematic of the r.m.s. of the lick mimic. Bottom, average VCN unit response to the lick mimic ($n = 6$). Thin lines are s.e.m. The lick mimic was delivered at 12 dB SPL in all experiments. (d, e) Same as b, c but for DCN unit responses to the mimic ($n = 13$). Traces in d are from same unit shown in Figure 1e. VCN and DCN unit responses to the mimic were not significantly different ($P = 0.32$, Wilcoxon rank sum test). (f) Responses to licking were highly correlated to those observed in the same units to the lick mimic for VCN ($n = 6$, $P = 0.00081$, $r = 0.95$, linear regression t -test) but not DCN recordings ($n = 13$, $P = 0.79$, $r = 0.0007$, linear regression t -test).

on characteristic reversals of tonotopy at the DCN/VCN border^{16,17} and verified by iontophoresis of a dextran-conjugated fluorescent dye (Fig. 1b, Supplementary Fig. 3 and Online Methods). Though unambiguous criteria for linking physiological response properties with morphological cell classes have not yet been established for the awake mouse DCN¹⁸, several properties of the recorded units indicate that they likely correspond to fusiform cells, including their high spontaneous firing rates and purely excitatory responses to acoustic stimuli (Supplementary Fig. 4 and Online Methods)^{19–23}. Units exhibiting complex spikes, putative cartwheel cell interneurons^{24,25} (Supplementary Fig. 1), were also encountered and analyzed separately. Results for complex-spiking units are reported in a separate section below.

Consistent with the possibility that licking behavior causes significant self-generated sounds, VCN units exhibited an overall firing rate elevation during licking, as well as firing rate modulations that tracked the r.m.s. amplitude of the licking sound (Fig. 1c,d). In contrast, DCN units exhibited substantially weaker firing rate modulations during licking (Fig. 1e–g). Though these results are consistent with cancellation of self-generated sounds in DCN, an alternative explanation is that differences between VCN and DCN responses during licking are due to systematic differences in their auditory response properties.

We evaluated this possibility in a subset of VCN and DCN units by comparing activity during licking to activity during delivery of an externally generated acoustic stimulus with temporal and spectral properties that roughly matched the licking sounds recorded across mice (Fig. 2a and Online Methods). This stimulus is referred

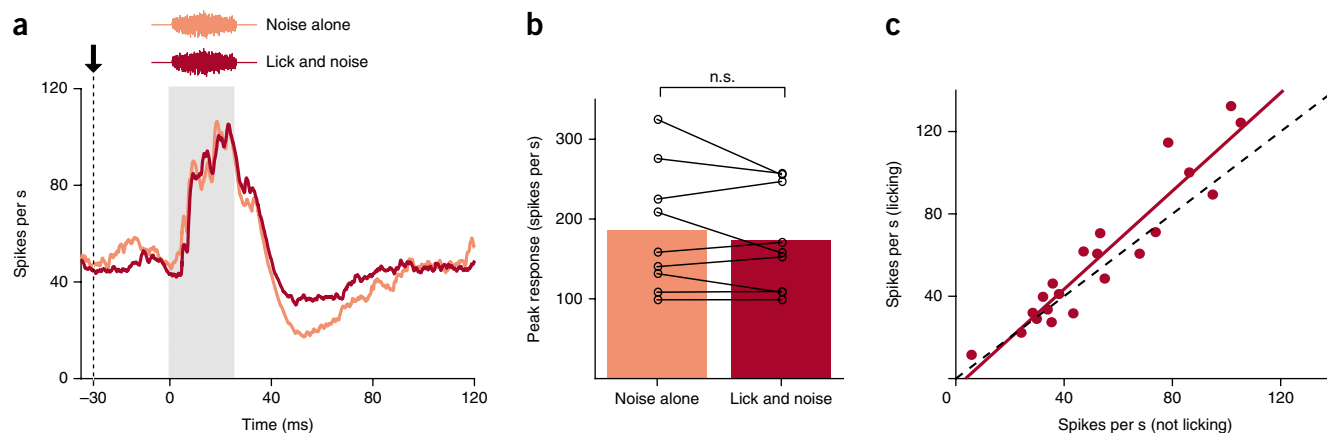


Figure 3 DCN responses to acoustic stimuli are not suppressed during licking. **(a)** Example DCN unit response to an acoustic stimulus (bandpass filtered 5–15 kHz, 15 dB SPL) played while the mouse was still versus during licking. Gray bar indicates stimulus duration. **(b)** No differences in responses were observed when the mouse was still versus licking ($n = 9$; n.s., not significant, $P = 0.49$, Wilcoxon signed rank test). **(c)** Overall firing rates for DCN units were similar when the mouse was licking versus still.

to henceforth as the lick mimic and was presented outside of licking bouts, when the mouse was still. Though the match between actual sounds generated by licking and the lick mimic is not expected to be perfect—for example, due to issues such as bone conduction—this stimulus nevertheless provided a principled means of comparing auditory responses in VCN and DCN. Strong responses to the lick mimic were observed in both VCN (Fig. 2b,c) and DCN units (Fig. 2d,e). The strength of responses during licking was highly correlated with the strength of responses to the lick mimic in VCN units (Fig. 2f). This is exactly what would be expected if VCN licking responses were indeed due to self-generated sounds. In contrast, there was no significant correlation between mimic and licking responses in DCN units (Fig. 2f): even units with strong responses to the lick mimic failed to respond during licking. These observations suggest that weaker responses to licking in DCN than in VCN cannot be explained by differences in auditory sensitivity between the two regions. What then is the mechanism underlying the apparent reduction of responses to self-generated sounds during licking behavior in DCN?

One possibility is that the overall sensitivity of DCN units to sound is reduced during licking behavior. Indeed, an overall suppression of auditory responsiveness during behavior has been reported in a variety of systems^{26,27}, including the mouse auditory cortex^{28,29}. To test this, we compared DCN unit responses to an externally generated acoustic stimulus (bandpass-filtered noise 5–15 kHz, 15 dB sound pressure level (SPL)) delivered either during licking or when the mouse was still (Fig. 3a). Responses to the acoustic stimulus were indistinguishable under the two conditions (Fig. 3a,b). In addition, overall firing rates in DCN units were similar when mice were licking versus still (Fig. 3c). Together, these results are inconsistent with an overall suppression of auditory sensitivity in DCN during licking and point instead to a mechanism for selectively canceling self-generated sounds.

Non-auditory signals related to licking revealed in DCN of deafened mice

In addition to auditory nerve input, DCN receives non-auditory, behavior-related signals conveyed by mossy fibers. Previous studies of cerebellum-like structures in fish have shown that behavior-related signals conveyed by mossy fibers serve to selectively cancel self-generated electrosensory input^{14,15}. Though electrophysiological correlates of non-auditory mossy fiber inputs to DCN have been

characterized in anesthetized or decerebrate preparations—for example, using electrical stimulation of somatosensory brain regions projecting to DCN^{11–13}—responses to non-auditory inputs have not yet been demonstrated in awake, behaving animals. To isolate non-auditory responses related to licking behavior, we recorded from DCN in deafened mice ($n = 3$). Deafening (see Online Methods) was confirmed by a lack of observable behavioral responses to acoustic stimuli and by recording auditory-evoked field potentials in DCN before and after deafening (Fig. 4a). For recordings in deafened mice, we focused exclusively on units that exhibited both isolated action potentials, known as simple spikes, and brief, high-frequency bursts of action potentials, known as complex spikes (Fig. 4b). Such complex-spiking units correspond to a class of DCN interneuron known as cartwheel cells (CWCs) that share many similarities with Purkinje cells in the cerebellum (Supplementary Fig. 1)^{24,25}. CWCs lack auditory nerve input, receive massive input from parallel fibers, and inhibit fusiform cells. Our reasons for focusing on CWCs were twofold: (i) the complex spike is a distinctive electrophysiological signature of CWCs, which allowed us to be confident that we were recording in the DCN even in the absence of sound-evoked responses in the deafened mice, and (ii) CWCs provide a convenient readout of non-auditory inputs conveyed by granule cells. Granule cells themselves are too small to be reliably isolated using conventional extracellular recording techniques.

In deafened mice, 9 of 11 complex-spiking units exhibited significant simple and/or complex spike firing-rate modulations related to licking (Fig. 4c–e). The overall rate of complex spike firing also increased slightly during licking (Fig. 4f). These results indicate that DCN receives non-auditory information related to licking behavior. Granule cells provide the main excitatory input to CWCs. Hence the non-auditory, licking-related responses we observed in CWCs are likely due to signals conveyed by parallel fibers.

We also recorded from complex-spiking units in hearing mice. Most complex-spiking units exhibited simple and complex spike firing-rate modulations related both to licking (Fig. 4g) and to presentation of the mimic when the mouse was still (Fig. 4h). Prominent responses to both licking and the mimic (Fig. 4i) are consistent with the notion that CWCs receive both non-auditory and auditory signals conveyed by granule cells. This is consistent with anatomical evidence for prominent non-auditory as well as auditory input to GCDs⁶ and previous electrophysiological evidence for prominent auditory responses in complex-spiking units in awake mice³⁰.

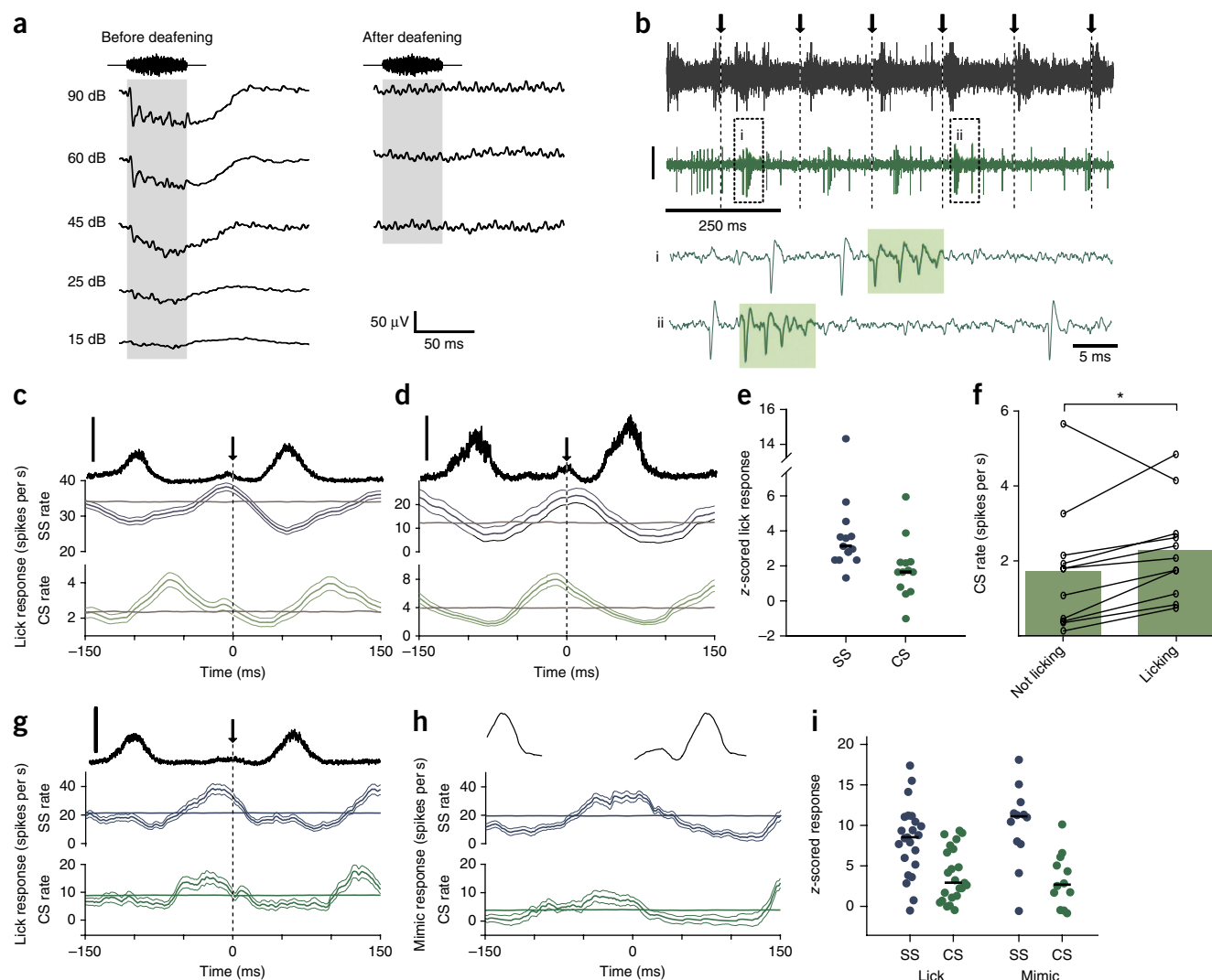


Figure 4 Non-auditory responses related to licking in DCN complex-spiking units. **(a)** Sound-evoked field potentials (50 ms, broadband noise, averaged over 50 presentations) recorded in DCN of the same mouse before (left) and after (right) surgical deafening. Note complete absence of sound-evoked field potentials after deafening. **(b)** Example DCN complex-spiking unit recorded during licking in a surgically deafened mouse. Arrows and dotted lines indicate times of tongue contact with the lick spout. Top trace, microphone recording. Second trace, extracellular voltage from a DCN complex-spiking unit (scale: 30 μ V). Bottom, expanded traces from boxed regions showing complex-spike (CS) (shaded rectangle) and simple-spike (SS) waveforms. **(c,d)** Lick-triggered SS and CS firing rates for two complex-spiking units recorded in deafened mice. Data in **c** are from same unit as example traces in **b**. Top trace (black) is the r.m.s. amplitude of the licking sound (scale bars, 1 a.u.). Thin lines are s.e.m. Gray traces show the average lick-triggered response of shuffled spike trains. **(e)** Summary of z-scored lick responses of 11 complex-spiking units recorded in 3 surgically deafened mice. Eight showed significant lick responses in their SS firing and 3 showed significant lick responses in their CS firing ($\alpha = 0.01$; see Online Methods). Median responses are indicated by solid lines. **(f)** Overall CS firing rates increased slightly during periods of licking in deafened mice ($n = 11$, $*P = 0.04$, Wilcoxon signed rank test). **(g)** Lick-triggered SS and CS firing rates for a complex-spiking unit recorded in a hearing mouse. Same display as in **c**. **(h)** Mimic-triggered SS and CS firing rates for a complex-spiking unit recorded in a hearing mouse. Same unit as shown in **g**. **(i)** Summary of licking and mimic responses in complex-spiking units recorded in hearing mice. Median responses are indicated by solid lines. Average z-score responses to licking ($n = 23$) were 8.2 ± 4.4 for SSs and 3.8 ± 4.4 for CSs. Average z-score responses to the mimic ($n = 12$) were 10.1 ± 4.9 for SSs and 3.2 ± 3.3 for CSs.

A role for the spinal trigeminal nucleus in cancelling self-generated sounds

Based on previous microstimulation and anatomical tracing studies, the spinal trigeminal nucleus (Sp5) is expected to be the main source of mossy fiber input to DCN conveying somatosensory information related to licking behavior^{12,31,32}. As expected from past studies in other mammals, injection of an anterograde viral tracer (AAV2-GFP) into mouse Sp5 resulted in labeled mossy fibers in the GCDs of DCN ($n = 3$; Fig. 5a), as well as in the cerebellum (data not shown). If non-auditory, licking-related inputs from Sp5 serve to cancel out responses

to self-generated acoustic stimuli in DCN, transiently silencing such inputs should reveal prominent licking-related responses in DCN neurons. Indeed, micropressure injection of the action potential blocker lidocaine into Sp5 led to an increase in overall firing in putative DCN output cells during licking, as well as an increased modulation of firing (Fig. 5b,d) that tracked the amplitude of the licking sound (Fig. 5b). No such changes were observed after saline injection (Fig. 5c,d). Furthermore, increases in licking responses after lidocaine injection could not be explained by differences in sensitivity to acoustic stimuli between lidocaine and saline groups (Fig. 5e), changes in

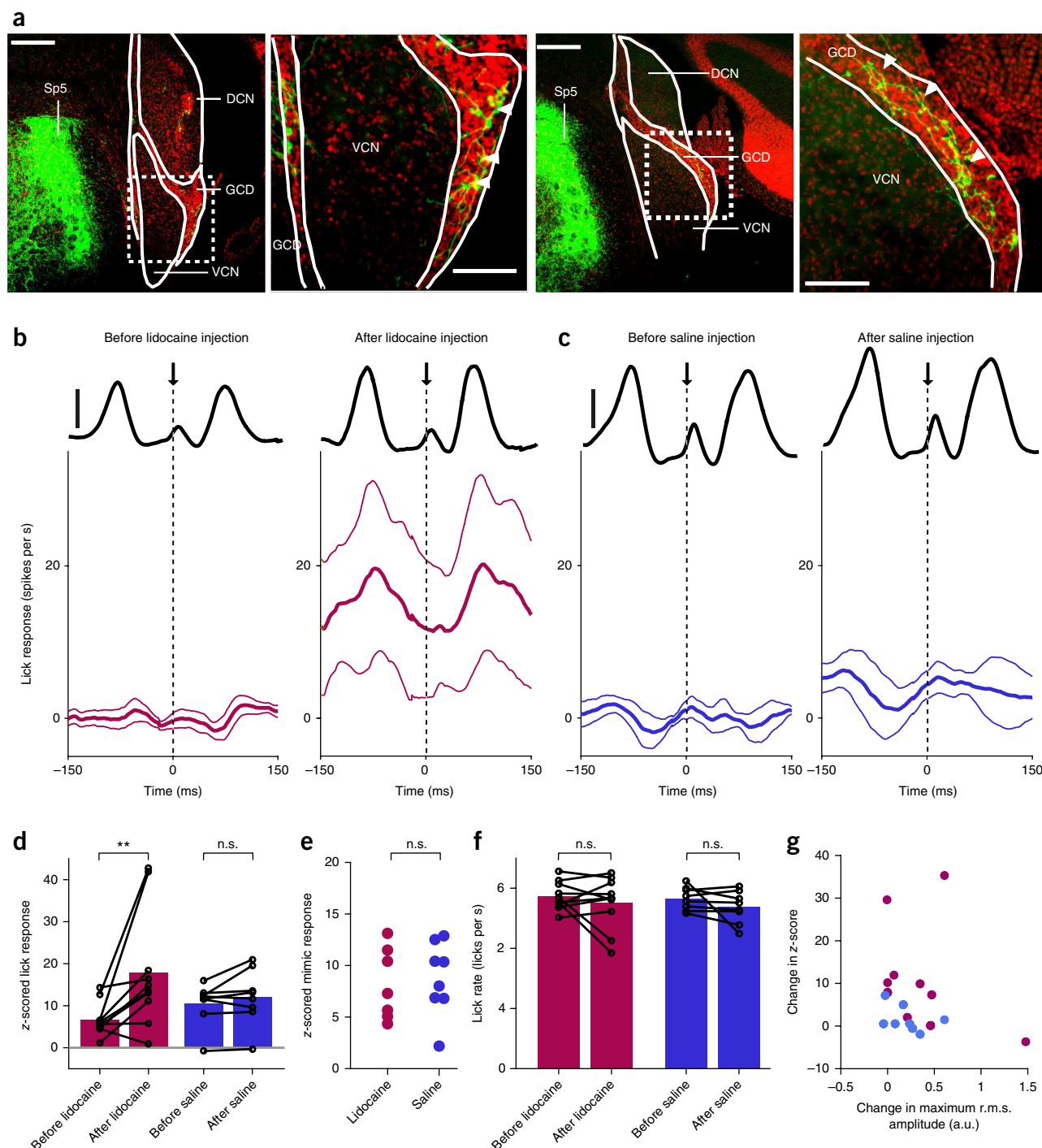


Figure 5 A role for the spinal trigeminal nucleus in cancelling self-generated sounds in DCN. **(a)** Labeled mossy fibers were observed in DCN GCDs after injection of an anterograde viral tracer (AAV2-GFP) into the ipsilateral Sp5. Transverse sections of the cochlear nucleus from two rostrocaudal locations are shown. Scale bars for left image in each pair: 200 μ m. Right in each pair, higher magnification views of areas indicated by dotted rectangle. Scale bars: 100 μ m. White arrowheads indicate labeled mossy fibers in GCDs. **(b,c)** Lick-triggered response of DCN cells before (left) and after (right) injection of lidocaine (**b**, $n = 10$) or saline (**c**, $n = 8$) into Sp5. Thin lines are s.e.m. Solid black lines show the r.m.s. amplitude of the licking sound (scale bars: 1 a.u.). **(d)** Lidocaine injection resulted in a significant increase in z-scored lick responses in DCN units (** $P = 0.0098$, Wilcoxon signed rank test, red) while no significant increases in z-scored lick responses occurred after saline injection (n.s., not significant; $P = 0.31$, Wilcoxon signed rank test, blue). **(e)** Auditory responses to the mimic were not significantly different in lidocaine and saline groups ($P = 0.87$, Wilcoxon rank sum test). **(f)** Lick rate did not differ before and after injection of lidocaine ($P = 0.77$, Wilcoxon signed rank test) or saline ($P = 0.25$, Wilcoxon signed rank test). **(g)** Changes in z-score lick responses were not correlated with changes in the maximum r.m.s. amplitude in arbitrary units (a.u.) of the licking sound after lidocaine injection (red, $P = 0.36$, linear regression t -test). Changes in the maximum r.m.s. amplitude of the licking sound did not differ between lidocaine and saline groups ($P = 0.51$, Wilcoxon rank sum test).

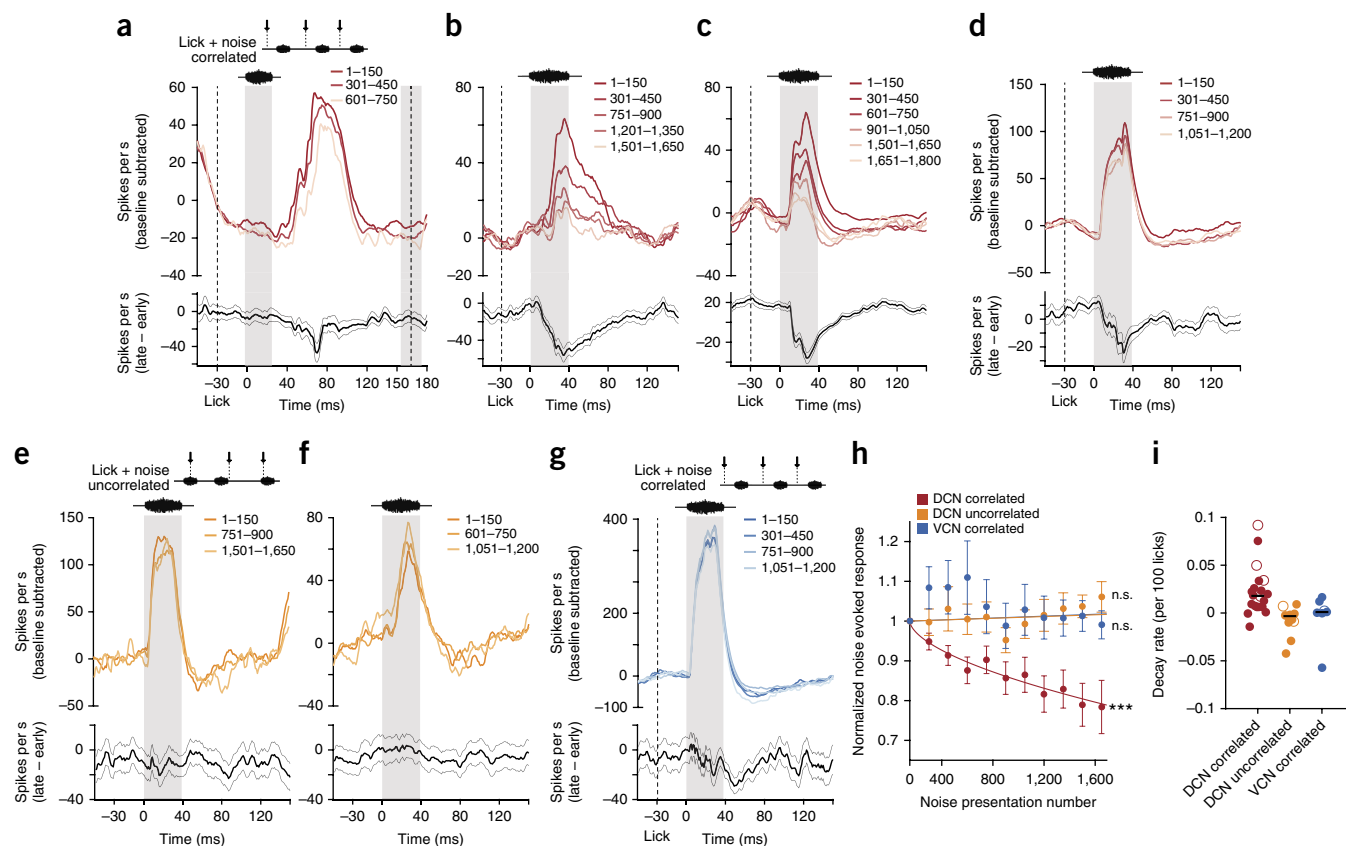


Figure 6 Adaptive cancellation of sounds correlated with behavior in DCN. (a–d) Top, response of an example DCN unit to an acoustic stimulus (broadband or bandpass-filtered noise 5–15 kHz) presented correlated with lick onset. Successively lighter traces show responses to later licks, averaged in bins of 150 licks. Bottom, final response of this cell minus initial response to the sound plus lick. Thin lines are s.e.m. Left gray area shows the stimulus presentation period. Left dashed line shows the time of lick onset. Right dashed line and gray box show the mean and s.d., respectively, of the time of the next lick. (e,f) Same display for two DCN units for which the acoustic stimulus was played uncorrelated with the onset time of a lick. (g) Same display for an example VCN unit in which the acoustic stimulus was presented correlated with the onset time of a lick. (h) Group data showing average changes in noise-evoked responses over the course of repeated stimulus presentations. For DCN correlated units (red), the best-fit decay rate was 0.0225 per 100 licks ($n = 20$, $***P = 6 \times 10^{-15}$, linear regression t -test); for DCN uncorrelated units (yellow), the best-fit decay rate was 0.001 but was not significantly different from 0 (n.s.; $n = 11$, $P = 0.42$, linear regression t -test); for VCN units (blue), the best-fit decay rate was 0.001 but was not significantly different from 0 ($n = 7$, $P = 0.56$, linear regression t -test). Error bars are s.e.m. (i) Scatter plot of decay rates of best-fit exponentials fit separately for every unit. Horizontal black lines show the median value for each group. Open symbols correspond to the units used as examples in a–g.

licking rate after lidocaine injection (Fig. 5f) or changes in the amplitude of licking sounds after lidocaine injection (Fig. 5g).

Adaptive cancellation of sounds correlated with behavior in DCN neurons

Studies of cerebellum-like structures in fish have shown that cancellation of self-generated inputs is not fixed but reflects an adaptive filtering process in which anti-Hebbian synaptic plasticity reduces correlations between principal cell activity and behavior-related signals conveyed by granule cells^{14,15}. Similar anti-Hebbian plasticity rules have been described at granule cell synapses in DCN^{7–9}. Adaptive filtering in DCN would explain both how diverse sources of mossy fiber input are sculpted into patterns of synaptic input that selectively cancel responses to self-generated sounds and how such patterns are updated if the auditory consequences of a given behavior change. To test whether DCN is capable of adaptive filtering, we delivered an external sound (broadband or bandpass-filtered noise 5–15 kHz) temporally correlated with licking (30 ms after tongue contact). The conditions were the same as those for the experiments shown in Figure 3, except that many more sound presentations were used.

Recordings were made using both glass microelectrodes and multi-site silicon probe electrodes (Supplementary Fig. 5). Use of the latter aided the maintenance of single-unit isolation through long bouts of licking. Responses of putative DCN output cells to the correlated sound declined over the course of several minutes of pairing (>1,000 paired lick–sound presentations) (Fig. 6a–d,h). Such declines were not due to overall changes in firing rate, but rather were specific to the period of the noise-evoked response (Fig. 6a–d,h). Decreases in DCN responses to sounds correlated with licking are unlikely to reflect adaptation of peripheral auditory input, as they were not observed in a separate group of DCN units in which an identical external sound was presented at the same rate but at random times relative to lick contact (Fig. 6e,f,h). Furthermore, no changes in sound-evoked responses were observed in VCN units when the external sound was temporally correlated with licking (Fig. 6g,h). The magnitude of the reductions in response to acoustic stimuli correlated with licking varied substantially across DCN units (Fig. 6i). We found no relationships between the magnitude of such reductions and a number of behavioral and neural parameters, including licking rate, licking variability, baseline firing rate and the initial magnitude of noise-evoked responses

(Supplementary Fig. 6). More definitive criteria for identifying DCN cell types, such as juxtacellular labeling and antidromic stimulation, along with a thorough characterization of auditory response properties may, in the future, provide insights into the source of this variation. Overall, these results are consistent with a plastic cancellation or adaptive filtering of self-generated stimuli in DCN similar to that described previously in cerebellum-like structures in fish.

DISCUSSION

Distinguishing between external and self-generated sensory stimuli is fundamental for perception and is thought to involve a comparison between external sensory input and internal reference signals related to the animal's own behavior—for example, motor corollary discharge or proprioception³³. The present study provides evidence that such a comparison takes place at the first central stage of mammalian auditory processing in the DCN. More specifically, our results suggest a scheme similar to that already well established in cerebellum-like structures in fish, in which behavior-related signals conveyed by a mossy fiber–granule cell–parallel fiber system cancel out responses to self-generated sensory stimuli in principal neurons^{14,15}. Several independent lines of evidence from the present study support such a function for DCN. First, responses to sounds generated by licking are substantially weaker in DCN than in VCN, and such differences are not accounted for by weaker responses to external acoustic stimuli in DCN or by an overall suppression of DCN responses during licking. Second, non-auditory responses to licking behavior are observed in putative CWCs, presumably due to non-auditory signals conveyed by mossy fibers and granule cells. Third, inactivation of Sp5, a prominent source of somatosensory mossy fiber input to DCN, revealed responses to self-generated sounds in DCN units that resembled those observed in VCN units, suggesting that such input normally functions to cancel DCN responses to self-generated sounds. Finally, repeated pairing of acoustic stimuli with licking resulted in a gradual reduction of DCN responses to the paired stimulus. Notably, such reductions were not observed when stimuli were presented at the same rate but uncorrelated with the time of lick contact.

Cancellation of self-generated sounds at an early processing stage could provide mammals with a dedicated channel through which salient or unexpected auditory signals can rapidly guide motor output, such as escape or orienting behavior. This interpretation is consistent with effects of DCN lesions, which disrupt orienting toward but not discriminating between sound source locations^{34,35}, and the fact that, in addition to projecting to the inferior colliculus, DCN projects directly to auditory thalamus³⁶, auditory cortex³⁷ and regions involved in the acoustic startle response³⁸. To our knowledge, responses to self-generated sounds have not been studied at the level of the inferior colliculus. Based on the present results, we would predict that a subset of inferior colliculus neurons selectively encodes external sounds and that this subset receives its dominant input from DCN rather than VCN. Though we focused on a single behavior and a single source of mossy fiber input, the fact that DCN receives mossy fiber inputs from numerous brain regions conveying a wide range of sensory and motor signals implies a much broader capacity for canceling predictable auditory input⁶. We also note that our results by no means rule out the possibility that additional sources of mossy fiber inputs (besides those originating from Sp5) help cancel self-generated sounds caused by licking behavior. For example, mossy fiber input to DCN GCDs originating from the pontine nuclei could provide motor-related signals relevant for cancelling the auditory consequences of the animal's own movements, including licking³⁹.

Though our results suggest that the integration of non-auditory and auditory inputs to DCN serves to cancel responses to self-generated

stimuli, they do not rule out other functions for multimodal integration in DCN. Numerous lines of evidence suggest that the DCN is important in processing spectral cues for sound localization^{6,10,34}. A recent study provided evidence that the integration of auditory and vestibular information in DCN could aid in distinguishing changes in auditory input due to motion of an external sound source from those due to self-motion¹³. Specifically, Wigderson *et al.*¹³ demonstrated that vestibular and auditory inputs are combined nonlinearly in putative DCN output cells. This is a different mode of integration from that suggested here and by studies of other cerebellum-like structures in fish in which behavior-related signals conveyed by mossy fibers are used to subtract out self-generated signals. Since vestibular inputs would not have been engaged during the head-fixed licking behavior we studied, no direct comparison between the two studies is possible. However, determining whether different sources of mossy fiber inputs to DCN, such as vestibular versus somatosensory, perform similar or different computations is an important question for future studies.

Key questions remain regarding the circuit mechanisms underlying the cancellation of self-generated sounds in DCN reported here. In cerebellum-like structures in fish, cancellation is due to the generation and subtraction of negative images of the responses of principal cells to self-generated inputs. Such negative images are formed by anti-Hebbian synaptic plasticity acting on corollary discharge, proprioceptive, and electrosensory signals conveyed by parallel fibers^{14,15}. Due both to limits on data collection imposed by satiation and the technical difficulty of maintaining stable single-unit recordings in brainstem through long bouts of licking, we focused exclusively on providing evidence for cancellation. A crucial next step will be to determine whether cancellation of self-generated sounds in DCN is due to the generation of negative images. Furthermore, genetic tools available in mice should make it possible to perform a detailed dissection of the mechanisms underlying the cancellation of self-generated sounds in DCN. Key questions include the functions of specific cell types, such as the CWCs, and the roles of specific sites and mechanisms of plasticity, such as spike timing-dependent plasticity at parallel fiber synapses onto fusiform cells and CWCs described *in vitro*^{7,8,40}.

Finally, our results are intriguing from evolutionary and comparative perspectives. The brains of most vertebrates contain both a cerebellum and one or more sensory structures with circuitry closely resembling that of the cerebellum^{2,6,15}. Though similarities between different cerebellum-like structures and the cerebellum are well established in terms of their evolution, development, gene expression patterns, circuitry and synaptic plasticity, the question of whether they perform similar functions has been more difficult to address. Cerebellum-like structures associated with electrosensory processing in three distinct groups of fish have been shown to act as adaptive filters^{14,15}, and several lines of evidence also support such a role for the mammalian cerebellum^{41,42}. In both cerebellum-like structures and the cerebellum, granule cells convey a rich variety of signals^{43–47} and a separate, non-plastic input (peripheral sensory input in the case of cerebellum-like structures and climbing fiber input in the case of cerebellum) instructs plasticity at granule cell synapses such that output that is predictable (in the case of cerebellum-like sensory structures) or associated with errors in motor performance (in the case of the cerebellum) is gradually reduced. Notably, adaptive cancellation of self-generated vestibular inputs has been demonstrated in neurons of the fastigial nucleus and vestibular nucleus in primates^{48,49}. Hence evidence provided here for sensory cancellation and adaptive filtering in DCN support the idea that core functions may be shared by cerebellum-like structures and the cerebellum across vertebrate phylogeny.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank L. Abbott, C. Bell and T. Jessell for comments on the manuscript. This work was supported by grants from the NIH (DC015449), the Alfred P. Sloan Foundation, the McKnight Endowment Fund for Neuroscience and the Irma T. Hirsch Trust to N.B.S. and an NIH grant (F30DC014174) to S.S.

AUTHOR CONTRIBUTIONS

S.S. and R.W. performed the experiments and analyzed the data. C.D. designed and performed the analysis. A.G.E. contributed custom software and hardware. N.B.S. and S.S. designed the study and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Cant, N.B. The cochlear nucleus: neuronal types and their synaptic organization. in *The Mammalian Auditory Pathway: Neuroanatomy* (eds. Webster, D.B. et al.) 66–116 (Springer, New York, 1992).
- Bell, C.C. Evolution of cerebellum-like structures. *Brain Behav. Evol.* **59**, 312–326 (2002).
- Berrebí, A.S., Morgan, J.I. & Mugnaini, E. The Purkinje cell class may extend beyond the cerebellum. *J. Neurocytol.* **19**, 643–654 (1990).
- Mugnaini, E., Warr, W.B. & Osen, K.K. Distribution and light microscopic features of granule cells in the cochlear nuclei of cat, rat, and mouse. *J. Comp. Neurol.* **191**, 581–606 (1980).
- Lorente de Nó, R. Central representation of the eighth nerve. in *Ear Diseases, Deafness and Dizziness* (ed. Goodhill, V.) 64–86 (Harper and Row, Hagerstown, Maryland, USA, 1979).
- Oertel, D. & Young, E.D. What's a cerebellar circuit doing in the auditory system? *Trends Neurosci.* **27**, 104–110 (2004).
- Fujino, K. & Oertel, D. Bidirectional synaptic plasticity in the cerebellum-like mammalian dorsal cochlear nucleus. *Proc. Natl. Acad. Sci. USA* **100**, 265–270 (2003).
- Tzounopoulos, T., Kim, Y., Oertel, D. & Trussell, L.O. Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nat. Neurosci.* **7**, 719–725 (2004).
- Zhao, Y. & Tzounopoulos, T. Physiological activation of cholinergic inputs controls associative synaptic plasticity via modulation of endocannabinoid signaling. *J. Neurosci.* **31**, 3158–3168 (2011).
- Young, E.D. & Davis, K.A. Circuitry and function of the dorsal cochlear nucleus. in *Integrative Functions in the Mammalian Auditory Pathway* (eds. Oertel, D. et al.) 160–206 (Springer, New York, 2002).
- Kanold, P.O. & Young, E.D. Proprioceptive information from the pinna provides somatosensory input to cat dorsal cochlear nucleus. *J. Neurosci.* **21**, 7848–7858 (2001).
- Shore, S.E. & Zhou, J. Somatosensory influence on the cochlear nucleus and beyond. *Hear. Res.* **216–217**, 90–99 (2006).
- Wigderson, E., Nelken, I. & Yarom, Y. Early multisensory integration of self and source motion in the auditory system. *Proc. Natl. Acad. Sci. USA* **113**, 8308–8313 (2016).
- Bell, C., Bodznick, D., Montgomery, J. & Bastian, J. The generation and subtraction of sensory expectations within cerebellum-like structures. *Brain Behav. Evol.* **50** (Suppl. 1), 17–31 (1997).
- Bell, C.C., Han, V. & Sawtell, N.B. Cerebellum-like structures and their implications for cerebellar function. *Annu. Rev. Neurosci.* **31**, 1–24 (2008).
- Luo, F., Wang, Q., Farid, N., Liu, X. & Yan, J. Three-dimensional tonotopic organization of the C57 mouse cochlear nucleus. *Hear. Res.* **257**, 75–82 (2009).
- Muniak, M.A. et al. 3D model of frequency representation in the cochlear nucleus of the CBA/J mouse. *J. Comp. Neurol.* **521**, 1510–1532 (2013).
- Ma, W.L. & Brenowitz, S.D. Single-neuron recordings from unanesthetized mouse dorsal cochlear nucleus. *J. Neurophysiol.* **107**, 824–835 (2012).
- Davis, K.A., Ding, J., Benson, T.E. & Voigt, H.F. Response properties of units in the dorsal cochlear nucleus of unanesthetized decerebrate gerbil. *J. Neurophysiol.* **75**, 1411–1431 (1996).
- Hancock, K.E. & Voigt, H.F. Intracellularly labeled fusiform cells in dorsal cochlear nucleus of the gerbil. I. Physiological response properties. *J. Neurophysiol.* **87**, 2505–2519 (2002).
- Rhode, W.S. Vertical cell responses to sound in cat dorsal cochlear nucleus. *J. Neurophysiol.* **82**, 1019–1032 (1999).
- Young, E.D. & Brownell, W.E. Responses to tones and noise of single cells in dorsal cochlear nucleus of unanesthetized cats. *J. Neurophysiol.* **39**, 282–300 (1976).
- Young, E.D. Identification of response properties of ascending axons from dorsal cochlear nucleus. *Brain Res.* **200**, 23–37 (1980).
- Manis, P.B., Spirou, G.A., Wright, D.D., Paydar, S. & Ryugo, D.K. Physiology and morphology of complex spiking neurons in the guinea pig dorsal cochlear nucleus. *J. Comp. Neurol.* **348**, 261–276 (1994).
- Zhang, S. & Oertel, D. Cartwheel and superficial stellate cells of the dorsal cochlear nucleus of mice: intracellular recordings in slices. *J. Neurophysiol.* **69**, 1384–1397 (1993).
- Eliades, S.J. & Wang, X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* **89**, 2194–2207 (2003).
- Poulet, J.F. & Hedwig, B. A corollary discharge maintains auditory sensitivity during sound production. *Nature* **418**, 872–876 (2002).
- Rummell, B.P., Klee, J.L. & Sigurdsson, T. Attenuation of responses to self-generated sounds in auditory cortical neurons. *J. Neurosci.* **36**, 12010–12026 (2016).
- Schneider, D.M., Nelson, A. & Mooney, R. A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* **513**, 189–194 (2014).
- Portfors, C.V. & Roberts, P.D. Temporal and frequency characteristics of cartwheel cells in the dorsal cochlear nucleus of the awake mouse. *J. Neurophysiol.* **98**, 744–756 (2007).
- Haenggeli, C.A., Pongstaporn, T., Doucet, J.R. & Ryugo, D.K. Projections from the spinal trigeminal nucleus to the cochlear nucleus in the rat. *J. Comp. Neurol.* **484**, 191–205 (2005).
- Zhou, J. & Shore, S. Projections from the trigeminal nuclear complex to the cochlear nuclei: a retrograde and anterograde tracing study in the guinea pig. *J. Neurosci. Res.* **78**, 901–907 (2004).
- Cullen, K.E. Sensory signals during active versus passive movement. *Curr. Opin. Neurobiol.* **14**, 698–706 (2004).
- May, B.J. Role of the dorsal cochlear nucleus in the sound localization behavior of cats. *Hear. Res.* **148**, 74–87 (2000).
- Sutherland, D.P., Glendenning, K.K. & Masterton, R.B. Role of acoustic striae in hearing: discrimination of sound-source elevation. *Hear. Res.* **120**, 86–108 (1998).
- Malmierca, M.S., Merchán, M.A., Henkel, C.K. & Oliver, D.L. Direct projections from cochlear nuclear complex to auditory thalamus in the rat. *J. Neurosci.* **22**, 10891–10897 (2002).
- Anderson, L.A., Izquierdo, M.A., Antunes, F.M. & Malmierca, M.S. A monosynaptic pathway from dorsal cochlear nucleus to auditory cortex in rat. *Neuroreport* **20**, 462–466 (2009).
- Lingenhöhl, K. & Friauf, E. Giant neurons in the rat reticular formation: a sensorimotor interface in the elementary acoustic startle circuit? *J. Neurosci.* **14**, 1176–1194 (1994).
- Ohlrogge, M., Doucet, J.R. & Ryugo, D.K. Projections of the pontine nuclei to the cochlear nucleus in rats. *J. Comp. Neurol.* **436**, 290–303 (2001).
- Tzounopoulos, T., Rubio, M.E., Keen, J.E. & Trussell, L.O. Coactivation of pre- and postsynaptic signaling mechanisms determines cell-specific spike-timing-dependent plasticity. *Neuron* **54**, 291–301 (2007).
- Dean, P., Porrill, J., Ekerot, C.F. & Jörntell, H. The cerebellar microcircuit as an adaptive filter: experimental and computational evidence. *Nat. Rev. Neurosci.* **11**, 30–43 (2010).
- Fujita, M. Adaptive filter model of the cerebellum. *Biol. Cybern.* **45**, 195–206 (1982).
- Chabrol, F.P., Arenz, A., Wiechert, M.T., Margrie, T.W. & DiGregorio, D.A. Synaptic diversity enables temporal coding of coincident multisensory inputs in single neurons. *Nat. Neurosci.* **18**, 718–727 (2015).
- Huang, C.C. et al. Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. *Elife* **2**, e00400 (2013).
- Ishikawa, T., Shimuta, M. & Häusser, M. Multimodal sensory integration in single cerebellar granule cells in vivo. *Elife* **4**, e12916 (2015).
- Kennedy, A. et al. A temporal basis for predicting the sensory consequences of motor commands in an electric fish. *Nat. Neurosci.* **17**, 416–422 (2014).
- Sawtell, N.B. Multimodal integration in granule cells as a basis for associative plasticity and sensory prediction in a cerebellum-like circuit. *Neuron* **66**, 573–584 (2010).
- Brooks, J.X., Carriot, J. & Cullen, K.E. Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nat. Neurosci.* **18**, 1310–1317 (2015).
- Roy, J.E. & Cullen, K.E. Selective processing of vestibular reafference during self-generated head motion. *J. Neurosci.* **21**, 2131–2142 (2001).

ONLINE METHODS

Animals. All experimental protocols were approved by the Columbia University Institutional Animal Care and Use Committee. Adult male wild-type mice (129S6/SvEvTac) were used for all experiments. Mice were purchased from Taconic Biosciences (Hudson, NY) and housed in an on-site animal facility on a 12-h light-dark cycle. Most experiments were performed during the light cycle. Data collection and analysis were not performed blind to the conditions of the experiments.

Surgery. Mice were anesthetized with isoflurane (1.5–2%) and placed in a stereotax equipped with zygomatic ear bars (Kopf Instruments). The skull was exposed and a small craniotomy (200–500 μm in diameter) was made over the right dorsal cochlear nucleus (5.5 mm posterior to bregma and 2.3 mm lateral to the midline). The craniotomy was covered with silicone elastomer (Kwik-Sil, WPI, Sarasota, FL). A custom headplate was attached to the skull using dental cement (C&B Meta-bond, Parkell, Edgewood, NY). Mice were allowed to recover for 3 d before the start of experiments.

Experimental apparatus and auditory stimulus presentation. All mouse behavior and neurophysiology experiments were performed in a double-walled sound-attenuating chamber (Double Deluxe Model, Gretchen Industries). The ambient noise in the chamber was <30 dB SPL as measured by a sound pressure level meter (Bruel and Kjaer type 2240). A custom head fixation device was used to secure the animal via two attachment points to a stainless steel headplate, allowing consistent positioning across multiple recording sessions. The animal's body was also secured between two pieces of Styrofoam molded to its body. A stainless steel lick spout was positioned in front of the animal's mouth and licks were detected using standard methods. Acoustic stimuli were generated using Spike2 software (Cambridge Electronic Design) and delivered using an electrostatic speaker (ES-1 Tucker Davis Technologies) positioned approximately 10 cm in front of the mouse just to the right of the midline. Sound pressure levels of acoustic stimuli as measured in dB SPL were calibrated to the location of the animal's right ear. The frequency response of the sound system was measured to be flat (± 4 dB) from 1 kHz to 50 kHz using a quarter-inch condenser microphone (377C01, PCP Piezotronics), attached to a preamplifier (426B03, PCP Piezotronics), positioned at the location of the mouse's right ear. Sounds caused by licking were monitored by a small electret microphone (Knowles model 23329N) placed just above the lick spout. Microphone signals were sampled at 100 kHz and digitized using an analog-to-digital converter (Power 1401, Cambridge Electronic Design).

The lick mimic was constructed from segments of microphone recordings 50 ms before tongue contact to 150 ms after tongue contact and bandpass filtered between 1 and 50 kHz ($n = 5$ mice). We transformed each segment to a spectrogram using a short-time Fourier transform (Hamming window with a width of 10.24 ms and a stride of 5.12 ms). We then constructed the mimic by performing principal component analysis on this set of lick-triggered spectrograms and making a weighted sum of the first five principal components. This resulted in a mimic spectrogram, which we used as a spectro-temporal filter to convolve with a random signal. This resulted in a stimulus (the lick mimic) that contained the most prominent spectro-temporal features of the licking sound (including distinct spectral peaks at 2, 8 and 30 kHz) with little power elsewhere. Due to issues such as bone conduction, we could not measure the exact loudness of natural licking sounds. The lick mimic was replayed at a loudness that evoked a response in VCN units that was, on average, similar to that evoked by licking. This same loudness (12 dB SPL) was used subsequently for all experiments involving the mimic.

Behavioral training. Mice were allowed to recover 3 d after surgery before to beginning water deprivation and habituation to head restraint in the experimental apparatus. Weight was monitored daily and extra water was given in the home cage if the animal's weight fell below 80% of its initial presurgical weight. Extracellular recordings from DCN and VCN units were then performed during daily sessions lasting 2–3 h. Mice licked roughly 3,000 times per session.

Extracellular recording and identification of DCN and VCN neurons. Standard procedures were used for extracellular recording using glass microelectrodes (5–20 M Ω resistance). Pipettes with a long taper were used to avoid tissue damage. On the day of recording, mice were placed into the head restraint, the silicone

elastomer was removed and 0.9% saline was placed over the exposed craniotomy. The microelectrode was lowered into the craniotomy vertically. As the electrode was advanced through the cerebellum, a series of 200-ms-long search tones from 5 kHz to 50 kHz (in 5 kHz steps) were delivered. Entrance into DCN was marked by a transient increase in electrode resistance along with the sudden appearance of tone-evoked multiunit activity that occurred ~2,700–3,200 μm below the surface of the cerebellum. The microelectrode was then advanced in 1- μm steps until a unit was isolated. Complex-spiking units were the first units encountered on an electrode penetration through DCN and could be unambiguously identified on the basis of their distinctive complex spikes. Complex spikes are stereotyped, high-frequency action potential bursts superimposed on a slower depolarization and are not observed in any DCN cell types except CWCs^{24,25}. Similarly to previous *in vivo* extracellular recording studies of DCN in a variety of species, including mouse¹⁸, we defined complex spikes as high-frequency bursts (ISIs < 3.5 ms) of 2–5 action potentials. Complex spikes were identified automatically in Spike2 using custom-written scripts and then confirmed individually. Within such bursts, action potentials successively widened and decreased in amplitude (Fig. 4b). Complex-spiking units were isolated 50–200 μm from the surface of the DCN. DCN units lacking complex spikes, referred to here as simple spiking units, were isolated 100–300 μm from the surface of the DCN. Complex-spiking units were never found ventral to simple spiking units on the same electrode penetration, consistent with the known cytoarchitecture of the DCN. Passage from DCN into VCN was determined by monitoring the tone frequency that most strongly drove multiunit activity for each 50 μm advance of the electrode. As the electrode advanced ventrally, the best frequency for driving multiunit activity progressively decreased. A sudden increase in the best frequency (generally from ~5 kHz to ~20 kHz and usually occurring 500–600 μm below the surface of DCN) signified entrance into the VCN. Units that were isolated at least 100 μm ventral to the best-frequency reversal (~800–1,000 μm below the surface of the DCN) and that showed clear tone-evoked responses were classified as VCN units. Units isolated less than 100 μm from the best-frequency reversal were not included in the analysis. Histological verification of DCN and VCN recording sites was performed by iontophoresis of dextran-conjugated Alexa Fluor 594 (D22913, Thermo Fisher Scientific) at recording sites between 100 and 300 μm below the surface of DCN (depths at which most DCN simple-spiking units were isolated) and at 900 μm (the depth at which most VCN units were isolated). Only units that remained well isolated through at least 75 licks were included in the analysis. Sounds associated with licking contain most power between 2 and 15 kHz, which corresponds to the lower portion of the mouse hearing range. For this reason we focused our recordings on regions of the cochlear nucleus that represent these frequencies. A subset of the recordings in Figure 6 (DCN correlated, $n = 10$ of 20; DCN uncorrelated, $n = 5$ of 11; VCN correlated, $n = 3$ of 7) were performed using a 16-channel silicon probe (Neuronexus, A1x16-5mm-25-177-A16). Silicon probe recordings proved superior to glass microelectrode recordings in terms of their stability during licking behavior. Probes consisted of a vertical linear array of 15- μm -diameter electrode sites spaced 25 μm apart. Impedances ranged from ~2 to 6 k Ω . Recording tracks were made in DCN or VCN until a well-isolated single unit emerged on at least one electrode site. Most sites exhibited only multiunit activity and were not analyzed. The same electrophysiological signatures described above were used to identify the DCN and VCN. Rank sum tests revealed no difference between probe and glass recordings in the median decay rate of cells in all three groups shown in Figure 6 (DCN correlated: $P = 0.09$, DCN uncorrelated: $P = 0.79$, VCN: $P = 0.63$).

Viral injections. A nanoliter injector (504126, WPI instruments) was used to inject adeno-associated virus expressing green fluorescent protein. The pipette was positioned over the coordinates 7.2 mm posterior to bregma and 1.8 mm right of the midline and lowered until the tip touched the surface of the cerebellum. The pipette was then lowered 3.5 mm below the surface of the cerebellum to the base of the spinal trigeminal nucleus. Virus (27 nL) was injected in three 9-nL pulses. Virus was also injected at 3.2, 2.9 and 2.7 mm below the surface of the cerebellum. The pipette was then slowly raised out of the cerebellum and the incision was closed using cyanoacrylate glue (Vetbond, 3M, Maplewood, Minnesota). Two weeks after surgery, mice were anesthetized with ketamine/xylazine and perfused with 4% formaldehyde. The brains were dissected from the skull and allowed to postfix in 4% formaldehyde overnight. They were then cryoprotected in a 30% sucrose solution and sectioned on a cryostat. Sections were then mounted on glass

slides (Superfrost, Fisher Scientific, Waltham, MA), counterstained with DAPI and imaged on a confocal microscope (Carl Zeiss Microscopy, Peabody, MA).

Deafening. Mice were deafened bilaterally. Surgery for deafening mice was performed using 2–4% isoflurane. An incision was made just posterior to the tragus and extended ventrally. The tympanum, malleus and incus were visualized through the auditory meatus. Using fine forceps, the tympanum was ruptured and the malleus and incus were removed. The stapes was removed, exposing the oval window, with care taken not to damage the stapedia artery. Using a 30-gauge needle, approximately 10–20 μ L of 1.0 mg/mL kanamycin was injected through the oval window and into the cochlea. The middle ear was packed with gel foam and the mouse was allowed to recover in its home cage. Deafening was verified by lack of observable behavioral responses to acoustic stimuli and by recording sound evoked field potentials to broadband noise (50 ms, 6–90 dB SPL) in DCN \sim 75 μ m below the first observed complex-spiking unit. This was done both before and 2 d after surgical deafening in each mouse. DCN recordings were performed 2–4 d after surgery. Recording locations in DCN were confirmed histologically using iontophoresis of dextran-conjugated Alexa 594 as described above.

Lidocaine injections into Sp5. A small craniotomy (\sim 300 μ m diameter) was made before attachment of the headplate at coordinates 7.2 mm posterior to bregma and 1.8 mm lateral to the midline and covered with silicone elastomer. On the day of the experiment, a glass micropipette with a long taper was pulled using a pipette puller (PC-10, Narishige Group) and manually broken to 3.5 μ m diameter under a microscope. The pipette was then filled with 2% lidocaine in 0.9% saline, with care taken to avoid air bubbles in the tip. The pipette was then coupled to a micropressure injector (Picospritzer MK III, Parker Instrumentation) and successful ejection of lidocaine was confirmed visually to ensure the tip was not clogged. The lidocaine pipette was advanced into Sp5 at an angle of 12.8 degrees. For Sp5 inactivation, DCN unit responses were recorded for \sim 200 licks before \sim 100 nL of lidocaine was injected in a single pulse. Location of the lidocaine pipette within DCN was verified histologically using iontophoresis of dextran-conjugated Alexa 594 as described above.

Lick-sound pairing. After isolation of a unit, access to water was given and contact to the lick spout by the animal's tongue was paired with a 30-ms noise (15–71 dB SPL, broadband or bandpass filtered 5–15 kHz). In the correlated condition the noise was presented 30 ms after contact with the lick spout. The pairing was conducted continuously until the animal stopped licking or unit isolation was lost. In the uncorrelated condition presentation of the noise during licking was unrelated to the tongue's contact with the spout and was instead presented at random intervals of 120–160 ms. Since these intervals are similar to inter-lick intervals the overall rate of sound presentations was similar in the correlated and uncorrelated conditions. Correlated versus uncorrelated conditions were tested in the same mice on alternating sessions. The condition to be tested during a given session was predetermined before isolating a unit.

Data analysis and statistics. All analyses were performed using custom-written scripts for Matlab (MathWorks, Natick, MA) and Spike2. No statistical methods were used to predetermine sample sizes. Comparisons between two groups were made by Mann–Whitney *U*-test or Wilcoxon signed rank test for paired groups. Tests of the significance of linear regression slopes used a linear regression *t*-test. For the linear regression *t*-test, residuals were assumed to be normally distributed but this was not formally tested. Differences were considered statistically significant at $P < 0.01$. Data are presented as mean \pm s.e.m. unless indicated otherwise.

Lick sound spectrograms. To compute the average spectrogram of the sound associated with licking, we first bandpass-filtered raw microphone traces, removing frequencies below 1 kHz and above 50 kHz (the highest frequency that could be detected by our equipment). 300-ms segments of the filtered microphone recording centered on the onset of each lick were transformed with a short time Fourier transform (Hamming window with a width of 10.24 ms and a stride of 5.12 ms) to obtain a set of lick-centered spectrograms. These were averaged to obtain a lick-triggered average spectrogram. Time-frequency peaks were found by first applying a 2D median filter (widths of 290 Hz and 3 ms) to individual spectrograms and then convolving with a 2D Gaussian kernel with widths of 1.5 kHz and 20 ms. We then calculated local time–frequency maxima by finding local maxima of the filtered spectrograms.

R.m.s. amplitude of microphone traces. To compute the r.m.s. amplitude of the sound associated with licking, microphone recordings were first bandpass filtered (1–50 kHz). We then computed the r.m.s. amplitude of this filtered microphone trace by convolving the squared trace with a moving average kernel of width 1 ms and taking the square root of the result. These recordings were then aligned to the time of tongue contact with the lick spout and averaged across licks.

Average and z-scored electrophysiological responses during licking and mimic presentation. To compute average responses to licking or during delivery of the mimic, spike trains were convolved with a normalized sum-of-two-exponentials kernel, with a rise time of 5 ms and a decay time of 20 ms. Averages were aligned either on tongue contact with the lick spout or mimic delivery and average baseline firing was subtracted. Baseline firing rates were taken to be the average firing rates in periods at least 25 ms before the next lick or mimic onset and at least 150 ms after the previous lick or mimic onset. Peak-to-trough firing rates were computed by taking the average licking or mimic response in a 200-ms window centered on the tongue-to-spout contact or mimic onset and determining the difference in the maximum to minimum firing rates. To compute z-scores we first took the maximum of the average licking or mimic response in a 200-ms window centered on tongue-to-spout contact or mimic onset. We then created shuffled spike trains of approximately the same length as the original spike train by randomly sampling from the inter-spike-interval distribution of the real spike train. Each shuffled spike train was convolved with the same kernel as the real spike train, its lick- or mimic-triggered average computed, and the maximum firing rate of this triggered average taken in the same 200-ms window. This was repeated 500 times and the maximum of the triggered average of the real spike train was expressed in units of the s.d. from the mean of the shuffled distribution; i.e., z-scored based on the shuffled distribution. We determined the significance of neural responses by computing approximate *P*-values for the recorded maximum lick-triggered rate, which we estimated using an approximate permutation test; that is, by the fraction of shuffled-spike trains showing maximum lick-triggered responses greater than that of the real spike train.

Correlated and uncorrelated sound-lick pairings. The noise-evoked response was defined in bins of 150 stimulus presentations. For each 150 presentations, the response was defined as the maximum of the average noise-evoked response during that stimulus period minus the baseline rate during that period. For each unit the response was normalized to equal 1 in the first bin. We performed a linear regression between the stimulus bin and the log of the noise-evoked responses for each population to extract a decay rate for each population.

A **Supplementary Methods Checklist** is available.

Data and code availability. The data and analysis code pertaining to this study are available from the corresponding author upon reasonable request.