The zebra finch auditory cortex reconstructs occluded syllables in conspecific song

Authors: Margot C Bjoring[1] and C Daniel Meliza[1,2,*]

Affiliations: [1] Department of Psychology, [2] Neuroscience Graduate Program, University of Virginia, Charlottesville VA 22904, USA

[*] Corresponding author. Email: cdm8j@virginia.edu

Sensory input provides incomplete and often misleading information about the physical world. To compensate, the brain uses internal models to predict what the inputs should be from context, experience, and innate biases (Liberman and Mattingly, 1985; Komatsu, 2006; Gilbert and Sigman, 2007; Heald and Nusbaum, 2014). For example, when speech is interrupted by noise, humans perceive the missing sounds behind the noise (Miller and Licklider, 1950; Warren, 1970; Samuel, 1996), a perceptual illusion known as phonemic (or auditory) restoration. The neural mechanisms allowing the auditory system to generate predictions that override ascending sensory information remain poorly understood. Here, we show that the zebra finch (*Taeniopygia guttata*) exhibits auditory restoration of conspecific song both in a behavioral task and in neural recordings from the equivalent of auditory cortex. Decoding the responses of a population of single units to occluded songs reveals the spectrotemporal structure of the missing syllables. Surprisingly, restoration occurs under anesthesia and for songs that the bird has not heard. These results show that an internal model of the general structure of conspecific vocalizations can bias sensory processing without attention.

Auditory restoration is one of a number of ways the brain extracts signals of interest from complex auditory scenes (Bregman, 1999). When phonemes are deleted from a speech stream and replaced with noise, listeners hear the missing phonemes behind the occluding noise (Miller and Licklider, 1950; Warren, 1970). The restored phonemes are appropriate to their phonological and lexical context (Warren and Sherman, 1974; Samuel, 1996), and restoration is stronger for words in the

listener's native language compared to foreign or nonsense words (Samuel, 1996; Ishida and Arai, 2016). These observations are consistent with cognitive models of speech perception in which top-down predictions based on context and experience actively influence processing in primary auditory areas (Heald and Nusbaum, 2014; Leonard et al., 2016). Alternatively, recurrent connectivity within auditory areas may be such that local neural circuit dynamics are sufficient to fill in missing information. Indeed, many illusions, including the visual analog of auditory restoration, can occur without attention (Komatsu, 2006), so it is possible that expectations about the acoustic structure of speech are implemented preattentively, within the auditory system.

Neural recordings have the potential to elucidate where and how auditory restoration occurs. In humans presented with occluded speech, the auditory cortex exhibits illusory responses: patterns of activity that closely resemble those evoked by the phoneme the subject reports hearing (Leonard et al., 2016). These recordings also show activity in higher-order language areas that precedes the missing syllable and predicts perception, suggestive of top-down influences but not excluding local computations. Nonhuman animals also experience a form of auditory restoration (Braaten and Leary, 1999; Miller et al., 2001; Petkov et al., 2003; Seeba and Klump, 2009), but neural responses have only been tested with simple tones (Sugita, 1997; Petkov et al., 2007), which lack the complex acoustic structure of speech and may involve different mechanisms. We therefore turned to the zebra finch, a social songbird with an acoustically rich song used for communication in complex, noisy social environments (Singh and Theunissen, 2003; Elie and Theunissen, 2016). After establishing that zebra finches experience illusory continuity using a behavioral task closely modeled on human psychophysical studies (Samuel, 1996), we recorded neural responses to occluded stimuli from the avian homolog of auditory cortex. Recordings were under anesthesia, thereby silencing top-down processes requiring attention. We employed a linear decoding model to test whether these neurons respond to the illusory percept and to determine where and when the illusion emerges within the auditory processing hierarchy.

Zebra finches experience illusory continuity of occluded songs

We tested the perception of auditory restoration in zebra finches using an oddball-detection paradigm (Fig. 1a). Birds were trained to detect a brief discontinuity in a sequence of otherwise identical conspecific song motifs. To create these discontinuities, we elided the sound in a 100ms critical interval, replacing it with white noise (*Replaced* condition). The other motifs in the sequence were continuous versions of the same motif with white noise added on top of the critical interval (*Added* condition). The amplitude of the white noise was varied so that the signal-to-noise ratio (SNR) ranged between 20 and –15 dB. The motifs were recorded from males in our colony, and the experimental birds were socialized with four of those males prior to training (Fig. 1b) so that half of the stimuli were familiar to them and half were unfamiliar.

At high SNR, trained birds ($n = 5$) were proficient at detecting the discontinuity, despite the difficulty of learning the task (7 of 12 birds were excluded for failing to reach baseline performance criteria). At 20 dB SNR, the mean hit rate was 0.50 ± 0.07 with a false alarm rate of 0.06 ± 0.03 (Fig. 1c). As the noise intensity within the critical interval increased, detecting the gap became more challenging, and hit rate decreased. The slope of the drop-off in hit rate was similar across all subjects, but there were individual differences in baseline performance and perceptual thresholds. The false alarm rate stayed constant across noise levels and varied little among subjects. It was also well below 0.2, the prior probability that a given motif would contain a gap, indicating a strong bias against responding unless the bird was confident it could hear the discontinuity.

A hallmark of perceptual illusions is that performance in behavioral tasks drops below chance. This is because subjects are not merely guessing, but are fooled into perceiving something that is not actually there (Samuel, 1996; Petkov et al., 2003). We therefore expected birds to make systematic errors at noise levels that induce restoration. To test this prediction, we fit the data with a generalized linear mixed-effects model (GLMM; see Materials and Methods) that included effects for familiarity, SNR, and the position of the oddball within the sequence. Although we were primarily interested in the effects of SNR and familiarity, there was a strong, nonlinear interaction with position (Supplementary Fig. 2). Animals performed considerably better on the task when the *Replaced* motif came last in the sequence, suggesting the influence of perceptual anchoring (Braida

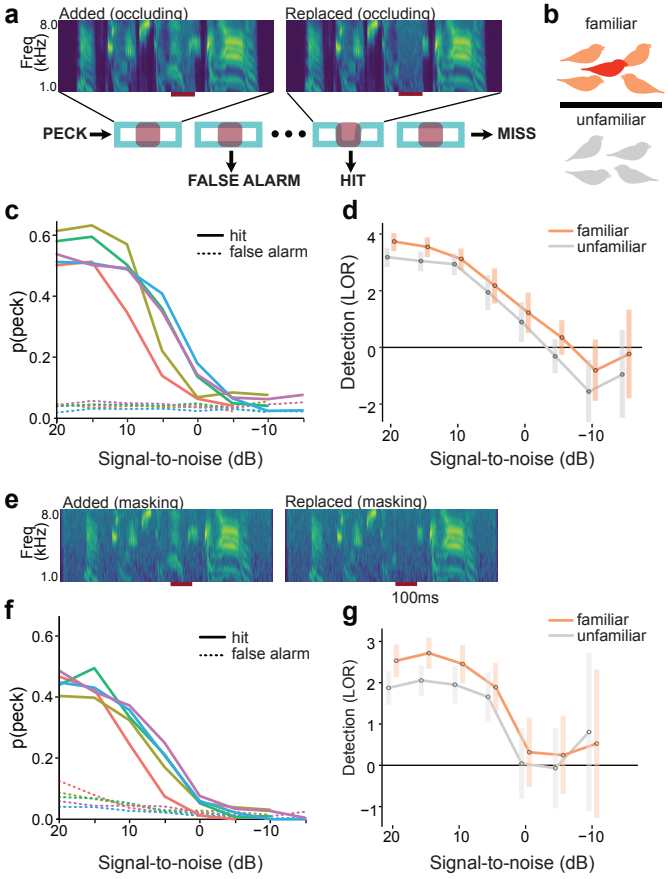


Fig. 1. Illusory perception of continuity during oddball detection task. (**a**) Example stimuli and trial structure for the oddball detection task with occluding noise. Red lines under the spectrograms indicate critical intervals. The position of the *Replaced* motif within the sequence was varied randomly. (**b**) Prior to the behavioral task, the subject (red) was familiarized with four of the males whose songs were used in the task (orange). The other four motifs were from birds with whom the subject had never been housed (gray). Familiarity was counterbalanced across subjects. (**c**) Average hit and false alarm rates for each of the subjects, with color indicating subject identity. (**d**) Estimates of average performance (log odds ratio, LOR) as a function of noise intensity and stimulus familiarity, with 0 corresponding to equal odds of pecking to *Replaced* and *Added*. Error bars show 90% confidence intervals, and estimates are only for the first interval in which the oddball could appear (see Supplementary Fig. 2 for all intervals). Confidence intervals become wider with increasing noise due to fewer trials at higher noise levels. Ignoring familiarity, performance is significantly greater than chance with SNR ≥ 0 dB ($p < 0.003$) and worse than chance at –10 dB SNR (LOR $= -1.19$, $z = -2.20$, $p = 0.028$, 90% CI $= \{-2.24, -0.13\}$). Comparing familiar (orange) to unfamiliar motifs (gray), performance is better for familiar stimuli across all noise levels ($F_{1,\infty} = 21.9$, $p < 0.001$), including specifically at –10 dB (LOR $= 0.74$, $z = 2.49$, $p = 0.013$, 90% CI $= \{0.25, 1.23\}$). (**e**) Examples of masked stimuli. Noise was added to the entire motif instead of just the critical interval to eliminate the restoration illusion. (**f**) Hit and false alarm rates for masked stimuli. Hit rates were lower than for the occluded stimuli even at the highest signal-to-noise ratio (solid lines). False-alarm rates decreased as SNR decreased, indicating a different response strategy than for the auditory restoration task (dotted lines). (**g**) Estimates of average performance for masked stimuli. Task performance remains at or above chance for all SNR values for both familiar and unfamiliar stimuli. Data for –15 dB SNR are omitted because the model could not otherwise converge, given the near-zero hit and false-alarm rates for that noise level.

et al., 1998; Zokoll et al., 2007) or the use of alternative, guessing-based strategies (for example, if no gap is heard in the first four motifs, then the conditional probability of it being in the last motif is much higher than the marginal probability). Thus, the results presented here are taken from the estimates of performance for the first interval when the gap could occur. This is the most difficult condition, because the bird has the least information and will be least affected by perceptual anchoring.

The effect estimates from the model provided strong evidence that zebra finches are susceptible to the auditory restoration illusion (Fig. 1d). As expected, performance decreased as the noise intensity increased, dropping significantly below chance to an estimated LOR of -1.19 at –10 dB SNR. On the response scale, this means the average bird was over 3 times less likely to peck to the *Replaced* stimulus than would be expected from the baseline false-alarm rate.

As a control, the auditory restoration test trials were interleaved with trials in which the noise in the *Added* and *Replaced* conditions extended the entire length of the motif (Masked condition; Fig. 1e). In human and other animal studies, this kind of masking does not induce an illusory percept (Petkov et al., 2003), so performance is expected to drop to chance but not below. Overall, zebra finches performed worse on Masked trials, and there was a tendency for the false alarm rate to decrease at the loudest noise levels, perhaps because the noise itself was aversive (subjects essentially refused to respond at all at the highest noise level). Importantly however, performance on the masked task remained at or above chance (Fig. 1e,f).

In human studies, familiarity with the stimuli tends to strengthen the effect of phonemic restoration. Surprisingly, although we saw an effect of familiarity on performance, it was of the opposite sign to what we expected. The birds had a harder time detecting gaps in unfamiliar stimuli across the range of noise intensities, and performance dropped off more quickly as SNR decreased (Fig. 1d). At –10 dB SNR, although there was still a clear trend for performance to be worse than chance, the estimated LOR for familiar stimuli was not significantly less than zero (LOR $= -0.81$, $z = -1.47$, $p = 0.14$).

Overall, we found that by manipulating zebra finch song following the principles of phonemic restoration, we could detect behavioral evidence of an illusory percept of continuity. The perfor-

mance of zebra finches trained to detect discontinuities in conspecific song dropped significantly below chance as the intensity of the occluding noise increased, while their performance on control stimuli designed not to induce an illusory percept dropped to chance but not below.

Neural responses to the illusory percept

The behavioral results imply that at some level of the zebra finch's auditory system, neurons are responding to an illusion rather than to the stimulus that was physically present. To test this hypothesis, we made extracellular recordings of 407 single units ($n = 14$ birds) across the avian auditory cortex, including the caudal mesopallium (CM, $n = 56$ units), field L subunits L1 ($n = 25$ units), L2a ($n = 33$ units), and L3 ($n = 59$ units), and the caudomedial nidopallium (NCM, $n = 90$ units) (Fig. 2a). We made these recordings under anesthesia to further test whether illusory responses require top-down processing.

Zebra finches were presented with the illusion-inducing *Replaced* stimulus (RS) as in the behavioral experiment and with three control variants: the Continuous unmodified stimulus (CS); a Discontinuous stimulus (DS) where the note in the critical interval was elided but not replaced by any noise; and a Noise-only stimulus (NS) (Fig. 2b). Note that the CS and DS stimuli are similar to *Added* and *Replaced*, respectively, with noise at the lowest level.

We observed a wide variety of response patterns to these stimuli. Many individual neurons were highly sensitive to elision of the critical interval (DS) but responded to the RS stimulus as if it were continuous (Supplementary Fig. 5a,b). Across the population, each unit's average firing rate during the critical interval of the RS stimuli was nearly identical to its response in the corresponding interval of the CS stimuli, such that almost 90% of the variance in RS response rates could be predicted by the responses to CS stimuli. In contrast, the firing rates elicited by RS stimuli were significantly less like the responses to DS and NS (Supplementary Fig. 5c,d).

To test whether these similarities were indicative of an illusory response (i.e., a pattern of activity similar to the response evoked by the missing syllable), we used a linear decoder to predict what stimulus was being presented from the responses to the RS motifs. The parameters of the decoder were fit using ridge regression on data from the three control variants (CS, DS, and NS)

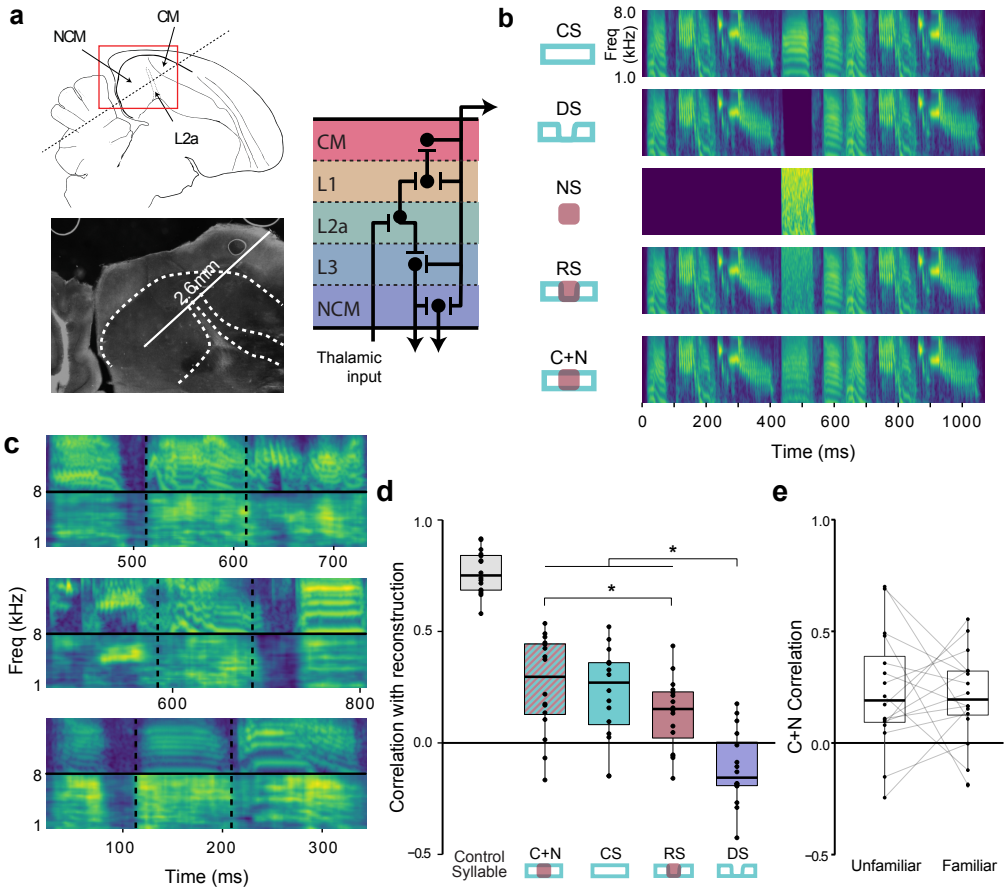


Fig. 2. Neural responses to occluded stimuli. (**a**) Location and microcircuitry of recording site. Counterclockwise from top: saggital diagram of zebra finch brain, with dashed line showing approximate orientation of the recording probe (adapted from Chen and Meliza, 2018); micrograph of auditory cortex showing DiI tracks; diagram of the subdivisions of the avian auditory cortex and their interconnections (adapted from Calabrese and Woolley, 2015). (**b**) Spectrograms of the four variants of an example motif used in extracellular recordings and a fifth variant used in analyses only (C+N). CS: continuous, unmanipulated motif; DS: discontinuous, with critical syllable deleted; NS: noise, a burst of white noise with the same duration as the critical syllable; RS: replaced, consisting of NS added to DS; C+N: an approximation of the expected illusory percept combining the CS syllable and the NS noise in the critical interval. (**c**) Reconstruction of three exemplar motifs from responses to RS stimuli. In each panel, the top row is a spectrogram of the CS stimuli surrounding the critical interval (dotted vertical lines), and the bottom row is the reconstruction. Times are relative to the start of the motif. (**d**) Similarity (correlation coefficient) of the reconstructed stimuli to the presented stimuli (CS, DS, and RS) and the expected illusory percept (C+N). The control syllable was outside the critical interval and was the same for CS, DS, and RS. The other groups show similarity for the critical interval in each of the four variants. Individual points and lines correspond to motifs. Boxplots show medians (horizontal line), interquartile ranges (boxes), and 10–90 percentile ranges (whiskers). Reconstructions were more similar to C+N than to RS (Kenward-Roger: $t_{60} = 2.24$, $p = 0.03$). The difference in similarity to CS and RS was not significant ($t_{60} = 1.59$, $p = 0.12$). Similarity to DS was lower than to C+N, CS, and RS ($p < 0.0001$). (**e**) Data were split by stimulus familiarity and the decoder was fit separately to each subset. Individual points show correlation coefficient between reconstructions and the C+N variant (as in **b**), with lines connecting each stimulus from the subset where it was unfamiliar to the subset where it was familiar. Across all stimuli, there was no significant effect of familiarity on reconstruction quality (paired t-test: $t_{28} = 0.29$, $p = 0.77$).

and then used to reconstruct the expected stimulus from the RS responses. The decoder yielded reconstructions that appeared to be a superposition of the missing syllable and the occluding noise, as if both the physical stimulus and the illusion were simultaneously present. In some cases the reconstruction looked remarkably similar to the CS motif (Fig. 2c, top panels), and in other cases the noise appeared to dominate (Fig. 2c, bottom panel). However, even the noisy reconstructions had spectrotemporal structure that bore some resemblance to the missing syllable.

To quantify similarity, we calculated the correlation coefficient between the reconstructed spectrogram and the spectrograms of the corresponding CS, DS, and RS variants (Fig. 2d). We also compared the reconstructions to a synthetic variant (C+N) comprising a 1:1 mixture of the missing syllable and the occluding noise (Fig. 2b, bottom). The baseline performance of the decoder was excellent despite its simplicity: for an unmanipulated syllable outside the critical interval, the correlation between the prediction and the actual stimulus was high ($r = 0.85 \pm 0.05$). Within the critical interval, the reconstructions were more similar to the C+N spectrograms than to the RS spectrograms, indicating that the population activity contains information about the acoustic structure of the occluded syllable, even though it was not physically present in the stimulus. The reconstructions were also more similar to the C+N, CS, and RS spectrograms than to DS, showing strong evidence against discontinuity.

As in the behavioral experiments, familiarity was counterbalanced such that half of the birds we recorded from were familiar with a different half of the motifs (Fig. 1b). We tested whether reconstruction was affected by familiarity by fitting the decoder separately to neurons from each of the two groups of subjects. Thus, one model was "familiar" with half of the motifs while the other was familiar with the other half. For each motif, we then compared the reconstructions from the familiar and the unfamiliar model to C+N. The average difference was not statistically distinguishable from zero (Fig. 2d). Thus, the avian auditory cortex is apparently able to reconstruct the acoustic structure of occluded syllables in songs that a bird has never heard.

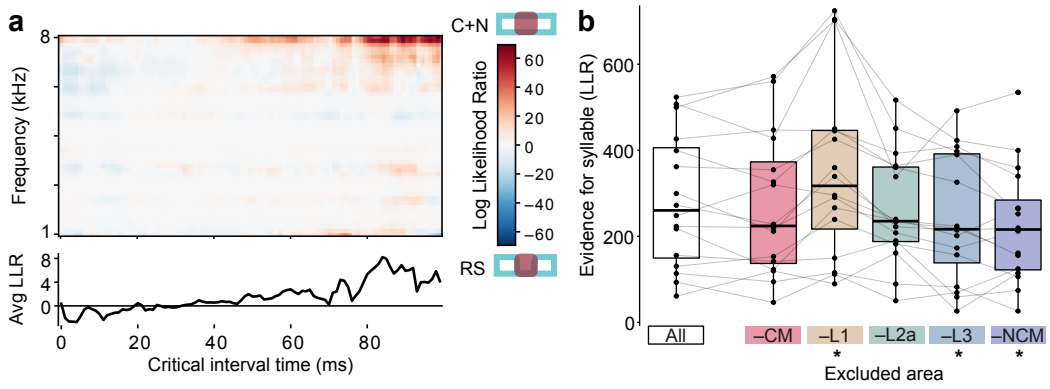


Fig. 3. Evidence for the missing syllable increases with time and is distributed throughout the cortical hierarchy. (**a**) Log of the posterior likelihood ratio for C+N over RS during the critical intervals as a function of frequency and time, averaged across all stimuli. Red indicates that given the response, the stimulus was more likely to be C+N than RS, and blue indicates RS was more likely than C+N. Bottom trace shows the average across all frequencies. (**b**) Change in the log likelihood ratio of C+N vs RS after fitting decoding models to data with all the units from one area left out compared to the evidence for C+N against the full model (see Materials and Methods). Note the sign reversal: a decrease in LLR indicates that the area was contributing evidence toward C+N. L1 units reduced the evidence for C+N relative to the full model ($p < 0.001, t_{60} = -6.55$), and L3 and NCM units increased the evidence for C+N ($p = 0.034, t_{60} = 2.38$ and $p = 0.002, t_{60} = 3.59$, respectively).

Responses become more illusory over the occluded interval

We hypothesize that the illusory percept of the missing syllable is generated by a predictive internal model and that the neural activity associated with this internal model is in conflict with ascending sensory input. The strength of the illusion could therefore depend on how effectively the response to noise is suppressed, and suppressing this activity could take time. To test this idea, we used the decoding model to calculate the amount of evidence in the neural response for the missing syllable. We quantified this as the log ratio of the (posterior predictive) likelihood that the presented stimulus was C+N over the likelihood that it was RS, for each time point and spectral channel during the critical interval (Fig. 3a).

Consistent with our hypothesis, we saw a strong change in the evidence for the missing syllable over the course of the critical interval. Initial responses were biased towards noise, perhaps reflecting the appearance of a new and potentially salient signal. By 20 ms, however, the evidence shifted toward the missing syllable and continued to strengthen over the course of the critical interval, supporting the view of a predictive internal model driving the neural response after the initial onset of noise.

There were also differences in evidence across the frequency spectrum, with particularly strong

evidence for the missing syllable in the highest frequencies near the limit of the zebra finch's hearing range (Fig. 3a). This result likely reflects the fact that the biggest differences in the power spectra of zebra finch song and white noise are in the higher frequencies (compare NS to CS in Fig. 2b). There were also distinct bands of evidence at lower frequencies, which could correspond to species-typical spectral structure.

Illusory responses are distributed throughout the auditory pallium

Where is the illusion of the missing syllable created? The avian auditory cortex comprises several laminar subdivisions organized in a hierarchy of increasingly complex functional properties (Sen et al., 2001; Meliza and Margoliash, 2012; Calabrese and Woolley, 2015). The illusion could be generated in one of these areas and then propagate to downstream areas. Alternatively, it could emerge in a more distributed manner or propagate from higher-order areas through feedback. To help distinguish between these possibilities, we tested how removing each of the five areas from the decoder impacted the evidence for the missing syllable. The advantage of this subtractive approach is two-fold: first, it allowed for similar numbers of units in each decoder, producing reconstructions comparable in quality to the full model; and second, it allowed us to identify the unique contributions of each auditory area. Because the decoding model accounts for correlated activity between neurons, removing redundant units will result in a shift of the linear weights to other cells, with little change to the model estimates. Therefore, if the posterior predictive likelihood of C+N decreases relative to RS after removing data from one area, we can conclude that area is making a non-redundant contribution to the illusory response.

With the full model, the evidence from the neural response was strongly in favor of the presence of the missing syllable for every motif tested (Fig. 3b). This result is consistent with the spectrogram correlations (Fig. 2B), but quantitatively different because the likelihood ratio accounts for posterior uncertainty. Removing specific areas from the model did not change the overall picture: evidence was still overwhelmingly for the missing syllable. However, some areas made significant, non-redundant contributions to the evidence for (or against) the missing syllable. L3 and NCM both increased the likelihood of C+N relative to RS, whereas L1 biased the reconstruction more toward the noise (Fig. 3b). CM and L2a did not make a statistically significant contribution in either

direction. Thus, although the illusory response appeared to be distributed throughout the auditory cortex, the deeper subdivisions (Fig. 2a) were the most responsive to the illusion compared to the physical stimulus.

Discussion

Perceptual illusions present a unique opportunity to expand our knowledge of how the brain uses internal models in sensory processing. Normally, internal models are aligned with incoming sensory input, but in an illusion, perception diverges from reality, and the neural activity produced by a model can be dissociated from the activity produced by the physical stimulus. In this study, we saw both behavioral and direct neural evidence that the zebra finch auditory cortex has an internal model of how conspecific songs should sound. This model biases auditory responses even under anesthesia and appears to be based on the general acoustic structure of song rather than specific memories.

The perceptual illusion we observed is analogous to phonemic restoration in humans (Warren, 1970; Warren and Sherman, 1974; Samuel, 1996; Bregman, 1999; Ishida and Arai, 2016). As in humans, it occurs when the occluding noise is brief and about the same intensity as the vocalization. Like all illusions, it is characterized by performance that becomes worse than chance, a sign that the subject is not simply guessing in the absence of evidence but convinced to some degree that the illusory percept is real (Petkov et al., 2003). By comparison, we found that when noise encompasses the entire stimulus, obscuring not only the gap but the all the acoustic features of the motif, performance remains at or above chance. These results are similar to those obtained in macaques using pure tones (Petkov et al., 2003) and consistent with observations in European starlings using a simpler and less stringent test (Seeba and Klump, 2009), but not with behavior in frogs (Seeba et al., 2010).

The strongest evidence for auditory restoration in zebra finches is that the brain responds as if the missing syllable were present (Bregman, 1999). We were able to detect this using a simple decoding model consisting of the best linear projection from the pattern of neural activity in the population to the power spectrum of the stimulus at a given instant in time (Mesgarani et al., 2009; Crosse et al., 2016). Note that the decoder has no information about the structure of the

stimulus except via the neural response. Moreover, it is simply a tool to analyze the patterns of activity in the population, not a hypothesis about the computation that allows the brain to infer or predict what incoming sensory inputs should look like. More complex, nonlinear decoders will surely do a better job of prediction, but at the cost of introducing more untested assumptions, and with a greater risk of overfitting (Glaser et al., 2020). We minimized overfitting here by using regularization and by training the model on responses to a large corpus of birdsong and noise bursts while holding out the trials with occluded stimuli for testing. When we used the model to infer what stimulus is being presented during the occluded portion of a song, the distribution of likely stimuli is overwhelmingly biased toward a mixture of the noise that was physically present and the syllable that should have been there. Human subjects report hearing a similar mixture of illusory and physical percepts in phonemic restoration.

In humans, the sensitivity of phonemic restoration to lexical and phonological context (Warren and Sherman, 1974; Samuel, 1996) is suggestive of top-down modulation actively shaping how lower-level auditory areas process ambiguous inputs. Such processes are generally assumed to be active, requiring attention and other cognitive resources (Heald and Nusbaum, 2014), but our results show this is not necessarily the case. Our observation of illusory responses in anesthetized animals suggests that local circuit dynamics can fill in missing information. These dynamics could be entrained by a lifetime of experience with zebra finch song through Hebbian plasticity of the dense network of recurrent connections within the auditory pallium (Wang et al., 2010; Calabrese and Woolley, 2015) and could bias the response of the network towards familiar or species-typical patterns. We observed this bias emerge quickly but not instantaneously: the response to the ambiguous stimulus segment is initially more consistent with noise but shifts to the missing syllable after about 20 ms, increasing throughout the critical interval (Fig. 3a). We hypothesize that peripheral adaptation to broadband noise may allow the cortical network dynamics to gradually dominate. These dynamics may be able to persist on their own for some time: in humans, the limit of illusory continuity in speech is approximately 300 ms (Bashford and Warren, 1987), and further work could characterize the full time-course of the neural illusion by recording responses to increasingly longer occlusions.

Surprisingly, both the behavioral and neural experiments indicate zebra finches do not require

familiarity with specific songs for auditory restoration to occur. European starlings show evidence of illusory continuity only for familiar stimuli (Braaten and Leary, 1999; Seeba and Klump, 2009), and in humans, phonemic restoration is stronger for words in the listener's native language compared to foreign or nonsense words (Samuel, 1996). These findings indicate that experience shapes the brain's internal models of what is being occluded by the noise. However, listeners do restore phonemes in non-native speech, just to a lesser degree (Ishida and Arai, 2016), suggesting that the internal models are not simply memorized images of specific phonemes or syllables. Instead, the auditory cortex may have a more general model of conspecific syllable and spectral patterns, and it may even be filling in a superposition of multiple candidates, as has been proposed for human speech recognition in the Trace model and its successors (McClelland and Elman, 1986; Heald and Nusbaum, 2014). Compared to starlings, zebra finches have a fairly limited set of vocal gestures, so it may be easier for them to learn a more general model of conspecific song. This effect may have been magnified in these experiments because all the songs came from the same colony.

We found evidence for the missing syllable distributed over all the areas we recorded, though the deeper auditory areas L3 and NCM show a somewhat greater bias towards the missing syllable (Fig. 3b). NCM is a secondary auditory area with high selectivity for conspecific song (Meliza and Margoliash, 2012), and it contains a subset of noise-invariant neurons (Schneider and Woolley, 2013) that could play a prominent role in sustaining the response to song through the occluding noise. Nevertheless, the entire auditory pallium appears to respond to the illusory syllable, even the thalamorecipient area L2a, which has highly linear response properties (Calabrese and Woolley, 2015). This implies that all these areas either participate in computing the illusory response or respond to recurrent connections from the neurons that do. The present study aggregates data from multiple subjects and averages across trials, obscuring individual- and trial-level dynamics that may yield more insight into the origin of the illusory activity. Recordings from animals engaged in the behavioral task could also reveal patterns of activity and specific areas that predict whether an individual animal responds to the illusion (Leonard et al., 2016).

Although language is a uniquely human behavior, these results imply that other animals employing complex vocalizations for communication have evolved similar mechanisms to compensate for acoustic noise and interference. Indeed, the use of internal models that actively modulate neural

activity at early stages of processing (Petkov and Sutter, 2011; Heald and Nusbaum, 2014) may be a general feature of neural systems (Rao and Ballard, 1999). The same internal models may also account for how populations of neurons throughout the avian auditory cortex are able to efficiently filter out "cocktail-party" noise from conspecifics (Narayan et al., 2007; Schneider and Woolley, 2013). Given the link between poor speech perception in noise and a number of learning disabilities (Bradlow et al., 2003), auditory restoration in zebra finches represents a powerful experimental system to investigate how internal models for acoustic communication are formed and tuned by experience.

Materials and methods

Animals

All animal use was performed in accordance with the Institutional Animal Care and Use Committee of the University of Virginia. Adult zebra finches were obtained from the University of Virginia breeding colony. Eight zebra finches (all male) were used for song familiarization, 12 (5 female) were trained on the behavioral experiment, and 14 (7 female) were used for extracellular recording.

Song recording and social familiarization

Recordings were made of the songs of eight adult male zebra finches to use as stimuli in the behavioral and electrophysiological experiments. Each singer was housed individually in a sound isolation box (Eckel Industries, Cambridge, MA) with *ad libitum* food and water on a 16:8 h light:dark schedule. A lavalier microphone (Audio-Technica Pro 70) was positioned in the box near a mirror to stimulate singing. The microphone signal was amplified and digitized with a Focusrite Scarlett 2i2 at 44.1 kHz, and recordings to disk were triggered every time the bird vocalized using Jill (https://github.com/melizalab/jill; version 2.1.4), a custom C++ real-time audio framework. A typical recording session lasted 1–3 days. From each bird's recorded corpus, a single representative motif was selected and high-pass filtered using a 4th-order Butterworth filter with a cutoff frequency of 500 Hz.

Subsequent to song recording, the eight males were randomly assigned to two groups of four and housed in group cages in separate rooms in the breeding colony. Experimental birds were housed in one of the group cages for at least one week to become familiar with the songs of the recorded males (Fig. 1B). Experimental birds were assigned essentially at random, but with the constraint that they had no prior social contact with the males in the group cage that they were not placed in. Thus, familiarity was counterbalanced, with half of the motifs familiar to a different half of the experimental subjects.

Behavioral experiment

Operant apparatus Behavioral experiments were run on a single-board computer (Beaglebone Black) with a custom expansion board (https://meliza.org/starboard, revision A2A) that interfaced with the operant manipulanda, cue lights, house lights, and feeder. The experiments were implemented using an event-driven framework our lab has developed for controlling behavioral experiments (https://github.com/melizalab/decide; version 3.2.1). Each subject was housed individually in an acoustic isolation box (Eckel Industries) with its own apparatus and single-board computer, which sent the trial data it collected to a centralized database (https://github.com/melizalab/django-decide-host; version 0.3.0).

Acoustic stimuli were presented by the single-board computer through an Altec Lansing Orbit iML227 USB speaker. The subject interacted with the apparatus by pecking an opening in a custom printed circuit board fitted with infrared beam-break detectors and cue lights. Reinforcement was standard finch seed, delivered through a custom 3D printed inlet housing a motorized screw shaft, which was advanced by a stepper motor for 500 ms to deliver approximately 3 seeds (median; range 0–13).

Shaping Following social exposure, behavioral subjects were moved to an acoustic isolation box and allowed to acclimate for 1–3 days. Throughout the subsequent shaping, training, and testing stages of the experiment, birds were maintained in the box on an semi-open economy. They received seed from the feeder at 5–10 minute intervals throughout the day, but the feeder was shut off at least 30 minutes before beginning a training or testing session. During sessions, food was

only available by completing trials. Behavior was monitored to ensure birds received adequate food, feeding intervals were adjusted to ensure the birds maintained surplus seed, and sessions were terminated if the bird went for more than 4 hours without eating.

Subjects were trained to peck the response panel using a standard autoshaping paradigm. First, a cue light located near the opening was lit just before automatic food delivery. Once the bird started pecking at the opening in anticipation of food, automatic food delivery stopped, and reinforcement was only given after pecking. It was often helpful to suspend a small piece of string in the opening during this initial shaping stage to encourage exploration. There were three blocks of 100 trials: in the first block, the bird had to peck the lit opening once; in the second, it had to peck twice; and in the third, the bird had to peck twice, but the cue light was eliminated. One of the 12 birds was excluded during this stage because it failed to acquire the pecking behavior.

Stimuli For each of the eight motifs, eight variants were constructed using a $2 \times 2 \times 2$ design (fig. 1). The first factor was the timing of the critical interval chosen for manipulation. Two non-overlapping intervals were chosen for each motif, avoiding both the first and last notes of the motif. The second factor was whether the motif was continuous or discontinuous. The discontinuous variants were constructed by deleting the sound in the critical interval. The third factor was the duration of white noise added to the motif. In the occluding case, the noise was only present during the critical interval; in the masked case, the noise was present throughout the motif. Following Petkov et al. (2003), edge artifacts were minimized by applying a 3 ms cosine ramp to the onsets and offsets of the noise and the gaps in the occluding case. In the masked case, noise onset and offset used a 25 ms cosine ramp.

The stimuli were amplified so that the unmodified motifs all had a RMS amplitude of 50 ± 2 dB SPL at the location where the bird interacted with the operant response panel, as measured by an NTi Audio XL2 Sound Level Meter. For each variant, the amplitude of the white noise was varied relative to the song in 5 dB increments between 20 and -15 dB signal-to-noise ratio (SNR), corresponding to SPLs between 30 and 65 dB. Thus, there were a total of 512 different stimuli. For an example of the full set of stimuli generated for one motif, see fig. 1.

Behavioral task Zebra finches had to be trained in several stages to learn the experimental task. In the first stage, the bird pecked to initiate a trial in which a single song motif was presented. A second peck within 1 s of the end of the motif was rewarded. In the second stage, the bird listened to two motifs and was rewarded for withholding a peck until the second motif was presented.

After shaping was completed, birds were trained on the main task, which was to detect a motif with a gap in a sequence of otherwise identical motifs without gaps (Fig. 1A), presented with 200 ms inter-stimulus intervals. The position of the discontinuous motif in the sequence was random, but the first motif was always continuous. The finches had a short window from the start of the gap in the discontinuous motif to peck for a correct response, which was rewarded with seed. Pecking at any other time during the trial was a false alarm, and failure to peck when there was a discontinuous motif was a miss. False alarms and misses were punished with a 2 s "time-out" during which the house lights were extinguished and trials could not be initiated. Stimulus playback ended immediately after a peck response, whether correct or incorrect. On 20% of the trials, all of the motifs were continuous, so the correct response was to withhold a peck, and a reward was delivered at the end of the motif set. Supplementary Movie 1 is an annotated video of a well-trained bird performing the task.

The initial training only used familiar motifs, and only the variants with occluding noise. To make the starting difficulty as low as possible while still allowing the birds to learn the task, the SNR was 20 dB, the critical interval was 200 ms, the sequence was three motifs long, and the finches had a 2 s window from the start of the gap in the discontinuous motif to peck for a correct response, which was rewarded with food. The difficulty of the task was progressively increased by incrementally shortening the response window from 2 s to 1 s, then reducing the critical interval from 200 ms to 150 ms to 100 ms, then increasing the number of motifs from 3 to 4 to 5. For the purpose of tracking performance, the log-odds ratio (LOR) was calculated empirically from sliding blocks of 50 trials as

$$LOR = \frac{1}{2}\log\left(\frac{H}{M} \times \frac{CR}{FA}\right),$$

where H is the proportion of trials with hits, M is the proportion with misses, CR is the proportion with correct rejections, and FA is the proportion with false alarms. Note that this LOR is calculated

on a per-trial basis and does not account for the fact that on many trials, birds have to correctly reject multiple *Added* stimuli. To be included in the study, birds had to achieve and maintain performance above an LOR of 1. Birds that failed to show a systematic increase in performance over a two week period were excluded. Five of the 12 birds were excluded during task-specific shaping. The remaining six subjects learned the task within 15700 ± 7800 trials and achieved a final LOR of 1.15 ± 0.09 (see table 1 for full breakdown of trials).

Five of the 12 birds had previously been trained on a similar oddball detection task where the oddball was an entirely different motif. Of those birds, three successfully learned the new task and were included in this study, and two were excluded during task-specific shaping.

Testing auditory restoration

Auditory restoration is characterized by an illusion of continuity when a gap in a stimulus is occluded by noise of similar amplitude. Thus, as the noise amplitude increases, the odds of responding during the correct interval is expected to decline while the false alarm rate remains constant. We tested this prediction by systematically varying the amplitude of the occluding noise while holding the motif amplitude constant, so that SNR ranged from 20 to –15 dB. The noise level was the same for all motifs within a trial. Trials with louder noise were introduced in stepwise blocks. That is, we started with all the trials at 20 dB SNR, then switched to a new block with 30% trials at 15 dB SNR. In the next block, there were 22% trials at 15 dB, 22% trials at 10 dB, and so forth. We stopped on the block where the lowest SNR was 0 dB (50 dB SPL) or where the performance on the lowest-SNR stimuli dropped below chance, whichever came later. This procedure ensured that a large proportion of the trials in each block were relatively easy for the bird, which helped to maintain baseline performance and reduced the likelihood that birds would become frustrated and switch to a guessing-based strategy.

During testing sessions, the unfamiliar motifs were included for the first time, and trials with occluded variants were randomly interleaved with trials with masked variants. The noise intensity distribution in the masked trials was matched to that of the occluded trials. Masking noise is an important control, because it also makes it more difficult for the birds to detect gaps in the motif, but without inducing an illusion of continuity, so the decline in performance with noise amplitude

is expected to be shallower. Moreover, without an illusory percept to fool the bird into thinking the stimulus is continuous, performance is expected to decline to chance but not below it.

Data analysis The data from the behavioral experiment comprised a total of 63,134 trials from six birds. After an initial exploratory analysis, we excluded one bird because its false alarm rate on the catch trials (with no gap) was greater than its hit rate on the other trials. This left 50,809 trials (range 5085–13,611 per bird). Each trial was split into 1–5 intervals, one for each motif the bird heard (recall that trials terminated immediately after the bird responded). Each interval was coded with a single binary dependent variable, *peck*, which was 1 if the bird pecked during the interval and 0 otherwise, and with seven independent variables: *dB*, the amplitude of the noise (coded as a factor with 8 levels); *gap*, the presence of a gap in the motif; *position*, the position of the interval in the sequence (coded as a factor); *familiarity*, whether the subject had been exposed to the motif in social housing; *condition*, whether the noise was occluding or masking; *bird*, the identity of the subject (coded as a factor); and *song*, the identity of the motif variant (two gap positions per motif, coded as a factor).

A generalized linear mixed-effects model (GLMM) was used to infer the effects of noise intensity and familiarity on the subjects' ability to detect the motif with a gap in it. *Peck* was modeled as a binomial random variable with log odds that depended on a linear function of *dB*, *gap*, *position*, *familiarity*, and their interactions. Two of the higher-order interactions (*dB:position:familiarity* and *dB:gap:position:familiarity*) had to be removed for the parameter estimation to converge. Random effects were included to account for variations in psychophysical curves associated with subject and motif identity (fig. 3 and fig. 4), with motif identity nested in subjects. The parameters were estimated in R using lme4 (version 1.1-23). The model specification was as follows:

```
peck ~ 0 + dB*gap*familiarity + dB*gap*position + (0 + ndB |
    subject/song) + (0 + ndB:gap | subject/song)
```

In this formulation, the parameters have the following interpretations: *dB* gives the false alarm odds for unfamiliar stimuli at each of the eight noise levels; *dB:gap* gives the log odds ratio of pecking when a gap is actually present, again for unfamiliar stimuli and at each of the noise levels; *dB:familiarity* gives the effect of familiarity on the log odds of false alarms; and *dB:gap:familiarity*

gives the effect of familiarity on the log odds ratio for detecting a gap. The *dB:position* and *dB:gap:position* parameters correspond to the effect of position within the sequence on false alarm rate and performance respectively. These are essentially nuisance parameters, but they need to be included because the intervals are otherwise not independent (the probability of pecking in later intervals depends on the bird not pecking earlier in the trial). Because the model is nonlinear and has many interactions, effects and confidence intervals are reported as estimated marginal means, calculated using the *emmeans* R package (version 1.5.4).

The data for trials with masking noise were analyzed using the same model, but separately. This was to avoid introducing another interaction into an already complex model.

Extracellular recordings

Extracellular stimuli The stimuli for the extracellular recordings had the same basic structure, but we allowed the duration of the critical interval to vary and included some additional variants. Critical intervals were selected to overlap completely with a single syllable in the motif but were never longer than 100ms. The design was 2×4: two critical intervals per motif, and four variants (Fig. 2B). The variants comprised the continuous stimulus (CS), which was unaltered; the discontinuous stimulus (DS), which replaced the song note within the critical interval with silence; the noise-only stimulus (NS), which was a segment of white noise spanning the critical interval; and the replaced stimulus (RS), which replaced the note within the critical with white noise to produce the illusory perception of the song continuing behind the noise. As with the behavioral stimuli, 2 ms ramps were applied to the edges of the noise and gaps. The noise amplitude was +15 dB relative to the motif amplitude and was not varied.

Surgery Birds were anesthetized with isoflurane inhalation (1–3% in O2) and placed in a stereotaxic apparatus (Kopf Instruments). An incision was made in the scalp, and the skin was retracted from the skull. The recording site was identified using stereotaxic coordinates relative to the Y-sinus. A metal pin was affixed to the skull rostral to the recording site with dental cement, and the skull over the recording site was shaved down but not completely removed. The bird was allowed to recover completely for several days prior to recording.

On the day of recording, the bird was anesthetized with three intramuscular injections of 20% urethane spaced half an hour apart. The bird was placed in a 50 mL conical tube, and the head pin was attached to a stand in the recording chamber. The thin layer of skull remaining over the recording site was removed along with the dura, and a well was formed around the recording site and filled with phosphate-buffered saline.

Stimulus presentation Stimuli were presented with the `sounddevice` python library (version 0.3.10) through a Samson Servo 120a amplifier to a Behringer Monitor Speaker 1C. The RMS amplitude of the unmodified motifs was 70 dB SPL. Stimuli were presented in a pseudorandom order to minimize stimulus adaptation, with 1 s between each song. Each stimulus was presented 10 times.

Data acquisition Neural recordings were made using a NeuroNexus 32-channel probe in a four-shank, linear configuration (A4x8-5mm-100-400-177-A32) connected to an Intan RHD2132 Amplifier Board. Data were collected by the Open Ephys Acquisition Board and sent to a computer running Open Ephys GUI software (version 0.4.6).

The recording electrode was coated with DiI (Invitrogen). The electrode was inserted at a dorso-rostral to ventro-caudal angle that allowed for recording of all auditory forebrain regions with a single penetration (Fig. 2A). The probe was lowered into the brain until the local field potentials across channels and shanks showed coordinated responses to birdsong, and the probe was allowed to rest in place for half an hour to ensure a stable recording. Recordings of responses were made across all 32 channels. After the recording, the probe was moved to successively deeper regions of the auditory pathway and additional recordings were made.

Histology After recording, birds were administered a lethal intramuscular injection of Euthasol and perfused transcardially with a 10 U/mL solution of sodium heparin in PBS (in mM: 10 Na2HPO4, 154 NaCl, pH 7.4) followed by 4% formaldehyde (in PBS). Brains were immediately removed from the skull, postfixed overnight in 4% formaldehyde at 4 °C, cryoprotected in 30% sucrose (in 100 mM Na2HPO4, pH 7.4), blocked saggitally into hemispheres or on a modified

coronal plane (Chen and Meliza, 2018), embedded in OCT, and stored at –80 °C. 60 µm sections were cut on a cryostat and mounted on slides. After drying overnight, the sections were rehydrated in PBS and coverslipped with Prolong Gold with DAPI (ThermoFisher, catalog P36934; RRID:SCR_015961). Sections were imaged using epifluorescence with DAPI and Texas Red filter cubes to located DiI-labeled penetrations. Images of the electrode tracks were used to identify the locations of recorded units (Fig. 2A).

Spike sorting Spikes were sorted offline using MountainSort 4. Single units were further curated by visual inspection for spheroid PCA cluster shape, very low refractory period violations in the autocorrelogram, and stability of the unit throughout the recording. These high-quality single units were included in the dataset if they showed a clear, phase-locked auditory response to at least one stimulus.

Stimulus reconstruction

A linear stimulus decoding model was used to predict stimulus spectrograms from recorded neural responses (Mesgarani et al., 2009; Crosse et al., 2016; Leonard et al., 2016). The model is similar to the spectrotemporal receptive field (STRF), in which the expected firing rate of a single neuron at a given time point t is modeled as a linear function of the stimulus spectrogram immediately prior to t. In the linear decoding model, the relationship is reversed, and the expected stimulus at time t is modeled as a linear function of the response that follows. Using a discrete time notation where s_t is the stimulus in the time bin around t, and r_t is the response of a single neuron in the same time bin, then the expected value of the stimulus is given by

$$\mathrm{E}(s_t) = r_t g_0 + r_{t+1} g_1 + \cdots + r_{t+k} g_k,$$

where k is the number of time bins one looks into the future, and $\mathbf{g} = (g_0, g_1, \ldots, g_k)$ are the linear coefficients of the model. If the errors are independent and normally distributed around the expectation with constant variance σ^2, then this is a ordinary linear model. If there are n time bins

in the stimulus, then the stimulus is a vector $\mathbf{s} = (s_0, \ldots, s_n)$ drawn from a multivariate normal distribution. In vector notation,

$$\mathbf{s}|\mathbf{g}, \sigma^2, \mathbf{R} \sim N(\mathbf{Rg}, I\sigma^2),$$

where $\mathbf{R}$ is the $n \times k$ Hankel matrix of the response. Without any loss of generality, the model can be expanded to include the responses of multiple neurons. If there are p neurons, then r_t becomes a p-element vector $(r_{1,t}, \ldots, r_{p,t})$, $\mathbf{R}$ becomes a $n \times pk$ matrix formed by concatenating the Hankel matrices for each of the neurons, and $\mathbf{g}$ becomes a pk-element vector.

Because the model is simply linear regression, standard tools can be used to estimate the parameters $\mathbf{g}$ and σ^2. Using a ridge penalty for regularization, the maximum likelihood estimate of $\mathbf{g}$ is

$$\hat{\mathbf{g}} = \left(\mathbf{R}^\top \mathbf{R} + \lambda \mathbf{I}\right)^{-1} \mathbf{R}^\top \mathbf{s},$$

where $\lambda\mathbf{I}$ is the identity matrix multiplied by the shrinkage penalty for the ridge regression.

Additional regularization can be achieved by projecting the response matrix $\mathbf{R}$ into an alternative basis set, such as a non-linearly spaced series of raised cosines (Pillow et al., 2005). The width of each basis function increases with lag, which gives the model high temporal resolution at short lags and lower resolution at longer lags. This allows the inclusion of longer lags without exploding the number of parameters.

In this study, the response matrix was constructed from the peristimulus time histograms (PSTHs) averaged over 10 trials for each of the 407 auditory single units from all 14 birds, using a bin size of 1 ms. Stimuli were converted to time-frequency representations using a gammatone filter bank, implemented in the Python package `gammatone` (version 1.0) with 50 log-spaced frequency bands from 1–8 kHz, a window size of 2.5 ms, and a step size of 1 ms. Power was log-transformed with a constant offset of 1, giving the transformed signal a lower bound of 0 dB.

The parameters were estimated with data from the Continuous (CS), Discontinuous (DS), and Noise-only (NS) stimuli, leaving out the Replaced (RS) responses for prediction, using the Python machine-learning library *scikit-learn* (version 0.23.0). We combined all 50 spectral bands into a single multivariate multiple regression, and then used 4-fold cross-validation to determine the best values for the ridge penalty λ, the number of time lags k, the number of raised-cosine basis functions, and a linearity parameter that controls the spacing of the basis functions. Basis functions were defined as in Pillow et al. (2005). Because the models differed in the number of parameters, Aikake information criteron (AIC) was used for scoring. For the reconstructions presented here, the optimal value for λ was 8.59, k was 300, the number of basis functions was 30, and the linearity factor was 30.

After fitting the model, we used the parameter estimates to decode the stimulus from the responses to the RS stimuli. Using $\tilde{\mathbf{R}}$ to denote these responses, the predicted stimulus is calculated as

$$\mathrm{E}(\tilde{\mathbf{S}}) = \tilde{\mathbf{R}}\hat{\mathbf{g}} \tag{1}$$

To quantify how similar the decoded stimulus was to the actual stimulus and the other variants, we calculated the correlation coefficient between $\tilde{\mathbf{S}}$ and $\mathbf{S}_{CS}$, $\mathbf{S}_{RS}$, $\mathbf{S}_{DS}$, and $\mathbf{S}_{C+N}$ within the critical interval for each of the motifs. $\mathbf{S}_{C+N}$ comprised a 1:1 mixture of $\mathbf{S}_{CS}$ and $\mathbf{S}_{RS}$, representing the expected illusory percept. As a baseline for how good the reconstruction could be, we calculated the correlation between $\tilde{\mathbf{S}}$ and $\mathbf{S}_{RS}$ in an interval outside the critical interval (which was the same for all variants).

Unless otherwise noted, all comparisons were fit with a mixed-effects linear model using `lme4` (version 1.1-23) in R and `emmeans` (version 1.5.2-1) was used to calculate the effects, confidence intervals, and significance.

Posterior predictive likelihood

Spectrogram correlations are a simple and easily interpretable way to quantify the similarity of the decoded stimulus to the physical stimulus and the illusion, but what we really want to know is, given the response to RS, how likely is it that the stimulus included the missing syllable? Equivalently, how much evidence is there in the neural response that the missing syllable was present? To answer this question, we used the decoding model to compute the posterior probability that the predicted stimulus $\mathrm{E}(\tilde{\mathbf{S}})$ was the expected illusory percept $\mathbf{S}_{C+N}$, relative to the probability that it was the actual stimulus physically presented to the animal $\mathbf{S}_{RS}$.

For a linear regression model, the posterior predictive distribution conditional on the observations used to fit the model (Gelman et al., 2020) is a multivariate t distribution with a mean given by Equation (1), $n-k$ degrees of freedom, and scale matrix $s^2(\mathbf{I}+\tilde{\mathbf{R}}\mathbf{V}_g\tilde{\mathbf{R}}^\top)$, where s^2 is the sample variance of the residuals,

$$s^2 = \frac{1}{n-k}(\mathbf{S}-\mathbf{R}\hat{\mathbf{g}})^\top(\mathbf{S}-\mathbf{R}\hat{\mathbf{g}}),$$

and $\mathbf{V_g}$ is the posterior variance of the parameter estimates,

$$\mathbf{V_g} = (\mathbf{R}^\top\mathbf{R})^{-1}.$$

Thus, the posterior predictive uncertainty reflects both the unexplained variance in the model (s^2) and the posterior uncertainty in the parameter estimates ($\mathbf{V_g}$). Because $n-k$ was very large (27,725), we approximated this distribution with a multivariate normal with mean $\mathbf{R}\hat{\mathbf{g}}$ and variance s^2 and calculated $\Pr(\tilde{\mathbf{S}}=\mathbf{S}_{C+N}|\mathbf{S})$ and $\Pr(\tilde{\mathbf{S}}=\mathbf{S}_x|\mathbf{S})$. These probabilities (or likelihoods) are not meaningful on their own, but the log of their ratio (ℓ) quantifies how much more likely it was (given the response) that the stimulus was the combined continuous and noise stimulus (C+N) or just the noise (RS). By analogy to Bayes Factors, this ratio can also be thought of as the amount of evidence

in the neural response for the presence of the missing syllable. Because $p(\tilde{\mathbf{S}}|\mathbf{S})$ is multivariate normal, with each dimension corresponding to each time-frequency point in the stimulus, we can either evaluate the marginal likelihood for each point separately to produce a spectrotemporal evidence plot (Fig. 3A) or use the joint distribution to obtain a single value of ℓ for each critical interval (Fig. 3B). Note that outside the critical interval, ℓ is always zero because $\mathbf{S}_{RS}$ and $\mathbf{S}_{C+N}$ are identical.

Area-level analysis

To understand how different auditory areas in the zebra finch auditory cortex influence the overall reconstruction, we performed a leave-one-area-out analysis. Using the same cross-validated parameters as the full reconstruction, we fit the model to the entire dataset minus the units from one area. This technique allowed us to identify unique contributions made by each of the auditory area because if the responses of one area were redundant with those of another, the model would shift its weights but the estimated reconstruction would remain unchanged. Changes in the reconstruction seen with this method mean that the model lost irreplaceable evidence when the units from one area were omitted.

To quantify the changes in reconstructions, we calculated the log likelihood ratio between $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{C+N}|\mathbf{S})$ and $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{RS}|\mathbf{S})$, using the area-specific reconstructions. We used the joint distribution and compared this value $\ell^{(i)}$ to the log likelihood ratio ℓ for the full model. A decrease in ℓ relative to the full model therefore implied that the removed area caused $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{C+N}|\mathbf{S})$ to increase relative to $\Pr(\tilde{\mathbf{S}} = \mathbf{S}_{RS}|\mathbf{S})$, and that neurons in that area were non-redundantly contributing evidence for the missing syllable.

Acknowledgments

We would like to thank Samantha Moseley for assistance with the behavioral experiment, Leah Kiely for assistance with histology, and Ayush Sagar and Crystal Gong for their work developing and testing the operant apparatus used in this study. This work was supported by NSF IOS-1942480 and NIH R01DC018621

References

Bashford JA, Warren RM. Multiple phonemic restorations follow the rules for auditory induction. Percept Psychophys 1987 Aug;42(2):114–121. doi: 10.3758/bf03210499.

Braaten RF, Leary JC. Temporal Induction of Missing Birdsong Segments in European Starlings. Psychological Science 1999 May;10(2):162–166. doi: 10.1111/1467-9280.00125.

Bradlow AR, Kraus N, Hayes E. Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear Res 2003 Feb;46(1):80–97. doi: 10.1044/1092-4388(2003/007).

Braida LD, Lim JS, Berliner JE, Durlach NI, Rabinowitz WM, Purks SR. Intensity perception. XIII. Perceptual anchor model of context-coding. The Journal of the Acoustical Society of America 1998 Jun;76(3):722. doi: 10.1121/1.391258.

Bregman AS. Auditory scene analysis. 2nd edition ed. The perceptual organization of sound, MIT Press; 1999.

Calabrese A, Woolley SMN. Coding principles of the canonical cortical microcircuit in the avian brain. PNAS 2015 Mar;112(11):3517–3522. doi: 10.1073/pnas.1408545112.

Chen AN, Meliza D. Phasic and Tonic Cell Types in the Zebra Finch Auditory Caudal Mesopallium. J Neurophys 2018 Mar;119(3):1127–1139. doi: 10.1152/jn.00694.2017.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front Hum Neurosci 2016;10:604. doi: 10.3389/fnhum.2016.00604.

Elie JE, Theunissen FE. The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. Anim Cogn 2016 Mar;19(2):285–315. doi: 10.1007/s10071-015-0933-6.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Introduction to regression models. In: Bayesian Data Analysis Chapman & Hall; 2020.p. 353–378.

Gilbert CD, Sigman M. Brain states: top-down influences in sensory processing. Neuron 2007 Jun;54(5):677–696. doi: 10.1016/j.neuron.2007.05.019.

Glaser JI, Benjamin AS, Chowdhury RH, Perich MG, Miller LE, Kording KP. Machine Learning for Neural Decoding. eneuro 2020 Jul;7(4). doi: 10.1523/ENEURO.0506-19.2020.

Heald SLM, Nusbaum HC. Speech perception as an active cognitive process. Front Syst Neurosci 2014;8:35. doi: 10.3389/fnsys.2014.00035.

Ishida M, Arai T. Missing phonemes are perceptually restored but differently by native and non-native listeners. Springerplus 2016;5(1):713–10. doi: 10.1186/s40064-016-2479-8.

Komatsu H. The neural mechanisms of perceptual filling-in. Nat Rev Neurosci 2006 Mar;7(3):220–231. doi: 10.1038/nrn1869.

Leonard MK, Baud MO, Sjerps MJ, Chang EF. Perceptual restoration of masked speech in human cortex. Nat Commun 2016 Dec;7(1):13619–9. doi: 10.1038/ncomms13619.

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition 1985 Oct;21(1):1–36.

McClelland JL, Elman JL. The TRACE model of speech perception. Cogn Psychol 1986;18(1):1–86.

Meliza CD, Margoliash D. Emergence of selectivity and tolerance in the avian auditory cortex. J Neurosci 2012 Oct;32(43):15158–15168. doi: 10.1523/JNEUROSCI.0845-12.2012.

Mesgarani N, David SV, Fritz JB, Shamma SA. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J Neurophys 2009 Dec;102(6):3329–3339. doi: 10.1152/jn.91128.2008.

Miller CT, Dibble E, Hauser MD. Amodal completion of acoustic signals by a nonhuman primate. Nat Neurosci 2001 Aug;4(8):783–784. doi: 10.1038/90481.

Miller GA, Licklider JCR. The Intelligibility of Interrupted Speech. J Acoust Soc Am 1950 Jun;22(2):167–173. doi: 10.1121/1.1906584.

Narayan R, Best V, Ozmeral E, McClaine E, Dent ML, Shinn-Cunningham B, et al. Cortical interference effects in the cocktail party problem. Nat Neurosci 2007;10(12):1601–1607.

Petkov CI, O'Connor KN, Sutter ML. Illusory sound perception in macaque monkeys. J Neurosci 2003 Oct;23(27):9155–9161. doi: 10.1523/JNEUROSCI.23-27-09155.2003.

Petkov CI, O'Connor KN, Sutter ML. Encoding of illusory continuity in primary auditory cortex. Neuron 2007 Apr;54(1):153–165. doi: 10.1016/j.neuron.2007.02.031.

Petkov CI, Sutter ML. Evolutionary conservation and neuronal mechanisms of auditory perceptual restoration. Hear Res 2011 Jan;271(1-2):54–65. doi: 10.1016/j.heares.2010.05.011.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ. Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. J Neurosci 2005 Nov;25(47):11003–11013. doi: 10.1523/JNEUROSCI.3305-05.2005.

Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 1999 Jan;2(1):79–87. doi: 10.1038/4580.

Samuel AG. Does lexical information influence the perceptual restoration of phonemes? J Exp Psychol Gen 1996;125(1):28–51. doi: 10.1037/0096-3445.125.1.28.

Schneider DM, Woolley SMN. Sparse and background-invariant coding of vocalizations in auditory scenes. Neuron 2013 Jul;79(1):141–152. doi: 10.1016/j.neuron.2013.04.038.

Seeba F, Klump GM. Stimulus familiarity affects perceptual restoration in the European starling (Sturnus vulgaris). PLoS ONE 2009 Jun;4(6):e5974. doi: 10.1371/journal.pone.0005974.

Seeba F, Schwartz JJ, Bee MA. Testing an auditory illusion in frogs: Perceptual restoration or sensory bias? Anim Behav 2010 Jun;79(6):1317–1328. doi: 10.1016/j.anbehav.2010.03.004.

Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. J Neurophysiol 2001 Sep;86(3):1445–1458.

Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. J Acoust Soc Am 2003 Dec;114(6):3394–3411.

Sugita Y. Neuronal correlates of auditory induction in the cat cortex. Neuroreport 1997 Mar;8(5):1155–1159. doi: 10.1097/00001756-199703240-00019.

Wang Y, Brzozowska-Prechtl A, Karten HJ. Laminar and columnar auditory cortex in avian brain. PNAS 2010 Jul;107(28):12676–12681. doi: 10.1073/pnas.1006645107.

Warren RM. Perceptual restoration of missing speech sounds. Science 1970 Jan;167(3917):392–393. doi: 10.1126/science.167.3917.392.

Warren RM, Sherman GL. Phonemic restorations based on subsequent context. Percept Psychophys 1974 Jan;16(1):150–156. doi: 10.3758/BF03203268.

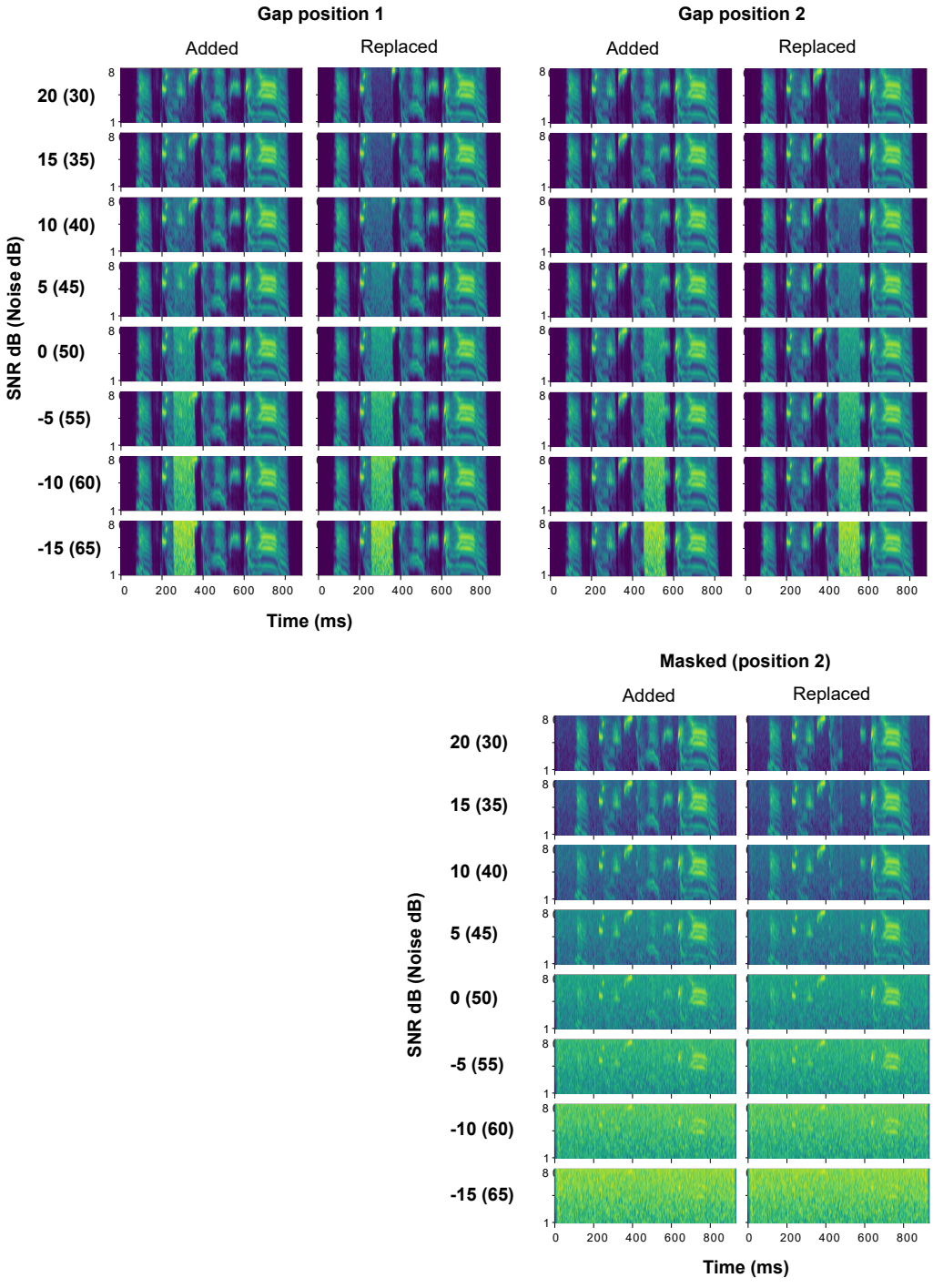
Zokoll MA, Klump GM, Langemann U. Auditory short-term memory persistence for tonal signals in a songbird. J Acoust Soc Am 2007 May;121(5):2842. doi: 10.1121/1.2713721.

Supplementary Materials

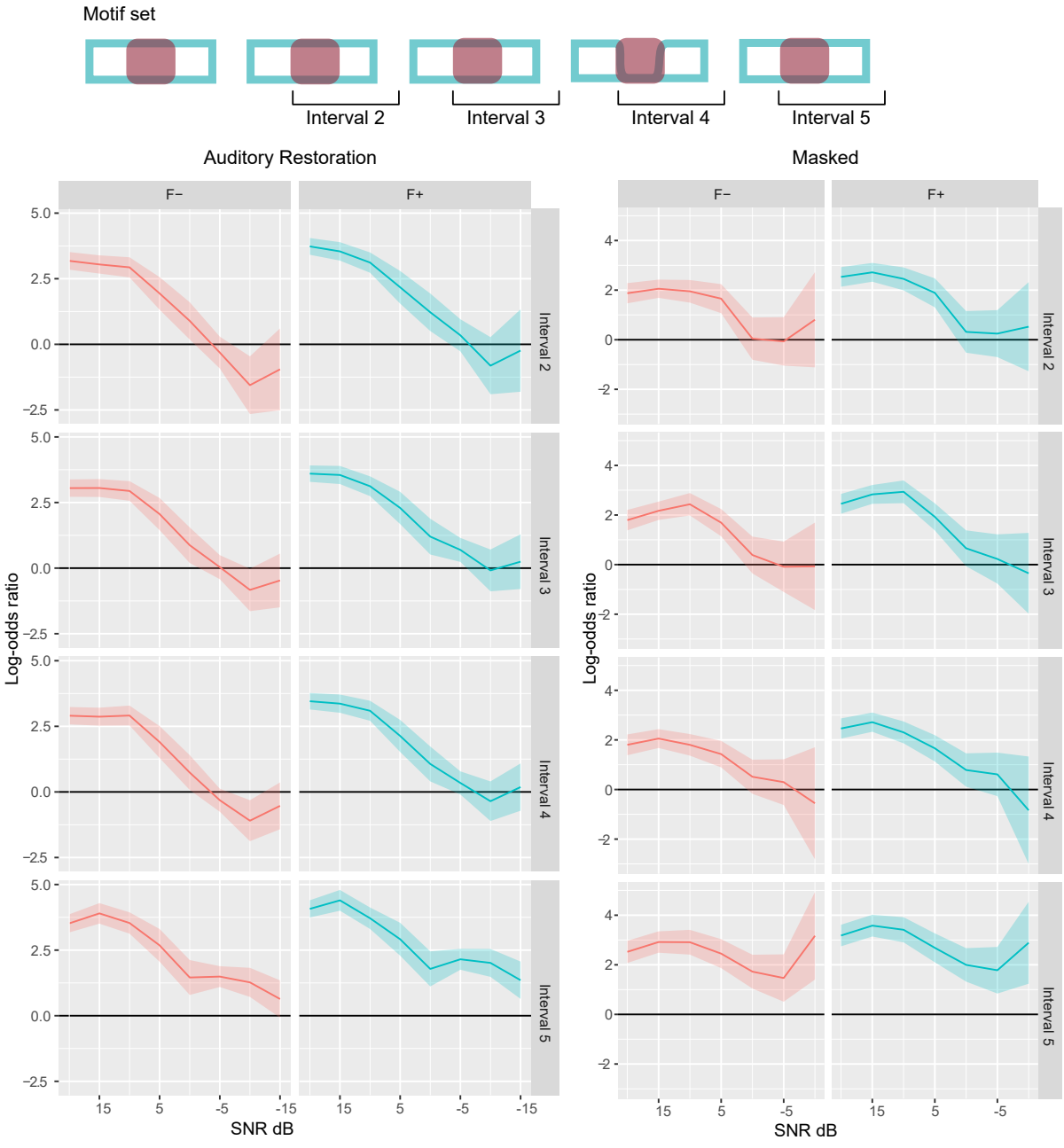
Supplementary Figures 1–5

Supplementary Table 1

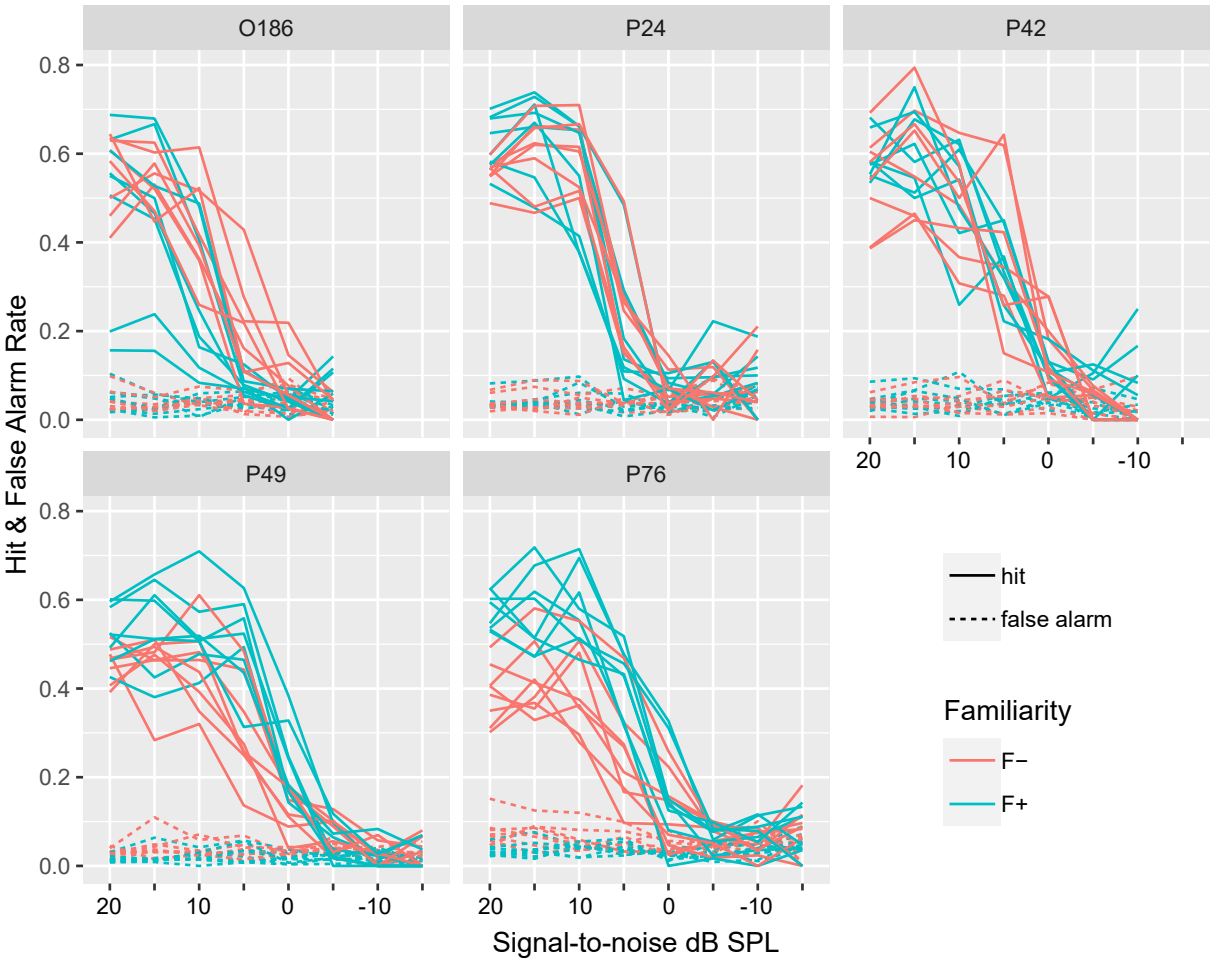
Supplementary Movie 1



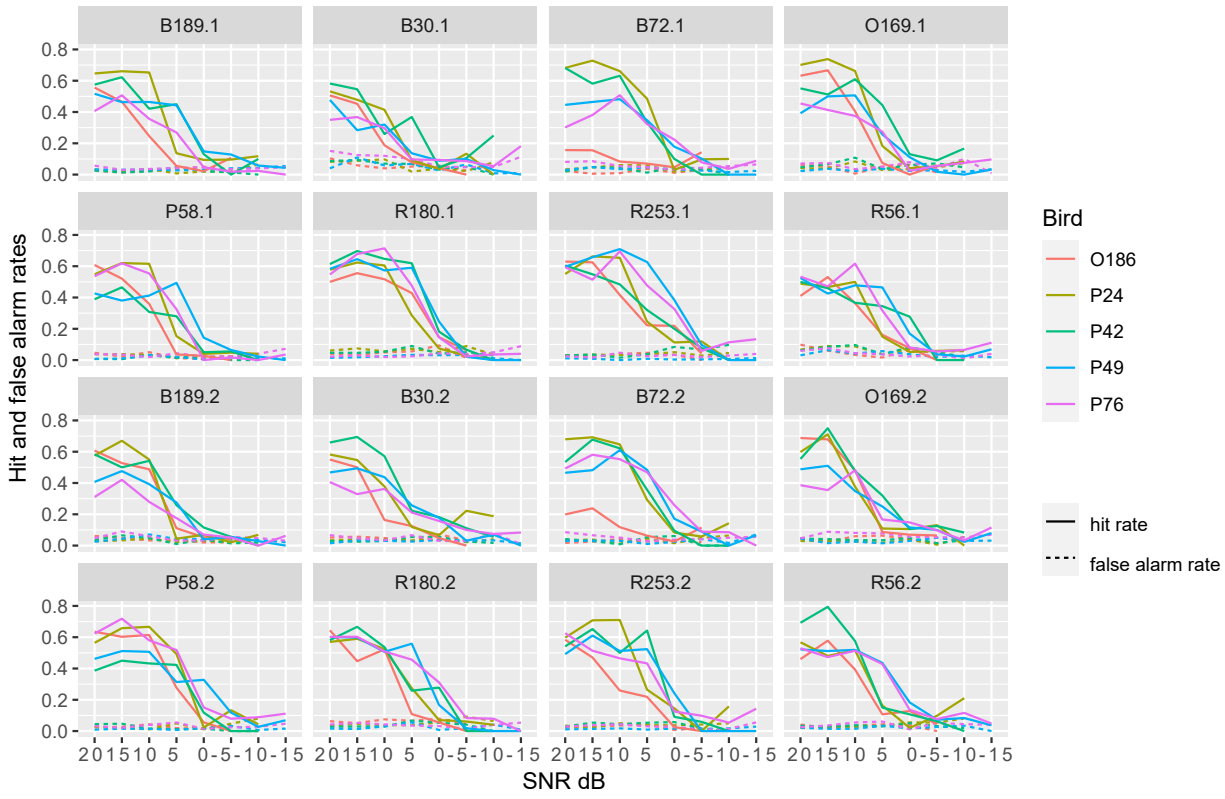
Supplementary Fig. 1. All the stimulus variants for a single motif.



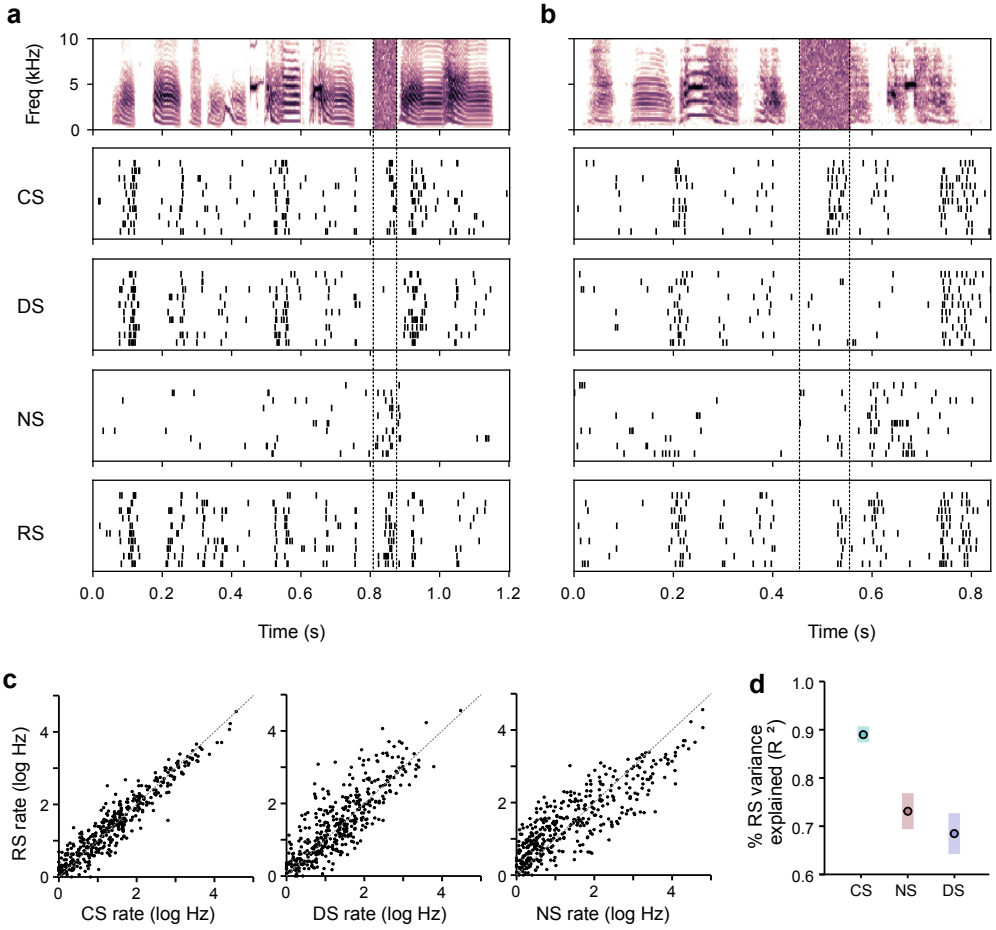
Supplementary Fig. 2. GLMM estimates of performance for all intervals. Top: An example motif sequence illustrating different response intervals. In this example, a peck during interval 4 would be a correct detection, and a false alarm in any other interval. Left: GLMM estimates and 90% confidence intervals for the auditory restoration task faceted by interval and familiarity (F-: unfamiliar; F+: familiar). The increase in performance in interval 5 likely indicates both a perceptual anchoring effect and a learned guessing strategy for the last motif in the sequence. Right: GLMM estimates and 90% confidence intervals for the masked task.



Supplementary Fig. 3. Within-subject variability. Hit and false alarm rates for each subject, with each trace corresponding to one motif variant and color indicating familiarity.



Supplementary Fig. 4. Between-subject variability. Hit and false alarm rates for each motif variant, with each trace corresponding to one subject.



Supplementary Fig. 5. Single-unit responses to occluded motifs. (**a**) Raster plot of an example single unit responding to the continuous (CS), discontinous (DS), replaced (RS) and noise (NS) variants of a motif. Top spectrogram shows the RS variant. Dotted lines indicate the critical interval. Note that the response to RS during the critical interval is more similar to CS than to DS or NS. (**b**) Raster plot of another example single unit, same format as **a**. (**c**) Average firing rates (log scale) during the critical interval of RS for all units compared to response rates to CS, DS, and NS during the same interval. Each point corresponds to an individual neuron ($n = 407$), and the dotted line indicates equality. CS response rates are highly predictive of RS response rates, while the rates for DS and NS show more scatter. (**d**) R^2 values (black dots) and 90% confidence intervals (colored bars) for linear regressions of RS firing rates against CS, NS, and DS rates.

Supplementary Table 1. Trial numbers per block during task-specific training.

Subject	3 Motifs	4 Motifs	5 Motifs	Ending LOR	Sex	Group
O186†	3268	3044	839	1.15 ± 0.19	M	B
P17	2703	8193	12340	1.24 ± 0.19	F	B
P24	2832	10195	13344	1.02 ± 0.26	M	B
P29	2807	13371	12588*		M	A
P30	1360	6104*			M	A
P35	2155	7671	3923*		M	B
P42	1876	6772	10203	1.20 ± 0.22	F	B
P49†	10375	2237	9327	1.20 ± 0.17	F	A
P52†	191	687‡			F	A
P76†	3731	2063	1729	1.06 ± 0.19	M	A
P8†	4980	3920*			F	A

*bird was excluded during this stage
†bird was previously trained on a similar same-different task
‡bird was excluded for low trial initiation

Movie 1. Well-trained zebra finch running trials to detect motifs with gaps in a sequence of otherwise identical motifs.