Motor control by precisely timed spike patterns

Kyle H. Srivastava[a,1], Caroline M. Holmes[b,c,1], Michiel Vellema[d], Andrea R. Pack[c,e], Coen P. H. Elemans[d], Ilya Nemenman[b,c], and Samuel J. Sober[c,2]

[a]Biomedical Engineering Doctoral Program, Georgia Institute of Technology and Emory University, Atlanta, GA 30322; [b]Department of Physics, Emory University, Atlanta, GA 30322; [c]Department of Biology, Emory University, Atlanta, GA 30322; [d]Department of Biology, University of Southern Denmark, DK-5230 Odense, Denmark; and [e]Neuroscience Doctoral Program, Emory University, Atlanta, GA 30322

Edited by William Bialek, Princeton University, Princeton, NJ, and approved December 5, 2016 (received for review July 17, 2016)

A fundamental problem in neuroscience is understanding how sequences of action potentials ("spikes") encode information about sensory signals and motor outputs. Although traditional theories assume that this information is conveyed by the total number of spikes fired within a specified time interval (spike rate), recent studies have shown that additional information is carried by the millisecond-scale timing patterns of action potentials (spike timing). However, it is unknown whether or how subtle differences in spike timing drive differences in perception or behavior, leaving it unclear whether the information in spike timing actually plays a role in brain function. By examining the activity of individual motor units (the muscle fibers innervated by a single motor neuron) and manipulating patterns of activation of these neurons, we provide both correlative and causal evidence that the nervous system uses millisecond-scale variations in the timing of spikes within multispike patterns to control a vertebrate behavior—namely, respiration in the Bengalese finch, a songbird. These findings suggest that a fundamental assumption of current theories of motor coding requires revision.

motor systems | neurophysiology | computational neuroscience | information theory | songbird

The brain uses sequences of spikes to encode sensory and motor signals. In principle, neurons can encode this information via their firing rates, the precise timing of their spikes, or both (1, 2). Although many studies have shown that spike timing contains information beyond that in the rate in sensory codes (3–5), these studies could not verify whether precise timing affects perception or behavior. In motor systems, rate coding approaches dominate (6, 7), but we recently showed that precise spike timing in motor cortex can predict upcoming behavior better than spike rates (8), showing that spike timing carries information in motor as well as sensory cortex. However, as in sensory systems, it remains unknown whether spike timing in motor systems actually controls variations in behavior (9, 10). Resolving this question, therefore, requires examining the spike code used by the neurons that innervate the muscles, because discovering an apparent spike timing code in any brain area upstream of motor neurons is subject to the same ambiguity about whether spike timing patterns actually affect behavior.

A spike timing-based theory of motor production predicts that millisecond-scale fluctuations in spike timing, holding other spike train features constant, will causally influence behavior. We tested this prediction by analyzing the activity of single motor units (that is, the muscle fibers innervated by a single motor neuron), focusing largely on the minimal patterns that have variable spike timing but fixed firing rate, burst onset, and burst duration: sequences of three spikes ("triplets"), where the third spike is a fixed latency after the first, but the timing of the middle spike varies. We examined timing codes in songbirds by focusing on respiration, which offers two key advantages. First, breathing is a relatively slow behavior (cycles last ∼400–1,000 ms), so the existence of timing codes is not a priori necessary; however, the precise control of breathing during singing (11, 12) suggests that timing may play a role. Second, we developed an electrode system which allowed us to collect spiking data over >50,000 breaths, yielding the large dataset sizes necessary to decipher the neural code.

We recorded electromyographic (EMG) signals from the expiratory muscle group (EXP) (13) using a flexible microelectrode array (Fig. 1*A*) to isolate spikes from single motor units. Precise timing codes might be implemented by individual spikes (1, 14, 15) or timing of spikes within a multispike pattern. Thus, we first verified whether single motor unit spike trains contain multispike features at high temporal resolution. Analysis of interspike intervals (ISIs) revealed that spiking was more regular than expected from a Poisson process (*SI Appendix*). Although not sufficient to establish the existence of a spike timing code, such regularity is crucial if the brain were to use spike timing patterns to control behavior, and hence output spikes in a controlled fashion. By showing that ISIs are more regular than expected in a Poisson process (in which the ISIs are independent), we establish the possibility of a timing code, and therefore the necessity of performing the more detailed analyses described below.

Next, we quantified the timescale on which the nervous system controls spikes within triplets by measuring the mutual information between consecutive ISIs in anesthetized birds. The nonzero value of information (Fig. 1*C*, red circle) suggests that consecutive ISIs (and hence, spike triplets) are controlled in the neural code. To understand the characteristic timescale of this control, we jittered the timing of each ISI by a Gaussian random number with SD σ, and again estimated the consecutive ISIs mutual information. We found that the information only approaches its unjittered values for $\sigma \sim 1$ ms (Fig. 1*C*, blue), showing that spike trains have millisecond-scale features. Similar findings were obtained in awake birds (*SI Appendix*, Fig. S1).

We then asked whether these millisecond-scale features predict behavior by simultaneously recording single motor units and air

Significance

A crucial problem in neuroscience is understanding how neural activity (sequences of action potentials or "spikes") controls muscles, and hence motor behaviors. Traditional theories of brain function assume that information from the nervous system to the muscles is conveyed by the total number of spikes fired within a particular time interval. Here, we combine physiological, behavioral, and computational techniques to show that, at least in one relatively simple behavior—respiration in songbirds—the precise timing of spikes, rather than just their number, plays a crucial role in predicting and causally controlling behavior. These findings suggest that basic assumptions about neural motor control require revision and may have significant implications for designing neural prosthetics and brain–machine interfaces.

Author contributions: K.H.S., C.M.H., C.P.H.E., I.N., and S.J.S. designed research; K.H.S., C.M.H., M.V., and A.R.P. performed research; K.H.S., C.M.H., M.V., A.R.P., C.P.H.E., I.N., and S.J.S. analyzed data; and K.H.S., C.M.H., C.P.H.E., I.N., and S.J.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper are available via https://figshare.com (DOI: 10.6084/m9.figshare.4546486).

[1]K.H.S. and C.M.H. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: samuel.j.sober@emory.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1611734114/-/DCSupplemental.

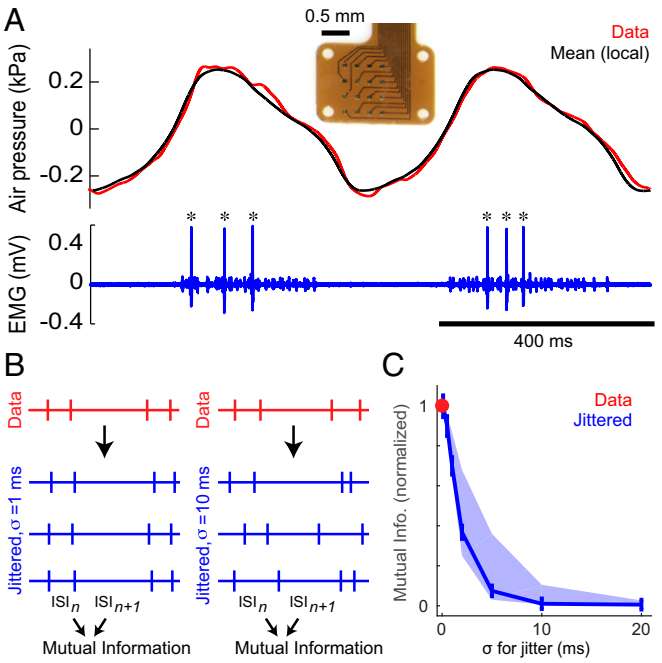


Fig. 1. The structure of the neural code. (*A*) Flexible microelectrode arrays (*Inset*) were used to record single motor units in the EXP. A sample recording shows (*Lower*) a well-isolated single motor unit (spikes marked by asterisks) and (*Upper*) the corresponding instantaneous and trial-averaged pressures (red and black, respectively). (*B*) To identify the temporal scale of precision of spike patterns, we jittered the ISIs and studied the mutual information between consecutive ISIs as a function of the jitter magnitude. (*C*) The mutual information in jittered spike trains approached that in the original recordings only for jitters on the scale of ~1 ms. The blue line shows data for bird EMG1, for which we had the most spikes (>350,000); the band shows the range across eight anesthetized birds (EMG1–EMG8). Unnormalized values of information at $\sigma = 0$ ms were 0.057–0.146 bits.

pressure within the respiratory system. Because respiration is controlled by ensembles of motor units, we did not expect a single motor neuron to drive the breathing cycle and instead expected it to only affect fluctuations around the mean. We therefore subtracted the mean respiratory pressure waveform from the recorded pressure (Fig. 1*A*) and investigated the relationship between such pressure residuals and the preceding spike train (Fig. 2*A*) using an estimator of mutual information (16) that we developed specifically for this purpose. This method separates the total mutual information between spikes and pressure residuals into contributions from spike count and spike timing:

$$I(spikes, pressure) = I(spike\ count, pressure) + I(spike\ timing, pressure).$$

Seven of eight birds tested (all but EMG3) had statistically significant information in spike timing, which was of the same order of magnitude as the information in spike rate (Fig. 2*A*), indicating that precise spike timing in motor units predicts the ensuing behavior.

We then verified directly that specific spike patterns predict behavior by selecting all pressure residuals that followed particular patterns and evaluating their means [pattern-triggered averages (PTAs)] and variances. Specifically, we focused on triplets preceded by ≥30 ms of silence, where the first and third spikes occurred 20 ms apart (at 2-ms accuracy), and that differed only by the timing of the middle spike (10 vs. 12 ms after the first spike or "10–10" and "12–8" triplets, respectively, with ISIs quantified at a temporal resolution of 2 ms) (green and blue tick marks in Fig. 2*B*, respectively). Such patterns had identical firing rates (three spikes in 20 ms) and burst onset/offset times and were among the most commonly observed patterns (e.g., N = 23,991 and 11,558 for the two patterns in bird EMG1, or 11 and 5% of all spike triplets of ≤ 20-ms duration, respectively).

We found that PTAs after the 10–10 and 12–8 spike triplets were significantly different (Fig. 2*B*). We quantified discriminability of the PTAs using the d' statistic (17) and found $d' = 0.108 \pm 0.011$ (SD) 17 ms after triplet onset (Fig. 2*C*). The same effect is present across all six birds (Fig. 2*D*). Notably, although d' traces are similar across animals, the PTAs themselves are not (*SI Appendix*, Fig. S2). Therefore, although the discriminability of different patterns is consistent (Fig. 2*D*), the encoding of pressure differs across individuals. Wavelet-based functional ANOVA (wfANOVA) (18) (*SI Appendix*) revealed a consistent significant effect between the PTAs across birds after accounting for intersubject variability (*SI Appendix*, Fig. S3 and Table S1). Therefore, millisecond-scale changes in timing of a single spike in a multispike pattern at a fixed firing rate predict significant changes in air sac pressure. This result agrees with our previous findings that cortical neurons upstream of vocal and respiratory muscles also use spike timing to encode behavior (8).

Although the above results show that precise spike timing predicts behavioral variations, they cannot reveal whether timing affects muscle output. To test this hypothesis, we extracted muscle fiber bundles from EXP and measured force production

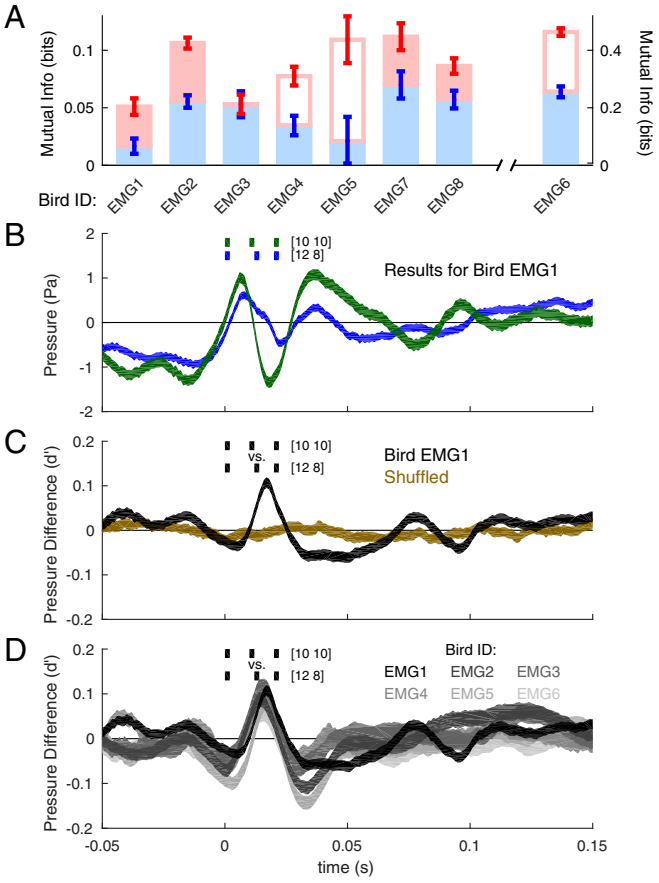


Fig. 2. Spike timing predicts respiratory air pressure. (*A*) Mutual information (±1 SD) between 20 ms of spike timing (red)/spike rate (blue) and 100 ms of pressure residuals. Empty bars indicate underestimated values. (*B*) Pressure residuals (the bands represent mean ±1 SEM; bootstrapped) (*Materials and Methods*) differed significantly for 10–10- (green) and 12–8-ms (blue) spike triplets. Negative residuals before pattern onset likely reflect activity of other correlated, nonrecorded motor units. (*C*) Discriminability (d'; mean ±1 SD; bootstrapped) between pressure residuals for the triplets shown in *B* compared with the reshuffled control. (*D*) The d' values of 10–10 and 12–8 triplets in six birds (plotting conventions are same as in *C*).

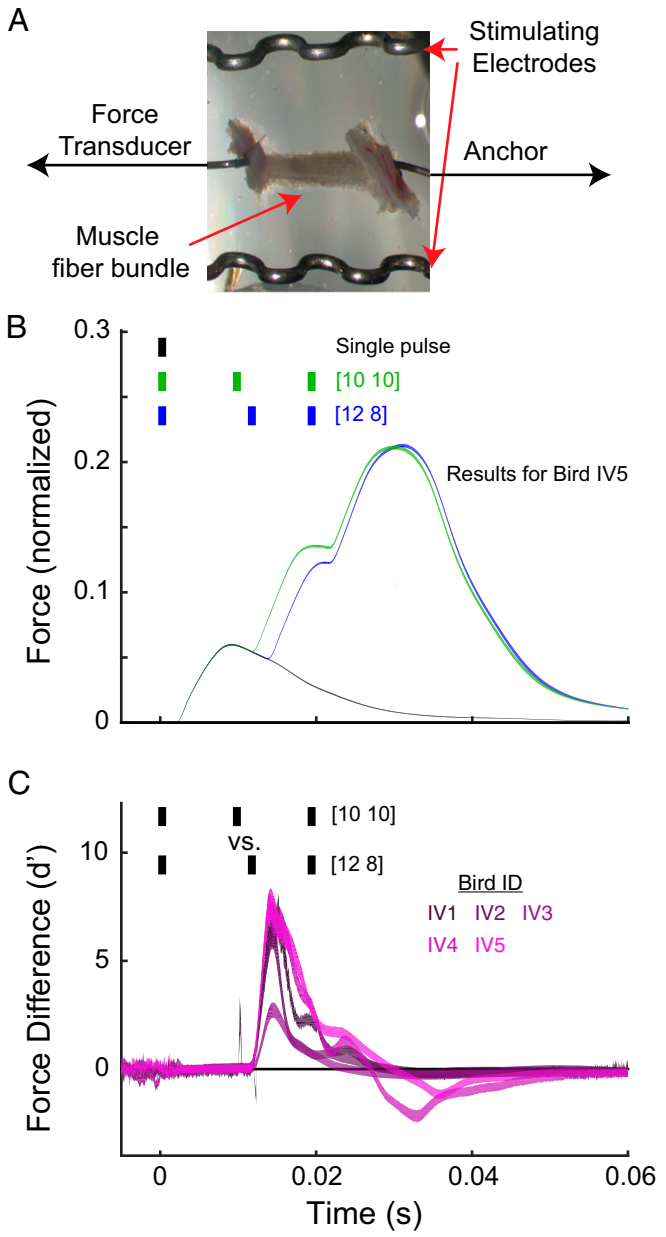


Fig. 3. Millisecond-scale stimulation patterns affect force production. (*A*) In vitro muscle fiber preparation (*Materials and Methods*). (*B*) Force output (full width of the line indicates mean ± SEM; normalized to peak force during tetanic contraction) differs significantly between 10–10- and 12–8-ms IPI stimulation triplets (wfANOVA) (*SI Appendix*, Fig. S4*A* and Table S1). (*C*) Discriminability (d'; full width of the line indicates mean ± SD; bootstrapped) of force profiles in five birds (birds IV1–IV5) after stimulation with the same stimuli as in *B*.

in vitro while stimulating using three-pulse patterns with 10–10- and 12–8-ms interpulse intervals (IPIs). Changing the timing of the middle pulse by 2 ms significantly altered force output (Fig. 3*B*); wfANOVA identified significant differences in the force evoked by these patterns (*SI Appendix*, Fig. S4 and Table S1). Furthermore, d' values evoked by 10–10 and 12–8 triplet stimulations are several SDs from zero (Fig. 3*C*). The same effect holds for other pairs of similar triplets, such as 2–18- and 4–16-ms IPIs (*SI Appendix*, Fig. S5 *A* and *B*). Therefore, our in vitro experiments establish that the small, precisely regulated differences in motor neuron spike patterns observed in vivo cause muscles to produce different forces.

We next explored whether different spike patterns not only correlate with behavior and drive distinct force production but also cause different behaviors in vivo. We recorded air sac pressure while simultaneously applying temporally patterned electrical stimulation to EXP, again using 10–10- and 12–8-ms stimulation triplets (Fig. 4*A*). Moving the middle pulse from 10 to 12 ms after the first evoked distinct pressure waveforms (Fig. 4*B*) consistently across all six birds tested (Fig. 4*C*). wfANOVA identified significant differences between the effects of these triplets (*SI Appendix*, Fig. S6*A* and Table S1). Finally, we comprehensively investigated the effects of moving the middle pulse from 2 to 18 ms after the first pulse (in steps of 2 ms), which resulted in significant differences in the mean pressure (Fig. 4*D*) ($P < 0.001$ for all 36 pairs of these stimulation patterns). These experiments thus show a causal link between millisecond-scale timing of muscle activation and the ensuing behavior.

Overall, we have shown that respiratory motor unit activity is controlled on millisecond timescales, that precise timing of spikes in multispike patterns is correlated with behavior (air sac pressure), and that muscle force output and the behavior itself are causally affected by spike timing (all on similar temporal scales) (Figs. 2*D*, 3*C*, and 4*C*). These findings provide crucial evidence that precise spike timing codes casually modulate vertebrate behavior. Additionally, they shift the focus from coding by individual spikes (1, 14, 19) to coding by multispike patterns and from using spike timing to represent time during a behavioral sequence (20, 21) to coding its

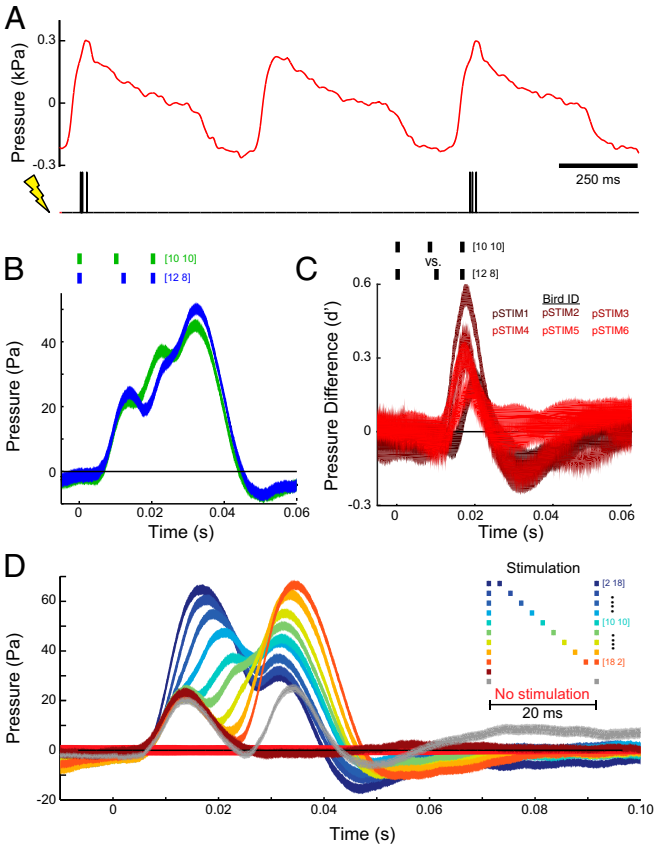


Fig. 4. Millisecond-scale differences in stimulation patterns modulate air pressure in vivo. (*A*) Three-pulse stimulation was delivered during respiration. (*B*) The 10–10- and 12–8-ms triplets (as in Figs. 2 and 3) caused significant differences in air sac pressure residuals (mean ± SEM) (*SI Appendix*, Fig. S6*A* and Table S1). (*C*) Discriminability (d'; mean ± SD; bootstrapped) of pressure residuals after 10–10- and 12–8-ms triplets across birds pSTIM1–pSTIM6. (*D*) Responses to all stimulation patterns tested (mean ± SEM). In *B*–*D*, time is relative to the onset of the stimulation pattern. Note that the reddish brown trace is the response to only one stimulus pulse and that the gray trace is the response to two pulses separated by 20 ms.

structural features. Put another way, although it is clear that earlier activation of neurons would lead to earlier activation of muscles, this relationship only accounts for encoding when a behavior happens (10, 22). Here, we show that changing the timing of a single spike within a burst by ∼1 ms can also affect what the animal will do, not just when it will do it. Furthermore, we showed that the effect of moving a single spike is stable across animals (Fig. 2). We believe that this precise spike timing code reflects and exploits muscle nonlinearities: spikes less than ∼20 ms apart generate force supralinearly (*SI Appendix*, Fig. S12), with stronger nonlinearities for shorter ISIs. Thus, changing the first ISI from 12 to 10 ms significantly alters the effect of the spike pattern on air pressure (Fig. 2*B*). Such nonlinearities in force production as a function of spike timing have been observed in a number of species (23–25), highlighting the necessity of examining the role of spike timing codes in the motor systems of other animals. Importantly, our findings show that the nervous system uses millisecond-timescale changes in spike timing to control behavior by exploiting these muscle nonlinearities, even though the muscles develop force on a significantly longer timescale (tens of milliseconds as shown in Fig. 3*B*).

The surprising power of spike timing to predict behavior might reflect synchrony between motor units in the respiratory muscles (26), so that timing variations in one motor unit co-occur with timing variations in others. Resolving this question requires examining temporal population codes in motor systems, a subject not yet explored. Furthermore, respiration is driven by a brainstem central pattern generator but modified by descending inputs from the forebrain (12, 27). It remains unknown which of these is the source of timing precision/variability. Because respiration is critical to vocalization in songbirds, it will be of special interest to record respiratory timing patterns during singing and determine how the temporal code used by neurons in upstream area RA (the robust nucleus of the arcopallium, which sends inputs to respiratory networks) is transduced to the motor periphery (8). Furthermore, our findings suggest that spike timing may contribute to motor control in other species and systems. Extracting additional information from spike timing (for example, in brain–machine interfaces, which typically rely on firing rates to drive action) may thus help us better decode motor activity.

Materials and Methods

Surgical Procedure. We used EMG and electrical stimulation to determine the importance of motor timing in the EXP for avian respiration. All procedures were approved by the Emory University Institutional Animal Care and Use Committee. Before surgery, adult male Bengalese finches (>90 d old) were anesthetized using 40 mg/kg ketamine and 3 mg/kg midazolam injected i.m. Proper levels of anesthesia were maintained using 0–3% (vol/vol) isoflurane in oxygen gas.

Subjects. Our studies used a total of 24 adult (>90 d old) Bengalese finches. Eight birds (which we refer to as birds EMG1–EMG8) underwent EMG recordings of single motor units, in which one single unit was isolated in each animal (that is, birds EMG1–EMG8 each contributed a single unit to the study, and data from all eight are used in the analyses in Fig. 1*C*). Concurrent with EMG recordings, the pressure within the air sac was continuously monitored (see below for detailed descriptions of procedures). Of these eight birds, six (EMG1–EMG6) yielded sufficient neural and pressure data, such that we were able to compute pressure waveforms conditional on the occurrence of a particular spike pattern (these subjects are the birds shown in Fig. 2 *A* and *D*). Furthermore, we recorded EMG spike trains (but not air pressure, because we were unable to find a pressure meter that was both sensitive enough to detect small respiratory pressure changes and lightweight enough for an awake bird to carry) from four awake birds, which we refer to as birds EMG9–EMG12, and they provided the data shown in *SI Appendix*, Fig. S1.

Additionally, muscle fiber bundles (one bundle per bird) were extracted from each of five birds for in vitro testing of force production. We refer to these birds (data are shown in Fig. 3) as birds IV1–IV5. We also examined the effect of electrically stimulating the expiratory muscles in vivo in six anesthetized birds, which we refer to as birds pressure stimulation 1 (pSTIM1) to pSTIM6. Results from these subjects are shown in Fig. 4. Lastly, four birds (C1–C4) were used to examine the effect of curare on both EMG activity and the efficacy of muscle stimulation. The results are shown in *SI Appendix*, Fig. S7. All birds were male, except for three of the birds used in the in vitro experiments.

Finally, we note that three birds were used in multiple experiments. Specifically, two birds were used in both EMG unit recordings and patterned electrical stimulation, and one bird was used in both EMG recordings and (subsequently) a curare experiment. To (possibly) reduce the complexity of our numbering scheme, the first of these subjects is referred to as both EMG7 and pSTIM1 (i.e., a single bird has two names), the second is both EMG4 and pSTIM3, and the third is both EMG8 and C1.

EMG Recordings. To optimize our ability to isolate individual motor units, we developed microscale, flexible, high-density electrode arrays that sit on the surface of individual muscles to record EMG signals (Fig. 1). The gold electrodes were fabricated on 20-μm-thick photo-definable polyimide with a range of contact sizes and spacing (Premitec). The electrode exposures ranged from 25 to 300 μm in diameter and were separated by as little as 25 μm. Several alignments of electrodes (16 per array) were fabricated, including a 4 × 4 grid and 4 tetrodes. To record from the EXP, an incision was made dorsal to the leg attachment and rostral to the pubic bone. After spreading fascia on the muscle group, an electrode array was placed on its surface. The other end of the array connects to a custom-designed Omnetics adapter to interface with a digital amplifier (RHD2132; Intan Technologies). The Intan evaluation board delivered the EMG signals to the computer at 30 kHz.

With these arrays, we were able to acquire high-quality EMG recordings from 16 locations simultaneously during quiet respiration in eight male Bengalese finches. The increased number of channels allows the experimenter to decide which channels should be subtracted from each other to create bipolar signals. Because of the high specificity and impedance of individual electrodes, we were able to extract single-motor unit data in some cases (*Materials and Methods*, *Data Analysis* and Fig. 1). Because we can record 16 unipolar signals in a very small area, we have increased the probability of recording a single motor unit while allowing us to test how different i.m. segments are differentially recruited. Although EXP is made up of three sheet-like overlapping muscles [musculus (m.) obliquus externus abdominis, m. obliquus internus, and m. transversus abdominis], we presume that we are recording motor units from the most superficial muscle: m. obliquus externus abdominis. However, because all three muscles have similar functional roles involving contraction during respiration (28), recording a motor unit from any of these muscles would not affect our interpretation.

Pressure Recording. Thoracic air sac pressure was monitored using a Silastic Tube (Dow Corning) inserted in the same manner as previous studies (11, 29, 30) with the pressure sensor 20INCH-D-4V (All Sensors). Briefly, a small incision was made inferior to the rib cage. A 6-cm Silastic tube (0.03-in i.d. and 0.065-in o.d.) with a beveled end was inserted through the incision and sutured to the rib cage. The other end of the tube was then connected to the sensor. For recording experiments, the Intan evaluation board delivered the pressure signal to the computer at 30 kHz. For the in vivo muscle stimulation experiments (below), pressure recordings were acquired using an NIDAq Board (BNC-2090A; National Instruments) at 32 kHz. In both cases, the Intan evaluation board provided the voltage supply for the pressure sensor.

In Vitro Muscle Stimulation. The in vitro muscle preparation was conducted using five Bengalese finches (two male and three female; none overlapping with the birds used for in vivo recordings and stimulation), such as we have done previously (31). We briefly describe the technique here. Animals were euthanized with an overdose of isoflurane (Baxter), and EXP was exposed as in the in vivo experiments. Fiber bundles were then isolated from the surface of m. obliquus externus abdominis (the most superficial muscle in the EXP group). The fiber bundles were then mounted in a test chamber while continuously being flushed with oxygenated Ringers solution at 39 °C. One end of the muscle was fixed to a servomotor (although it was not used) using silk suture, whereas the other end was mounted on a force transducer (Model 400A; Aurora Scientific). The muscle fibers were then stimulated through the solution using parallel platinum electrodes (Model 701C; Aurora Scientific). For each muscle preparation, both stimulation current and preparation length were optimized for maximum force generation. A single 300-μs pulse was used for stimulus optimization followed by a 200-Hz, 100-ms tetanic stimulation for length optimization. To test the importance of motor timing, we stimulated the muscle with three 300-μs pulses at optimal current, with the first and third pulses separated by 20 ms and the middle pulse being placed 2, 4, 10, or 12 ms after the first pulse. Additional trials were conducted with only a single pulse as a control. These five stimulation patterns were repeated in random order with 60 s between each trial and after five such iterations, followed by a 200-Hz, 100-ms tetanic stimulation. After this procedure, we measured a total of 25 iterations, taking ∼135 min. To account for muscle fibers dying over the course of the experiment, force measurements were normalized to the fiber bundles maximum tetanic force at 200 Hz and linearly interpolated for each stimulus. Force transducer and stimulation signals were digitized at 20 kHz with an NIDAq Board (PCI-MIO-16E4; National Instruments).

In Vivo Muscle Stimulation. Stimulation of EXP was performed in six male Bengalese finches using two fine-wire electrodes made of insulated multistranded alloy (50-μm diameter, Phoenix Wire Inc.). Pressure recordings were used to trigger stimulation with custom-written LabVIEW code when the pressure crossed a user-defined threshold. Generally, the stimulation pulse train was targeted for 100 ms after the pressure crossed zero. In birds in which the initial upswing of the respiratory cycle was slower than normal, stimulation was delayed up to 50 ms longer to prevent stimulation from occurring during the upswing itself. We wanted to avoid stimulating during that interval because it was more difficult to extract the pressure effects of stimulation. Stimulating between 100 and 150 ms after the zero crossing also mirrored the timing of spikes found in EMG recordings. The LabVIEW code then sent a stimulation pattern to an external stimulator (Model 2100; A-M Systems), which was connected to the fine-wire electrodes in EXP.

To test the importance of motor timing on behavior, three stimulation pulses (biphasic, 250-μs pulse duration, and 250-μA current) were delivered, with the first and third separated by 20 ms and the middle pulse placed in 2-ms increments across the duration (nine different patterns) in addition to a single pulse and no pulse control stimuli. All 11 patterns were randomly interleaved during the experiment. Pressure and trigger times were recorded at 32 kHz using the LabVIEW code.

The selection of an appropriate current was important for interpreting the results of these experiments. To properly compare them with EMG recordings, we wanted to stimulate using a current that activates the axons of motor neurons but does not activate muscle fibers directly. One previous study that stimulated songbird muscles used currents as great as 2 mA (31), whereas a more recent one posited that currents below 500 μA were likely activating nerve fibers (32). We therefore selected a current of 250 μA for EXP stimulation for the figures in the text to ensure robust effects on air sac pressure. To test that we were only stimulating the axons of motor neurons, we applied curare, which locally blocks the neuromuscular junction, to EXP and compared both EMG and stimulation effects with those when saline was applied to the muscle. Curare eliminated the recorded EMG signal observed when only saline was applied to the muscle (*SI Appendix*, Fig. S7 *A* and *B*). Stimulation at currents as low as 100 μA produced clear effects on air sac pressure when saline was applied to the muscle (*SI Appendix*, Fig. S7*C*), but those effects were abolished when curare was applied to the muscle (*SI Appendix*, Fig. S7*D*). The same strong stimulation effects observed at 250 μA when saline was applied (*SI Appendix*, Fig. S7*E*) were greatly reduced when curare was introduced to the muscle (*SI Appendix*, Fig. S7*F*). Although we were unable to completely eliminate stimulation effects at 250 μA using curare, we believe that the result was caused by the current spreading farther than the spatial span of the drug. Applying too much curare and fully paralyzing EXP would endanger the wellbeing of the animal. However, conducting our full three-pulse stimulation experiment at 100 μA produced quite similar air sac pressure effects (*SI Appendix*, Fig. S7*G*) as those observed using 250 μA (Fig. 4*D*). Therefore, we believe that our muscle stimulation experiments were only activating the axons of motor neurons and were not activating muscle fibers directly. This finding allowed us to make insightful comparisons between the results of our spike pattern and stimulation analyses.

Data Analysis. All pressure recordings were converted from voltages to kilopascals via calibration measurements taken with a manometer, and then band-pass filtered between 1 and 50 Hz. EMG recordings were band-pass filtered between 300 and 7,500 Hz. Pairs of EMG channels were subtracted to optimized motor unit isolation. After a good pair was selected, motor unit spikes were sorted using custom-written MATLAB (Mathworks) code (7).

To analyze the EMG and pressure recordings together (Fig. 2), we searched through the pressure for occurrences of particular spiking patterns and compared the pressures after (triggered by) those patterns. We assumed that the individual spikes of our single motor unit did not drive the overall pressure cycle. Therefore, instead of comparing raw pressure measurements, we subtracted out mean cycles and analyzed statistical dependences between spiking and such pressure residuals. To define the mean waveform, we segmented individual breathing cycles (defined by a stereotypical rise and fall in pressure; segmentation by maxima, minima, or zero crossings gave similar results). The cycles had different durations, and we renormalized time within each cycle to the breathing phase $(0,2\pi)$. Because the structure of breathing changes over the hours of recordings, we averaged pressures at the same phase in a 21-cycle sliding window (other windows were tried with no significant changes), resulting in a time-dependent mean pressure waveform. The mean was then subtracted from the phase-rescaled local pressure to produce a residual in each individual cycle. We then compared these residuals triggered by the chosen spiking patterns using the usual d' discriminability metric:

$$d'(t) = \frac{\bar{x}(t) - \bar{y}(t)}{\sqrt{\frac{1}{2}\left(\sigma^2_{x(t)} + \sigma^2_{y(t)}\right)}},$$

where $\bar{x}(t)$ and $\bar{y}(t)$ are the sample means of the pressures triggered by the two spike patterns at time t after the onset of the patterns, and $\sigma_{x(t)}$ and $\sigma_{y(t)}$ are the corresponding SDs. To calculate the error bars on d', we bootstrapped the entire analysis pipeline (33) by resampling with replacement the pressure traces residuals after the analyzed spike patterns 500 times and estimating the SD of the set of d' resulting from the bootstrapping. In addition to comparing patterns that are the same to 2 ms, we also were able to compare the same patterns for a single-millisecond shift in a single spike (*SI Appendix*, Fig. S8), and significant differences were seen in some birds.

To isolate the effects of EXP stimulation on air sac pressure, the mean pressure waveform of 20 previous unstimulated (catch) respiratory periods was similarly subtracted from the stimulated pressure waveform. A trailing window was used to eliminate the possibility of future effects of stimulation affecting our mean subtraction. For this pressure residual calculation, waveforms were aligned to the phase of the respiratory pattern at which the stimulation, instead of the spike pattern, occurred. All catch-subtracted pressure waveforms were averaged within a given stimulation pattern, with the SE calculated at every time point. To compare responses from two different stimuli, d' and its estimated error were calculated as above.

In vitro force measurements were compared following different stimulation patterns using the same d' analysis as both for the recording and the in vitro stimulation analyses above. Because of difficulty obtaining Bengalese finches in Denmark (where our in vitro studies took place), we used two male and three female subjects for analysis. Despite other experiments only being conducted on male Bengalese finches, no qualitative differences were observed between sexes aside from normal intersubject variability (*SI Appendix*, Fig. S9). Because the sample size was small for each sex, we could not perform a statistical comparison between the two groups.

Mutual Information: Consecutive ISIs. To estimate the scale of temporal structures in the neural code, we evaluate the mutual information between subsequent ISIs that are ≤ 30-ms long (and hence fall into the same breathing cycle) and also between these ISIs corrupted by a Gaussian noise with various SDs (Fig. 1*C*). Mutual information between two continuous variables x and y is defined as (34)

$$I(x,y) = \int dx\, dy\, P(x,y) \log_2 \frac{P(x,y)}{P(x)P(y)}, \quad [1]$$

and a sum replaces the integral for discrete variables. Mutual information is a measure of statistical dependency that does not assume normality of the underlying distribution in contrast to the more familiar correlation coefficient, and it measures all statistical dependences between the two variables, such that it is zero if and only if the two variables are completely statistically independent. Because ISIs are non-Gaussian distributed, using mutual information is more appropriate than simpler dependency measures. Mutual information is measured in bits. Measurement of x provides one bit of information about y (and vice versa) if the measurement of x allows us to answer one binary (yes/no) question about y.

Estimation of mutual information from empirical data is a complex problem (16, 35). To solve this problem for mutual information between two real-valued consecutive ISIs, we use the k-nearest neighbors estimator (36). The method detects structures in the underlying probability distribution by estimating distances to the k-nearest neighbors of each data point. By varying k, one explores structures in the underlying data on different scales. We choose which k to use by calculating the mutual information for varying amounts of data (using different size subsets of the full data) and detecting the (absence of) the sample size-dependent bias (5, 8, 37, 38). The joint distribution of consecutive ISIs is smooth, and hence a broad range of k near $k = 10$ produces unbiased information estimates. To identify possible sample size-dependent biases and calculate the error bars, we divided the dataset of N samples into nonoverlapping subsamples of size N/m, with the inverse data fraction $m = 2, \ldots, 10$. We calculated mutual information in each subsample and then evaluated the SD of the estimates for a given m, where 10 independent partitions were done for each m (*SI Appendix*, Fig. S10*A*). We then fitted these empirical variances to the usual 1/(sample size) law by performing a linear regression $\log \sigma^2(m) = A + \log m$. We then estimated the variance for the full dataset by setting $m = 1$ (*SI Appendix*, Fig. S10*B*). The same analysis was performed for both the original dataset and the jittered datasets. To the extent that mutual information estimates for different sample sizes (different values of m) agree with each other within the error bars, the estimate of the mutual information likely does not have a sample size-dependent bias (*SI Appendix*, Fig. S10*A*).

Mutual Information: Spikes and Pressure. We calculated the mutual information between 20-ms-long spike trains and 100-ms-long pressure residuals (Fig. 2*B*). These timescales were chosen because spikes have to be closer to each other than about 20 ms to cause supralinear effects in muscle activation (*SI Appendix*, Fig. S12) and because effects of spikes on pressure appear to last less than 100 ms (Fig. 2*C*). We focused on spike trains that began at the phase $\varphi \approx 0.8\pi$ in the breathing cycle (*Materials and Methods*, *Data Analysis*), because the most spiking occurs near that phase (nearby choices reveal similar results). For the pressure patterns, we started 10 ms after 0.8π phase point, because it takes about 10 ms for spiking to affect behavior (Fig. 4*D*). Importantly, although the starting point of patterns/pressures was chosen based on the phase, the time within the patterns and the pressure traces were not rescaled. The autocorrelation time of pressure residuals is about 10 ms for all EMG birds, and therefore, we chose to describe the 100-ms pressure residuals using $p = 11$ real-valued data points spread 10 ms apart (*SI Appendix*, Fig. S11*A*).

With this choice, we needed to calculate the mutual information between an 11D pressure vector (y in Eq. **1**) and the spiking vector (x in Eq. **1**). For this calculation, we modified the k-nearest neighbor mutual information estimator (36) in the following way. We rewrote the mutual information between the spikes and the pressure as

$$I(x,y) = I(n,y) + \sum_n P(n)\ I(x,y|n), \qquad [2]$$

where n is the number of spikes in the 20-ms spike train. The first term on the right-hand side of Eq. **2** is the information between the firing rate alone and the pressure, and the second is the information between the timing alone and the pressure. This partitioning allowed us to automatically estimate the relative contribution of each of these terms.

We estimated each of the information quantities $I(n,y)$ and $I(x,y|n)$ using the k-nearest neighbors estimator [recall that $I(x,y|0) = 0$], where x is now an n-dimensional vector of spike timings for a fixed spike count n (we did not discretize time for this analysis). Because mutual information is reparameterization-invariant, we rescaled the number of spikes to have zero mean and unit variance, and then additionally reparameterized the spike times and the pressure values to have normal marginal distributions, so that the i th value of the variable in a set of N samples was mapped into the value that corresponds to the cumulative distribution of a unit variance normal being equal to $(i-1/2)/N$. Having unit variances ensured that every variable contributes similarly to determining nearest neighbors of data points. Furthermore, making variables to have exactly normal marginal distributions decreased the influence of outliers (we have verified that this reparameterization has a negligible effect on data with no outliers).

As in *Materials and Methods*, *Mutual Information: Consecutive ISIs*, we found the value of k that produced no sample size-dependent drift in the estimate of $I(x,y)$; for EMG1, this value was $k = 3$ (*SI Appendix*, Fig. S11 *A* and *B*). We similarly estimated error bars by subsampling the data, estimating the variance of each subsample, and extrapolating to the full dataset size. For some birds, no value of k produced estimates with zero sample size-dependent bias within error bars. In these cases, we chose the k that resulted in the smallest sample size-dependent drift. In all such cases (bird identifications EMG3–EMG6) (indicated by empty boxes in Fig. 2*A*), the drift was upward as the sample size increased, indicating that the value of the mutual information calculated at the full sample was an underestimate. Underestimating the mutual information makes it harder to show that spike timing contains information about the behavior, and yet, seven of eight birds in Fig. 2*A* showed statistically significant information in spike timing.

We additionally verified that $p = 11$ points for characterization of the pressure is the right choice by estimating $I(x,y)$ at different values of p. As p increases from one, more features in the pressure get sampled, uncovering more information. Information plateaus near $p \sim 10$, indicating that all relevant features in the data have been recovered, and in some cases, it finally leaves the plateau at larger p because of undersampling, indicating that such fine discretization should be avoided.

Finally, we note that n in $I(n,y)$ is a discrete variable, making the use of the k-nearest neighbors estimator problematic. We address the issue by injecting each discrete datum with small Gaussian random noise. All data shown used the noise with SD of 10^{-4}, but other values in the range $10^{-8} \ldots 10^{-2}$ were tried with no discernable differences.

ACKNOWLEDGMENTS. This work was supported by NIH Grants P30NS069250, R01NS084844, F31DC013753, and 5R90DA033462; National Science Foundation Grants 1208126 and 1456912; James S. McDonnell Foundation Grant 220020321; the Danish Research Council and Carlsberg Foundation; and the Woodruff Scholarship at Emory University.

1. Rieke F (1999) *Spikes: Exploring the Neural Code* (MIT Press, Cambridge, MA).
2. Theunissen F, Miller JP (1995) Temporal encoding in nervous systems: A rigorous definition. *J Comput Neurosci* 2(2):149–162.
3. Fairhall A, Shea-Brown E, Barreiro A (2012) Information theoretic approaches to understanding circuit function. *Curr Opin Neurobiol* 22(4):653–659.
4. Reinagel P, Reid RC (2000) Temporal coding of visual information in the thalamus. *J Neurosci* 20(14):5392–5400.
5. Strong SP, Koberle R, van Steveninck RRR, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.
6. Churchland MM, et al. (2012) Neural population dynamics during reaching. *Nature* 487(7405):51–56.
7. Sober SJ, Wohlgemuth MJ, Brainard MS (2008) Central contributions to acoustic variation in birdsong. *J Neurosci* 28(41):10370–10379.
8. Tang C, Chehayeb D, Srivastava K, Nemenman I, Sober SJ (2014) Millisecond-scale motor encoding in a cortical vocal area. *PLoS Biol* 12(12):e1002018.
9. DiLorenzo PM, Victor JD, eds (2013) *Spike Timing: Mechanisms and Function*, Frontiers in Neuroscience (CRC/Taylor & Francis Group, Boca Raton, FL).
10. Stevens CF, Zador A (1995) Neural coding: The enigma of the brain. *Curr Biol* 5(12): 1370–1371.
11. Wild JM, Goller F, Suthers RA (1998) Inspiratory muscle activity during bird song. *J Neurobiol* 36(3):441–453.
12. Suthers RA, Goller F, Wild JM (2002) Somatosensory feedback modulates the respiratory motor program of crystallized birdsong. *Proc Natl Acad Sci USA* 99(8):5680–5685.
13. Fedde MR, DeWet PD, Kitchell RL (1969) Motor unit recruitment pattern and tonic activity in respiratory muscles of Gallus domesticus. *J Neurophysiol* 32(6):995–1004.
14. Bialek W, Rieke F, de Ruyter van Steveninck RR, Warland D (1991) Reading a neural code. *Science* 252(5014):1854–1857.
15. Thorpe SJ (1990) Spike arrival times: A highly efficient coding scheme for neural networks. *Parallel Processing in Neural Systems and Computers*, eds Eckmiller R, Hartmann G, Hauske G (North Holland, Amsterdam), pp 91–94.
16. Paninski L (2003) Estimation of entropy and mutual information. *Neural Comput* 15: 1191–1253.
17. Wickens TD (2001) *Elementary Signal Detection Theory* (Oxford Univ Press, New York).
18. McKay JL, Welch TDJ, Vidakovic B, Ting LH (2013) Statistically significant contrasts between EMG waveforms revealed using wavelet-based functional ANOVA. *J Neurophysiol* 109(2):591–602.
19. Fairhall A (2014) The receptive field is dead. Long live the receptive field? *Curr Opin Neurobiol* 25:ix–xii.
20. Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419(6902):65–70.
21. Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. *J Neurosci* 25(3):652–661.
22. Borst A, Theunissen FE (1999) Information theory and neural coding. *Nat Neurosci* 2(11):947–957.
23. Drzymała-Celichowska H, Kaczmarek P, Krutki P, Celichowski J (2016) Summation of slow motor unit forces at constant and variable interpulse intervals in rat soleus muscle. *J Electromyogr Kinesiol* 30:1–8.
24. Moritz CT, Barry BK, Pascoe MA, Enoka RM (2005) Discharge rate variability influences the variation in force fluctuations across the working range of a hand muscle. *J Neurophysiol* 93(5):2449–2459.
25. Perreault EJ, Day SJ, Hulliger M, Heckman CJ, Sandercock TG (2003) Summation of forces from multiple motor units in the cat soleus muscle. *J Neurophysiol* 89(2):738–744.
26. Chagnaud BP, Zee MC, Baker R, Bass AH (2012) Innovations in motoneuron synchrony drive rapid temporal modulations in vertebrate acoustic signaling. *J Neurophysiol* 107(12):3528–3542.
27. Gracco VL, Abbs JH (1988) Central patterning of speech movements. *Exp Brain Res* 71(3):515–526.
28. Fedde MR (1987) Respiratory muscles. *Bird Respiration*, ed Seller TJ (CRC Press, Boca Raton, FL), Vol I, pp 3–37.
29. Goller F, Larsen ON (1997) A new mechanism of sound generation in songbirds. *Proc Natl Acad Sci USA* 94(26):14787–14791.
30. Ashmore RC, Wild JM, Schmidt MF (2005) Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J Neurosci* 25(37):8543–8554.
31. Elemans CP, Mead AF, Rome LC, Goller F (2008) Superfast vocal muscles control song production in songbirds. *PLoS One* 3(7):e2581.
32. Srivastava KH, Elemans CP, Sober SJ (2015) Multifunctional and context-dependent control of vocal acoustics by individual muscles. *J Neurosci* 35(42):14183–14194.
33. Efron B, Tibshirani RJ (1994) *An Introduction to the Bootstrap* (CRC, Boca Raton, FL).
34. Shannon CE, Weaver W (2015) *The Mathematical Theory of Communication* (Univ of Illinois Press, Urbana, IL).
35. Nemenman I, Shafee F, Bialek W (2002) Entropy and inference, revisited. *Advances in Neural Information Processing Systems 14*, eds Dietterich TG, Becker S, Ghahramani Z (MIT Press, Cambridge, MA), pp 471–478.
36. Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* 69(6 Pt 2):066138.
37. Nemenman I, Lewen GD, Bialek W, de Ruyter van Steveninck RR (2008) Neural coding of natural stimuli: Information at sub-millisecond resolution. *PLOS Comput Biol* 4(3):e1000025.
38. Nemenman I, Bialek W, de Ruyter van Steveninck R (2004) Entropy and information in neural spike trains: Progress on the sampling problem. *Phys Rev E Stat Nonlin Soft Matter Phys* 69(5 Pt 2):056111.

Supporting Information (SI Appendix):

Supporting Text 1: Analysis of regularity of spike trains

In order to find at what timescale the spiking was controlled (and hence could be used in motor control), we looked at regularity and predictability within the single motor unit spike trains. Spikes generally occurred at a similar phase within each breathing cycle, separated by gaps on the order of hundreds of milliseconds, and so we only analyzed spikes falling within one breathing cycle. For EMG1, the mean interspike interval (ISI) for such spikes (where the same cycle was defined as ISI $\leq$ 30 ms) was 10.6 ms, while the standard deviation was 4.1 ms, with the refractory period of 2.5 ms. In contrast, for a refractory Poisson process with the same firing rate and refractory period, we would expect the ISI standard deviation of 7.1 ms. Thus the recorded spike train is more regular than a refractory Poisson process, suggesting that spike timing is controlled and can be used for motor coding.

Supporting Text 2: Precise timing observed in awake subjects

In this study, we focused mostly on motor unit recordings from anesthetized, breathing birds (with simultaneous pressure recordings) in order to generate very large data sets. Several factors prevented us from performing all of our studies in awake animals. First, the pressure sensor we employed (see Methods: Pressure Recordings) was too large (10 g) to attach to an awake, freely-moving bird. While previous studies have recorded pressure chronically by attaching a smaller sensor to a backpack on a bird (*1-3*), we needed a sensor with better resolution in order to measure the small changes in air pressure driven by individual motor units. In contrast, previous studies often used pressure to simply identify expiration onset or qualitatively look at changes in

respiration patterns. Secondly, we could not isolate single motor units over long durations of time in awake birds. We were, however, able to record single motor units in four awake birds (without concurrent pressure measurements) for shorter durations, which we used to perform an analysis similar to that done in Figure 1c in the main text. In these awake recordings, as in the anaesthetized birds, we observed similarly strong correlations between consecutive ISIs, which disappeared following jitter on the scale of 1-2 ms or larger (Figure S1).

Supporting Text 3: Individuality of motor behavior in individual birds

Figure 2d in the main text shows that moving a single middle spike by 2 ms produces effects of a similar structure in all birds. Figure S2 illustrates that it is only these changes that are stable across the birds, while the motor behavior itself (as quantified by PTAs) varies substantially among the birds.

Supporting Text 4: Reconstructed wavelets from wfANOVA analysis

Figures S3-S6 illustrate the statistical significance of moving a single spike or a single pulse in a triplet by 2 ms in different experiments employed in our analysis. See below ("Supporting Text 7: *Wavelet-based Functional ANOVA*") for a full explanation of this analysis method.

Supporting Text 5: Muscle stimulation activates motor neurons

Another important consideration for this study is how electrical stimulation actually excites the muscle tissue. Electrical stimulation of muscle tissue (see Methods) might evoke muscle contraction in one of two (non-exclusive) ways: either by directly depolarizing muscle fibers and/or by inducing action potentials in the motor neurons innervating the muscles. To distinguish these possibilities, we examined the effect of applying curare (which blocks synaptic transmission at the neuromuscular junction). As expected, curare treatment largely abolished

EMG signals (Figure S7a,b). We then compared the effects of patterned electrical stimulation on air pressure with and without curare. We found that curare completely abolished the air pressure changes induced by 100 mA current (Figure S7c,d) and mostly abolished air pressure changes induced by 250 mA current (Figure S7e,f). Therefore, although we cannot rule out some contribution of direct muscle fiber activation in our *in vivo* stimulation experiments, it seems likely that any such contribution is minimal. Nevertheless, to evaluate whether such a contribution of direct muscle fiber activation might account for the results shown in main text Figure 4d, we repeated this experiment with 100 mA stimulation current. As shown in Figure S7g, this smaller current, the effect of which is entirely mediated by motor axon activation, produces qualitatively identical results as the 250 mA current used in Figure 4d, providing strong evidence that our *in vivo* stimulation paradigm affect behavior via neuromuscular transmission.

Supporting Text 6: Determining relevant time intervals

To determine the salient duration to test 3-pulse stimulation experiments, two stimulation pulses (each pulse was biphasic with a 250 μs pulse width at 250 μA) were delivered with the interpulse interval (IPI) varied to include 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 40, and 60ms. As controls, a single pulse and a null stimulus were also delivered. All 21 patterns were interleaved during the experiment. We suspected that a precise timing code would rely on muscle nonlinearity at short interspike or interpulse intervals to produce significant differences in forces due to changes in spike timing on the scale of a few milliseconds. Thus we designed experiments to find interpulse intervals that would have a nontrivial, supralinear effect on the motor output. Such interpulse and interspike intervals ($\leq$ 20 ms in length, Figure S12) were then used in the rest of the research (e.g., 10-10 and 12-8 ISI patterns in main text Figure 2).

To find the characteristic time scale for supralinear force production, we measured nonlinear summation by comparing the actual pressure response to a given stimulation pattern with the response constructed by adding the pressure response to a single stimulation pulse at the same times. At a small IPI of 2 ms, the nonlinear summation was large (Figure S12b), while the difference was negligible at an IPI of 20 ms (Figure S12c). We averaged the area difference between these two responses across trials, then took the absolute value for IPIs between 1 ms and 60 ms. This metric showed that the supralinear effect disappeared at IPIs above 20ms (Figure S12d).

Supporting Text 7: ***Wavelet-based Functional ANOVA***

To compare stimulation patterns while removing inter-bird variation, we implemented an analysis technique called wavelet-based functional ANOVA (wfANOVA), which does not treat each time point in a waveform independently since comparisons are performed in the wavelet space (*4*). While the above *d'* analysis provides fine temporal resolution for comparing pressure waveforms from different stimulation patterns, it treats each time point as independent from other time points, when in fact that is not the case. These two types of analyses, therefore, complement each other. The wfANOVA technique was described in detail in (*4*), but we outline it briefly here. Our analysis used an adapted version of the MATLAB code used in (*4*).

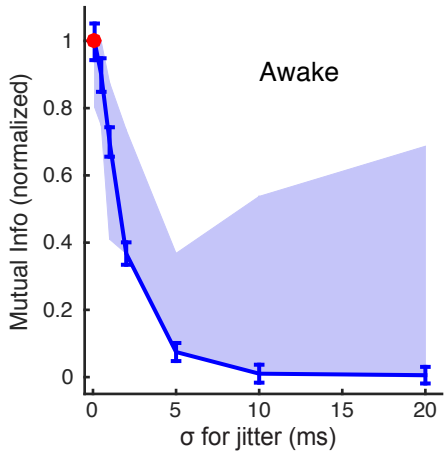
In our approach, each trial was zero-padded to make the total number of time points equal to an exponent of two, as is required for a wavelet transform. All trials were then transformed into the wavelet domain using the discrete wavelet transform, which is similar to the Fourier transform, as it produces coefficients for a family of wavelets that can be linearly transformed back into the time domain. We chose to use the third-order coiflet family, as in (*4*), because their symmetry does not introduce phase shifts into the data and their shapes approximated those of

the pressure waveforms. The wavelets in this family are orthogonal to each other. This is important for implementing multivariate ANOVA, which performs poorly when there are high correlations between data points, as is the case in the time domain. Because wavelets are localized in time unlike Fourier components of a signal, wavelet transform of our signals featured many small or zero magnitude wavelets. The coefficients determined for each of the wavelets in the family were then used for subsequent ANOVA.

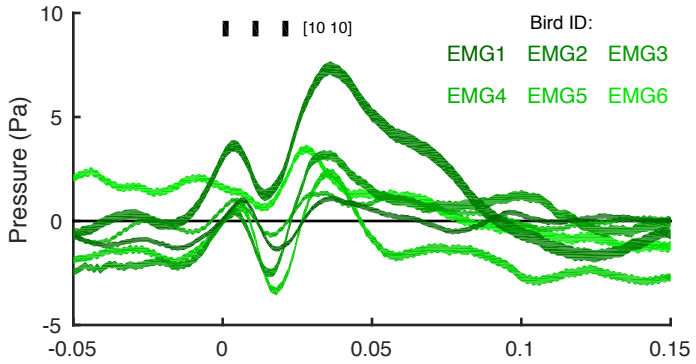
To perform the ANOVA, all trials were grouped by two factors: stimulation (or spike) pattern and bird. Fixed-effect, two-factor ANOVA was performed on the wavelet-decomposed pressure waveforms from the EMG and the stimulation experiments. This analysis therefore determined whether different spike/stimulation patterns evoked different effects on air pressure by quantifying whether one or more wavelet coefficients of the response difference was significantly nonzero (α = 0.05, F test with separate post hoc Scheffé tests, with the null hypothesis being that the pressure waveforms – represented by the coefficients of the wavelet decomposition – from the pairs of stimulation patterns were equal). These post hoc tests were conducted using a significance level that was Bonferroni-corrected for the number of significant F-tests corresponding to the initial ANOVA with respect to the stimulation pattern factor (i.e., 0.05 divided by the total number of post hoc tests being conducted). Wavelet coefficients that were significantly different between stimulation patterns were subtracted to find the coefficient corresponding to the difference between the two patterns. Even having one wavelet with significant differences in their coefficients would indicate a significant difference between the two input patterns. All nonsignificant coefficients were set to zero, and the entire decomposition was transformed back into the time domain. This produced a contrast curve between two stimulation patterns across birds that removed the effects of inter-subject variability on the result.

This analysis technique was also used for the *in vitro* muscle experiments, with the input being the force waveforms instead of the pressure waveforms.

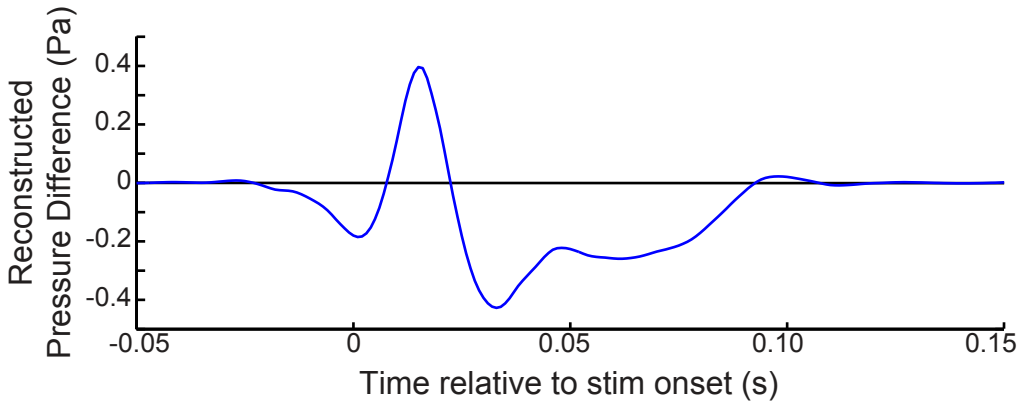
Supporting Figures:



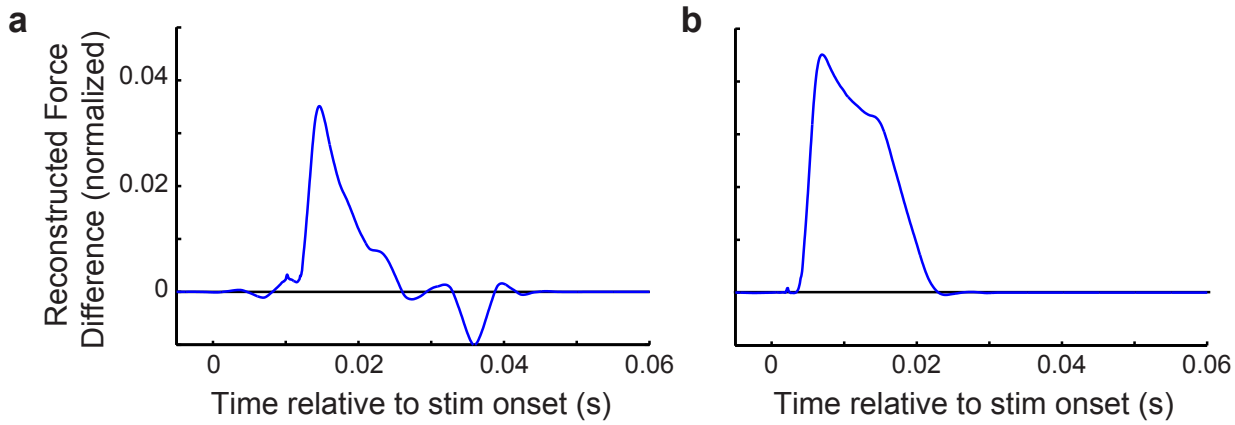
Supporting Fig. S1: Timescales in the spike trains in awake birds. As in anesthetized birds (main text Figure 1), mutual information in jittered spike trains recorded from awake birds approached that in the original recordings only for jitters on the scale ~1 ms. The blue line shows results from the longest spike train recorded for awake birds (EMG 9), for which we had over 9,000 recorded spikes; the band shows the range across 4 awake birds. The similarity to Figure 1c suggests that the observed millisecond-scale regularities in the spike train were not due to anesthesia. The upper limits of the range shown were driven by a bird with large error bars on its mutual information measurements (relative error of up to 27%), resulting in a statistically insignificant increase in the mutual information for larger jitters. Un-normalized values of mutual information at $\sigma = 0$ ranged from 0.094 to 0.147 bits across awake birds, which is similar to anesthetized birds.



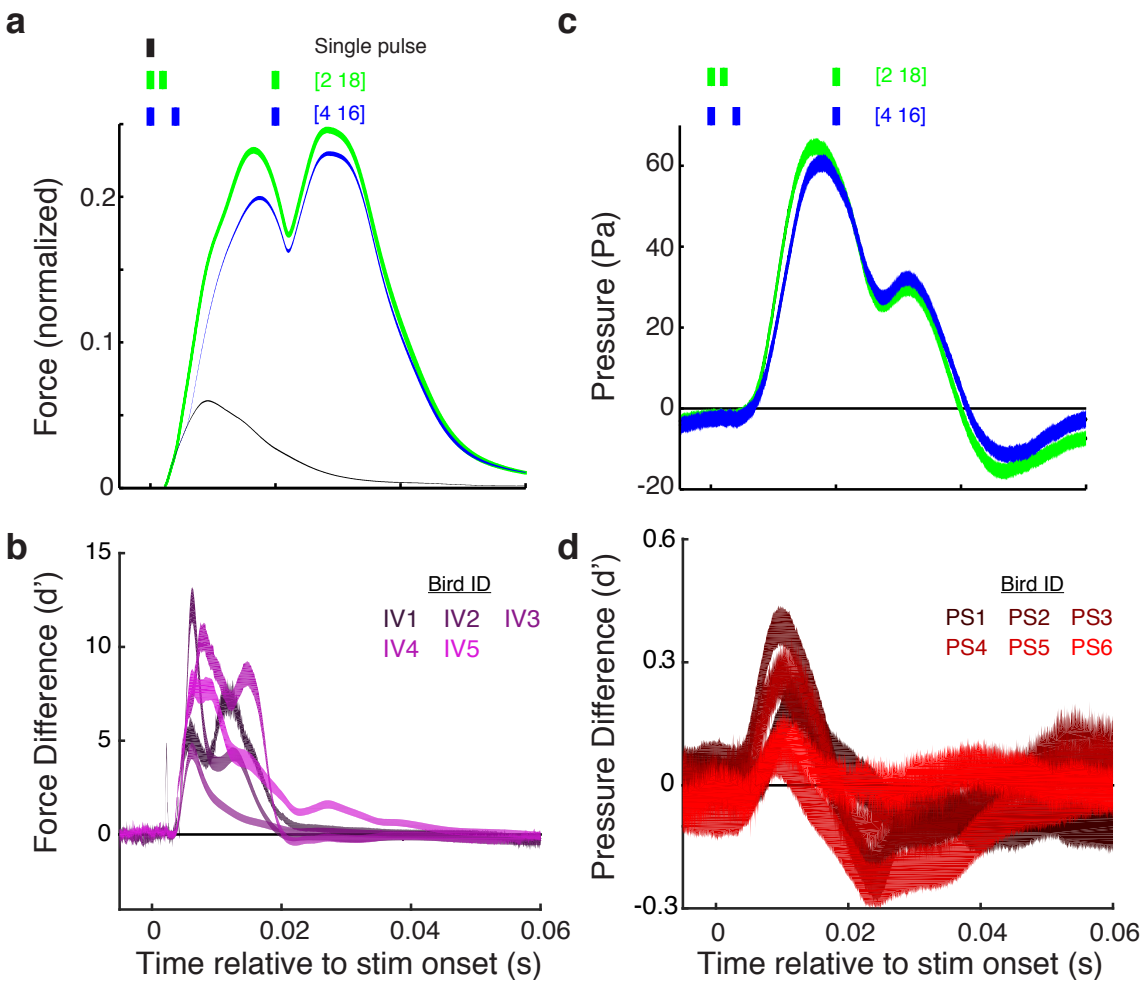
Supporting Fig. S2: Differences in PTAs across birds. In Figure 2d, we show that differences in PTAs across birds are remarkably consistent. This is not, however, a result of each bird having the same PTAs for each pattern, and does not mean that all of the motor units in the muscle have the same effect on air pressure. Here we show the PTAs for the 10-10 pattern treated in Figure 2; these correspond to the green trace in Figure 2b, but for all six EMG birds instead of just EMG1. Note that the scales on the vertical axes of Figure S2 and Figure 2b are different, and that bird EMG1 (whose data is also shown in Figure 2b in the main text) corresponds to the thin dark green line with the range of about +/- 1 Pa."



Supporting Fig. S3: wfANOVA analysis of air pressure. Reconstructed signal from wfANOVA (18 significant wavelets, post-hoc $\alpha = 6.66 \times 10^{-4}$) revealed the difference in air sac pressure between 3-spike patterns over 20 ms, with the middle spike either occurring at 10 ms or 12 ms after the first spike (i.e., 10-10 vs 12-8 spike triplets), independent of inter-subject variability (see main text Figure 2b-d).

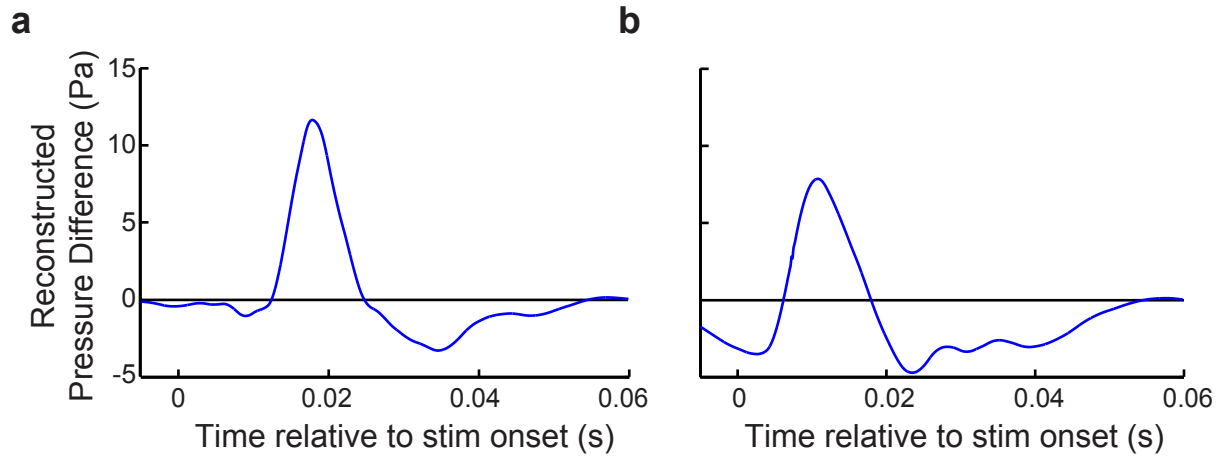


Supporting Fig. S4: wfANOVA analysis of *in vitro* muscle force. Reconstructed signal from wfANOVA revealed the effect of moving the *in vitro* stimulation pulse on muscle force, independent of inter-subject variability. Note that as in main text Figure 3b, force measurements are normalized to the maximum force during tetanic contraction. For 3-pulse *in vitro* stimulations taking place over 20 ms (see Figure 3), significant wavelets demonstrated the difference between patterns with the middle pulse at (a) 10 ms and 12 ms (i.e., 10-10 vs. 12-8 triplets, 29 significant wavelets, post-hoc $\alpha = 2.71 \times 10^{-4}$) and at (b) 2 ms and 4 ms (2-18 vs. 4-16 stimulation triplets, 47 significant wavelets, post-hoc $\alpha = 2.97 \times 10^{-4}$).

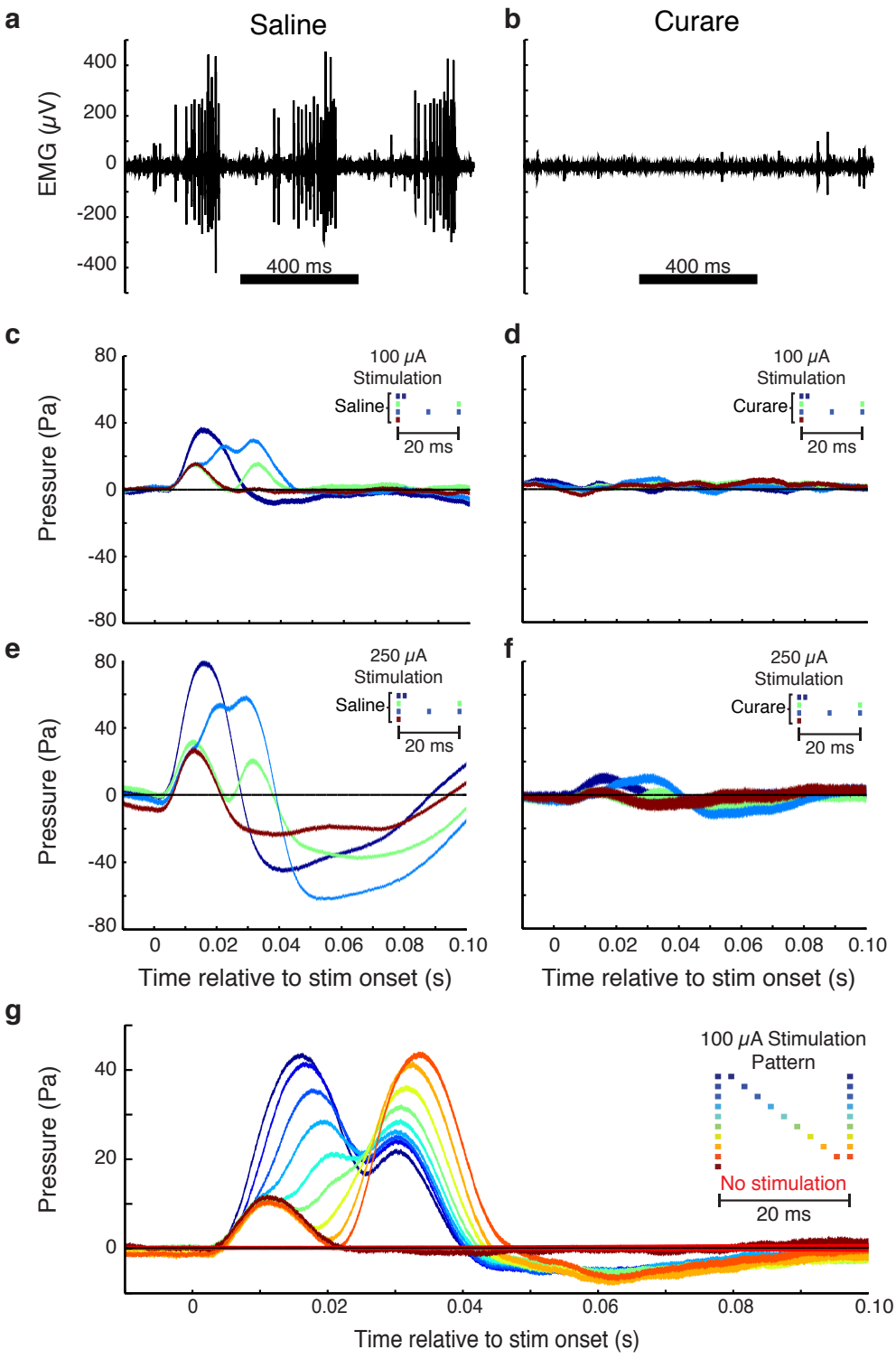


Supporting Fig. S5: Motor timing showed significant effects on force and behavior when middle pulse is close to the beginning of interval (in the main text, we studied cases when the middle pulse was near the middle of the interval). (a) Force waveforms were significantly different when moving the middle stimulation pulse from 2 ms to 4 ms after the first spike in a 20 ms 3-spike pattern (47 significant wavelets, post-hoc $\alpha = 2.97 \times 10^{-4}$, see Figure S4b). (b) This significant effect, as measured by *d*', was observed across all 5 subjects. (c) The same stimulation patterns caused a significant difference in air sac pressure (16 significant wavelets, post-hoc $\alpha = 1.64 \times 10^{-4}$, see Figure S6b below). (d) This effect, measured by *d*', was also

observed across all 6 subjects. Note that these stimulation patterns tested the limits of the system, but they were not observed as spike patterns during EMG recordings in unstimulated birds.

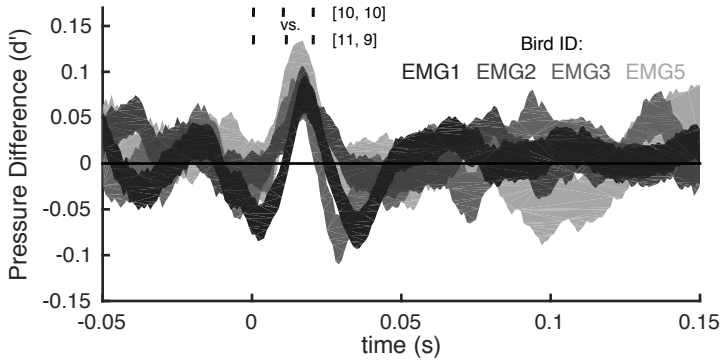


Supporting Fig. S6: Reconstructed signal from wfANOVA revealed the effect of moving the *in vivo* stimulation pulse on air sac pressure, independent of inter-subject variability. For 3-pulse *in vivo* stimulations taking place over 20 ms (main text Figure 4), significant wavelets demonstrated the difference between patterns with the middle pulse at (a) 10 ms and 12 ms (i.e., 10-10 vs. 12-8 pulse triplets, 14 significant wavelets, post-hoc $\alpha = 1.18 \times 10^{-4}$), and at (b) 2 ms and 4 ms (i.e., 2-18 vs. 4-16 triplets, 16 significant wavelets, post-hoc $\alpha = 1.64 \times 10^{-4}$).

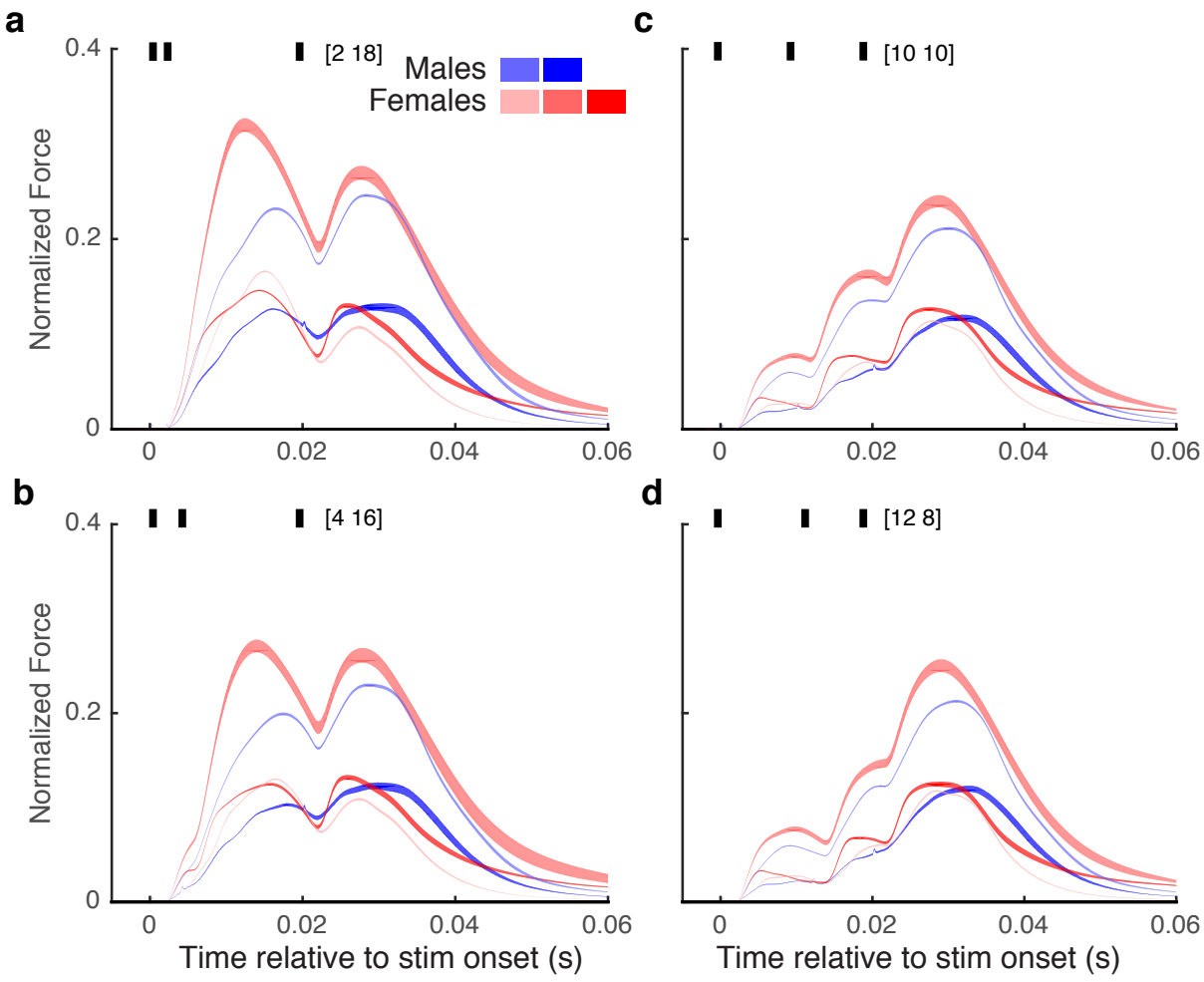


Supporting Fig. S7: Curare experiments suggest that EXP stimulation activates the axons of motor neurons, and not the muscle fibers directly. (a) EMG recordings showed strong

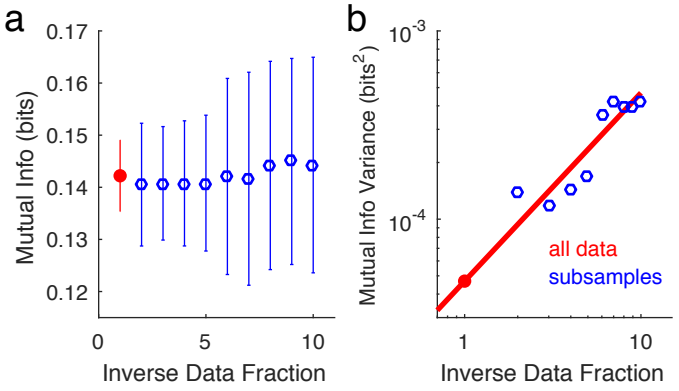
activity when saline was applied to the muscle, (b) but that activity quickly disappeared when curare, which locally blocked synaptic transmission at the neuromuscular junction, was applied to the muscle. (c) Various stimulation patterns at 100 μA had clear effects on air sac pressure, (d) but curare abolished those effects at the same stimulation current. (e) The same stimulation effects at 250 μA were (f) greatly reduced by curare. (g) Our experiments in the main text described the effects of EXP stimulation at a current of 250 μA. We saw similar effects at currents such as 100 μA shown here, where stimulation effects were abolished by curare as in (c). The color bars represent mean +/- s. e. m.



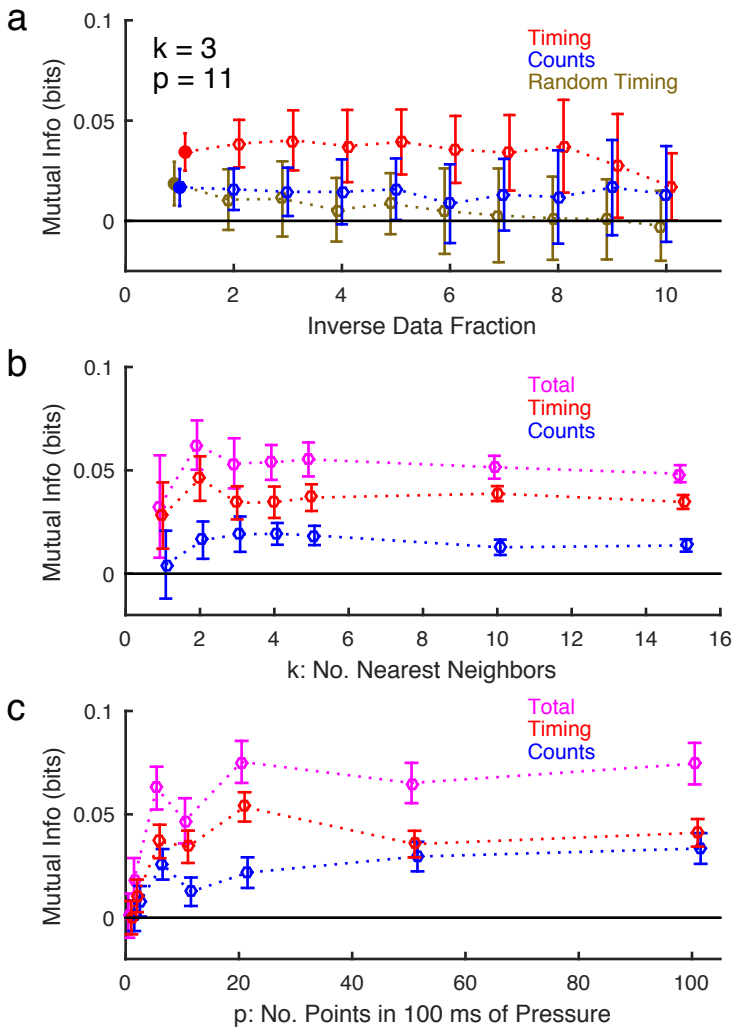
Supporting Fig. S8: Distinguishability of Pattern Triggered Averages (PTAs) on 1 ms scale. Similar to main text Figure 2d, here we plot *d*' (mean +/- s.e.m, bootsrapped) for the distinguishability of the pressure traces triggered by 10-10 and 11-9 ms ISI patterns — different only by the position of the middle spike in a spike triplet by 1ms. Of the six birds analyzed in Figure 2d, four (EMG1, EMG2, EMG3, and EMG5) had statistically significantly nonzero *d*' curves triggered by these 1-ms different spike patterns. At peak, EMG1 showed *d*'=0.08+/-0.02. The decrease in statistical significance likely reflects having many fewer equivalent patterns if viewed at a higher temporal resolution (N = 23,991 and 6,507 for 10-10 triplet for EMG1 at 2 ms and 1 ms resolution, respectively).



Supporting Fig. S9: *In vitro* stimulation of EXP did not show any qualitative differences across sexes. Across all four stimulation patterns (a-d), all subjects showed similar force trajectories, with no subjects showing differences outside of normal inter-subject variability. Sample size for each group was too small to perform a statistical comparison, but the data shown above suggest that using muscle fibers from both male and female birds for *in vitro* studies (see Methods) did not greatly affect our results.

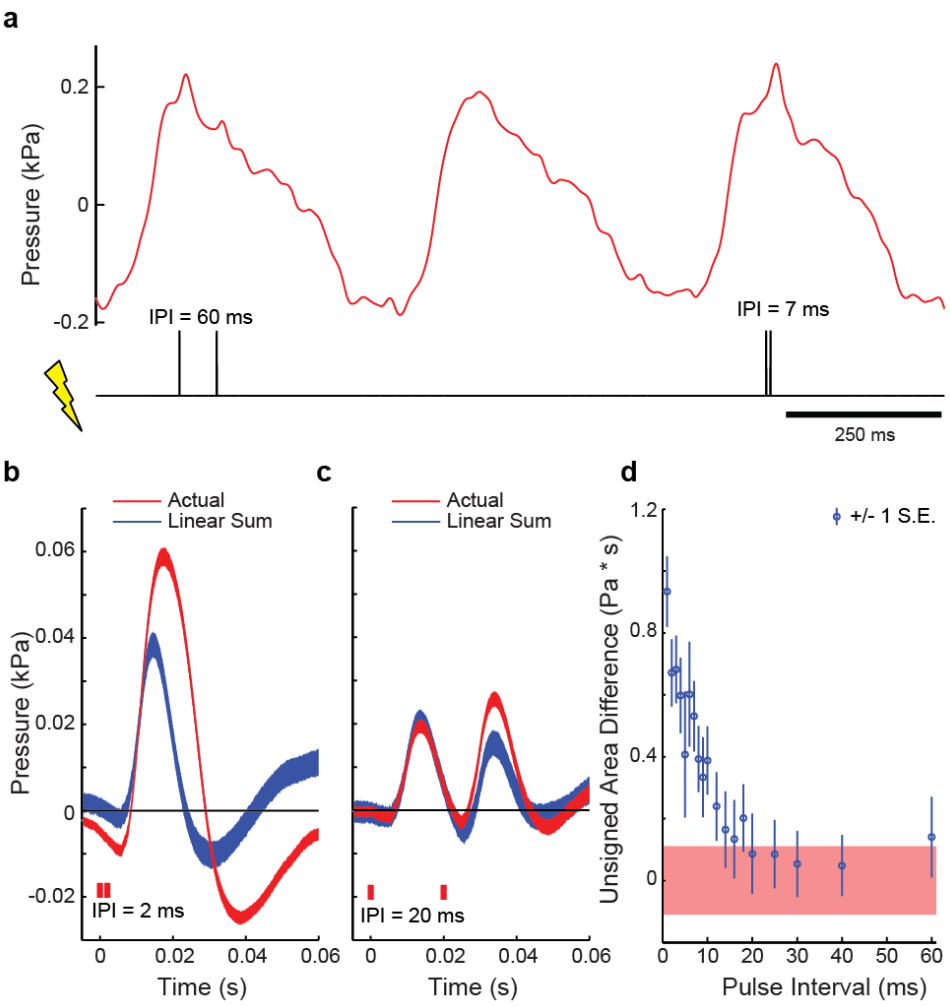


Supporting Fig. S10: Estimation of mutual information between consecutive ISIs for EMG1. (a) The plot of mutual information vs. data set size shows stability of the calculated values, and hence absence of the sample-size dependent bias in the estimation of information. The full data set here includes N = 154,548 consecutive ISIs pairs (red dot). Estimates for other data set sizes (blue circles) are obtained by taking $1/m$ fraction of the total amount of data, with m shown on the horizontal axis (see *Methods: Mutual Information: Consecutive ISIs*). (b) For subsampled data (blue), we estimate the variance, σ^2, of the mutual information estimates (σ is also shown as error bars in panel (a)) as the variance of m non-overlapping subsamples. We repeated this ten times at each m and averaged the variances. As discussed in Methods: Mutual Information: Consecutive ISIs, we then extrapolated the variance to the full data set using linear regression (red line and dot), and the extrapolated standard deviation is shown as the error bar in panel (a).



Supporting Fig. S11: Estimation of mutual information between spikes and pressure for EMG1. (a) We plot mutual information as a function of inverse data size, similar to Figure S10a for the chosen k, the number of nearest neighbors in the mutual information estimator (see panel (b)), and p, the number of data points used to represent the pressure (see panel (c)). Here the full data set is N=16,516 breathing cycles (solid circle, inverse data fraction of 1). Information between the spike count and the pressure (blue) and the spike timing and the pressure (red) are shown separately (also reproduced in main text Figure 2a for p = 11, k = 3, and the full data set). To further establish statistical significance of the timing mutual information estimate, we show the mutual information between the pressure traces and randomly shuffled spike timings, while

keeping the number of spikes fixed (brown). This information falls within error bars from zero, as it should because random spike timing does not carry information about the pressure. (b) As explained in Methods, we chose the number of nearest neighbors, k, for the estimation algorithm (k = 3 for EMG1) such that k is large enough to have small estimated error bars, and yet small enough so that fine features in the distribution are not averaged out and the information does not decrease a lot from the maximum. We verify the choice of k by insisting on the smallest possible sample size dependent drift in the estimated information, as in panel (a). Spike timing (red), spike count (blue), and total (magenta) mutual information values are shown. (c) As elaborated on in Methods, we chose the same number of points (p = 11, 10 ms apart) to represent 100 ms of pressure in all EMG birds. The choice was guided by the requirement of using the smallest p that leads to both best sampling and smallest overall error bars, while large enough to result in a mutual information that has reached its large-p plateau value across all birds simultaneously, within error bars. As in panel (b), spike timing (red), spike count (blue), and total (magenta) mutual information values are shown.



Supporting Fig. S12: Nonlinear summation in thoracic air sac pressure in response to *in vivo* stimulation. (a) During respiration, stimulation of a varying interpulse interval (IPI) was delivered after air sac pressure crossed a user-defined threshold (see *Methods: In Vivo Muscle Stimulation*). (b) At small IPIs ~2 ms, the actual pressure response was much greater than that produced from summing two single pulse responses separated by 2 ms. (c) On the other hand, the responses were quite similar for an IPI of 20 ms. In these plots, the width of the color bands represent the mean +/- s.e.m. (d) The difference between the actual and the summed response dropped as IPI increased, leveling off after 20 ms. The red band denotes the mean +/- s.e.m. when calculating the (unsigned) difference of areas between responses to individual pulses and

the mean single pulse response; the area of this region provides an estimate of the measurement variability introduced by computing the difference between many individual pressure traces and the mean trace. As expected, the uncertainty of measurements of area difference between two-pulse responses and a sum of two single pulse responses at different pulse intervals (blue error bars) are of roughly the same magnitude as the red area. At >20 ms IPI, the area differences are indistinguishable from zero. Data shown are from bird pSTIM3.

Supporting Tables:

Supporting Table S1: Results of wfANOVA

Type of Experiment	**Number of Significant Wavelets**	**Bonferroni-Corrected Alpha**	**Smallest p-value**	**Largest p-value**	**Figure Reference**
EMG	18	1.10×10^{-3}	3.36×10^{-86}	6.66×10^{-4}	Fig. 2, Fig. S3
In vitro stim (10-10 vs 12-8)	39	3.70×10^{-4}	4.04×10^{-105}	2.71×10^{-4}	Fig. 3, Fig. S4a
In vitro stim (2-18 vs 4-16)	47	3.47×10^{-4}	3.47×10^{-82}	2.97×10^{-4}	Figs. S4b, S5a-b
Air Sac Stim (10-10 vs 12-8)	14	1.18×10^{-4}	5.31×10^{-56}	6.79×10^{-5}	Fig. 4, Fig. S6a
Air Sac Stim (2-18 vs 4-16)	16	1.64×10^{-4}	2.73×10^{-50}	1.43×10^{-4}	Figs. S5c-d, S6b

References

1. J. M. Wild, F. Goller, R. A. Suthers, Inspiratory muscle activity during bird song. *J Neurobiol* **36**, 441-453 (1998).
2. R. C. Ashmore, J. M. Wild, M. F. Schmidt, Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *J Neurosci* **25**, 8543-8554 (2005).
3. F. Goller, O. N. Larsen, A new mechanism of sound generation in songbirds. *Proc Natl Acad Sci U S A* **94**, 14787-14791 (1997).
4. J. L. McKay, T. D. J. Welch, B. Vidakovic, L. H. Ting, Statistically significant contrasts between EMG waveforms revealed using wavelet-based functional ANOVA. *J Neurophysiol* **109**, 591-602 (2013).