

RESEARCH ARTICLE

Closed-loop stabilization of the Jamming Avoidance Response reveals its locally unstable and globally nonlinear dynamics

Manu S. Madhav^{1,*}, Sarah A. Stamper¹, Eric S. Fortune² and Noah J. Cowan¹

¹Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218, USA and

²Department of Biological Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA

*Author for correspondence (manusmad@jhu.edu)

SUMMARY

The Jamming Avoidance Response, or JAR, in the weakly electric fish has been analyzed at all levels of organization, from whole-organism behavior down to specific ion channels. Nevertheless, a parsimonious description of the JAR behavior in terms of a dynamical system model has not been achieved at least in part due to the fact that ‘avoidance’ behaviors are both intrinsically unstable and nonlinear. We overcame the instability of the JAR in *Eigenmannia virescens* by closing a feedback loop around the behavioral response of the animal. Specifically, the instantaneous frequency of a jamming stimulus was tied to the fish’s own electrogenic frequency by a feedback law. Without feedback, the fish’s own frequency diverges from the stimulus frequency, but appropriate feedback stabilizes the behavior. After stabilizing the system, we measured the responses in the fish’s instantaneous frequency to various stimuli. A delayed first-order linear system model fitted the behavior near the equilibrium. Coherence to white noise stimuli together with quantitative agreement across stimulus types supported this local linear model. Next, we examined the intrinsic nonlinearity of the behavior using clamped frequency difference experiments to extend the model beyond the neighborhood of the equilibrium. The resulting nonlinear model is composed of competing motor return and sensory escape terms. The model reproduces responses to step and ramp changes in the difference frequency (df) and predicts a ‘snap-through’ bifurcation as a function of df that we confirmed experimentally.

Key words: *Eigenmannia*, electrosensory system, feedback, dynamical systems modeling.

Received 16 April 2013; Accepted 8 August 2013

INTRODUCTION

Weakly electric fish emit an electric organ discharge (EOD) that is used for electrolocation (Caputi and Budelli, 2006; von der Emde, 2006) and communication (Hopkins, 1974; Fortune, 2006; Hupé and Lewis, 2008; Triefenbach and Zakon, 2008). In wave-type electric fish, the EOD is quasi-sinusoidal and has a relatively stable baseline frequency when undisturbed (Bullock, 1969; Bullock et al., 1975; Moortgat et al., 1998; Zakon and Dunlap, 1999; Zakon et al., 2002). When two or more fish are in close proximity, their EODs interact to produce emergent amplitude modulations (AM) at the difference frequency (df) between the fish. *Eigenmannia virescens* shift their baseline EOD frequency in response to low (<15 Hz) df s (Watanabe and Takeda, 1963; Bullock et al., 1972), which have been shown to impair aspects of electrolocation (Heiligenberg, 1973; Bastian, 1987). The direction of the frequency shift is determined by the sign of the df and results in an increase in the magnitude of the df . This behavior is known as the Jamming Avoidance Response (JAR).

The JAR has been analyzed at all levels of organization, from whole-organism behavior down to specific ion channels (Fortune and Rose, 2003; Heiligenberg, 1991; Bastian and Heiligenberg, 1980; Heiligenberg and Bastian, 1980). Despite the fact that the JAR is among the best understood sensorimotor circuits, the sensorimotor responses have not been modeled as a dynamical system. One challenge to modeling the temporal dynamics arises from the intrinsically ‘unstable’ nature of the JAR. This is because the fish shifts its EOD frequency (f_1) in the direction away from the

frequency of the conspecific (f_2), resulting in an increase of $|df|$, where $df=f_2-f_1$. We overcame this challenge by closing a feedback loop around the natural behavior: the frequency of a conspecific-like signal was calculated and adjusted in real-time to stabilize the response and drive it to any desired frequency in a neighborhood of the fish’s original baseline frequency.

Perturbation experiments on the stabilized closed-loop system were used to characterize the dynamics of the JAR. These perturbations included sinusoids, sums of sinusoids, chirps and band-limited noise. Responses to these stimuli were used to estimate a non-parametric frequency response function (FRF). The FRF was then used to infer the frequency response of the open-loop behavior, i.e. the JAR itself. A first-order delayed parametric model was fitted to the behavior near its equilibrium.

This local model does not, however, capture the nonlinear features of the behavior: the biological relevance of the JAR lies in its escape from the unstable equilibrium. To address this, we extended the linear model using additional experiments in which the df s were ‘clamped’ to furnish a complete nonlinear model. The nonlinear model was parsed into terms that capture competing avoidance and return responses and was validated by comparison with responses to open-loop stimuli. The model was also used to predict a saddle–node bifurcation in the vector field of the system. The bifurcation causes one stable equilibrium to vanish, ultimately manifesting as the ‘snap-through’ of the fish’s frequency (f_1) from one stable equilibrium (low $|df|$) to another (high $|df|$ of opposite sign).

MATERIALS AND METHODS

Experimental setup

Adult *Eigenmannia virescens* (Valenciennes 1836) (10–15 cm in length, EODs between 346 and 452 Hz) were obtained from commercial vendors. The fish were housed in group aquarium tanks that had a water temperature of $\sim 27^\circ\text{C}$ and a conductivity in the range $150\text{--}500\ \mu\text{S cm}^{-1}$ (Hitschfeld et al., 2009). All experimental procedures were approved by the Johns Hopkins Animal Care and Use Committee and followed guidelines established by the National Research Council and the Society for Neuroscience.

The experimental tank was maintained at a temperature of $25\pm 3^\circ\text{C}$ and a conductivity of $150\pm 25\ \mu\text{S cm}^{-1}$. Each fish ($N=7$; $n=5$ reference-tracking trials, $n=2$ clamp trials) was tested individually. Each fish was acclimated to the testing tank for a minimum of 24 h prior to the start of the experiment. After the initial acclimation period the fish was restricted in a chirp chamber for 2–3 h to allow the EOD frequency to stabilize. The chirp chamber served to restrict the movement of the fish and prevent changes in orientation during the experiment, resulting in more consistent measurements of the EOD frequency. A pair of measurement electrodes (red) were placed longitudinally (near the fish's head and tail) to record the EOD and a second pair of stimulus electrodes (black) were placed transverse to the fish to provide a frequency-controlled sinusoidal stimulus (Fig. 1A). The distance between the electrodes in each pair was 25 cm.

The EOD of the fish, recorded *via* the head-to-tail electrodes (Fig. 1A; red circles), was filtered and amplified (0.1 Hz to 1 kHz bandpass, gain 100; A-M Systems Model 1700, Sequim, WA, USA)

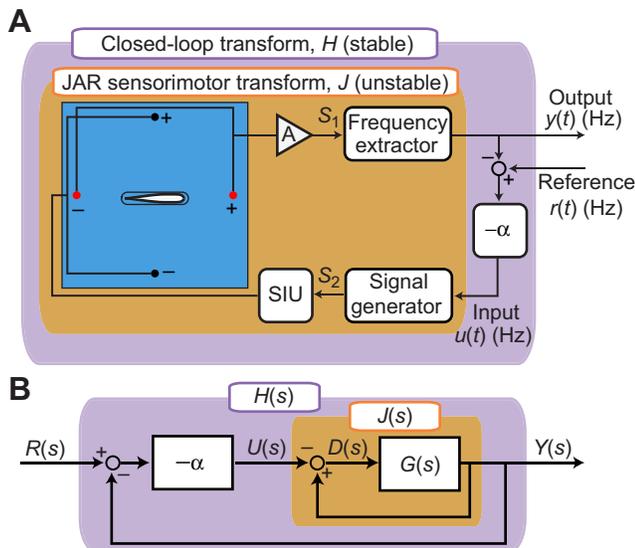


Fig. 1. Experimental setup. (A) Schematic diagram of the closed-loop system. The reference frequency $r(t)$, electric organ discharge (EOD) frequency $y(t)$ and the frequency of the input signal $u(t)$ are all baseline subtracted, so that 0 Hz corresponds the fish's baseline EOD frequency. The EOD is measured (recording electrodes, red), amplified, and its frequency $y(t)$ is extracted. The input frequency $u(t)=\alpha[y(t)-r(t)]$ is fed to a signal generator, which outputs a sinusoid of that frequency. This sinusoid is played back through a stimulus isolation unit (SIU) to the fish (stimulating electrodes, black). (B) A block diagram representation of the experimental system. Here, $R(s)$, $U(s)$, $Y(s)$ and $D(s)$ are the Laplace transforms of the reference, input, output and difference frequencies, respectively. $J(s)$ represents the (open-loop) JAR behavior, namely the transfer function from $U(s)$ to $Y(s)$. $G(s)$ is the open-loop transfer function from the computed difference $D(s)$ and output $Y(s)$. The closed-loop transfer function $H(s)$ relates $R(s)$ to $Y(s)$.

and input to a frequency-to-voltage converter (F2V; FV-1400, Ono-Sokki, Yokohama, Japan). The F2V calculates the frequency of the signal using measured time differences between zero crossings, digitally averages these values over four cycles, and outputs a voltage that varies linearly with frequency over a set range (e.g. the range $\pm 5\ \text{V}$ would correspond to 475–525 Hz). This creates a time lag of no more than four cycles of the EOD. The F2V output was further filtered (Chebyshev low-pass, 30 Hz cutoff). All told, these delays correspond to a worst-case phase lag of 5.2 deg relative to the highest reference stimulus frequency (1 Hz), at the lowest EOD baseline (346 Hz) tested. Both the amplified signal and the filtered F2V output were fed into a Power1401 Mk.II signal acquisition device (CED, Cambridge, UK) which ran a custom-written sequencer script that read the input signal, performed the feedback calculation and generated a sinusoid with the desired output frequency. Using this setup enabled regular temporal sampling intervals and provided more deterministic computation time than including a standard computer in the feedback. The signal acquisition device received parameters and reference signal for each trial from the Spike2 software (CED), which ran simultaneously on a computer. During the trials, this software received and recorded data from the input, output and intermediate channels. The amplitude of the stimulating sinusoid was $100\ \mu\text{V cm}^{-1}$ (unless otherwise noted), which produced 14–43% contamination across individuals (EOD amplitudes $232\text{--}715\ \mu\text{V cm}^{-1}$) measured at a 1 cm dipole placed adjacent to the head of the fish. We performed an identification experiment on the feedback system to assess what its characteristics were, especially the delay introduced by the equipment. The contribution was minimal ($\sim 2\ \text{ms}$ delay) due to fast instrumentation. As such, the feedback delay was not incorporated into subsequent calculations.

The closed-loop approach

The fish EOD frequency $f_1(t)$ and the stimulus frequency $f_2(t)$ are functions of time. Under constant lighting, temperature and conductivity, and without conspecific stimulation, the EOD frequency remains relatively stationary over long periods of time (Bullock, 1969). The initial time $t=0$ for each trial was preceded by a period of no stimulation for at least 300 s, and we defined the baseline (initial) frequency $f_1(0)$ as the EOD frequency at that initial time.

The following control variables were chosen as frequencies relative to $f_1(0)$:

$$\begin{aligned} y(t) &= F_1(t) - F_1(0), \\ u(t) &= F_2(t) - F_1(0), \\ d(t) &= y(t) - u(t). \end{aligned} \quad (1)$$

We refer to $u(t)$ as the 'input' and $y(t)$ was the measured 'output' of the behavior. Note that $u(t)$ and $y(t)$ are frequencies relative to the baseline frequency, and are thus an abstraction of the stimulus and response, and not the raw signals themselves. The signal $d(t)$ is the difference between these signals, also referred to as the d_f .

Given a reference signal $r(t)$, a simple proportional controller was able to stabilize the system:

$$u(t) = -\alpha(r(t) - y(t)). \quad (2)$$

The frequency of the applied signal S_2 was simply $f_2(t)=u(t)+f_1(0)$. The controller gain was typically selected as $\alpha=2$ (unless otherwise noted) and was positive for all experiments.

In the frequency domain, $J(s)$ denotes the input–output transfer function corresponding to the behavior at frequency $s=j\omega$. Thus $J(s)$

is the behavior transfer function from $U(s)$ to $Y(s)$, where $U(s)$ and $Y(s)$ are the Laplace transforms of $u(t)$ and $y(t)$, respectively. $G(s)$ is the open-loop transfer function between $D(s)=[Y(s)-U(s)]$ and $Y(s)$. Similarly, $H(s)$ is the closed-loop transfer function between $R(s)$ and $Y(s)$ (Fig. 1B). The open-loop transfer function, $J(s)$, can be converted to closed loop, $H(s)$, and *vice versa*. The following equations relate these three transfer functions:

$$\frac{Y(s)}{R(s)} = H(s) = \frac{-\alpha J(s)}{1 - \alpha J(s)} = \frac{-\alpha G(s)}{(1 - \alpha)G(s) - 1}, \quad (3)$$

$$\frac{Y(s)}{U(s)} = J(s) = \frac{H(s)}{\alpha(H(s) - 1)} = \frac{G(s)}{G(s) - 1}, \quad (4)$$

$$\frac{Y(s)}{D(s)} = G(s) = \frac{J(s)}{J(s) - 1} = \frac{H(s)}{H(s)(1 - \alpha) + \alpha}. \quad (5)$$

Stimulus types

Single sines

Sinusoidal stimuli were of frequencies 0.01, 0.055, 0.215 and 0.995 Hz and of durations 1000, 1000, 500 and 200 s, respectively. The stimulus durations were chosen to have a sufficient number of beat cycles for spectral analysis.

Sum of sines

These stimuli were the sum of 10 logarithmically spaced sinusoids with randomized phase, in the range 0.01–1 Hz. The sum-of-sine stimuli included the four single-sine frequency components. The stimulus duration was 1000 s.

Chirps

The chirp stimulus was a sinusoid of increasing frequency, from 0.01 to 1 Hz over 1000 s. The increase of frequency was exponential, ensuring sufficient stimulus power across all frequencies. Several weakly electric fish also exhibit EOD frequency changes termed ‘chirps’, not to be confused with the chirp stimulus used in this paper.

Long chirps

The long-chirp stimulus was similar to the chirp stimulus except that the frequency increased from 0.001 to 1 Hz. Consequently, the stimulus duration was increased to 10,000 s.

Band-limited pseudo-random noise

Band-limited noise stimuli consisted of non-overlapping, 2 Hz wide frequency bins from 0 to 20 Hz. For each bin, a stimulus was generated by summing together sinusoids associated with all uniformly spaced frequencies as dictated by the sampling rate. This ensured that the stimulus has uniform power over all the frequencies analyzed within each bin. Each component had randomized phase and was scaled equally so that the sum would have a maximum magnitude of 1 Hz. The trial duration was 300 s, except for the lowest frequency band of 0–2 Hz, which was 1000 s long.

Trial types

Closed-loop reference-tracking trials

For closed-loop reference-tracking trials, we provided a reference signal $r(t)$ and applied the controller as described in Eqn 2. At the start of each trial, the fish’s baseline was measured. Subsequently, there was a ‘balancing’ period of 100 s wherein the controller aimed to keep the fish at the initial EOD frequency [$r(t)=0$]. To avoid startling the fish, the amplitude of the stimulus signal was

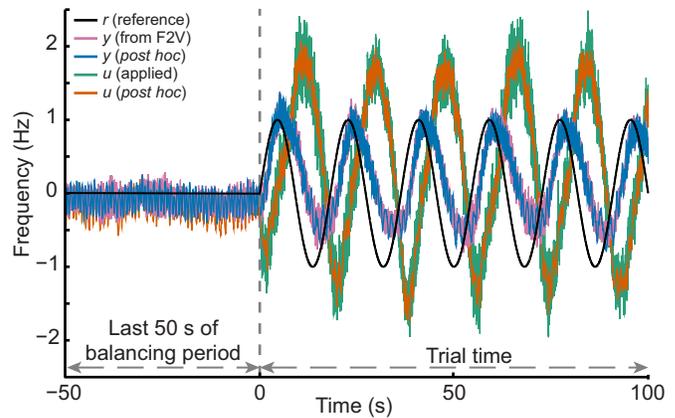


Fig. 2. Sample closed-loop trial and verification of applied signals *via post hoc* analysis of measured signals. The fish’s baseline-subtracted output frequency $y(t)$ from the F2V converter (pink) was used during the experiment and its value was verified with a *post hoc* estimate of frequency (blue) based on a spectrogram of the measured EOD signal. The reference signal $r(t)$ for the trial (black) is predetermined. The baseline-subtracted input frequency $u(t)=\alpha[y(t)-r(t)]$ (green) was verified against a *post hoc* estimate of the applied frequency (orange) based on a spectrogram of the measured input signal.

ramped up linearly from 0 to 100 $\mu\text{V cm}^{-1}$ (unless noted) over the first 50 s of the balancing period. After the balancing period, a stimulus (single sine, noise, etc.) was introduced as the reference. Trials were pseudo-randomized. All stimuli were pre-generated at 1 kHz and input to the sequencer at the start of the trial. A sample interval of a closed-loop trial with a single-sine stimulus is shown in Fig. 2.

Each fish ($N=5$) completed closed-loop reference-tracking trials ($n=49$) in a randomized order within a single testing session. The trials consisted of the following: (i) single sine stimuli ($n=8$), the four frequencies of which were replicated for magnitudes of 1 and 2 Hz; (ii) sum-of-sines stimuli ($n=4$), with two different component magnitudes (0.2 and 0.3 Hz), with two different sets of randomized component phases each; thus, the maximum stimulus magnitude was 2 or 3 Hz; (iii) chirps ($n=2$), with magnitudes of 1 and 2 Hz; (iv) white noise stimuli ($n=30$), with three identical replicates of each of the 10 frequency bands; (v) chirp ($n=2$) stimuli with signal amplitude 50 and 200 $\mu\text{V cm}^{-1}$ (typical value=100 $\mu\text{V cm}^{-1}$); these were to examine sensitivity of the identified system to signal amplitude; and (vi) chirp ($n=3$) stimuli with controller gains 1.5, 2.5 and 3 (typical value=2); these stimuli were, similarly, to examine sensitivity to the feedback gain. Sample responses to four stimulus types are shown in Fig. 3.

In addition, a subset of three individual fish were presented with long chirp stimuli at two magnitudes (1 Hz, 2 Hz). After each trial, the stimulus was turned off and the EOD frequency was allowed to stabilize over a period of 300 s before the next trial began.

Closed-loop clamp trials

Instead of driving the fish frequency towards a goal, the clamp trials applied a stimulus such that $d(t)$ was maintained at a desired value. Two types of clamp trials were run: (i) static clamps, where the clamp was a constant value $d=d_s$ for 300 s; and (ii) dynamic clamps, where the clamp was kept at a particular value $d=d_s$ for 100 s and then oscillated around the value according to a reference trajectory $r(t)$, such that $d(t)=d_s+r(t)$.

Static- and dynamic-clamp experiments were performed on $N=2$ individuals that did not complete the reference-tracking trials. These

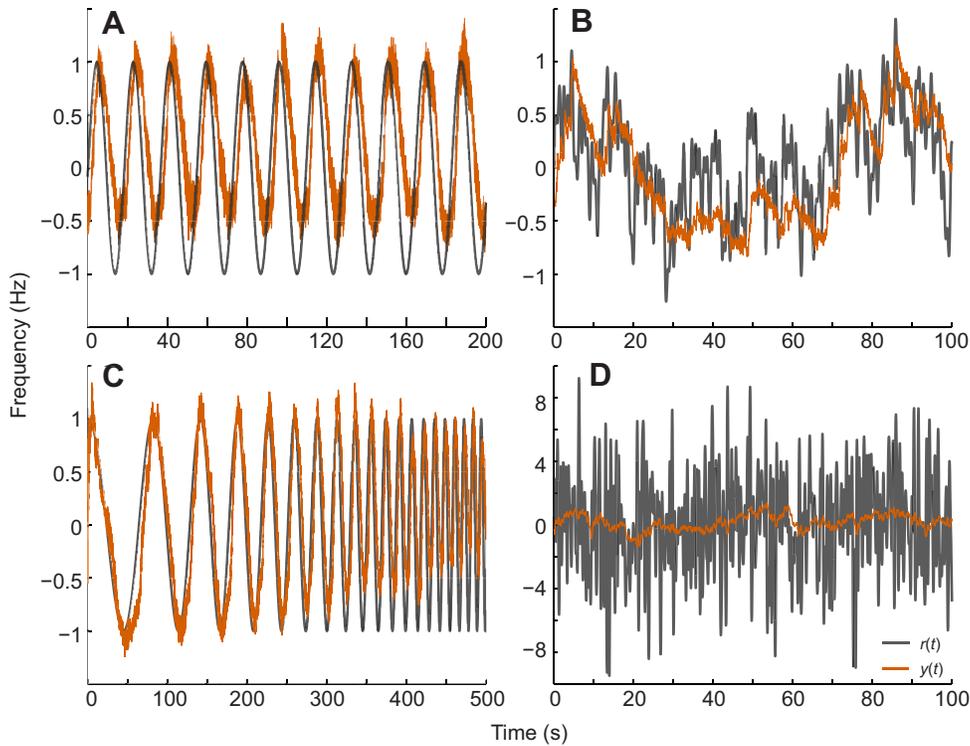


Fig. 3. Stimulus–response plots for closed-loop reference-tracking trials for different types of reference signals. (A) single sine, (B) sum of sines, (C) chirp and (D) band-limited noise in the 0–2 Hz range. The reference trajectory is gray, the response is orange. All these trials used a stimulus amplitude of $100 \mu\text{V cm}^{-1}$ and controller gain $\alpha=2$.

included $n=39$ static clamps with d_s from -50 to 50 at higher resolution closer to 0. The dynamic clamp trials had single sine, sum-of-sines and chirp stimuli as the reference trajectories.

Open-loop trials

The objective of these trials was to observe the response of the fish to a stimulus whose frequency trajectory $u(t)$ was predetermined, and not tied to $y(t)$. Two types of open-loop trials were performed. (i) Step inputs: initially, the fish was driven towards 0 Hz in closed loop for the balancing period of 100 s. The amplitude was linearly ramped up to $100 \mu\text{V cm}^{-1}$ for 50 s as described previously. At 100 s, the stimulus switched to open loop, and $u(t)$ was commanded to a fixed value for a further 100 s. We performed trials at steps with magnitudes from -5 to $+5$ Hz. (ii) Ramp inputs: the magnitude of u was ramped down from $+30$ to -30 Hz or up from -30 to $+30$ Hz in 300 s at a constant rate of change of 0.2 Hz s^{-1} . The stimulus amplitude was ramped to $100 \mu\text{V cm}^{-1}$ in the first 10 s. The initial difference frequency was large, $|df| \approx 30 \text{ Hz}$, well outside the typical range of the JAR.

Data analysis

All data analysis was carried out using scripts custom-written in MATLAB (The MathWorks Inc., Natick, MA, USA). For each trial, we recorded the reference, EOD, F2V output, and output waveform, sampled at 10 kHz. The voltage signal from the F2V was scaled and offset to convert it into frequency. Extremely rapid, transient changes in frequency (commonly caused by fish movement) were eliminated. The known baseline frequency of the fish was subtracted from all signals, so that a frequency of 0 Hz represents the fish's pre-stimulus (baseline) signal. The processed F2V signal was then subsampled to 100 Hz, and used as the output signal $y(t)$. The input for analysis depended on the type of trial: $r(t)$ was used for reference-tracking trials, $d(t)$ was used for the dynamic clamp trials, and $u(t)$ was used for the open-loop trials. These inputs were pre-generated trajectories, as mentioned previously.

Estimating FRFs for sinusoidal inputs

The frequency domain representations of the input and output signals were calculated using a fast Fourier transform (FFT) and peaks corresponding to the known number of frequency components in the input were determined (1 for single sines, 10 for sums of sines). The frequency response at a particular frequency was calculated as the ratio of the Fourier transform of the output to the input of the signal at that particular frequency. Thus, we measured eight data points from eight single-sine trials and 40 points from four sums-of-sines trials. Each data point was represented as a phasor, namely a number in the complex plane. The gain (distance of the phasor to the origin) and phase (angle of the phasor from the positive real axis) for all data were computed.

Estimating FRFs for chirp inputs

The input and output signals were filtered, subsampled and transformed *via* FFT as described above. The data points in this case were the input–output ratios of all the frequency components in the chirp frequency range. Thus, with a single trial, we obtained many data points, but each individual data point is somewhat more susceptible to noise. The data points were binned and averaged at 10 bins per decade of frequency, giving us 20 data points per chirp trial and 30 data points per long chirp trial.

White noise

White noise stimuli were used to evaluate the range over which the behavior is linear. This was done by evaluating the difference of the square root of the response–response coherence [$\sqrt{(RR)}$] and the stimulus–response coherence (SR) (Roddey et al., 2000). SR is the coherence between each stimulus and its corresponding response. $\sqrt{(RR)}$ is the square root of the coherence between two responses to the repeated presentation of the same stimulus. In our experiment, as there were three replicates of each frequency band, we obtained three data points per frequency per individual for SR as well as $\sqrt{(RR)}$.

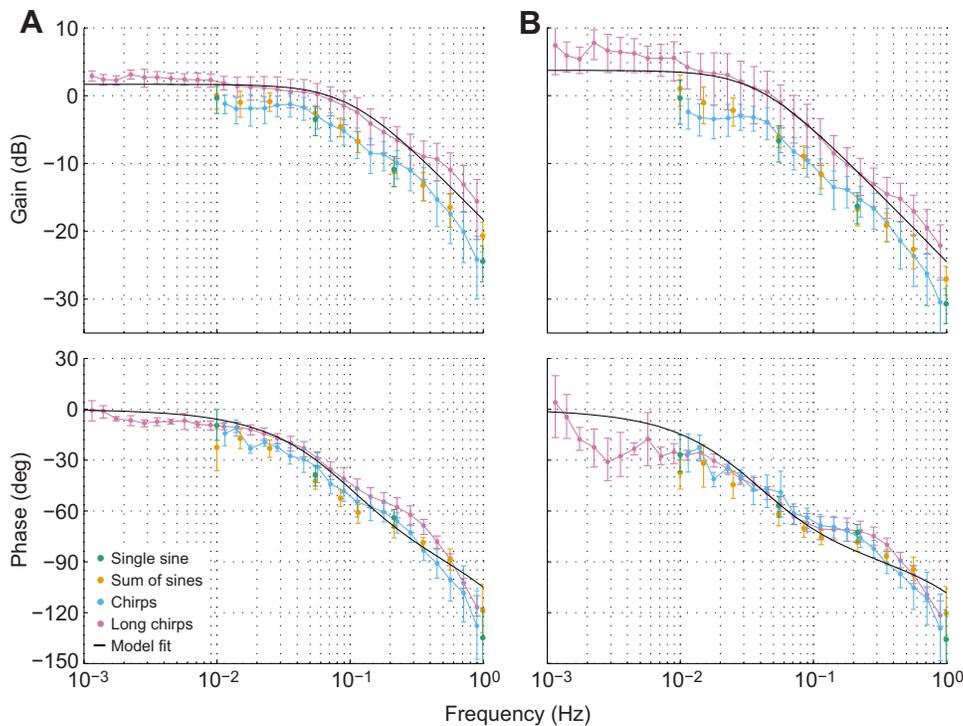


Fig. 4. Bode diagrams of experimental frequency response functions (FRFs) from single sine, sum of sines, chirp and long chirp trials. In a Bode diagram, the gain of each FRF point, $z=a+ib$, is expressed in decibels [$20\log_{10}(|z|)$, top] and its phase in degrees [$(180/\pi)\angle z$, bottom]. The single sine and sum of sines data points are slightly offset horizontally to avoid overlapping. (A) Closed-loop FRF. (B) The same data, mapped to open-loop via Eqn 5. The model was fitted to the open-loop FRF and mapped back to closed loop via Eqn 3.

Steps and ramps

These trials were open loop and were used to compare the fish response with the model response in the time domain. Hence, no frequency domain analysis was done on these trials.

RESULTS

The JAR is approximately linear around the unstable equilibrium

The FRF for single sine, sum of sines, chirp and long chirp closed-loop tracking trials were plotted in a Bode diagram as shown in Fig. 4A. The data were converted to open loop according to Eqn 4 as shown in Fig. 4B. The agreement between single sine, sum of sines and chirp data indicates that the behavior can be modeled approximately as a linear system in the neighborhood of the equilibrium, i.e. near the baseline frequency $f_1(0)$. There are small differences in the closed-loop data among different trial types, particularly between long chirps and the others. Sensitivities of opening the loop tend to amplify these, creating significant differences in the open-loop data. The modest difference in closed-loop responses could come from a variety of sources, including nonlinearities and time dependencies in the behavior.

SR and \sqrt{RR} coherences from white noise trials are shown in Fig. 5A. \sqrt{RR} represents the maximum theoretically possible coherence for a linear system in the presence of additive noise. The difference:

$$e = \sqrt{RR} - SR, \quad (6)$$

is an indicator of the nonlinearity of the system (Roddey et al., 2000). This difference (Eqn 6, Fig. 5B) does not exceed 0.4 over the white noise stimulation range of 0 to 20 Hz. However, past approximately 6 Hz, \sqrt{RR} drops to around 0.4. In that range and beyond, even the best nonlinear model will only be able to capture a fraction of the behavior. However, the frequency range over which we performed frequency response analysis and modeling (0.01–1 Hz) is far below this, and would likely encompass most naturally occurring frequency modulations among conspecifics.

As an important control, we investigated the sensitivity of the open-loop dynamics to experimental parameters. Amplitudes were compared at both half and twice the value of $100 \mu\text{V cm}^{-1}$ used in all of the other experiments described in this paper. Gains were tested at values of 1.5, 2.5, 3 and 4 compared with the value of 2 used in all other experiments. There was little-to-no effect of changing either the amplitude of the signal (Fig. 6A) or the feedback gain (Fig. 6B) on the open-loop frequency response. The amplitude insensitivity results in an experimental advantage, as the size of the fish or the specific placement of the fish between the stimulus electrodes would have little effect on the dynamics.

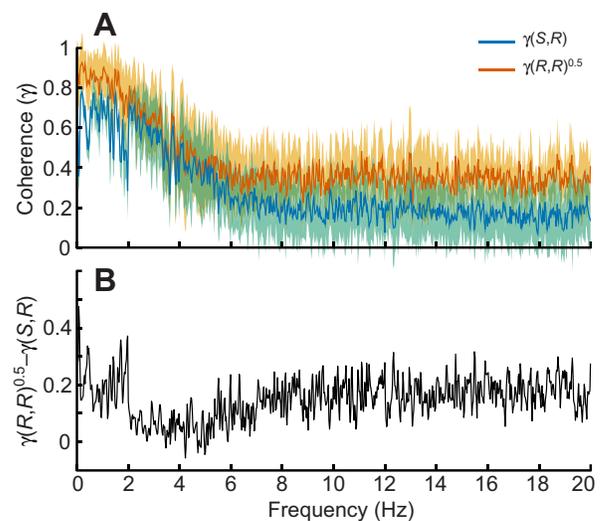


Fig. 5. Noise coherence plots. (A) Stimulus–response coherence SR and square root of response–response coherence \sqrt{RR} , as calculated from band-limited white noise trials over all individuals. The solid lines indicate the mean value for each frequency and the lighter bands show one standard deviation above and below the mean. (B) The difference between the coherences is an indicator of the linearity of the system (Roddey et al., 2000).

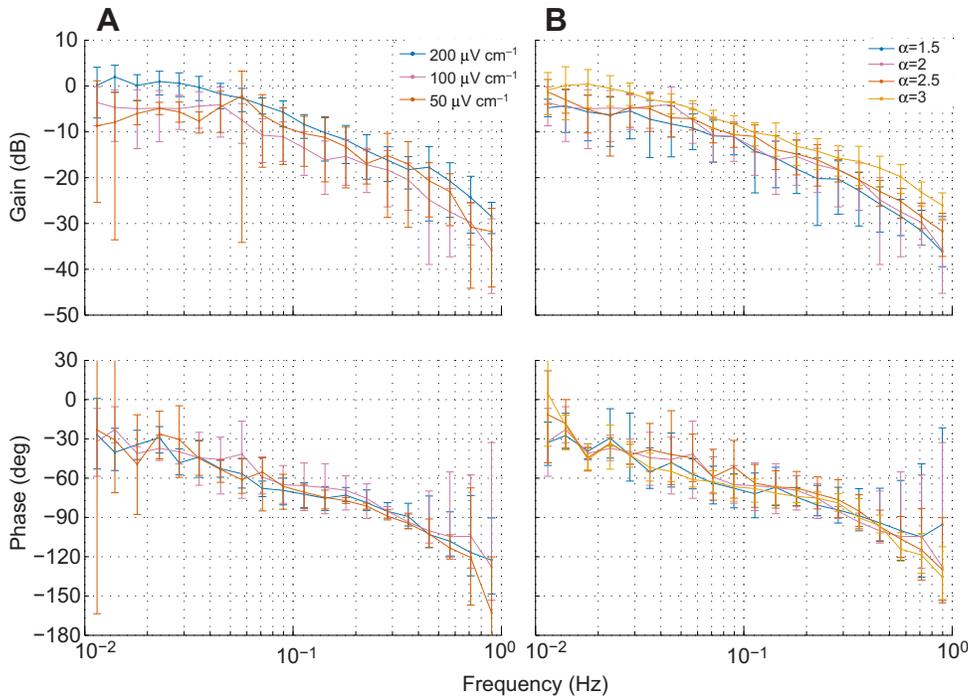


Fig. 6. Sensitivity to experimental parameters. (A) Open-loop frequency responses to chirp stimuli, for stimulus amplitudes of 50 and 200 $\mu\text{V cm}^{-1}$ along with the usual experimental value of 100 $\mu\text{V cm}^{-1}$. The stimulus amplitudes were measured across a 1 cm dipole placed adjacent to the head of the fish. (B) Open-loop responses to the feedback gain α being set to 1.5, 2.5 and 3 as opposed to the usual experimental value of 2. Changing the feedback gain causes divergence in the closed-loop response as expected; however, the computed open-loop responses do not appear to be sensitive to stimulus amplitude or feedback gain. In both Bode plots, the top plot is the gain response and the bottom plot is the phase response.

Determining the local linear model

To fit a model to the JAR behavior at the equilibrium, each of the closed-loop data points was transformed, in the complex plane, to the corresponding open-loop data point using Eqn 4 (see Fig. 4B). A model was fitted to the open-loop response; various model structures were tested and the best fit (see Appendix, Table A1) was obtained from a simple first-order model with delay:

$$G(s) = \frac{k}{s-p} e^{-sT}, \quad (7)$$

where $k=0.38$ Hz, $p=-0.24$ Hz and $T=57$ ms. Recall that $G(s)$ represents the transfer function from the df , $d(t)$, to the fish's EOD frequency, $y(t)$ (both signals were baseline subtracted). The negative real pole ($p < 0$) confirms that this is a stable system. The delay for the JAR behavior has not been previously reported, but our estimated value of $T=57$ ms lies within the range of sensorimotor delays reported for other animals, sensory modalities and motor behaviors (Franklin and Wolpert, 2011). The transfer function is plotted as the black curve in Fig. 4B. The behavior transfer function $J(s)$ (computed *via* Eqn 4) is unstable if there are zeros of the denominator $G(s)-1$ in the open right-half plane. Similarly, the closed-loop transfer function $H(s)$ (computed *via* Eqn 3) is stable if there are no zeros of $1-\alpha J(s)$ in the open right-half plane. Based on the Nyquist stability criterion (Astrom and Murray, 2008), it is easy to confirm that $J(s)$ is unstable (there is exactly one pole in the right-half plane). Furthermore, it is trivial to show that $H(s)$ is stable so long as α is larger than α_{\min} :

$$\alpha_{\min} = \frac{p}{k} + 1. \quad (8)$$

For the model parameters in Eqn 7, the minimum stable gain is $\alpha_{\min}=0.35$. The experimental value of $\alpha=2$ is much higher than this threshold, thus robustly stabilizing the closed-loop system.

Determining the global nonlinear model

The model fit using the closed-loop experiment is, effectively, a local linearization of the behavior about the unstable equilibrium $y=0$ (corresponding to the fish's pre-stimulus baseline frequency).

Thus, the transfer function described in Eqn 7 is unable to reproduce the full extent of a naturalistic response, which, in addition to avoiding a jamming frequency, involves achieving a steady-state frequency at a higher final df magnitude. Using the linear model as a starting point, and known features of the JAR neural circuit, we fitted a nonlinear model as described below.

The model structure we propose parses JARs into competing sensory and motor components. The sensory component captures the primary functional computation in the JAR circuit, namely an 'escape' term, $e(d)$, that depends on the df , $d=y-u$. The delay in the linear model fit is lumped into the computation of the df , $d=(t-T)$, which is in turn passed into the escape term, namely $e[d(t-T)]$. The motor component captures the known tendency of the pacemaker nucleus to 'return' to the pre-stimulus (baseline) EOD frequency upon removing a transiently applied jamming stimulus. The combination can be thought of as a leaky nonlinear integrator. When stimulated by a jamming signal, this model settles down to an equilibrium frequency where the sensory and motor components are equal but opposite. By further assuming that the strength of the motor return depends linearly on the deviation from baseline, we arrive at the following model structure:

$$\tau \dot{y}(t) = -y(t) + e(d(t-T)). \quad (9)$$

In this equation, τ is the characteristic time constant of the return to baseline, $-y(t)$ represents the return to baseline, and $e[d(t-T)]$ represents the repulsive escape. In the time domain, Eqn 9 is expressed as the following differential equation:

$$\dot{y}(t) = py(t) + kd(t-T), \quad (10)$$

with $p=-0.24$ Hz and $k=0.38$ Hz. Upon linearizing the nonlinear dynamics in Eqn 9 around the baseline, $d=0$, $y=0$, we compare terms with the linear model in Eqn 10 and obtain the following parameters for the nonlinear model:

$$\begin{aligned} \tau &= -p^{-1} = 4.17 \text{ s}, \\ \left. \frac{\partial e}{\partial d} \right|_{d=0} &= \tau k = 1.58. \end{aligned} \quad (11)$$

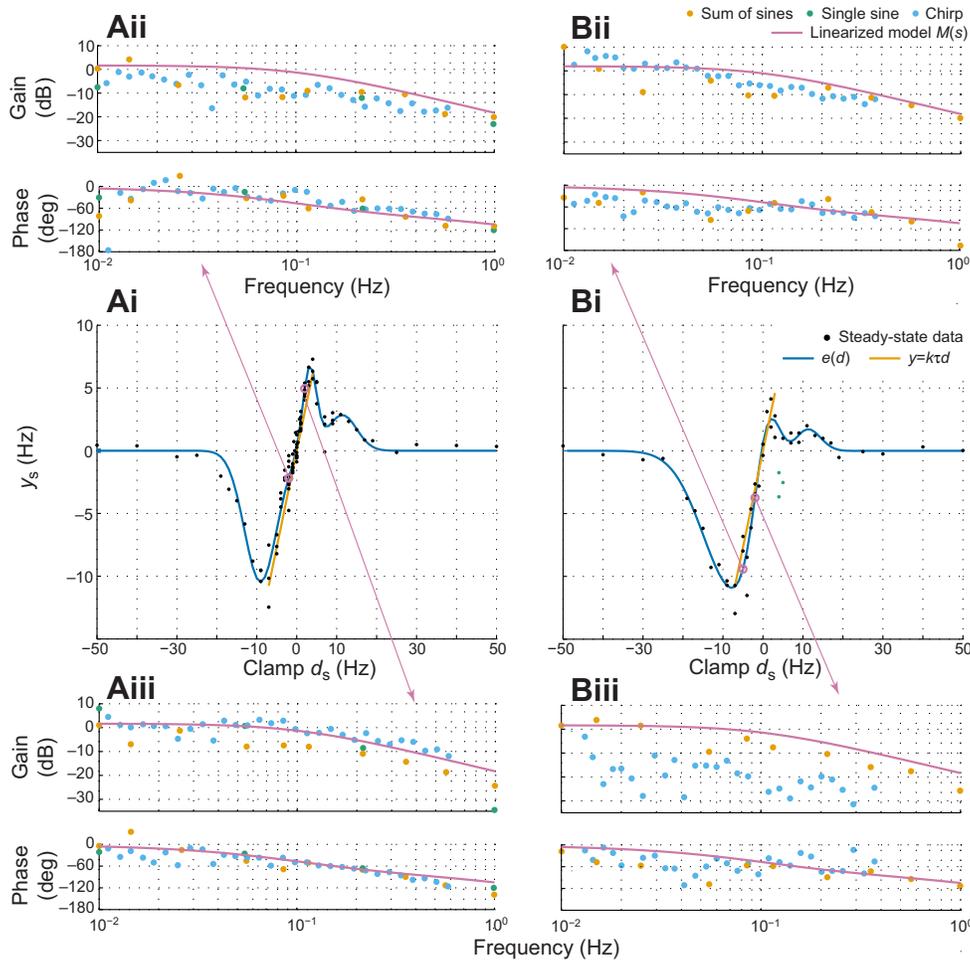


Fig. 7. Static and dynamic clamps. (Ai,Bi) The steady-state frequency, y_s (black dots), of two individuals vary nonlinearly as a function of clamped frequency difference, d_s . These data were used to fit distinct escape functions, $e(d)$ (blue curve), for each of the two distinct fish: (A) see Eqn 12 and (B) see Eqn 13. Green dots represent trials in which the fish frequency shifted in the direction of the input; these data points were removed from fitting. The linearization at origin obtained from reference-tracking trials $y=1.58d$ is shown (orange line) to compare with the slope of $e(d)$ at the origin. (Aii,iii, Bii,iii) Bode plots for the dynamic clamp trials for the selected clamp values $d=-2.0$ (Aii) and $d=2.0$ (Aiii) for the first individual, and $d=-5.0$ (Bii) and $d=-2.0$ (Biii) for the second individual. On each Bode plot, the response of the linearized model (Eqn 10) is also shown (pink curve). The dynamic clamps were not very robust, so we show the best pair of frequency data that we were able to obtain for each individual. This is evident e.g. in Bii, where the sum-of-sines response is well predicted by the model but the chirp responses have a lower gain response.

The value of the time constant τ is in the range of JAR return time constants (approximately 3–12 s) previously reported (Hitschfeld et al., 2009). To recover the function $e(d)$, note that if d is kept constant at d_s , Eqn 9 predicts that y should settle down to the equilibrium corresponding to $\dot{y}=0$, which is $y_s=e(d_s)$. Hence, the steady-state values of the clamp trials determine the escape function. We can use both the final values of the static clamp trials and values from the dynamic clamp trials at the end of the static clamp period, before the reference trajectory begins.

Fig. 7A,B shows the steady-state values of both static and dynamic clamp trials as a function of the clamp d , for two individuals. For the purpose of generating theoretical predictions from our model, we characterized the details of each of these two individuals' escape curves using the a sum of three Gaussians (Matlab curve-fitting toolbox):

$$e(d) = 2.85e^{-\left(\frac{d-11.39}{5.12}\right)^2} + 6.71e^{-\left(\frac{d-3.26}{2.40}\right)^2} - 10.45e^{-\left(\frac{d+9.05}{5.68}\right)^2}, \quad (12)$$

$$e(d) = 2.55e^{-\left(\frac{d-10.43}{4.98}\right)^2} + 8.70e^{-\left(\frac{d-1.08}{4.53}\right)^2} - 11.18e^{-\left(\frac{d+6.91}{11.23}\right)^2}. \quad (13)$$

The units for all numerical values in the above expression are s^{-1} . While the Gaussian mixture approximation provided a good fit for $e(d)$, no mechanistic insights can be drawn from its specific form, which is somewhat arbitrary.

The (dimensionless) slopes at the origin for the two curves are remarkably similar (1.68 and 1.73). Both slopes agree well with

the predicted value of 1.58 (Fig. 7, orange line), particularly given that the linear system identification data used to predict the escape function slope of 1.58 was categorically different from the clamp experiments, fitting and analytical differentiation used to determine each individual's slope. Note that three trials (Fig. 7, green points), for which the responses were qualitatively different from the other responses, were removed from fitting; including those trials biased the curve toward these outliers, and away from the typical behavior.

As shown in Fig. 7A,B and Eqns 12,13, the escape curves reflect variability between individuals despite the fact that the linearization at baseline is consistent.

Dynamic clamps validate nonlinear model

As a first validation of the nonlinear model, we examined the dynamic clamp trials, whose FRF should, in theory, match the prediction of the linearization of Eqn 9 at (d_s, y_s) . We examined this prediction for two fish (see Fig. 7).

For our prediction, we approximated the nonlinear function $e(d)$ as a straight line through the origin, namely $y_s=tkd_s$. With this approximation, the FRFs should match the prediction from Eqn 10. The FRFs in Fig. 7Aii,iii and Bii,iii show the frequency responses to dynamic clamps at two points along $e(d)$, marked in Fig. 7Ai and Bi. The response of the linearized model is shown in all four plots and is in good agreement with the frequency responses from the clamp trials. The data shown in Fig. 7 were selected to most clearly illustrate the prediction, but among the experimental paradigms we used, clamp trials were the least robust and repeatable.

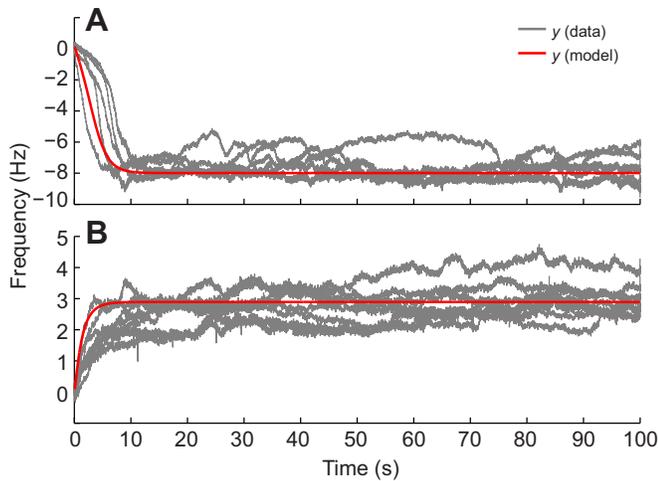


Fig. 8. Step responses. (A) Responses of one individual to open loop steps of $u=4$ Hz (gray) with the model response to the same stimulus (red). (B) Responses of the same individual to steps of $u=-3$ Hz (gray) with the model response to the same stimulus (red).

Open-loop steps validate nonlinear model

For the same two individuals for which the escape functions were fitted, we simulated the model in Eqn 9 in response to (open-loop) step inputs. The simulated step responses matched the data from the open-loop step trials remarkably well, particularly given the qualitatively different nature of open-loop step experiments compared with the experiments used to furnish the nonlinear model. Two examples, showing all the data from one individual for steps $u=+4$ Hz and $u=-3$ Hz, along with the model response, are shown in Fig. 8.

Snap-through bifurcation predicted for ramp stimuli

The model predicts that if the external stimulus frequency were slowly increased or decreased from outside the frequency range typically thought to elicit the JAR, then it should drive the fish's frequency up or down, respectively, until the fish frequency 'snaps over' to the other side of the stimulus and then moves away from the stimulus in the opposite direction. Using the open-loop ramp

trials (see Materials and methods), we tested this prediction of the model.

We now describe the model's snap-through prediction in more detail. Assuming $u(t)=\text{constant}$, the solution(s) to $y(t)=e[u-y(t)]$ are the equilibria of the system, i.e. the intersections of functions y and $e(u-y)$. The stability of the equilibrium depends on the derivative of the vector field on the right-hand side of Eqn 9 with respect to y : if the derivative is negative, positive or zero, the equilibrium is stable, unstable or a saddle-node, respectively. The subplots along the top in Fig. 9 decompose the vector field at different values of u that produce qualitatively different nonlinear dynamics in the sense that the number and type of equilibria change. Assuming u changes sufficiently slowly, the dynamics should track the nearest stable equilibrium.

The number and type of equilibria depend explicitly on the structure of the escape function $e(d)$ (where, recall, $d=u-y$). For the specific fit of $e(d)$ in Fig. 7A we tested the predictions of the nonlinear model. For this escape function the dynamics are punctuated by two bifurcation points at critical levels of the stimulus input, namely $u\approx-2.21$ and $u\approx 3.71$. For constant stimuli below -2.21 there is one equilibrium (Fig. 9A). The single equilibrium (green dashed line) here is stable, as \dot{y} crosses from positive to negative at the equilibrium. At the critical level $u\approx-2.21$, an additional equilibrium is introduced, namely a saddle-node (Fig. 9B). For $-2.21 < u < 3.71$, one stable and one unstable equilibrium branch out of the saddle-node, while the original stable equilibrium persists; thus, in this region there are three equilibria (Fig. 9C). But, at $u=3.71$ the unstable branch intersects with the original stable branch creating a saddle-node (Fig. 9D), which vanishes for $u > 3.71$, leaving just one lower stable equilibrium (Fig. 9E). The locus of equilibria at each u is shown in Fig. 9F (green curve); note that the middle branch of this inverted S-shaped curve is the unstable branch.

We simulated the dynamics for increasing and decreasing ramps between -30 and $+30$ Hz. We started the simulation with an initial condition of $y(0)=0$, which was near the only equilibrium for the initial values of $u=\pm 30$ Hz. The red curve in Fig. 9 shows that, indeed, at each time t , $y(t)$ remains near the closest stable equilibrium point until the output $y(t)$ snaps through to the other branch. This occurs when the input $u(t)$ reaches a saddle-node bifurcation. In the case of the increasing ramp from -30 , both the simulated (red)

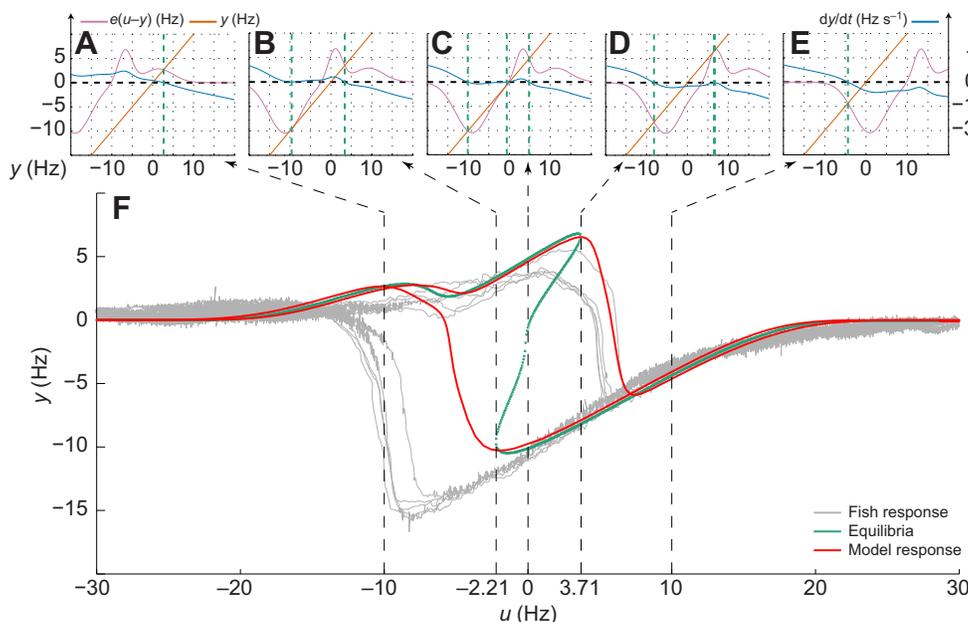


Fig. 9. Snap-through bifurcation of the dynamics $y(t)=e^{-1}\{e[y(t-T)-u(t-T)]-y(t)\}$. (A–E) The right-hand side of the dynamics (blue, right axis) is the scaled difference between two components, $e(y-u)$ (pink, left axis) and y (orange, left axis). The intersection points of y and $e(y-u)$ comprise the equilibrium frequencies (green dashed lines), corresponding to $dy/dt=0$. The graphs in A–E correspond to five distinct values of u , indicated by the black dashed lines originating in F. (F) As the stimulus $u(t)$ was ramped up or down from ± 30 Hz, the fish response (gray curves) initially followed the nearest stable equilibrium. For frequencies between approximately -2.21 and 3.71 there are three possible equilibria (green curve); the center branch is unstable. When the input reached a bifurcation frequency, the output snapped through to the only remaining equilibrium, as predicted by the model output (red curve). The general hysteretic structure of the responses is well captured by the model.

and actual (gray, multiple trials) $y(t)$ track the stable equilibrium on the upper branch of the green curve in Fig. 9F, even as the two additional equilibria are introduced at the bifurcation point of $u \approx -2.21$. Eventually, the unstable equilibrium combines with the stable equilibrium that $y(t)$ is tracking, causing both of them to vanish after forming a saddle–node. At that point, the closest (and only) stable equilibrium is the continuation of the lower branch, which causes $y(t)$ to snap through, ultimately converging to the remaining equilibrium. Note that the sign of the df for a given stimulus frequency depends on whether the EOD frequency was initially above or below the stimulus frequency. The fact that the JAR circuitry is driven by the df (including its sign), not merely the stimulus frequency, manifests itself as the hysteresis loop shown in Fig. 9. Additionally, for the decreasing ramps, the snap-through for the actual data are delayed relative to the model. The overshoot of the model is indicative that either $e(d)$ was underestimated in $d > 0$ or there is a velocity dependence on d not captured by our model.

DISCUSSION

The JAR is a behavior in several species of weakly electric fish that allows an individual to shift its EOD frequency away from an interfering conspecific, whose frequency differs from its own by a low (but non-zero) value. The neural computation of the JAR is conceptually simple yet mechanistically complex: the fish achieve the JAR without an internal reference to their own EOD frequency (Heiligenberg et al., 1978; Bullock et al., 1972) and instead integrate information from receptive fields across the body surface, ultimately evaluating the single parameter df to raise or lower the EOD frequency so as to increase the magnitude of df (Heiligenberg, 1991).

Despite the mechanistic complexity of the JAR, our goal was to capture its conceptual simplicity and express it as a low-order dynamical system. To achieve this goal, we stabilized the JAR using a computer-controlled feedback system. This stabilization allowed us to reduce the complete computational algorithm of the JAR into a simple parsimonious model comprising a delay, a sensory escape function and a motor return (see Fig. 10).

Given the structure of the model described above, comparatively simple experimental measurements can now be used to inform the parameters of a model that in turn predicts the responses to novel naturalistic or artificial stimuli. For example, this model captures a snap-through bifurcation that was not previously appreciated. Further, this model can be used to simulate social interactions between multiple individuals, without considering the mechanistic details of the internal dynamics underlying each individual's response.

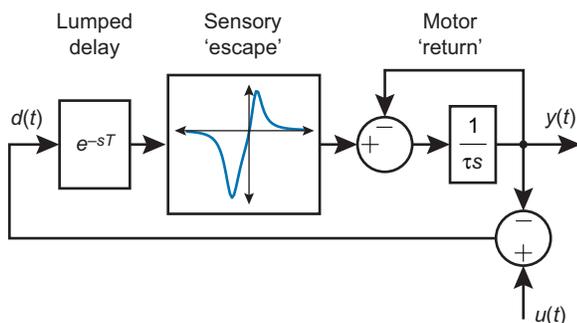


Fig. 10. A parsimonious nonlinear model for the JAR. The motor output of the EOD is represented in terms of the frequency of the pacemaker, $y(t)$. As is well known, the nervous system extracts the df , $d(t) = y(t) - u(t)$, where $u(t)$ is the frequency of an external stimulus (e.g. from a nearby conspecific). The sensorimotor transform includes a delay T , an sensory escape function $e(d)$ and autogenous motor return with time constant τ .

Local and global modeling

A stimulus whose frequency perfectly matches that of the fish's own steady-state EOD would not elicit a JAR response; there would be no amplitude modulations or changes in the timings of zero crossings, as is necessary to drive the JAR circuit. This situation is an unstable equilibrium, because any deviations of either the stimulus or response frequency would cause the fish to shift its frequency away from it. The EOD frequency does not escape indefinitely, however, as it settles down into a new equilibrium that is a function of the applied stimulus frequency. The existence of additional stable equilibria reveals the inherent nonlinear nature of the behavior.

Further, it is known that there are two parallel motor pathways, one that shifts the EOD frequency up and the other down (Metzner, 1993). This implies that there could well be a significant non-linearities or even non-smoothness of the dynamics at the switching point between the two circuits, i.e. the baseline. Thus, it is imperative that we understand the local dynamics at baseline, in addition to capturing the nonlinear escape dynamics. Indeed, it is the global asymmetric shape of the nonlinear escape curve $e(d)$ – and not a local discontinuity – that captures the asymmetry that was described previously by Metzner. The local linear model implies that the acceleration and deceleration of the pacemaker signal initiated by these two parallel pathways are similar in magnitude around baseline, causing a 'smooth' transition between the circuits. The mechanism of the JAR is based on spatial integration and voting, and not all units agree on the direction of the df (Heiligenberg et al., 1978). Thus, as the stimulus transitions from negative to positive df (or *vice versa*), the balance in the differential recruitment of the 'up *versus* down' JAR circuits tips, in a manner that is, evidently, graded and linear.

Closed-loop stabilization of unstable behavior

The JAR is an escape response, which makes the examination of the behavior around the baseline challenging. It is the instability of the equilibrium at baseline that underlies this difficulty. However, the stabilization of an unstable system using feedback is a classical application of control theory. In the case of the fish, we stabilized the unstable equilibrium by setting up a closed-loop feedback system that used the error signal between the fish's own frequency and a predefined reference frequency trajectory as the basis for the feedback. This system allows us to 'dictate' the frequency of the fish. We showed experimentally (Fig. 2) and confirmed analytically after modeling (Eqn 8) that this feedback did indeed stabilize the open-loop system.

This artificially closed-loop system was systematically perturbed (Fig. 3), and the resulting FRF data were numerically transformed to open loop using the known feedback transformation (Eqn 4, Fig. 4B). This allowed us to identify a model structure and verify it on the experimentally obtained closed-loop FRF data (Fig. 4A). In this manner, by knowing the feedback parameters and the closed-loop response, we were able to fit a frequency domain model to an unstable behavior. This procedure is relevant and applicable to a wide range of unstable biological behaviors, especially robust escape behaviors.

Evaluating the JAR in relation to the model

In each of the fish that we tested (five for reference tracking and two for clamp), the linearization at baseline was found to be identical. This was an unexpected result, as jamming avoidance is possible with different slopes for $e(d)$ at the origin, as long as the slopes are negative. Assuming similar motor dynamics, this suggests that small excursions around baseline will be similar across individuals.

Further, the linearizations were also found to be independent of stimulus amplitude. Previous work on the JAR (Bullock et al., 1972) has clearly shown that stimulus amplitude affects the ‘strength’ of the JAR in the sense that the magnitude of the frequency excursion is positively correlated with the depth of modulation. This would suggest that, in view of our model, the effect of stimulus amplitude would manifest as a change in the ‘height’ of the escape curve (e.g. see Fig. 7) but that the slope at the origin remains unchanged. The consistency of the slope at the origin also means that the escape curve, which can be identified by clamped- df experiments, serves as a possible ‘signature’ for the JAR behavior. By quantifying the escape curve for two individuals, jamming interactions between them can be modeled and predicted.

Reducing the JAR to this curve also allows us to examine social aspects of the JAR without considering the mechanistic details underlying its dynamics. For instance, previous research, interpreted in the context of our model, suggests that the escape curves may vary across development and be dependent on sex (Kramer, 1987). This means that the distribution of frequencies in a social group might be tailored by the individuals over time by modification of their escape curves. Future work might include identifying a general model structure that works across individuals and across stimulus amplitudes so that a complete interaction between freely swimming individuals in the wild can be modeled.

Finally, the snap-through response shown in Fig. 9 was first reported as part of the work that led to the discovery of the JAR (Watanabe and Takeda, 1963) (Fig. 5), and was described as the “effect of ‘chasing’ the discharge frequency by changing the stimulus frequency towards the discharge frequency”. This sudden reversal of the direction of df can now be understood as a bifurcation in the dynamics of the nonlinear model described in this paper.

Refining and extending the model

The effect of stimulus amplitude on the dynamics has not been fully explored in this work. While we found no significant dependence on constant amplitude (see Fig. 6), the JAR certainly depends on changing amplitude as a changing amplitude is simply an AM. In a natural environment, the amplitude of EOD signals from nearby conspecifics will depend on the time-varying distance and orientation between the animals. If the goal is to model the interaction of groups of fish, it becomes necessary to quantify the effect of stimulus amplitude dynamics on the JAR.

In addition, the escape function may depend on higher order time derivatives of d , taking the form $e(d, \dot{d}, \ddot{d}, \dots)$. An asymmetry in this kind of velocity dependence may explain the difference between the model response and data in the case of the decreasing ramps. Experiments similar to the dynamic clamp trials described in this paper are required to fully understand the contribution of these higher order terms.

The ‘return’ term was chosen to be linear based on the fact that, after a temporary JAR stimulus is removed (that is, there is no longer an exogenous stimulus), the EOD frequency exhibits nearly exponential returns (Hitschfeld et al., 2009). The fact that exponential decays are exhibited by linear dynamics suggests that a linear model for the unstimulated return to baseline is a good approximation. That said, this paper makes no systematic attempt to discover nonlinearities in the return and upon more detailed studies, one may find that nonlinearities of the return are important and may vary between individuals much like the escape. If this were the case, the model would be as follows:

$$\dot{y} = -\rho(y) + e(d), \quad (14)$$

with $\rho(y)$ and $e(d)$ denoting return and escape, respectively. This would imply that the steady state frequencies are $y_s = \rho^{-1}[e(d_s)]$. Further behavioral or neural experiments will then be required to tease apart ρ and e , as this cannot be done with input–output behavior alone.

In their natural habitat, weakly electric fish are often found in groups of two or more individuals (Stamper et al., 2010; Tan et al., 2005). In these groups, the complex interaction of electric fields can give rise to modulations called envelopes (Stamper et al., 2013), through the relative movement between individuals (Yu et al., 2012), or the higher order interaction of their EODs (Stamper et al., 2012). Evidence of envelope processing has been revealed in the electrosensory system in weakly electric fish (Middleton et al., 2006; Longtin et al., 2008; McGillivray et al., 2012; Savard et al., 2011). In addition, *Eigenmannia* have a behavioral response to low-frequency ‘social envelopes’ (Stamper et al., 2012). Extending our JAR model not only to incorporate pair-wise differences between individuals but also to capture envelope responses would ultimately provide a powerful tool to predict and interpret complex social behavior in these fish.

This analysis can also be used to further explore the neural circuitry that controls the JAR. These behavioral dynamics are, of course, manifest in CNS circuits. Specifically, our model parses escape as a sensory-driven phenomenon. Based on this, we might expect to discover neural correlates of the escape component in electrosensory areas, such as the electrosensory lateral line lobe, midbrain torus semicircularis, and perhaps the nucleus electrosensorius. Indeed, it is well known that these areas encode and perform the necessary computations for the control of the JAR (Heiligenberg, 1991). In contrast, our model suggests that the return is mediated by motor processes. If so, then we would expect to discover neural correlates of the return in the properties of neurons in the pre-pacemaker nucleus or pacemaker nucleus. This is consistent with previous work in *Apteronotus* which examined how the baseline EOD frequency is controlled and modulated through activity in the pacemaker (Oestreich and Zakon, 2002). Indeed, our approach could be extended to examine longer term modulation of the motor system by introducing biases or offsets into the feedback controller.

APPENDIX

This section describes the modeling approach for the closed-loop reference tracking trials. Three of the four fish used in the reference frequency tracking trials were tested using long chirp stimuli whose reference frequency started at 0.001 Hz and increased to 1 Hz; the resulting FRF data binned into 30 frequency bands (see Materials and methods). Two chirp magnitudes (1 and 2 Hz) were tested for each individual for a total of six trials that were used for fitting.

The FRF of the closed-loop system was determined as described in Materials and methods. These data were converted to open loop as per Eqn 5. In open loop, each trial was represented by complex numbers for each frequency bin. At a given frequency ω , the ratio of output $\hat{Y}(j\omega)$ to its corresponding input $\hat{D}(j\omega)$ can be represented by its gain and phase components:

$$\hat{G} = \frac{\hat{Y}}{\hat{D}} = \frac{|\hat{Y}|}{|\hat{D}|} e^{i(\angle\hat{Y} - \angle\hat{D})} = |\hat{G}| e^{i\angle\hat{G}}. \quad (A1)$$

The logarithm of this complex number is:

$$\log(\hat{G}) = \log|\hat{G}| + \angle\hat{G} = \log\frac{|\hat{Y}|}{|\hat{D}|} + i(\angle\hat{Y} - \angle\hat{D}), \quad (A2)$$

which is also a complex number, but weights gains logarithmically and phases linearly. This is important as we are interested in the

ratio of magnitudes and the difference of phases. This is also analogous to computing error directly in the Bode plot, e.g. Figs 4, 6. Distances in this ‘log-complex plane’ thus provide a good measure of error for model fitting, as previously used by Roth and colleagues (Roth et al., 2012).

We also have to account for the fact that the open-loop response will have increased sensitivity to noise at some frequencies, corresponding to the equation for opening the loop (Eqn 4). The sensitivity of opening the loop can be computed in log-complex plane coordinates:

$$S(\omega) = \left| \frac{d \log G(\omega)}{d \log H(\omega)} \right| = \frac{1}{|H(\omega)(1 - 1/\alpha) - 1|}. \quad (\text{A3})$$

We use the inverse of this sensitivity to weight the error. The error between model response $M(\omega)$ and the system response $G(\omega)$ at a given frequency ω is then defined as follows:

$$E(\omega) = \left| \log \left(\frac{M(\omega)}{G(\omega)} \right) \right|^2 \left(|H(\omega)(1 - 1/\alpha) - 1| \right). \quad (\text{A4})$$

The total error between model M and response G is the sum of this error over all frequencies.

Determining model structure

The next step was to determine the model structure. A model-fitting criterion such as reduced χ^2 , or information criteria such as Akaike information criterion (AIC) or Bayesian information criterion (BIC) can be generally minimized to determine the model order. However, these criteria assume independent data points. As the chirp is a time-varying frequency stimulus, the response data are clearly covariant, and the relatively small number of trials (six) is insufficient to determine the covariance structure between 30 bins. So, we implemented a selection criterion decision technique that takes into account two factors: model fit and model consistency.

Model fit

Leave-one-out cross-validation was used as the technique to fit a model while penalizing over-fitting. This technique is asymptotically equivalent to AIC (Stone, 1977). For each model, the Matlab nonlinear optimization function `fminsearch` was run with the error function (Eqn A4) to fit parameters to data from five trials, leaving the sixth trial out. This optimization was initialized 100 times (using Matlab to generate pseudo-random initial parameter values), and the set of parameters with the lowest fit error among them was then compared with the left-out trial, using the same error function. The leave-one-out error determined for one trial was averaged over all six trials being left out, providing a combined leave-one-out error for the model structure (see Fig. A1, y-axis).

Model consistency

Even though a certain model might have low leave-one-out error, the parameters fitted during each of the six leave-one-out minimizations may vary significantly. This variance, which is a measure of the uncertainty of the parameters based on our data, needs to be penalized, as we desire to have a consistent model fit with all six trials. To measure model consistency, we calculate the maximum singular value of the $K \times 6$ matrix where each of the six columns contains an estimate of all K parameters, with one column for each fit (one for each of the six trials left out). The poles and zeros of each fit are sorted, and the mean of each parameter across trials is subtracted, before finding the maximum singular value.

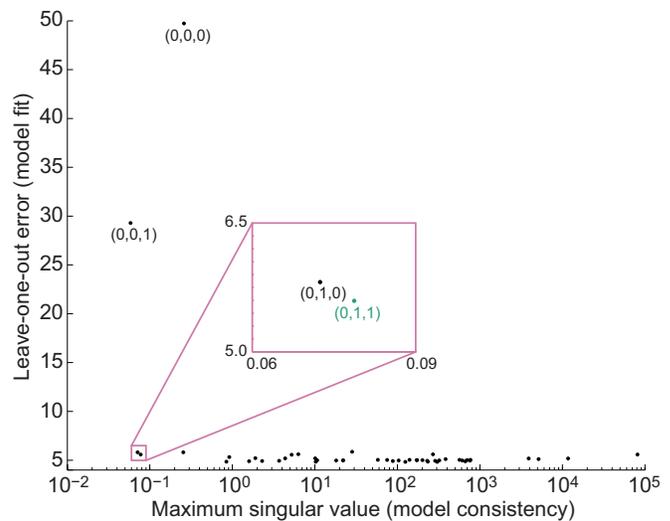


Fig. A1. Model fit versus model consistency. Each data point (black dot) represents a particular model structure (n_z, n_p, n_d). The average leave-one-out error for a particular model structure is on the y-axis. The maximum singular value of the $K \times 6$ parameter matrix for each model structure is shown on the x-axis. The first four models are labeled for clarity, and the region containing the two best models is magnified (inset). The model order (0,1,1) was ultimately chosen as the best compromise between fit and consistency (inset; green dot).

Altogether, 72 transfer function models of up to seventh order were tested. This includes all combinations $n_z \leq n_p \leq 7$ where n_z is the number of zeros and n_p is the number of poles. The delay was turned on ($n_d=1$) or off ($n_d=0$). The models are referred to as (n_z, n_p, n_d) , and correspond to the transfer function:

$$(n_z, n_p, n_d): G(s) = k \frac{a_0 + a_1 s + \dots + a_{n_z} s^{n_z}}{b_0 + b_1 s + \dots + b_{n_p} s^{n_p}} e^{-n_d s T}. \quad (\text{A5})$$

In this equation, the n_z zeros are the roots of the numerator polynomial, and the n_p poles are the roots of the denominator polynomial. The total number of parameters for a model is $K = n_z + n_p + n_d + 1$, where the extra parameter is due to the gain. The maximum K tested is then $7 + 7 + 1 + 1 = 16$. Fig. A1 shows leave-one-out error versus maximum singular value. For clarity, only models with maximum singular values below 10^5 are shown. It is clear from the figure that the two initial models (0,0,0), namely a pure gain, and (0,0,1), namely a gain with delay, both have unacceptably high errors. The next two models, both containing a gain and pole, provide low-error consistent fits, either without (0,1,0) or with (0,1,1) a delay.

The higher order models have lower error in some instances, but are much less consistent. As shown in the inset of Fig. A1, among the two ‘good’ models, we chose the model closest to the origin, (0,1,1), corresponding to a delayed first-order system.

Model fitting

After determining the model structure, the parameters for each model were ultimately estimated without cross-validation (i.e. using all available data). Optimization was performed 100 times for each model structure with random initial parameter guesses. The lowest error fit among these was designated the fit error, the parameters associated with which were used. For the two ‘good’ models (see Table A1), we inspected the histogram of all 100 final parameters and found that most initial parameter guesses converged to the same final parameter values. This suggests that, given the wide set of

Table A1. Fit parameters, fit error, leave-one-out error and maximum singular value for the first 12 model structures

Model	K	Gain	Zeros	Poles	Delay	Fit error	Leave-one-out error	Maximum singular value
(0,0,0)	1	0.45				295.17	49.73	0.258
(0,0,1)	2	0.45			0.49	172.34	29.30	0.058
(0,1,0)	2	0.37		-0.23		28.77	5.81	0.071
(0,1,1)*	3	0.38		-0.24	0.057	27.12	5.59	0.077
(1,1,0)	3	-0.021	16.14	-0.24		26.97	5.56	5.23
(1,1,1)	4	0.085	-3.65	-0.19	0.25	25.42	5.32	0.92
(0,2,0)	3	7.52		-0.25, -19.90		27.33	5.61	6.30
(0,2,1)	4	2.52×10 ⁴		-0.24, -6.72×10 ⁴	0.057	27.12	5.59	8.12×10 ⁴
(1,2,0)	4	0.39	-0.08	-0.04, -0.38		25.67	5.86	28.32
(1,2,1)	5	0.41	-0.16	-0.06, -0.54	0.08	23.03	5.22	1.90
(2,2,0)	5	-0.03	-0.13, 13.18	-0.06, -0.48		23.02	5.19	4.37
(2,2,1)	6	0.06	-0.09, -5.78	-0.04, -0.37	0.21	22.75	5.12	383.72

Poles correspond to the roots of the denominator polynomial, and zeros correspond to the roots of the numerator polynomial of Eqn A5. The 'best' models (0,1,0) and (0,1,1) are highlighted in bold. The (0,1,1) model (starred) and its parameters were used in Eqn 7.

initial guesses, our final estimate was not trapped in a local minimum. For clarity, we present the best parameters and errors of only the 10 lowest order models shown in Fig. A1 and Table A1. Fig. A2 compares experimental FRFs with the four lowest order

models. The (0,0,0) and (0,0,1) models are obviously poor fits to the data, whereas the two-parameter model (0,1,0) fits the data well except at the highest frequencies. The (0,1,1) model compensates for this by introducing a phase lag due to the delay. There is no discernible improvement in the frequency responses of models of higher order than (0,1,1), so they are not shown. The best-fit parameters for the (0,1,1) model are reported in Results (see Eqn 7).

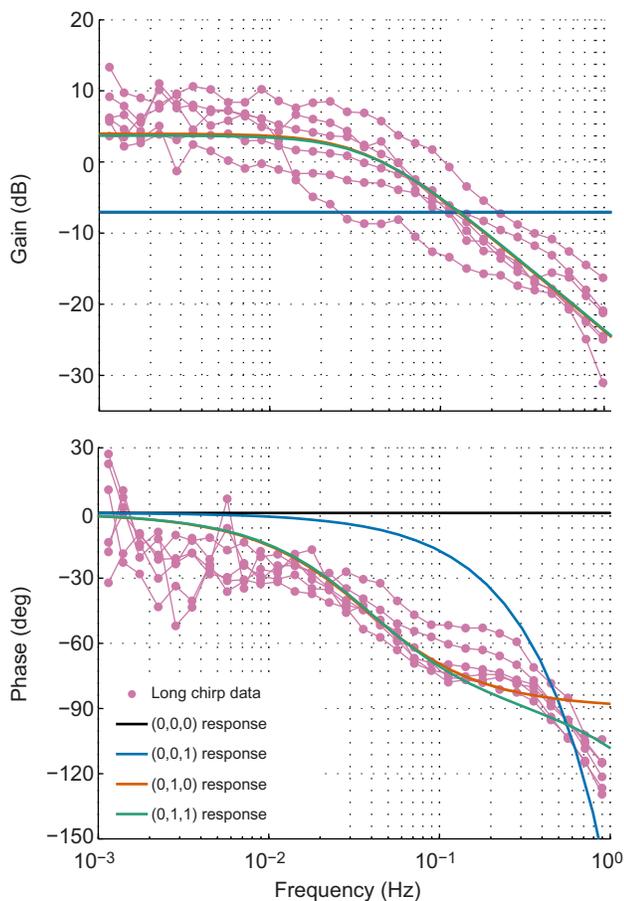


Fig. A2. Candidate model fits to open-loop data. The FRF data from six long-chirp trials transformed to open loop using Eqn 5 are shown, along with the best parameter fits for model structures (0,0,0), (0,0,1), (0,1,0) and (0,1,1). Models (0,0,0) and (0,0,1) have identical gains and thus the black curve is occluded by the blue curve in the gain plot. Models (0,1,0) and (0,1,1) differ slightly at the highest frequencies tested. Frequency responses of higher order models were not substantively different from the (0,1,1) model within this frequency range, and thus are not shown.

LIST OF SYMBOLS AND ABBREVIATIONS

AIC	Akaike information criterion
AM	amplitude modulation
BIC	Bayesian information criterion
$D(s)$	Laplace transform of difference signal
$d(t)$	difference signal (Hz)
df	difference frequency (Hz)
$e(d)$	escape function
EOD	electric organ discharge
f_1	EOD frequency (Hz)
f_2	stimulus frequency (Hz)
FRF	frequency response function
$G(s)$	behavior transfer function
$H(s)$	closed-loop transfer function
$J(s)$	open-loop transfer function
JAR	Jamming Avoidance Response
K	number of parameters in candidate model
k	transfer function gain (Hz)
$M(s)$	model transfer function
N	number of fish
n	number of trials
n_d	existence of delay (0/1) in candidate model
n_p	number of poles in candidate model
n_z	number of zeros in candidate model
p	pole of transfer function (Hz)
$R(s)$	Laplace transform of reference signal
$r(t)$	reference signal (Hz)
RR	response–response coherence
s	Laplace complex frequency variable
SR	stimulus–response coherence
T	delay of transfer function (s)
t	time (s)
$U(s)$	Laplace transform of input signal
$u(t)$	input signal (Hz)
$Y(s)$	Laplace transform of output signal
$y(t)$	output signal (Hz)
α	feedback gain
$\rho(y)$	return function
τ	time constant (s)
ω	frequency (rad s ⁻¹)

ACKNOWLEDGEMENTS

The authors would like to thank Eatai Roth for insightful discussions on system identification, and also Jon Dyrh and Sean Carver for discussions on developing the model fitting scheme.

AUTHOR CONTRIBUTIONS

M.S.M., S.A.S., E.S.F. and N.J.C. designed the experiments. M.S.M and S.A.S. constructed the experimental rig, conducted experiments and analyzed the data. M.S.M., S.A.S., E.S.F. and N.J.C. wrote the manuscript.

COMPETING INTERESTS

No competing interests declared.

FUNDING

This material is based upon work supported by the National Science Foundation (NSF) under grants IOS-0817918, CMMI-0941674 and CISE-0845749, and the Office of Naval Research under grant N000140910531.

REFERENCES

- Astrom, K. and Murray, R. (2008). *Feedback Systems: an Introduction for Scientists and Engineers*. Princeton, NJ: Princeton University Press.
- Bastian, J. (1987). Electrolocation in the presence of jamming signals: behavior. *J. Comp. Physiol. A* **161**, 811-824.
- Bastian, J. and Heiligenberg, W. (1980). Neural correlates of the jamming avoidance response of *Eigenmannia*. *J. Comp. Physiol.* **136**, 135-152.
- Bullock, T. (1969). Species differences in effect of electroreceptor input on electric organ pacemakers and other aspects of behavior in electric fish. *Brain Behav. Evol.* **2**, 85-101.
- Bullock, T. H., Behrend, K. and Heiligenberg, W. (1975). Comparison of the jamming avoidance responses in gymnotoid and gymnarichid electric fish: a case of convergent evolution of behavior and its sensory basis. *J. Comp. Physiol.* **103**, 97-121.
- Bullock, T. H., Hamstra, R. H. and Scheich, H. (1972). The jamming avoidance response of high frequency electric fish. *J. Comp. Physiol.* **77**, 1-22.
- Caputi, A. A. and Budelli, R. (2006). Peripheral electrosensory imaging by weakly electric fish. *J. Comp. Physiol. A* **192**, 587-600.
- Fortune, E. S. (2006). The decoding of electrosensory systems. *Curr. Opin. Neurobiol.* **16**, 474-480.
- Fortune, E. S. and Rose, G. J. (2003). Voltage-gated Na⁺ channels enhance the temporal filtering properties of electrosensory neurons in the torus. *J. Neurophysiol.* **90**, 924-929.
- Franklin, D. W. and Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron* **72**, 425-442.
- Heiligenberg, W. (1973). Electrolocation of objects in the electric fish *Eigenmannia* (Rhamphichthyidae, Gymnotoidei). *J. Comp. Physiol.* **87**, 137-164.
- Heiligenberg, W. (1991). *Neural Nets in Electric Fish*. Cambridge, MA: The MIT Press.
- Heiligenberg, W., Baker, C. and Matsubara, J. (1978). The jamming avoidance response in eigenmannia revisited: the structure of a neuronal democracy. *J. Comp. Physiol. A* **127**, 267-286.
- Heiligenberg, W. and Bastian, J. (1980). The control of *Eigenmannia's* pacemaker by distributed evaluation of electroreceptive afferences. *J. Comp. Physiol. A* **136**, 113-133.
- Hitschfeld, E. M., Stamper, S. A., Vonderschen, K., Fortune, E. S. and Chacron, M. J. (2009). Effects of restraint and immobilization on electrosensory behaviors of weakly electric fish. *ILAR J.* **50**, 361-372.
- Hopkins, C. D. (1974). Electric communication: functions in the social behavior of *eigenmannia virescens*. *Behaviour* **50**, 270-304.
- Hupé, G. J. and Lewis, J. E. (2008). Electrocommunication signals in free swimming brown ghost knifefish, *Apteronotus leptorhynchus*. *J. Exp. Biol.* **211**, 1657-1667.
- Kramer, B. (1987). The sexually dimorphic jamming avoidance response in the electric fish *Eigenmannia* (Teleostei, Gymnotiformes). *J. Exp. Biol.* **130**, 39-62.
- Longtin, A., Middleton, J. W., Cieniak, J. and Maler, L. (2008). Neural dynamics of envelope coding. *Math. Biosci.* **214**, 87-99.
- McGillivray, P., Vonderschen, K., Fortune, E. S. and Chacron, M. J. (2012). Parallel coding of first- and second-order stimulus attributes by midbrain electrosensory neurons. *J. Neurosci.* **32**, 5510-5524.
- Metzner, W. (1993). The jamming avoidance response in *Eigenmannia* is controlled by two separate motor pathways. *J. Neurosci.* **13**, 1862-1878.
- Middleton, J. W., Longtin, A., Benda, J. and Maler, L. (2006). The cellular basis for parallel neural transmission of a high-frequency stimulus and its low-frequency envelope. *Proc. Natl. Acad. Sci. USA* **103**, 14596-14601.
- Moortgat, K. T., Keller, C. H., Bullock, T. H. and Sejnowski, T. J. (1998). Submicrosecond pacemaker precision is behaviorally modulated: the gymnotiform electromotor pathway. *Proc. Natl. Acad. Sci. USA* **95**, 4684-4689.
- Oestreich, J. and Zakon, H. H. (2002). The long-term resetting of a brainstem pacemaker nucleus by synaptic input: a model for sensorimotor adaptation. *J. Neurosci.* **22**, 8287-8296.
- Roddey, J. C., Girish, B. and Miller, J. P. (2000). Assessing the performance of neural encoding models in the presence of noise. *J. Comput. Neurosci.* **8**, 95-112.
- Roth, E., Reiser, M. B., Dickinson, M. H. and Cowan, N. J. (2012). A task-level model for optomotor yaw regulation in *Drosophila melanogaster*: a frequency-domain system identification approach. In *Proceedings of the IEEE International Conference on Decision Control*, pp. 3721-3726.
- Savard, M., Krahe, R. and Chacron, M. J. (2011). Neural heterogeneities influence envelope and temporal coding at the sensory periphery. *Neuroscience* **172**, 270-284.
- Stamper, S. A., Carrera-G, E., Tan, E. W., Fugère, V., Krahe, R. and Fortune, E. S. (2010). Species differences in group size and electrosensory interference in weakly electric fishes: implications for electrosensory processing. *Behav. Brain Res.* **207**, 368-376.
- Stamper, S. A., Fortune, E. S. and Chacron, M. J. (2013). Perception and coding of envelopes in weakly electric fishes. *J. Exp. Biol.* **216**, 2393-2402.
- Stamper, S. A., Madhav, M. S., Cowan, N. J. and Fortune, E. S. (2012). Beyond the Jamming Avoidance Response: weakly electric fish respond to the envelope of social electrosensory signals. *J. Exp. Biol.* **215**, 4196-4207.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Stat. Soc. Series B Stat. Methodol.* **39**, 44-47.
- Tan, E. W., Nizar, J. M., Carrera-G, E. and Fortune, E. S. (2005). Electrosensory interference in naturally occurring aggregates of a species of weakly electric fish, *Eigenmannia virescens*. *Behav. Brain Res.* **164**, 83-92.
- Triefenbach, F. and Zakon, H. (2008). Changes in signalling during agonistic interactions between male weakly electric knifefish, *Apteronotus leptorhynchus*. *Anim. Behav.* **75**, 1263-1272.
- von der Emde, G. (2006). Non-visual environmental imaging and object detection through active electrolocation in weakly electric fish. *J. Comp. Physiol. A* **192**, 601-612.
- Watanabe, A. and Takeda, K. (1963). The change of discharge frequency by a.c. stimulus in a weak electric fish. *J. Exp. Biol.* **40**, 57-66.
- Yu, N., Hupé, G., Garfinkle, C., Lewis, J. E. and Longtin, A. (2012). Coding conspecific identity and motion in the electric sense. *PLoS Comput. Biol.* **8**, e1002564.
- Zakon, H. H. and Dunlap, K. D. (1999). Sex steroids and communication signals in electric fish: a tale of two species. *Brain Behav. Evol.* **54**, 61-69.
- Zakon, H., Oestreich, J., Tallarovic, S. and Triefenbach, F. (2002). EOD modulations of brown ghost electric fish: JARs, chirps, rises, and dips. *J. Physiol. Paris* **96**, 451-458.