SUPPLEMENTARY INFORMATION

Text

Computational simulation

1 System model

The only state that the Kalman filter concerns is the produced call frequency of the bat. Following Shadmehr and Mussa-Ivaldi (1), the Kalman filter maintains a sliding window of the last \( N + 1 \) call frequency values to deal with the delayed sensory feedback. The uncertainties and correlation of the states are represented using a joint normal distribution.

The state vector could be written as

\[
x_k = [x_{k-N} \ldots x_{k-2} x_{k-1} x_k]^T, \quad P_k
\]

where \( x_{k-N}, \ldots, x_{k-2}, x_{k-1}, x_k \) are frequencies of the last \( N + 1 \) history calls, and \( P_k \) is the covariance matrix of the states.

When the bat produces a new call, the frequency prediction of the call follows

\[
x_k = x_{k-1} + u_k + w_k
\]

\[
w_k \sim \mathcal{N}(0, \sigma_{\text{pred}}^2)
\]

where \( x_{k-1} \) is the previous call frequency before the new call, \( u_k \) is the incremental frequency adjustment conducted by the vocal production system, \( w_k \) is the process noise of the prediction step including the uncertainty of the frequency adjustment. Here, we assumed the bat follows the simple principle of correcting the estimated frequency offset (error) from the previous call. Thus, the frequency adjustment follows

\[
u_k = -\hat{x}_{k-1}
\]

where \( \hat{x}_{k-1} \) is the estimated value of the previous call frequency.

The system model of the state vector follows

\[
x_k = F_k x_{k-1} + u_k + w_k
\]
where $F_k$ is the state transition matrix which is time-invariant in our case.

As is shown in the system model, the call frequency states in the sliding window are highly correlated, making it possible that delayed auditory feedback could help correct the estimation of the latest call frequency, thus adjusting the vocal production with low latency. The magnitude of the process noise $w_k$ determines how tightly the history call frequency states are correlated.

### 2 Observation model

The Kalman filter estimates and corrects its prediction of the system states through observations. There are two kinds of observations taken into consideration in our implementation, namely the somatosensory observation (feedback) and the auditory observation (feedback). Generally speaking, the auditory observation is more likely to be affected by the ambient environment or potential perturbations, while the somatosensory observation is more internal and may work more independently and stably.

#### 2.1 Auditory observation model

The auditory observation model is described as

$$z_{k, aud} = x_{k-\lambda_{aud}} + v_{aud,k-\lambda_{aud}}$$

$$v_{aud,k-d_{aud}} \sim \mathcal{N}(0, \sigma_{aud}^2)$$

where $z_{k,aud}$ is the auditory observation obtained at epoch $k$, $\lambda_{aud}$ is the delayed epochs of the auditory observation due to the time required for the auditory system to process the feedback information, $x_{k-\lambda_{aud}}$ is the call frequency of the bat at epoch $k - \lambda_{aud}$, $v_{aud,k-\lambda_{aud}}$ is the noise (uncertainty) of the auditory observation. The auditory observation noise is assumed to have a zero-mean normal distribution with a standard deviation of $\sigma_{aud}$.

Moreover, an exponential function is introduced to approach an increase in the auditory uncertainty with the progression of the repeated auditory feedback perturbations, as there
is a general trend for all *H. armiger* to decrease the size of frequency adjustments after a few perturbations. The function is described as

\[ \sigma_{aud} = \sigma_{aud,0} \cdot f^\tau(\Delta) \]

where \( \sigma_{aud,0} \) is the initial auditory observation noise level of the bat, \( f(\Delta) \) is an index about the auditory uncertainty corresponding to shift size \( \Delta \) of the perturbed call, \( \tau \) is the count of the perturbed calls.

### 2.2 Somatosensory observation model

The somatosensory observation model is described as

\[ z_{k,som} = x_{k-\lambda_{som}} + v_{som,k-\lambda_{som}} \]

\[ v_{som,k-\lambda_{som}} \sim \mathcal{N}(0, \sigma_{som}^2) \]

where \( z_{k,som} \) is the somatosensory observation obtained at epoch \( k \), \( \lambda_{som} \) is the delayed epochs of the somatosensory observation due to time required for the somatosensory system to process the feedback information, \( x_{k-\lambda_{som}} \) is the call frequency of the bat at epoch \( k-\lambda_{som} \), \( v_{som,k-\lambda_{som}} \) is the noise (uncertainty) of the somatosensory observation that has a zero-mean normal distribution with a standard deviation of \( \sigma_{som} \).

As the somatosensory system is presumed to work more independently and did not receive any perturbation in our case, the uncertainty level of the somatosensory observation is assumed to be time-invariant.

### 2.3 Combined observation model

Based on the two observation (sensory feedback) models, the combined observation model at epoch \( k \) is written as

\[ z_k = H_k x_k + v_k \]

\[
\begin{bmatrix}
  z_{k,aud} \\
  z_{k,som}
\end{bmatrix} =
\begin{bmatrix}
  0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\
  0 & \cdots & 0 & \cdots & 1 & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
  x_k - N \\
  x_k - 2 \\
  x_k - 1 \\
  x_k
\end{bmatrix} +
\begin{bmatrix}
  v_{aud,k-\lambda_{aud}} \\
  v_{som,k-\lambda_{som}}
\end{bmatrix}
\]
where $H_k$ is the observation matrix, which is related to the delays of the two sensory feedback models.

### 3 Details of the Kalman filter processing

The details of the Kalman filter processing are presented as follows, based on the system and observation models given above.

#### 3.1 Prediction

During the prediction step, the state vector and the state covariance are propagated following

$$
\dot{x}_{k|k-1} = F_k \dot{x}_{k-1} + u_k
$$

$$
P_{k|k-1} = F_k P_{k-1} F_k^T + Q_k
$$

where

$$
Q_k = E(w_k w_k^T) = \begin{bmatrix} 0 \\ \sigma_{pred}^2 \end{bmatrix}
$$

is the process noise matrix, which reflects the uncertainty of the state prediction.

#### 3.2 Update

During the update step, new observations obtained at the latest epoch $k$ are used to update the system states. The innovation is written as

$$
y_k = z_k - H_k \dot{x}_{k|k-1}
$$

The Kalman gain is calculated following

$$
K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1}
$$

where

$$
R_k = E(v_k v_k^T) = \begin{bmatrix} \sigma_{aud}^2 & \sigma_{som}^2 \\ \sigma_{som}^2 & \sigma_{som}^2 \end{bmatrix}
$$

is the observation noise matrix, which reflects the uncertainty of the observation.

Then the state vector and the covariance are updated following
\[ \ddot{x}_k = \dot{x}_{k|k-1} + K_k y_k \]
\[ P_k = (I - K_k H_k) P_{k|k-1} \]

### 3.3 Details about the sensory delay

Bats intermittently produce echolocation calls, i.e., short discrete calls (pulses) are separated by silence gaps. Whether a bat can use the sensory feedback from the previous call to guide the production of a next call depends on both the time window between two successive calls, i.e., the inter-pulse interval (IPI), and the sensory processing delays. Thus, the implementation of the sensory (auditory and somatosensory) delays (in epochs) within our model are affected by the actual IPIs of the individual *H. armiger* and the assumed sensory feedback latency, following

\[ \lambda_{\text{aud}} = \text{ceil}(t_{\text{aud}}/t_{\text{call}}) \]
\[ \lambda_{\text{som}} = \text{ceil}(t_{\text{som}}/t_{\text{call}}) \]

where \( t_{\text{aud}} \) and \( t_{\text{som}} \) are the latencies of the auditory/somatosensory systems respectively, \( t_{\text{call}} \) is the average IPI of the bat during the perturbation window for a given condition.

At every epoch, delayed observations of the history call frequencies would be obtained by the sensory system and used to correct the frequency prediction of the previous call, thus affecting the frequency adjustment of this epoch’s call production.

Because the exact delays for the somatosensory and auditory feedback systems of *H. armiger* are not known, we considered multiple combinations of the delays in our models. Based on the vocal response latency data measured for two other echolocating bat species (2, 3), we considered 60 ms to be a likely auditory feedback delay for *H. armiger*. Because the somatosensory feedback delay is suggested to be between 30~50 ms for humans (4) and the much smaller size of the bats compared to humans, we considered 30 ms to be a likely somatosensory feedback delay for *H. armiger*. Nevertheless, we additionally tested multiple combinations of the delays in our models, including the 0, 15, 30, and 45 ms somatosensory feedback delays in combination with a 60 ms auditory feedback delay, and the 0, 30, 60, and 90 ms auditory feedback delays in combination with the 30 ms somatosensory feedback delay. It is noted that *H. armiger* shortened the IPI for most up-
shifted feedback conditions during the perturbation window (Fig. 2C; Fig. S4). However, here we aimed to focus on the frequency parameter only, and the effects of auditory feedback perturbation on the temporal parameters of the calls will be examined in detail in the future. Hence, in the model, the auditory delay $d_{aud}$ is treated as a dynamic value based on the average IPIs of each perturbation condition of each bat. During the period that the bat shortens its call interval (typically, from 80 ms to 30 ms) as the perturbation begins, there would be several epochs (typically, 1 or 2) of call production when no new auditory observations are obtained. At these epochs, the previous auditory observation (observation of the first perturbed call) would be used to update the state estimation.

References

1. R. Shadmehr, S. Mussa-Ivaldi, *Biological Learning and Control: How the Brain Builds Representations, Predicts Events, and Makes Decisions* (MIT Press, 2012).
2. J. Luo, C. F. Moss, Echolocating bats rely on audiovocal feedback to adapt sonar signal design. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10978–10983 (2017).
3. C. Geberl, S. Brinklov, L. Wiegrebe, A. Surlykke, Fast sensory–motor reactions in echolocating bats to sudden changes during the final buzz and prey intercept. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4122–4127 (2015).
4. B. Parrell, V. Ramanarayanan, S. Nagarajan, J. Houde, The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology* 15, e1007321 (2019).
Fig. S1 Time-related frequency adjustment by two individual bats in response to auditory feedback of different shift sizes and directions.
Fig. S2 Effects of somatosensory feedback delay on the model performance. During the optimization, the auditory feedback delay was fixed at 60 ms.
Fig. S3 Effects of auditory feedback delay on the model performance. During the optimization, the somatosensory feedback delay was fixed at 30 ms.
Fig. S4 Inter-pulse interval (IPI) of individual bats in different auditory feedback perturbation conditions. Each figure panel shows the distribution of the IPIs for the pre-perturbation (green line), perturbation (red line), and post-perturbation (blue line) window, with a 10 ms bin size and based on the data from the first experiment. The shortening of IPIs occurred mostly in the perturbation window of the positive-shifted conditions (i.e., the 20 cents and 50 cents conditions).