Threshold-Free Neural Network Models for Swim Bout Detection of Larval Zebrafish

Haochen Wang[1], Grant Kroeschell[2], Yunlong Zhu[3], Jeff S. Mumm[2], Ji Yi[1 2]

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

[2]Department of Ophthalmology, Johns Hopkins University, Baltimore, MD, USA

[4]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, USA

Abstract

Accurate identification of swim bouts is essential for decoding motor responses in larval zebrafish, yet conventional threshold-based methods rely on subjective cutoffs and per-experiment tuning. To overcome these limitations, we developed threshold-free deep learning models for *post-hoc* data processing (offline model) and an online model optimized for real-time detection. Both models showed superior precision over threshold-based methods, especially for low-amplitude bouts. Using the offline model, we validated that our head-fixing protocol with closed-loop visual feedback largely preserves naturalistic swimming kinematics. We further showed that repeated trials of moving grating stimulus led to lower bout frequency and longer interbout intervals (IBI). By contrast, transparent pigmentation mutants (*nacre* and *casper*) showed similar bout frequencies and IBIs but shorter duration compared to *wildtype* (*AB*) larvae. Together, these findings highlight the versatility of our models, enabling reproducible, high-throughput, and biologically relevant analysis of zebrafish behavior across a wide range of experimental paradigms.

Introduction

The brain evaluates stimuli to generate and refine behavioral responses. Decoding this sensorimotor transformation is a central goal of neuroscience. The larval zebrafish (*Danio rerio*) is an established model organism for such studies, offering a vertebrate central nervous system amenable to high-throughput experiments and system-wide neural imaging, with findings that can inform translational biomedical research[1-8]. To investigate the neural circuitry governing sensorimotor behaviors, a common paradigm involves immobilizing the larval head region for neural activity recording under microscopes, while leaving the tail free to move for behavioral tracking in response to controlled sensory stimuli[9].

When analyzing motor outputs of larval zebrafish, a fundamental step is accurate identification of swim bouts, which are discrete, oscillatory tail movements separated by periods of quiescence[9, 10]. Conventionally, bouts are identified using either manual annotation or, more commonly, a threshold applied to kinematic parameters like tail angle[11-22]. While widely adopted within established analysis packages[14, 23-25], the threshold-based approach has an intrinsic limitation: reliance on an arbitrary user-defined cutoff. A low threshold risks misclassifying noise as a genuine bout, while a high threshold will systematically miss subtle, low-amplitude events. Even if the threshold is optimized, it still requires per-experiment tuning that can complicate reproducibility and even biological relevance of the data.

Here, we overcome this limitation by developing a threshold-free deep learning framework for bout detection, which is objective and robust across experimental setups. We trained a Convolutional Neural Network with a Bidirectional Long Short-Term Memory layer on a ground-truth dataset, in which bouts were manually annotated based on video recording. We first demonstrated that our threshold-free model maintains high accuracy across diverse signal regimes, including low-amplitude tail movements whose magnitudes were similar to that of background noise—conditions under which conventional threshold-based methods consistently fail. We then validated our closed-loop behavioral platform, showing that it preserves naturalistic visuomotor kinematics. Finally, by applying our robust detection model to data from the validated closed-loop platform, we revealed that repeated exposure to the same visual stimuli (trial effect) and genetic background (strain effect) have dissociable impacts on zebrafish larvae: prior trial experience modulated all bout kinematic parameters, whereas strain effects presented primarily in bout duration. This demonstrates the broad applicability of our approach, enabling researchers to address diverse questions and scopes extending beyond the experimental paradigms explored here.

Results

Behavioral Assays that Robustly Elicit Bouts for Automated Data Acquisition

The optomotor response (OMR) is a reflexive behavior, in which animals like insects[26], fish[27], and humans[28] move to stabilize their position relative to perceived optic flow. In zebrafish larvae, OMR behavioral assays using moving grating stimuli increase swim bout frequencies[27, 29]. We therefore leveraged this behavioral paradigm to efficiently produce a large collection of bouts for analysis.

We performed OMR assays for free-swimming 6-7 dpf larvae with the setup depicted in **Figure 1A**. A highspeed camera recorded kinematics at 200 Hz, and the captured frames were processed using open-source software Stytra[30]. Stytra reported the angle difference between the rostral and the caudal regions of the tail, which was recorded as *tail_sum* (θ). The software then detected bouts with its built-in, threshold-based detection method that we used as benchmark for detection performance. For visual stimulation, a custom PsychoPy[31] script projected moving red and black gratings—a robust stimulus for zebrafish[32, 33]—onto a silicon diffusion pad below the fish. Tracking and stimuli generation were synchronized as illustrated in **Figure S1**.

OMR assays for head-fixed larvae were performed as shown in **Figure 1B**. The optic components were secured to a 3D-printed housing (**Figure S2**). To mount larvae, we refined protocols for immobilizing the head in agarose[34-37], while leaving both the eyes and tail free to move (**Figure 1C**, protocols are illustrated in **Figure S3**). Stimuli were generated in two distinct modes. In open-loop (OL) mode, the grating moved at a constant, predefined speed, decoupling sensory inputs from motor outputs. In closed-loop (CL) mode, the grating speed was dynamically updated based on the intended motion of the fish, providing real-time visual feedback (**Figure 1D**).

Behavioral assays featured grating movement cycles of 20 second ON (moving), 20 second OFF (stationary). For free-swimming assays, larvae completed six trials with alternating grating directions (**Figure 1E**). For head-fixed assays, larvae underwent four trials, each comprising three closed-loop (CL) and three open-loop (OL) ON-OFF cycles, with the grating consistently moving in a tail-to-head direction during the ON phase (**Figure 1F**). Alteration between CL and OL cycles was used to induce behavioral diversity thus enriching the dataset[27].

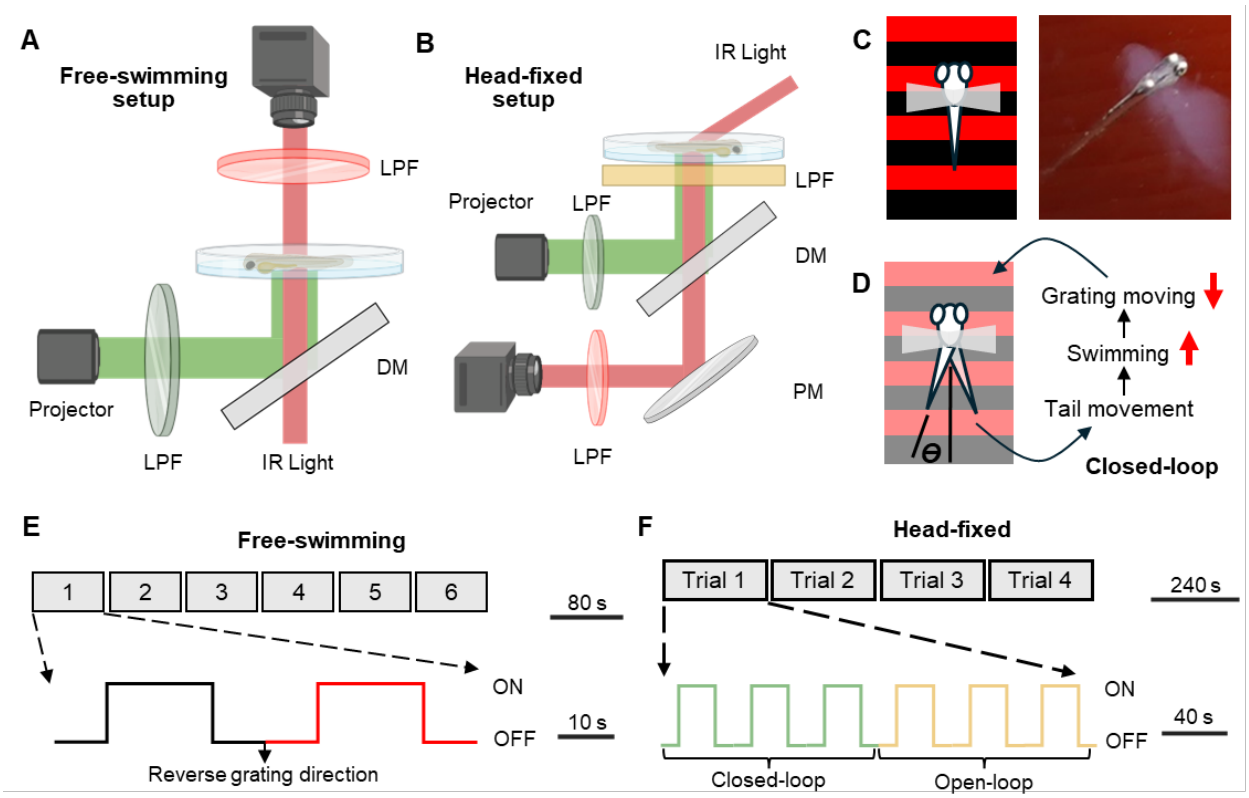

Figure 1. Experimental setups and behavioral paradigms for studying visuomotor responses.

(A) Schematic of the free-swimming behavioral setup. Larval zebrafish are tracked from above by a high-speed camera using infrared (IR) illumination. Visual stimuli are projected from below onto a diffusion screen. Long-pass filters (LPF) are used to separate the imaging and stimulus light paths.

(B) Schematic of the compact, head-fixed closed-loop setup. A folded optical path with a cold mirror (DM) and plane mirror (PM) directs IR light to the camera for tail tracking, while the stimulus is projected onto the same cold mirror and reflected to the fish.

(C) Left panel: Schematic of grating stimulus display beneath the head-fixed fish with movement in the head-to-tail direction. **Right panel:** The photo shows fish mounted with our protocol that has eyes and tail freed from agarose.

(D) Schematic of closed-loop visual feedback. Images captured by the camera are processed by Stytra, which converts the tracking parameter of the fish (*tail_sum*, θ) to the intended forward motion of the fish. Forward swims slow down the grating moving speed from the perspective of the fish to mimic their free-swimming experience.

(E) Protocol for the free-swimming OMR assay. The 480 s session consists of six trials of alternating 20 s grating movement (ON) and 20 s static pause (OFF) cycles. The direction of movement is reversed between consecutive cycles within each trial.

(F) Protocol for the head-fixed OMR assay. The 960 s session is divided into four identical trials. Each trial consists of three closed-loop (CL) cycles followed by three open-loop (OL) cycles. In CL, grating speed is adjusted to the swimming behavior of the fish; in OL, it moves at a constant velocity.

Panels A and B were created with Biorender.

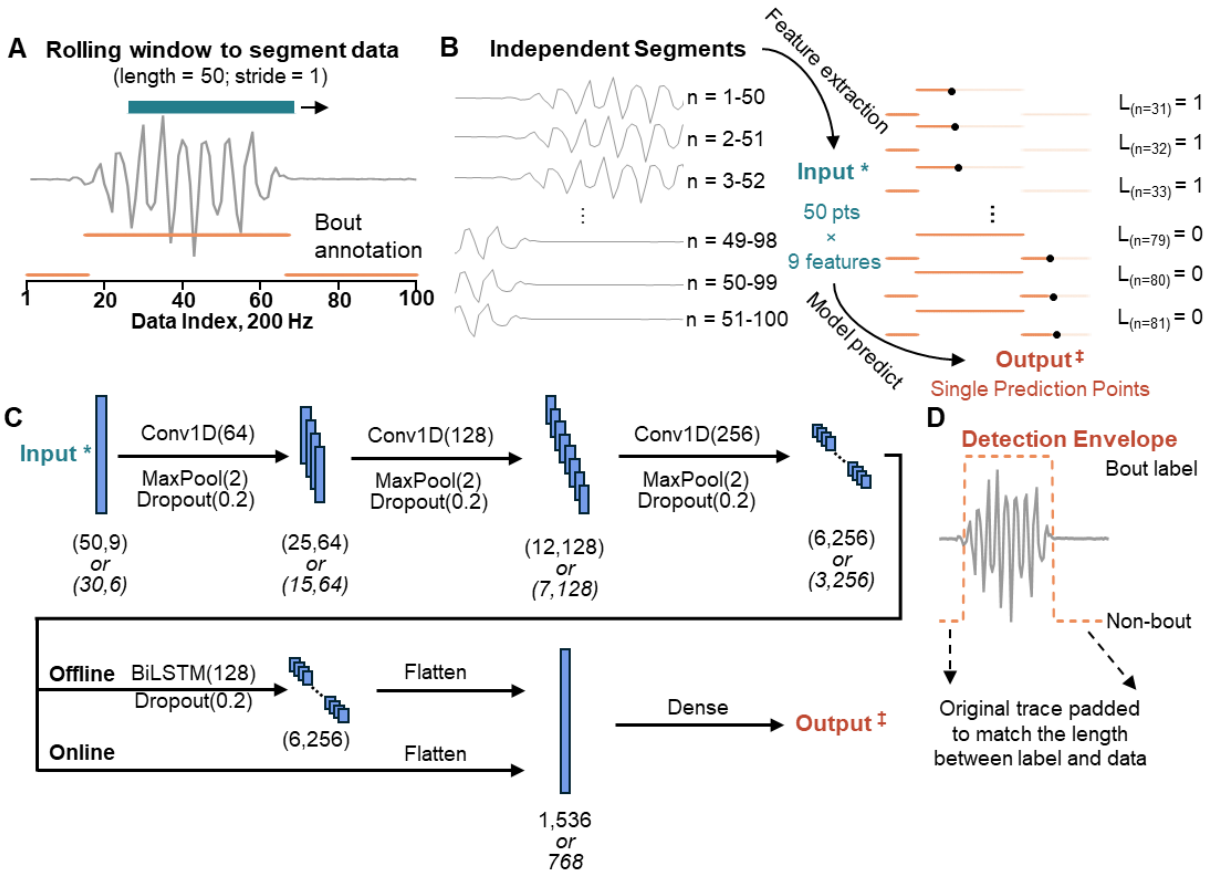

Figure 2. Illustration of data segmentation, inference, and structure of the threshold-free model.

(A) Illustration of data segmentation. A rolling window of 50 datapoints (depicted by the dark teal bar) moves one step at a time (stride = 1) over the entire time series dataset. Training and testing datasets are manually annotated to provide a positive label at time points where a bout occurred (depicted by the orange bout envelope).

(B) Left and center panel: Feature vectors are extracted for each data segment as the input (Input*) to the model (50 × 9 input shape). **Right panel**: The model produces a probability score (Output‡) for the 31st timepoint within the data segment as the label (1 for bout events and 0 for non-bout). Model structure converting Input* to Output‡ is further elaborated in **Panel C**.

(C) Architecture of offline (for *post hoc* data processing) and online (for real-time application) neural network models. Both models feature three hierarchical blocks. Each block consists of one 1D Convolutional layer (Conv1D; number of channels in parentheses), one Max Pooling layer (MaxPool; pooling size in parentheses), and one Dropout layer (Dropout rate in parentheses). The offline model includes an additional LSTM block containing one bi-directional LSTM layer to capture temporal dependencies within the segment. Both models terminate with a Flatten layer followed by a Dense (fully connected) layer with a sigmoid activation function to produce the final output score. Segments with scores greater than 0.5 are classified as positive (bout), while others are classified as negative (non-bout). Output shapes for each layer are indicated below them (top row = offline model; bottom row if applicable = online model). Predictions are connected to form continuous detection envelop for the whole dataset as shown in **Panel D**.

(D) Inference schematic for the model. The original data are zero padded to ensure length of the bout envelop matches that of the raw dataset. As the rolling window advances across the data, labels generated for each segment concatenate to produce a continuous detection envelope.

A Deep Learning Model Combining Convolutional and Recurrent Neural Network Layers for Context-Aware Bout Detection

For supervised learning and validation on bout detection performance, we first generated ground-truth labels for 8 training datasets and 6 testing datasets through manual annotation (**Figure S4**). The ground-truth was a precise binary time series, in which every timepoint was identified as either "bout" (1) or "non-bout" (0) (**Figure 2A**, orange trace).

To address the intra- and inter-larval variations in bout kinematics[38, 39], our deep learning model does not classify signals based solely on instantaneous values (content-based), rather, it is designed to assess a window with preceding and succeeding data, therefore determining the appropriate label for the timepoint (context-aware bout detection). To achieve this, the continuous *tail_sum* data (θ) is extracted using a 50-point rolling window (**Figure 2A**, teal bar). The prediction target was the label of the 31st point within each 50-point segment, using prior and subsequent data as the context (e.g., predicting the label $L_{(n=31)}$ for segment n = 1-50 in **Figure 2B**).

With the time series of 50 *tail_sum* values, we extracted a matrix of nine features—including the raw data (*tail_sum*) and eight derived features—designed to capture bout characteristics (**Figure 2B**, model for *post hoc* bout detection in **Table S1**). Derived features were computed by applying a 3-point rolling window to zero-padded *tail_sum* time series, which preserves the original data length. These features include the sum of squared *tail_sum* values (*sq*), mean of absolute values (*mean*), first derivative of *tail_sum* (*speed*), local maximum (*max*), local minimum (*min*), mean of *tail_sum* in the rolling window (*roll_mean*), gradient of *sq* (*sq_grad*), and *sq* of filtered data (*filt_sq*), where a passing band of 20-40 Hz was applied to *tail_sum* to highlight bout kinematics before extracting the *sq* values[40]. To validate our feature selection, we performed a permutation feature importance (PFI) analysis[41]. **Table S2** summarizes all features used in the initial version of the model along with their scores, and **Table S3** shows the importance score for the current version. PFI analysis confirmed that all selected features contributed to bout prediction, although their relative importance varied.

The architecture of this model begins with three hierarchical convolutional blocks (**Figure 2C**). Each block consists of a one-dimensional CNN (Conv1D) layer, one Max Pooling (MaxPool) layer, and one Dropout layer. The number of channels doubled sequentially (64 → 128 → 256) across the CNN blocks, while the kernel size decreased (9 → 3 → 1) to refine feature extraction. The output is then passed to a Bidirectional Long Short-Term Memory (BiLSTM) layer, enabling the model to learn long-range temporal dependencies in data. A final dense layer outputs a probability score, and a cut off at 0.5 generates a binary classification.

This process, applied iteratively across the entire dataset, produces a continuous "detection envelope" (**Figure 2D**). To avoid overfitting, the model was trained using 8-fold cross-validation, a binary cross-entropy loss function, and scheduled early stopping and learning rate reduction. Loss and F1-scores of each epoch are shown in **Figure S5** for training and validation. The training set contained 1.21 million labeled 50-point segments, where 14.2% were labeled as bouts. Since this model was designed for post-processing, we hereafter refer to it as the offline model.

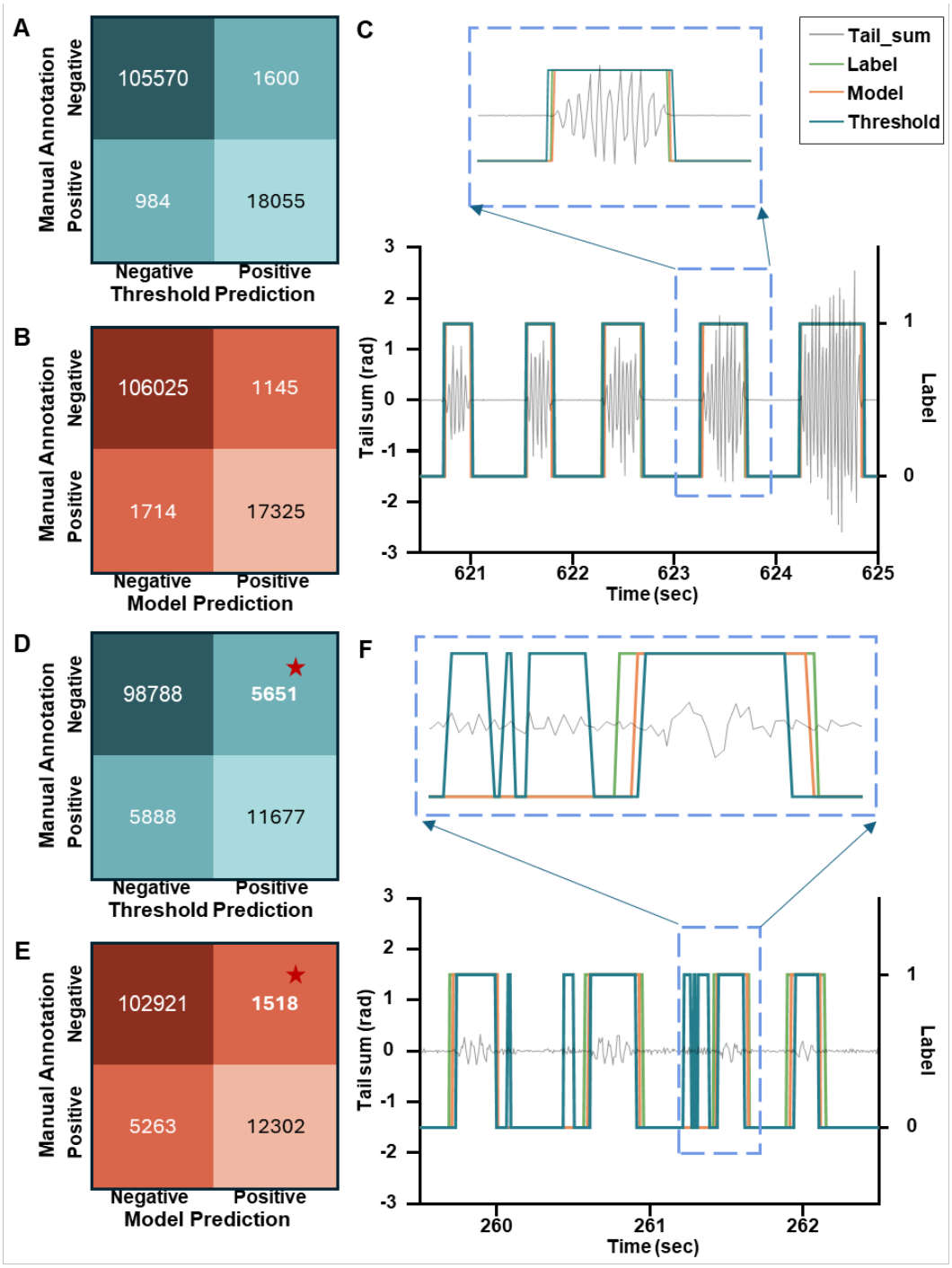

Figure 3. Comparison of threshold-based detection and the *post hoc* bout detection (offline) model on high- and low-amplitude datasets. Head-fixed 6–7 dpf *WT AB* larvae complete one session of four trials; each trial comprises three CL cycles followed by three OL cycles. Bouts were manually annotated from videos synchronized to tail traces and used as ground truth. The threshold-based detection method from Stytra, using the default threshold at 0.050, serves as the benchmark. In all confusion matrices, the horizontal axis is the predicted label (threshold-based or model prediction), and the vertical axis is the ground truth. Positive labels are bout, while negative labels are non-bout (label = 1 and 0 in **Panels C** and **F**). Results shown are raw outputs from each method without post-processing.

(A–C) Performance comparison on a high-amplitude dataset. High-amplitude datasets are defined as recordings in which more than 50% of manually annotated bouts exceed a maximum *tail_sum* amplitude of 1.0 rad. In this representative dataset, 56.12% of bouts meet this high-amplitude criterion. **Panel A** shows the confusion matrix of the threshold-based method, which achieves 97.95% Accuracy, 91.86% Precision, 94.83% Recall, and an F1-score of 93.32%. **Panel B** shows the same dataset analyzed with the offline model, yielding comparable performance with 97.73% Accuracy, 93.80% Precision, 91.00% Recall, and an F1-score of 92.38%. **Panel C** (620.5-625 s) illustrates that both methods closely match the ground truth for these high-amplitude bouts; the zoom-in (623-624 s) reveals that minor differences between methods primarily arise from small shifts in predicted bout onset and offset.

(D–F) Performance comparison on a low-amplitude dataset. Low-amplitude datasets are defined as recordings in which fewer than 50% of manually annotated bouts exceed a maximum *tail_sum* amplitude of 1.0 rad. In this example dataset, only 7.87% of bouts are high amplitude. **Panel D** shows the confusion matrix of threshold-based method, which yields an Accuracy of 90.54%, Precision of 67.39%, Recall of 66.48%, and F1-score of 66.93%. **Panel E** shows confusion matrix of the offline model applied to the same dataset, achieving higher Accuracy (94.44%), Precision (89.02%), Recall (70.04%), and F1-score (78.39%). **Panel F** (259.5-262.5 s) shows that both methods detect true bouts; however, the zoom-in view (261.2-261.7 s) highlights that the threshold-based method frequently misclassifies noise as bouts (false positives), whereas the offline model more effectively reduces these errors.

Notes: Grey traces = tail_sum tracking data from Stytra; Green envelops = manually labeled bouts; Orange envelops = bouts detected with offline model; Dark teal envelops = bouts detected with threshold-based method. See Method section for detailed definition of tail_sum and the operations of bout detection.

The Offline Model Outperforms a Threshold-based Method for Low-Amplitude Bout Detection

The threshold-based method from Stytra, with its default threshold level at 0.050, was used as benchmark to evaluate offline model performance. We further classified each testing dataset as "high-amplitude" or "low-amplitude" according to whether more than half of its manually annotated bouts reached a θ_{max} greater than 1.0 rad.

On a representative high-amplitude dataset where 56.12% of bouts were high amplitude, both methods performed well, achieving comparable accuracy (97.73% vs. 97.95%, offline model vs. threshold-based method) (**Figure 3A-B**). The offline model achieved higher precision (93.80% vs. 91.86%) while having lower recall (91.00% vs. 94.83%). For these distinct, high-amplitude events, predictions from the offline model, the threshold-based method, and manual annotations are largely in agreement (**Figure 3C**).

However, when analyzing low-amplitude data where the signal-to-noise ratio is poor, our offline model was more precise and showed higher sensitivity (recall score) than the threshold method. On a challenging dataset where only 7.87% of bouts were high-amplitude, our model outperformed the threshold-based method across all metrics (**Figure 3D-E**). Precision saw the greatest improvement, increasing from 67.39% to 89.02% when using the offline model. This was accompanied by significant gains in accuracy (from 90.54% to 94.44%), recall (from 66.48% to 70.04%), and the overall F1-score (from 66.93% to 78.39%). Visual inspection provides an explanation for the difference between the two methods: the simple threshold method flagged fluctuating noise as bouts, generating thousands of false positives that our context-aware model correctly ignored (**Figure 3F**).

By varying threshold levels in the threshold-based method for this low-amplitude dataset, we observed that lower values (≤ 0.060) produced more false positives, while higher values (≥ 0.065)

failed to capture the subtle onset and offset of low-amplitude bouts (**Figure S6A**). Although, based on the confusion matrix and performance metrics (**Figure S6B-C**), a threshold of 0.060 performed better than the default 0.050 used in Stytra (benchmark) and was the best of all tested thresholds. This demonstrates the need for per-experiment tuning using threshold-based bout detection methods. In contrast, the offline model utilized contextual information to detect bouts, leading to higher accuracy and F1-score than all tested thresholds for this dataset. On average for low-amplitude testing datasets, our model delivered a precision increase of 15.56% compared to the benchmark with a threshold of 0.050 (**Tables S4**). Averaged across all test data, the offline model achieved 97.82% accuracy, 94.26% precision, 90.44% sensitivity (recall), and an F1-score of 92.12%.

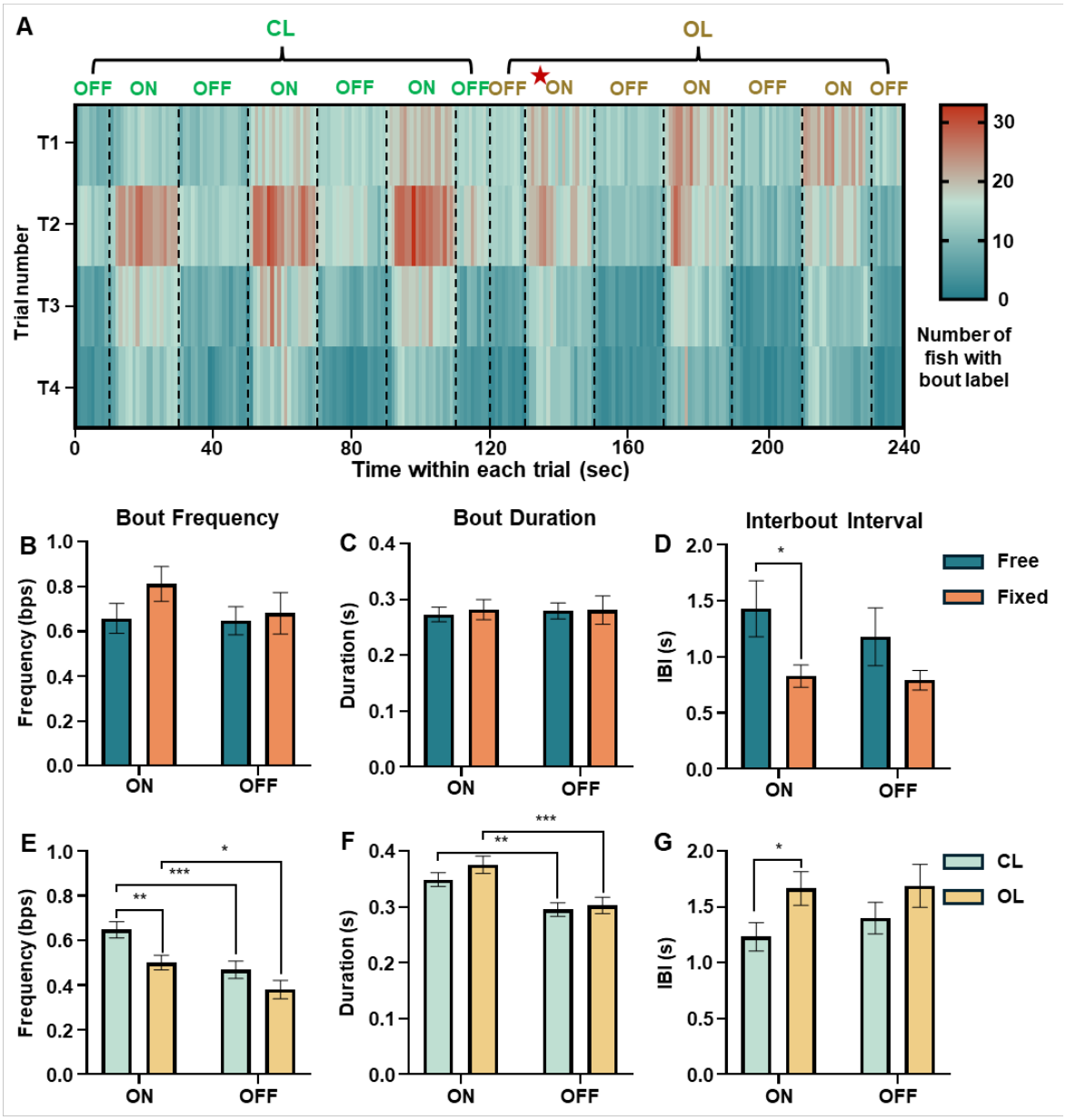

Figure 4. Grating movement (ON) and pause (OFF) evoke similar bout kinematics in free-swimming and head-fixed (closed-loop) *wildtype* larvae; closed-loop (CL) visual feedback preserves swim bout vigor in *WT* (*AB*) zebrafish larvae during the OMR. Free swimming fish (6-7 dpf, n = 13) undergo six rounds of ON/OFF stimulus cycles, with alternating grating movement direction within each round. Each head-fixed fish (6-7 dpf, n = 55) experiences a session containing four trials, where each trial contains three CL cycles followed by three OL cycles. Bouts are labeled by the offline model. To account for influence of OL rounds on response during subsequent CL rounds, comparison between free swimming fish and fixed-head fish only includes the closed-loop data from the first trial (Panels **B-D**).

(A) Heatmap for bout occurrences over time. Labels from fifty-five *WT* (*AB*) larvae are summed for 1s bins. For each fish, every timestep is labeled 1 (within-bout) or 0 (non-bout). The four trials within a session are shown as separate rows; the horizontal axis is elapsed time within each trial. Warmer (red) colors indicate higher numbers of larvae scoring positive for bout occurrences at that time; cooler (teal) colors indicate fewer bouts across larvae.

(B) Bout frequency comparison under grating movement/pause for free swimming and head-fixed fish. Fixed fish showed similar frequency to free swimming fish both during grating movement ($p = 0.1405$) and grating pause ($p = 0.7675$). Grating movement and pause result in similar bout frequency for free swimming fish ($p = 0.9104$). The same effect is observed for the CL rounds of head-fixed fish ($p = 0.2860$). Free-ON = 0.6583 ± 0.0667 bps; Free-OFF = 0.6480 ± 0.0626 bps; Fixed-ON = 0.8118 ± 0.078 bps; Fixed-OFF = 0.6812 ± 0.0923 bps. Mean ± S.E.M.

(C) Bout duration comparison under grating movement/pause for free swimming and head-fixed fish. Duration of bouts during grating movement phase ($p = 0.6992$) and during grating pause ($p = 0.9622$) are similar in fixed-head fish compared to in free swimming fish. For free swimming fish, grating movement triggers similar bout duration to grating pause ($p = 0.7394$). The same effect is observed for head-fixed fish ($p = 0.9797$). Free-ON = 0.2732 ± 0.0129 sec; Free-OFF = 0.2796 ± 0.0142 sec; Fixed-ON = 0.2817 ± 0.0180 sec; Fixed-OFF = 0.2810 ± 0.0256 sec. Mean ± S.E.M.

(D) Interbout interval comparison under grating movement/pause for free swimming and head-fixed fish. During grating movement, the interbout interval (IBI) of head-fixed fish is significantly shorter than free-swimming fish (Fixed-ON < Free-ON, $p = 0.0283$). IBI of fixed-head fish is similar to free-swimming fish during grating pause phase ($p = 0.1586$). For both free-swimming and fixed-head fish, the interbout interval is similar for grating ON and OFF bouts (Free $p = 0.4888$; Fixed $p = 0.7832$). Free-ON = 1.4291 ± 0.2497 sec; Free-OFF = 1.1796 ± 0.2580 sec; Fixed-ON = 0.8277 ± 0.099 sec; Fixed-OFF = 0.7912 ± 0.0867 sec. Mean ± S.E.M.

(E) Bout frequency comparison under grating movement/pause during closed/open-loop condition. CL leads to higher bout frequency than OL during grating movement (CL-ON > OL-ON, $p = 0.0030$), while the frequency is similar during grating pause phase ($p = 0.1202$). Grating movement induces significantly more bouts (OMR triggered bouts) than grating pause (spontaneous swims) during closed-loop (CL-ON > CL-OFF, $p = 0.0009$) and open-loop (OL-ON > OL-OFF, $p = 0.0238$) conditions. CL-ON = 0.6478 ± 0.0362 bps; CL-OFF = 0.4686 ± 0.0388 bps; OL-ON = 0.5005 ± 0.0329 bps; OL-OFF = 0.3803 ± 0.0411 bps. Mean ± S.E.M.

(F) Bout duration comparison under grating movement/pause during closed/open-loop condition. Bout duration is similar between CL and OL for both grating movement ($p = 0.1841$) and pause phase ($p = 0.6927$). Grating movement induces longer bout duration than grating pause during CL (CL-ON > CL-OFF, $p = 0.0025$) and during OL (OL-ON > OL-OFF, $p = 0.0008$). CL-ON = 0.3490 ± 0.0126 sec; CL-OFF = 0.2952 ± 0.0122 sec; OL-ON = 0.3756 ± 0.0156 sec; OL-OFF = 0.3028 ± 0.0146 sec. Mean ± S.E.M.

(G) Interbout interval comparison under grating movement/pause during closed/open-loop condition. CL reduces IBI during grating movement (CL-ON < OL-ON, $p = 0.0295$) but not during grating pause ($p = 0.2289$). Grating movement has no significant effect on IBI during CL condition ($p = 0.3766$) or OL condition ($p = 0.9245$).

CL-ON = 1.2318 ± 0.1276 sec; CL-OFF = 1.4004 ± 0.1410 sec; OL-ON = 1.6653 ± 0.1506 sec; OL-OFF = 1.6885 ± 0.1921 sec. Mean ± S.E.M.

Notes: Free = free swimming fish; Fixed = fixed-head fish during closed-loop experimental condition; ON = grating moving; OFF = grating paused. CL = closed loop. OL = open loop.

The Validated Offline Model Confirms Head-Fixed Larvae Largely Recapitulate Behaviors Observed in Free-Swimming Larvae

Having validated our offline model, we next applied it to a collection of datasets from 55 *wildtype* (*WT AB*) head-fixed 6-7 dpf larvae to investigate behavioral effects of our head-fixed, eye and tail free preparation and closed-loop visual feedback.

While swim bout activity initially increased from Trial 1 (T1) to Trial 2 (T2), likely reflecting adaptation to the arena, it steadily declined in subsequent trials–a behavioral signature of habituation[42, 43] (**Figure 4A**). We also observed a sharp, immediate drop in swimming behavior at the onset of the first OL cycle in each trial (**Figure 4A**, red star). This indicates that larvae rapidly detected the sensory mismatch when concordant visual feedback was removed, which is consistent with previous studies[29, 44, 45].

A critical question is whether our head-mounting protocol alters behavior. We compared bout kinematics between head-fixed larvae during the CL period of Trial 1 (prior to any exposure to the visual stimuli) and a cohort of freely swimming larvae (n = 13, *WT AB* larvae, 6-7 dpf). This analysis showed that bout frequency (measured in bouts per second, bps) was not significantly different between the free-swimming and head-fixed groups, during either grating movement (Free vs. Fixed, $p = 0.1405$) or pause ($p = 0.7675$) (**Figure 4B**). This consistency also extended to bout duration, with no significant difference found during movement ($p = 0.6992$) or pause ($p = 0.9622$) (**Figure 4C**). These results validate that our head-fixed protocol combined with closed-loop visual feedback largely preserves free swimming behavioral repertoires in head-fixed fish.

We did uncover a difference in interbout intervals (IBIs) between free and fixed fish. During grating movement, free-swimming fish exhibited significantly longer rest periods between bouts than their head-fixed counterparts ($p = 0.0283$). This difference was not apparent when the grating was paused ($p = 0.1586$) (**Figure 4D**). We speculate this is due to the energetic cost of locomotion. Free-swimming animals must overcome hydrodynamic drag, potentially leading to fatigue and longer recovery times[46], which is a physical constraint absent in the head-fixed preparation.

Finally, by comparing CL and OL conditions within the head-fixed cohort, we investigated how real-time visual feedback modulates the OMR. Data from all four trials in the fixed-head sessions were included for statistical analysis. A moving grating acted as a potent stimulus, significantly increasing bout frequency in both CL (ON vs. OFF, $p = 0.0009$) and OL ($p = 0.0238$) conditions, and significantly increasing bout duration in both CL ($p = 0.0025$) and OL ($p = 0.0008$) (**Figure 4E-F**).

The crucial role of visual feedback emerged only when larvae were actively swimming in response to the stimulus. During grating movement, CL feedback drove more persistent locomotion, resulting in a significantly higher bout frequency (CL vs. OL, $p = 0.0030$) and significantly shorter IBIs compared to OL ($p = 0.0295$). These differences were not present when the grating was paused,

with no significant differences observed in frequency (p = 0.1202), duration (p = 0.6927), or IBI (p = 0.2289) (**Figure 4E-G**). This supports that the neural circuits governing the reflexive OMR are highly dependent on visuomotor contingency, while those controlling spontaneous swimming are not, providing additional evidence for functionally distinct motor control pathways[27, 47].

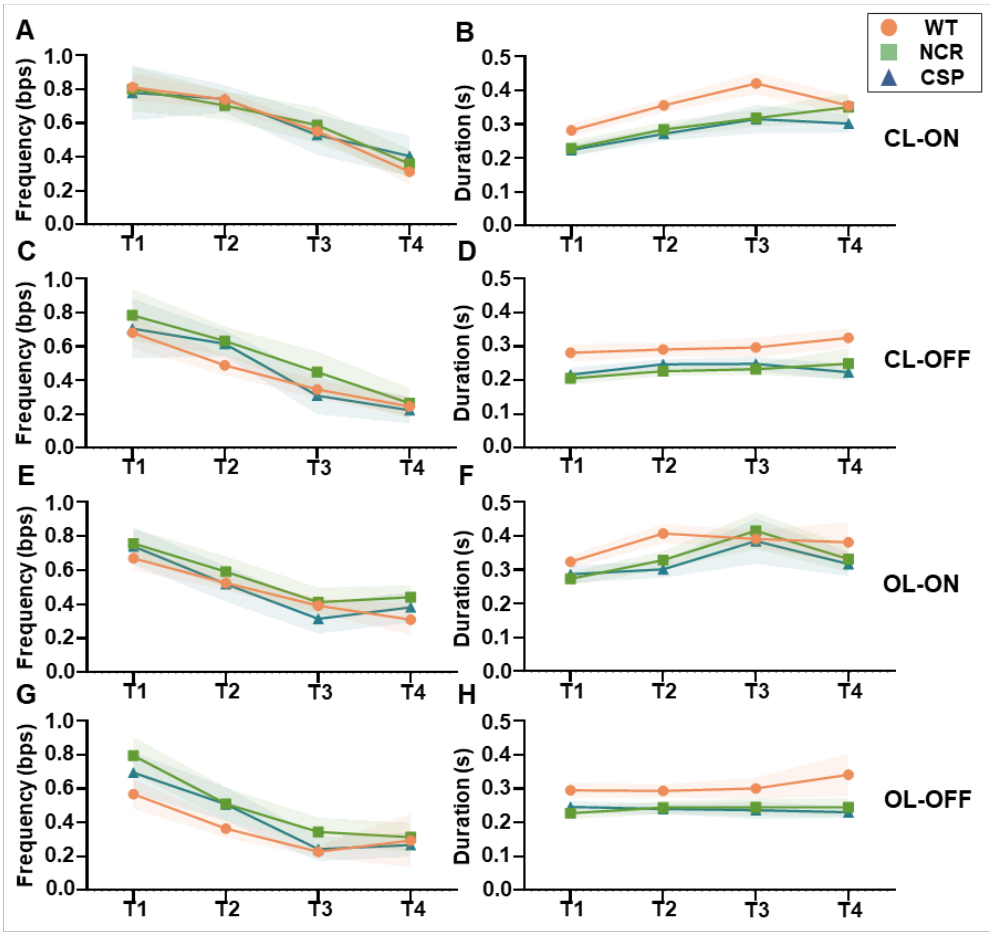

Figure 5. Bout frequency of zebrafish larvae is affected by prior trial experience, and genetic background strain effect contributes to variation in bout duration. 6-7 dpf fish of known genetic background including *wild-type AB* (*WT*, n = 55), *nacre* (*NCR*, n = 31), and *casper* (*CSP*, n = 26), each undergo a four-trial session, where each trial contains three CL rounds followed by three OL rounds. Bouts are labeled by the offline model. Data from the same experimental condition (CL/OL, ON/OFF) are averaged per trial per fish. In all panels, bout kinematics are presented as mean (labels in the panels) ± SEM (shaded area in the panels). Trial (T1-T4) and strain (WT, NCR, and CSP) effects are determined by 2-way ANOVA, then pairwise Tukey test is used to compare bout kinematics within the same session or across strains.

(A) Comparison of bout frequency across trials and for fish of different genetic backgrounds under the CL-ON condition. Trial effect contributes significantly to bout frequency in the CL-ON condition (2-way ANOVA, $p < 0.0001$). Pairwise comparison reveals decrease in bout frequency for WT and NCR but not for CSP fish. WT-T1 > WT-T4 (p = 0.004), WT-T2 > WT-T4 (p = 0.0024), NCR-T1 > NCR-T4 (p = 0.0040), and NCR-T2 > NCR-T4 (p = 0.0265).

(B) Comparison of bout duration across trials and for fish of different genetic backgrounds under the CL-ON condition. Bout duration is affected by both trial effect (2-way ANOVA, $p < 0.0001$) and strain effect ($p = 0.0002$). For trial effect, pairwise comparison reveals increase in bout duration for WT and NCR but not for CSP fish. WT-T1 < WT-T3 ($p = 0.0002$) and NCR-T1 < NCR-T4 ($p = 0.0088$). Bout duration in T2 and T3 reveals differences by strain. Namely WT-T2 > CSP-T2 ($p = 0.0484$), WT-T3 > NCR-T3 ($p = 0.0183$), and WT-T3 > CSP-T3 ($p = 0.0233$).

(C) Comparison of bout frequency across trials and for fish of different genetic backgrounds under the CL-OFF condition. Trial effect contributes significantly (2-way ANOVA, $p < 0.0001$). Pairwise comparison reveals decrease in bout frequency for all three strains. WT-T1 > WT-T3 ($p = 0.0310$), WT-T1 > WT-T4 ($p = 0.0065$), NCR-T1 > NCR-T4 ($p = 0.0013$), NCR-T2 > NCR-T4 ($p = 0.0277$), and CSP-T1 > CSP-T4 ($p = 0.0226$).

(D) Comparison of bout duration across trials and for fish of different genetic backgrounds under the CL-OFF condition. Bout duration is affected by strain effect (2-way ANOVA, $p < 0.0001$). Pairwise comparison shows difference in T4: WT-T4 > CSP-T4 ($p = 0.0400$).

(E) Comparison of bout frequency across trials and for fish of different genetic backgrounds under the OL-ON condition. Trial effect contributes significantly (2-way ANOVA, $p < 0.0001$). Pairwise comparison reveals decrease in bout frequency for all three strains from T1 to T3 and from T1 to T4. WT-T1 > WT-T3 ($p = 0.0241$), WT-T1 > WT-T4 ($p = 0.0047$), NCR-T1 > NCR-T3 ($p = 0.0149$), NCR-T1 > NCR-T4 ($p = 0.0276$), CSP-T1 > CSP-T3 ($p = 0.0066$), and CSP-T1 > CSP-T4 ($p = 0.0443$).

(F) Comparison of bout duration across trials and for fish of different genetic backgrounds under the OL-ON condition. Bout duration is affected by trial effect (2-way ANOVA, $p = 0.0087$). Pairwise comparison reveals difference for NCR with NCR-T1 < NCR-T3 ($p = 0.0307$).

(G) Comparison of bout frequency across trials and for fish of different genetic backgrounds under the OL-OFF condition. Trial effect contributes significantly (2-way ANOVA, $p < 0.0001$). Pairwise comparison reveals decrease in bout frequency from T1 to T3 and from T1 to T4. Including WT-T1 > WT-T3 ($p = 0.0131$), NCR-T1 > NCR-T3 ($p = 0.0037$), NCR-T1 > NCR-T4 ($p = 0.0013$), CSP-T1 > CSP-T3 ($p = 0.0136$), and CSP-T1 > CSP-T4 ($p = 0.0323$).

(H) Comparison of bout duration across trials and for fish of different genetic backgrounds under the OL-OFF condition. Bout duration is affected by strain effect (2-way ANOVA, $p = 0.0001$). Differences occur in T4, where WT-T4 > NCR-T4 ($p = 0.0431$) and WT-T4 > CSP-T4 ($p = 0.0387$).

Notes: CL = closed loop. OL = open loop. ON = grating moving; OFF = grating pause. WT = wild-type AB fish. NCR = nacre fish. CSP = casper fish. T1-T4 = Trials 1-4.

Analysis with Offline Model Reveals Trial and Strain Effects on Swim Bout Kinematics

Optical access to the brain is crucial to record neural activities simultaneously with visuomotor behaviors. Popular zebrafish mutants like *nacre* (*mitfa*$^{w2/w2}$), which lack melanophores, and *casper* (*mitfa*$^{w2/w2}$; *roy*$^{a9/a9}$), which lack both melanophores and iridophores, are ideal for this purpose[48]. However, the underlying genes may have pleiotropic effects[49]. If the mutations alter behavior, extrapolating correlations between neural activity and motor outputs to the *wildtype* strain becomes problematic.

We compared the bout kinematics of head-fixed *wild-type* (*WT AB*), *nacre*, and *casper* larvae (n = 55, 31, and 26, respectively) across four trials with alternating CL/OL and ON/OFF conditions (**Figure 1F**). 2-way ANOVA analysis identified a significant trial effect on bout frequency across all four experimental conditions ($p < 0.0001$ each). Visually, bout frequency decreased in later trials (**Figure 5A, C, E, G**). However, *post hoc* pairwise Tukey's tests ($\alpha = 0.05$) showed this decrease was not statistically significant for every strain across every condition or trial. For instance, *casper* fish did not show a significant frequency drop in the CL-ON condition from Trial 1 to Trial 4 (**Figure 5A**).

A similar pattern was observed for interbout interval (IBI), where ANOVA analysis reported a significant trial effect ($p < 0.004$ for all conditions), but *post hoc* tests showed a trend of increasing rest period was observed primarily in *WT* fish and was not consistently significant for mutant larvae (**Figure S7**).

A trial effect was also found for bout duration during grating movement (CL-ON, $p < 0.0001$; OL-ON, $p = 0.0087$), where bouts tended to become longer in later trials, though again, this was not consistent for all strains or for all trials (**Figure 5B, F**).

We also observed a significant strain effect on bout duration, with ANOVA indicating significance in three of the four conditions (CL-ON, $p=0.0002$; CL-OFF, $p<0.0001$; OL-OFF, $p=0.0001$). *Post-hoc* comparisons specified that where significant differences existed, *WT AB* fish elicited longer bouts than the mutant strains. Bout duration was not significantly different between *nacre* and *casper* fish (**Figure 5B, D, H**).

A Computationally Efficient Online Model Enables Real-Time Bout Detection without Sacrificing Tracking Speed

While our offline model demonstrates high accuracy, its computational demands, particularly from the BiLSTM layer (~559k parameters), create a processing bottleneck. When embedded in the Sytra-PsychoPy workflow (**Figure S1**), the offline model slowed down the real-time tracking rate from 200 Hz to ~50 Hz. This is insufficient for closed-loop experiments, where timely feedback is essential for maintaining larval visuomotor responses[20]. To overcome this, we developed an online model optimized for speed.

We preserved the three-block CNN backbone (~278k parameters) while removing the computationally expensive LSTM layer (**Figure 2C**). We further optimized the pipeline by selecting the six most important features from our earlier analysis, extracting them over a shorter 30-point window (predicting the label at the 27th point), and accelerating computation with a just-in-time (JIT) compiler[50]. A permutation importance analysis confirmed our selection of six features

was both effective and non-redundant (**Table S3**). Training and validation histories are shown in **Figure S8A-D**.

The online model retained the key advantage of the deep learning approach: it outperformed the benchmark threshold-based method on noisy, low-amplitude datasets, while showing comparable performance to the benchmark for high-amplitude datasets (**Table S4**). Compared to the offline model, the online version showed only a minimal degradation in performance. Mean accuracy remained high at 97.42% (a 0.4% decrease relative to the offline model), while the F1-score saw a 1.58% decline to 90.58%. This shows that, if computational time is not a strict restraint, the BiLSTM layer boosts model performance and is thus preferred (**Tables S4**).

When integrated into the live Stytra framework, the online model did not impose any noticeable overhead for the software workflow, operating seamlessly at the full ~200 Hz tracking rate (**Figure S8E**). We measured the end-to-end latency in a live setting for three *WT AB* fish (5 dpf, n = 3), finding an average delay of 212 ms for bout onset and 178 ms for offset. This real-world latency originated from the model's input window (15 ms for 3 datapoint future context), the feature extraction pipeline (measured 86.3 ms on average), and other system-level processes (such as data I/O on the shared memory to execute bout detection). Ultimately, the online model successfully balances computational speed and classification power, providing a robust and efficient solution for real-time, closed-loop experiments, while the latency warrants further improvement.

Discussion

In this study, we present two threshold-free swim bout detection models for zebrafish larvae, including a high-accuracy offline version for post-processing and a computationally efficient online version for real-time applications. With a threshold-free design, both models exhibit similar performance to the benchmark threshold-based method for high-amplitude datasets. However, our threshold-free models outperformed the benchmark for processing low-amplitude datasets. Due to similarities between noise and low-amplitude bouts such as those shown in **Figure 3F** and **Figure S6**, a simple amplitude-based threshold inevitably results in either missing true low-amplitude bouts (false negatives) or misclassifying noise as bouts (false positives), both of which our models minimize.

Using our validated model, we demonstrated that our head-fixed, eyes-and-tail-free preparation combined with closed-loop visual feedback largely preserves the natural free-swimming behavior of zebrafish larvae. Bout frequency and duration remained comparable between head-fixed and freely swimming cohorts, with the only significant difference being a shorter interbout interval in head-fixed fish during grating movement, which we speculate may be due to the absence of hydrodynamic drag.

We noticed that the bout frequencies in our head-fixed larvae were lower than those reported in one optomotor response (OMR) study, which reached 1.5-1.9 Hz[46]. Additionally, moving grating in free swimming larvae did not induce higher bout frequencies as previously reported[27, 29]. This may be due to our choice of visual stimulus: a grating with a 4 mm spatial period moving at 4 mm/s. While effective for inducing the OMR, these parameters generate a relatively slow optic flow, which likely initialized fewer and less vigorous bouts compared to faster stimuli. This observation aligns with previous findings that distinguish between distinct "slow" and "fast" bout types based on certain bout kinematics, such as maximum tail beat frequency (maxTBF), and the underlying neural activities of the fish[51, 52].

We also extended the application of our offline model to investigate the influence of genetic background (strain effect) and prior stimuli exposure (trial effect) on bout kinematics in larval zebrafish, which directly addresses longstanding calls to consider strain-specific behavior in experimental design[2, 53]. The impact of strain variation on zebrafish behavior is well-established for other assays, including variations in drug responses[54-56], longevity[57], disease susceptibility[58], social interactions[59-61], anxiety levels[60], metabolism[62], escape responses[8, 63], navigation[64], hunting[65], and swim bladder inflation[49]. Surprisingly, the influence of genetic background on the OMR—one of the most widely used assays for visuomotor function—has remained largely unexplored.

One previous study did compare the OMR of *wildtype* and pigmentation mutant larvae, which focused on bout frequency and reported no significant differences[48]. Here, we also incorporate bout duration and interbout interval (IBI) of *nacre* and *casper* strains, which are used for neural imaging due to their transparency. We found that bout frequency and IBI in all three strains, including *WT AB*, were influenced by prior exposure to the behavioral assay and, concurring with the study from Antinucci and Hindges, these metrics were not affected by genetic background. The similarities suggest that behavioral conclusions drawn from bout frequencies of *nacre* and *casper* can, with appropriate caution, be generalized to the *wildtype AB* line.

However, our analysis also uncovered a statistically significant strain effect on bout duration, where *post hoc* tests confirmed that the mutants consistently displayed shorter bouts than *wildtype*

larvae across multiple conditions (**Figure 5B, 5D, 5H**). The underlying cause remains an open and intriguing question. Future work should aim to disentangle whether this stems from variations in the central neural circuits processing optic flow, differences in motor command strength, or more peripheral factors such as muscle physiology and propensity for fatigue.

Our iterative development process has illuminated several clear pathways for enhancing model performance. First, we will enhance data diversity by expanding our training and testing sets to include more animals with a wider range of bout kinematics, thereby improving the model's generalization capabilities[66].

For the offline model, where detection accuracy is essential, we will explore architectures that better leverage temporal context, such as using longer time bins to maximize the effectiveness of the LSTM layer. For the online model, the primary challenge is balancing accuracy with inference speed. Simply deepening the network could negate its real-time advantage; therefore, we will focus on optimizing its existing architecture by exploring kernel sizes and feature selection to improve performance without increasing latency.

Finally, by converting the parameters of the online model from 32-bit floating-point precision level to more computationally efficient 16-bit or 8-bit representations, we anticipate a significant reduction in inference time with faster data transfer on shared memories. Achieving this will transform the model from a semi-real-time into a truly real-time detection method, unlocking its full potential for closed-loop visuomotor studies.

In summary, our threshold-free neural network models provide an accurate and precise framework for bout detection of larval zebrafish, overcoming the limitations of threshold-based methods and enabling more reproducible behavioral quantification. Because the models operate on tail-curvature dynamics, our approach is a broadly applicable tool for high-throughput behavioral assays and, when tail-angle measurements can be obtained, for studies involving other animal species and experimental paradigms.

Methods

Larval Zebrafish

All experimental procedures were conducted in accordance with the National Institutes of Health (NIH) guidelines and were approved by the Johns Hopkins University Animal Care and Use Committee (ACUC). Larval zebrafish (*Danio rerio*) of *wild-type* (*WT AB*), *nacre*(*mitfa*$^{w2/w2}$), and *casper* (*mitfa*$^{w2/w2}$; *roy*$^{a9/a9}$) genetic backgrounds were raised and maintained at 28.5 °C on a 14:10 hour light-dark cycle. Fish were tested at 5 days post-fertilization (dpf) for experiments utilizing the online bout detection model to detect bouts in real time, or at 6-7 dpf for all other behavioral experiments. Sex is not a variable relevant to larval stage studies as evidence of sexual differentiation does not occur until 3-4 weeks of age.

Control and Experimental Groups

Thirteen *wild-type* (*WT*) *AB* zebrafish larvae (n = 13) were used to establish baseline free-swimming behavior. A total of fifty-five *WT AB* (n = 55), thirty-one *nacre* (n = 31), and twenty-six *casper* (n = 26) larvae were used for head-fixed experiments. The head-fixed *WT AB* group served a dual purpose: (1) as the experimental group for comparison against the free-swimming condition to assess the effects of mounting, and (2) as the control group for evaluating kinematic differences to the *nacre* and *casper* mutant strains.

High-Framerate Visual Stimulus Generation

To achieve stimulus refresh rate exceeding the 60 Hz limit of Stytra[30], we set up inter-process communication pairing Stytra (Python 3.8) with PsychoPy[31] (Python 3.10+). Communication was managed using *multiprocessing.shared_memory* module in Python. The primary Stytra process allocated shared memory buffers to relay commands and real-time data to the concurrent PsychoPy process, enabling instantaneous updates to the visual stimulus (**Figure S1**, **Table S5**). The Stytra-PsychoPy workflow was used for both free-swimming and head-fixed OMR protocols.

Free-Swimming Setup

Experiments on free-swimming larvae were conducted in the arena depicted in **Figure 1A**. The arena was illuminated from below by a custom infrared (IR) LED array (Waveform Lighting, USA). A high-speed camera (Teledyne, USA) positioned above the arena captured dorsal-view images of the fish at up to 226 Hz (1440 × 1080 pixel resolution). A varifocal lens (Computar, USA) and a long-pass filter (LPF_1, $\lambda p \geq 830$ nm, Schott, Germany) were mounted to the camera. The camera field of view was adjusted to encompass the entire 35 mm petri dish (FABLAB, USA) containing the larva. Visual stimuli, generated by PsychoPy, were displayed using a projector (KODAK, USA) onto a cold mirror (DM, Techspec, USA) then reflected to a silicone diffusion pad beneath the petri dish. A long-pass filter (LPF_2, $\lambda p \geq 590$ nm, Thorlabs, USA) in the projection path created a high-contrast red-on-black stimulus, known to effectively elicit visuomotor responses[32, 33].

Free-Swimming Optomotor Response (OMR) Protocol

A single larval zebrafish (6-7 dpf) was placed in a 35 mm petri dish with E3 medium at 29 °C. After a 1–2 minute acclimation period with a static grating, a 480 s protocol was initiated (**Figure 1E**). The protocol consisted of six trials of a grating movement-pause cycle. Each cycle comprised 20 s of grating movement (spatial period ≈ 4 mm, speed ≈ 4 mm/s) followed by a 20 s static pause. The direction of movement alternated within each trial (e.g., 90°± 15° then 270° ± 15°) to prevent the fish from dwelling at the dish edge.

Head-Fixed Setup

For experiments with head-fixed fish, a compact version of the behavioral setup was designed for integration with microscopes[7], hereafter referred to as the "closed-loop setup" (**Figure 1B**, **Figure S2**). Tail beat kinematics were tracked using a folded infrared (IR) imaging path. An IR LED array (Waveform Lighting, USA) was positioned beside the microscope objective to illuminate the tail of the fish. The IR light first passed through the cold mirror (DM, Techspec, USA) and then was redirected by a plane mirror (PM, Thorlabs, USA). Images were captured by a high-speed camera (Teledyne, USA) through a custom 4f optical system (containing a camera lens from Kowa, Japan, focused at the infinity, followed by a 175 mm biconvex lens L1 from Thorlabs, USA) and a long-pass filter (LPF_1, $\lambda p \geq 830$ nm, Schott, Germany). To accommodate the different working distances of the 10× and 20× OPM objectives, L1 was mounted on a slider for manual focus adjustment. Same as in the free-swimming setup, a long-pass filter (LPF_2, $\lambda p \geq 590$ nm, Thorlabs, USA) was used to create a high-contrast red-on-black visual stimulus, which were presented by reflection light from a projector (Sony, Japan) by the cold mirror (DM) onto a projection screen. The projection screen was constructed from weighing paper (Fisherbrand, USA) and featured a rectangular aperture, ensuring the fish's tail remained visible to the camera while allowing the visual stimulus to be presented on the surrounding surface. The position of the aperture was adjusted with a slider on the 3D-printed dish holder.

Zebrafish Larva Mounting Protocol for Head-Fixation

To ensure consistent and minimally invasive head-fixation for behavioral assays, a mounting protocol was inspired by established methods[34-37].

Larval zebrafish (6-7 dpf) were selected for vigor, identified by their active escape response to gentle suction with a transfer pipette. A selected larva was transferred into 2% (w/v) low melting point agarose (LMPA, Fisher Scientific, USA) held at 42 °C. Approximately 35 µL of agarose containing the larvae was then pipetted into the center of a petri dish (FABLAB, USA). Before solidification, the larva was quickly oriented into a dorsal-up position by gently tapping the petri dish or using a soft, fine-tipped tool (e.g., a fire-polished pipette tip). The agarose was then allowed to solidify for approximately one minute (**Figure S3A**).

Next, a custom 3D-printed tool was used to cut the agarose and expose the tail and eyes (**Figure S3I**). To free the tail, four symmetrical, angled cuts (~45° distally) were made: two at the mid-caudal point of the body and two cuts originating near the pectoral fins (**Figure S3B**). The ideal agarose texture allows for a clean incision without exudation of liquid; if the cut is not clean, additional solidification time is recommended. Following the incision lines, the blade was inserted

vertically and used to gently peel the agarose block away from the body, freeing the entire caudal region of the fish (**Figures S3C-S3E**). A similar procedure was used to expose the eyes. Angled cuts were made proximal to the eyes (**Figure S3F**), and the blade was used to carefully lift and remove the resulting agarose pieces, ensuring an unobstructed field of view (**Figures S3G-S3H**). Note that all cuts were made at ~1 millimeter away from the fish, the rigidity of LMPA will allow complete removal of the excess piece along the cuts.

Immediately after agarose removal, a small drop (~35 µL) of E3 medium was applied to the fish to prevent drying out. Prior to behavioral assay, the petri dish was then filled with approximately 2 mL of warm (29 °C) E3 medium, which was added gradually to avoid dislodging the larva from the agarose.

Fixed-Head Optomotor Response (OMR) Protocol

Head-fixed fish (6-7 dpf) were acclimated for 1-2 minutes with still a grating before starting a 960s session (**Figure 1F**). The session consisted of four identical trials, each containing three closed-loop (CL) cycles followed by three open-loop (OL) cycles. Each cycle included a 20 s stimulus movement period (grating ON) and a 20 s stationary period (grating OFF).

- **Closed-Loop (CL):** The grating speed (spatial period = 4 mm, base moving velocity ~4 mm/s) was dynamically updated based on the fish tail movements to create a virtual swimming environment, driven by real-time bout detection described below.
- **Open-Loop (OL):** The grating moved forward at a constant velocity of ~4 mm/s during ON phase and was stationary during OFF phase.

Real-time Tail Tracking and Bout Detection

Larval tail kinematics were tracked using Stytra, which models the tail as six segments (with tracking angles θ_0 to θ_5, **Figure S4A**). From these, we modified a cumulative tail curvature metric in Sytra, *tail_sum* (θ), defined as the change in the sum of the rostral-most and caudal-most tail segment angles:

$$\theta(t) = \theta_5 + \theta_4 - \theta_1 - \theta_0 \qquad \textbf{Eq. 1}$$

Each tail segment angle θ_i was the difference between the instantaneous angle and the angle from the previous timestep, $\theta_i = \theta_i(t) - \theta_i(t-1)$, where $i \in \{0, 1, 4, 5\}$.

For closed-loop feedback for head-fixed fish, this *tail_sum* value was fed in real-time into the threshold-based function in Stytra (to generate datasets) or our online deep learning model (to test the model performance) for bout detection. Upon detection, a signal was sent via shared memory to PsychoPy, which then updated the grating velocity according to a custom function, *v(t)*, designed to determine the intended forward motion of fish:

$$v(t) = \left[\sum_{i=0}^{m} v_{t-i} e^{-ki}\right] + a\sqrt{E} \qquad \textbf{Eq. 2}$$

This function models inertia by integrating previous velocities (m = 4, k = 1.0366) and adds instantaneous thrust proportional to the current tail beat energy (E), with a scaling factor (a =

0.33681). Energy (E) was calculated from the power of key frequency components ($0.25f_s$ to $0.5f_s$) of the *tail_sum* signal, derived from an 8-point Short-Time Fourier Transform (STFT). The parameters were empirically derived from free-swimming behavioral data.

The Stytra user interface (monitoring at 60 Hz) and the PsychoPy stimulus window were screen-recorded in parallel using OBS Studio (Lain Bailey, USA) at a 60 Hz frame rate. These video recordings were later used for manual, post-hoc annotation of swimming episodes (**Figure S4A**).

Manual Annotation of Bouts

To generate high-fidelity ground truth labels for model training and testing, a two-stage annotation process was developed using custom MATLAB scripts (MathWorks, USA, version 2024b). First, a graphical user interface (*videoAnalysisTool*, **Figure S4A**) was used for coarse, frame-by-frame annotation of swim bouts from video recording of the fixed-head sessions. Second, these coarse labels were refined with high temporal precision using a second script (*boutExtraction*, **Figure S4B**). For each coarsely labeled bout event, this tool displayed the raw *tail_sum* data alongside its moving sum of squares and the derivative of its moving variance, allowing the annotator to precisely define the start and end of each bout based on quantitative signal changes. The resulting curated bout events served as ground truth labels. To further balance the training datasets, we included all bout labels and the 25 non-bout labels immediately before and after the bouts.

Feature Extraction and Model Architecture

We developed two bout detection models: a high-accuracy offline model and a low-latency online model. The training set was constructed from eight independent recordings of *WT* fish, with a separate test set of six recordings. Prior to feature extraction, missing values in the *tail_sum* data were imputed via quadratic spline interpolation. Interpolated data was zero padded to match the length between raw traces and the prediction sequence. A feature vector was then constructed using a sliding window (offline model: 50 points; online model: 30 points). **Table S1** summarizes the features for both models, which were calculated over a 3-point sub-window. The model architecture consists of three sequential 1D-CNN blocks (64, 128, and 256 channels; ReLU activation), each followed by Max Pooling and Dropout (rate = 0.2) (**Figure 2C**).

For the offline model, output of the CNN blocks was fed into a Bidirectional Long Short-Term Memory (BiLSTM) layer (128 units) to capture long-range temporal dependencies, prioritizing accuracy. The BiLSTM layer was omitted in the online model to minimize computational latency for real-time application. In both models, the final output was passed through a dense layer with a sigmoid activation function. An output value greater than 0.5 classified a frame as part of a bout.

Model Training and Validation

Models were trained using a binary cross-entropy loss function and the Adam optimizer (learning rate = 1×10^{-4}) for up to 100 epochs (batch size = 64). We used 8-fold cross-validation and implemented early stopping (patience = 5 epochs) and a learning rate scheduler to prevent overfitting and to optimize performance, monitored via the F1-score.

Inference and Threshold-Based Benchmark

To identify bouts in new datasets, the same data processing pipeline (interpolation, padding, feature extraction) was applied. For analyzing free-swimming datasets with intermittent tracking failures, data was segmented into contiguous, successfully tracked periods, and each segment was processed independently by the model. Predictions for periods of tracking failure were set to not-a-number (NaN).

For performance benchmarking, our models were compared against the threshold-based method used in Stytra, where a bout is detected if the standard deviation of *tail_sum* over a 10-point window exceeds a fixed threshold of 0.050. Comparison of bout detection performance did not involve post-processing described in the following procedures.

Detection Post-Processing and Kinematic Parameter Extraction

Swim bouts were first identified from raw *tail_sum* data using the high-accuracy offline model. A series of post-processing steps were applied to refine these events: detected segments shorter than 75 ms were discarded as noise, and segments longer than 2.0 s were either split into multiple bouts at points of inactivity or discarded[19, 67]. From the final curated bouts, the following kinematic parameters were calculated:

- **Bout Frequency:** Calculated as the total number of valid bouts initiated within a given condition. To account for periods of tracking failure in free-swimming recordings, bout frequency was normalized by the total duration of successfully tracked data during the condition.
- **Bout Duration:** The time in seconds from the start to the end of a single bout. The bout had to start and end within the same condition (e.g., entirely during a "grating ON" period) to be included.
- **Interbout Interval (IBI):** The time elapsed from the end of one bout to the start of the subsequent bout. Both bouts had to start and end within the same condition.

Statistics and Reproducibility

Statistical analyses were performed using MATLAB (R2024b, MathWorks, USA) and Prism (v10.6.1, GraphPad, USA), with a significance level of $\alpha = 0.05$. In all analyses, individual fish were treated as biological replicates. Kinematic parameters were first averaged for each fish within a given experimental condition, and these per-fish averages were used for statistical comparison. Trials that did not contain any bout were discarded, and thus all T-tests were unpaired. Data are presented as mean ± s.e.m. unless otherwise noted. The sample size (n) reported in figures and text refers to the number of individual fish.

- **Effect of Head-Fixation (Figure 4B-D):** Kinematics from free-swimming *WT* fish (n = 13) were compared to the first trial of head-fixed *WT* fish (n = 55) in the closed-loop condition using an unpaired, two-tailed Student's T-test.

- **Effect of Visual Stimuli (Figure 4E-G):** Comparisons between CL and OL conditions, and between grating ON and OFF periods, were made using data from all trials of head-fixed *WT AB* fish (n = 55) with unpaired, two-tailed Student's T-test.
- **Effect of Genetic Strain and Trial (Figure 5, Figure S7):** A two-way ANOVA was performed to assess the main effects of strain (*WT*, *nacre*, *casper*, n = 55, 31, and 26, respectively) and prior exposure to the stimuli (Trial 1-4), as well as their interaction. Where the ANOVA revealed significant effects, Tukey's multiple comparison test was used for post-hoc analysis of pairwise differences.

Data Availability

The datasets recorded and analyzed during the current study are available from the corresponding author on reasonable request.

Code Availability

Code for data acquisition and analysis are available at https://github.com/Hao-Chen-Wang/Threshold-Free-Bout-Detection

Author Contributions

J.Y. and H.W. conceptualized the study; H.W. and Y.Z. established the free-swimming behavioral arenas and revised the tracking code; Y.Z. developed algorithm to transform tracking data to fish velocity; H.W. designed and implemented the fixed-head behavioral setup, experimental protocols, and stimuli display; G.K. bred and raised the larval zebrafish; H.W. and G.K. jointly established the fish mounting protocol; H.W. performed the *in vivo* larval zebrafish experiments; H.W. designed, trained, and validated the deep learning models; H.W., G.K., and J.Y. performed data analysis; H.W. and J.Y. prepared the figures and wrote the manuscript; all authors contributed to manuscript revision; J.Y. and J.S.M. coordinated the collaboration, and J.Y. provided overall supervision for the study.

Competing Interests

The authors declare no competing interests.

Acknowledgement

This manuscript was supported in part by the National Institutes of Health (NIH, EBEB034272; AG072305; NS041435) and subject to the NIH Public Access Policy. NIH has been given the right to make this manuscript publicly available in PubMed. We also acknowledge the Boons Pickens Endowment Fund, and an unrestricted grant to the Wilmer Eye Institute from Research to Prevent

Blindness. GK received support from the Johns Hopkins Cross-Disciplinary Graduate Program in Biomedical Sciences. JSM received support from the Helen Larson and Charles Glenn Grover endowment funds.

References

1. Ahmad, F., Noldus, L., Tegelenbosch, R. & Richardson, M. Zebrafish embryos and larvae in behavioural assays. *Behaviour* **149**, 1241–1281 (2012).

2. Vignet, C. *et al*. Systematic screening of behavioral responses in two zebrafish strains. *Zebrafish* **10**, 365–375 (2013).

3. Li, F. *et al*. Characterization of the locomotor activities of zebrafish larvae under the influence of various neuroactive drugs. *Ann. Transl. Med.* **6**, 173 (2018).

4. Lin, T. *et al*. Negative optokinetic afternystagmus in larval zebrafish demonstrates set-point adaptation. *Sci. Rep.* **9**, 19039–4 (2019).

5. Li, X., Dong, G., Han, G., Du, L. & Li, M. Zebrafish Behavioral Phenomics Links Artificial Sweetener Aspartame to Behavioral Toxicity and Neurotransmitter Homeostasis. *J. Agric. Food Chem.* **69**, 15393–15402 (2021).

6. Zhang, L. *et al*. Large-scale phenotypic drug screen identifies neuroprotectants in zebrafish and mouse models of retinitis pigmentosa. *Elife* **10**, 10.7554/eLife.57245 (2021).

7. Shao, W. *et al*. Mesoscopic oblique plane microscopy with a diffractive light-sheet for large-scale 4D cellular resolution imaging. *Optica* **9**, 1374–1385 (2022).

8. Caperaa, M. *et al*. Development of sensorimotor responses in larval zebrafish: A comparison between wild-type and GCaMP6s transgenic line. *Behav. Brain Res.* **481**, 115412 (2025).

9. Portugues, R. & Engert, F. Adaptive locomotor behavior in larval zebrafish. *Front. Syst. Neurosci.* **5**, 72 (2011).

10. Budick, S. A. & O'Malley, D. M. Locomotor repertoire of the larval zebrafish: swimming, turning and prey capture. *J. Exp. Biol.* **203**, 2565–2579 (2000).

11. Borla, M. A., Palecek, B., Budick, S. & O'Malley, D. M. Prey Capture by Larval Zebrafish: Evidence for Fine Axial Motor Control. *Brain Behav. Evol.* **60**, 207–229 (2002).

12. Green, M. H., Ho, R. K. & Hale, M. E. Movement and function of the pectoral fins of the larval zebrafish (Danio rerio) during slow swimming. *J. Exp. Biol.* **214**, 3111–3123 (2011).

13. Huang, K., Ahrens, M. B., Dunn, T. W. & Engert, F. Spinal projection neurons control turning behaviors in zebrafish. *Curr. Biol.* **23**, 1566–1573 (2013).

14. Mirat, O., Sternberg, J. R., Severi, K. E. & Wyart, C. ZebraZoom: an automated program for high-throughput behavioral analysis and categorization. *Front. Neural Circuits* **7**, 107 (2013).

15. Trivedi, C. A. & Bollmann, J. H. Visually driven chaining of elementary swim patterns into a goal-directed motor sequence: a virtual reality study of zebrafish prey capture. *Front. Neural Circuits* **7**, 86 (2013).

16. Semmelhack, J. L. *et al*. A dedicated visual pathway for prey detection in larval zebrafish. *Elife* **3**, 10.7554/eLife.04878 (2014).

17. Portugues, R., Haesemeyer, M., Blum, M. L. & Engert, F. Whole-field visual motion drives swimming in larval zebrafish via a stochastic process. *J. Exp. Biol.* **218**, 1433–1443 (2015).

18. Jouary, A., Haudrechy, M., Candelier, R. & Sumbre, G. A 2D virtual reality system for visual goal-driven navigation in zebrafish larvae. *Sci. Rep.* **6**, 34015 (2016).

19. Mearns, D. S., Donovan, J. C., Fernandes, A. M., Semmelhack, J. L. & Baier, H. Deconstructing Hunting Behavior Reveals a Tightly Coupled Stimulus-Response Loop. *Curr. Biol.* **30**, 54–69.e9 (2020).

20. Markov, D. A., Petrucco, L., Kist, A. M. & Portugues, R. A cerebellar internal model calibrates a feedback controller involved in sensorimotor control. *Nat. Commun.* **12**, 6694–0 (2021).

21. Randlett, O. pi_tailtrack: A compact, inexpensive and open-source behaviour-tracking system for head-restrained zebrafish. *J. Exp. Biol.* **226**, jeb246335. doi: 10.1242/jeb.246335. Epub 2023 Nov 22 (2023).

22. Kim, S. S. *et al*. Adverse effects of diethyl phthalate and butyl benzyl phthalate on circadian rhythms and sleep patterns in zebrafish larvae. *Ecotoxicol. Environ. Saf.* **288**, 117350 (2024).

23. Guilbeault, N. C., Guerguiev, J., Martin, M., Tate, I. & Thiele, T. R. BonZeb: open-source, modular software tools for high-resolution zebrafish tracking and analysis. *Sci. Rep.* **11**, 8148–x (2021).

24. Jouary, A. *et al*. Megabouts: a flexible pipeline for zebrafish locomotion analysis. (2025).

25. Teicher, G. *et al*. Marigold: a machine learning-based web app for zebrafish pose tracking. *BMC Bioinformatics* **26**, 30–2 (2025).

26. Srinivasan, M. V. Visual control of navigation in insects and its relevance for robotics. *Curr. Opin. Neurobiol.* **21**, 535–543 (2011).

27. Naumann, E. A. *et al*. From Whole-Brain Data to Functional Circuit Models: The Zebrafish Optomotor Response. *Cell* **167**, 947–960.e20 (2016).

28. Warren, W. H. J., Kay, B. A., Zosh, W. D., Duchon, A. P. & Sahuc, S. Optic flow is used to control human walking. *Nat. Neurosci.* **4**, 213–216 (2001).

29. Mu, Y. *et al*. Glia Accumulate Evidence that Actions Are Futile and Suppress Unsuccessful Behavior. *Cell* **178**, 27–43.e19 (2019).

30. Štih, V., Petrucco, L., Kist, A. M. & Portugues, R. Stytra: An open-source, integrated system for stimulation, tracking and closed-loop behavioral experiments. *PLoS Comput. Biol.* **15**, e1006699 (2019).

31. Peirce, J. *et al*. PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).

32. Orger, M. B. & Baier, H. Channeling of red and green cone inputs to the zebrafish optomotor response. *Vis. Neurosci.* **22**, 275–281 (2005).

33. Zhou, M. *et al*. Zebrafish Retinal Ganglion Cells Asymmetrically Encode Spectral and Temporal Information across Visual Space. *Curr. Biol.* **30**, 2927–2942.e7 (2020).

34. Arrenberg, A. B. Fiber Optic-Based Photostimulation of Larval Zebrafish. *Methods Mol. Biol.* **1451**, 343–354 (2016).

35. Dehmelt, F. A., von Daranyi, A., Leyden, C. & Arrenberg, A. B. Evoking and tracking zebrafish eye movement in multiple larvae with ZebEyeTrack. *Nature Protocols* **13**, 1539–1568 (2018).

36. Hirsinger, E. & Steventon, B. A Versatile Mounting Method for Long Term Imaging of Zebrafish Development. *J. Vis. Exp.* **(119):55210. doi**, 10.3791/55210 (2017).

37. Vladimirov, N. *et al*. Light-sheet functional imaging in fictively behaving zebrafish. *Nat. Methods* **11**, 883–884 (2014).

38. Jouary, A. & Sumbre, G. Automatic classification of behavior in zebrafish larvae. *bioRxiv*, 052324 (2016).

39. Fitzgerald, J. A., Kirla, K. T., Zinner, C. P. & vom Berg, C. M. Emergence of consistent intra-individual locomotor patterns during zebrafish development. *Scientific Reports* **9**, 13647 (2019).

40. Roussel, Y., Gaudreau, S. F., Kacer, E. R., Sengupta, M. & Bui, T. V. Modeling spinal locomotor circuits for movements in developing zebrafish. *eLife* **10**, e67453 (2021).

41. Ewald, F. K. et al. *A Guide to Feature Importance Methods for Scientific Inference* (Explainable Artificial, Springer Nature Switzerland, Cham, 2024).

42. Roberts, A. C. *et al*. Long-term habituation of the C-start escape response in zebrafish larvae. *Neurobiol. Learn. Mem.* **134 Pt B**, 360–368 (2016).

43. Lamiré, L. A., Haesemeyer, M., Engert, F., Granato, M. & Randlett, O. Functional and pharmacological analyses of visual habituation learning in larval zebrafish. *Elife* **12**, 10.7554/eLife.84926 (2023).

44. Duque, M. *et al*. Ketamine induces plasticity in a norepinephrine-astroglial circuit to promote behavioral perseverance. *Neuron* **113**, 426–443.e5 (2025).

45. Chen, A. B. *et al*. Norepinephrine changes behavioral state through astroglial purinergic signaling. *Science* **388**, 769–775 (2025).

46. Holman, J. G. *et al*. A behavioral and modeling study of control algorithms underlying the translational optomotor response in larval zebrafish with implications for neural circuit function. *PLoS Comput. Biol.* **19**, e1010924 (2023).

47. Ahrens, M. B. *et al*. Brain-wide neuronal dynamics during motor adaptation in zebrafish. *Nature* **485**, 471–477 (2012).

48. Antinucci, P. & Hindges, R. A crystal-clear zebrafish for in vivo imaging. *Sci. Rep.* **6**, 29490 (2016).

49. D'Agati, G. *et al*. A defect in the mitochondrial protein Mpv17 underlies the transparent casper zebrafish. *Dev. Biol.* **430**, 11–17 (2017).

50. Lam, S. K., Pitrou, A. & Seibert, S. *Numba: a LLVM-based Python JIT compiler* (Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, Association for Computing Machinery, New York, NY, USA, 2015).

51. Severi, K. E. *et al*. Neural control and modulation of swimming speed in the larval zebrafish. *Neuron* **83**, 692–707 (2014).

52. Berg, E. M. *et al*. Brainstem circuits encoding start, speed, and duration of swimming in adult zebrafish. *Neuron* **111**, 372–386.e4 (2023).

53. Guryev, V. *et al*. Genetic variation in the zebrafish. *Genome Res.* **16**, 491–497 (2006).

54. Dlugos, C. A. & Rabin, R. A. Ethanol effects on three strains of zebrafish: model system for genetic investigations. *Pharmacol. Biochem. Behav.* **74**, 471–480 (2003).

55. Arenzana, F. J. *et al*. Teratogenic effects of ethanol exposure on zebrafish visual system development. *Neurotoxicol. Teratol.* **28**, 342–348 (2006).

56. Loucks, E. & Carvan, M. J. Strain-dependent effects of developmental ethanol exposure in zebrafish. *Neurotoxicol. Teratol.* **26**, 745–755 (2004).

57. Gerhard, G. S. *et al*. Life spans and senescent phenotypes in two strains of Zebrafish (Danio rerio). *Exp. Gerontol.* **37**, 1055–1068 (2002).

58. Bretaud, S., MacRaild, S., Ingham, P. W. & Bandmann, O. The influence of the zebrafish genetic background on Parkinson's disease-related aspects. *Zebrafish* **8**, 103–108 (2011).

59. Mahabir, S., Chatterjee, D., Buske, C. & Gerlai, R. Maturation of shoaling in two zebrafish strains: a behavioral and neurochemical analysis. *Behav. Brain Res.* **247**, 1–8 (2013).

60. Parker, M. O., Brock, A. J., Millington, M. E. & Brennan, C. H. Behavioural phenotyping of casper mutant and 1-pheny-2-thiourea treated adult zebrafish. *Zebrafish* **10**, 466–471 (2013).

61. Clements, K. N., Miller, T. H., Keever, J. M., Hall, A. M. & Issa, F. A. Social Status-Related Differences in Motor Activity Between Wild-Type and Mutant Zebrafish. *Biol. Bull.* **235**, 71–82 (2018).

62. Zhou, Z., Sun, Y., Yang, J. & Abliz, Z. Mapping the Metabolic Characteristics and Perturbation of Adult Casper Zebrafish by Ambient Mass Spectrometry Imaging. *Metabolites* **14**, 204. doi: 10.3390/metabo14040204 (2024).

63. Ren, J. Q., McCarthy, W. R., Zhang, H., Adolph, A. R. & Li, L. Behavioral visual responses of wild-type and hypopigmented zebrafish. *Vision Res.* **42**, 293–299 (2002).

64. Stowers, J. R. *et al*. Virtual reality for freely moving animals. *Nat. Methods* **14**, 995–1002 (2017).

65. O'Malley, D. M. *et al*. Optical physiology and locomotor behaviors of wild-type and nacre zebrafish. *Methods Cell Biol.* **76**, 261–284 (2004).

66. Xie, J., Sun, L. & Zhao, Y. F. On the Data Quality and Imbalance in Machine Learning-based Design and Manufacturing—A Systematic Review. *Engineering* **45**, 105–131 (2025).

67. Bollmann, J. H. The Zebrafish Visual System: From Circuits to Behavior. *Annu. Rev. Vis. Sci.* **5**, 269–293 (2019).