

## Automated Tracking of Zebrafish Shoals and the Analysis of Shoaling Behavior

Noam Miller and Robert Gerlai

### Abstract

Zebrafish spend the majority of their time in groups, called shoals. Shoaling behavior is complex and dynamic: fish leave and rejoin the shoal, distances between shoal-members oscillate, and the speed and polarization of the shoal change on timescales of seconds to minutes. All these features of shoals can be modified by various pharmacological and environmental—and possibly also genetic—manipulations and a thorough characterization of shoaling behavior can therefore be used as an effective assay for complex aspects of vertebrate social behavior. We present methods for acquiring and analyzing detailed trajectory data from shoals of zebrafish and demonstrate how these methods can be used to distinguish episodes of shoaling under different conditions. These methods could be further developed to create a standardized assay of shoaling behavior that will allow for an in-depth exploration of social behaviors in zebrafish.

**Key words:** Shoaling, Schooling, Automated video tracking, Time-series analysis, Zebrafish

---

### 1. Automated Tracking of Zebrafish Shoals and the Analysis of Shoaling Behavior

Zebrafish spend the majority of their time in small, cohesive groups termed shoals. The term shoal is used to describe an aggregation of individuals. Shoaling does not necessarily imply polarization, i.e., synchronized swimming direction. Shoaling is a complex, dynamic behavior (1) and can serve as a sensitive assay for a wide range of genetic (2, 3) or pharmacological manipulations (4–7). In general, a detailed analysis of zebrafish shoaling may lead to a better understanding of vertebrate social behaviors. Here, we present experimental examples and computational methods for efficient and detailed quantification of zebrafish shoaling.

Moving in shoals (and herds or flocks) provides multiple advantages in foraging and avoiding predation (8) and almost half of all known fish species shoal at some stage of their lives (9). Fish in shoals may be better able to detect oncoming predators (this is known

as the “many-eyes” effect, (10)). This allows each individual member of the shoal to spend less time on vigilance (11). In addition, a fish in a large shoal is less likely to be targeted by a predator (called dilution) and predators attacking a large, homogeneous shoal may experience sensory or cognitive “confusion,” limiting their ability to capture the fish in the shoal. Fish in shoals may also be able to locate and utilize food sources otherwise undetectable or unavailable to singletons. Last, shoaling presents an obvious advantage in cooperative breeding systems and/or in cases where synchronized spawning confers selective advantages.

Analyzing shoaling behavior relies on the acquisition of moment-to-moment trajectories of all members in the shoal. Recently, several researchers have begun collecting such trajectories for a number of species (12–14) and these data have allowed an unprecedented level of analysis of the nature of collective motion and detection of changes resulting from increasingly subtle pharmacological and environmental manipulations. We and others have previously shown how zebrafish shoaling is disrupted by, for instance, exposure to alcohol (6) or the threat of predation (15).

Below we present basic automated video tracking techniques for acquiring shoal trajectories and a framework for analyzing the resulting data. Most traditional statistical techniques are ill-equipped to uncover the regularities in extensive time-series data and we hope that the methods presented here will form the nucleus of a standardized method for quantifying zebrafish shoaling. All formulas and descriptions of behavioral measures are provided in Table 1.

The techniques we present apply to tracking fish in 2 dimensions (2D). We test fish in shallow tanks, which is not only ecologically

**Table 1**  
**Basic measures and their formulas**

Name	Formula	Description
Nearest neighbor distance (NND)	$\text{NND}_i = \min(D_{ij}), \text{ where}$ $D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$	Distance between fish $i$ and its nearest neighbor. The mean value is reported (Note 6)
Inter-individual distance (IID)	$\text{IID}_i = \frac{1}{N-1} \sum_{j \neq i} D_{ij}, \text{ where}$ $D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$	Mean distance from fish $i$ to all others. The mean value is reported
Speed	$S_i = \sqrt{\Delta x^2 + \Delta y^2}, \text{ is the distance}$ <p style="text-align: center;">moved in time <math>\Delta t</math>, usually 0.5 s</p>	Momentary speed of fish $i$ . The mean value is reported
Polarization	$P = \frac{1}{N} \left  \sum_{i=1}^N v_i \right , \text{ where } v_i \text{ is the}$ <p style="text-align: center;">movement vector of fish <math>i</math></p>	Magnitude of the mean vector of all fish

relevant (zebrafish live in small streams and shallow lakes) but also biases the distribution of fish in the shoal to the observable two dimensions. Results from 3D tracking of shoals (e.g., (12)) are sufficiently similar to the results obtained with 2D tracking that we do not believe the additional experimental and computational effort is necessary. For a discussion of techniques for 3D tracking see Wu et al. (16).

---

## 2. Materials

Adult zebrafish<sup>1</sup> should be experimentally naïve and allowed to acclimate to the laboratory environment before testing. Additionally, we house groups of fish together that are scheduled to be tested together for at least a week before the experiment. There is evidence to show that zebrafish prefer to shoal with familiar conspecifics (18). Groups should consist of equal numbers of males and females. We have tested groups of various sizes from 8 to 50 ((15); unpublished data) and the techniques below apply equally to any group size.

The design of the testing environment is particularly important for later video tracking of the fish (Fig. 1). The tank should be

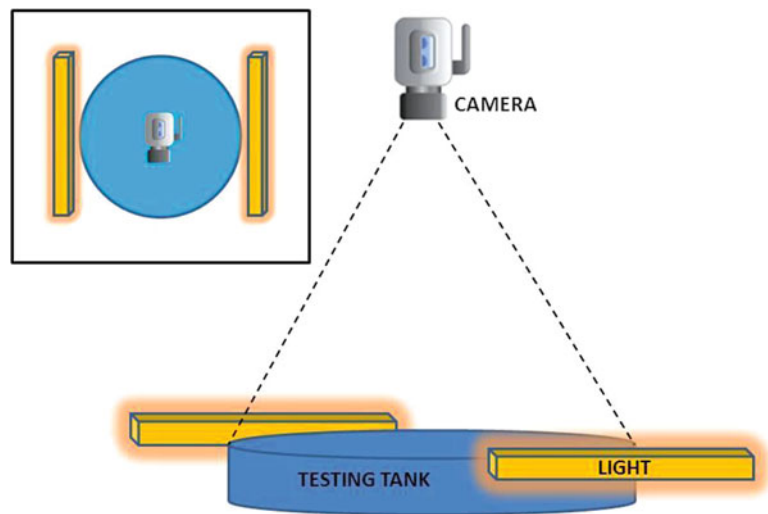


Fig. 1. Experimental setup. The inset shows the tank from above. Lights should be placed just above the height of the lip of the tank. The camera is placed so that the entire tank is visible in the image.

---

<sup>1</sup>Zebrafish shoaling develops gradually from around 7 days of age to adulthood (17). We do not give any data on juvenile shoaling but the techniques we present are broadly applicable to any shoaling fish.

white if possible, to provide high contrast with the fish. This color choice is purely technical as there is some controversy in the literature as to light/dark preferences in zebrafish. However, zebrafish—which are mildly counter-shaded—are more visible from above against a light background. A white, brightly lit tank may also induce low levels of fear, which will enhance shoaling tendencies. The tank should be large enough for the fish to swim freely and to allow the shoal to disperse. We have found that a tank approximately 1 m in diameter is sufficiently large for groups of up to 50 adult zebrafish. A video camera is mounted above the tank such that the entire tank is visible in the frame of the video. We recommend recording videos at high resolution (ideally  $1920 \times 1080$  pixels) and at no less than 10 frames per second (fps).

The tank should be lit by fluorescent lights. To avoid reflection from the lights in the video image and to ensure an even level of illumination in all parts of the tank, the lights are placed at the sides of the tank, just above the lip of the tank (Fig. 1). The tank is filled with water to a depth of about 10 cm.

---

### 3. Procedure

Fish are gently netted from their home tank into a plastic beaker. If the experiment involves exposure to a pharmacological agent, the fish are placed into a small tank in which they are exposed to the drug (6). After exposure (and recovery, if necessary), the fish are transferred in the beaker to the testing tank. The camera is started before the fish are placed in the tank, to provide a reference image of the tank for tracking purposes. The fish are then carefully released into the center of the tank. It is important to allow the fish time to habituate to the testing tank. We have found that shoaling characteristics stabilize after about 5 min in the testing tank and we therefore ignore the trajectories of the fish for the first 5 min of each session. Different experiments will require different session lengths but we have found that shoaling can be characterized in sessions as short as 10 min. The water in the testing tank is replaced periodically (usually after every two groups) to control for possible odor cues left by previous groups. At the end of the session, the fish are gently netted back into the beaker and returned to their home tank.

---

### 4. Video Tracking

Video files downloaded from the camera will often need to be converted into a standard format to be tracked. This can be achieved

using a number of video converter programs (we recommend using open-source video editors such as ffmpeg or VirtualDub; Note 1). Once videos have been converted, the positions of all the fish in each frame of the video are pinpointed using a tracking program. Below, we present a step-by-step guide to constructing an automated tracking program. Details of similar applications have been described in the literature (13, 16, 19). A more specialized application, specifically for tracking pairs of moving zebrafish, is described by Kato et al. (20). Computer applications that can reliably track a single moving individual are widespread (e.g., Noldus's Ethovision) and are not discussed here as most of them cannot track multiple identical targets in the same arena.

Several general principles of video tracking, conserved across almost all current applications, may be enumerated. In most cases, target locations are identified by subtracting the image of interest (a single frame of video) from a reference image of the empty testing enclosure, without the targets. Reference images are either frames of video filmed before the subjects are introduced into the arena, or are created digitally *post hoc* by erasing the targets using image editing software (e.g., (21)) or averaging multiple images together (e.g., (12); Note 2). Any pixels in which the difference between the reference and target images exceeds some threshold are flagged as part of a target.

Once the locations of the targets in a single frame have been identified, most current applications attempt to reconstruct the trajectory of each target through successive frames, i.e., through time. A detailed (and technical) discussion of this process is given by Wu et al. (16). The computation required may be thought of as mapping each target in frame T to the corresponding target in frame T-1. Most systems use a regression that finds the mapping giving the smallest summed spatial deviation of targets between frames. The density (in time) of tracked frames must be high so that targets do not move very far between successive frames.

Finally, the most computationally difficult problem faced by all video tracking systems arises when targets occlude. Occlusion occurs when two or more targets move under one another or are so close that they appear as one target (Fig. 2). The difficulty lies in deciding which of the targets after separation corresponds to each target before the occlusion (Fig. 2). Experimenters have sought to reduce the severity of this problem by testing fish in very shallow water (sometimes as little as 2 cm; e.g., (22)) or by tracking targets in all three dimensions using multiple cameras (e.g., (12, 14)), a solution that creates its own problems. The most common solution to this problem, which we also adopt, is presented below.

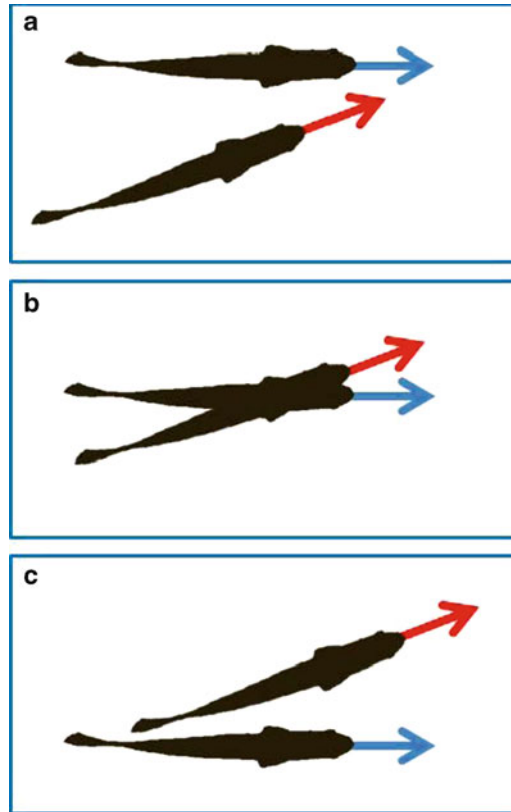


Fig. 2. Resolving target occlusions. Occlusions occur when two or more fish occupy the same position in the video image. The figure shows the positions of two fish before (a), during (b), and after (c) an occlusion. To determine which fish in (c) corresponds to the fish in (a), we assume that each fish does not turn during the occlusion (see text for details).

One simple solution to the difficulties of automated tracking is to allow a human user to perform the most computationally difficult parts of the tracking. Many manual tracking applications exist and, where the number of targets and duration of video to be tracked are not too large, have been used to great effect (e.g., (15)). Most applications automate all the intervening steps to increase the efficiency of the coding (an excellent free example is ImageJ's MTrackJ plugin; Note 1).

## 5. Step-by-Step Instructions for Writing an Automated Tracking Program

1. *Image subtraction*: the red, green, and blue values are summed for each pixel of the reference image and for the corresponding pixel in the target image. The difference between the values is

computed. A user-defined threshold is used to identify pixels that may belong to a target.

2. *Clumping*: the computational challenge here is to decide which pixels belong to the same target (fish) and which belong to an adjacent target. Each clump is seeded from a single flagged pixel and “grows” outwards by recruiting other above-threshold pixels that are close enough to be considered part of the same target. The growth process is iterated until the target no longer grows with repeated iterations. The next unassigned pixel is used to seed the next clump and the process repeats until all above-threshold pixels have been assigned to a clump.
3. *Modifying clumps*: some clumps will be too large or too small to be one single fish. The minimum and maximum number of pixels that a “real” target can occupy are determined by the user. Clumps consisting of fewer pixels than the minimum are removed from further consideration. Clumps consisting of more than the maximum number of pixels are assumed to have resulted from partial occlusion and are successively split until each clump contains fewer than the maximum number of pixels. Each over-sized clump is split along its shortest axis. If any resulting clump is still made up of more than the maximum number of pixels, it is split again. The minimum and maximum values must be carefully chosen to avoid splitting single fish into multiple targets or ignoring small fish altogether (Note 3).
4. *Target identification*: the mean location of all the pixels constituting a particular clump is taken as the location of a target in the current frame.
5. *Trajectories*: this step assigns an identity to each target in each frame and tries to ensure that identities are consistent across frames. Identities are assigned based on the smallest summed deviation of target locations between successive frames. For each frame, the summed distance moved by all targets is calculated for every possible mapping of target identities. The minimal solution is selected. Some applications limit the regression results using a trajectory prediction algorithm, which assumes a maximal turning rate of the fish (e.g., (13)). We do not apply trajectory prediction since zebrafish frequently turn abruptly and quickly.

In addition to the above basic steps, we suggest using one more simple technique to ensure that targets are not misidentified during tracking. In some frames, one or more fish may be missed completely. We therefore include a user-defined movement threshold in our tracking algorithm. If the closest match of a target is further than the threshold distance (in other words, the target “moved” more than the threshold distance in one frame), it is

assumed that the target was not tracked in that frame. Target identities are remapped (Step 5), excluding the missing target. The closest of any remaining unassigned targets after the mapping is complete is tentatively presumed to be the missing target. The application pauses and the user is required to manually identify the location of the missing target. The correct application of this method requires that the movement threshold be carefully selected. The threshold should be just larger than the distance a fish could move in a single frame.

Finally, we discuss how to deal with occlusions. A simple method is to allow the user to manually resolve all occlusions (e.g., (12)). Alternatively, we suggest one automated method (Fig. 2; see also ref (16)). It is important to note that occlusions often involve more than two fish (sometimes all the fish), and even an experienced observer is sometimes unable to unambiguously assign target identities. During an occlusion, we assign all participating fish the same coordinates (which may be modified later when the trajectories are smoothed). We then assume that fish continue on their path with minimal turning during an occlusion (as in Fig. 2). We calculate the angle of turn of each participating fish under each possible combination of target identities and select the option that minimizes the summed total turn angle. Delcourt et al. (13) have recently described an application that uses the same method.

An automated tracking system designed for a particular application will be more effective than a generic program. Any automated system embodies certain assumptions about the experimental parameters it is trying to uncover, and ours is no exception. Specifically, the clumping and occlusion resolution methods we propose will work best when occlusions occur only some of the time. The assumption that fish turn minimally during an occlusion, for example, becomes increasingly less valid as the duration of an occlusion increases. The ability to distinguish the paths of fish also decreases as more fish participate in any given occlusion. We have found that shoals of zebrafish do not occlude too frequently for the method described here to be used effectively, though the number and duration of occlusions increase with increasing group density, for obvious reasons.

In addition, our method is designed to work well when the targets are small on the image. We do not use the size of a target or its outline to distinguish between individuals, a method that may be useful when tracking smaller numbers of larger (and more diverse) organisms. In our experimental setup an average-sized zebrafish (about 3.5 cm in length) takes up about 150–250 pixels on the video image (at a resolution of 1920 × 1080 pixels).



## 6. Analysis

Trajectory data consist of multiple long time-series in which adjacent—and also more distant—data points are nonindependent. In addition, the relationships between time-series' (i.e., between the trajectories of different fish) are of interest. Standard statistical methods commonly employed in psychology research are unable to effectively analyze such data and we therefore present a set of techniques that extract the basic relationships in trajectory data. It is our hope that this set of methods will be enlarged, improved, and eventually standardized as more trajectory data are accumulated and the most informative tests are identified. Table 2 presents a list of the endpoints we suggest using to characterize shoaling, discussed in detail below. A *Mathematica* (Wolfram Technologies) notebook to calculate all measures is available from the corresponding author.

### 6.1. Resampling

A common method of analyzing dense time-series' is to resample the data at longer intervals, thus reducing the nonindependence of adjacent points. The resampling interval chosen is usually the first zero-crossing of the autocorrelation function of the data, i.e., the interval at which consecutive values are maximally uncorrelated (e.g., using behavioral data see ref (23)). This method is less than ideal, since autocorrelations in time-series' represent two independent sources of correlation: one resulting from the proximity in time of consecutive measurements; the second from behavioral consistency in the subjects, which we wish to uncover, not remove. We suggest, instead, to simply reduce the effective N for each statistical test by a fixed proportion (we commonly use the number of frames coded per second).

### 6.2. Smoothing

The first step in analyzing trajectory data is smoothing the trajectories. Raw trajectories will be noisy due to changes in the fish's body shape during swimming (which may affect where the tracking system places the center of the target), the limited spatial and temporal resolution of the video, and errors of the tracking system.

**Table 2**  
**Behavioral endpoints**

Number and duration of excursions away from the shoal (26)
Cross-correlation coefficients of NND, IID, polarization, and speed (Note 7)
Comparison of distributions of NND, IID, polarization, and speed
Typical oscillation frequencies of NND, IID, and polarization (28)

Trajectories are therefore smoothed using a weighted moving average with a window width of 0.5 s (24). In addition, all the coordinates are rescaled—using the known width of the arena—so that all values are in centimeters. The origin of the axes is placed in the lower left-hand corner so that all the positions have positive coordinates for simplicity.

### 6.3. Shoal Membership

To develop a clear picture of a given episode of shoaling it is obviously important to exclude from analysis fish that are not members of the shoal. We (25) have recently devised an algorithm for determining shoal membership that identifies excursions of individuals or subgroups away from a shoal. In addition to permitting the exclusion of fish that have temporarily left the shoal, the number and duration of excursions provide additional information about the characteristics of the shoal (see Fig. 3 and ref (25)).

We focus our analysis on four basic descriptive statistics (Table 1): the Nearest Neighbor Distance (NND), Inter-Individual Distance (IID), Polarization, and Speed. Each measure describes a univariate time-series and we utilize both the distribution and dynamics of each one (below). In addition, it is often informative to examine the correlations between the different measures. For instance, it has been reported that shoal speed and polarization correlate (e.g., (12)), and that polarization increases with the size of the group (26). In addition, NND and IID often correlate strongly. Simply comparing the correlation coefficients of these measures between experimental conditions may uncover interesting effects on shoaling behavior.

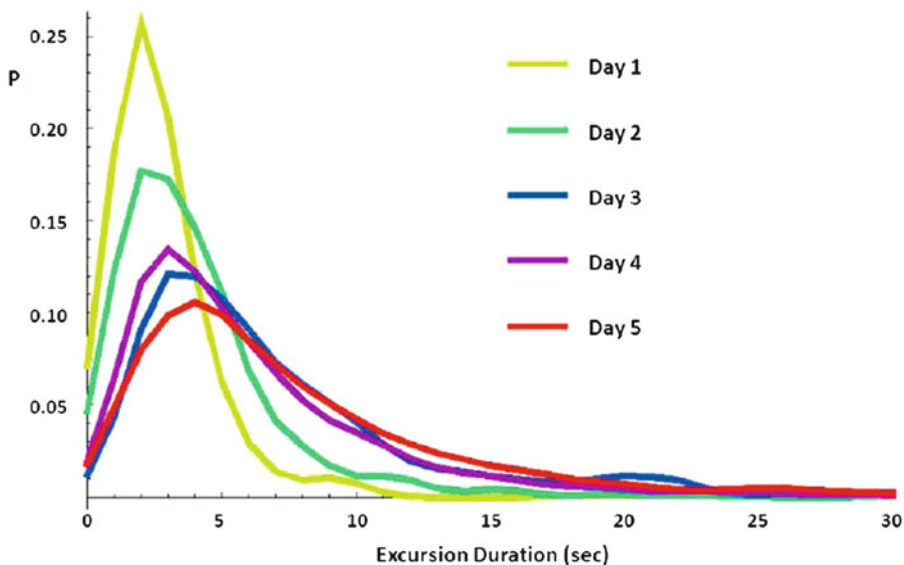


Fig. 3. The effects of repeated exposure to the testing tank on excursion durations. Distributions of excursion durations (episodes where one or more fish left the main shoal; (25)) across multiple exposures to the testing tank. As fish habituate to the testing environment the shoal gradually dissolves and excursions become longer. See (25) for more details.

#### **6.4. Measure Distributions**

The density distribution of each measure is constructed using a kernel density estimator (essentially a smoothed histogram; see ref (24)). Distributions from different experimental conditions are compared using a two-sample Kolmogorov–Smirnov test (with  $N$ 's modified as described above). This method is comparable to examining the mean of each measure but has the added advantage that it can detect differences in the distributions even when their means are very similar and takes into account differences in variance between distributions. For example, Fig. 3 presents the distributions of durations of excursions away from a shoal as a function of repeated exposures to a testing tank (25), clearly showing the distribution spreading despite little change in the location of the distribution's peak.

#### **6.5. Measure Dynamics**

We have previously shown that the IID of zebrafish shoals oscillates with a distinctive period (27). Aoki (28) showed similar oscillations in the NND of field gudgeon shoals and Viscido et al. (12) showed that the polarization of shoals of giant danios (*Danio aequipinnatus*) oscillates as well. Thus, none of these measures is constant across time, even within a single short testing session. We therefore examine the time-series of each measure for periodic components. Periodic oscillations in the time-series' are examined using the discrete Fourier transform of the data (27). The Fast Fourier Transform (FFT) algorithm is used to calculate the periodogram of each time-series (Note 4). Peaks in the periodogram represent potentially significant oscillation periods in the time-series. In order to determine which of the periodogram peaks are significant, we use the Lomb–Scargle test ((29, 30); Note 5). For each measure, the distribution of significant peaks (pooled from all sessions under the same experimental condition) is constructed. Distributions are compared to each other using a KS test (with  $N$ 's unmodified in this case, as the peaks of a periodogram are independent of each other; (27)).

In summary, we suggest that the measures above provide, in combination, a multi-dimensional characterization of an episode of shoaling. There are, of course, many more measures that could be extracted from shoal trajectories and which might reliably distinguish conditions of interest from each other. For instance, some researchers have examined the distances and bearings to further neighbors (i.e., not just the nearest neighbor; e.g., (31)). The set of measures we have selected (Table 2), however, are easy to calculate and compare across groups and, in our experience, characterize shoaling episodes thoroughly. Comparing the behavioral profiles of shoals under different genetic, environmental, or pharmacological conditions will allow researchers to determine the effects of these manipulations on social behavior in zebrafish. The complexity of shoaling behavior and the sensitivity and depth of the suite of measures suggested here will permit more subtle effects to be detected than with traditional behavioral assays.

## 7. Notes

1. There are several powerful video- and image-processing programs that are open-source and available for download free of charge. These can be of tremendous help in acquiring trajectory data: *ffmpeg*: a powerful command-line video converter and editor (<http://www.ffmpeg.org>); *VirtualDub*: a video editor; can also convert videos to AVI (<http://www.virtualdub.org>); *ImageJ*: image editor; videos can be loaded as stacks of images (<http://rsbweb.nih.gov/ij>); *MtrackJ*: a plugin for manual tracking in ImageJ (<http://www.imagescience.org/meijering/software/mtrackj>).
2. If a reference image of the empty testing enclosure is not filmed there are two simple methods for generating one *post hoc*. First, image editing software (such as Adobe's Photoshop) can be used: take two images from the video in which the fish are in different parts of the tank; copy the white space from one image and paste it over the fish in the second image; include part of the tank edge in the image to assist in correctly aligning the two images. A second, somewhat more robust method is to average together multiple images from the video. This has the effect of smoothing out the moving targets and retaining the background: take multiple images from the video (at least 100) in which the shoal occupies different positions in the tank; for each pixel position, average together the red, green, and blue pixel values from all the images; use these means for the red, green, and blue values of the pixels in the reference image.
3. When two or more fish are partially occluding a single pixel clump may contain the images of all of them. When the clump is divided, some pixels may be incorrectly assigned. Since the targets are very close together by definition, this will have a minimal effect on the final position of the target. Nonetheless, to minimize this effect we split clumps along their shortest axis (the shortest line that goes through the clump centroid) which will often separate them into their component targets.
4. The periodogram is a Fourier power spectrum of a discretely sampled time-series. It expresses the power of each harmonic in the data as a function of its period. The periodogram of a time-series  $X(t)$  at period  $\omega$  is given by (31),
 
$$I(\omega) = \frac{1}{N} \left[ \left( \sum_{t=0}^{T-1} X(t) \times \cos[\omega t] \right)^2 + \left( \sum_{t=0}^{T-1} X(t) \times \sin[\omega t] \right)^2 \right],$$
 where  $N$  is the sample size.
5. The Lomb–Scargle test evaluates the significance of peaks in a periodogram. The following is taken from Hernandez ((29);

see also (30)). Define a test statistic  $\gamma = I(\omega) / 2\sigma_r^2$ , where  $\sigma_r^2$  is the variance of the time-series and  $I(\omega)$  is the value (power) of the periodogram at period  $\omega$ . Then  $P(\gamma > z) = 1 - (1 - e^{-z/2})^{N/2}$ , where  $N$  is the sample size and  $z$  is any value to be exceeded. Thus, a significant peak is one for which  $\gamma > -2 \times \log(1 - P_0^{1/n})$ , where  $n = N/2$ ,  $N$  is the sample size, and  $P_0$  is the required significance level (e.g., 0.05). Note that the value of  $\gamma$  depends on the variance of the data (and not only on the power of the individual periodogram peak) and thus the critical value for significance will vary between datasets.

6. During occlusions, the NND of all participating individuals will be 0 (as they are all given the same coordinates). We suggest excluding such cases from the analysis.
7. Each of the measures we present describes a univariate time-series. Thus, for each measure it is also possible to construct an autocorrelation function, though we have found these to be largely uninformative. However, the cross-correlations of the time-series' may reveal the structure of the shoal in greater detail. For instance, NND and IID will likely correlate more highly in a more evenly spaced shoal.

## References

1. Miller NY, Gerlai R (2011) Shoaling in zebrafish: what we don't know. *Rev Neurosci* 22:17–25
2. Wright D, Rimmer LB, Pritchard VL, Krause J, Butlin RK (2003) Inter and intra-population variation in shoaling and boldness in the zebrafish (*Danio rerio*). *Naturwissenschaften* 90:374–377
3. Wright D, Nakamichi R, Krause J, Butlin R (2006) QTL analysis of behavioral and morphological differentiation between wild and laboratory zebrafish (*Danio rerio*). *Behav Genet* 36:271–284
4. Dlugos CA, Rabin RA (2003) Ethanol effects on three strains of zebrafish: model system for genetic investigations. *Pharmacol Biochem Behav* 74:471–480
5. Echevarria DJ, Hammack CM, Pratt DW, Hosemann JD (2008) A novel behavioral test battery to assess global drug effects using the zebrafish. *Int J Comp Psychol* 21:19–34
6. Gerlai R, Ahmad F, Prajapati S (2008) Differences in acute alcohol-induced behavioral responses among zebrafish populations. *Alcohol Clin Exp Res* 32:1–11
7. Kurta A, Palestis BG (2010) Effects of ethanol on the shoaling behavior of zebrafish (*Danio rerio*). *Dose Response* 8:527–533
8. Krause J, Ruxton GD (2002) *Living in groups*. Oxford University Press, Oxford
9. Shaw E (1978) Schooling fishes. *Am Sci* 66:166–175
10. Pitcher TJ, Parrsiah JK (1993) Functions of shoaling behavior in teleosts. In: Pitcher TJ (ed) *Behavior of teleost fishes*. Chapman & Hall, London
11. Magurran AE, Pitcher TJ (1983) Foraging, timidity and shoal size in minnows and goldfish. *Behav Ecol Sociobiol* 12:147–152
12. Viscido SV, Parrish JK, Grunbaum D (2004) Individual behavior and emergent properties of fish schools: a comparison of observation and theory. *Mar Ecol Prog Ser* 273:239–249
13. Delcourt J, Becco C, Vandewalle N, Poncin P (2009) A video multitracking system for quantification of individual behavior in a large fish shoal: advantages and limits. *Behav Res Methods* 41:228–235
14. Ballerini M, Cabibbo N, Candelier R, Cavagna A, Cisbani E, Giardina I et al (2008) Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim Behav* 76:201–215
15. Miller N, Gerlai R (2007) Quantification of shoaling behavior in zebrafish. *Behav Brain Res* 184:157–166
16. Wu HS, Zhao Q, Zou D, Chen YQ (2011) Automated 3D trajectory measuring of large numbers of moving particles. *Opt Express* 19:7646–7663

17. Buske C, Gerlai R (2011) Shoaling develops with age in zebrafish (*Danio rerio*). *Prog Neuropsychopharmacol Biol Psychiatry* 35(6): 1409–1415
18. Gerlach G, Lysiak N (2006) Kin recognition and inbreeding avoidance in zebrafish, *Danio rerio*, is based on phenotype matching. *Anim Behav* 71:1371–1377
19. Cavagna A, Giardina I, Orlandi A, Parisi G, Procaccini A (2008) The STARFLAG handbook on collective animal behaviour: 2. Three-dimensional analysis. *Anim Behav* 76:237–248
20. Kato S, Nakagawa T, Ohkawa M, Muramoto K, Oyama O, Watanabe A et al (2004) A computer image processing system for quantification of zebrafish behavior. *J Neurosci Methods* 134:1–7
21. Tien JH, Levin SA, Rubinstein DI (2004) Dynamics of fish shoals: identifying key decision rules. *Evol Ecol Res* 6:555–565
22. Becco Ch, Vandewalle N, Delcourt J, Poncin P (2006) Experimental evidences of a structural and dynamical transition in fish school. *Physica A* 367:487–493
23. Inman AJ (1990) Group foraging in starlings: distributions of unequal competitors. *Anim Behav* 40:801–810
24. Chatfield C (2002) *The analysis of time series: an introduction*, 6th edn. Chapman & Hall, London
25. Miller N, Gerlai R (2011) Redefining membership in animal groups. *Behav Res Methods* 43(4):964–970
26. Buhl J, Sumpter DJT, Couzin ID, Hale JJ, Despland E, Miller ER et al (2006) From disorder to order in marching locusts. *Science* 312:1402–1406
27. Miller N, Gerlai R (2008) Oscillations in shoal cohesion in zebrafish (*Danio rerio*). *Behav Brain Res* 193:148–151
28. Aoki I (1980) An analysis of the schooling behavior of fish: internal organization and communication process. *Bull Ocean Res Inst Univ Tokyo* 12:1–65
29. Hernandez G (1999) Time series, periodograms, and significance. *J Geophys Res* 104:10355–10368
30. Frescura FAM, Engelbrecht CA, Frank BS (2007) Significance tests for periodogram peaks. *NASA Astrophys Data Sys Arxiv: 0706.2225v1*
31. Lukeman R, Li Y-X, Edelstein-Keshet L (2010) Inferring individual rules from collective behavior. *PNAS* 107:12576–12580