An interactive beat-pattern analysis of spermatozoa

Cambridge FertiliTEAM

18 February 2024

1 Introduction

Our solution meets the requirements of the Hackathon in the following ways:

1. Specific - We have brought to life the figures of the paper titled "*Computer-assisted beat-pattern analysis and the flagellar waveforms of bovine spermatozoa*", authored by Benjamin J. Walker, Shiva Phuyal, Kenta Ishimoto, Chih-Kuan Tung, and Eamonn A. Gaffney, providing interactive versions of all figures. These are easy to explore and give the user ample options to study and compare different spermatozoa waveforms.

2. Measurable - In addition to reproducing results from the source manuscript, we conduct further quantitative analyses. We investigate waveform reconstruction, wherein our live script outputs a 'score' based on the accuracy of waveforms generated by the user via an interactive interface. We also use the dataset for biophysical modelling of swimming spermatozoa, quantifying their kinematics based on flagellar geometries from the dataset.

3. Achievable - All the scripts included in our submission can run easily on standard laptops available to the public, students and researchers. The accompanying reports make it easy to understand the scientific contents of the manuscript.

4. Relevant - Our submission straddles multiple scientific disciplines–biology, statistics, fluid mechanics–and will help researchers in these fields communicate with each other. By reproducing the results of a scientific journal on an interactive interface, we are making that information more accessible to the research community.

5. Timebound - We have successfully reproduced and explained the manuscript's results. We have gone further and dissected the key features present in each of the figures. We now have a framework which can be used to arbitrary datasets on spermatozoa geometry, hence our work is easily extendable to a wide-range of flagellar beat-pattern analyses.

2 Summary of solution

Images of beating flagella of sperm cells, if decoded correctly, can provide crucial information about mammalian infertility, its causes, and cures. An important step toward understanding flagellar beat-patterns is to identify key morphological features that describe most of the observed data. While mathematical techniques to perform this identification exist, it is important to make these available to laboratories working with sperm cell imaging and analysis. By reproducing the results of an existing scientific manuscript and using publicly available datasets, we have developed a computer program for the analysis of experimentally observed flagellar waveforms. Our program accepts data extracted from images of sperm cells and lets the user interact with said data, by: (i) performing comparison between flagella of different sperm samples, (ii) suggesting the most important shape features of a given flagellar waveform and (iii) performing a comparative swimming analysis across selected sperm cells. In this way, our digital tool aims to simplify an otherwise involved process of investigating mammalian sperm cells, which can be used by a diverse scientific community.

3 Main components of the submission

3.1 Large-scale quantification of flagellar beating

We first reproduce Fig. 1 of Walker et al. (2020), illustrating how a real flagellar waveform is extracted from experimental data. We bring this to life through animation. We allow the user to explore the dataset further by watching the beating waveform of any subset of the 216 spermatozoa, enabling easy visual comparison of the sperm waveforms. We also animate synthetic waveforms, generated via the sampling method of Walker et al. (2020).

3.2 Quantitative analysis of waveforms

Walker et al. (2020) distills the waveforms of the 216 spermatozoa into a set of 'modes' or principal components using the method of principal component analysis (PCA). We create an interactive plot which allows the user to visualise how the the first principal component varies over time and across spermatozoa. The user is able to select the sample – A (fresh), B (thawed), or both – on which the analysis is based and toggle which statistics to display.

An alternative way to visualise the variation of the first principal component across the sample is via histograms. The histogram shows the user the frequency of the coefficient c_1 across the chosen sample at a specific moment in time, chosen with a slider. We allow the user also to visualise all the data at once as a pseudocolour plot.

3.3 Quantitative waveform comparison of samples

The livescript allows the user to view histograms of

- the constant contribution α and the sinusoidal contribution β to the decomposition $c_1 = \alpha + \beta\cos(kt - \phi)$ of the coefficient of the first PCA mode.
- beat period
- flagellar length
- maximum distal curvature
- swimming speeds

The 216 spermatozoa in the study are divided into two groups: Sample A (fresh sperm) and Sample B (frozen and thawed sperm, from a different bull). Sample B is further divided into two subgroups, 'blebbed' and 'unblebbed', depending on whether visible cytoplasmic blebbing is present in the spermatozoa.

In all histograms, the user has the option to select the samples from which the samples are constructed. The user may display results for samples A and B on the same histogram, thus reproducing Fig. 3 and Fig. 5 of Walker et al. (2020). The user may also choose to display results for only sample A or for only sample B separately for clarify of view.

Furthermore, the user may similarly view histograms for blebbed vs. unblebbed spermatozoa, allowing visualisations of quantitative results for these subsamples.

3.4 Waveform reconstruction

Consider the dataset of flagellar waveforms of bull spermatozoa cells, represented schematically as:

$$\left[\begin{array}{c} \{x_1, x_2, \,..\,, x_{1000}\}_1^1,\ \{y_1, y_2, \,..\,, y_{1000}\}_1^1 \\ \{x_1, x_2, \,..\,, x_{1000}\}_2^1,\ \{y_1, y_2, \,..\,, y_{1000}\}_2^1 \\ \vdots \\ \{x_1, x_2, \,..\,, x_{1000}\}_T^1,\ \{y_1, y_2, \,..\,, y_{1000}\}_T^1 \\ \hline \{x_1, x_2, \,..\,, x_{1000}\}_1^2,\ \{y_1, y_2, \,..\,, y_{1000}\}_1^2 \\ \{x_1, x_2, \,..\,, x_{1000}\}_2^2,\ \{y_1, y_2, \,..\,, y_{1000}\}_2^2 \\ \vdots \\ \{x_1, x_2, \,..\,, x_{1000}\}_T^2,\ \{y_1, y_2, \,..\,, y_{1000}\}_T^2 \\ \hline \vdots \\ \hline \{x_1, x_2, \,..\,, x_{1000}\}_1^N,\ \{y_1, y_2, \,..\,, y_{1000}\}_1^N \\ \{x_1, x_2, \,..\,, x_{1000}\}_2^N,\ \{y_1, y_2, \,..\,, y_{1000}\}_2^N \\ \vdots \\ \{x_1, x_2, \,..\,, x_{1000}\}_T^N,\ \{y_1, y_2, \,..\,, y_{1000}\}_T^N \end{array}\right]. \tag{1}$$

Here, $\left(\{x_1, x_2, \,..\, , x_{1000}\}_j^i, \{y_1, y_2, \,..\, , y_{1000}\}_j^i\right)$ are arrays containing the coordinates of 1000 points along the flagellum, relative to the sperm-head. The super-script i denotes the time, in the time-period of the waveform, at which the data was captured. This time-series runs from time-step, $t = 1$ to $t = T$. The sub-script j identifies the sperm cell number for which the data has been recorded (a total of N distinct sperm cells). The dataset can be found here, and corresponds to $N = 216$ and $T = 100$.

A principal component analysis (PCA) of the above data yields *time-independent* principal modes and *time-dependent* coefficients. The modes can be linearly combined–in relative proportions set by the coefficients–to reproduce all the waveforms in the time-series dataset. An example is shown in Fig. 1, where Fig. 1a represents the true (experimentally imaged) waveform of a sperm cell (identified by the 'Sperm ID' 68 in the dataset), and the rest of the figures represent waveforms obtained via the first mode (Figs. 1b), and via linear superposition of the first 2, 3 and 12 principal modes (respectively, Figs. 1c to 1e).

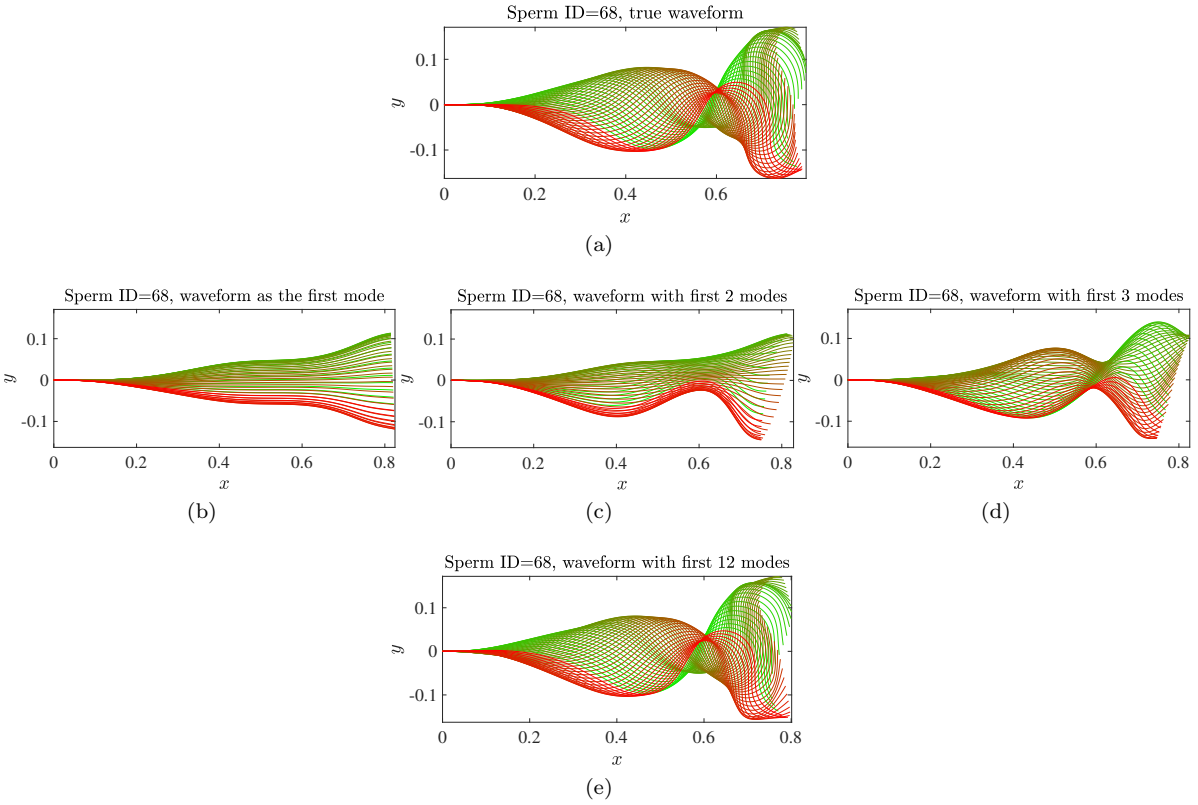

Figure 1: Improving reconstructions (panels (b) to (e)) of the true flagellar waveform given in panel (a), for a sperm cell corresponding to Sperm ID 68. The color coding denotes the temporal location of the waveform in its time period: green being the beginning of the time period and red, the end.

It can be clearly seen that increasing the level of 'modal approximation' (i.e., increasing the number of modes used to represent the flagellum) leads to a better reconstruction of the true flagellar waveform. More importantly, it can be shown using standard PCA concepts that the first 3 modes (Fig. 1c) can capture $\approx 90\%$ of the true waveform. Therefore, from here on we will give special emphasis to the '3-mode approximation' of a prescribed flagellar waveform:

$$\begin{aligned}\left(\{x_1, x_2, \,..\, , x_{1000}\}_j^i, \{y_1, y_2, \,..\, , y_{1000}\}_j^i\right) \approx & c_j^{\mathbf{1}}(t_i)\left(\{x_1, x_2, \,..\, , x_{1000}\}_{\mathbf{1}}, \{y_1, y_2, \,..\, , y_{1000}\}_{\mathbf{1}}\right) + \\ & c_j^{\mathbf{2}}(t_i)\left(\{x_1, x_2, \,..\, , x_{1000}\}_{\mathbf{2}}, \{y_1, y_2, \,..\, , y_{1000}\}_{\mathbf{2}}\right) + \\ & c_j^{\mathbf{3}}(t_i)\left(\{x_1, x_2, \,..\, , x_{1000}\}_{\mathbf{3}}, \{y_1, y_2, \,..\, , y_{1000}\}_{\mathbf{3}}\right) + \\ & \left(\{x_1, x_2, \,..\, , x_{1000}\}_{\text{mean}}, \{y_1, y_2, \,..\, , y_{1000}\}_{\text{mean}}\right),\end{aligned} \tag{2}$$

where, now $\left(\{x_1, x_2, \,..\, , x_{1000}\}_{\mathbf{k}}, \{y_1, y_2, \,..\, , y_{1000}\}_{\mathbf{k}}\right)$ represents the k-th principal mode/shape-form and the $c_j^{\mathbf{k}}(t_i)$ are coefficients that determine the amount of this mode's contribution to the true waveform. Notably, the coefficients change depending on the time t_i and the Sperm ID j of the waveform. The main advantage of the PCA is that it reduces the rather large dataset in eqn. (1) to a more compact form of eqn. (2): instead of now 'handling'/analysing 2000 data-points at each time instant (and for each sperm cell), one can just work with a set of three numbers $\left(c_j^{\mathbf{1}}(t_i), c_j^{\mathbf{2}}(t_i), c_j^{\mathbf{3}}(t_i)\right)$!

We allow the user to analyse, in three parts, the PCA decomposition of the true waveform. In the first part, we ask the user to specify the Sperm ID whose waveform they would like to analyse. This script then produces movies comparing the true (i.e., experimentally obtained) waveform to the '3-mode approximation'. For example, if the user were to enter '68' as the Sperm ID, they would be able to see movie versions of Figs. 1a and 1c and compare how the two waveforms differ at different instants in their time periods. We further give the user an option to specify the number of time periods for which they would like to observe the flagellar oscillations.

The second part allows the user to explore further the beat-pattern of the Sperm ID chosen at the previous step. At this stage, the user chooses a time-step (from the 100 possible time-steps for a given Sperm ID) and gets a chance to approximate the true waveform by changing the coefficients of the first three principal modes. A score is assigned dynamically to the user based on the average separation between their approximation and the best-possible 3-mode fit to the true waveform (see Fig. 2). The objective of this 'waveform-fitting game' is to minimize the score appearing on the figure.

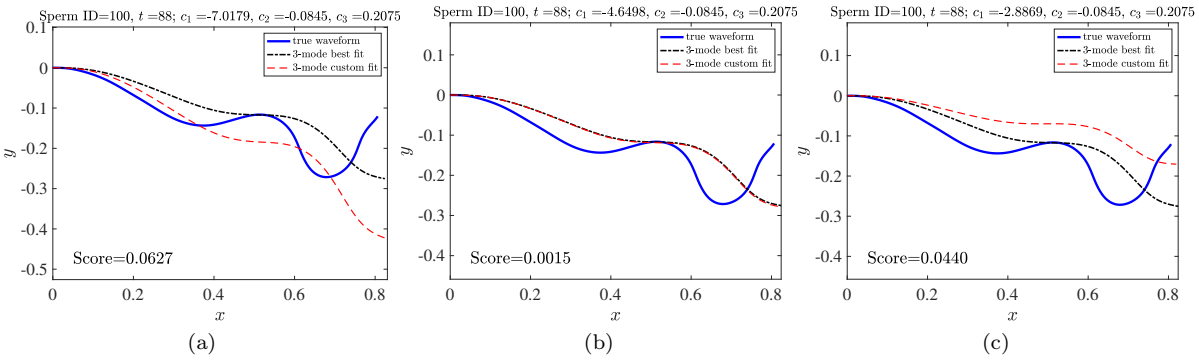

Figure 2: Demonstration of user-selected custom 3-mode waveforms (red, dashed line) and their respective scores (bottom left), depending on the 'average distance' of the custom waveform from the best possible 3-mode waveform fit (black, dash-dotted line). The solid blue line in all panels is the true waveform that is being approximated by the different 3-mode waveforms.

The final part of the livescript is an interactive demonstration of how increasing the number of principal modes leads to progressively better fits between the principal-modal-approximation and the true waveform (see Fig. 3). This is a step-by-step analysis of the equivalances shown in Fig. 1, i.e., the user can focus on a single waveform (from a host of waveform data based on their Sperm ID and time-step number) and analyse how it is reconstructed.

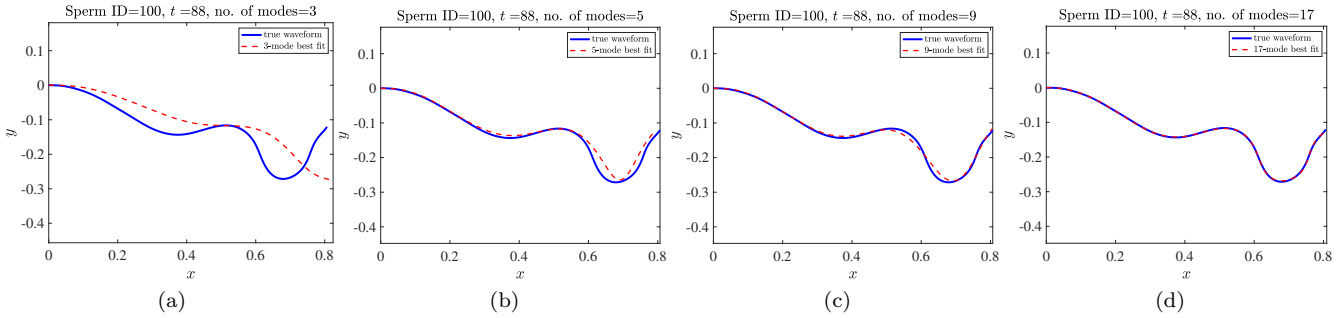

Figure 3: An illustration of a progressively better fit to the true waveform, obtained by increasing the number of modes of the PCA.

In this way, by choosing different waveforms, the user can explore their relation to the principal modes, and see the effectiveness of PCA in reconstructing time-varying flagellar beat-patterns.

3.5 Biophysical modelling

As a demonstration of biophysical modelling using the dataset, we reconstruct the swimming kinematics using a calculation based on resistive force theory. This is a well-established framework to solve for the low-Reynolds number fluid dynamics of a slender filament, such as the flagellum of a spermatozoon, and involves an anisotropic

proportionality relationship between the force density $\mathbf{f}(s,t)$ along the flagellum and the local velocity $\mathbf{u}(s,t)$,

$$\mathbf{f}(s,t) = \xi_{\parallel}\mathbf{u}_{\parallel}(s.t) + \xi_{\perp}\mathbf{u}_{\perp}(s,t), \tag{3}$$

where $\mathbf{u}_{\parallel}$ and $\mathbf{u}_{\perp}$ are components of $\mathbf{u}$ tangential and normal respectively to the flagellum. We use the resistance coefficients $\xi_{\perp}$ and $\xi_{\parallel}$ reported for swimming sperm in Friedrich, B M et al. (2010) The Journal of experimental biology vol. 213,Pt 8: 1226-34. doi:10.1242/jeb.039800.

We model the sperm head as a prolate spheroid of major axis 10 µm and minor axis 5 µm, and use drag coefficients as calculated in Chwang & Wu (1976) Journal of Fluid Mechanics. 75:677-689.

The local velocity (in the lab frame) $\mathbf{u}(s,t)$ of the flagellum is a sum of undulations in the swimmer frame (as computed from measured waveform data) and a resultant global translation and rotation (to be determined). Using RFT for the flagellum and the drag coefficients for the head, we compute the forces and torques exerted by the flagellum and the head, and solve the conditions of no net force and no net torque to determine the kinematics at each time instant.

We present these results in an interactive livescript. We animate the swimming of the sperm, firstly for a relatively small number of periods and on length scales comparable to the spermatozoa, illustrating the details of the oscillatory swimming trajectories along a mean swimming path.

We further implement 'a spermatozoa swimming race', in which we animate, on a larger length scale, the swimming from the same starting point of any subset, chosen by the user, of the 216 swimmers. This allows the user to explore the longer-timescale swimming dynamics of different spermatozoa, and for instance, identify especially competent swimmers. We thus create a fun and interactive platform for the user to explore the results of rigorous scientific modelling.

We calculate for each sperm in the dataset the average swimming speed, defined as the swimmer displacement over a beat period divided by the beat period. These results are presented as a histograms in the live script, for sperm samples/subsamples of the user's choice.