Reviewer 1

I applaud the philosophy adopted by this type of ms – that more can be made of the signals we have from biologgers than we currently do – so that's a great start. Against that, the actual length of time when a (single) captive whale could be assessed was just seconds, and it occurred under optimal conditions. The free-living whale data is, in some senses, better (though the captive work is needed to make sense of them) but is noisy (the authors recognise this as an issue). Overall therefore, the ms documents a rather complex procedure that may or may not work in other cetaceans and situations. I think that it is still worth publishing (because people ought to try to maximize the value of their data) but it would be helpful if the authors were more forthright about the shortcomings of their work and provided more detail. Specific comments are detailed below.
Thank you for recognizing the aims of this paper and raising concerns about generality. We addressed this issue by providing more details about the decisions involved in signal processing and adding a section to the discussion ("Limitations and considerations for future applications") that we hope clarifies things.

Line 27-28. – 'related variables' is rather vague.
Added examples like wing beat frequency and feeding rate (line 28)
Line 53 – Perhaps better to say; 'the so-called IJK complex' (which is…), and which occurs during systole.
Added more details about the IJK complex (lines 54-58)
Line 59 – 'when cetaceans are biologged in the wild' is perhaps better. However, some studies on smaller cetaceans have captured animals to place tags so I suggest you put in 'most cetaceans'.
Rewritten for clarity (lines 62-63).
Line 53-4 – 'produce heartrates that are statistically equivalent to instantaneous ECG heartrates'.
Changed (lines 67-68)
Line 74 – Where did the ECG tag come from?
We now state the ECG tag was custom-built and described by Bickett et al. (2019) (lines 78-80).
Line 76 – Please state that you recorded tri-axial acceleration (orthogonal) and give measures of accuracy and precision of the sensors.
Updated. We added "tri-axial" on lines 81-82, provided the IMU part number on lines 83-84, and specified the range, sensitivity, and accuracy of the accelerometer on lines 84-85.
Line 86 – with regard to the whale resting at the surface and the animal in the wild, what about noise affecting the accelerometer signal (see comments later)?
See response with "**" on the next page.
Line 99 – Please be more specific about 'windowed operations'.
More detail added (lines 107-112)
Procedure – I think this would be better if written in the past tense as part of the methods. Thus, 1. Would be 'We removed… etc.'
Changed tense.
Line 104 – Please specify here that you differentiated up to all 3 signals. Over how many data points (time interval) did you differentiate? Also, please state that you therefore derived 'jerk' (and give units).
Added more details about the differentiation procedure (lines 122-132). Clarified the distinction between differencing and the derivative, so the resulting quantity is not

technically jerk. Regarding multiple axes, we specify which steps of the procedure were applied tri-axially on lines 158-160.
Line 110 – Please give more information on this procedure.
Expanded (lines 138-144).
Line 111 – Please be specific about your heuristic removal of noisy peaks. We understand that it won't be perfect but we still need to know.
Expanded (lines145-157).
Line 124 – I guess that the value of gyros (and accelerometers) for detecting strokes will depend critically on the position of the tag. You only have one instance of data from a free-living (blue) whale and one from a captive (killer) whale – where you could position your tag precisely - so your ability to define the value of your IMU sensors to define body motions (such as fluking), will be accordingly limited. This should be noted somewhere. Note also that your Gough et al. citation is incomplete.
**Detecting fluke strokes from gyroscope signals is considered robust to tag position, which we now clarify in the text (lines 172-174). Thank you for catching the error in the citation, now fixed.
Lines 127-130 – This seems ok but the Goldbogen et al. (2019) study was only conducted on 1 individual so your approach allows for very little inter-individual variability. I'm not disputing the value of what you have done but it is good to have the community appreciate potential limitations.
We added a caveat about sample size and inter-individual variation. (lines 180-182)
Discussion – This is extraordinarily thin. There needs to be some serious consideration of how conditions on the whales(s) affects the acceleration noise. This is important for others aspiring to use the approach but also to speculate (at least) how fluking/waves/speed/body position etc. might affect the viability of the technique.
Thank you for this feedback. We added a section ("Limitations and considerations for future applications") to the discussion to address these issues.
Fig. 1 – There are only scale units for the ECG in the y-axes. This needs correcting. The 'nearly identical heart beat predictions' between the two systems could be displayed in an x-y regression (e.g. of heart rate beat period or timing match/mismatch).
Scale units were deliberately excluded. In the figure captions, we provide the rationale, "Y-axis scale units excluded because the filtering process introduces magnitude distortion and only the relative shape of the signal is relevant to the analysis."
We added an x-y regression figure (now Fig. 3C)
Fig. 2 – There are no scale units anywhere in the y-axes.
We deliberately excluded y-axis scale units because the signal processing procedure introduces magnitude distortion; only the shape of the signal is relevant to the analysis. This is specified in the figure caption ("Y-axis scale units excluded because filtering introduces magnitude distortion and only the relative shape of the signal is relevant to the analysis.").

The references are not standardized and some are incomplete.
We corrected the incorrect bibliography entries.

Reviewer 2
Major comments
This is a really exciting paper, for two reasons: (i) it opens up a possibility for existing datasets to have more value squeezed out of them across the broader disciplines of animal biologging and physiology, (ii) it provides a totally novel option for recording heart rate in

wild animals at liberty, and (iii) it has the potential to catalyse a change in animal welfare with regards to physiological data collection. I think this work is likely to be very impactful, to make people think a bit differently about what they have already collected, and what they plan to collect in the future. This is certainly the overall direction our field must move in, and I find this an exciting first step. I really commend the authors for the work. I also commend their open practice and code sharing – great work for the larger community and transparent science!

Thank you for these kind comments!

I have little to request as amendments on this, but my main comment is that there are some plots I'd really like to see, and some others I would move around. The main plot I was expecting to see was a scatterplot of ECG-logger derived heart rate and accelerometry-derived heart rate, with the deviation from slope=1 shown. I realise on reading carefully that this is probably because the sample size from the orca was just 14 seconds (which is 14 beats, or I suppose 14 equivalent heart rate estimates) but I still think it would be good to see, and the text at line 116 refers makes you want to see such a plot. Also, as below, it would be good to see the signal and noise plotted from the analyses, and two of the supplemental figures are so good, I'd like to see the in the main text.

We reworked the figures to include these items.

How does the IJK complex differ from the QRS complex? I am familiar with the latter but never heard of the former – perhaps just a few sentences on this would help.

We clarified the similarities and differences between the two waveforms (lines 53-59)

Figures S2 and S3 – these figures are great and are quite essential to see the blue whale data for ourselves. I think they could be combined and added to the main document and would strengthen it well. I think they're just great.

Figures S2 and S3 are now part of figure 2. Fig. S2 is now Fig. 1A and Fig. S3 is now Fig. 1B-C. Panels A, B-C, and D-G are temporally nested, which is now indicated with red dashed boxes.

Minor comments

A few statements throughout that could use supporting references, e.g. line 33, 43, 55, 105, 177.

Supporting references added to 43 (now 45), 55 (now 59), 105 (now 129), and 177 (now 265).

Re: 33. We think the assertion "simpler IMU tags have fewer logistical constraints and provide access to more species and larger sample sizes" is supported by the subsequent text (lines 36-41). For additional context, ECG and heart rate have only been recorded in one medium-sized wild odontocete, *Monodon monoceros* (Williams et al. 2017 *Science*), and one small odontocete, *Phocoena phocoena* (McDonald et al. *submitted*). Both required capture and release of the study organism. Conversely, one study (Goldbogen et al. 2019 *Science*) analyzed 300 IMU tag deployments on 13 cetacean species, and the majority of tags were deployed on unrestrained animals.

Line 59 – I think this could be written more clearly / rephrased.

Thank you for the feedback. We now distinguish the 1D and 3D BCGs more explicitly and describe why uncertain tag placement is relevant. Hopefully this is more clear (now lines 60-69).

Materials and methods – please add how the tags were attached to both the killer whale and blue whale (suction cups?)

Correct, suction cups. Now see lines 81 and 94

Line 76 – just a thought rather than a request, but this did make me wonder what resolution the majority of studies out there record at – there will be many that don't record at a high enough frequency to extract BCG, and would be good to give a threshold cut off, but that would require you to review the whole body of literature and I wouldn't expect you to for this article.

We address this and other concerns for future studies in the added section "Limitations and considerations for future applications". Specific to this thought, see lines 238-247.

Line 82 – give detail of tag used (it just says "tag" but give make, model etc).

Added details about the ECG tag (lines 78-80)

In both killer and blue whale cases, I assume these are archival tags that you had to physically recover and download the data from? Please add this detail.

Correct, they were archival. Added detail on lines 77 and 94.

Line 99 – "windowed operations" might need either a reference or a very quick explainer.

Added more details on lines 107-112.

Lines 99 and 102 – why different data length and band pass filter for the killer and blue whales?

The different window sizes are now justified on lines 108-112. We also added more details about the band pass filter cutoff frequencies on lines 116-121.

Line 110 – triangular moving average smoother – why this type of smoother? And maybe add a supporting reference? To keep this article succinct, this detail could be added to the supplementary material, but I think it's important to state why somewhere.

More details about the triangle moving average smoother added in lines 138-144.

Line 111 – does this mean some were manually removed?

Details of this procedure were described in the supplemental, so we moved the most relevant details into the main text (lines 145-157)

Line 114 – as per line 110 – I think perhaps a slight expansion of the rationale for some of this could go into supplementary materials perhaps?

Our explanation is in lines 123-130 of the new manuscript

Line 117 – insert "heart rate" after 'ECG-derived'.

Done (now line 163)

Line 130 – This made me want to see something like smoothed histograms for signal and noise so we could see for ourselves how well the data fit the expected range. Can you add a figure like this (could also annotate on the band pass filter etc)?

This is similar to what we show in Fig. 3a. Except instead of smoothed histograms we showed the power spectrum, which we think is stronger evidence. The shaded portion under the curves corresponds to the signal frequency band.

Line 136 – as above, figure S2 is just great, and I think it is needed in the main document.

We moved figure S2 into the main manuscript. Now panels B and C in Fig. 2.

Line 146 – this is why we need a scatterplot of the ECG and accel derived heart rates I think.

Added to Fig. 3C.

Line 146 – why was the period of data collection so short? I do appreciate how very hard this work is to do, and 14 seconds appears to have been more than enough, but I wonder if a sentence about it might not be a good idea?

Valid point. Access to cetaceans in managed care is, of course, limited and the experiment for this project had to be scheduled around other, higher priority projects. We were able to collect one session's worth of data, which we present here. We planned for additional sessions but, unfortunately, logistical issues prevented us from collecting more data. We added a sentence acknowledging this limitation (203-205).

Line 160 – the heart rate relationship in the plot that uses the Theil-Sen regression line clearly isn't linear, so perhaps not best to use a linear model (at the same time, thank you for introducing me to this regression, which looks like we should all be using it for linear relationships in place of standard lm().)
We agree that a polynomial or other model would fit these data better. However, our goal for this analysis was not to make out-of-sample predictions; rather, we intended to demonstrate a positive trend in heartrate over the course of dives. Therefore, we believe a linear model is valid for this purpose. (We are happy to hear you appreciate Theil-Sen regression. Credit to Markus Horning for using it in his 2012 paper in Frontiers in Physiology.)
Line 164 – I would delete from "Publishing data and code…critical step towards" to just shorten this bit a little.
This section cut from the results and combined with the related section in the methods.
Line 174 – as per my comments above, I think the ethical / 3Rs gains from this study are considerable too, and I wonder whether it's worth adding a sentence to that effect?
Thank you for this feedback. Related to the subject of animal welfare, we added text about applications to conservation physiology, using sonar disturbance as an example (261-264). Please advise if you'd like to see something more explicitly about the 3Rs.
Figure 2 legend – I'm not quite sure I understand "forward differencing acceleration" – as above this could go in supplementary methods.
We added "(see section **Procedure**, step 2)" to the legends of figures 1 and 2 to point the reader to the relevant part of the methods.
Figure 3 – I would add "3D" and "1D" to the top left of the plots in (A).
Done

Lucy Hawkes, January 2022

Additional comments from Lab member:
Possibly highlight high frequency accelerometer - high resolution mentioned in the introduction but could define a threshold (i.e. over 20 Hz?)
We did not perform an analysis to quantitatively identify a threshold for accelerometer frequency, but we added a paragraph about what to consider in future studies (238-247).
Line 82: change reference formatting to "followed the methods in Bickett et al (2019).
Thank you for catching that, done.
Line 98: date in brackets for the reference.
Done
Line 99: Why the difference in windowed operations, what do they mean? What is the smoothing window value? Maybe add Shepard et al 2008 ref.
We did not provide sufficient detail about windowed operations (and several other parts of the methods) in our initial submission. Please see revisions at lines 107-112.
Line 101-106: Maybe more description to what the filters are/how they work?
We added details about the bandpass and S-G filters. See lines 114-132.
Line 107: which acceleration axis is k (x, y, z?). All 3 axes for the 3D model?
k represents the set of axes in the summation, so in the 1D case it's {surge} and in the 3D case it's {surge, sway, heave}. We didn't explain that sufficiently in the original submission. Please see lines 134-137.
Methods: Duration of tag deployment for both the orca and blue whale could be moved here?
We are willing to do so if you think it's necessary.
Don't mention how Jerk is calculated from acceleration data, more details required here as cannot be reproduced

Excellent point. On review, we changed the terminology from “jerk” to “differenced acceleration” because we didn’t take the derivative with respect to time. In other words, jerk would technically be $(a_{t+1} - a_t)/\Delta t$, but we didn’t divide by Δt. Furthermore, we used a Savitzky-Golay filter to do the differencing step $(a_{t+1} - a_t)$. The method details and rationale are now better described on lines 122-132.

Two reproducibility sections - just merge both at the end of the manuscript?

We merged them into the methods. We can move them to the Results+Discussions if that reads easier.

Would be useful to know if could get similar results at lower acceleration frequencies (what minimum required to get a valid signal?). How replicable is this study if only requires acceleration to be recorded at extremely high frequency? Would it work at 100 Hz? 50 Hz? Would this only work on cetaceans, or would this method work on teleost/elasmobranchs?

Please see the new paragraph on lines 238-247 for thoughts about sampling frequencies. Whether or not the results are replicable at lower frequencies is an interesting question and the subject of future work. In theory, it should work with other marine vertebrates, since the physical phenomenon that generates the BCG waveform (i.e., heart contraction and ejection of blood into the aorta) is shared throughout the clade. That said, most teleost/elasmobranch bio-logging studies that I’m familiar with attached the bio-logger to the dorsal fin. I doubt the BCG waveform is readily detectable when the accelerometer is on an appendage rather than the torso.

Maybe more details about practicalities of understanding heart rates of cetaceans, physiology, energetics…

Y axes values missing in the figures?

Scale units were deliberately excluded. In the figure captions, we provide the rationale, “Y-axis scale units excluded because the filtering process introduces magnitude distortion and only the relative shape of the signal is relevant to the analysis.”

Supplementary figure 1: A) x axis label missing

Fixed, thank you for catching that. We also moved some other text in the figure to fix overlap and clipping issues.