

Reviewer comments and our point-to-point response

Reviewer B:

The authors use a method based on the Region Of Interest (ROI) technique that had been previously developed to automatically detect whale calls on Passive Acoustic Monitoring (PAM) spectrograms. Here the ROI algorithm is tuned to focus solely on detecting T-phase events generated by underwater earthquakes, from hydrophone spectrograms.

The authors did a good job at analyzing a challenging hydroacoustic dataset containing different noises due to active acoustic experiments (transmission of M-sequence pings) and periods with strumming likely caused by deep currents as well as other sources (whale calls and shipping noise). In addition, the PAM dataset is not continuous but follows a preset duty-cycle : hydrophones only record for 45 minutes every 12 hours. Within the 1476 hours (~2 months) of raw data, 96 T-phase events were manually labelled by an expert annotator (the main author). The automatic detector was then tuned on this set of events. A compromise was chosen between limiting the number of False Positives (100) and False Negatives (71). This new detection method is proposed for preliminary screenings of hydrophone recordings in order to accelerate annotation of T-phase events.

In addition, the authors extract acoustic features for the extent and centroid of each ROI shape ($tstart$, $tend$, T , $fmin$, $fmax$, Δf , tc , fc , Atf) and present an analysis of the resulting parameters for their PAM dataset.

This article shows a good description of the state of the art on passive acoustic studies and explain existing detection methods used for seismological purposes. The authors compare their new methodology (ROI) to two existing automatic methods for detecting T-phases: the seminal work based on the Gradient Boosted Decision Trees technique published by Sukhovich et al. (2014) and a recent CNN TISSNet model based on spectrograms published by Raumer et al. (2024).

Conclusively, the authors propose this new method that they refer to as “a robust solution to further explore earthquake properties generating T-phase signals, particularly for remote regions with diverse seismic activity.”

We would like to thank Sara Bazin, Reviewer B, for thoroughly reading through our manuscript and providing very useful suggestions. We also appreciate her summary above, acknowledging our efforts in handling difficult hydroacoustic data, repurposing the ROI algorithm for seismic signal detection and our review of the state of the art of hydroacoustic detection of seismic activity.

In particular, Sara’s first four comments (two major concerns, two minor) have highlighted weaker parts of the original draft. These suggestions have helped us greatly in strengthening the manuscript, including some additional analysis and expanding on topics in both the main text and Supplementary Material. Other comments have been helpful in addressing parts of the manuscript where we perhaps needed to elaborate further on our methodology and results, along with adjusting sections to better suit the target audience (seismologists).

We have replied to each comment individually below. Line numbers provided in the replies refer to the file “[Seismica_Revised_MAIN.pdf](#)”.

I have **two major** concerns:

1) **The first concerns earthquake magnitudes.** Unlike for bioacoustics studies based on whale calls, the amplitude of a T-phase events is an important feature in seismology. Biologists can recognize whale species based on the shape and frequency content of

individual calls. Indeed, the amplitude in dB of a whale call does not help to classify the animal. On the contrary, size matters in seismology. Surprisingly, the amplitude of the T-phase (in dB) is not exploited in the proposed analysis. I will suggest several additions to the current study.

Authors' reply:

We thank Reviewer B for raising this subject. We have included some additional analysis in Section 4.6 (Lines 394-413) presenting Received Levels measured for every T-phase signal identified as an ROI, both annotated (*TP*) and newly discovered (*TP_new*). We have also included some remarks on how Received Level values compare across the signal qualities assigned to the T-phase signals during the manual annotation stage, along with commenting on the ranges of Received Levels encountered.

Received Levels are currently all that we can report on, as we are unable to provide estimates of earthquake magnitudes. This is due to the unique acoustic propagation conditions of the Arctic Ocean and the lack of a robust model that we can use to locate the earthquakes with certainty using the time-of-arrival of the T-phases alone, especially considering the unfavourable geometry of the three moorings considered in the study. This latter aspect will be the focus of an upcoming study. We have elaborated further on acoustic propagation in the Arctic Ocean in the Introduction and Discussion sections (Sections 1 and 5), along with commenting on the implications the local bathymetry and currents may have on the T-phase signals in the Nansen Basin where the recordings were taken (Lines 62-72, Lines 477-485).

2) The **second concerns benchmarking**. The authors mention the superiority of their new method compared to the existing ones, but there is **no real comparison**. Furthermore, the proposed method has not been tested in a different soundscape context. Lastly, the authors claim the new analysis is **computationally efficient** but there is no details provided (computer type, duration of each run per data length, etc). I would suggest the authors to use existing benchmarks to quantify the superiority of the new ROI method. One simple way would be to apply the proposed ROI method on a publicly available hydroacoustic dataset such as the one that was set up by Raumer et al. (2024) in "An Open Source Hydroacoustic Benchmarking Framework for Geophonic Signal Detection" in **doi:10.26443/seismica.v3i2.1344**. It would not take much work as the data format is similar to the NERSC format and it would be a valuable addition.

Authors' reply:

We thank Reviewer B for suggesting to test our ROI algorithm on the manually labelled acoustic datasets, provided by IFREMER (Brest, France), and for providing the instructions on where to access the .wav files. The outcomes of this test have been included in the Supplementary material (Section S8, Figures S5-7).

The ROI algorithm has been run across the HYDROMOMAR-2013 and OHASISBIO-2020 datasets (accessible online). **All 82 T-phase signals that were manually annotated confidently by the original inspectors of the datasets were spotted by the detector as ROIs.** We have plotted the distributions of the acoustic features of the T-phases detected, as shown in Figure S7: duration, maximum frequency, bandwidth and spectral centroid. **These match the distributions presented in the main text on the CAATEX dataset** (Figure 6). Note that we didn't make any changes to the setup of the detector (same script and parameters), using the exact same procedure described in Section 3.3.

Many of the ‘uncertain’ events in the manual annotation datasets that looked like T-phases were captured by the ROI algorithm; we have mentioned this and included an example of a spectrogram featuring these signals (Figure S5). The ROI acoustic features for these ‘uncertain’ events have been computed, but further analyses have not been completed.

We noticed that the original amplitude thresholds used for the binarisation mask stage of the detector, when applied to the CAATEX dataset, were perhaps too low compared with the HYDROMOMAR-2013 and OHASISBIO-2020 datasets (Figure S6). As a result, the detector was more sensitive to background noise, which was not fully suppressed during the Gaussian filter stage. Remnants of background noise across the spectrogram were often picked up as ROIs, many of which were very short duration (< 10 s) and narrowband (< 5 Hz). Likely, as amplitude thresholds were unoptimised for the IFREMER datasets, there were also several instances of annotated T-phases being merged with other signals (unverified T-phases or non-seismic sounds) (Figure S6).

A summary of this benchmark study has been included in the Discussion section (Lines 547-554). We will also include findings from this benchmark study as additional supplementary material as a .csv file containing the acoustic features of the known T-phases (+ notes of peculiarities of certain events, e.g. T-phases merging with other sounds when processed).

Details about the computer used for the main study and processing time for an individual 45-min recording from the CAATEX dataset have been added in Section 3.3 (Lines 221-224).

I have two **minor** concerns:

1) I think the authors should take advantage of the limited number of automatic detections (301) to make a detailed analysis of the *FP* (100). Figure 8 would be quite useful for that.

Authors’ reply:

We have included a table describing the various sounds that caused the False Positive detections in the Supplementary Material (Section S4, Table S1). We have also added more information in the Results (Lines 369-379) and Discussion (Lines 498-512) sections. Overall, we have only managed to identify strumming, which we now more conservatively call “flow noise”, and a single instance of wind-ice related noise resulting in a False Positive. The sources responsible for the rest of the False Positive detections remain unknown.

2) **16** earthquakes were reported by the ISC whilst hydrophones were actively recording, but this information **is not exploited in the article**. For example, the authors do not explain how many of the 96 T-phase events relate to the ISC earthquakes. Adding the ISC events (with their magnitude information) to Figure 4 would be useful and would improve the present analysis.

Authors’ reply:

Thank you for this suggestion. We compared the CAATEX recordings a second time against the 16 earthquakes reported by the ISC and have now been able to associate T-phase signals to 5 of the ISC-reported earthquakes, rather than just 2 stated previously. Interestingly, of these 5 events, only 3 had T-phase signals that were assigned Event IDs, the other 2 events had T-phase signals that were discovered by the ROI algorithm. We have included an extra section in the Supplementary Materials (Section S5, Table S2) detailing the data provided by ISC for these earthquakes and added extra lines in the Discussion section

(Lines 464-477) and in the caption of Figure 1. We think that Figure 4 already includes enough information, and we preferred not to change it.

Detailed comments:

- Line 21

The authors should specify what they mean by “minimal computational resources”.

Authors’ reply:

We added a paragraph at the end of Section 3.3 (Lines 221-224) to detail the computer used and its processing power, along with how long an individual file took to process.

- Lines 31 to 34

The authors explain that deploying OBSs is challenging because of ice coverage and that hydrophone arrays are a good alternative tool. Can you explain why hydrophones are less challenging to deploy than OBSs in this context?

Authors’ reply:

We added extra lines in the introduction section (Lines 33-36) to clarify the potential limitations in deploying and recovering OBS units in an ice-covered environment.

- Line 37

I do not think “mainly” is necessary here.

Authors’ reply:

This sentence has been modified (Lines 44-45).

- Lines 59-60

I would replace “movements of the ridge” by “ridge activity”.

Authors’ reply:

Amended with “activity along the Gakkel Ridge” (Lines 81-82).

- Lines 63

Raumer et al. (2024) is not an appropriate reference as it reports only automatic detections.

Authors’ reply:

The reference has been removed here (Lines 84-85).

- Lines 65-66

I may suggest to rephrase “To identify earthquake phases recorded by seismometers” by “To pick time-arrivals of seismic waves on seismograms”.

Authors’ reply:

This sentence has since been removed as we decided to just focus on hydrophone networks in this paragraph (Lines 84-90).

- Line 120:

maybe “depth” is better than “deep”.

Authors’ reply:

This has been corrected (Lines 144-145).

- Lines 119-121

I do not understand why the authors write “was selected”. Are there several hydrophones on each mooring?

Authors’ reply:

The CAATEX dataset is part of an acoustic experiment which included an array of multiple hydrophones placed in each of the 3 moorings. For our study, only the recordings from the deepest among the hydrophones have been used. The justification for that is to focus on signals with better SNR, reducing the noise from the water surface and ice. It has been previously shown that the ice canopy is one of the drivers of sound reverberation and refractions in the Arctic Ocean; hence, T-phase signals may appear more obscured by reverberant trails recorded at shallower depths.

- Line 154

The filtering step smoothing the spectrogram means that the method can detect events that have various shapes (for example solely T-phase or a P and T phase). But it will assign it the same label : earthquake. One drawback is that the *tmin* feature could be a P arrival or a T arrival. Therefore, *tmin* cannot be directly used for earthquake localization. Further analysis would therefore require an additional manual step. It could be worth mentioning this drawback here.

Authors’ reply:

We have added an extra paragraph to expand on this (Lines 203-211). With spectrogram-based detectors, a compromise may be needed in preserving temporal resolution vs spectral resolution; hence, a 1 s temporal resolution is likely no good for time of arrival estimates to begin with. We mentioned that while the ROI algorithm was tuned to identify T-phase signals, it was possible for P- phase signals to merge into the ROI during the binarisation mask stage, stemming in part from the Gaussian smoothing stage, which slightly shifts energy across bins in the spectrogram. We have also summarised these points in the Discussion section (Lines 529-532) that the detections made by the algorithm cannot be used for phase-picking by itself.

- Line 170

The authors applied a trick to suppress any non-seismic sounds of shorter duration or narrower bandwidth (e.g. whale calls). What is the shortest T-phase that can be detected then?

Authors’ reply:

The Gaussian filter is a key preprocessing step rather than a trick before running the binarisation mask step of the ROI algorithm. The current setup of the Gaussian filter may partly suppress T-phase signals with a duration < 5 s or bandwidths < 2.5 Hz. However, it would be difficult to verify whether such events are indeed seismic, especially if the event was both short in duration and bandwidth.

The Gaussian filter is applied across both axes of the spectrogram, so a short-duration T-phase signal with a wider bandwidth, suggesting a close, strong event, would likely not be erased entirely. Admittedly, in our CAATEX dataset, we do not have an example to really define the lower limit of duration for detectable T-phase signals. The current pre-processing procedure effectively suppresses most sharp transient sounds (e.g., from sea ice) and tonal

sounds (even if they are frequency modulated). Additional lines have been added in Section 3.3 of Materials and Methods (Lines 199-201) and Discussions (Lines 491-493).

- Line 191

The performance of the automatic detection is presented with the usual metrics (precision and recall) regularly used to estimate the success of ML methods. However, here, because of the small size of the dataset, there is no separation between the training dataset and evaluation dataset. This situation is quite unusual and the authors should better warn the readers about this.

Authors' reply:

Precision and recall are performance metrics from information theory and general confusion matrix problems, and predate Machine Learning. We appreciate that there is a focus on ML methods in research at the moment, so these metrics are highly associated with ML, but they are and have been used alongside non-ML methods.

Additionally, we have made a statement about True Negative events, which are indeed missing in our analyses (Lines 234-236). In our case, they don't serve a statistical purpose and it is unreasonable to count them, as they constitute for all non-seismic sounds, including background noise, the bounds of which cannot be discretely defined in a spectrogram.

Regarding the confusion about training/validation dataset, we feel that there is not a point in the text where the reader will get confused about our method, it is not Machine Learning based, so there is no need for a training dataset. We have reiterated at the start of Section 3.4 (Line 226) that we run the algorithm across our entire hydrophone dataset, then compare our findings to the manually-annotated dataset.

-Line 283

The explanation about the strumming noise and the correlation with *FP* is interesting and it might indicate a seasonal effect. It may be worth mentioning what is known about the local oceanic currents and the effect of the bathymetry in the area.

Authors' reply:

We have included some new sections in the Supplementary Material: S6, where we describe the water conditions and currents around the measurement region; and S7, where we summarise how we associated these continuous tonal sounds with flow noise. The presence of 'flow noise' coincided with increases in hydrostatic pressure, measured by sensors on the hydrophone moorings.

The CAATEX moorings were placed near the Yermak Plateau (NERSC2 being the closest), where there is some complex water movement, but we do not discuss this in detail; it is outside the scope of this study and will be covered in future publications by our co-author, Vår Dundas. In the main text, when mentioning flow noise, we refer to what is written in S6 and S7, in Section 4.5 of Results (Lines 350-358) and in the Discussion (Lines 502-508).

- Lines 358-359

It is uneasy to discuss about the "weakest earthquake signals" when there is no amplitude quantification. The authors mention that small magnitude "would appear on spectrograms with shorter durations and narrower bandwidths". However, we know that the correlation between T-phase duration and earthquake magnitude is not straightforward. Instead, I would really advise to quantify amplitude here.

Authors' reply:

The amplitude of a recorded T-phase depends on source strength, source depth, and source-receiver distance. In previous studies (e.g., Dziak, 2001; Fox et al., 2001; Pulli and Upton, 2002; Sukhovich et al., 2014; Royer et al., 2015; Guardato et al., 2022), T-phases could be linked to located earthquakes, allowing event locations, source-receiver distances, and source levels (*SL*) to be estimated. *SL* is a measure of the energy released into the water column at the point of seismo-acoustic conversion and serves as a proxy for hydroacoustic event magnitude.

In our case, this is not possible because the CAATEX data consist of only two 45-minute recordings per day. As a result, only 5 of the 16 detected T-phases could be associated with earthquakes listed in the ISC catalogue. Since locations are unavailable for most events, T-phase amplitudes cannot be used to infer earthquake magnitudes.

Although we calculated the maximum received level (*RL_{max}*) for each true-positive detection, source levels cannot be estimated without knowing the source-receiver distance and corresponding transmission loss. *RL* values nonetheless provide a measure of relative signal strength. The weakest detected signal had an *RL_{max}* of 94.7 dB re 1 μ Pa (Lines 413, 491). Methods for locating the earthquakes responsible for these T-phases will be investigated in a future study.

- Line 400

The noise coming from the cryosphere was filtered out using a lower frequency to compute the ROI's centroid. It could be interesting to illustrate this observation by showing typical spectrograms from **cryogenic sources** in the Supp Mat.

Authors' reply:

We have included an example of ROIs along a spectrogram that were rejected using the spectral centroid feature in the Supplementary Material. Here we used examples of intermittent increases in background noise due to wind-ice interaction withstanding the pre-processing of the ROI algorithm. We refer to Figure S8 in Section 3.3 (Line 218).

Paragraph on Results

- The ROI method is based on edge detection in images: the amplitude of the spectrogram is therefore not kept in the processing. However, it is well-known that the amplitude of T-phases (in dB) are strongly correlated with earthquake magnitudes. This important characteristic that is not kept in the present ROI analysis. There might be a good reason why the authors did not exploit the amplitude, and they should explain why they did not do it. One solution to easily **exploit the amplitude information would be to extract the amplitude from the background PSD spectra** (shown in Figure 7) **at the *tc* of each event**. I would be valuable to add the amplitude to the 9 extracted features (*tstart*, *tend*, *T*, *fmin*, *fmax*, Δf , *tc*, *fc*, *Att*).

Authors' reply:

As previously mentioned, we have included analysis on Received Level (*RL*) in the Results Section 4.6, and described our approach for computing it in Section 3.4 (Lines 253-256). Rather than measuring *RL* at the temporal centroid t_c of the ROI, which is where the centre of energy lies within the ROI, not necessarily where the maximum level lies, we have computed the maximum *RL* within the spectro-temporal bounds of the ROI. This is derived using equations for Broadband Sound Pressure Level, from Merchant et al 2015, whereby Sound Pressure Level can be obtained along each time step (1 s), restricting calculations to the frequency range [*f_{min}*, *f_{max}*].

As for correlation with earthquake magnitudes, we feel we can't determine these yet, as we don't know enough about how the propagation of T-phase signals is affected by the water conditions, ice cover, bathymetry and currents in the region. We believe our paper presents enough information regarding the potential of our detection method for T-phase signals and that a full analysis of earthquake localisation and estimates of magnitudes will be the focus of another entire paper in itself.

At the moment, *RL* measurements have not been implemented as part of the set of acoustic features automatically captured when an ROI is computed; they are computed afterwards using a separate script. This could potentially be implemented in the future, though, once the paper is published.

- I find the *TP_new* events interesting but they are unfortunately not discussed in the main text and they are only visible as teal diamonds in one figure in Sup Mat. Which of the 9 ROI features are the *TP_new* more sensitive to?

Authors' reply:

We have attempted to highlight findings of *TP_new* signals in the main text. Originally, *TP* and *TP_new* signals were combined together in the plots, but now they are separated in the PCA, histogram and the PSD empirical probability density plots. We feel the plots are still easy to interpret. We have presented more information on *TP_new* signals in the Results (Lines 281-286, 320-325, 333-338, 390-393, 408-412) and Discussion sections (Lines 434-447, 452-454, 475, 558-560), particularly when discussing the distributions of acoustic features, and stating that they have similar features to the weaker True Positive events. In Figure 7, we opted to leave *TP* and *TP_new* combined as this figure already has a lot of information, and the temporal placement of *TP_new* is not discussed.

- Figure 4

I do not understand the white label called "absent".

Authors' reply:

We tried to clarify what is meant by 'absent' signals (Lines 266-269). There were six occasions where an earthquake signal was observable at NERSC1+3, but not NERSC2. We mentioned that these events are 'absent' at NERSC2, i.e. an event was perhaps expected in the file but was not observable. We preferred to avoid speculating on why these events are observable at NERSC1+3 but not NERSC2, but it's likely a combination of acoustic attenuation, propagation, and perhaps shadowing from local bathymetry.

How are the "new events" presented in Figure 4 and in Table 1?

In this section, we focus on comparing the detections from the ROI algorithm to what was in the dataset of manually annotated events. We're looking at what the detector has correctly detected, the TPs, and what was missed, FNs, so we're not looking at the new TP events just yet. We have added some additional information about the new TP events in Section 4.2 (Lines 279-286).

- Figure 6

I do not understand how the "density" is computed here. Are the plots just histograms?

Authors' reply:

We have replaced density on the y-axis with count in Figure 6 and S3. Hopefully this removes some unnecessary complexity in interpreting the plots.

- Figure 7

This figure is very interesting but it may need some more labeling/explanation. It shows the correspondence between *FP* and potential causes. The *FP* caused by strumming noise are obvious. You may help the reader to spot the whale chorus around 20 Hz and the bands of shipping noise. It could be interesting to mention if the *FP* in NERCSC2 during sep-oct-nov 2019 are caused by the whale chorus. It could be also interesting to mention if the *FP* on all 3 hydrophones in July 2020 are caused by shipping noise. The shipping noise maybe be seen also on Figure 8c in the 60-100Hz frequency band.

Authors' reply:

We have added labels to Figure 7 to help the reader. We have revised our findings for the *FP*s; no *FP*s were caused by shipping tones, but the cause remains unknown (intermittent low-frequency sounds encountered mostly towards the end of deployment). We have removed mention of shipping in Section 4.5; we cannot be certain of the sources that caused the other *FP*s (Lines 373-378).

- Figure 8

The PSD figures are also very interesting. Such plots are rarely presented in the literature although they allow good representation of the site/conditions. It could be interesting to add on it a reference curve such as the ones presented by Brown et al. (2014) in The IDC Seismic, Hydroacoustic and Infrasound Global Low and High Noise Models. *Pure and Applied Geophysics*, 171, 361–375, DOI 10.1007/s00024-012-0573-6

Authors' reply:

We thank Reviewer B for suggesting this paper. However, despite being interesting, we think that Figure 8 does not need extra information or reference lines, since we neither consider nor superimpose global noise models/noise curves. Moreover, many of the presented curves in that paper are just < 10 Hz, and we wanted to give a broader overview of frequency ranges. We have separated *TPnew* signal spectra into a separate panel now, following recommendations in previous comments regarding further analysis of *TPnew*.

Supplementary Material

Page 1, last sentence: A figure number is missing.

Authors' reply:

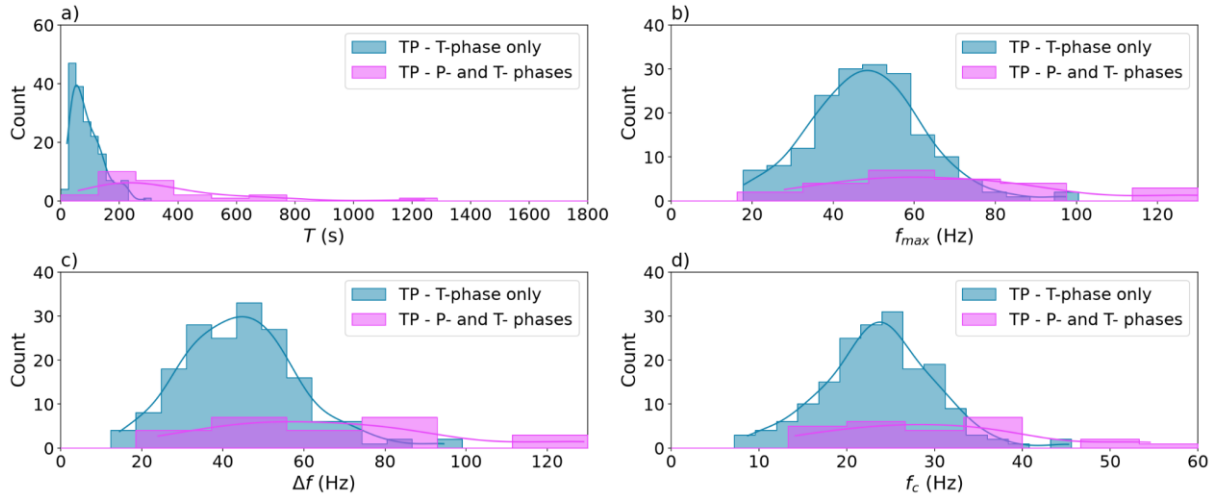
This has been amended.

The **multi-phase detection (P-T)** is an interesting outcome but it complexifies the analysis of 4 of the acoustic features (*tstart*, *tend*, *T*, *Atf*). I can see that they were added in the Supp Mat but very little is said in the main text about that.

Authors' reply:

We agree with the fact that it complicates the analysis of 4 of the acoustic features (*tstart*, *tend*, *T*, *Atf*), and due to the lack of more multi-phase detection (P-T) signals, we prefer to not run any further analysis on that aspect. What we say in Section 4.3 (Lines 318-320) also stands for the findings in Section 4.4.

We have an extra figure of acoustic feature distributions below, where T-phase and P-T signals are presented separately. We see that P_T signal ROIs had increased T , but similar spectral features. While the seismic ROIs with the widest bandwidth and highest f_{max} and f_c were P-T signals, this is more due to signal strength. We felt this figure was not interesting enough to discuss.



Moreover, the Supplementary Figure S2 has been changed, and it has only one panel now.

Similarly, the **new_TP** are also interesting to separate from the other **TP**. I was going to ask a different symbol for the PCA projection in figure 5, but then I saw it illustrated in the Supp Mat. The authors should probably discuss these observations in the main text as well.

Authors' reply:

Following your previous comment on **TP_new** signals, we have added more details in the Results (Lines 281-286, 320-325, 333-338, 390-393, 408-412) and Discussion (Lines 434-447, 452-454, 475, 558-560) sections on how seismic events newly discovered by the ROI algorithm differ from those featured in the manual dataset. We have also changed Figures 5, 6 and 8 in the main text to treat **TP_new** as a separate category. Turns out the **TP_new** acoustic features were quite different, so worth discussing separately.

Recommendation: Revisions Required

Reviewer C:

Review of “T-phase detection within spectrograms of hydroacoustic data from the Arctic Ocean” by Anichini et al.

Summary:

Study topic/area – central Arctic seismicity, Regions Of Interest algorithm, T-phase detection
Main goal(s) – develop a better method for detecting T-phase signals in hydroacoustic data
Methods/analyses – regions of interest algorithm, catalog comparison (e.g., precision & recall), principal component analysis, spectral analysis and signal characterization
Results – 85% of manual events detected with some new weak events added; overall precision of 71% and recall of 74% though varied significantly by hydrophone; fmax, df, and fc were main contributors to the first principal component (45.8%); false positives tended to have different features than true positives; many false positives occurred during periods with strong strumming noise, but false negatives had no notable correlation with noise levels and were mostly weaker events/signals
Interpretations/conclusions – the ROI algorithm was able to perform well at identifying and characterizing T phases in hydroacoustic data

This is an interesting study that proposes using an image processing technique to identify T phases in hydroacoustic data. With all the attention on machine learning detectors right now, **I think simpler options like this are also valuable for the community to see and consider.** Overall, this manuscript is very well written with clear explanations. There are a few places that could use additional information, clarification, or corrections, primarily in the introduction. There are also some comparisons of the proposed detector to more commonly used ones, though these seem to be entirely speculation based on known limitations rather than direct comparisons. The figures are all of good quality and support the arguments in the main text. Links for the data and code are provided and accessible. I've provided some general comments and thoughts below as well as more specific line comments. I think this manuscript will be a **great** addition to the literature once revised.

We would like to thank Reviewer C for their thorough evaluation of our manuscript and for their very useful suggestions. We are grateful for the summary above and the compliments given towards the manuscript, especially the comment on our methodology (and its relative simplicity) having value for the seismic community.

The comments have helped us reinforce several sections of the manuscript, especially the Introduction section where we needed to elaborate further on the nature of T-phase signals and the complexities in acoustic propagation in the Arctic Ocean. There were a few suggestions on improving definitions we used throughout the manuscript, such as T-phase and 'detection'. We feel these comments have added greater clarity to our manuscript and will help readers (who may not have strong knowledge in acoustics) navigate through the paper.

We have replied to each comment individually below. Line numbers provided in the replies refer to the file “[Seismica_Revised_MAIN.pdf](#)”

General comments:

While the introduction overall fits the manuscript nicely, there are several places where the information is somewhat incorrect or could be misleading (see line comments), or where additional info could be useful to the reader. It may be helpful to define how you are using the term “T phase”. Initially, it was used for earthquake phases that started and ended as

seismic waves but primarily propagated as a converted hydroacoustic wave. Although it is often used to include other situations now, some researchers still use it in the original way, using a term like “**H phase**” to indicate a primarily hydroacoustic phase detected by a hydrophone (rather than a hydro-to-seismic converted phase on a land-based station).

Authors’ reply:

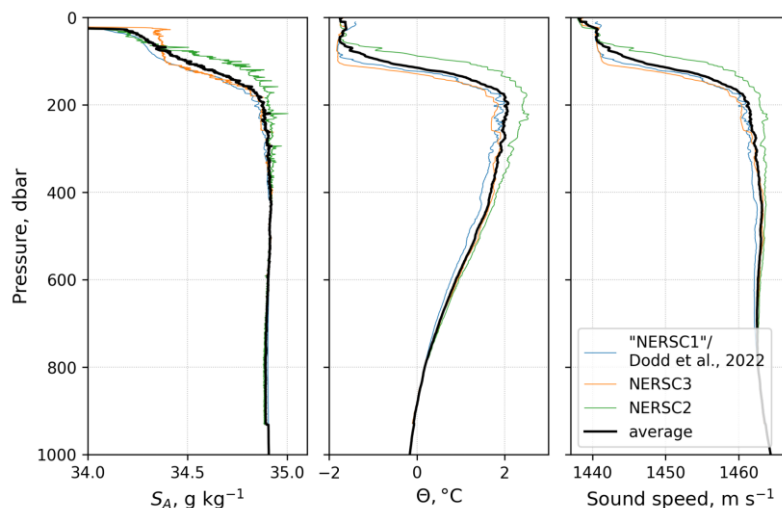
We have added a short statement at the end of the Introduction to clarify that we use only one term, T-phase, to refer to seismic-to-hydroacoustic converted phases generated by an earthquake recorded by a hydrophone (Lines 98-99). In recent literature (e.g. Raumer et al 2024), the “H” is used to describe mainly acoustic emissions from volcanoes, so there are inconsistencies with the use of this term as well. In our work, it would be more confusing than helpful. We prefer to avoid over-categorising the recorded signals and just use T-phase to describe the hydroacoustic signals we observe.

Similarly, the introduction makes no mention of the SOFAR channel, which is sometimes assumed to be the primary propagation path for T-phases. I did notice later on that section 5 had a nice discussion about T-phase generation and propagation. I don’t think this needs to be repeated in the introduction; however, it might be good to at least mention the **SOFAR** and how the arctic typically has a polar half-space (+ small surface duct) velocity profile instead. Lines 49-51 sort of allude to it.

Authors’ reply:

We have added extra information in the Introduction section (Lines 62-72) and in Suppl. material (Section 0.6) to report what is known about the water stratification in the Arctic Ocean.

Moreover, below are reported profiles of pressure, temperature, salinity and sound speed from the CAATEX moorings and adjacent CTD profiles.



Profiles of absolute salinity (S_A), conservative temperature (Θ), and the sound speed estimated based on these variables using the GSW toolbox, following convention by TEOS-10. The function used for the sound speed is described in https://www.teos-10.org/pubs/gsw/html/gsw_sound_speed.html. Deployment CTDs for NERSC3 (orange) and NERSC2 (green) and a CTD profile taken by Dodd et al., 2022, which matches the temperature at NERSC1 relatively well (blue), are shown. The deployment CTDs represent the mooring observations well. The thick black line in each panel is the average based on the

three CTD profiles. Before averaging, the data are interpolated onto a regular vertical grid with spacing 0.5 dbar.

Although these profiles give a brief overview of the oceanographic conditions in the region during the CAATEX experiment, we prefer to omit them from the current manuscript. The analysis of the vertical ocean structure and their relation to the T-phase signal recorded go beyond the scope of the current study.

You should also consider defining “**detection**” at some point. It can be used for both classification problems and phase-time picking or a combination of the two. This manuscript seems focused on classification (or identification), although there are times where it discusses methods that are more relevant to the time picking, like STA/LTA (though those often do both in a sense). I think having this clearly and explicitly defined where applicable could help to clarify and highlight some of the strengths and weaknesses of the different methods and how they may best be used. In some places it is, but others it’s not.

Authors’ reply:

We added further clarification lines in the Background (Lines 114-115 and 135-139) and in the Discussion sections (Lines 529-532). We have not set up the ROI algorithm to work as a phase-time picking tool; rather, it’s just an exploratory tool for automatically logging earthquake events across large datasets. We have better defined what ‘detection’ means in the current work and also made it clear that classification of events, seismic vs. non-seismic, T phase vs. P phase, was done manually after running the algorithm.

Could you add a little more detail about your decisions regarding P & T phases. In section 4.1, the text implies that the ROI detector counted the P phases as distinct for several of the P_T events, and that those P phases were ignored. I guess it’s not clear to me why you chose to lump the P & T together for analysis. I’d expect that if the earthquake was near the receiver, the P & T arrivals would be close enough to effectively merge into one ROI, whereas for more distant events, the arrival times would be farther apart, leading to the “separate” detections. Seems to me that those ROIs could be grouped together and counted as P_T even though the phases were technically separate ROIs – the detector did what it was supposed to (identify ROIs that may be P/T phases); the “problem” to solve simply may not have been well-designed for the scenario.

Authors’ reply:

The ability of the ROI algorithm to detect the P and T phases is an interesting outcome, and, indeed, the distance between the receiver and the source of the earthquakes may play a role in how the P and the T phases may be grouped in the same ROI. However, as the source-receiver distances remain unknown in the current work, we focused the analysis only on those events having the P and T phases merged in a single ROI. We tried to clarify this aspect in the Results section (Lines 272-277). Rather than speculate why those ROI with P phase only resulted in being separated from their corresponding T phase, we mention the link between P_T phase events and the quality of the signal in the Discussion section (Lines 428-433).

Regarding comparisons to other detectors, I think the broader discussions comparing the strengths and weaknesses of the different methods is a valuable part of this paper. However, there are a couple statements (e.g., lines 27, 382) about how the ROI detector outperforms other detectors that seem to be based entirely on **speculation** rather than on testing. I’d suggest either removing these statements or actually trying some other methods to see how they perform. While the latter would take more time and perhaps be considered beyond the

scope of this study, it would certainly be interesting to see how other methods perform on this dataset in **comparison**.

Authors' reply:

We have removed "outperforms comparable seismic signal detectors" as we have not directly compared our method to others (Lines 513-515), and we changed the statement in the abstract (Lines 27-28).

Line comments:

- Lines 30-31: Why would year-round ice coverage be an issue for OBSs but not hydrophones? They both need to be deployed and retrieved by boat.

Authors' reply:

We have added extra lines in the introduction section (Lines 38-34) to clarify the potential limitations in deploying and recovering OBS units in an ice-covered environment.

- Line 37-38: The wording here implies that the energy is being released directly into the water column; however, this only happens if the earthquakes originate very near the surface. Much more often, energy initially in the form of seismic body waves converts to a hydroacoustic wave at the seafloor, via different processes. The conversion can happen near the earthquake's origin or farther away, for example on a nearby seamount that provides a more efficient conversion point. Please revise this line as appropriate to clarify this distinction.

Authors' reply:

We have added extra lines in the Introduction section to describe the complexity of the T-phase signal generation and propagation, and what is known about their effect on the temporal and spectral properties of the signal (Lines 57-61).

Lines 41-42: I think the key here is the earlier arrival times. While T phases are often thought of as the typical "spindle" shape, I've seen some that closely resemble the source characteristics (e.g., with distinct P & S phases). Their amplitudes and durations can also be quite variable depending on the situation (conversion/generation & propagation/detection), though frequency content may be a useful discriminator in most cases. Of course, there may be clearer trends within a specific dataset that make amplitude & duration more reliable for discrimination.

Authors' reply:

Thank you for these hints; we agree that the interplay of factors makes the scenario of T-phase complex. We added extra lines in the Introduction section (Lines 57-61) to expand on what is known about generation and spectral-time T-phase characterisation. However, since we do not focus much on the arrival time in the current work, we preferred to keep this part short.

Lines 44-45: **"commonly presenting a sharp onset of 30-40 dB above background noise" – Most of the signals in this study are classified as weak events. Are those mostly 30-40 dB above background with a sharp onset?** Often T phases are described as emergent without sharp onsets (including in at least some of the cited studies), which is why STA/LTA detectors can struggle with them. So, I'd disagree with the statement in the text that those characteristics are common (though they certainly can occur).

Authors' reply:

In these introduction lines, we have attempted to generalise the broad range of the T-phase signals encountered in the CAATEX dataset (from impulsive broadband short to long-lasting spindle-shaped phases), describing how they can appear in hydrophone recordings, mentioning that the onset of T-phase signals can often be shorter than the tail-off (which is what we meant by “sharp onset” originally) (Lines 50-53).

Line 62: This is not entirely true. There are automated methods that are commonly used for detecting T phases, typically those based on beam-forming and cross-correlation/coherence. However, those methods require a small aperture array, so more widely distributed hydrophones may require analyst picking or adoption of “seismic” methods like sta/Ita. Perhaps just specify **“For widely spaced hydrophone networks” or something similar at the beginning. Also, “T-phase” should be plural.**

Authors’ reply:

Thank you for this suggestion; we added this at the beginning of the paragraph (Lines 84-85).

Lines 65-68: Is this only talking about automated methods for seismic data, or is it also referring to methods developed for seismic data being applied to hydroacoustic data? It’s not entirely clear.

Authors’ reply:

Following the previous comment, we felt this paragraph should focus on methods adopted for processing data from hydrophone networks, so we have removed comments on seismograms and trigger algorithms (Lines 84-90).

Line 69: “T-phase detection” - typo

Authors’ reply:

This sentence was originally not worded clearly; it has now been modified (Lines 91-92).

Lines 94-95: “T phases” – typo

Authors’ reply:

This has been corrected.

Lines 96-98: Is this method for classification/identification or for phase-time picking or both?

Authors’ reply:

What was important for us to cite from both Sukhovich et al 2014 and Raumer et al 2024 was their method for classification/identification of earthquake signals, not time-picking. We have attempted to make this clearer (Lines 87-90) also in the Material and Methods Section 3.3 (Lines 203-204).

Line 112: I know of at least one study that used a similar technique (edge/line detection) on seismic data:

Tepp, G., Stubailo, I., Kohler, M., Guy, R., & Bozorgnia, Y. (2024). Shake to the beat: Exploring the seismic signals and stadium response of concerts and music fans. *Seismological Research Letters*, 95(4), 2179-2194. <https://doi.org/10.1785/0220230385>

Authors’ reply:

Thank you for the recommendation; this is an interesting paper! We are aware of Hough's Transform, but it is best used to identify lines or ellipsoids in images, like the harmonic tremors described in the paper. While there is a generalised Hough's Transform for non-geometric shapes, the Canny-derived ROI algorithm is more straightforward. There's an example of Hough's Transform being used on 20-Hz fin whale calls too, which do not vary much (<https://doi.org/10.1016/j.econinf.2025.103499>). We are not convinced this method easily translates to detecting sounds that can vary in duration, bandwidth, and shape.

As the tremors described are anthropogenic in origin, and the paper is using data from accelerometers, our statement "The ROI algorithm has yet to be applied explicitly to geophonic signals." still stands.

Line 120: Approximately how deep is the water column in the hydrophone location?

Authors' reply:

In Dzieciuch et al. 2026, the reported water depths are 4018, 4067 and 4021 m, respectively for NERSC1, NERSC2 and NERSC3.

Line 148: "mooring name, and date and start" – missing word

Authors' reply:

This has been corrected (Lines 174-176)

Line 315: Wouldn't the high EPD at 0 dB in Fig 8e, suggest that most FPs have similar levels above 10 Hz to the background noise? The biggest difference is around the strumming frequencies, as expected if that's the cause of most FPs.

Authors' reply:

Yes, that's correct; we have added some text in the section mentioning the increase in *EPD* along the peak caused by the strumming (now referred to as 'flow noise'), and how there's little change in spectra above 10 Hz (Lines 388 -390).

We have updated Figure 8 and removed the panels with the ΔPSD values, replacing them with a panel of *EPD* for the *TP_new* signals by themselves. We were balancing comments between reviewers in this section, and felt that the original Fig. 8e and 8f did not add much information, and perhaps were also difficult to interpret at a glance.

Line 324: Did you do any checks with standard earthquake catalogs to confirm this? There wasn't any mention of comparing the hydroacoustic catalogs to, e.g., the ISC catalog, though I do see some events plotted in Fig 1.

Authors' reply:

Thank you for this observation. As mentioned in response to Reviewer B, we compared the CAATEX recordings a second time against the 16 earthquakes reported by the ISC and have now been able to associate T-phase signals to 5 of the ISC-reported earthquakes, rather than just 2 stated previously. Of these 5 events, only 3 had T-phase signals that were assigned Event IDs; the other 2 events had T-phase signals that were discovered by the ROI algorithm. We have included an extra section in the Supplementary Materials (Section S5, Table S2) detailing the data provided by ISC for these earthquakes and added extra lines in the Discussion section (Lines 464-486) and in the caption of Figure 1.

Line 339: "of" should probably be "by"

Authors' reply:

This has been corrected.

Lines 344-346: Other factors can also play into this.

Authors' reply:

We have mentioned inefficient T-phase generation, which results from the various factors we summarise at the end of the paragraph, citing Bottero et al 2020, which was recommended below (Lines 483-486).

Line 355: This would be another good reference here:

Bottero, A., Cristini, P. & Komatitsch, D. On the Influence of Slopes, Source, Seabed and Water Column Properties on T Waves: Generation at Shore. *Pure Appl. Geophys.* **177**, 5695–5711 (2020). <https://doi.org/10.1007/s00024-020-02611-z>

Authors' reply:

Thank you for the suggestion; we have included this reference here and added “seabed properties” to the list of other factors that can affect T-phase generation (Lines 482-486).

Lines 366-367: Shouldn't the strumming frequency be primarily dependent on the length of and material used for the mooring cable? Perhaps you're suggesting using current speeds to help determine when the filter should be applied or something similar?

Author's reply:

Indeed, the strumming noise frequency, which we now conservatively refer to as ‘flow noise’ in the paper, is dependent on the length and material of the mooring cable, but the severity of the noise is also driven by current speeds. Across the deployment, there are occasions where flow noise is weak, resulting in tones only below ~10 Hz, but when flow noise is particularly severe, there are multiple harmonics. So when we refer to an adaptive filter, it would need to be designed to not only filter out the fundamental frequency but also the secondary harmonics when present; we have added this small clarification in the Discussion section (Lines 508-511).

When addressing comments from Reviewer B, we have included some contextual information in the Supplementary Material covering the oceanic setting of the measurement region (S6) and a summary of how we related the continuous low-frequency tones to flow noise (S7).

Lines 382-392 (also line 27): It would be interesting to see how those detection methods actually perform on this dataset. Much of this paragraph seems like it's generalizing known limitations of those methods and then extrapolating that they wouldn't work well on this dataset.

Authors' reply:

We weighed up comments from both reviewers on benchmarking our method: either running the ROI algorithm on another, publicly available dataset or using the STALTA algorithm. We opted to run the ROI algorithm on datasets from IFREMER (Raumer et al. 2020), and we were able to detect all T-phase events confidently labelled by the authors of the respective annotation datasets. Please see reference to this matter in the Discussion section (Lines 548-555) and in the new section in Suppl. Material S8.

Moreover, we have overall toned down the language of the related paragraph in the Discussion when comparing ours with other methods (Lines 514-525).

Line 389: “only a few” - typo

Authors’ reply:

This has been corrected.

Lines 436-437: You could simply provide the DOI from Zenodo here rather than the long url link.

Authors’ reply:

The DOI from Zenodo (<https://doi.org/10.5281/zenodo.17939966>) has been added to the revised version of the manuscript. However, the page will be public only once the paper is ready for publication.

Table 1: The last column was a little confusing at first glance. Maybe label it “Total True Positives” or similar.

Authors’ reply:

We have followed your suggestion for renaming the last column of Table 1.

Figure 2: Please note in the caption what the boxes and +’s mean (I assume those are detections from your algorithm?).

Authors’ reply:

We have included this detail in the figure caption.

Figure 4: In the caption, I’d recommend adding a “(TP)” after “True Positive” to define the abbreviation in the figure, and to do similarly for “FN” so that it’s defined in the caption. This is keeping in mind some readers may look at figures without or before reading the main text.

Authors’ reply:

We have amended the caption to feature both *True Positive (TP)* and *False Positive (FP)*.

Supplemental Material:

“Additional PCA projections”: Missing figure number (marked as ??) in first line.

Authors’ reply:

This has been amended.

Recommendation: Revisions Required