Validation of F-POD Cetacean Detections   21/11/2023

This is an essential step in delivering sound results. Before diving into the file you need to know:

What detection metric are you using?

If that is DPM – detection positive minutes – or any larger unit then you will be able to pass a minute (or the larger) unit as correct if it has at least one correct train classification and you don't have to consider every train in that minute. This saves a lot of time.

What is an acceptable level of FPs (False Positives)?

Don't say 'Zero'! You need to assess how big the difference is that you are looking for. If you wanted to show there is a difference between sites that are giving detection rates that differ by a factor of 3 then an error rate of 10% FPs is very unlikely to affect your conclusion.

If you are looking for a trend over years in a population you will want a much lower FP rate. You might aim for a rate per year in your whole set of files that was below 5% or below 1% - discuss this with your friendly statistician!

What is an acceptable level of FNs (False negatives)?

Actually these are something you can generally ignore – they are very common - nearly all the cetacean clicks in the sea are missed by your F-POD, and many trains that can be seen in the data are also missed – this includes all trains with less than 5 clicks in series. The FN rate is properly handled by a 'detection function' that we will not go into here. Few studies have one, and make the assumption that variations in the FN rate are small enough to ignore. That has worked out well in practice but keep your wits about you – consider changes like rising levels of boat sonars or fishery pingers that could be relevant.

'Train fragments' are some of these false negatives – they are short groups of clicks that look likely to be from a train source to an assessor, and they are useful in the validation process.

Upgrading FNs to true positives is very rarely a good idea. It does two bad things – it makes your new results subjective, which means you have to introduce assessor assessment, and it takes ages. If your FP rate is above your target the right approach is usually to find some filtering within the FPOD app that deals with that effectively, and then use that consistently in data from those sites.

Were the KERNO-F settings appropriate for the context?

In general the default settings are satisfactory, but if you know there are no NBHF species, then you can make the classifier work a bit better by classifying using 'No NBHF' and the same applies to other cetaceans and boat sonars.

If in doubt stick with the defaults, because in the future what you picked might no longer apply - boats have moved in with sonars etc - and your older data would have to be re-analysed. Ouch!

You can recall the settings used in any FP3 file by viewing or listing the 'classification warnings'.

So we're ready to go…

Validation has two stages:

1. A*ssess the whole file* by looking at when the POD was immersed, the angle to vertical, and the noise level.

2. *Validate a sample of trains*.

> That means correctly placing a click train in one of the train source categories below. It does not matter if the train incorporates some clicks that were not cetacean clicks – they do not invalidate the source of the other clicks.

Chance trains

> Sources that do not make click trains and are generally creating clicks independently e.g. sediment particles in suspension striking each other, rain drops hitting the water surface, many shrimps clicking independently, things brushing the hydrophone housing etc. Chain noise is a very rare source and there is generally no need to avoid deployments near to chains.

> By chance these random clicks can fall into a neat, or perfect, train sequence. There is rarely any difficulty in being confident about deciding a train is not a chance train by looking at its context.

> Chance trains are put in the '*unclassified*' category, along with trains from cetaceans and sonars where the classifier was in doubt about the species (classifiers do a lot of worrying).

NBHF Trains

> Narrow-band High Frequency trains are made by all porpoises and a few dolphin species and consist of rather long clicks with typically 4 or more similar cycles at a frequency of 100 – 150kHz.

> The main errors here come from:

> - SONARS at NBHF frequencies
> - DOLPHINS can sometimes make surprisingly narrowband high frequency click trains and cause false positives.
> - Rarely: sediment transport noise from fine sand in suspension.
> - Rarely: WUTS – weak unknown train sources.

'Other cetacean' trains

> These are BBTs – Broad-band transients. They are typically less than 4 considerably different cycles and are made by all the other cetaceans that make non-NBHF clicks. Beaked whales make long clicks that have a frequency sweep within them. They are also placed in this category. The main problems here are

> - SONARS and other man-made sources. At source these clicks would be easy to distinguish, being long and regularly timed, but after multiple reflections and long transmission paths the trains can look very different.
> - NBHF species can be, rarely, misclassified as other cetaceans.
> - WUTS – weak unknown train sources. This one is hard, but rare.

Boat sonar trains …. etc

> Boat sonars and other man-made sources. These typically make long clicks with very regular timing.

> The FP rate for sonars from KERNO-F is not as low as for cetaceans, but this identification is largely correct and can be used to give information on boat presence. These are put in the '*sonar*' category.

> The presence of these sources usually reduces the sensitivity to cetacean trains (the classifier is worrying). In signal detection terms they are interference rather than noise.

> You should avoid deployment in harbours and busy traffic lanes if possible.

> Acoustic fish tags are logged, and can be seen and extracted from the data, but rarely get identified as train sources.

1. Assessing the whole file

A view of the whole time range of a file:
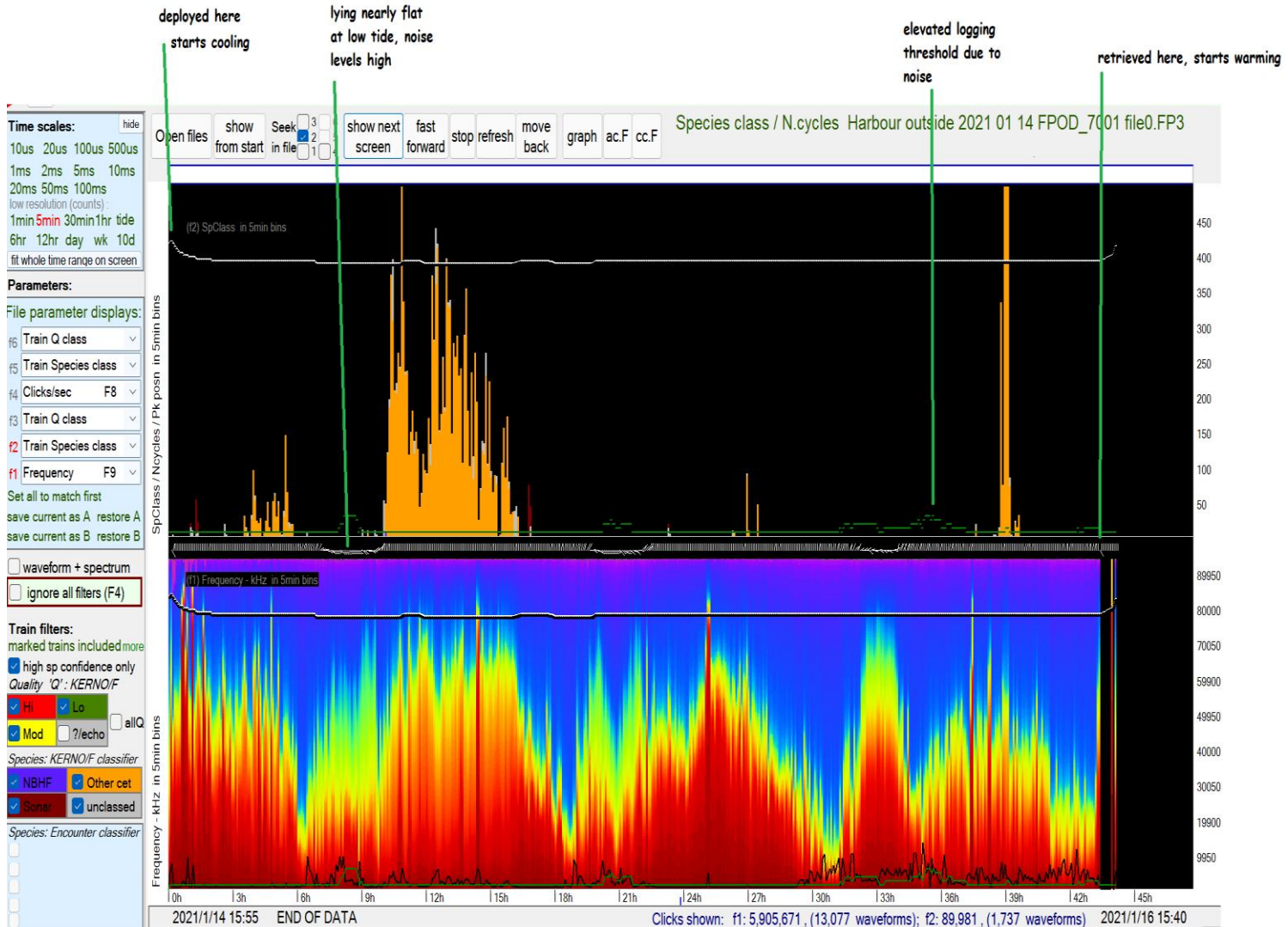
In most projects you will crop the file at the points where it has been deployed and before it is retrieved.

Logging threshold / Noise patterns

These vary with sites and high levels i.e. more than 2 steps up on the green line showing the logging threshold, will potentially reduce the detectability of weak cetacean clicks. Sometimes the noise is due to cetaceans in which case there are also cetacean detections. So you need to know whether the noise patterns are normal for your site or not. The noise source can be assessed in the FP1 file.

The black line shows the click count, regardless of source.

Angles are useful

… are useful. Here the big shifts are expected, but they may tell you the deployment was not what you expected!

Temperatures

… the black and white line – typically show big diel swings before deployment. After deployment it settles to sea temperature over several house and you may subsequently see evidence of mixing water bodies and especially of a thermocline moving across the POD which can bring strong changes in species mix.

Classification warnings = Error risk reports

These are a useful warning of files that may have a lot of classification errors. Where there are no cetaceans the likely error rate must be either 100% (if any 'cetaceans' are 'detected') or 0% % (if no 'cetaceans' are 'detected'), so files with few or no cetaceans will often be flagged up by the warnings, and sometime files with many detections …   See this video on the Chelonia website: *Using F-POD error reports.mp4*

Classification warnings can be shown for open files, or exported for batches of files, via the Filters +files page.

Here they are, bottom left, for the same file as above: Harbour outside 2021 01 14 FPOD_7001 file0.FP3

As experience accumulates you may be able to skip train validation at this point for NBHF or Other Cet if there are no warnings related to them.

You need to be viewing the version 2 warnings and th video demonstrates how to get these and view them: *Using F-POD error reports.mp4*

2. Assessing the click trains

A fille may contain 100,000 trains and validation requires assessing only a sample - the default is 100 trains but you can change that if your target max FP rate is 3% or less.

The sample is obtained by setting validation points:

- Right-click in the display area of the FP3 (or CP3) file you want to validate and select 'Set this file for action'.
- Set the filters on the left to the selection you will be using. For dolphins these are the normal filter settings:
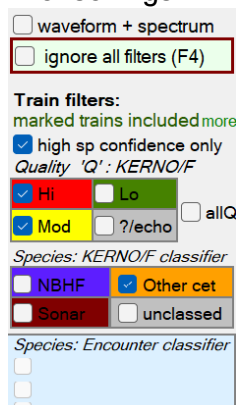
-
- Click 'set validation sampling points' (low on the right of the Filters +files page shown above)
- That will pick 100 minutes evenly spaced across all the 'other cet' clicks, with the starting point at a randomised position in the first 1% of that set. If you repeat it you will get a different set, and you can also change the number of validation points.
- Set the display time resolution to 20ms or similar.

Using validation points:

- Click 'Show from start'
- Use Ctrl + S to step through the validation points
- If your chosen metric is DPM you only have to decide if each validation minute contains at least one dolphin click train. With experience this is usually very quick! Then use Ctrl + S to move to the next validation minute i.e. you only check the validation minutes.
- If you find no false positives (i.e. validation minutes with no correct trains) a reasonable assessment is that you can be 95% confident that the true FP rate for DPM is less than 3%. This is called the 'Rule of Three' and there are many descriptions, of varying clarity, online!

Checking the source classification

In the file shown above there were 31 NBHF clicks and they were all false - they were in one train that came from a dolphin - it was a 'weak porpoise train' surrounded by dolphin trains.
There were 49k dolphin clicks and 372 DPM. All the validation points for dolphins were DPMs.
There were 836 sonar clicks and 8 sonar DPM, all correct.
So the next skill is:

Did it come from a cetacean? Which type?

The main super-power you bring to this task is your ability to weigh up the *wider scene* - typically a few minutes - around the click train in question. In high-res which you will use you can scroll backwards and forwards with the mouse wheel. If you accidentally skip to the next train, which could be days away, you can get back with Ctrl Z twice. It's also useful if you zoomed in a lot, as it will go back through your last 16 screen displays.

Another power you have is to see that the train picked out by KERNO-F is part of longer and rather different real train.

CETACEAN versus SONAR

THE give-away feature of sonars, that KERNO-F does not capture well, is this:
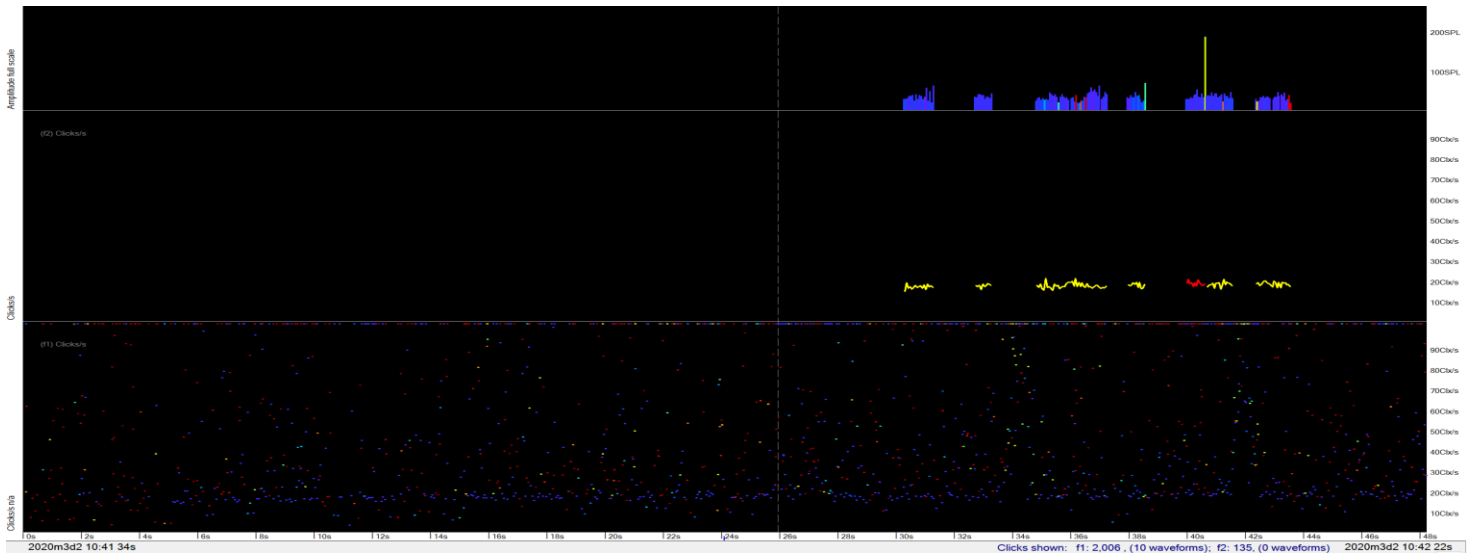

The lowest panel is showing *clicks/s* – that's just 1 / interval between successive clicks in the raw data. What can be seen is *a long fuzzy line spanning more than 1 minute* showing repeated inter-click intervals. So it's a sonar! And the FP3 the click trains have been wrongly identified as coming from an NBHF species because the clicks are the right length (number of cycles), the right frequency, and narrowband.

By filtering the clicks – not trains – to those over 110kHz that long flat line in clicks/s in the FP1 becomes sharper:
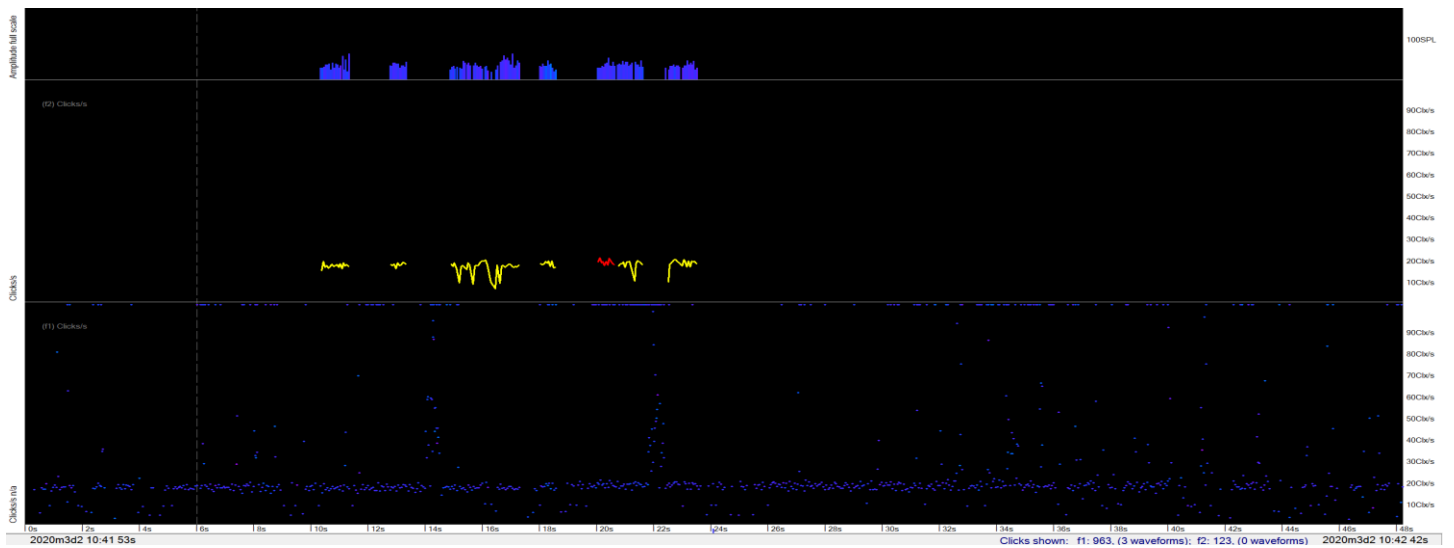
Long nearly-horizontal lines like that are THE diagnostic feature of sonars. They may have a gentle slope because of multipath effects. You do not see such long distinct lines in cetacean data.

There are other much weaker features, but you rarely need to look for them: KERNO-F handles the features of the clicks reasonably well, so you won't often get much power to detect any errors from them, but for completeness here are the main discriminatory features:

Feature	SONAR	CETACEANS
Number of cycles in click	Sometimes over 70cycles	NBHF occasionally up to 70cycles
kHz	Sometimes over 160kHz	Rarely over 160kHz
Wavenumber of loudest cycle	Sometimes over 10, and sometimes late in a long click	Higher for NBHF than dolphins but not often >10. Rarely occurs late in a long click.
Amplitude profile	Tends to be flat	From dolphins it is rarely flat
Bandwidth	Often low	Rarely low from dolphins
wider scene	Boats often go past in a few mins	

TRAIN SOURCE versus CHANCE TRAIN

This is a tough computational problem because there are so many distinct possible sequences of clicks in almost every minute.

…. but the KERNO-F results on F-POD data are good, and much better than the results from the KERNO classifier processing C-POD data. The KERNO classifier missed many dolphin click trains – actually it did *find* the trains, but it could not be confident that they were cetacean trains. KERNO-F has a unique (currently) input of very high-precision time-domain data on each click, with wave period values (250ns resolution) consistently referenced to the loudest cycle in the click. It also receives the cycle number of the loudest cycle, the last wave period and other time-domain features.

KERNO-F uses this high-resolution time-domain data to measure the *coherence* of each train and that becomes a major element in identifying and rejecting chance trains. Coherence is an aggregated measure of how much each click and each interval resemble its neighbours in the sequence…

However you will see errors occasionally and here are their features:

Feature	Cetaceans	Chance trains
wider scene	Cetacean detections are typically of *encounters* in which the animals are within detection range for a few minutes, producing trains that can be seen by eye even if they have not been classified as trains. A pattern of increasing amplitudes early on, with a more rapid fall at the end is typical.	These are mainly isolated, or within minutes that have true cetacean trains where they have 'survived' as a result of positive feedback provided by their true cetacean neighbours.
Amplitude profile within minute - at 100ms resolution or higher.	Cetacean trains typically form discrete, neatly rounded or prominent, humps on the amplitude display. Sometimes the spacing of the peaks is showing you the cetaceans tail beat rate.	Chance trains and sonars are often not prominent, or discrete, and the amplitude envelope is ragged rather than smooth.
Amplitude profile within clicks	Lots of sequences of similar profiles within a train - see images below. Big step changes in the profile don't matter - it's the sequences that count.	The profile usually jumps around from click-to-click but may be similar when source is sediment transport noise with a narrow frequency e.g. fine sand in suspension.
Inter-click-interval profile i.e. click rate	Often has a smoothly varying profile. Don't worry about infrequent very brief up/down spikes on the graph.	Overall, a more irregular graph with sharp transitions in rate and few smooth sections.
Click rate of train		Often a brief, irregular rise to high rates >100/s
Other click features - bandwidth, NBHFindex, number of cycles	More coherent	Less coherent
Multipath cluster features	Where these are logged they are generally the most powerful discriminatory feature and are described in more detail below.	

Below: highly coherent NBHF trains
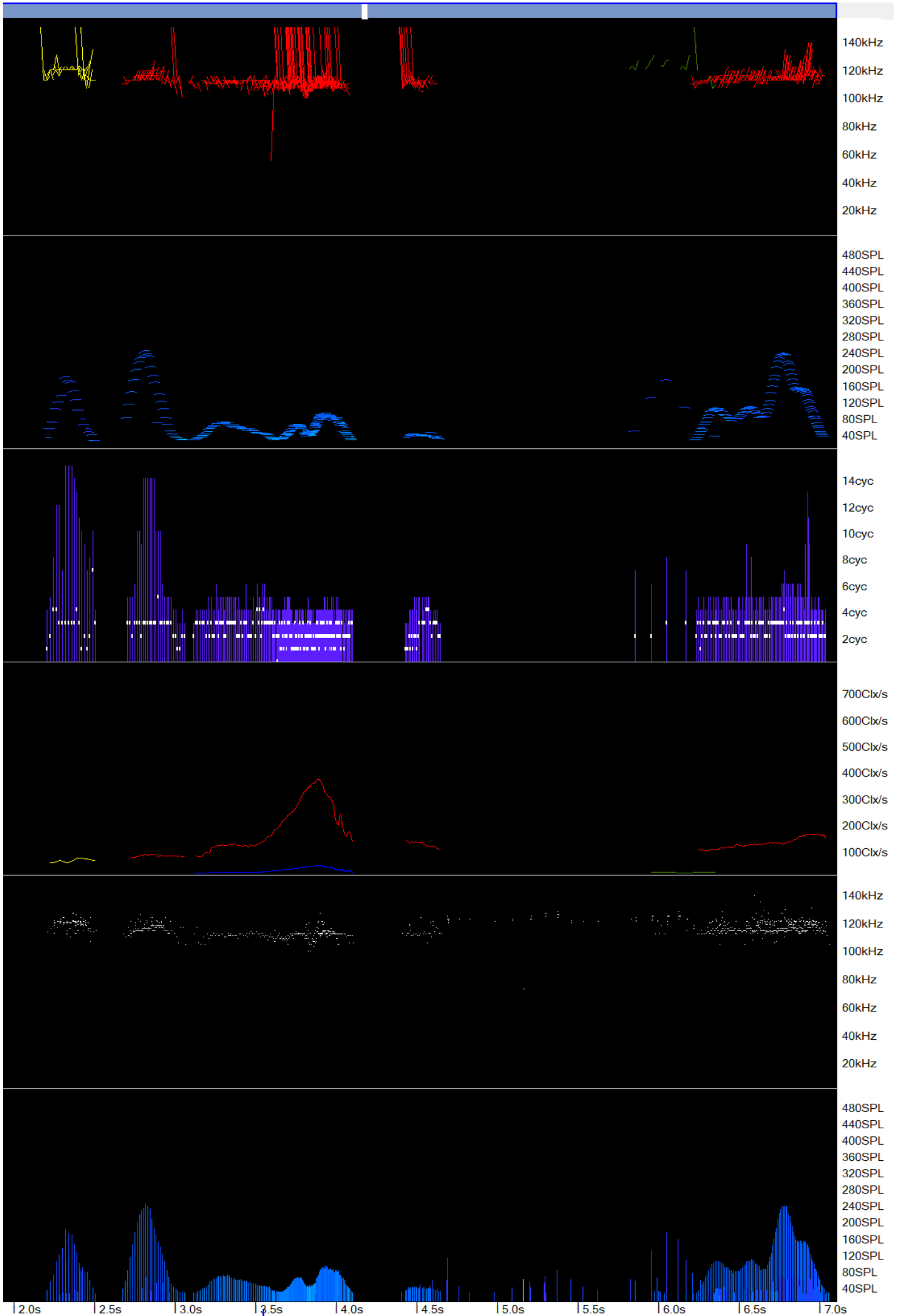
Below: a chance train from a brief noise burst and overlapping loud clicks from a porpoise
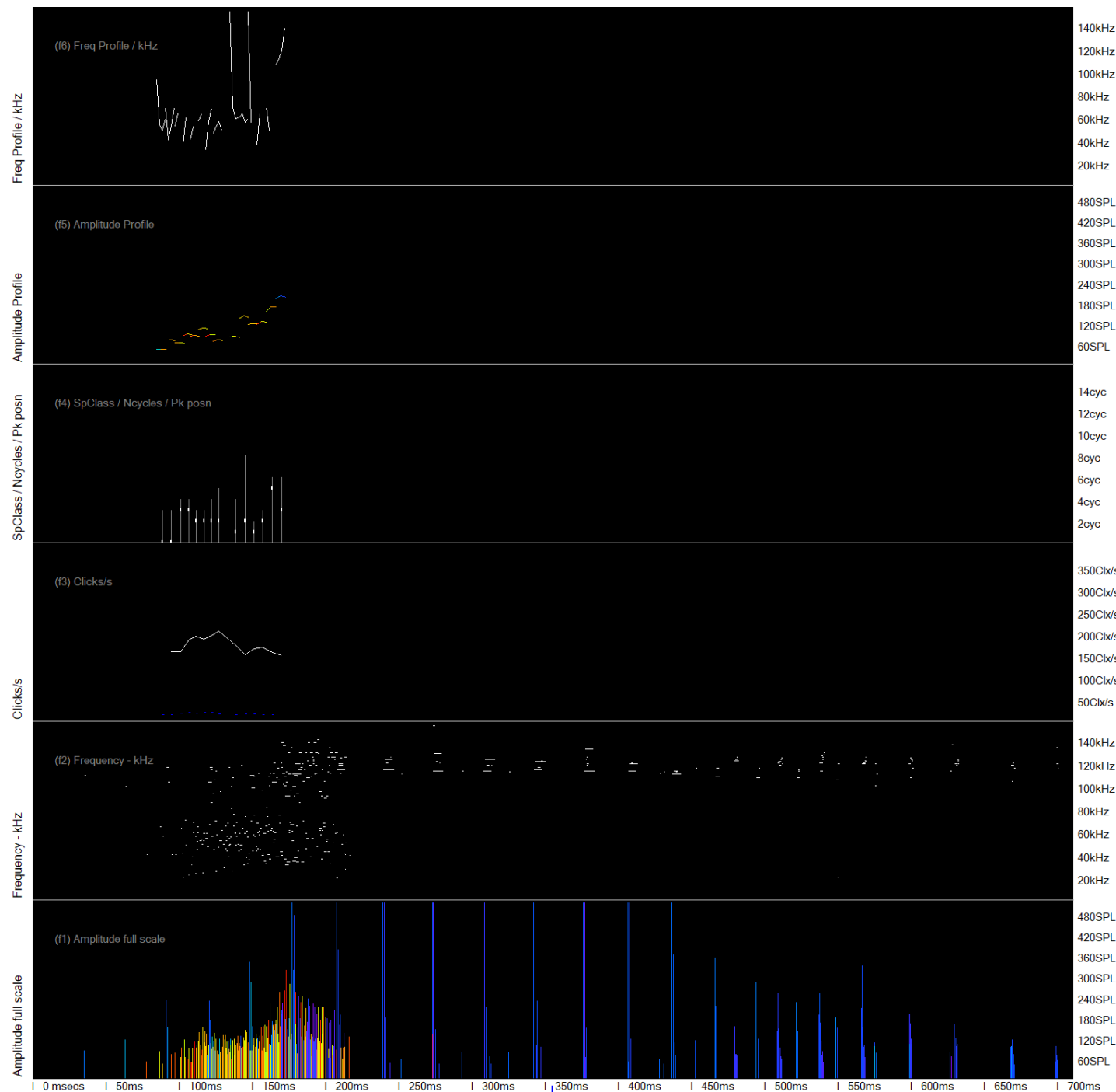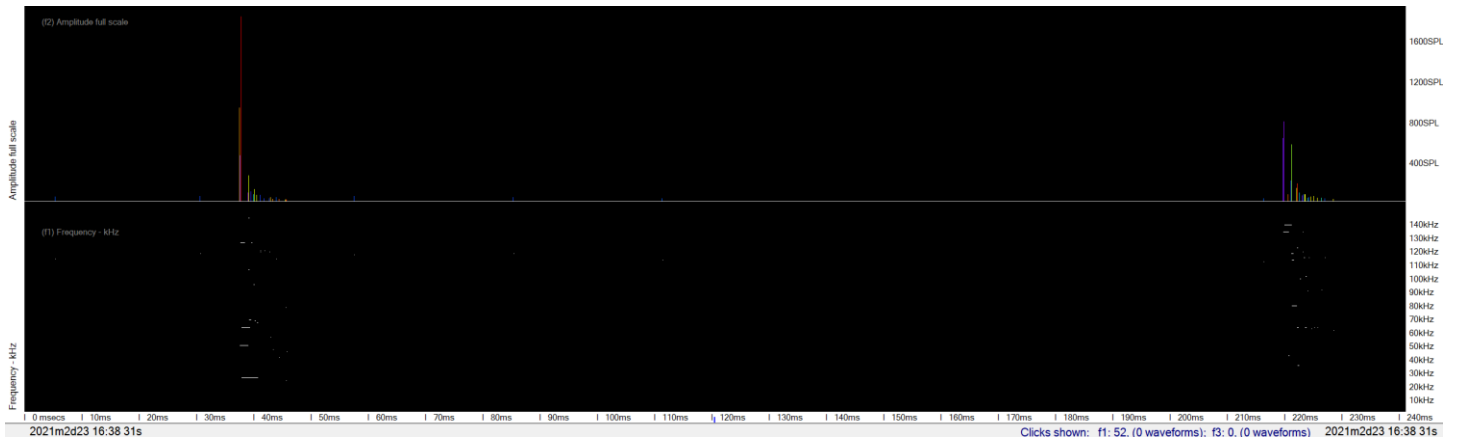

Multipath cluster features

Multipath clusters are highly significant because they carry information about *where* the source is. They are created by reflection of the click from surfaces – mainly the sea surface, and by refraction. Refraction is the bending of the click path by variation in the speed of sound along the sound path caused by small differences in water temperature and salinity.

Chance trains come from different sources – i.e. multiple individual shrimps etc. arrive along different pathways so their multipath clusters, if any, are highly varied.

Even more significantly, only clicks that are loud at their origin travel far enough to acquire multipath clusters without becoming so attenuated, by spreading and absorption, that they are no longer detectable by a POD.

Note: It is possible to make a 'virtual F-POD' out of a conventional sound file if it was sampled at a sufficiently high rate, but only if there was no click selection process to create click snippets as that gets rid of the weaker clicks that form the cluster.

The graphic above shows the exponential decay in amplitude that tells you this is multipath cluster from a very loud click. The two clusters are so similar that they are very likely to belong to a train. That's an awesome demo of the power of multipath – only two primary clicks but you're already confident it's part of a train.

The graphic below shows the same clusters forming multi-coloured lines in the amplitude display and vertical bands in the frequency display. These are very characteristic of clicks that were very loud at source.
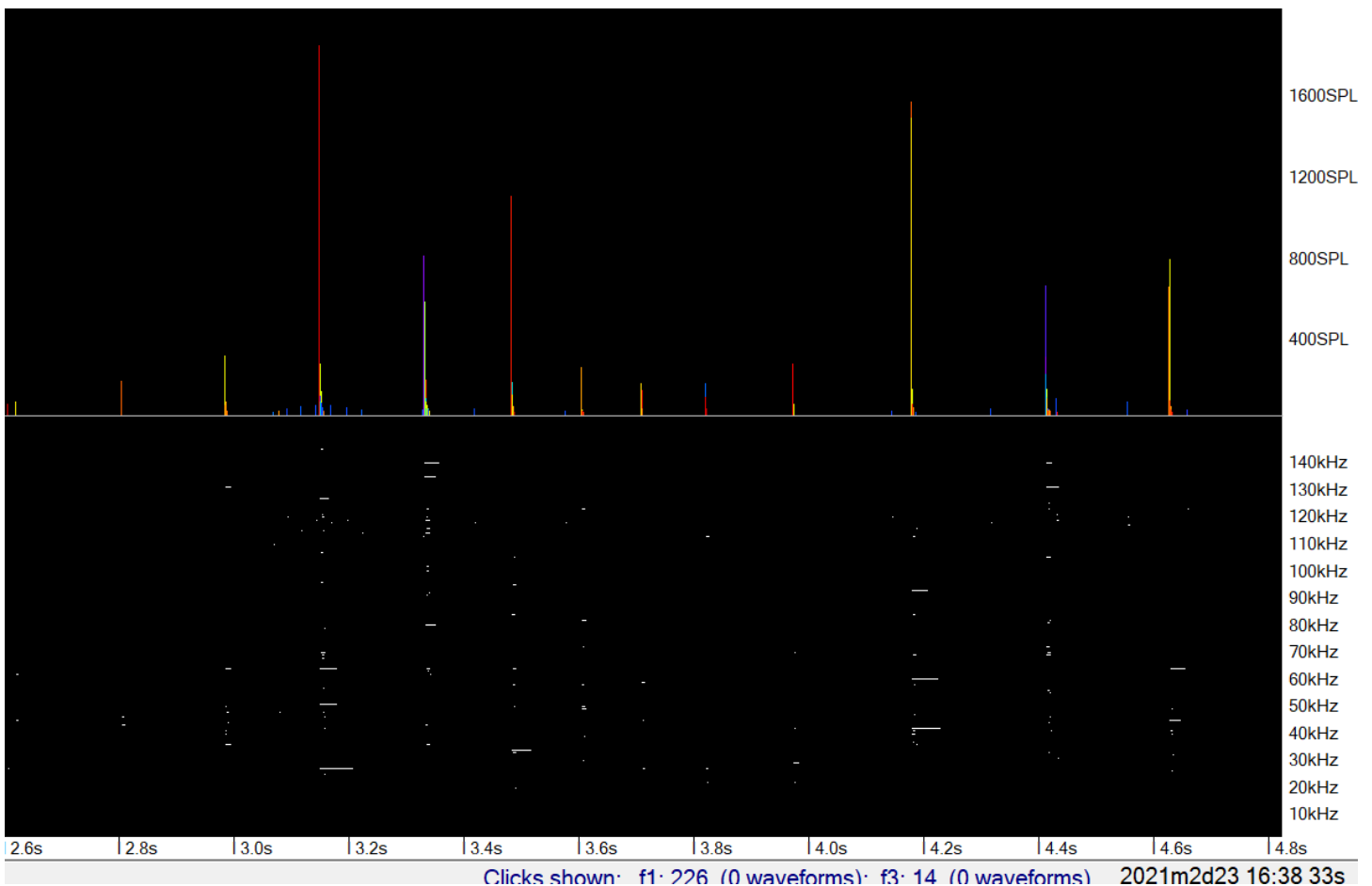
In the graphic below the multipath frequency content forms structured lines in the lowest panel - the frequency display of the FP1 file. This is typical of fast dolphin click trains because the pathway does not change much during the short inter-click intervals, so the clicks get split and sometimes reunited in similar ways as they travel.
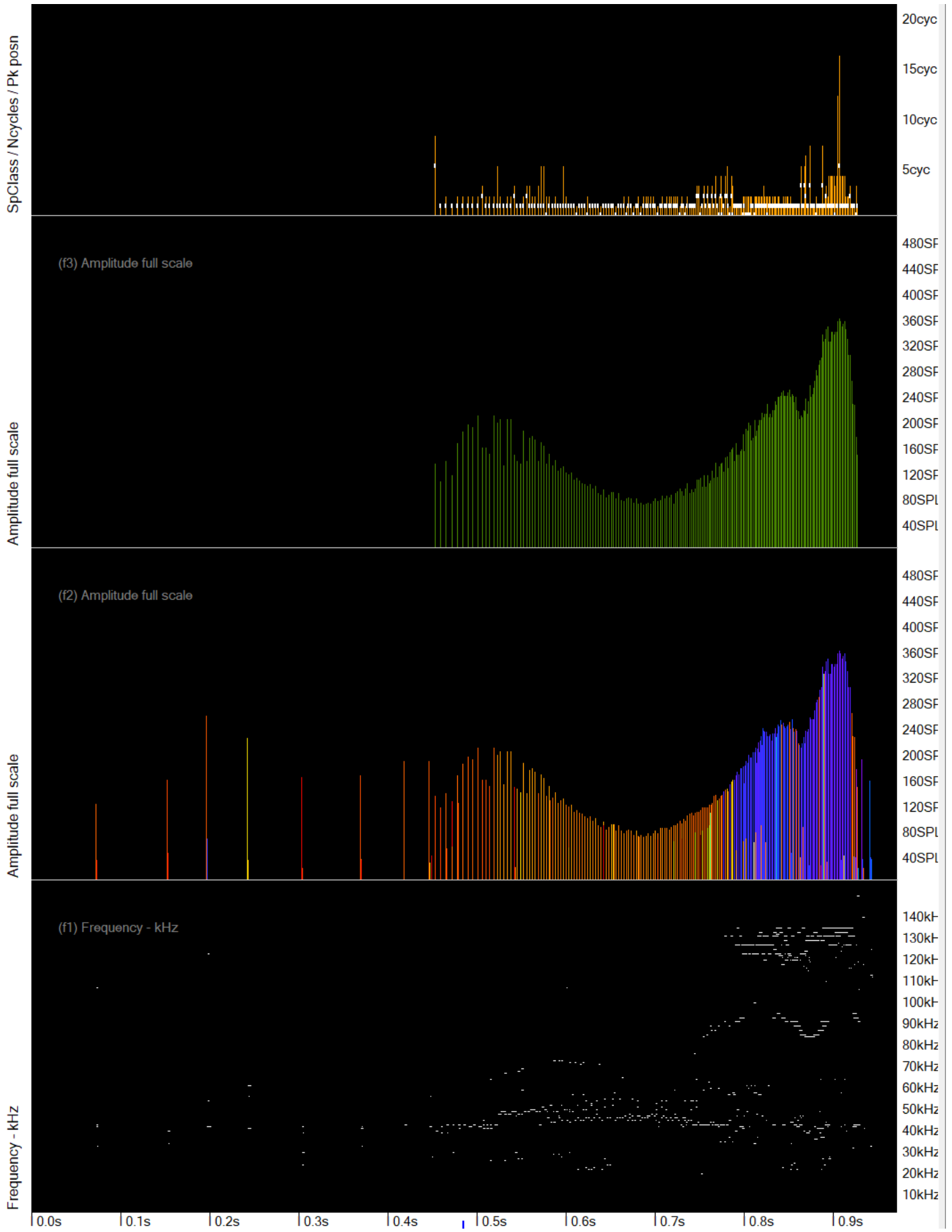
Below there is an NBHF train and then a false train from a very brief noise burst.

It shows no such structure in the lowest panel which shows the frequencies , in the FP1 display (the lower 2 panels here), and the amplitude profile is ragged.
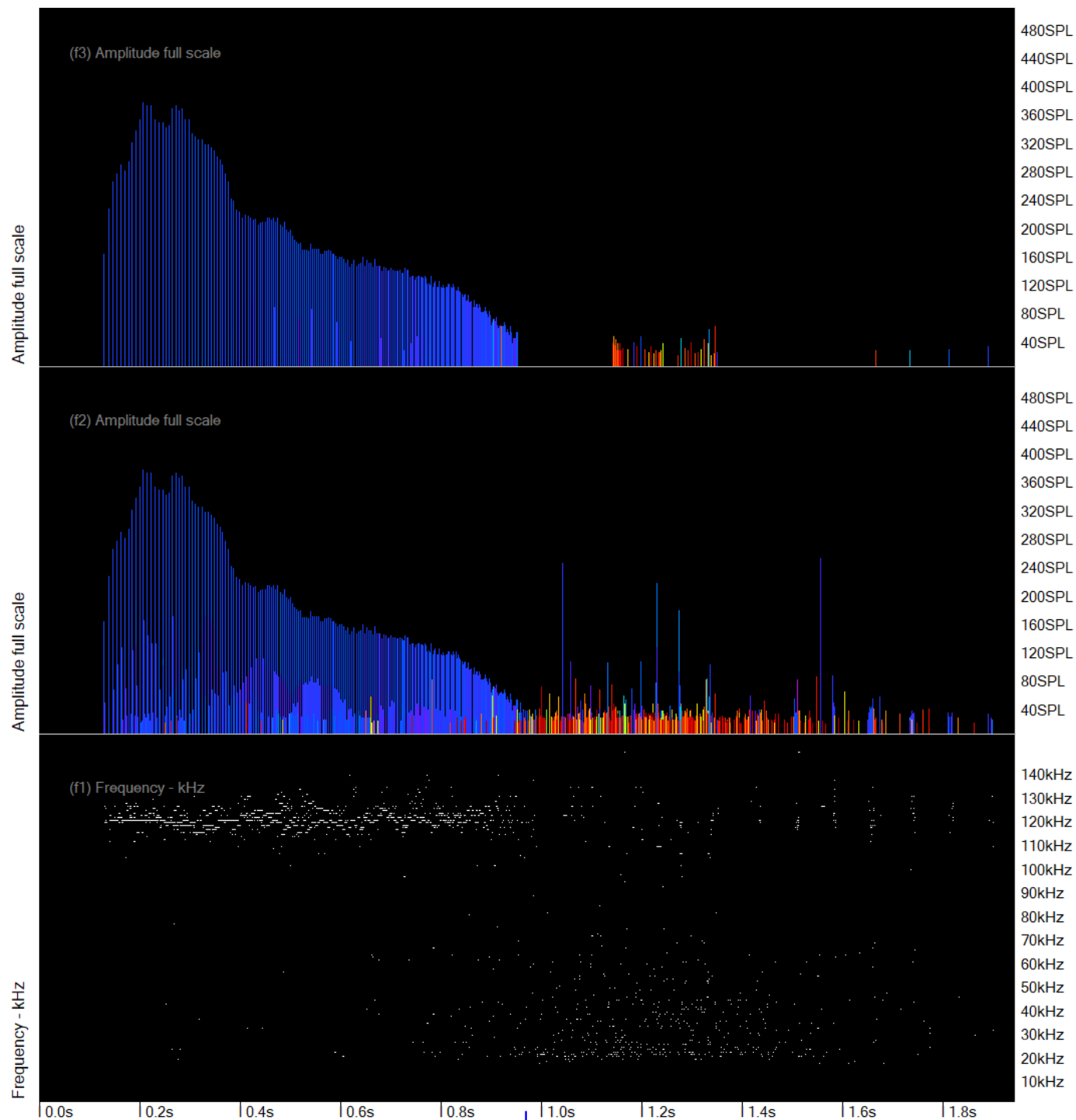
NBHF versus OTHER CETACEANS

The main problems here are:

- Some, possibly many, dolphin species can produce narrow-band high frequency clicks. This does not happen often but is seen in the large volumes of data collected by PODs. This gives false NBHF trains.
- Some 'NBHF species' make clicks that are not typical of the group and give false 'other cetacean' trains. The report from the KERNO-F classifier gives an indication of this by showing the actual modal kHz of the NBHF trains found in the whole file. (It is possible to shift the target frequency, duration and position of the loudest cycle to optimise detections for particular species and populations.)
- Possibly some 'NBHF species' make clicks that are not typical of the group and may contain much lower frequencies that appear in the multipath clusters. This gives false 'other cetacean' trains.
- All features vary and are affected by ambient noise, so there are no perfectly sharp dividing lines.

'NBHF index' is an arbitrary value derived from several click features (the code is given at the end) including the number of cycles and bandwidth, so you often don't need to look at them individually. It is often useful in this species discrimination challenge. NBHFi also uses the 'target frequency' set – the default is 120kHz, but there is variation between species and regions. If the target is too high it will tend to generate errors in which NBHF trains are put into 'other cet'. Fixing that requires re-processing through KERNO-F with the adjusted target settings.

Feature	NBHF species	Other cetaceans
wider scene	For both groups cetacean detections are of encounters in which the animals are typically within detection range for a few minutes. You may be able to see a gradual rise in amplitude as the animals approach and more rapid fade when they have gone past. You can assess whether there is a likely encounter of each species and in doing that you can include 'unclassified' trains which you may find fit nicely into an encounter by one species. Your intelligence in doing that is well above the KERNO-F classifier which has no concept of an encounter at all. So you can reject a train of one species if there is nothing much of the same species around it to see, and especially if there is good evidence of the other species guild.	
Multipath clusters	Composed of clicks within NBHF frequency range (105 - 150kHz)	May include clicks outside the NBHF frequency range
NBHFi	Often above 3. Mode = 1	Usually low values < 3, mode = 0
… you rarely need to look at the features below as they are represented in the NBHFindex		
Click duration = number of cycles	Mostly above 4	Often less than 4
Click bandwidth	Mostly below 5, mode = 1	Mostly above 5, mode = 31
Peak At	Mostly above 1, mode = 3	Mostly below 2, mode = 1

The graphic below shows some Beluga trains that are classified wrongly as NBHF train:
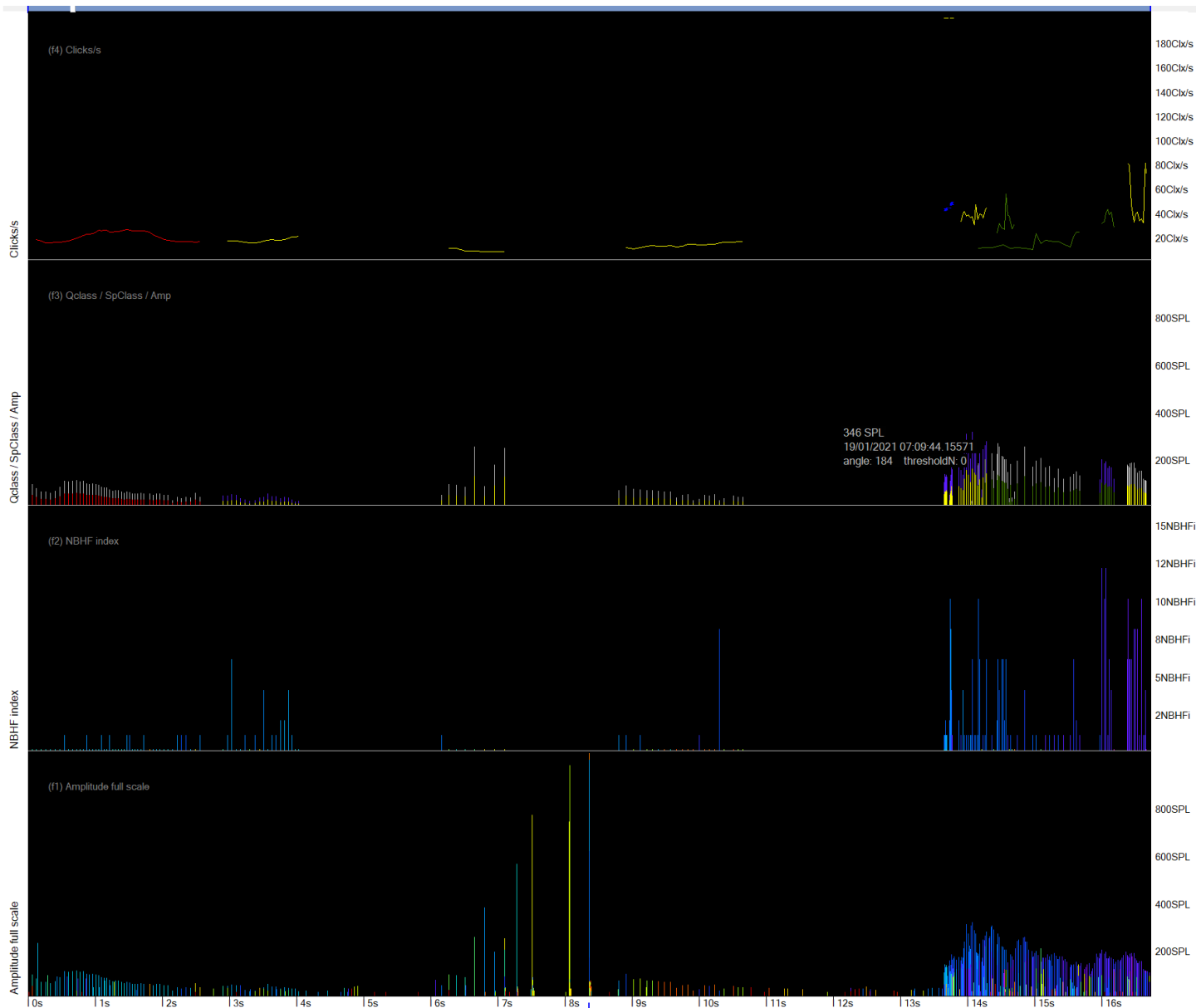
There are good 'other cetacean' trains just before the misclassified NBHF train.

Zooming in on the first two 'NBHF' trains shows:
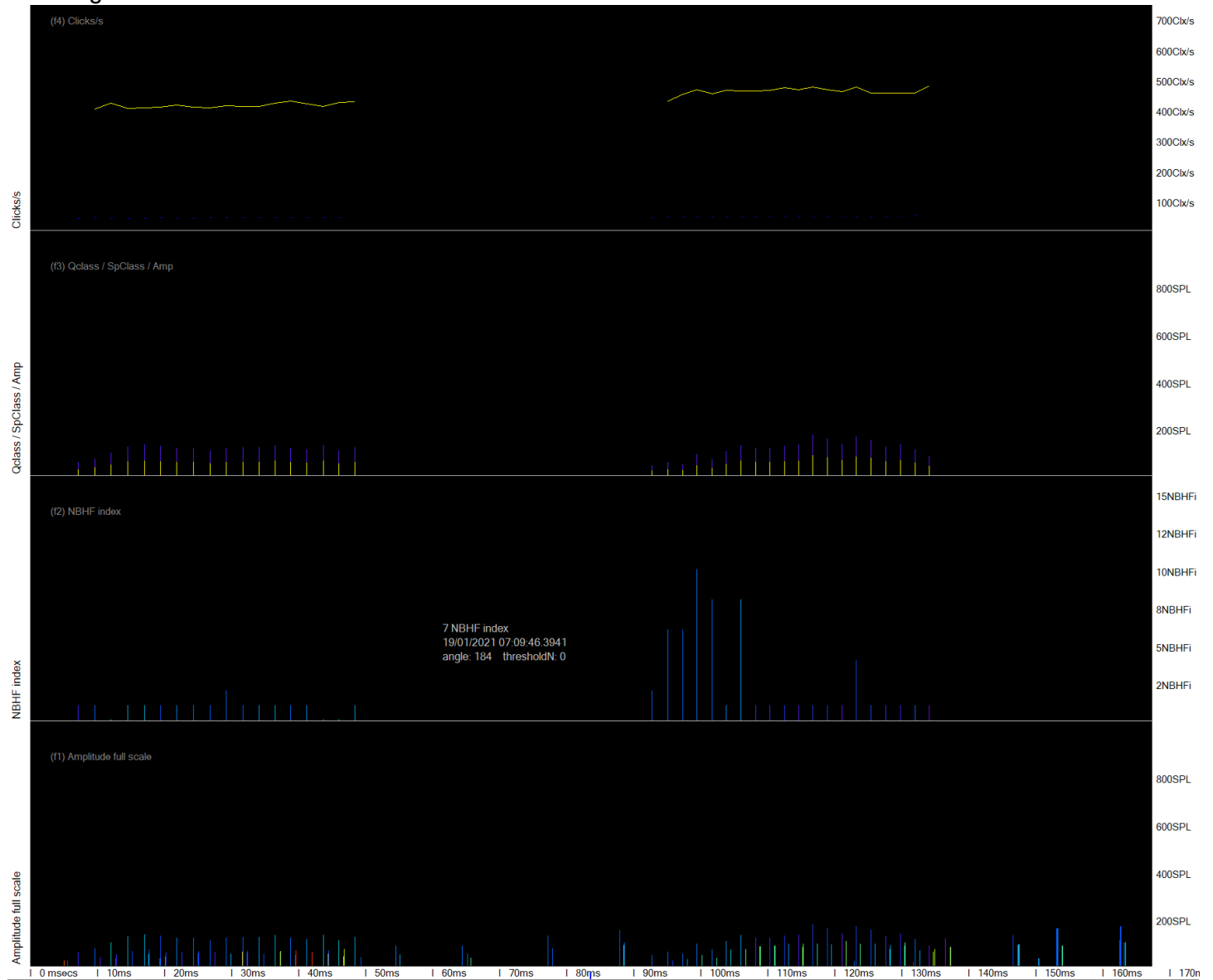
The echoes here – the lower amplitude lines between the louder clicks – show a range of frequencies with many outside the kHz range of NBHF clicks. They are clearly part of the train because of their match in timing and amplitude. So this cannot be an NBHF source.

The explanation here is that the clicks emitted from the Beluga's melon on the acoustic axis were just acceptable as an NBHF train, although many had low NBHFi values. At the same time the sound emitted off the axis included lower frequencies than would come from a porpoise, and these are logged after reflection by the sea surface.

WEAK UNKNOWN TRAIN SOURCES

WUTS

WUTS are not classified as such by KERNO-F because we have too little data, they overlap other species classes, and KERNO-F does not take a sufficiently wide view of the pattern of detections.

Instead KERNO-F gives trains a WUTS risk, and trains can be filtered by that. A high risk does not mean that it does come from a WUTS, only that it has some features of that.

Concluding it is a WUTS depends on your analysis of wider scene as described in the table below.

I'm confident WUTS are biological and that there are many species producing these sounds. Their features are quite diverse and can overlap both dolphin and NBHF trains, so they are a challenge. Suspect sources include small pelagic crustaceans, mollusc radulas, and polychaete worms in sediments.

They were first recognised in T-POD data from a ria in the SW of Britain, then in mangrove swamps, and they seem to be more numerous in places with high nutrient levels. Other risky areas and PODs among kelps – large seaweeds, or lying on the sea bed.

Feature	WUTS	The rest
wider scene	These are generally isolated trains but where the POD is on the sea bed many may be recorded, and sometimes this happens when the POD is higher in the water column. The absence of any 'good' trains of dolphins or porpoises in the surrounding minutes is the most powerful feature.	'Encounters' are usual and often there is a recognisable approach phase as clicks get louder, then trains are identified, and the end of the encounter is more abrupt (at least for cetaceans – boat sonars being vertical may fade in the same way they grew)
Multipath clusters – very important	Rare, and if present very limited i.e. one weak replicate very close in time to the primary path click.	Multipath is common with more clicks in the cluster in the middle of the real train than near the start or finish.
Amplitude	Never loud (>240), mostly below 180.	Sometimes loud
Amplitude profile of train	Mostly fairly flat but some do have rounded amplitude profiles, which is normally a feature of cetacean trains	Rounded amplitude profiles are the norm
Frequency - kHz	From the lowest logged to about 140kHz. A useful feature is a sweep in frequency through the train.	Trains below 25kHz are, by default, not classified as dolphins by KERNO-F NBHF trains don't show weak frequency sweeps, and dolphin trains rarely show smooth frequency sweeps (although they might in broadband data)
Click rate profile of train	Often monotonic. Sometimes there is a series of linked trains with a progression of click rates through the series. Smooth exponential decay of rate in downsweeps is very characteristic if present.	Varied
Click rate range	Can be very fast – near 2,000/s or down to 2/s …	The Boto uses social click trains at similarly high rates, but so far WUTS have not been identified in data from rivers.

Click features	There are, unfortunately, no features that are seen only in WUTS.	

See also:

'Validation of the F-POD - A fully automated cetacean monitoring system'
PLOS ONE   https://doi.org/10.1371/journal.pone.0293402
2023   Julia Ivanchikova, Nicholas Tregenza.

Guides and videos on www.chelonia.co.uk