C-POD cetacean click train logger

CPOD.exe is a program to operate C-PODs, and to process and display the data from them.

C-PODs are self contained ultrasound monitors that select tonal clicks and record the time and duration of each click to 5microseconds resolution.

C-PODs continuously search for periods of sound that are tonal within the range 20kHz - 160kHz. Tonal means that a narrow band of frequencies (the tone) contains more energy than all the rest of the range.

CPOD.exe analyses the data from a C-POD on a PC to find click trains and identify those trains characteristic of dolphins and porpoises.

Getting Started

The software

.... directories ....

CPOD.exe requires no installation program and can be directly run from any location including a CD or memory stick. CPOD.exe can simply be deleted along with any other POD-related files.

To read the data on the SD card from a C-POD you must have a directory with the word 'CPOD' (no space or dash) somewhere in the directory name.

To view data files they must not be read-only, so they cannot be read directly from a CD.

Download CPOD.exe, and the sample data files from www.chelonia.co.uk/cpod_downloads.htm

... viewing some data ....

Run CPOD.exe and explore it. Many items of text, buttons etc. have useful pop-up hints attached if you hold the pointer over them for a few seconds.

The menu is a multi-page dialog that is shown, or hidden, when move the mouse pointer over the dark blue bar at the top of the screen. 'Page' in this help file means a page of the menu.

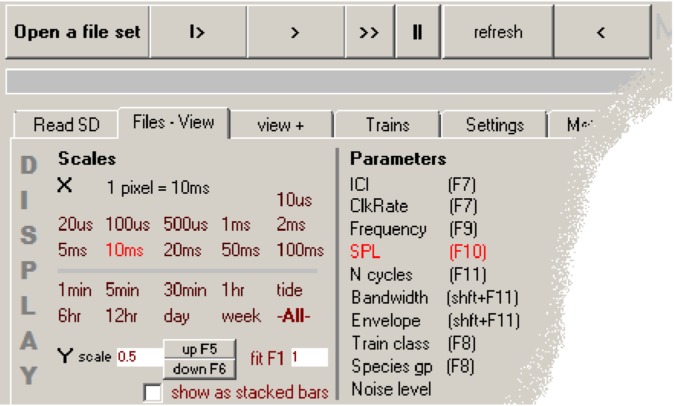
Data read from the SD card into raw data .CP1 files.

Go to the Trains page of the menu and click 'Find trains ...'

Select one or more .CP1 files

Each is then processed :

1. Collects and displays information on the frequency distribution and clustering of clicks logged in each minute.
2. Finds and classifies trains of clicks.
3. Summary data is placed on the results page of the menu, and it is useful to copy this to the clipboard and past it into a spreadsheet.



Now you can click 'Open file set' and select any of the files you have just processed.

Click the |> button to move to the start of the files, and then the > button to move onwards. You will now move through the data from one identified cetacean train to the next, skipping periods with no clicks in the file shown in the 'seek' box below the display area. Normally you will seek clicks in a CP3 file which only has the interesting clicks - those that are in trains.

Try right-clicking on the display area of the screen to see some useful tools.

The Scales section of the Files- View page has a list of available time scales:

Low-resolution displays (above the grey bar) show counts of clicks, classified according to the parameter selected.

High-resolution displays (below the grey bar) show individual clicks. At coarser scales many clicks may have occurred within the time span covered by one pixel of the screen.

Try viewing different parameters.

Hover the mouse pointer over anything that interests you on each page of the menu ....

Close the files via the Files-View page.

It will take a few days to get familiar with this program. It will allow you to analyze, access and validate data at huge speed - years of data containing many millions of clicks can be accurately analysed in a day.

Files

File Types

on SD card:

SD cards have a number of blank data files written to them in fixed locations and the POD writes real data within these shells. A 4Gb card will have DATA0.CHE, DATA1.CHE, DATA2.CHE, DATA3.CHE. Each is 1GB except the last which depends on the actual available card space. A small file SETUP.CHE will also be present and carries …. wait for it …. the setup. SD cards have a 'wear levelling algorithm' that moves file components whenever anything is written to the SD card, so any write process makes the card unusable by a C-POD. Actual data can usually be recovered, and the card can be reformatted.

on PC:

CP1 click data files

These contain all the clicks logged and they can be re-analysed by future, better, versions of CPOD.exe. You should archive these or a 'pruned' copy that has had periods at the start and finish when the POD was not in the water removed.

.CP2 files exist briefly during train detection but may remain in a directory if there is a failure during that process. You can safely delete them.

CP3 train data files - These contain only a small sub-set of the clicks in the CP1 file, and the characteristics of some of these are changed to better reflect the characteristics of a cluster of clicks.

c1m, c2m, c3m files are 'map files' that enable much faster navigation through the data files. They are generated when a file is first created from the SD card, or created by the train detection process. They can be safely deleted, as CPOD.exe can always produce a new one by re-mapping the file, and will do this automatically.

.txt - exporting data may produce large text files. Get 'EditPad' from www.just-great-software.com/ - it's brilliant! You can also paste large chunks of data into it, and then select parts to paste into Excel, much more quickly than you could have moved the stuff around in Excel itself.

File Names

To read the data that is held within some part of the If you follow the steps 1 to 4 on the Read SD page CPOD.exe will construct a filename like

Fish Factory 2009 04 30 POD414 .CP1

location + date data recovered + POD number + file number if there are multiple files + CP1 (file type.)

This allows alphabetic sorting in Windows Explorer to order files by location and date. You can edit this name before saving the file.

Keep all your files in one or a few folders so that you can carry out batch processes on them all. Keep backups of either the original CP1 files or a pruned copy.

Selection points and copying

If you set a Start or End point the software excludes all that is before, or after, that. To set these points find the end point you require and zoom in to a high resolution display in which individual clicks are shown. Right-click in the display area and select 'set selection start' ... or end ...

You can now view or export from just this selection, but you may find the most useful thing is to be able to create a new shorter copy file using the 'copy selected time period into new file(s)' button the Files - View page.

Or you can save the endpoints in the file for future use by clicking the 'save endpoints' text.

File Sets

If you click 'Open file set' (or right-click on the screen when no file is open) then currently open files are closed and you can open either a CP1 or CP3 file and the program will open the corresponding pair of files, with the CP1 as File1, and the CP3 as File2. It will also set 'Skip on' to File 2 so that use of the > button will skip along to the next train in the CP3 file that is passed by whatever filters you have set. The corresponding time period of the CP1 file will be shown.

Clicks and Trains

Clicks are the raw material in the CP1 file and can be filtered by any of their parameters –

- Frequency in kHz is measured over the first 10 cycles of the click (or all of it if it is shorter)
- End frequency in kHz estimated from the last zero-crossing-interval.
- Amplitude = sound pressure level = maximum peak to peak of the loudest cycle in the click = Pp-p (pressure, peak to peak)
- Duration – expressed as the number of cycles in the click.
- Envelope = two 4 bit values representing the slope from the first to the 5^{th} wave and the first to the maximum.
- Bandwidth – a measure of bandwidth on an arbitrary scale 0-31

Click data from any screen can be put on the clipboard by right-clicking with the mouse over the file you want to export, and selection 'Paste raw data to clipboard'. Then go to a spreadsheet and paste it using Ctrl + V.

Trains Trains are defined as sequences of clicks in which the variation between successive intervals is constrained. CP3 files contain only clicks that are in trains and are produced when a CP1 file is processed – go to the Trains page to do this. The train detection algorithm finds and classifies these trains in two ways:

Train Quality: how likely it is that this series of clicks came from an actual train source as opposed to a series arising by chance from unrelated or random sources. Q classes are Hi, Mod, Low and ? (doubtful).

Species Class: trains sources are identified here. Species classes are Porpoise, Dolphin, Other cetacean (unused at present), Unclassified, Boat Sonar. The Porpoise class includes all species producing long narrow-band clicks in the 110 – 160kHz range. The Dolphin class includes all other cetaceans and is less reliable. In general trains of low or doubtful quality should not be used in identifying dolphin trains. More below…

The right-click menu allows you to remove the species class from the first train after the pointer in click displays.

Display

In CPOD.exe the colour of most text is: **Clickable, Selected, Data, L**abels and instructions.

The main graphical display can show -

- **up to 3 files at the same time.** Files are shown with their times in register, so it helps greatly if PODs that will be close together are started as accurately as possible against the same clock.
- **different time scales** set on the X axis. These are set on Files - View page, or by zooming with the ^v arrow keys, or by left dragging across the display.
- Scales with a unit of 1 minute or above show ***counts of clicks*** in the time period selected. These can be shown as a set of graphs representing counts in different classes (bins) or in some cases as a 'stacked bar chart'. Stacked bars can be selected on the Files-View page or by holding Shift while pressing the shortcut key for the parameter to be displayed.
- Time scales of 100ms and below show ***individual click data*** with 1 pixel representing various time durations from 10microseconds up to 100milliseconds.
- **different parameters:** duration, frequency, amplitude, bandwidth, envelope of clicks. Quality, species, inter-click-interval, click rate of trains.
- **filtered data:** *trains* can be filtered by train quality, source classification, or maximum click rate. *Clicks* can be filtered by amplitude, duration, frequency, or bandwidth.
- **different Y axis scales**, changed by the Yscale Up/Down buttons or F5/F6 keys, or entering numbers into the edit box on the Files – View page. CP1 files are scaled lower than CP3 so that the two types can be 'on scale' at the same time, and the difference can be adjusted on the Files – View page.

All these can be set on the Files -View page, and are also accessible via function keys that are shown on that page and on the menu bar at the top.

Parameters

Choosing parameter to display:

F1	next screen
F7	show train click rate or inter-click-interval : toggles between these.
F8	show train quality/species class
F9	show click frequency classes +Shift for stacked bars
F10	show amplitude + Shift for stacked bars. The display shows the maximum sound pressure measured peak-to-peak, and is expressed in Pascals. It is roughly accurate for 130kHz
F11	show duration as number of cycles + Shift for stacked bars.
	F11 + Shift shows bandwidth or envelope displays and toggles through these three options

These can also be selected from the Files – View page.

Filters

These are set on the Files – View page, but can be toggled in and out:

F3	ignore/apply Quality filters
F4	ignore/apply all filters

Navigation

Moving through files:

F1	next screen
F2	show/hide results page
Drag time bar slider	to move to a date.time
Left button drag to right	Zoom in to selection
Down arrow	time zoom in, centred on mouse pointer position
Up arrow key	time zoom out, centred on mouse pointer position
Left arrow key	show data to left of pointer
Right arrow key	show data to right of pointer

Selections

Endpoints for a time period can be set, and can be saved into the file, or deleted from the file. They allow you to analyses sub-periods of the data, and to create shorter files containing only the selected time period.

They are managed via the popup menu on right-clicking on the display area, and via the clickable text and buttons at the bottom of the Files – View page.

Right-click allows you to set the time of selection start or end, which is taken as the time of the mouse pointer, so it is best to do this in the 100ms display where you see the clicks at the lowest time resolution.

Note taking

The right-click menu allows you to paste file name and times into the Results page text memo, where you can add or edit it.

Other tools

Shift C compare CP3 files

Shift F key F then steps thro' kHz classes

Shift T puts time in results memo

Shift S appends selection to a spike file

Up arrow + Shift: changes the display to 1 min, kHz, stacked bars

Filtering data

Clicks can be filtered using simple threshold filters for click parameters can be set on the Files - View page. These are mainly of use in development, but if you wanted to look for detections of, say, Vemco fish tags, you could set frequency limits a little above and little below the output frequencies of the device, and look at what has been logged in the CP1 file.

The power of C-PODs to reject false positives is due to a ‘train filter’ that detects and classifies trains of clicks within the raw data of the CP1 file and creates a CP3 file holding information only on clicks that are identified as in a train. A PowerPoint explaining the train filter is available on http://www.chelonia.co.uk/cpod_downloads.htm.

The train filter classifies trains in two distinct ways:

- Train quality is a measure of how unlikely it is that the source was not a train producing source as a train can occur by chance from random sources such a sounds from shrimps clicking independently, or collisions between sediment particles. Four quality classes are identified (and correspond to different thresholds on a ROC curve). They are high, moderate, low, doubtful.
- Species identification attempts to distinguish between boat sonars, cetacean species and some other sources. ‘Porpoise type’ clicks are produced by various dolphin species (Cephalorhynchus, Lissodelphis) and by Kogia, and the term ‘porpoise’ is used to cover all these!!

For most purposes use only High and Moderate quality detections. For porpoises it may be appropriate to use a wider range of quality classes provided some validation testing has been done.

Managing and Exporting data to Excel, Access, Word

Quick export of click data is described under ‘clicks’ above.

Choice of statistic:

For studies of habitat use and changes over weeks or longer N of detection positive minutes (DPM) per day is likely to be most useful. Larger time units, detection positive ten minutes (DP10M) or detection positive hours (DPH) are useful where the values are all below around 30% as they minimise the effects of any variation in loggers. Detection positive days (DPD, or PPD for porpoise positive days) may be useful as a way of explaining the results to non-scientists, or in some other roles.

For studies of behaviour, e.g. diel patterns N of detection positive minutes per hour is useful, plus train detail data to reveal changes in click rates in trains.

For short term EIAs you may need to look at the data in relation to each impact event.

Results can go to the Results page of the menu, from which you can copy them and paste them directly into a spreadsheet, or into a text editor (Get EditPad from www.just-great-software.com it’s a text editor that is vastly better than the microbial NotePad.). From there they can go into a file that can be imported into Access, or you can save the results directly into such a file.

Access

Importing train data into Access is easy and does not require you to set up a table first. Open a blank database and go to:

- File>Get External Data>Import> set 'files of type' to text files .txt > Import
- Then select:
- Delimited
- ... next...
- Tab
- First Row contains filed names
- ...next...
- Finish (this skips several pages)

File	Name
Trn N	Identifying number for train
Time	Time in microseconds within minute
Min	Time as minutes since the start of 1900. Makes some calculations easier.
SpClass	Dolphin, porpoise type, boat sonar, unclassified
TrClass Start	High, Medium, Low, doubtful
TrDur_us	Duration of train in microseconds
NofClx	Number of clicks in the train
avICI_us	Average inter-click-interval in microseconds
avF	Average frequency of clicks in the train kHz
avSPL	Average of peak sound pressure level of clicks, uncorrected for the frequency response of the system
MaxSPL	Maximum peak sound pressure level of any click in the train
MaxICI_us	maximum inter-click-interval in microseconds, excluding the first and last ICI
MinICI_us	minimum inter-click-interval in microseconds, excluding the first and last ICI
MMM	Mean multipath minimum – mean of the SPL of the weakest clicks in the multipath clusters identified in the CP1 file

These data are primarily intended for export to a database, e.g. Access, where they can be analysed efficiently. As every train generates all these data values the entire set may be too large for Excel.

Working with MinuteNinYear ('Min') in Access

Some handy functions that can be copied and pasted are below.

INT() is a function that converts a number into an integer - the number of whole units, so INT(5.93) = 5

Square brackets [] are used by Access to identify data fields from a table.

Hour: INT(([Min] - [Day] * 1440) / 60)

TenMinofDay: INT(([Min] - [Day] * 1440) / 10)

MinofDay: INT([Min] - [Day] * 1440)

Putting click rate values into bins:

Add a field for click rate and use an Update query to calculate it from avICI_us

CLKRATEbin: INT([ClkRate] / 10) * 10 .. gives categories of 10 / s with the lower limit as the value for trains in that ClkRate class

Or, if your prefer

CLKRATEbin: INT([ClkRate] / 10) * 10 + 5 .. gives categories of 10 / s with the central value of that ClkRate class

Larger classes can be produced by substituting 20 or 50 for 10 in the expressions above , and changing 5 to 10 or 25..

Tidal phase:

Position in cycle, as a fraction of 1: ((Min + offset) / CycleMins) - INT((Min + offset) / CycleMins)

CycleMins = the duration of the tidal cycle = approx 744 minutes. Offset = minutes between file start and the start of the tidal cycle.

Access : some more handy functions

MeanClkDur: 10 * [TotalDurClx] / [NofClx]

Criteria for daytime: (uses Hour as above)

Daytime - Hour (>5 And <18)

Nightime - Hour Not (>5 And <18)

Criteria to select trains by class:

Exclude very doubtful and boat sonar train:

Not (Like '?' Or Like 'Fxd')

Trains

Trains are defined as sequences of clicks in which the variation between successive intervals is constrained.

The level of constraint defines the maximum acceptable irregularity of the train. The POD train filter comes from an algorithm that uses approximately 40% increase or decrease in interval as the constraint. The true value for small odontocete trains is occasionally much higher but cannot be implemented in practice without very complex processing and/or a high level of false positive trains.

Trains come from:

- cetaceans - clicks made for echolocation purposes.
- boat sonars - highly regular long sounds are used by boat sonars for depth detection, fish finding etc. They mainly appear on the POD as regular bunches of clicks as they are often received as echoes because the source itself is highly directional.
- chance trains - clicks arriving at random from non-train-producing sources have a definite probability of falling into trains by chance alone.
- WUTS – weak unknown train sources.

A period of time with many random clicks will contain a huge number of possible trains. To deal with this the train filter uses a probability model that calculates the probability of trains arising by chance if clicks were arriving at the prevailing rate from non-train sources. At high click rates the filter will inevitably fail to extract 'real' trains or give an unacceptably high rate of false positives, so it is temporarily suspended.

Clicks very close together are treated as a **cluster** of echoes or multi-path duplicates and only the first of the cluster is processed.

Autocorrelation

The [ac.f]button will show an autocorrelation if the display is in low-resolution i.e. mins or longer intervals. The values graphed, and sometimes some more, will be in the Jotter and can be copied (right click, Select All, Ctrl + V) and pasted into your favourite spreadsheet.

The data used is the count of clicks in time bins that can be from 1min to 6 hours in size and are taken from the current display settings. The current filters and any end-points on the time logged are also used.

The formula is taken from 'The Analysis of Time Series' by Chris Chatfield, 6th edition, published by Chapman and Hall, 2004. p23:

Given N observations $x_1, \ldots, x_N$, on a time series, we can form $N-1$ pairs of observations, namely, $(x_1, x_2), (x_2, x_3), \ldots, (x_{N-1}, x_N)$, where each pair of observations is separated by one time interval. Regarding the first observation in each pair as one variable, and the second observation in each pair as a second variable, then, by analogy with Equation (2.2), we can measure the correlation coefficient between adjacent observations, x_t and x_{t+1}, using the formula

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\left[\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2\right]}} \qquad (2.3)$$

where

$$\bar{x}_{(1)} = \sum_{t=1}^{N-1} x_t/(N-1)$$

is the mean of the first observation in each of the $(N-1)$ pairs and so is the mean of the first $N-1$ observations, while

$$\bar{x}_{(2)} = \sum_{t=2}^{N} x_t/(N-1)$$

The formula 2.3 is used in CPOD.exe. Various modifications of this formula are often used for the same purpose, and generally represent minor changes that simplify the calculation with little effect on the output when the number of values is high. As in POD data the number of values can be low with the larger bin sizes the full version of the formula has been used here.

The formula, for r1, gives the correlation between each minute and the next one (except that it is not every minute, as the last must be omitted as there is no subsequent one).

r2 is the correlation between each time unit and the one two time units later (in this case the last two must be omitted). In CPOD.exe the number of values in the series of lag values r1,r2... is limited to the lower of 1000 or 20% of the number of bins. The larger bin sizes smooth the output where data is sparse, but as it reduces the number of data points the length of the autocorrelation may fall to 80% of the length of the data file. The same length of autocorrelation is used for all lag values.

The graph drawn is the series of values of r1,r2,... etc and is called the correlogram. The horizontal limits marked 2/SqRt(N) represent approximate 5% p values - points outside them are 95% likely to indicate a real temporal correlation between values separated by that time difference. Values commonly exceed these levels because the data is usually non-random as there are days, tides and storms as well as animal migrations, patchily distributed prey etc.

Have fun with this. Try a lot of things. Tidal patterns will match the tidal cycle grid lines without drifting and this helps to distinguish tidal cycles from diurnal cycles which is often difficult as cetaceans are often reacting to both and the cycles are of similar length x2.

Acoustic test tank notes

Tank testing is tricky! Inter-POD comparisons at sea are generally more powerful, reliable, informative and useful, but require deployments long enough to capture many hundreds of cetacean trains. Tank testing is quicker and can provide assurance that instruments are still working normally.

Tank testing notes:

- tanks are mostly subject to significant reverberation and testing has to deal with that by confining study to the period before the first echo arrives. Echoes can be louder than the direct path signal and if this is the case the tank is not suitable for POD testing.
- sound sources vary in their radial output and an accurate measurement of the sound field must be made at the POD hydrophone position and must take into account the radial variation in sensitivity of the calibration hydrophone as this will comparable with the level of standardisation of the POD.
- tests of reproducibility of results are essential.
- PODs must be soaked for 24 hours before testing.
- The temperature of the POD hydrophone must be known.
- The radial variation of the POD must be tested using at least 8 positions, preferably 16.

clean signal time

Any tank has a 'clean signal time' that is the difference in arrival time at the receiver between a signal arriving by a direct path and the earliest echoes (reflections) of that signal.

For a tank 2.5m in diameter with a source 25 cm below the surface and the receiver 1m away at the same depth the clean signal time is 42 microseconds. Then the surface echo starts to arrive and interfere with the directly received signal if it is continuing. If the signal is a pulse at 130kHz it will be continuing if the pulse produced by the acoustic transmitter is more than 5 cycles long. If the horizontal separation is reduced to 10cm the clean signal time increases to 30 cycles at 130kHz.

At the end of the clean signal time the receiver is in an interference field of echoes, that evolves over time in a complex pattern determined by all the acoustic surfaces and properties of the tank and its environment. It is extremely difficult to work with this interference or echo field because -

• at points during the time evolution of a signal it may be more intense than the directly received signal.

• at any time the intensity varies strongly between points that may be only a wavelength apart. For a 130kHz signal this wavelength in water is 12mm. The accuracy of positioning of the POD transducer in the small cylinder that contains it is plus or minus 1mm, so this could be positioned consistently within the sound field, but the field itself will only be reproducible if the water depth, source position and other acoustic parameters are comparably consistent, and that is much harder to achieve.

To maximise clean signal time put the source near the centre of the tank, and put the receiver near to it.

PODs will log echoes and if any echo is louder than the direct signal, which is a common situation, this will be the last detected signal as the transmitted intensity is reduced to find the threshold. In that case the clean signal time is irrelevant because detection is being made later in the echo field. At this threshold level the POD data will show a nice clean record of single detections at the transmitted rate, but do not be deceived! To minimise echo effects put the receiver near to the transmitter, and absorb as much sound as possible in the tank.

C-PODs, unlike T-PODs, are not calibrated to produce a standard threshold, but to give an accurate measurement of the sound pressure level of a 130kHz signal.

Electromagnetic and electrostatic interference

RF interference can be huge nuisance - it may appear as anything from big bursts of clicks to variability of the duration of logged clicks, and it varies over time. RF interference penetrates water to tank depths but is reduced if the water is saline. Screening the POD within a steel cylinder produces striking improvements in the evenness of the durations of detected acoustic signals.

Electrostatic coupling between the source and receiver can occur over surprisingly long distances because of the dielectric properties of water and may cause puzzling effects when you are exploring the sounds in your tank. A hydrophone then shows a perfect replica of the transmitted signal without any detectable time delay, and without all the distortions that are found in the acoustic transmission route. Move the receiver or transmitter around and no change in timing occurs (this assumes you have the source driver as a trigger for capturing the waveform).

other things

C-POD transducer housings must be soaked in water for several hours before testing.

Transmitting and receiving hydrophones impose their own characteristics on the electrical signal - once hit they ring like a bell, and their resonances affect the duration, frequency and bandwidth of the acoustic signal. The best signal comes from a high frequency transmitter but power will be low. In a good tank the C-POD will log 4000 clicks per second. From electronic sources bypassing the transducer the rate is much higher.