Basic Use of Filters in AnalookW

Chris Corben
Dated 9 April 2009

In AnalookW, it is possible to use filters and scans with Anabat files without having to revert to the old DOS version of Analook. The new filtering routines are very similar, and often identical to those which have been available for many years in DOS Analook, but they differ in subtle ways and they are accessed through a much better user interface.

CAVEATS

The new filter software has worked well for a number of purposes, but you should be aware that it is not yet thoroughly tested and may contain bugs which affect the accuracy of the results from it. So it is VITAL to check on any output you get from this software, to make sure it makes sense and is giving sensible results.

The second point to note is that I am currently trying to come up with better ways of measuring bat calls, so the measurements and filtering algorithms currently used are likely to be improved in the future. I will endeavour to ensure that currently available measures will work the same, but this may not be possible in all cases. For example, I would like to find a more robust way to measure Fc, and if this succeeds, it will take over the name Fc and will not necessarily generate the same values as the current version. Keep this in mind for any long-term projects, and be prepared to rerun them in the future when better parameters are available. For the moment, I have used measures which are essentially identical to those in the DOS version - as indicated by the "Legacy Mode" at the top of the measures screen (accessible by pressing "m").

File Formats

The new software uses different file formats to the older DOS software, and these are not compatible. Filter files are now *.ABF (were *.ANS) and Scan files are now *.ABS (were *.ANS). These are text files and may be read and written using a standard text editor such as Notepad. Unfortunately, I haven't had time to write a converter, so you will have to do the conversion manually.

Output files are now *.APL, *.ANL and *.AML. The uses of these files are detailed below. All these files are text files and may be read and written using a standard text editor such as Notepad.

Compatability

There are many changes to the routines which calculate the various parameters. For the most part, these will generate the same values for the same data as the old software, but this is not guaranteed, and there are likely to be small discrepancies. Under some conditions, larger discrepancies might occur, especially with complicated parameters such as Fc. This will hopefully reflect greater robustness of the newer parameters, which should be less likely to be tricked by the strange things which happen to bat calls, but this does mean that results will not be exactly the same as those obtained using older versions. So the results of different versions should never be used in the same study.

If you want to reconstruct filters used in previous versions, this should not be difficult, as at this time, the parameters are essentially the same as those used earlier. However, it cannot be guaranteed that every parameter will generate exactly the same values as in the older software, as explained above.

How Filters Work

The basic process works as follows:

1. The data from a file is scanned for strings of data points which are smoothly connected. These are called fragments. A fragment must meet certain criteria before it is processed further.
2. The fragments are tested and grouped together to form calls, provided this grouping meets certain criteria. A call can consist of several fragments, but a fragment can only belong to one call.
3. The calls are tested to see if they meet other criteria. If they do, they will be considered further by the filter. Otherwise they are rejected.
4. If required, groups of calls can be tested for parameters which depend on how calls vary within a sequence. If they pass this test, they will pass the filter and be processed further (which could involve displaying them on the screen, for example). Otherwise, they will be rejected.

Generating Filters

A filter is written using the Filter Editor, which can be accessed in various ways from the main AnalookW menu. The first step is to make a filter file by clicking on Filter, New Filter. This will open the Filter Editor with a set of parameters which have no effect on the calls to which they are applied. In other words, the default Filter which automatically comes up in the Filter Editor will do nothing, so it is then necessary to edit the parameters of the filter to make it useful. The best way to proceed from there is to immediately save the filter using File, Save As from the Filter Editor menu and then close the Filter Editor. This saves the newly named filter which does nothing. Next, load the new filter (while a suitable Anabat file is open) with Filter, Load. The name of the new filter should appear in the Status Bar at the bottom of the AnalookW screen. If it is not visible, turn on the Status Bar by clicking on View, Status Bar. You can turn the filter on and off as desired by pressing the Page Up and Page Dn keys respectively, or by clicking on Filter, Reload or Filter, None from the main menu of AnalookW.

With a filter active (see the Status Bar to confirm) you can now edit the filter by clicking on Filter, Edit. This will open the Filter Editor. If at any time you wish to see the effect of changing a parameter, you can do so by clicking on Apply at the bottom of the Filter

Editor, though this process often occurs automatically when you click the mouse in a different part of the Filter Editor dialog. You can reset the parameters on the currently open page of the Filter Editor by clicking on Set Page Defaults, and this will set all parameters on that page to values which have no effect. When you have finished editing the filter, you can shut the Filter Editor with Apply and Close, and this will also save the new values to the filter, so it is not necessary to use any other approach to save the filter file.

Choosing Filter Parameter Values

There are many possible values which can be used in the filters, and in this section, I will go over these, including noting what sensible values might be. In general, many parameters will be ignored unless the Use box beside the parameter range values is checked. Most parameters allow you to set both minimum and maximum values, which means a call recognised by the filter will only be passed if the relevant parameter is between those values. Again, these values will be ignored if the appropriate Use box is not checked.

Calls page

- Smoothness

  Establishing smoothness is the first step in recognising a bat call from other sounds. For the most part, only bat calls generate smoothly connected groups of points on the screen. This parameter decides just how smooth they must be to be considered potentially part of a bat call.

  A good value for smoothness would be 30 to 50. Higher values allow more slack, so they let worse quality calls through.

- Low Start

  If this option is checked, any call will be constrained to start at the lowest frequency of the call. Any preceding data points of higher frequency will be excluded from the call. There will be very few occasions to use this parameter.

- High Start

  If this option is checked, any call will be constrained to start at the highest frequency of the call. Any preceding data points of lower frequency will be excluded from the call. This can be very useful, as often the start of a call is detected poorly and appears at a lower frequency than it should. For the many species which start high in frequency and drop rapidly, this would be a good choice.

- All Drop

  If this option is checked, each data point must be lower than the previous data point. Therefore, the call must continually descend in frequency.

Only ever use this option for steep calls, because calls of relatively low slope will likely contain points which jump around enough in frequency that some will be higher than the previous, even if the trend of the call as a whole is to drop continually in frequency.

- Keep dot status from file

  If this option is checked, points which are turned off in each file loaded will remain off and will not be processed, irrespective of whether or not they would have formed calls which passed the filter.

  If this option is not checked, all data points in each file loaded will be tested as if they were turned on.

  This option should normally be turned off. For most purposes, the filter being used should decide what points should be turned on. A good case for using this option would be if you had previously edited the file, for example, to get rid of calls of the wrong bat. In that case, by going Alt-f, s you can save the state of the dots so that next time the file is loaded, the dots turned off will not be seen.

- Maximum Change
  - +ve

    If used, this will set the maximum increase in frequency from one data point to the next, which will be tolerated as part of a bat call. One example of where this might be useful is that the first point in a call is often improperly plotted, because it is usually in a part of the call which is of low amplitude, and which is therefore likely to be detected poorly and plotted at a lower frequency than is real.

    Any jump in frequency which violates this parameter will terminate a fragment of data points. However, it may not terminate a call, because a call can consist of multiple fragments with gaps between them due to temporary signal loss.

    The effect of this parameter will be affected by the division ratio, since points are closer together if lower division ratios are used.

  - -ve

    If used, this will set the maximum decrease in frequency from one data point to the next, which will be tolerated as part of a bat call. One example of where this might be useful is that the last point in a call is often improperly plotted, because it is usually in a part of the call which is of low amplitude, and which is therefore likely to be detected poorly and plotted at a lower frequency than is real.

Any jump in frequency which violates this parameter will terminate a fragment of data points. However, it may not terminate a call, because a call can consist of multiple fragments with gaps between them due to temporary signal loss.

The effect of this parameter will be affected by the division ratio, since points are closer together if lower division ratios are used.

- Allow harmonics
    - NOT CURRENTLY IMPLEMENTED

- Call fragments
    - NOT CURRENTLY IMPLEMENTED

- If gap between calls less than
    - NOT CURRENTLY IMPLEMENTED

Body page

The Body Page is all about measuring call parameters which are associated with the Body of the call. The Body is that portion of the call with the lowest absolute slope. In most bat calls, the slope of the call varies, because the frequency-time relationship is not precisely linear. The region where the call is "flattest" (ie closest to zero slope) is called the Body. The Body is likely to have some significance biologically, as the bat spends more time emitting frequencies in the band covered by the Body than elsewhere in the spectrum, and can therefore be expected to be most sensitive to those frequencies. There is a close connection between the Body of the call and the Frequency of Peak Energy, commonly measured by those using spectral analysis of bat calls, because both reflect the slope of the call.

Automated recognition of the Body can be very misleading. This comes about for two main reasons: firstly, noise results in short-term fluctuations in exactly what frequency is displayed at any given point of the call, and secondly, the bat may add flatter regions to other parts of the call where the slope is usually higher, perhaps in response to social contexts. Therefore, the automated measurements of Body parameters should always be regarded as estimates which may be associated with substantial errors.

- Body Over

    The Body Over parameter has a great deal of influence over the determination of the other parameters on this page.. These parameters are partly found by sliding a straight-line segment along the call starting at each individual data point, until the segment of lowest slope is found.

    Choosing a longer Body Over value has the effect of smoothing over the small irregularities in the data points so that the result is less affected by call quality. However, if the Body Over parameter is too large, the result is also likely to be less accurate. At the extreme, a Body Over value equal to the duration of the call would result in just one estimate of slope over the whole call, and it could be quite meaningless unless the call is linear.

In practice, there will usually be little reason to change the default value of 1000 microseconds (1 millisecond).

- Fc

  Fc is the Characteristic Frequency, ie the frequency at the right hand end of the portion of the call with the lowest absolute slope (the Body). Fc is the single most important parameter for species identification, as it shows a considerable robustness compared to other frequency measures, and an inherent linkage to the actual shape features of the call.

  One reason for the robustness of the Fc value is that it usually occurs where the call has a relatively high amplitude, so it is in a part of the call which can still be detected when the Bat is far enough away that much of the call is lost.

- Sc

  Sc is the Characteristic Slope, or the slope of the Body of the call. It is a vitally important feature for call characterisation, and a major way in which different species of similar Fc may differ.

  However, one feature of Sc is that it often acts as a surrogate for clutter. So a Bat in high clutter (detecting echoes from close objects) will typically raise its Sc value in response to that clutter. Therefore, the critical distinctions between species which differ inherently in Sc may only be useful if it is known under what degree of clutter those calls were emitted.

- Dc

  Dc is just the duration of the Body of the call. If the body is well-defined (ie corresponds to a substantial, nearly linear portion of the call) then Dc will have some meaning as the length of that portion. However, if the Body is poorly defined, as in a call with a continuous change in slope, the value of Dc may not mean anything. Dc = Tc -Tk.

- Tc

  Tc is the time between the start of the call and the point at which Fc is measured (ie the right hand end of the body).

- Fk

  The Body of a call is said to start at the Knee, which is usually a point where a dramatic change of slope occurs. However, many calls are gently curved, with a continuous change in slope, so that the Knee is very poorly defined. Fk is just the frequency at the Knee.

- Tk

  Tk is the time from the start of a call to the Knee. As a feature, it is often not very meaningful, because the start of the call is usually where the call is of low amplitue and its detection is highly dependent on conditions (eg how far away is the Bat).

Frequencies page

The Frequencies Page gives access to a small number of other frequency parameters. Most of these are not especially useful, as they are significantly dependent on observing conditions. This is especially true of Fmax for the many calls which start at low amplitude and their highest frequency, then sweep downwards in frequency as they build up in amplitude. The point at which such a call is first detected is highly dependent on the distance from Bat to detector, and other factors which will determine how much of the initial downsweep will be detected. This is a common reason why bat calls in a sequence appear to rise in Fmax towards the middle of the sequence, then fall again.

- Fmin

  This is simply the lowest frequency recorded in the call.

- Fmax

  This is simply the highest frequency recorded in the call.

- Sweep

  This is just the difference between the highest and lowest frequencies in the call. Sweep = Fmax - Fmin.

- Fmean

  This is the average frequency of a call, found by dividing the area under the call by the call duration.

Slopes page

- S1

  S1 is the initial slope of a Bat call. Most Bat calls start at their highest frequency then sweep steeply downwards in frequency. S1 is an attempt to quantify the steepness of that initial downsweep.

  S1 is subject to the same sort of limitations as Fmax, in that its value will depend on how much of the initial downsweep is actually detected. While this parameter may be of importance in separating some species, it has the same problem as Sc, in that most bats raise S1 in response to increasing clutter. However, unlike Sc, it is further varied by the degree to which a Bat call is detected. So for most purposes, it is not a very useful parameter.

Times page

- Dur

  The duration of a call is just the time from the beginning of a call to its end. In AnalookW, it is simply found as the difference in time between the first data point (dot) displayed and the last data point displayed.

Sequence page

- Minimum call rate required

  This parameter can be very useful for weeding out partial calls and other poorly recorded fragments. A good way to illustrate this is to consider a steep call which sweeps downwards over an extended frequency range. If the call is lost at some point in the middle of a sweep, it may be that the first part of the sweep will appear as an entire call of another species which sweeps through a smaller range of frequencies. While it may be common to encounter this sort of signal loss, it will not usually happen consistently for many consecutive calls. Therefore, a requirement that several calls must be encountered within a certain time frame will often eradicate such events.

  Another common use of this parameter is to eliminate the emergence of fake calls from the random data points associated with noise. If noise (eg insects) is sufficiently prominent, then occasionally, groups of dots will form which resemble calls of species in a similar frequency range to the noise. Again, however, it is very unlikely this will happen repeatedly in a manner closely resembling what a real Bat would produce.

  One important point to remember is that the time interval specified is that between the start of the first call and the start of the last call. So if a Bat was calling at a rate of 10 calls per second, the time interval between calls would be 100 milliseconds, but that means 5 calls would be detected in 400 ms.

Checking a Filter

An essential process in the development of filters is checking them against existing recordings to observe the effects they have. This should involve both checking that the filter does not reject calls which should pass, and that it doesn't pass calls which it should reject. These are two very different situations, and it is important to understand the consequences of both types of error.

Development of filters will usually be a compromise between rejecting too many calls or passing too many calls. Which type of error is preferable will depend on what the filter is being used for. It may be that the filter is intended to scan over data sets looking for signs of a rare species. In that case, you would be more interested in detecting any instances of the rare species but would be willing to live with many false positives. On the other hand, if the filter is being used to determine the temporal distribution of a frequently detected species, it might be more important to have a filter which is very specific to that species and throwing away many of the more ambiguous examples of that species might be of no consequence - so the emphasis would be on avoiding false positives.

Either way, it's important to understand exactly what the filter is doing. A significant aid to this is enabling Offdots (right click on the screen and check Offdots). If the Offdots are still not visible, it could be that they have been set to a colour which does not show up, so try Tools, Options, Display, and adjust the colours until the Offdots are clearly apparent. Note that for many purposes you do not want to see Offdots, since they take up space which could be better used to display calls which pass the filter. But for development of filters, this is an important tool.

One problem that often arises is that the filter behaviour is difficult to understand because of the interacting nature of the parameters being used. To avoid this, it is usually best to use as few parameters as feasible, emphasising just those of most significance to that specific filter. The less parameters you use, the easier it will be to see why a change has occurred. Frequent practice is the best way to gain a good understanding of filters and the pitfalls they present.

Generating Scans

A Scan File (*.ABS) has the basic function of specifying how several filters can be run over the same set of files. Each file is loaded and then scanned consecutively by any number of filters. An output may result from scan by a particular filter, but options can be set to only output the result of a combination of filters, as explained below.

To make a new scan, follow this procedure. First, start a new scan by clicking on Filter, New Scan. This opens a dialog asking for a location and name for the ABS file which will be produced. Make sure that any filters (*.ABF) you wish to use in the scan are already located in the folder which will contain the ABS file. Subsequently, you will see a special Scan editor dialog. This allows you to store commands in the scan file. At present, its interface is a little flakey, but following the procedure below should get you through.

1) Click on the arrow to the right of the Filter Files and choose one of the filters available. Only filters already present will be offered.

2) Choose one of the PreActions (see below).

3) Provide an Output Label if desired.

4) Click on Insert under "To add a new line". This will enter a single line into the list at the bottom of the screen, specifying the PreAction, Filter and Output you have chosen.

5) While a line is selected in the list, changing values in the boxes in the top half of the dialog will immediately change the values in the list. For this reason, to enter a new line, you should first Insert it, select it, then edit it in the boxes at the top.

6) Click OK to close the box and save the file. After doing that, you can return to the editor just by clicking on Filter, Edit Scan.

The four PreActions work as follows:

Consider the scan specified below

Prior Action	Filter Name	Output Label
reload	filter1	Filter1
reload	filter2	Filter2
exclude	filter3	Filter3
subset	filter4	Filter4
keep	filter5	Filter5

reload

After a file has been scanned by the previous filter, it is reloaded before scanning by the next filter. This means that the operation of filter2 is completely independent of the operation of filter1, and the operation of filter3 is completely independent of the operations of filter1 or filter2. So the same call could be recorded as passing both filter 1 and filter2, for example. The first filter in the list of filters must always be loaded using “reload”. The next result will be the number of calls which pass the next filter.

exclude

The next filter will scan only those calls which weren’t detected by the previous filter. Any call which has already been detected will be excluded from further consideration. The process is reset by a “reload”. This means the output of filter3 will not count any calls already detected by filter2, but could count a call already detected by filter1, since filter2 is loaded by “reload”. Therefore, filter2 and filter3 are exclusive - they cannot count the same calls. The next result will be the number of calls which pass the next filter, but didn’t pass the previous filter.

subset

The next filter will scan only those calls which were detected by the previous filter. This means that the Filter4 output will count the number of calls which passed both filter3 and filter4. In this way, the Filter4 output will count that subset of the calls passed by filter3 which also passed filter4. The next result will be the number of calls which pass the next filter, having already passed the previous filter.

keep

The next filter will scan only those calls which weren't detected by the previous filter, but the output of the next filter will contain the sum of the the outputs of that filter and the previous filter. This means, the Filter5 output will contain the sum of the calls which passed filter 4 and those which passed filter5 but not filter4. In this way, it is possible to make two different filters to detect different types of calls of the same species, but sum their results to a single figure. The next result will be the number of calls which pass the next filter, but didn't pass the previous filter, plus the number of calls which passed the previous filter.

Running Scans

To scan a bunch of files, run Filter Scan. This brings up the Scan dialog, which allows you to run the actual scan. A scan can be run without a scan file, using instead a single filter.

There are three different output types.

APL
Output is a file which counts the pulses which pass each filter summed over a period of time. The period is selected in the box to the right of "APL periods". If "Count Files" is checked, there will also be an output for the number of files recorded which contained pulses passing the filter.

ANL
Output is a list of files which contained any pulses passing a filter. An ANL file can be opened as if it is a folder of Anabat Files, using File, Open Anabat List. The advantage is that it lets you scan quickly through a list of files which passed a filter without having to find and load each one separately. These files could be in many different folders.

AML
Output is a file of measurements of pulse parameters, for all pulses which passed a particular filter. This file could be huge! It makes no sense to produce an AML file from a scan, so it can only be produced if a single filter is used.

To operate, follow the procedure below:

1) Choose the source for the files to be scanned. Typically, you would check the Tree box, then select a folder by pressing the wide button to the right of Tree. Only Anabat files can currently be scanned. A single file could be scanned, but this would only make sense with the AML output feature.

2) Choose the Scan or Filter. Check either Scan or Filter, then browse to the file desired by pressing the wide button to the right. A single filter can be used with all three output file types.

3) Choose an output location and base filename. The different output types will modify the output filename in different ways.

4) Choose the output type/types.

5) Click run to start the scan. A large scan could take a long time, so start with short scans until you get comfortable with how it works.

Philosophy of Filter Use

The use of filters is a step towards automated identification of Bat calls - the Holy Grail of Bat detection. Automated identification of some species will be easy, but many species will prove difficult to distinguish from other species which occur in the same place. The use of filters is based on the concept that most species will produce some calls which are diagnostic of that species in a given locality. But it also assumes a comprehensive knowledge of identification criteria.