Annotation of AAD Data for the SORP Annotated Library of Antarctic Recordings

Contents

Accessing the data set

Acoustic recordings from are hosted by the Australian Antarctic Data Centre at:
https://data.aad.gov.au/metadata/AcousticTrends_BlueFinLibrary

Classification scheme, call types, and metadata

Table 1 contains a list of all the call types of blue and fin whales that should be annotated within the library. Table 1 also includes the labels for each annotation, and references describing each type of call. Appendix A contains example spectrograms of each call type.

Table 1 - Classification and labelling system for blue and fin whale sounds in the SORP Library of Annotated Recordings

Call Type	Label	References	Description
Antarctic blue whale unit A;	Bm Ant-A	(Rankin et al. 2005, Miller et al. 2014)	Antarctic blue whale unit A (i.e. 26 Hz tone) without other units. Usually repeated at 60-70 s intervals.
Antarctic blue whale unit AB	Bm Ant-B	(Rankin et al. 2005, Miller et al. 2014)	Antarctic blue whale unit A tone followed by partial or full inter-tone interval downsweep (unit B)
Antarctic blue whale z-call; (AKA 3 unit vocalisation)	Bm Ant-Z	(Širović et al. 2004, Rankin et al. 2005)	Antarctic blue whale 'z-call' with units A and C present (and unit B either present or absent)
Blue whale FM (AKA D-calls)	Bm D	(Rankin et al. 2005)	Any frequency modulated calls from blue whales. Typically, but not always longer in duration and lower in frequency than FM calls from fin and minke whales.
Fin whale 20 Hz pulse	Bp 20Hz	(Watkins et al. 1987)	20 Hz fin whale pulse without substantial energy at higher frequencies

Fin whale FM calls (AKA 'high frequency' downsweep; AKA 40 Hz pulse)	Bp Downsweep	(Gedamke & Robinson 2010, Širović et al. 2013, Ou et al. 2016)	Frequency modulated, usually downswept calls believed to be produced by fin whales. Usually, but not always shorter in duration and slightly higher in frequency than FM calls produced by blue whales.
Fin whale 20 Hz pulse secondary energy at higher frequencies	Bp Higher_calls	(Gedamke 2009, Gedamke & Robinson 2010)	Fin whale 20 Hz pulses that have secondary energy at higher frequencies (for example between 80-100 Hz).
Unidentified bioacoustic sounds	Unidentified	-	Sounds that are believed to be biologically produced, but are not recognisable. That is, the analyst believes they are biological, but is not confident in assigning them to one of the above labels.

A metadata spreadsheet that has a row for each hour long chunk, and columns for qualitative assessment of different types of sounds or species has been created and included with the dataset. For now, the possible options for each entry are *absent*, *possible*, *likely*, *present*, and *not_inspected*. Absent means the analyst is certain that sound wasn't there, and present means the analyst is certain that it was there. On a scale of 0-3, absent would be 0, possible 1, likely 2, and present 3. Noise fields are for a qualitative description of the main types of noise sources and overall intensity, e.g. regular seismic noise with moderate SNR, or intermittent ice cracking with high and moderate SNR.

Workflow and settings for manual inspection and annotation of audio

Presets and Workspaces

In addition to the acoustic data files and metadata spreadsheet, a zip file called 'Raven Pro 1.5.zip' has been included on the FTP server. This contains Raven Workspaces, Presets, and Selection Tables. I believe these Raven settings files can be installed by unzipping 'Raven Pro 1.5.zip' into %homedrive%%homepath%\Raven Pro 1.5 (for example for me it would be: "c:\Users\brian_mil\Raven Pro 1.5\"). After downloading the wav files, you should be able to load the *SORP Acoustic Trends - Casey2014.wsp* or *Kerguelen2014.wsp* workspaces. Upon loading the workspace the analyst will then need to direct Raven to the location where they have stored the downloaded wav files.

Workflow

Because the sounds of blue and fin whales are very low frequency and many calls, such as 20 Hz and Z calls, are not audible, it is expected that the analyst will spend most of the time viewing spectrograms rather than listening to the audio. Thus, the selection and tracking of spectrogram settings is of paramount importance in creating a reproducible workflow and result (see section below). However, it is expected that the analyst will listen to the sound (possibly played back at a rate faster than originally recorded) as part of the classification process if they have uncertainty about the classification of a potential detection.

Annotation of calls should be done by first creating a separate Selection Table in Raven for each classification (already included in the Casey2014 and Kerguelen2014 workspaces). To create a new

annotation, the analyst should first activate the corresponding Selection Table with the correct classification, and then simply mark the bounds of call on the spectrogram using Raven's *Create Selection Mode*. They should continue creating annotations for all of those type of calls that they see on the screen. When finished annotating all of the calls of one classification type, they would then repeat the process for the next call type. The analyst would only advance to the next page of data after annotating all detections on the page. This will generate a separate Measurement file plaintext for each classification and workspace dataset. Raven's Measurement files appear to be tab-delimited text files. Within each Measurement file and only the filename provides an indication of the classification.

Call types that are not listed in Table 1 need not be marked on the spectrogram. However, the analyst should note in the spreadsheet whether they noticed any of those sounds in that hour. For example, Casey 2013-12-25_06-00-00, is dominated by leopard seal calls. For that hour, the analyst would annotate the calls from blue and fin whales listed in Table 1, and at the end of the hour would then note in the excel spreadsheet that they were certain that there were pinnipeds calls. Another example: Kerguelen 2014-08-05_11-00-00 has lots of intense minke whale calls, so the analyst would mark the boundaries of all the blue and fin calls, and then enter 'certain' for bioduck in the spreadsheet. When labelling blue and fin whale calls in that file, the analyst might also find that there are calls that are possibly minke whale downsweeps too, so they would enter 'possible' for minke downsweeps in the spreadsheet. The purpose of these extra columns in the spreadsheet is to start to give us a better idea of prevalent sound/noise sources in each file.

If the analyst is not confident or experienced at recognising minke, sei, pinnipeds, humpback, seismic, or ice sounds then they should enter *'not_inspected'* for these columns in the metadata spreadsheet. Ideally, columns should not be left blank, and it is likely that blank columns will be treated as 'not_inspected' rather than 'absent'.

Measurements made for each annotation

The default measurements for a Raven Selection Table do not provide sufficient information to re-create Selections when reloading a workspace. Instead, we have included a Measurement Preset called "SORP_Annotated_Library_Raven_Measurement_List_1.0". The measurements contained in this list provide sufficient information to re-create Annotations in the workspace, and also provide sufficient information to load the portion of audio from a third-party program such as Matlab or R. These measurements are: *Selection, View, Channel, Begin File, End File, Begin Time (s), End Time (s), Beg File Samp (samples), End File Samp (samples), Begin Date Time, Delta Time (s), Low Freq (Hz), High Freq (Hz)*

For these workspaces Raven has been configured to *Use clock-time axis labels* (Figure 1) so that the measurements for Begin Date Time will reference the actual date and time of sounds at each site, rather than just the relative time since the start of the first file in the 'Raven Sound Folder'. Using this option will also provide handy additional options in the page browser for navigating to a specific year, month, day, hour, etc.

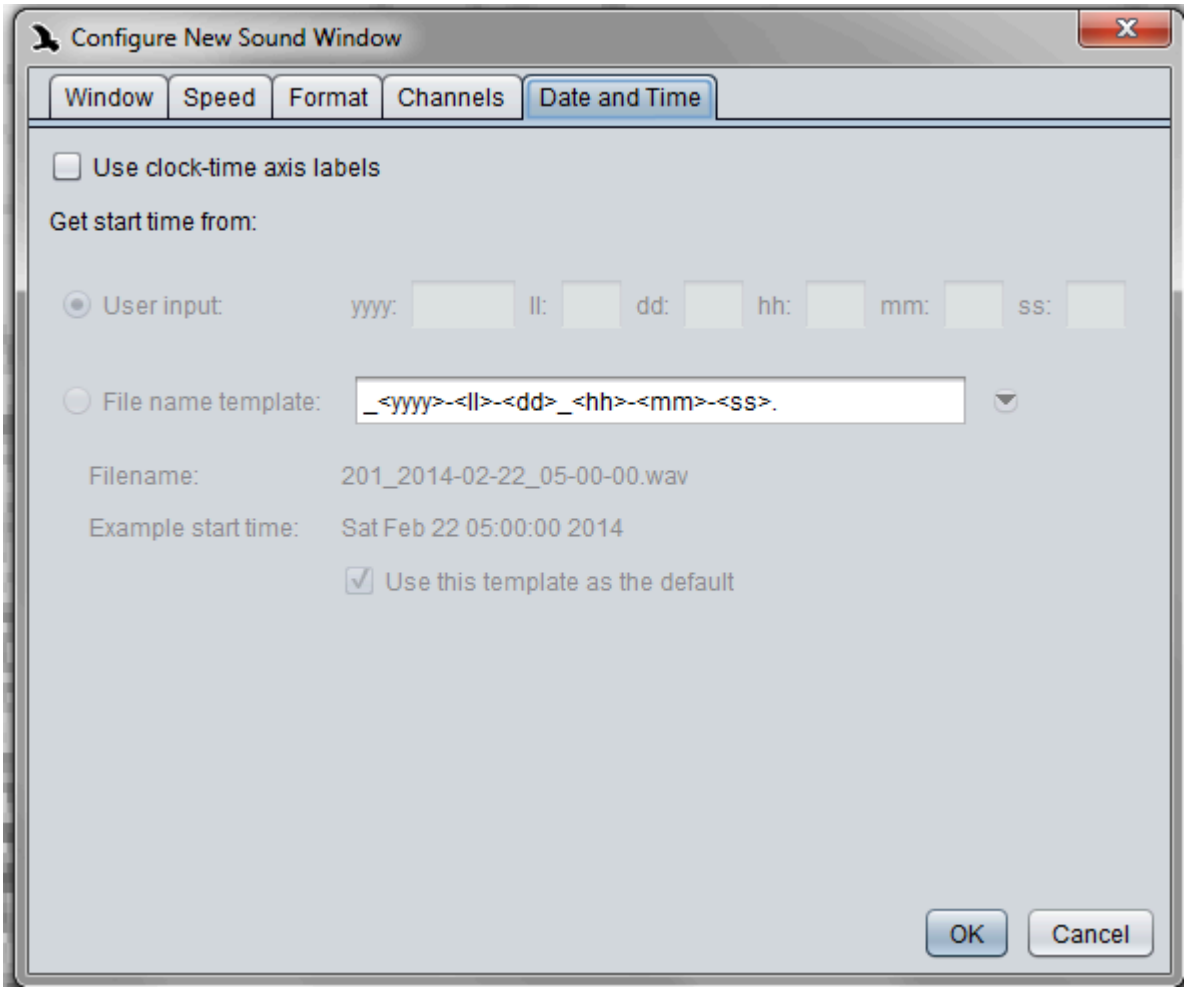

Figure 1 - Configuring Raven to use clock-time for the time axis.

Spectrogram parameters

The spectrogram settings in each workspace have been chosen to achieve approximately 1 Hz frequency spacing, and 85% overlap between spectrogram time-slices. For the 1000 Hz wav files used in this analysis this corresponds to a *Window size* and *DFT size* of 1024 samples, *Hop size* of 154, and the *3 dB Filter bandwidth* of 1.40 Hz. The Raven setting for *Clipping* was not used, nor was the setting for *Averaging*. Raven's *Page Size* sets the maximum duration of data that can be viewed on screen. Recommended *Page Size* for annotation of blue and fin whale calls is 60-120 s. The analyst should feel free to zoom in further to inspect potential detections.

It appears that Raven 1.5 does not provide a mechanism to apply hydrophone or system calibrations so that the on-screen units are in dB re 1 µPa. Rather, the amplitude of a given sample in the waveform is scaled to the bit-depth (e.g. between -32768 and 32768 for 16-bit files). Power measurements (e.g. spectrum and spectrogram amplitudes) use these scaled amplitudes with a reference value of 1.0 (unity) for calculating power in decibels.

The Raven toolbar includes options for Brightness and Contrast that control how color scaling is mapped to the intensity of the spectrogram. It is recommended that analysts DO NOT directly adjust the Brightness or Contrast to control the color scale. Instead, the spectrogram scale should be adjusted via the Power Threshold Floor and Ceiling (Figure 2). This can be achieved by selecting *View->Configure Brightness and Contrast*, which will open a dialog (Figure 2) that providing access to

the Power Threshold Floor and Ceiling. While almost certainly related via some mathematical formula, the dB measures of the floor and the ceiling seem more useful to me than the seemingly arbitrary values of brightness and contrast on the slider.

To achieve consistent and reproducible annotations, it is recommended that the analyst adjust the floor and the ceiling at the start of each hour chunk/file as follows. When initially adjusting the sliders, the analyst should try to keep the difference between the *Floor* and the *Ceiling* between 30 and 40 dB. Adjust the floor aiming to keep the background levels in the band 0 to 125 Hz are about 20% white. This should be roughly equivalent to setting the noise floor for the 0-125 Hz band to the 20th percentile for that time window. The analyst should then adjust the *Ceiling* so that it is between 30 to 40 dB above the background. This will result in sounds with intensities more than 30-40 dB above the floor appearing solid black. If within a file the noise changes drastically and the analysts finds large portions of the screen are black, or more than 20% is white, then they should to readjust the floor and ceiling as above. The analyst should note in the metadata spreadsheet the floor and ceiling values for each file, and any time you change them in the metadata spreadsheet. Since the values for floor and ceiling are logged, I believe that logging the brightness and contrast might be redundant, so these need not be recorded.

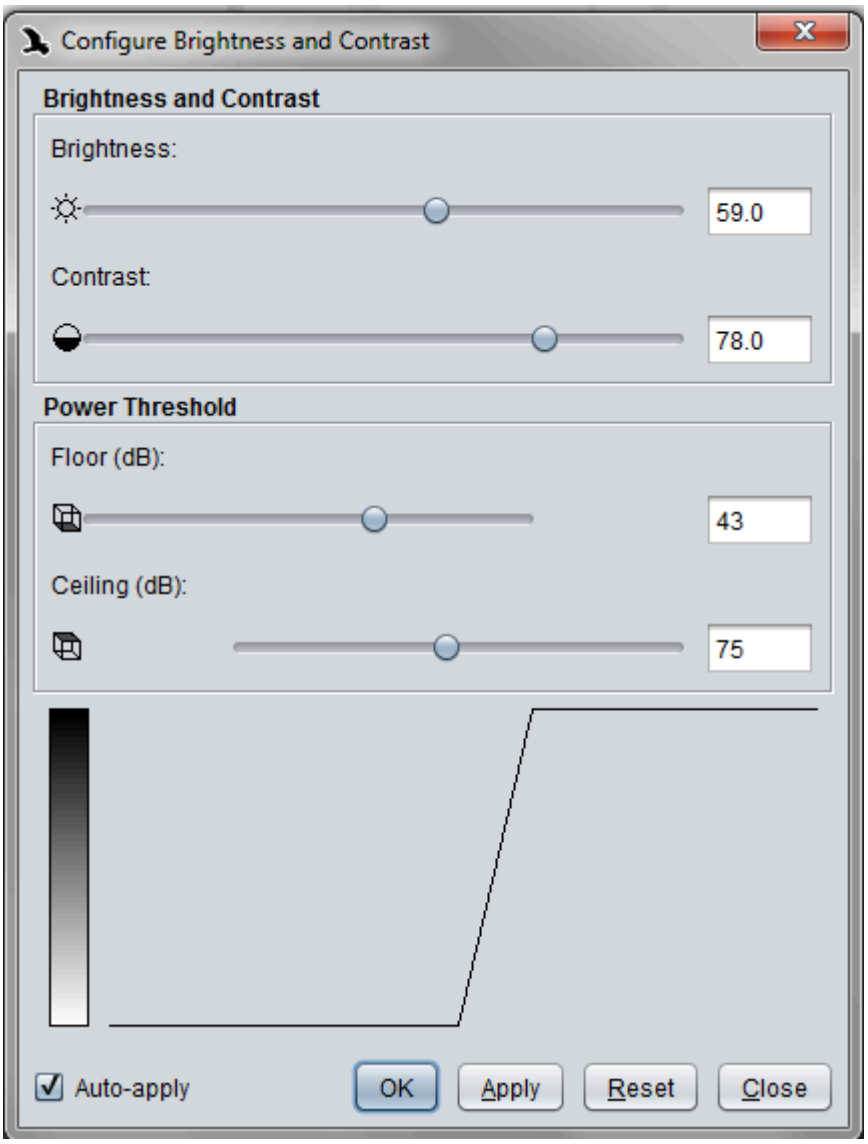

Figure 2 - Scaling Raven spectrograms using Power Threshold Floor and Ceiling instead of Brightness and Contrast

Appendix A – Spectrograms of blue and fin whale calls

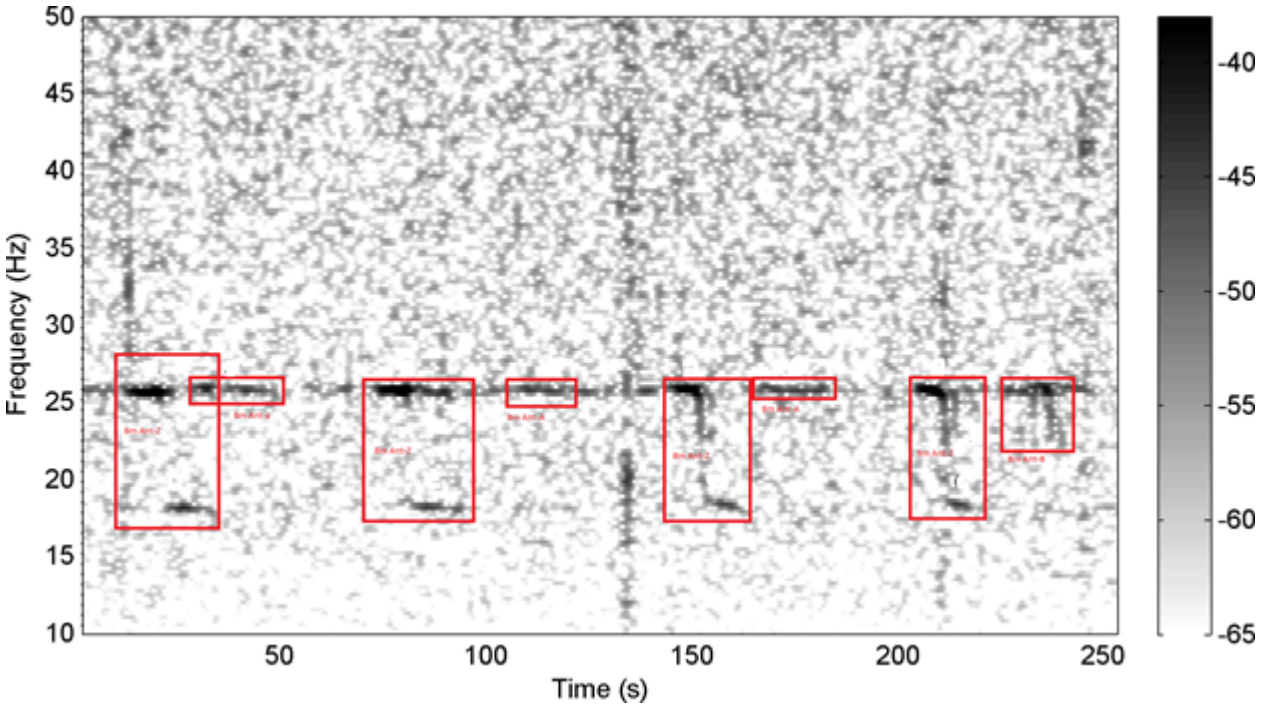

Figure 3 - Example showing Bm Ant-Z (1st, 3rd, 5th & 7th red boxes starting from the left), Bm Ant-A (2nd,4th, and 6th red boxes from left), and Bm Ant-B (8th red box from left)

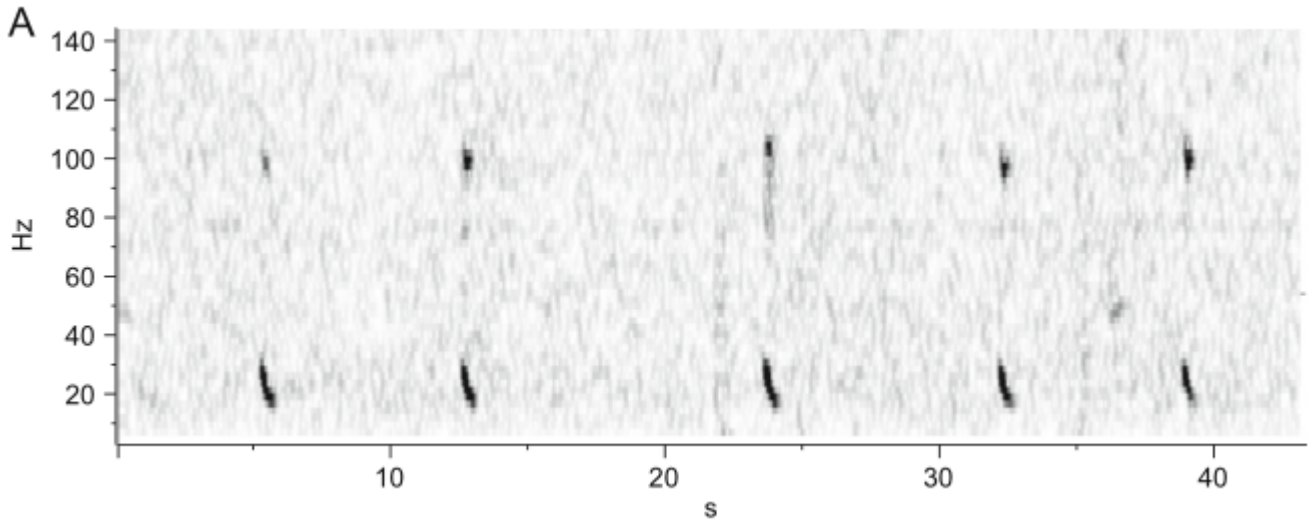

Figure 4 - Example of Bp Higher_calls: fin whale 20 Hz pulses with secondary energy at 100 Hz. Reproduction of Figure 5 from Gedamke & Robinson 2010.

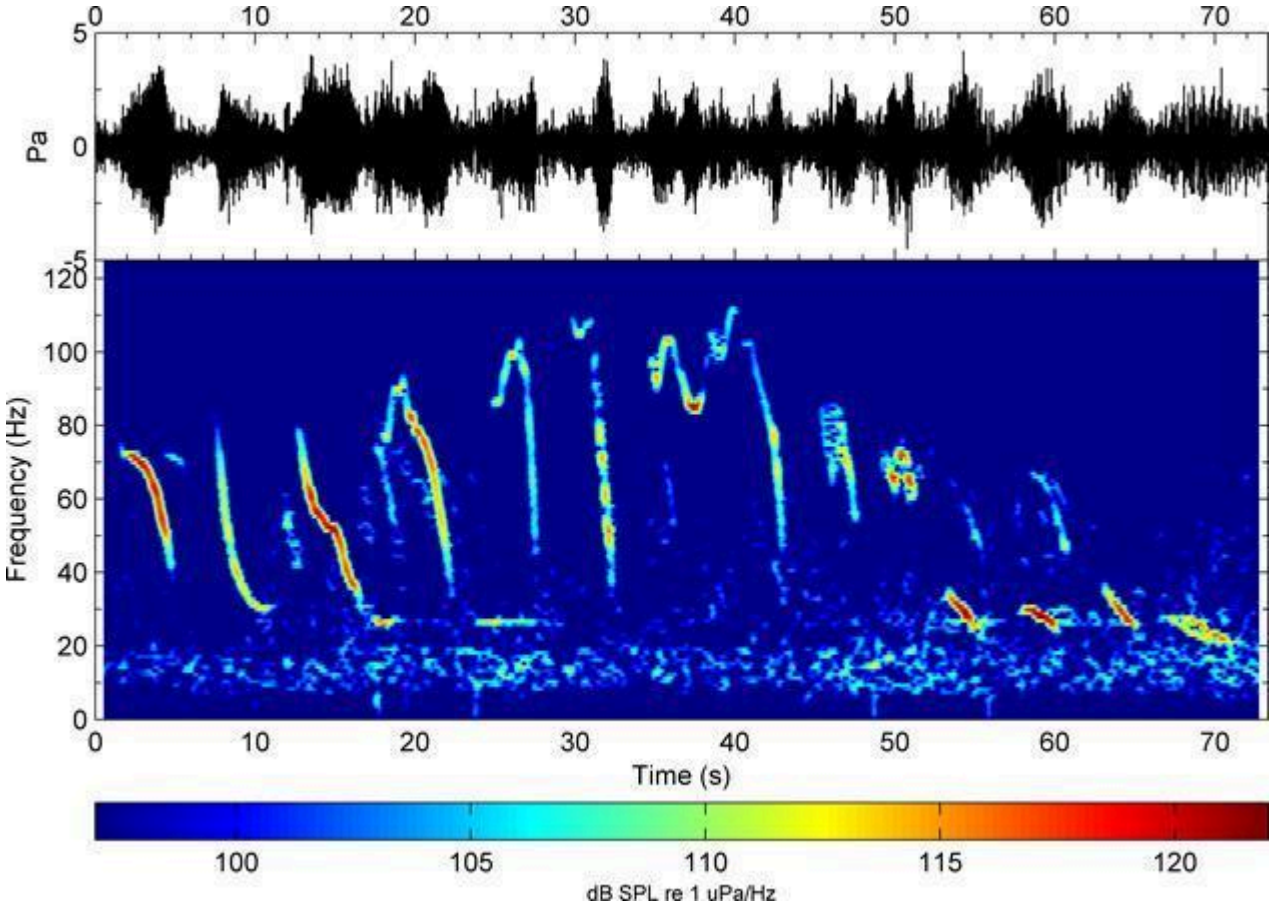

Figure 5 - Examples of the variety of forms of FM calls produced in the presence of blue whales during the 2015 Antarctic Ecosystems Voyage. Presentation comprises individual FM/D calls that have been concatenated into a single audio file to produce a single spectrogram (i.e. each FM sweep was recorded on separate occasions and the figure does not represent a sequence found in nature).

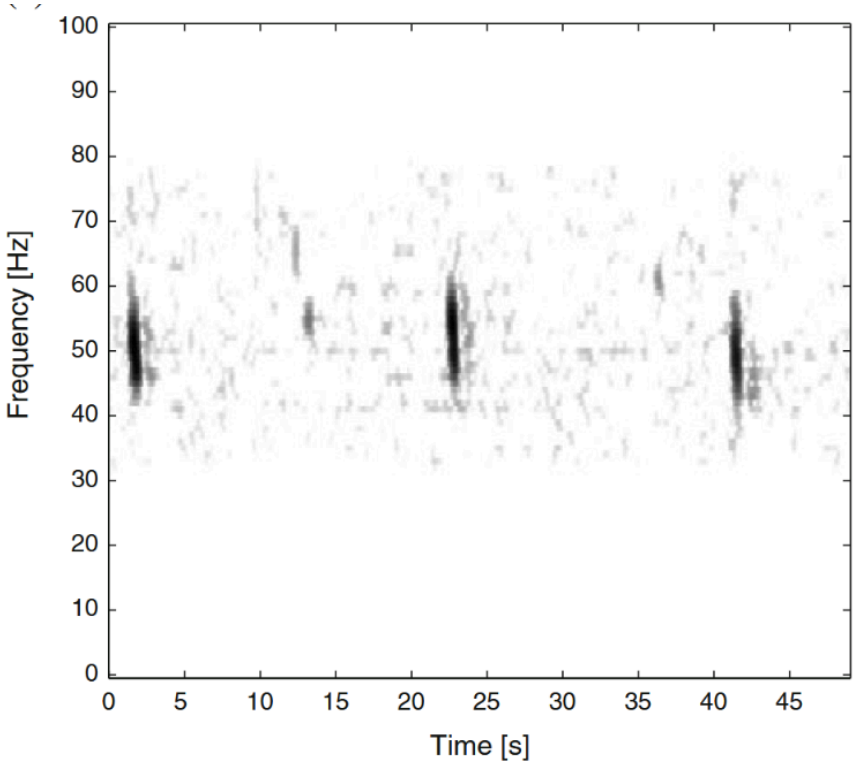

Figure 6 (left) - Examples of fin whale downsweeps from reproduced from Širović et al (2013) under a creative commons license. Original caption: 40-Hz calls (2000-point FFT with 95 % overlap, band-pass filtered between 30 and 80 Hz; sample rate 2 kHz) recorded on April 18, 2010. All recordings were collected off Southern California