

Exercise: Spectrogram Cross-Correlation

Summary: This exercise demonstrates the process of calculating the similarity between sounds using a process called spectrogram cross-correlation. We will walk through the process using a folder of sounds produced by a Red-eyed Vireo.

Learning objectives:

- Master the mechanics of applying spectrogram cross-correlation
- Be able to predict situations where spectrogram cross-correlation will be informative and when it is poorly suited and unlikely to perform well

Prerequisite knowledge:

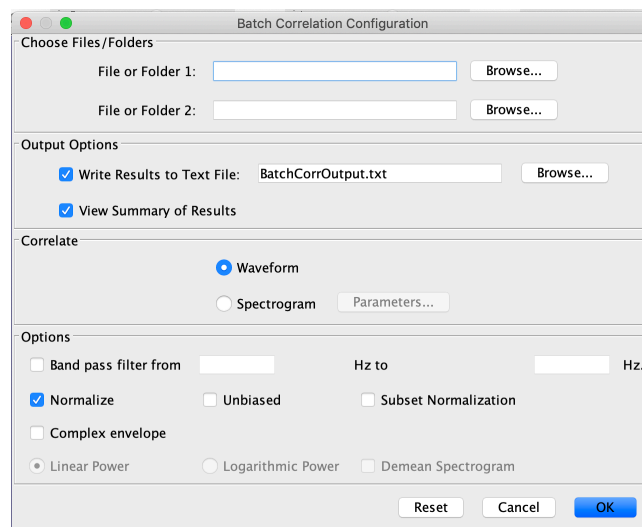
- Basic familiarity with spectrogram parameters and terminology

Instructions:

1. Open Raven.
2. Under the menu options, find Tools -> Batch Correlator.

Note: There is also a Tools -> Correlator option, but this can only compare two individual sound files.

3. You will now see an options window that looks like the image below.



4. To choose the first folder of clips to correlate, click the "Browse" button and navigate to the folder of Red-eyed Vireo selections. These selections were created using the Band Limited Energy Detector in a previous exercise. Here, we will start with this folder of 91 selections, each containing a Red-eyed Vireo vocalization.

- Next, use the browse button next to Folder 2 to select the second set of clips to correlate. In this case, we will use exactly the same set of red-eyed vireo recordings.

It is common to correlate all clips against all other clips from the same recording session in order to find clips that are similar to each other. Alternatively, you may have a set of pre-defined call types that you wish to correlate with new recordings. In this case, the two folders would be different, with one folder including reference clips and one folder containing segments from new recordings.

- By default, the output will be written to a file called "BatchCorrOutput.txt." If you wish, you can provide a more informative title for your output matrix (e.g. Red-eyed vireo cross-correlation).
- Under the correlate options, select Spectrogram. For spectrogram parameters, be aware that the options that you choose will affect the correlation values. Generally, the spectrogram parameters that you use to view the sounds are also appropriate for correlating the sounds.

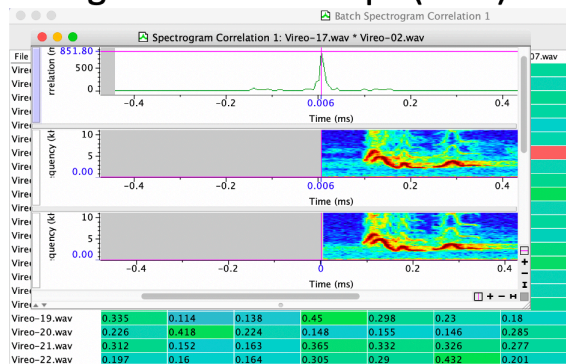
Note: If you wanted to correlate the time signature and ignore frequency differences, you can choose the Waveform option.

- Under the remaining options, use the default settings. Most of these options are meant for relatively specialized applications and are rarely used.
- After correlations have been calculated, click 'close' to complete the dialog window.
- You will now see a window that provides correlation values among detections. Click the "Colors" radio button in the center of the top panel to color cells by the correlation value.

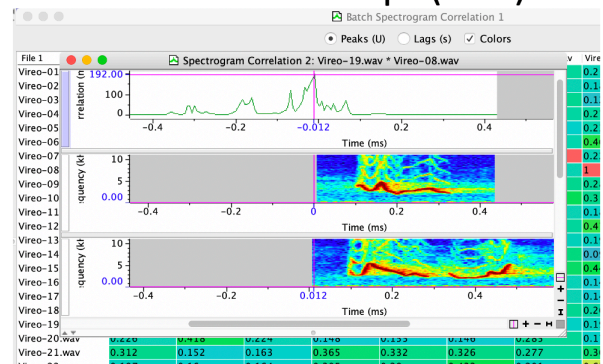
File 1	File 2 >>	Vireo-01.wav	Vireo-02.wav	Vireo-03.wav	Vireo-04.wav	Vireo-05.wav	Vireo-06.wav	Vireo-07.wav	Vireo-08.wav	Vireo-09.wav	Vireo-10.wav
Vireo-01.wav		1									
Vireo-02.wav		0.204	1								
Vireo-03.wav		0.223	0.183	1							
Vireo-04.wav		0.448	0.122	0.164	1						
Vireo-05.wav		0.315	0.221	0.266	0.453	1					
Vireo-06.wav		0.337	0.114	0.143	0.311	0.258	1				
Vireo-07.wav		0.283	0.224	0.305	0.269	0.17	0.231	1			
Vireo-08.wav		0.212	0.187	0.129	0.271	0.235	0.466	0.221	1		
Vireo-09.wav		0.322	0.166	0.205	0.342	0.256	0.336	0.309	0.289	1	
Vireo-10.wav		0.337	0.279	0.172	0.24	0.204	0.213	0.417	0.312	0.183	1
Vireo-11.wav		0.29	0.178	0.211	0.213	0.226	0.252	0.226	0.183	0.242	0.235
Vireo-12.wav		0.373	0.113	0.118	0.524	0.313	0.514	0.28	0.472	0.266	0.179
Vireo-13.wav		0.223	0.133	0.095	0.23	0.175	0.196	0.206	0.193	0.206	0.206
Vireo-14.wav		0.118	0.146	0.15	0.102	0.11	0.058	0.34	0.098	0.423	0.204
Vireo-15.wav		0.309	0.285	0.19	0.28	0.198	0.294	0.393	0.447	0.138	0.329
Vireo-16.wav		0.306	0.313	0.35	0.152	0.195	0.12	0.215	0.147	0.18	0.225
Vireo-17.wav		0.209	0.852	0.218	0.167	0.201	0.101	0.227	0.149	0.196	0.269
Vireo-18.wav		0.332	0.082	0.114	0.392	0.281	0.293	0.298	0.26	0.301	0.247
Vireo-19.wav		0.335	0.114	0.138	0.45	0.298	0.23	0.18	0.192	0.362	0.169
Vireo-20.wav		0.226	0.418	0.224	0.148	0.155	0.146	0.285	0.176	0.193	0.302
Vireo-21.wav		0.312	0.152	0.163	0.365	0.332	0.326	0.277	0.364	0.395	0.151
Vireo-22.wav		0.197	0.16	0.164	0.305	0.29	0.432	0.201	0.678	0.246	0.325
Vireo-23.wav		0.356	0.251	0.218	0.225	0.177	0.196	0.365	0.313	0.173	0.621
Vireo-24.wav		0.227	0.307	0.138	0.131	0.117	0.191	0.174	0.12	0.168	0.185
Vireo-25.wav		0.362	0.138	0.323	0.509	0.908	0.278	0.213	0.197	0.293	0.209

- Note that there is a diagonal line of cells with a correlation value of 1. This reflects the perfect correlation when a clip is correlated with itself. Note the presence of other high correlation values that reflects similarity between different clips.
- Double click on a cell in order to see the two clips that were correlated and the alignment between them. The top panel shows the correlation that is achieved at different time offsets between the clip. If there is one good alignment, there is typically one strong peak, but particularly for calls with repeated pulses, there may be multiple alignments that would produce relatively high correlations between the two sounds.

High correlation clips (0.85)



Low correlation clips (0.19)



- These correlation matrices can now be used with hierarchical cluster analysis or other analytical approaches to reveal groups of similar sounds. These approaches can be applied through a variety of software such as R, Matlab, and Python.

By correlating a library of labeled sounds against a folder of unlabeled sounds, it is also possible to assign sounds to existing categories using an objective criterion.

Synthesis

- What are some of the factors that may cause spectrogram cross-correlation to be artificially high?
- What are some factors that may lead to low correlation values, even when sounds appear similar?

Answers

- If two sounds have similar continuous background noise, spectrogram cross-correlation may be high simply because the backgrounds match and produce high correlation values.
- Because spectrogram cross-correlation relies on pixel-by-pixel comparison, if sounds are shifted slightly in frequency or stretched/compressed slightly in time, it can lead to low cross-correlation values between seemingly similar sounds.