

REPORT ON EXPEDITION / PROJECT / CONFERENCE

Expedition/Project/ Conference Title:	Bioacoustics of the long-range howl in mantled howler monkeys (<i>Alouatta palliata mexicana</i>): building a reviewable call database at the Instituto de Neuroetologia, Universidad Veracruzana
Travel Dates:	25/07/2026 – 05/09/2026
Location:	Xalapa and Los Tuxtlas, Veracruz, Mexico. Instituto de Neuroetologia, Universidad Veracruzana; field site at La Flor de Catemaco, on the edge of the Los Tuxtlas Biosphere Reserve.
Group member(s):	None, solo placement, hosted by the research team of Prof. Pedro A. D. Dias at the Instituto de Neuroetologia.
Aims:	1. To be trained in the field and analytical methods used to study howler monkey vocal behaviour: acoustic recording, behavioural observation protocols, and acoustic measurement. 2. To turn the team's raw recordings into a trustworthy, reviewable database of who called and when — the foundation any later work on vocal structure depends on. 3. To test, with unsupervised machine learning, whether the long-range howl is a single graded call or several distinct types, and to leave the lab with tools they can keep using after I left.
Photography consent form attached: (please refer to your award letter)	☐ Yes ☒ No (no identifiable individuals appear in the photographs included)

OUTCOME

With the support of the James Rennie Bequest and the SEB Travel Grant, which together covered my flights, I spent six weeks at the Instituto de Neuroetologia of the Universidad Veracruzana, working with Prof. Pedro A. D. Dias's team on the vocal behaviour of wild mantled howler monkeys (*Alouatta palliata mexicana*) in Los Tuxtlas, Veracruz. Howler monkeys are among the loudest land mammals, and their long-range howls carry for a kilometre or more through the forest; what those howls are for, and whether they are built from distinct call types, is still unsettled. My aim was to help move the team's recordings from hours of audio on a hard drive to a database that can be analysed.

A month of recordings

The team records with a Zoom F3 in 32-bit float, a format with effectively no clipping ceiling and no noise floor, so both a roar from a nearby male and a distant howl survive intact. Over the season this produced 61 recordings, 78.8 hours of audio across 17 days of fieldwork. I took part in the fieldwork, learned the observation protocols the group uses for scan and focal sampling, and then took responsibility for the analysis side. Every recording carries the recorder's own clock inside the file, which is what makes it possible to place a call not merely at “3,412 seconds into file 7” but at a real time of day that can be matched to the observers' notes on who was calling and what was happening.

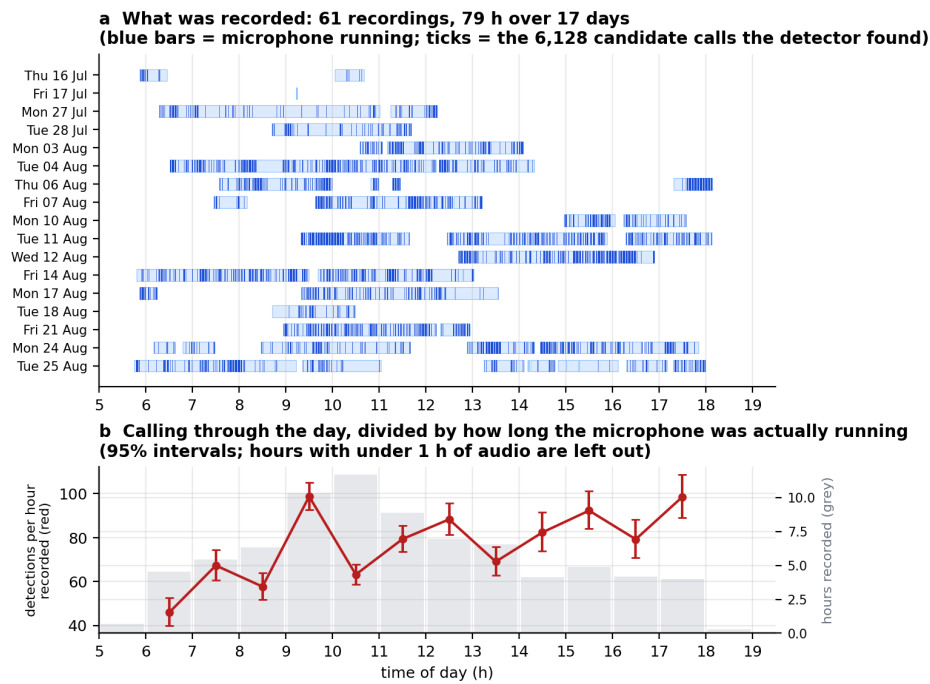


Figure 1: The month of recordings. (a) Each blue bar is one recording, drawn at the time of day it was made; the ticks are the 6,128 candidate calls found automatically. (b) Those candidates divided by how many hours the microphone was actually running in each hour of the day — without that correction the pattern would mostly show when we chose to record. These are automatic detections before human review, so they still include false alarms.

Finding the calls without telling the computer what a monkey sounds like

The first problem is that nobody can listen to 79 hours of forest. I built a detector on a deliberately simple principle: it measures the energy in the 250–1,200 Hz band where howler calls live, compares it with a band higher up where insects and birds dominate, and asks not “is this loud?” but “is this unusual for right here, right now?”, comparing every moment against a rolling median of the surrounding minute. A cicada chorus is very loud, but loud in the wrong place, and is rejected; a distant howl in a quiet dawn is kept. Two thresholds are used rather than one, so a call counts only if its loudest part is convincing, while its boundaries extend outwards into the quiet onset and decay that a single threshold would cut off.

The detector has no examples and no model of a howler call, which is the point: it cannot preferentially find calls resembling the examples we happen to own, so it cannot quietly bias the database towards the calls we already know about. Over the season it found 6,128 candidate events in 51 of the 61 recordings, about 1.3 per minute.

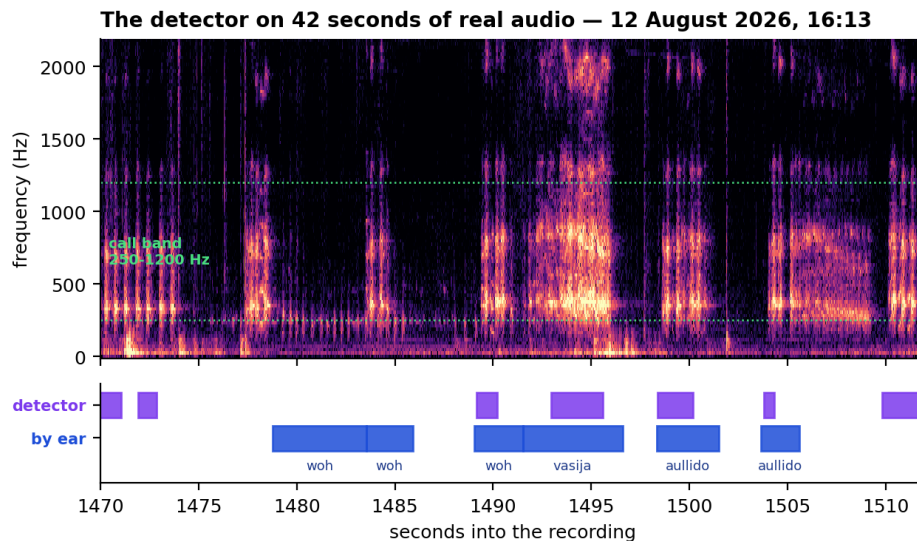


Figure 2: The detector on 42 seconds of real audio from 12 August. Above, the spectrogram: the stacked horizontal bands low in the frequency range are howls, and the dotted lines mark the band the detector measures. Below, what the detector proposed (purple) against the calls I marked by ear (blue), with the names used in the field. The loud howls line up well; the quieter calls are where the detector struggles.

From detections to a database people can trust

A detection is not a call, it is a place worth listening to. The second piece of work was therefore a review system. It is a single self-contained notebook that takes a folder of recordings, finds the calls, writes one master database ordered by date and time, and exports label files for Praat and Audacity, the two programs the lab already uses; it can be emailed to a colleague and run on their own laptop. Alongside it I built review pages, one per day of fieldwork, opened by double-clicking, with the spectrogram and audio of every detection embedded so they work offline. Each detection is judged with a single keypress and, where the automatic boundaries cut a howl short, its true extent is dragged across the spectrogram by hand. Those corrections are stored beside the detector's original values rather than replacing them, so the disagreement between machine and ear stays measurable.

Reviewing a first set of clips myself produced 144 hand-marked calls, and that comparison showed where the method stands: the detector finds the loud howls reliably, but the quiet calls exchanged at close range, the soft “oddles” and grunts, are close to invisible to it. That is now a documented limitation with a measured size, rather than an unknown.

Are there different kinds of howl?

The scientific question was whether the long-range howl is one graded call or several types. I measured 88 hand-marked howls in two deliberately independent ways: sixteen interpretable acoustic measurements (pitch, bandwidth, how tonal or noisy the call is, the shape of its rise and fall), and separately the spectrogram itself, time-normalised and stripped of loudness. Loudness was excluded throughout, because how loud a howl is mostly reflects how far away the animal was; clusters built on it would map distance from the microphone very convincingly and mean nothing biologically. Both descriptions were clustered by three different algorithms, with my own labels held out of the process and used only afterwards as an independent check.

The groups that came out are more separated than the same analysis finds in data with the structure deliberately destroyed (Figure 3), and they are demonstrably not a disguised measure of loudness. But the two independent descriptions did not agree with each other about which calls belong together, so any group I named would be a property of the measurement I happened to choose, not of the animal. The honest conclusion is that no howl sub-type can be claimed on this evidence, most likely because 88 calls from a handful of recordings is a thin sample for a question of this kind.

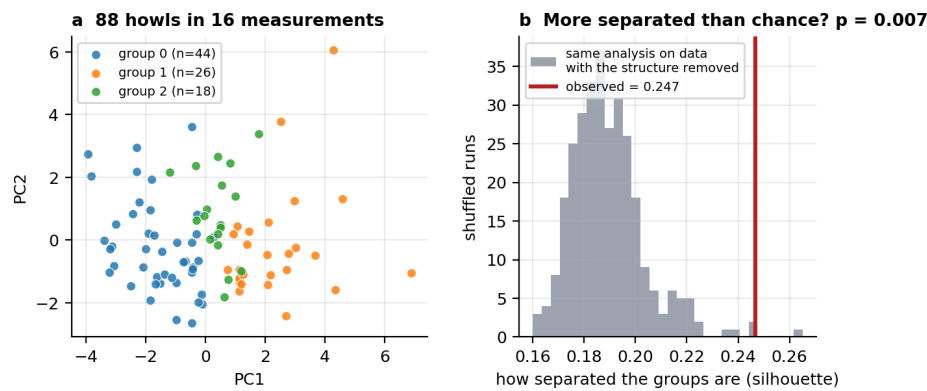


Figure 3: (a) The 88 howls described by sixteen acoustic measurements and grouped without reference to my own labels. (b) The test that decides whether that grouping means anything: the grey distribution is the same analysis run on data whose structure has been destroyed. The real data beat it ($p = 0.007$) — but a second, independent description of the same calls disagreed about which howls belong together, so no call type is claimed.

A second tool, for the rest of the lab

Working alongside the students I found a problem that had nothing to do with my own project: behavioural observations are recorded as spoken Spanish narration, and turning hours of that into coded spreadsheet rows is done by hand. I built a tool that transcribes those recordings into text time-stamped to the real clock, flagging the words it is unsure about for a human to check, and a second stage that drafts the rows of the lab's own coding sheet. Tested on a real 30-minute focal recording it transcribed the protocol phrases cleanly and agreed with the student's own coding on roughly three-quarters of behavioural categories, not a replacement for the observer, but a first draft that removes most of the typing.

What I take from it, and what happens next

I arrived expecting to learn field techniques, and did: how a focal follow is run, how vocalisations are recorded and scored, and how a long-term primate study works day to day. What I did not expect to learn, and value more, is how easily an analysis can produce a confident and wrong answer, and what it costs to check one properly. Every claim here is now paired with a test against data where the effect has been removed, and running those tests changed several of my conclusions. I can defend every step from a raw waveform to a clustering result, which was not true of me in July.

The work is not finished, and I have not presented it as though it were: the review of all 6,128 detections is built and running, sixteen pages, one per day of fieldwork, of which the first is complete (35 detections judged, 18 confirmed howls with their boundaries set by ear). Finishing the remaining days and re-running the clustering on that larger and less biased set is the next step. Everything needed to carry on without me is already in the team's hands: the detector and database tool with its documentation, the review pages, the transcription tool, the confirmed example calls, and a written record of which figure was measured on which data and when, so that nothing has to be taken on trust.

I am very grateful to the James Rennie Bequest. The grant paid for the flights that made the placement possible, and the experience has confirmed that research in animal communication is the direction I want to take.

Photographs

