Supplementary Materials for

Automated audiovisual behavior recognition in wild primates

Max Bain*, Arsha Nagrani, Daniel Schofield, Sophie Berdugo, Joana Bessa, Jake Owen, Kimberley J. Hockings, Tetsuro Matsuzawa, Misato Hayashi, Dora Biro, Susana Carvalho, Andrew Zisserman

*Corresponding author. Email: maxbain@robots.ox.ac.uk

Published 12 November 2021, Sci. Adv. 7, eabi4883 (2021)
DOI: 10.1126/sciadv.abi4883

The PDF file includes:

- Supplementary Text
- Figs. S1 to S3
- Table S1
- Legend for Movie S1

Other Supplementary Material for this manuscript includes the following:

- Movie S1
Fig. S1. Detection and Tracking Results: (A) Precision Recall Curves for the individual body detectors for Bossou (left) and Cantanhez (right), with their Average Precision (AP) scores labelled. (B) Qualitative examples of detections (detector confidences shown above each detection) and body tracks on both the Bossou and Cantanhez video footage. Photo credit: Kyoto University, Primate Research Institute; The Cantanhez Chimpanzee Project
Fig. S2. Screenshots of the annotation tool, the VGG Image Annotator (VIA). VIA is a browser based, open-source annotation software which we modified for this purpose, facilitating temporal annotation of video.
Fig. S3. Proportion of buttress-drumming bouts by context for Adult Males and Females. Context here consists of agnostic display, feeding and travel. Proportions for each context are approximately equal for both males and females.
Table S1. Training and Testing Data Statistics for (A) Nut-Cracking and (B) Buttress-Drumming for the Bossou and Cantahez video footage respectively.

| (A) | Split                        | Sampled Years | Duration of video (hours) | # Nut-cracking Sequences | # Tracklets | # Tracklets containing nut-cracking |
|-----|------------------------------|---------------|---------------------------|--------------------------|-------------|----------------------------------|
|     | Training                     | 2012          | 10.5                      | 864                      | 11,400      | 508                              |
|     | Training (Audio proposals)   | 2004, 2008, 2012 | 1.9                      | 1,455                    | 9,564       | 1,786                            |
|     | Testing                      | 2013          | 2.6                       | 417                      | 3,699       | 154                              |
|     | Analysis                     | 2012, 2013    | 13.1                      | 1,281                    | 15,099      | 662                              |

| (B) | Split                        | Sampled Communities | Duration of video (hours) | # Drumming Bouts | # Tracklets | # Tracklets containing drumming  |
|-----|------------------------------|---------------------|---------------------------|------------------|-------------|----------------------------------|
|     | Training                     | Cabante, Caiquene-Cadique | 32.8                      | 778              | 2,526       | 1,249                            |
|     | Testing                      | Lautchande          | 1.7                       | 166              | 521         | 191                              |
|     | Analysis                     | Cabante, Caiquene-Cadique, Lautchande, Cambeque | 10.8                      | 1,251             | -           | -                                |
Movie. S1. Demonstration of automatic behaviour classification pipeline to example videos. Videos are first processed by the body detection tracking stage, followed by audiovisual behaviour classification.