From Pretrain to Primate: Decoding Chimp Nocturnes for Wellbeing

Lara Narbona Sabaté
Research Department
Fundació Mona
17457 Girona, Spain
lara.narbona@gmail.com

Pablo R. Ayuso
Research Department
Fundació Mona
17457 Girona, Spain
pablo.bio.uah@hotmail.com

Abstract

Sanctuaries play a crucial role in ensuring primate welfare, yet limited economic and human resources constrain animal behavior from being systematically monitored. While daytime behavior can be observed directly, nocturnal behavior (equally critical for wellbeing, social dynamics, and management) remains largely inaccessible. Chimpanzees vocalize at night in high arousal contexts, shaping circadian rhythms, emotional states, and subsequent cooperation in management routines. Current approaches to accessing this information, such as manual audio and video inspection, are unsustainable for long-term sanctuary operations.

We propose an automated framework to analyze nocturnal vocal behavior using self-supervised speech representation learning. By retraining HuBERT, to which we added a sequential classification layer, we aim to detect and classify vocal interactions from audio recordings. This frugal and scalable approach provides ethologically meaningful insights, enabling informed decisions from keepers without requiring intensive human labor or costly infrastructure.

As a case study, we focus on Fundació MONA, a European primate sanctuary where manual night analyses have already informed management decisions. Automation will extend these efforts, offering timely insights that directly enhance chimpanzee wellbeing, and will allow to share the methodology ai to other sanctuaries worldwide.

1 Introduction

Primate sanctuaries are specialized facilities that provide sustained, high-quality, lifelong care and welfare for primates victims of wildlife trade or other illegal human activities [27]. Within these contexts, evidence-based husbandry and behavioral management are central to maintaining welfare: structured enrichment, social management, and cooperative training reduce stress, facilitate veterinary procedures, and support the expression of species-typical behaviors [5] [20] [24] [29]. Because sanctuaries often operate with limited financial and staffing resources, there is a need for scalable, low-overhead monitoring approaches that translate directly into improved management practices.

Animal wellbeing in captivity depends fundamentally on detailed knowledge of their behavior [17][7]. Behavioral indicators provide early-warning signals of stress, illness, or conflict, often before clinical symptoms appear [6]. They can also help keepers make more informed decisions, providing information not only for immediate, day-to-day husbandry, such as the order in which animals are moved or fed, but also for broader strategic decisions, such as whether certain individuals should be separated overnight.

Contemporary welfare frameworks emphasize that wellbeing is a 24/7 endeavor, extending beyond standard working hours [4]. While behavior can be monitored during the day with human observers,

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: .

nocturnal activity remains far less studied. Nighttime welfare is no less important than daytime, particularly for highly social species like chimpanzees, where their social lives continue during the night [1] [8]. During the night, they also perform species-specific behaviors such as nesting, crucial to chimpanzee wellbeing [19]. Neglecting the night risks missing critical components of their behavioral and social worlds.

The importance of the nights becomes especially evident when considering the ways past social interactions shape future behavior. Chimpanzees are highly sensitive to their social history: they track both competitive and cooperative exchanges across domains, and base future interactions on these prior experiences [11] [10] [15] [21] [30]. Nocturnal events may therefore strongly influence daytime cooperation, conflict, and even willingness to engage in management tasks. Social dynamics also affect animals' emotional states, which in turn shape participation in cooperative management routines such as positive reinforcement training (PRT). Voluntary management is critical for welfare, as animals' chosen participation in husbandry procedures reduces distress and creates safer conditions for both humans and animals [20] [23] [25]. Underpinning these practices is the notion of agency and control: providing primates with meaningful choice enhances wellbeing and partially restores autonomy lost in captivity [18] [26] [29].

Moreover, proper sleep–wake cycles (i.e. circadian rhythms) are essential for cognitive function, emotional regulation, and healthy aging in primates. Disrupted circadian rhythms or poor-quality sleep contribute to psychological stress and accelerate cognitive decline [1] [3] [13]. This point is especially pressing in European sanctuaries, where chimpanzee populations are aging and thus increasingly vulnerable to sleep disruption and its downstream health effects [12].

Reliable knowledge of nocturnal events is thus a key question for chimpanzee welfare. It can guide informed day-to-day decisions, aligning practice with evidence rather than guesswork. Ultimately, the goal is to adapt routines to enhance wellbeing in a holistic manner, integrating behavioral evidence across the full 24-hour cycle.

However, achieving this insight is far from straightforward. After-hours staffing is limited for obvious reasons of cost, practicality, and human welfare. To track nocturnal behavior, sanctuaries must rely on recording devices. Passive acoustic monitoring (PAM) represents a particularly attractive option: it requires less infrastructure than multi-camera systems, offers high sensitivity to socially salient sound events, and is far more affordable for under-resourced sanctuaries [9] [32]. Chimpanzees are an especially promising target for PAM, since they are highly vocal, particularly in high-arousal contexts such as aggression, alarm, or displays [31] [33]. By focusing on vocal behavior, it is possible to capture much of the behavioral landscape that most strongly impacts welfare.

Yet manual review of long-duration audio is infeasible. Sanctuaries could quickly accumulate hundreds of hours of overnight recordings, and detailed annotation would be infeasible. Automated pipelines offer a solution. Advances in deep learning for bioacoustics have reduced human effort by orders of magnitude while retaining ecologically relevant signals, making near-real-time nocturnal summaries technically and economically feasible (e.g. [22] [28]). Delivering automated overnight reports would allow keepers to adapt management proactively rather than reactively.

This is where self-supervised learning enters the sanctuary. Models such as wav2vec 2.0 and HuBERT have revolutionized speech processing by learning rich acoustic representations from unlabeled audio [2] [14] [16]. These methods discover latent sound units and temporal patterns without requiring costly annotation, and they transfer effectively to downstream tasks with limited data. For sanctuary bioacoustics, where labeled corpora are sparse, this paradigm is particularly well-suited. Having these models at hand is very useful for transfer learning in a situation where labeled data are sparse. Moreover, we reframe nocturnal vocalizations as a sequential problem, rather than isolated call classification, allowing us to capture the temporal dynamics of chimpanzee social life. By using a trained backbone (HuBERT) to which we added a classification layer, we can segment long overnight audio streams into units, detect bursts of high-arousal activity, and transitions between vocal states that may foreshadow chimpanzee behavior the following day. This combination of self-supervised representation learning with sequence modeling and classification bridges the gap between individual calls and behaviorally meaningful episodes, providing sanctuary keepers with an automated early-warning tool for welfare-relevant dynamics.

1.1 Case study: Fundació Mona

The potential of this approach can be illustrated with a case study from Fundació MONA, a European sanctuary housing 14 chimpanzees divided into three groups. In small groups, interpersonal relations are particularly important. Individuals sleep in indoor rooms at night while accessing outdoor enclosures by day. Rather than sleeping continuously, chimpanzees wake, interact, and resettle throughout the night.

Previous studies at Fundació MONA have demonstrated that these nocturnal interactions influence daytime behavior and management decisions [1]. For example, in one of the groups, Mutamba, nighttime recordings revealed systematic aggression by the beta male (Juan) against the lowest-ranking male (Marco). Although Marco did not retaliate, repeated episodes were detected through manual review of video and audio. Sanctuary staff responded by rearranging sleeping arrangements, pairing Juan with the alpha male (Bongo) instead of Marco. Following this change, group dynamics improved.

Importantly, this insight required extensive manual review, an unsustainable practice for a small team aiming for a systematic, continuous and long-term monitoring. Automating detection of such nocturnal events would make them visible in near-real time, providing managers with actionable knowledge without consuming scarce human resources.

Automated nocturnal monitoring thus offers a dual benefit: it enables sanctuaries like Fundació MONA to improve welfare immediately through better-informed decisions, and it establishes a scalable, transferable workflow for sanctuaries worldwide. Using self-supervised speech models for sequential bioacoustic analysis, we can transform how sanctuaries understand the night, shifting welfare monitoring from labor-intensive hindsight to proactive, data-driven foresight.

2 Methods

2.1 Study site and subjects

The study is being conducted at Fundació MONA, a sanctuary located in Riudellots de la Selva (Girona, Spain), dedicated to the rehabilitation of chimpanzees (*Pan troglodytes*) rescued from the entertainment industry and illegal pet trade. The study subjects will be 14 adult chimpanzees (7 females) living in 3 mixed-sex groups. During the day they have access to an outdoor enclosure (two naturalistic measuring 2420 m^2 and 3220 m^2 and a non-naturalistic measuring 25 m^2), and during the night to an indoor non-naturalistic rooms (measuring between 25–30 m^2).

2.2 Data collection

Nocturnal vocalizations are recorded using a digital stereo audio recorder (TASCAM®DR-07X, TEAC Corporation©, Tokyo, Japan) installed inside the chimpanzees' sleeping rooms. The device operates nightly between 8:30 p.m. and 8:30 a.m., with the automatic recording function triggering whenever sound levels exceeds -24 dBFS and stopping after five seconds below this threshold. This configuration was validated to capture a wide range of chimpanzee sounds, including drumming, alarm calls, screaming, and agonistic displays. Each recording is automatically timestamped with date and time, enabling systematic alignment with daily observations. This setup allows for the unobtrusive and continuous monitoring of chimpanzees' vocal communication and drumming without human interference during the recording period.

2.3 Data annotation

A subset of recordings was manually annotated by a trained ethologist using custom software developed by one of the authors. Each chimpanzee-related sound was classified into labeled events representing vocalizations or other acoustic behaviors of interest. The annotation schema included: *pant-hoot introduction*, *pant-hoot build-up*, *pant-hoot climax*, *waa-barks*, *drumming sequences*, and *other* [33]. These categories were selected because they represent the most frequent high-arousal vocalizations reported in previous studies [1]. The label *other* captured less common or previously unreported chimpanzee vocalizations. For each event, annotators specified onset and offset times in seconds relative to the start of the audio file, allowing alignment of behavioral labels with the temporal dynamics of the acoustic signal.

2.4 Data augmentation and class balancing

Given the extensive manual work required for annotation and the small team size, we implemented an extensive data augmentation strategy to increase dataset diversity and improve model robustness. We applied five distinct augmentation techniques to simulate natural acoustic variation: (1) time-stretching at rates of 0.9× and 1.1× introduced natural tempo variations that could occur in chimpanzee vocalizations, (2) pitch shifting by ±2 semitones simulated individual vocal differences and varying recording conditions, (3) background noise addition at 15 dB enhanced robustness to environmental interference, (4) temporal shifting by ±0.5 seconds with zero-padding and exclusion of events falling entirely outside the audio boundaries after shifting, and (5) gain adjustments of ±6 dB simulated recording level variations. Additionally, we created composite augmentations by chaining multiple transformations, further increasing acoustic diversity. Crucially, this augmentation framework enabled class balancing by generating more synthetic examples for underrepresented labels, ensuring roughly equal representation across vocalization types in the training set and mitigating potential biases from imbalanced data distribution.

2.5 Deep learning framework (HuBERT-based modeling)

We will implement an audio sequence labeling framework based on HuBERT (Hidden-Unit BERT), a self-supervised speech representation model originally trained on large-scale human speech corpora. HuBERT was chosen because its architecture is well suited to capturing the complex acoustic features of chimpanzee vocalizations, in much the same way it models human speech. In particular, it leverages the ability to learn robust representations from unlabeled audio while modeling long-range temporal dependencies, which provides an advantage over alternative architectures for this task. These properties make it an ideal backbone for adapting to non-human primate communication. Furthermore, HuBERT is distributed under the MIT license, which permits unrestricted use, modification, and distribution, making it both legally and practically feasible for research and applied work in sanctuary contexts.

The methodology will follow four key stages plus a post-processing step:

Feature extraction. Raw audio waveforms will be resampled to 16 kHz and processed using the *Hugging Face AutoFeatureExtractor* associated with HuBERT. This step will produce fixed-length input sequences of acoustic embeddings at the frame level.

Frame-wise labeling. For each audio file, the annotated onsets and offsets will be converted into frame-aligned binary vectors, indicating the presence or absence of each target label at each temporal frame. This transformation will enable the task to be formulated as a multi-label frame classification problem.

Model architecture. A custom neural architecture will be constructed by combining the HuBERT backbone (pretrained *facebook/hubert-base-ls960*) with an additional linear classification layer that mapped the hidden representation at each frame to the set of ethological labels. The model will be trained with a binary cross-entropy loss applied frame-wise, to account for overlapping events such as simultaneous drumming and vocalizations. The model will produce frame-level predictions indicating the presence or absence of each label.

Training and evaluation. The dataset will be split into training (70%), validation (10%), and test (20%) subsets, ensuring that the distribution of individuals, contexts, and dates is balanced across partitions. Training will be conducted for up to five epochs with gradient accumulation, a learning rate of 3e-5, and evaluation every 100 steps. Performance will be assessed using the micro-averaged *F1-score* on the frame-level predictions. Model training will be carried out on a Google Colab instance using a GPU T4 runtime with Python 3.

Event-level post-processing. To translate frame-level predictions into ethologically meaningful events, a post-processing step will be applied. Consecutive frames exceeding a probability threshold will be grouped into single events, yielding outputs defined by onset, offset, label, and confidence score. This step will ensure that model predictions can be directly interpreted in terms of discrete communicative behaviors. In practice, this transforms a frame-by-frame output—where each frame may encode one or more vocalizations, including drumming—into a structured list showing where events occur in the audio, their duration, and their associated confidence.

2.6 Use of Large Language Models (LMMs)

A LLM (GPT 5) was employed to support the curation of code for combining the HuBERT backbone with an additional linear classification layer. The LLM was used strictly as a coding assistant to write and refine scripts and improve reproducibility. All model design choices, dataset preparation, and methodological decisions were defined and implemented by the authors.

3 Expected results and impact

This project will generate a structured dataset of nightly chimpanzee vocalizations and drumming, including onset and offset times, acoustic categories, and confidence scores. When paired with existing video recordings, these outputs will allow keepers to identify the individuals involved and the direction of interactions (e.g., aggressor versus recipient). This timely information will support proactive husbandry decisions, enhancing welfare management across the full 24-hour cycle.

In addition, the dataset will enable systematic analyses linking nocturnal vocal behavior with daytime interactions, extending previous findings [1]. A logical next step will be to extend the model to incorporate caller identification. By retraining the pipeline with individual-level labels, we could automatically link nocturnal vocal events to specific chimpanzees. This development would further reduce reliance on video inspection and provide keepers with a more complete picture of nightly dynamics. Further automating the process incorporating who vocalized, when, and in what context would allow for even more precise and informed welfare interventions.

4 Code availability

The code is available in a figshare depository (`https://figshare.com/projects/From_Pretrain_to_Primate_Decoding_Chimp_Nocturnes_for_Wellbeing/268244`).

5 Acknowledgements

The authors thank Dietmar Crailsheim for helpful discussions, Guillem Rosselló for his help in the AI side, and Fundació MONA's keepers who took part in data collection.

This work received funding from Fundació MONA and The European Solidarity Corps.

References

[1] P. R. Ayuso, O. Feliu, D. Riba, and D. Crailsheim. Listening to their nights: Sleep disruptions in captive housed chimpanzees affect their daytime behavior. *Animals*, 13(4):696, 2023. doi: 10.3390/ani13040696.

[2] A. Baevski, W.-N. Hsu, Q. Xu, A. Babu, J. Gu, and M. Auli. data2vec: A general framework for self-supervised learning in speech, vision and language. In *Proceedings of the 39th International Conference on Machine Learning*, pages 1298–1312, 2022. URL `https://proceedings.mlr.press/v162/baevski22a.html`.

[3] S. Banks and D. F. Dinges. Behavioral and physiological consequences of sleep restriction. *Journal of Clinical Sleep Medicine*, 3(5):519–528, 2007. doi: 10.5664/jcsm.26918.

[4] S. Brando and H. M. Buchanan-Smith. The 24/7 approach to promoting optimal welfare for captive wild animals. *Behavioural Processes*, 156:83–95, 2018. doi: 10.1016/j.beproc.2017.09.010.

[5] L. Brent. Life-long well being: Applying animal welfare science to nonhuman primates in sanctuaries. *Journal of Applied Animal Welfare Science*, 10(1):55–61, 2007. doi: 10.1080/10888700701277626.

[6] M. Dawkins. Animal welfare and the paradox of animal consciousness. In M. Naguib, H. J. Brockmann, J. C. Mitani, L. W. Simmons, L. Barrett, S. Healy, and P. J. B. Slater, editors, *Advances in the Study of Behavior*, volume 47, pages 5–38. Academic Press, 2015. doi: 10.1016/bs.asb.2014.11.001.

[7] Marian S Dawkins. Using behaviour to assess animal welfare. *Animal welfare*, 13(S1):S3–S7, 2004.

[8] J. A. Funkhouser, H. Boostrom, H. Hellmuth, L. Lacy, S. Bekins, P. Joshi, W. Mayoukou, S. Ndassoba, C. Singono, C. Abedine, C. E. Ayina, E. Claisse, E. Eslinger, P. McElmurray, S. Musgrave, D. Morgan, and C. Sanz. Chimpanzee activity and behavioral diversity extends across 24 hours in both captive and wild settings. *American Journal of Primatology*, 87(1): e23729, 2025. doi: 10.1002/ajp.23729.

[9] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2):169–185, 2019. doi: 10.1111/2041-210X.13101.

[10] C. M. Gomes and C. Boesch. Reciprocity and trades in wild west african chimpanzees. *Behavioral Ecology and Sociobiology*, 65(11):2183–2196, 2011. doi: 10.1007/s00265-011-1227-x.

[11] C. M. Gomes, R. Mundry, and C. Boesch. Long-term reciprocation of grooming in wild west african chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, 276(1657): 699–706, 2008. doi: 10.1098/rspb.2008.1324.

[12] K. Havercamp, N. Morimura, and S. Hirata. Sleep patterns of aging chimpanzees (pan troglodytes). *International Journal of Primatology*, 42(1):89–104, 2021. doi: 10.1007/s10764-020-00190-3.

[13] C. Hozer and F. Pifferi. Relationships between endogenous circadian period, physiological and cognitive parameters and sex in aged gray mouse lemurs (microcebus murinus). *Chronobiology International*, 39(3):363–373, 2022. doi: 10.1080/07420528.2021.2001478.

[14] W.-N. Hsu, B. Bolte, Y.-H. H. Tsai, K. Lakhotia, R. Salakhutdinov, and A. Mohamed. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3451–3460, 2021. doi: 10.1109/TASLP.2021.3122291.

[15] K. Jensen, J. Call, and M. Tomasello. Chimpanzees are vengeful but not spiteful. *Proceedings of the National Academy of Sciences*, 104(32):13046–13050, 2007. doi: 10.1073/pnas.0705555104.

[16] Z. Jiang, A. Soldati, I. Schamberg, A. R. Lameira, and S. Moran. Automatic sound event detection and classification of great ape calls using neural networks. arXiv preprint arXiv:2301.02214, 2024.

[17] N. Jones, S. L. Sherwen, R. Robbins, D. J. McLelland, and A. L. Whittaker. Welfare assessment tools in zoos: From theory to practice. *Veterinary Sciences*, 9(4):170, 2022. doi: 10.3390/vetsci9040170.

[18] L. M. Kurtycz, K. E. Wagner, and S. R. Ross. The choice to access outdoor areas affects the behavior of great apes. *Journal of Applied Animal Welfare Science*, 17(3):185–197, 2014. doi: 10.1080/10888705.2014.896213.

[19] Ingelin Kyrrestad. *Nesting behaviour of the great apes*. PhD thesis, Faculty of Veterinary Medicine, Szent Istvan University, 2014.

[20] G. E. Laule, M. A. Bloomsmith, and S. J. Schapiro. The use of positive reinforcement training techniques to enhance the care, management, and welfare of primates in the laboratory. In *Training Nonhuman Primates Using Positive Reinforcement Techniques*. Psychology Press, 2003.

[21] A. Mielke, A. Preis, L. Samuni, J. F. Gogarten, R. M. Wittig, and C. Crockford. Flexible decision-making in grooming partner choice in sooty mangabeys and chimpanzees. *Royal Society Open Science*, 5(7):172143, 2018. doi: 10.1098/rsos.172143.

[22] J. Müller, O. Mitesser, H. M. Schaefer, S. Seibold, A. Busse, P. Kriegel, D. Rabl, R. Gelis, A. Arteaga, J. Freile, G. A. Leite, T. N. de Melo, J. LeBien, M. Campos-Cerqueira, N. Blüthgen, C. J. Tremlett, D. Böttger, H. Feldhaar, N. Grella, and Z. Buřivalová. Soundscapes and deep learning enable tracking biodiversity recovery in tropical forests. *Nature Communications*, 14 (1):6191, 2023. doi: 10.1038/s41467-023-41693-w.

[23] S. J. Neal Webb, J. Hau, and S. J. Schapiro. Relationships between captive chimpanzee (pan troglodytes) welfare and voluntary participation in behavioural studies. *Applied Animal Behaviour Science*, 214:102–109, 2019. doi: 10.1016/j.applanim.2019.03.002.

[24] S. H. Oppler, S. D. Palmer, S. N. Phu, and M. L. Graham. The role of behavioral management in enhancing clinical care and efficiency, minimizing social disruption, and promoting welfare in captive primates. *Veterinary Sciences*, 11(9):401, 2024. doi: 10.3390/vetsci11090401.

[25] M. J. Prescott and H. M. Buchanan-Smith. *Training Nonhuman Primates Using Positive Reinforcement Techniques*. Psychology Press, 2003.

[26] S. Ross. Captive chimpanzees. In *The ethics of captivity*, pages 57–76. Oxford University Press, New York, 2014.

[27] S. R. Ross and J. G. Leinwand. A review of research in primate sanctuaries. *Biology Letters*, 16 (4):20200033, 2020. doi: 10.1098/rsbl.2020.0033.

[28] Z. J. Ruff, D. B. Lesmeister, C. L. Appel, and C. M. Sullivan. Workflow and convolutional neural network for automated identification of animal sounds. *Ecological Indicators*, 124: 107419, 2021. doi: 10.1016/j.ecolind.2021.107419.

[29] S. J. Schapiro and S. P. Lambeth. Control, choice, and assessments of the value of behavioral management to nonhuman primates in captivity. *Journal of Applied Animal Welfare Science*, 10 (1):39–47, 2007. doi: 10.1080/10888700701277345.

[30] K. E. Slocombe and K. Zuberbühler. Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, 104(43):17228–17233, 2007. doi: 10.1073/pnas.0706741104.

[31] K. E. Slocombe, N. J. Lahiff, C. Wilke, and S. W. Townsend. Chimpanzee vocal communication: What we know from the wild. *Current Opinion in Behavioral Sciences*, 46:101171, 2022. doi: 10.1016/j.cobeha.2022.101171.

[32] L. S. M. Sugai, T. S. F. Silva, Jr Ribeiro, J. W., and D. Llusia. Terrestrial passive acoustic monitoring: Review and perspectives. *BioScience*, 69(1):15–25, 2019. doi: 10.1093/biosci/biy147.

[33] R. Tenaza and P. Marler. Signaling behavior of apes with special reference to vocalization. In *How Animals Communicate*, pages 965–1033. Indiana University Press, Bloomington, 1997.