

POSTPRINT: This paper has been peer reviewed at *PLOS Computational Biology*. The final published version will differ.

Title: A toolkit for the dynamic study of air sacs in siamang and other elastic circular structures

Short title: A toolkit for the dynamic study of elastic circular kinematics

Authors

Lara S. Burchardt (l.s.burchardt@gmx.de)^{1,2*}, Yana van de Sande¹, Mounia Kehy³, Marco Gamba³, Andrea Ravignani^{4,5,6}, Wim Pouw (wim.pouw@donders.ru.nl)^{1*}

* Corresponding authors

Affiliations

1. Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, Netherlands
2. Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany
3. Department of Life Sciences and Systems Biology, University of Torino, via Accademia Albertina 13, 10123, Turin, Italy
4. Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
5. Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music, Aarhus, Denmark
6. Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

Abstract

Biological structures are defined by rigid elements, such as bones, and elastic elements, like muscles and membranes. Computer vision advances have enabled automatic tracking of moving animal skeletal poses. Such developments provide insights into complex time-varying dynamics of biological motion. Conversely, the elastic soft-tissues of organisms, like the nose of elephant seals, or the buccal sac of frogs, are poorly studied and no computer vision methods have been proposed. This leaves major gaps in different areas of biology. In primatology, most critically, the function of air sacs is widely debated; many open questions on the role of air sacs in the evolution of animal communication, including human speech, remain unanswered. To support the dynamic study of soft-tissue structures, we present a toolkit for the automated tracking of semi-circular elastic structures in biological video data. The toolkit contains unsupervised computer vision tools (using Hough transform) and supervised deep learning (by adapting DeepLabCut) methodology to track inflation of laryngeal air sacs or other biological spherical objects (e.g., gular cavities). Confirming the value of elastic kinematic analysis, we show that air sac inflation correlates with acoustic markers that likely inform about body size. Finally, we present a pre-processed audiovisual-kinematic dataset of 7+ hours of closeup audiovisual recordings of siamang (*Symphalangus syndactylus*) singing. This toolkit (<https://github.com/WimPouw/AirSacTracker>) aims to revitalize the study of non-skeletal morphological structures across multiple species.

Keywords: Elastic Kinematics, Morphometrics, Acoustics, Siamang, Air Sac, Open Dataset

Author summary

Many animals move not only by activating skeletal muscles, but also by pushing air or fluids into cavities like air sacs. There is currently a lack of investigation and technology to track movement of these inflating balloon-like structures. Here we present a toolkit with open access data for tracking movement of inflatable air sacs in small Asian apes called siamang. The techniques overviewed here can be applied to other inflatable structures that are abundant in other animals too.

Introduction

Animal *skeletal* pose tracking from video data has undergone nothing short of a revolution through developments in computer vision and deep learning [1–6]. There are now many pretrained pose detection models for humans, and there are increasingly more pre-trained models for non-human animals (e.g., rhesus macaques, *Macaca mulatta*, [7], chimpanzees, *Pan troglodytes* [8,9]). These developments are key for animal research as they allow non-invasive monitoring based on video alone, which can further be used for automatic classification of behavioral patterns [10–12]. The tracking of *elastic* biological structures has received comparatively less attention, leaving out a major aspect of biological motion. Organisms and other biological morphologies combine elements that resist compression (e.g., bones, skeleton, cytoskeleton) and elements that resist expansion (e.g., muscles, connective tissues, membranes) which include elastic structures such as shown in Figure 1 [13]. Together these elements form coherent interconnected systems that define how animals move and communicate [13,14]. These elastic structures include vocal sacs in frogs, tissue inflation in puffer fish, the gular sacs of birds, or air sacs in primates, such as the very prominent laryngeal air sacs in the siamang (*Symphalangus syndactylus*).

Figure 1. Overview of elastic semi-circular biological structures Examples of different species with elastic endogenous and exogenous biological structures that are trackable as semi circles with the approaches described in this report. From top-left to top right, guineafowl puffer (*Arothron meleagris*), greater sage-grouse (*Centrocercus urophasianus*), green tree frog (*Hyla cinerea*), prairie chicken (*Tympanuchus cupido*), magnificent frigatebird (*Fregata magnificens*), siamang (*Symphalangus syndactylus*), elephant seal (*Mirounga angustirostris*). All photos are public domain.



Elastic structures are involved in a wide variety of behaviors and ecological functions. In many species the expandable, oftentimes semi-circular structures are involved in communication, both acoustically and visually [15,16]. Such expandable structure may serve a function in mate attraction [16–19]. In birds they might also be involved in thermoregulation and stem from respiratory adaptations [20]. Monitor lizards and other reptiles have an inflatable gular cavity that aids in respiration [21] and may serve social functions (e.g., mate attraction). Many of these suggested functions remain empirically untested. There are also no detailed studies of dynamic shape and size variation of these inflatable biological structures. Here we show how focusing on elastic structures provides a first step towards a new layer of morphometry analyses, thus far mostly focused on skeletal, bony structures, which in turn can bridge knowledge gaps.

In the particular case of laryngeal air sacs, there is a lack of tools and data that are needed for empirical research in this domain to thrive. This is not for a lack of interest and there seems to

be a gap of methodological and theoretical knowledge when it comes to air sacs. For example, understanding the function and functioning of air sacs is widely accepted as relevant to understanding the evolution of animal communication, including human speech [15,17,22–24]. In this paper we bridge this gap and provide a toolkit for tracking elastic kinematics via I) a data archive for audiovisual recordings of siamang air sac during their singing, II) a computer vision pipeline for analyzing spherical biological structures and air sacs, and III) proof of concept analyses linking air sac inflation with vocal acoustics. Our toolkit and proposed pipeline can enable the study of elastic biological structures in many more systems. They can provide indirect mechanistic insights via purely behavioral, non-invasive approaches.

Small asian apes, the siamang, and laryngeal air sacs

Small asian apes or gibbons (*Hylobatidae*) are phylogenetically closely related to humans and other great apes [25]. Like humans, they are highly vocal [26–30]. Gibbons produce daily duetting songs to maintain and advertise pair bonds within an area, regulating their socially monogamous and territorial lifestyles [31]. These vocalizations are produced at high intensities and show highly distinctive species specific traits [26]. The siamang, also a gibbon species, can sing louder than 120 dB [32], thereby exceeding the vocalization ranges of most humans in terms of amplitude. Important for our current purposes, the siamang has one of the largest air sacs in extant primates relative to body size [24,25,33], and it is also the only species with a very visible semi-circular air sac. The air sac consists of a soft-walled cavity connected to the vocal tract just above the vocal folds and below the false vocal folds [33]. The soft-wall cavity forms a membrane under tension and thereby can resonate and radiate sounds. The air sacs are inflated during, and possibly in preparation for certain types of calls during singing, suggesting their possible supportive role in vocal production [33,34]. Some suggest [22,35,36] that air sacs inflate due to lung exhalation while closing the nasal and nostril passages, others hypothesize that the false vocal folds may close completely to redirect exhaled air into the air sac even when the upper air passages are open ([22], p. 635 for a discussion).

Primate laryngeal air sacs likely evolved because of some adaptive function. This hypothesis is tentatively supported by a cost-benefit analysis: why would air sacs evolve without a function when they also entail the risk of infections? Infections have been reported in detail in a mountain gorilla [37], and observed more generally in primates [38,39]. There are however no systematic studies that quantitatively report how often air sacs get infected. On the benefit side, there are many hypotheses about the function of (siamang) air sacs [17,23,33,34,40,41]; they

may: be relatively functionless [42], help manage oxygenation [41], or help stabilize the thorax during brachiation and singing [40]. Moreover, it has been suggested that an inflated air sac increase the expiratory flow relative to expiring from a single air reservoir (the lungs) when the upper air passages open, leading to a “glottal-shock” that increases the amplitude of the vocalization after the air sac is loaded with air [33].

Limited empirical work and real-world data are available to evaluate these hypotheses. Based on biomechanical and acoustic modeling we would expect a higher air sac volume to statistically predict less energy at higher formant frequencies relative to the dominant frequency and increased overall amplitude [34]. The latter relates to dynamic anti-resonance properties of the air sac (a cancellation of energy at resonant frequencies of the air sac), as supported by physical models and simulations that assess different static sizes of air sacs [22,34]. Thus the laryngeal air sac may likely serve as an amplifying organ by changing the resonant properties of the vocal source, whereby energy at lower frequencies is increased relative to the harmonics, thereby aiding sound travel and “dishonestly” advertising a larger body size [18,23]. In cluttered environments where formant structure can be lost even at short distances, body size might rather be signaled through longer sound duration, which could be enabled by the extra air volume that the air sacs provide (next to the lungs).

In sum, laryngeal air sac mechanics, articulatory states, and vocalization acoustics, feature many key unknowns. Studying the dynamic variation of air sacs together with articulation and acoustics will provide novel insights of their possible adaptive functions [43]. It promises to better understand the development of singing in the siamang by allowing to track air sac use and growth, and inter-individual variability therein. Further, by accounting for vocal variation attributed to air sac dynamics, we can start to better account for variations in vocal acoustics across species [24,44,45]. Indeed it was suggested that understanding the adaptive functions of air sacs is key to understanding the evolution of vocal-articulatory communication in hominins [46]. It is currently unknown why the laryngeal air sacs seem to have been lost in the genus *Homo* [47,48]. The evolutionary vestiges of air sacs are still present in humans, as is evident in pathological cases of trumpet players that develop a highly similar laryngeal air cavity [49]. Fitch concludes on these open issues that “Understanding why we lost air sacs requires a clear understanding of their function” [46] (p. 266). To gain such understanding we believe I) computer vision methods need to be optimized for elastic tracking, II) audiovisual data of laryngeal air sacs need to become available, and III) multimodal signal processing methods need to then be applied to study the relation of vocalization acoustics, articulatory and air sac kinematics. In this article, we provide a complete toolkit that fulfills all these requirements.

Summary of our toolkit

The current toolkit includes a data archive, computer vision tools, and bioacoustic analysis. We first introduce I) an open dataset of more than seven hours, featuring tracking data, that allows for the detailed study of siamang air sacs. We then introduce II) a set of computer vision and data wrangling tools to track siamang air sacs and other spherical biological structures. Having introduced the toolkit, we report on promising findings that relate siamang air sac inflation with the acoustic properties of singing (III). The current paper provides a complete resource to promote a more in-depth study of the laryngeal air sacs and their functions (Figure 2).

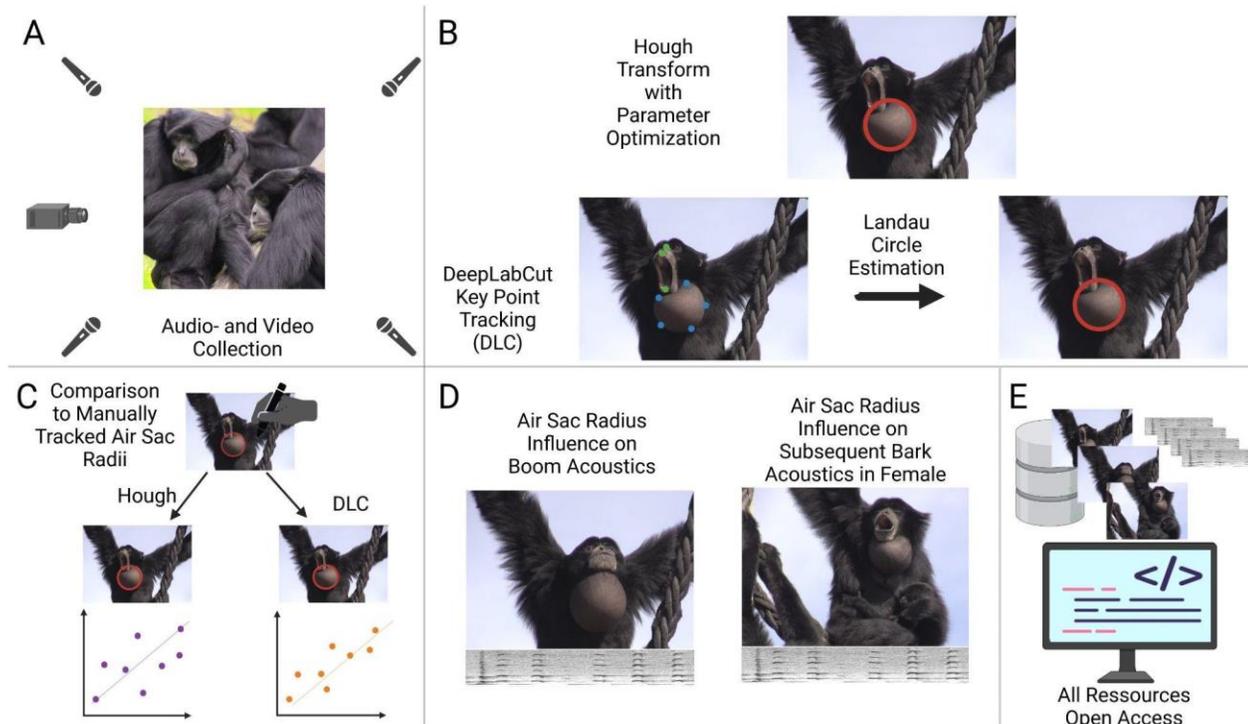


Figure 2: Overview experimental design. A) Audio and video data were collected in the Jaderpark Tier- und Freizeitpark an der Nordsee, Germany. B) Air sacs were automatically tracked with two approaches: the lesser performing Hough Transformation (see sample [here](#)) versus the very good performing DeepLabCut tracking (see here [sample](#)) with Landau Circle Estimation (DLC+: see [here](#) for a sample). C) For a subset of the data, air sacs were tracked manually and compared to the automatically tracked radii. DLC estimated radii had a high correlation of $r > 0.8$ with the manually tracked radii. D) Acoustic parameters of two different kinds of calls were analyzed and related with air sac inflation as a proof of concept. E) All data and code are shared [open access](#). The figure was created with BioRender. All images are photos by Wim Pouw and Mounia Kehy.

Materials and methods

Overview

I) Audiovisual recordings of siamang in captivity were made which forms the open dataset of the toolkit (see Audiovisual Dataset). **II)** We investigate two approaches to track the recorded air sac inflations automatically: 1) circle tracking through mathematical transformations with the Hough Transform, and 2) circle tracking through point tracking using a trained DeepLabCut model (version 2) combined with circle estimation using the Landau algorithm [50], which we refer to as DLC+. **III)** To show the toolkit in action, we provide two different proof-of-concept analyses on the corresponding acoustic and newly obtained kinematic data. The experimental design is summarized in Figure 1. Together, our toolkit and proposed analysis pipelines allow for the extensive study of air sac dynamics, with further investigations of articulatory rhythms (the dataset contains labial kinematics).

I) Audiovisual dataset

Audiovisual recordings of six siamang (Table 1) were collected by us at a single location in Germany, Jaderpark Tier- und Freizeitpark an der Nordsee, for 21 research days over two months in the summer of 2022. We used an opportunistic sampling scheme. The opportunistic sampling started whenever the apes began to sing. The apes usually sang in the mornings, after lunchtime, and/or occasionally around five p.m. For the sampling of singing events, we randomly picked one of two recording strategies for each recording session: 1) record whoever is best visible, or 2) record an individual chosen randomly. The first strategy maximized the amount of usable close-up video recordings, while the second strategy allowed for tracking the song's development as contributed by a single individual (although it did lead to a lot of unusable video data due to occlusions). The closeup audiovisual recordings can be accessed on the Donders Repository (https://data.donders.ru.nl/collections/di/dcc/DSC_2022.00071_151?3). The study was assessed and approved by the ethical committee of the Donders Institute (reference number: DCC.2022.00071), and the study was waived for further ethical approval for animal testing as the study did not change the circumstances of the captive siamang in any way.

Table 1. Information on individuals: Sex, age class, and age for all siamang at Jaderpark are given. Data were only used from five of the six individuals; Tristan, the newborn, was not considered. All images are photos by Wim Pouw and Mounia Kehy.

	Pelangi	Roger	Baju	Fajar	Jamil	Tristan
						
Sex	f	m	m	m	m	m
Age class	Adult	Adult	Subadult	Juvenile	Infantile	newborn
Age	18y10m	29y10m	7y8m	4y11m	3y6m	6w

Audiovisual recording. A tripod-operated camera was used (Canon Legria HF G30) with high-zoom capabilities due to an add-on lens (TL-H58 Telekonverter), sampling at 25fps (50i), and for our second visit, we sampled at 50fps to increase temporal resolution. A Sennheiser ME64-12 was fed as an audio input to the camera (using a DXA-2T audio adapter).

Audio recording. We used two types of microphones. Firstly, a cardioid boom microphone with a windjammer was directed at the center of the site (Sennheiser ME67 HDP2), which was connected to a DR-40 linear PCM recorder (TASCAM) sampling at 48Khz.

Additionally, we recorded multi source audio streams from four KE400 Sennheiser microphones with windjammers sampling at 48 kHz at four locations. We combined all channels into a single mono-channel source by synchronizing audio waveforms using Adobe Premiere Pro CS6. This multi source audio stream is more suitable for estimating acoustic measurements, such as amplitude, as amplitude is influenced by the distance and direction between the microphone and the sound source.

Enriched data. The data repository features information about the weather (humidity, temperature, cloudiness, etc.) during each recording session, next to the start and end times of the recordings. Additionally, pictures are provided with the names of the individuals. Finally, we have tracked all videos in the repository with our DLC+ model using a GPU-supported machine. These tracked videos and resulting time series data are also in the repository.

II) Computer vision tracking tools

Ground truth. To assess the success of automatic tracking, we first established a ground truth, where we asked a student assistant to manually track the radii of air sacs in images of siamang closeups. We created a subset of 1612 frames from nine scenes and three days (three

scenes per day) to account for different lighting and backgrounds. Radii were drawn on individual frames in the OpenSource Software Fiji, and the diameters and coordinates of the circle center were exported. If no air sac was visible, or deemed to be untrackable because of occlusion, low inflation, or any other reason, we drew a very small circle at the edge of the frame, clearly smaller than any tracked circle. This systematic strategy allowed to include those ‘untrackable’ frames in testing the automatic tracking. Before comparison, we transformed diameter to radius.

Unsupervised computer vision: Hough transform. We use a feature extraction technique called the Hough Transform to detect imperfect circles (or semi-circles) in individual video frames [50]. The complete process is implemented in Python and primarily supported by OpenCV (see script [here](#)). In the procedure, frames undergo a series of preprocessing steps that are commonly advised to increase the detection success of circles, which essentially serve to maximize the visibility of any edges in the scene. Our Hough detection and the preprocessing process is depicted in Figure 3, where circles detected with the Hough Transform are indicated for all processing steps to illustrate their necessity. First, frames were converted to grayscale and optimized for brightness and contrast. A median blur is applied as a next step, followed by a ‘Canny edge detection’, where the found edges are then dilated. The Canny edge detector inputs a lower and an upper threshold in which edges are detected. The issue arose that the threshold is easily set too high or too low for a particular scene. Generic thresholds for very different images increase this problem; ideally, the thresholds are tuned for each video separately. Therefore, we set the lower and upper limits by a normalization procedure, where the threshold depends on the mean intensity of the frame in question. After dilating, the image is blurred again with a median blur.

After image preprocessing the Hough Transform can be applied to the image. In our case the Hough transform is set to detect semi-circles with a minimal radius of five pixels and a maximum of 270 pixels can be found (these numbers depend on your data and aim). We found that 270 pixels as a maximum for detection was optimal for our dataset by trial and error and visual inspection. But for future users this needs to be adjusted on a case-by-case basis, depending on the expected circle size in the videos. To reduce any noise-related jitter, after per-frame circle detections in the full video, we applied a Kolmogorov-Zurbenko smoothing algorithm to the detected position x,y data (centroids) and radii of circles.

The preprocessing parameters were optimized with another Python script, changing brightness and contrast parameters, median blur, edge detection parameters, and dilation strength (Table 2). This optimization is necessary for each new dataset. Depending on the variability in the dataset, some parameter combinations might perform well for some and poorly for other images;

instead, other parameter combinations might deliver a satisfactory performance for all images. Another step for potential optimization of circle detection with the Hough Transform is the correct choice of the image section. Hand-selecting region of interest can increase tracking success. Thus, we advise cropping the video so that the circular structure is maximized in the frame. Figure 4 provides an example of the output of the Hough tracker.

Figure 3: Image processing to increase Hough Transform circle detection success. Four pre-processing steps in four different frames with various backgrounds are shown to illustrate the necessity of the preprocessing steps for the Hough Transform approach to work. Only after the last pre-processing step are air sacs correctly found with the Hough Transform. All images are photos by Wim Pouw and Mounia Kehy.



Table 2: Parameters iterated for parameter optimization in Hough Transform. After initial testing, six parameters were chosen to iterate over different setting combinations to find the best performance. The best performance was found by correlating the found radius results to the manually tracked radii and choosing the parameter combination, which showed the highest correlation over all videos.

Parameter	function	Range	steps	Best for our videos
Alpha	Brightness	0.5-3	0.5	2
Beta	Contrast	20-40	5	30
Blur	Median Blur	25,27,29,35		27
Dilation	Dilation	3-7	1	5
Canny 1	Parameter used to determine lower limit in Canny edge detection	4-12	2	5
Canny 2	Parameter used to determine upper limit in Canny edge detection	8,10,13,15,17		14

Figure 4: Example of our Hough Transform approach. Examples: Hough Transform tool applied to a video of a siamang (see [here](#)) and a green tree frog (*Hyla cinerea*; see sample [here](#)), with satisfactory performance for some purposes such as position tracking or large-scale changes in inflation (but after investigation of tracking performances we favor DLC+). The left image contains a photo by Wim Pouw and Mounia Kehy. The picture on the right is from the [youtube video](#) and analyzed by us for testing:



Supervised computer vision: DeepLabCut + Landau (DLC+)

DLC model info. We trained a Resnet-101 convolutional neural network at 500K iterations¹ using Deeplabcut 2.0 [3,4]. We first trained a shallower network that is set as the default in DLC, Resnet-50, which yielded poor air sac tracking and was soon abandoned. The key points that our model tracks are: ['UpperLip', 'LowerLip', 'Nose', 'EyeBridge', 'Start_outline_outer_left', 'Start_outline_outer_right', 'LowestPoint_outline', 'MidLowleft_outline', 'MidLowright_outline']. The model weights and metadata (e.g., training loss data) can be found on our github.

DLC labeling. A training dataset was created with a labeling approach that is optimized for 2D tracking of semi-circular 3D objects under variable camera angles. We found the following DLC labeling approach provides a robust method for tracking circular objects with DLC+. Firstly, we define the start of the outline on the left of the semi-circle, then the right end of the outline (figure 5). These are fixed points and can be isolated based solely on visual features in the frame as they do not require the context of other points to be determined. The other non-fixed points lie on the edge of the air sac but are further defined in relation to the fixed points. The lowest point is defined as a point on the edge of the air sac, on the middle of a line perpendicular to the line between the start left and the end right. Then, the middle points along the circle are defined for the left and right sides. Note, that this labeling method allows for a well-formed description of each point, regardless of the perspective of the siamang (see figure 5). Only three points are necessary to estimate a circle in the subsequent Landau estimation, described below. By aiming to

¹Given the deeper model architecture, we opted for more training iterations than the 250K suggested as default by the DLC team. However, to assess our optimized performance by excluding possible overfitting artifacts we also provide a model at a checkpoint trained at 280K iterations on our [github](#).

track five points while only three are necessary, the tracking becomes more error-robust, as the likelihood of two points dropping out of the frame is much lower than a single one dropping out.

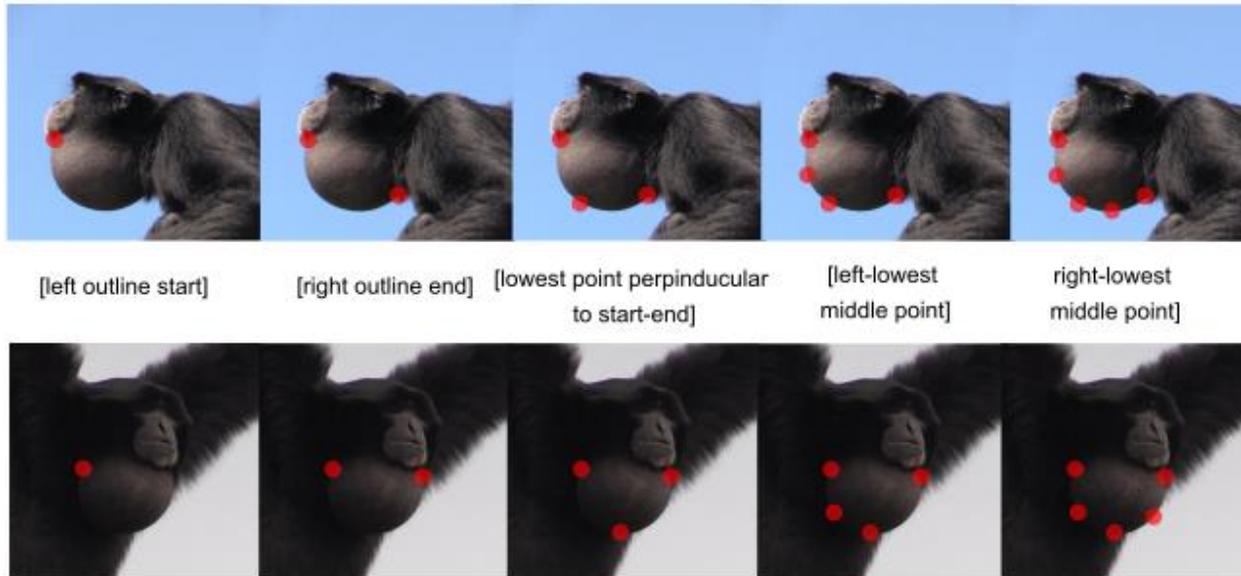


Figure 5. DLC labeling approach for DLC+. The five points to be tracked on the air sac outline are depicted. Note that two outline ends, which we call fixed points, are clearly defined. The other three points are defined relative to these fixed points. Firstly, it is the middle point between the start and end point (vertically defined by being on the edge of the air sac). The two other points are again the middle of the start or end and the adjacent middle point. In this way, we ensure that DLC can always determine the points, even when some are not determinable, without relating them to the two fixed start and end points. As such, some points are only relationally but nevertheless systematically definable.

DLC training set and internal DLC validation. The total number of hand-annotated images was 390, with only one animal prominently in the frame². This set of images was used as a training (90%) and test (10%) set for training of the DLC model. After training (500k iterations), we yielded a pixel error for 0.6 likelihood tracked points of 9.5 pixels for the test image set (for $1920 \times 1080 = 2.073.600$ pixel images).

DLC+: Radius estimate using Landau. We used Landau's geometric circle fit method, which is an ordinary least square estimation method to estimate circles from at least three points. In contrast to other least square estimation methods, it is circle specific, has been described as very robust and a benchmark circle-detection method [51], and it is implemented in proprietary software such as MATLAB [52]. The Landau method minimizes the mean square distance from a fitting curve to the data points, to find the best circle in a fixed-point iterative scheme. The geometric Euclidean distance from data points to the fitting curve was used [50]. We estimated

² Note that other models could be optimized for multi-animal tracking, which could further improve the tracking in our dataset provided.

the radius of a circle and the (x,y) coordinates of its centroid. In our analysis points tracked by DLC were only used if they have a likelihood value > 0.6 . For frames, where less than 3 points fulfill that criterion, we did not get a radius estimate and the frame was excluded from further analysis. DLC tracking was transformed into radii and (x, y)-centroid data with a custom made Python script ([link](#)). We also implemented the same routine in an R-script ([link](#)).

Kinematic-acoustic analysis approach

Data snippets for the kinematic-acoustic analysis were sampled randomly from the full dataset. To be included in the analysis and get reliable radii estimations to be matched to acoustic parameters, snippets for both analyses had to fulfill several criteria: 1) camera angle and zoom were not allowed to change within the snippet; 2) a snippet could contain only a single siamang vocalizing; 3) background noise had to be minimal. Individual vocal boom snippets data were sampled from all four vocalizing siamangs. As great call boom-bark sequences are only or primarily produced by females, samples for this analysis were taken from the only female, Pelangi. Booms produced by Pelangi in the analyzed great call sequences were not analyzed as individual booms but analyzed as part of a sequence.

All acoustic analyses were run in R 4.2.3 [53]. The window length for the acoustic analysis was chosen to match the duration of one video frame. This enables direct comparison of air sac inflation status to the resulting circle-specificities. Video frame data consistently had 25fps. Videos that were originally recorded with 50fps were downsampled to 25fps. This was done after the videos were tracked with DLC. Downsampling, therefore, was performed on the level of analyzed frames. Every second frame with accompanying radius information was kept for the acoustic analysis.

The function “analyze” from the soundgen R-package [54,55] was used to perform an extensive analysis, reporting on 47 standard acoustic parameters, such as amplitude, fundamental frequency (called f0 in the analysis), entropy, or spectral Centroids (see Documentation in soundgen package and codes for a comprehensive list of parameters).

Acoustic analyses were then conducted on two types of calls: a) boom calls and b) bark calls. Acoustic features of the boom calls were compared to the corresponding radius in the same frame, matched by filename and frame number. Pearson’s correlation coefficient (r) was calculated for all acoustic parameters and the radius of the air sac with the base R “cor” function. To analyze the influence of the air sac inflation of a boom call on the subsequent bark in a great call sequence (boom – bark — boom — bark — [...]) we extracted the last trackable radius during the boom. We compared it to the average, minimum and maximum of the acoustic parameters

across the subsequent bark. We used the same window length for the individual boom call analysis. Acoustics were not analyzed for the boom parts of the great call sequences.

Results

First we report the results of the external validation of the air sac tracking methods against ground-truth hand-estimated air sac radii. Secondly we report the results of the proof-of-concept analyses where we relate acoustic parameters to air sac inflation (referred to as part III of our contribution in this paper).

External validation automatic circle tracking

To evaluate the tracking success of both approaches (Hough Transform & DLC+), a subset of the data was manually labeled (serving as ground truth) to compare to automatic tracking. We then correlated the automatically tracked radii. The results of both methods, raw and smoothed, are reported in Table 2. The smoothed results are visualized in Figure 6. Automatic trackings of DLC+ are of sufficient quality, showing a correlation coefficient $r > .80$. DLC+ showed the highest correlation to manually tracked data, therefore this method is used for the subsequent analysis in part III.

Table 2: Reliability comparison with manually tracked ground-truth data between tracking algorithms. For comparison, only automatically tracked radii below 270 px were used, as this is the maximum radius tracked with the Hough Transform and approximately the maximum radius found manually (max: 266 px). This also applies to smoothed radii. While smoothing increases tracking success for Hough Transform, tracking works equally well, if not better, without smoothing using DLC+ tracking. Notice that these correlations feature the whole subset of data. The Hough Transform works very well for particular examples; DLC+ works better on average and, across the board, has a high performance.

Circle estimation method	Best Mean Correlation Coefficient (r)	sd**	min**	max**
Hough Transform, (best settings overall)	0.19	0.22	-0.18	0.57
Hough Transform, overall smoothed*	0.23	0.33	-0.31	0.8
DLC+, LAN	0.86	—	—	—
DLC+, LAN smoothed*	0.85	—	—	—

* Kolmogorov-Zurbenko, iterations = 2, window length = 3 | ** between videos

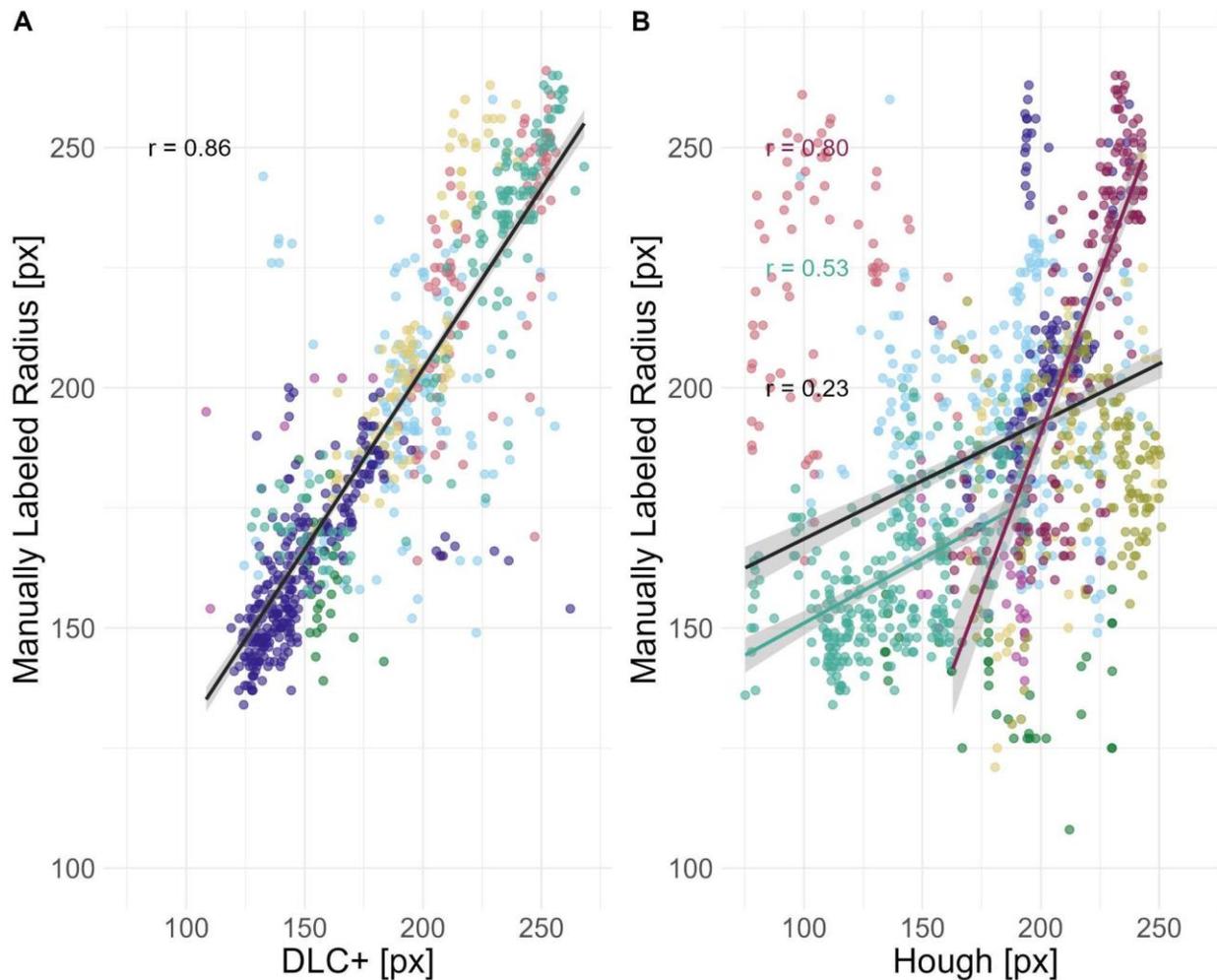


Figure 6: Automatically tracked radii in comparison to manually labeled radii for both approaches: DLC+ and Hough Transform.

A) Comparison of DLC+ trackings and manual tracking of air sac radii, both in pixel units, for a set of > 1000 frames from nine different video scenes. Perfect fit would mean $r = 1.0$ and direct match in pixels. We find that the DLC+ radii match very well with the manual trackings, $r = 0.86$, as calculated on all frames over nine videos. B) Comparison of Hough Transform tracking and manual tracking for the same set of > 1000 frames. The correlation coefficient (r) for the nine test videos was 0.23. Parameters can be optimized for each video specifically, and when set adequately for individual videos, we see correlations close to the one for DLC+ trackings. As a trendline, in turquoise, we see the second-best correlation for one video with $r = 0.53$; in red, the best correlation with $r = 0.8$.

Note that for the tracked radii with the Hough method we set the maximum radii to 270 px as this was the maximum tracked manually and the maximum trackable radius in the Hough Transform algorithm. Radii below 100 px were not regarded for any of the datasets. Colors denote the different video scenes.

Analyses relating acoustic parameters to air sac inflation (III)

To study the influence of the air sac inflation on the acoustic parameters of co-occurring vocalizations, 47 acoustic parameters were considered for two different call types. We analyzed acoustic parameters for the “boom” calls, produced during air sac inflation and for “bark” calls

produced in so-called “great call” sequences by the female siamang. Barks are produced directly after a boom in the sequences that we selected for analyses (see an example [here](#); note, barks would be extracted from this longer sequence). We matched acoustics to video data by frame, to compare air sac inflation and acoustic parameters in a meaningful way.

Air sac inflation influences acoustic parameters of boom call as predicted by the model in adults. We correlated air sac inflation as determined with DLC+ for two adult individuals (one female) with a set of acoustic parameters provided by the `soundgen` R-package [54,55]. A total of 25 call sequences were analyzed with 176 adequate frames (M[SD] frames per call sequence = 7[5]). This final set was the number of frames for which we could determine an air sac radius and concurrent acoustic parameters of the boom call.

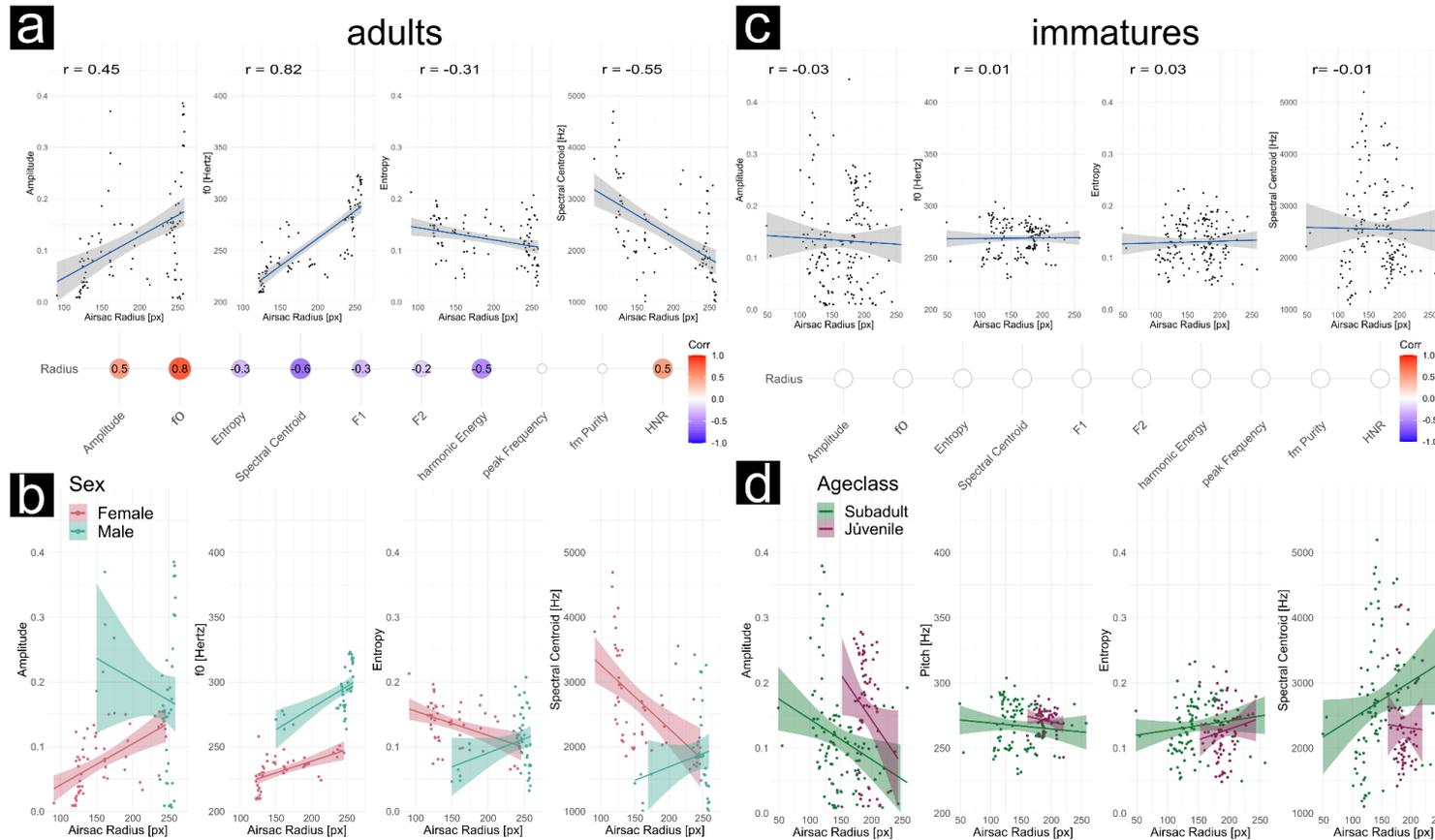
Several of the tested acoustic parameters showed a significant correlation with the radius of the air sac (Figure 7E, correlation plot adults). Figure 7, top panel, left shows the four parameters amplitude, Fundamental Frequency (f_0), entropy and spectral centroid, in more detail. Amplitude indicates the sound pressure level of the signal while f_0 as we use it here stands for the fundamental frequency representing the lowest frequency component of a complex sound signal. The spectral centroid is a weighted average or the “center of mass” of all frequencies in the signal, containing information about timbre and f_0 . Finally, (Shannon) entropy is a measure of the disorder in a signal. In bioacoustics, entropy is often used to characterize the complexity or diversity of sound signals. As predicted by theoretical models, we observed an increase in amplitude with an increase in radius ($r = 0.45$, $p < 0.0001^{***}$) [34]. The fundamental frequency also strongly correlated with an increasing air sac inflation status ($r = 0.82$, $p < 0.0001^{***}$). Entropy and spectral Centroid of the boom call were negatively correlated with radius inflation, though (entropy: $r = -0.31$, $p = 0.002^{**}$; spectral Centroid: $r = -0.55$, $p < 0.0001^{***}$).

Individual patterns emerge when study subjects are analyzed individually. The overall correlation appears to be driven by the adult female (Figure 7, bottom panel, left). Given our tiny sample size, this individual analysis only suggests that data for more individuals from both sexes is needed to achieve reliable conclusions and test for potential sex differences.

In addition to the adults, two male individuals of younger age classes were analyzed: Bajú, a subadult (7 years, eight months), and Fajar, a juvenile (4 years, 11 months). Their results contrast the results we found in adults and what theory predicts. No clear pattern emerged when we pool both age classes (Figure 7, top panel, right). When dividing further by age class (juvenile and subadult), we can see patterns opposite to what we see in adults and expected to see from adults and modeling work (Figure 7, bottom panel, right).

This difference between the patterns between age classes suggests an influence of ontogeny on the relationship between acoustic parameters and air sac inflation that should be considered in further analyses to see if the effect is stable with a larger sample size and what this could mean.

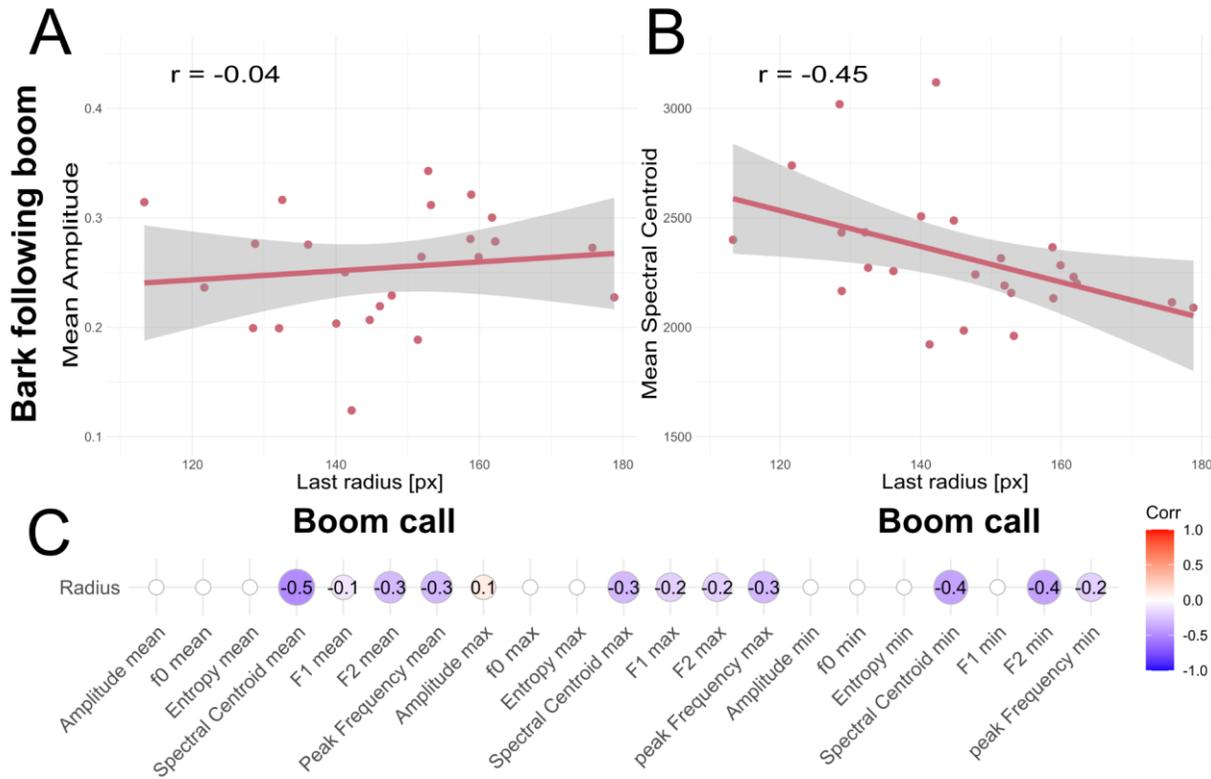
1 **Figure 7: Significant correlations between air sac inflation (as air sac radius in pixel) and acoustic parameters.**
 2 The figure is divided into results of adult siamangs (left; panel a and b) and immature siamangs (right; panel c and d). The top panel
 3 shows the pooled data. Air sac radius in pixel on the x-axis is shown against four different acoustic parameters (from left to right):
 4 sound amplitude, f0 or fundamental frequency in Hertz, Wiener entropy and spectral Centroid in Hertz. For adults, we find clear
 5 correlations: The more the air sac is inflated, the higher the sound amplitude produced. This is in line with model predictions (see [34]).
 6 The fundamental frequency (f0) is also positively correlated with the air sac inflation, and seems to be the only stable relationship
 7 reflected in both the male and female individual. Mean Entropy is negatively correlated with air sac inflation, meaning, the more
 8 inflated the air sac is, the more tonal the produced sound. The Spectral Centroid is negatively correlated as well, indicating the higher
 9 the inflation, the more energy in the lower frequencies. We do not see those relationships in immature not fully grown individuals. In
 10 the middle panel significant correlation coefficients for all tested acoustic parameters are shown. Notice, that none of the acoustic
 11 parameters showed a significant relation in immatures (indicated by white circles). The bottom panel divides the pooled data. For
 12 adults we divide by sex into male and female, for the immatures we divide by ageclass into subadult and juvenile.



14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Influence of air sac inflation on subsequent calls. In a second analysis, we studied how air sac inflation of a boom call influences the acoustics of the subsequent bark, using early phases of the great call sequences of the only female Pelangi (which consisted of these alternating boom-bark units, see [here](#)). We analyzed 16 boom-bark pairs from a total of 8 great call sequences. We found that the air sac inflation statistically predicts the average spectral centroid of the subsequent bark (Figure 8B): The more inflated the air sac was before the booming call (last trackable inflation radius of the boom), the lower the average spectral centroid of the bark ($r = -0.45$, $p = 0.001$). This matches the relation between these features within one boom call as described above, suggesting that air sac inflation can be a preparatory action for the subsequent call the bark) produced with an open airway. Furthermore, this lowering the spectral centroid aligns with the theory that air sacs attenuate higher formant energy to increase the acoustic appearance of body size and sound radiation [34]. None of the other relations found for boom calls (i.e., entropy and fundamental frequency) could be observed in the data (Figure 8C). No clear correlation to the sound amplitude of the bark call could be found ($r = -0.04$, $p = 0.14$, Figure 8A), which speaks against the glottal shock theory suggesting that there is an increased air pressure release due to an extra subglottal pressure producing air reservoir provided by the air sac [33].

32 **Figure 8: Relation between air sac inflation for a boom call, with acoustics of bark**
 33 **immediately following the boom.** The top panel shows the relationship between the last
 34 observed radius of the air sac during a boom call, plotted against the subsequent bark's acoustic
 35 parameters. A) shows the results for the acoustic parameter mean sound amplitude, which does
 36 not relate to the inflation of the air sac in this case. B) shows the relation of the mean spectral
 37 centroid (given in Hertz) of the bark relative to the inflation of the preceding boom. In C),
 38 correlation estimates are shown for the suite of acoustic parameters included in these exploratory
 39 analyses, but only statistically significant correlation coefficients are shown. Mean values and
 40 minimum and maximum values of the acoustic parameters in the bark were tested against the
 41 radius of the last boom.



42
 43 **Discussion**

44 The current report breaks terrain in tracking elastic kinematics in animals; it uses
 45 unsupervised and supervised computer vision methods to detect semi-circular soft-tissue
 46 structures. In our main use case, we focused on the laryngeal air sacs in the siamang
 47 (*Symphalangus syndactylus*). We thereby also provide a data archive of 7+ hours of siamang
 48 singing with closeup video data with multi-source audio, ideal for the study of articulatory,
 49 acoustic, and air sac states. Data sharing in primatology is rare, and we are confident that this
 50 openly accessible dataset enriched with DLC+ trackings will be of significant utility to
 51 forthcoming researchers due to its user-friendly nature. DLC+ was the best method as assessed
 52 with ground truth data. However, our toolkit provides both unsupervised (Hough Transform) and
 53 supervised computer vision tools (DLC+) to track elastic circular biological structures. Finally,
 54 summarized below, we provided a kinematic-acoustic analysis of air sac inflation and its relation

55 to the acoustics of calls. Together, this report breaks new terrain that we hope will invite further
56 research incorporating a complete morphometric analysis of animal behavior involving bones,
57 elastic and expandable soft tissue.

58 Our kinematic-acoustic analyses of air sac dynamics in siamang singing confirmed that
59 the lower frequencies are modulated while the higher frequencies are attenuated, captured by a
60 lower spectral centroid. We find this for adult boom calls and barks following booms in the
61 female boom-bark sequences that occur at the early phases of the so-called “great call”. We also
62 observed an increased amplitude of the boom calls for higher inflated air sacs. This finding
63 requires further scrutiny as all gibbons sing loudly, but only siamang have large laryngeal air sacs.
64 Further study is required to understand the many possible factors (e.g., see [27,56]) that shape the
65 gibbon’s capacity to sing loudly. These findings obtained from dynamic real-world data align
66 with modeling research where air sacs were initialized at different air sac volumes [34]. We
67 further obtain that adults and younger siamang individuals show different relationships between
68 air sac inflation and the assessed acoustic parameters. This developmental effect may have
69 indicated a role of ontogeny. This is not surprising as a physiological explanation: the cavities of
70 smaller air sacs will have different resonant properties due to morphological differences.
71 Furthermore, we observed other exciting patterns in acoustics and air sac inflation, where, for
72 example, boom f_0 increases with air sac inflation. With its small sample size, the current
73 exploratory analysis has only limited explanatory strength but already shows promising paths to
74 further the study of air sacs, especially in the siamang.

75 The current toolkit also provides a boon for collecting new audiovisual data on elastic
76 circular biological structures in the wild, as there are now tested approaches to study said
77 structures. The pipelines provided in our toolkit are fully reproducible and await large-scale
78 applications in a diverse set of elastic circular biological structures on this planet (Figure 1). The
79 combinations of point estimates using DLC and Landau circle approximation, as implemented in
80 our DLC+, provide the best performance; we recommend them for further research endeavors.
81 We encourage researchers to optimize further tools like the Hough Transform implementation,
82 which we found suboptimal for our current aim. With some adjustments in the pre-processing of
83 images, future studies can improve the Hough Transform approach to the point where it becomes
84 more efficient for large-scale use. In such a case, the need for training a neural network, such as in
85 DLC+, becomes obsolete, sparing labor. Our work incites one more development: training a
86 general-purpose model based on a range of elastic circular biological structures. This trained
87 model could then serve as an out-of-the-box approach for tracking these circular objects, similar
88 to models already developed (see DLC SuperAnimal, [57]). A drawback of the current approach

89 is that we have focused on single-animal tracking, and it invites further development of
90 applications for multi-animal tracking; this is relatively straightforward as a multi-animal DLC
91 approach already exists [58].

92 Researchers can build on the current DLC+ materials in a number of ways. Most
93 pertinently, the current trained model can be applied to a much bigger dataset containing many
94 more siamang individuals; data are relatively easy to collect, as captive siamang are not known to
95 sing less than siamang in the wild. We hope our contribution of an open dataset will eventually
96 lead to the gradual construction of a bigger repository of audiovisual recordings of siamang
97 singing. Furthermore, the current model trained on siamang air sacs can be retrained to be adapted
98 for circular kinematics in other animals. On the technical side, the current pipeline can be scaled
99 up to 3D tracking using triangulation methods whereby multiple calibrated cameras generate 2D
100 videos that can be tracked with the current DLC+ model, which then form the input for a
101 triangulation pipeline [59] which obtains 3D positions which can then be used to approximate air
102 sac volume with a higher precision.

103 Our approach, advancing morphometric studies from bony structures to elastic structures,
104 can potentially address diverse questions across various species, encompassing birds, primates,
105 pinnipeds, and frogs. Consider for example that the role of vocal sacs in anuran vocalizations is
106 well documented, but additional functions have been proposed related to respiration, buoyancy
107 control, chemical signaling, or even thermoregulation [19]. In sum, the ability to study vocal sac
108 dynamics in detail will help shed light on this multi-functionality [19]. Furthermore, tracking of
109 semi-circular biological objects could also be used to assess the health and development of the
110 species through pure observation, serving as an indicator of ecosystem health.

111 To conclude, the current open-source dataset, open-source computer vision tools,
112 benchmarking, and proof-of-concept analysis provide a way to study diverse biological structures
113 (see Figure 1). The current highlighted case of siamang air sacs can help us understand the
114 adaptive functions of these extreme biological modifications. We invite the community to study
115 the dynamic modulation of elastic circular structures in animals.

116

117

118 **References**

- 119 1. Cao Z, Simon T, Wei S-E, Sheikh Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields.
 120 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE; 2017. pp.
 121 1302–1310. doi:10.1109/CVPR.2017.143
- 122 2. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, et al. MediaPipe: A Framework for Building
 123 Perception Pipelines. arXiv; 2019. doi:10.48550/arXiv.1906.08172
- 124 3. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose
 125 estimation of user-defined body parts with deep learning. *Nat Neurosci.* 2018;21: 1281–1289.
 126 doi:10.1038/s41593-018-0209-y
- 127 4. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose
 128 estimation across species and behaviors. *Nat Protoc.* 2019;14: 2152–2176. doi:10.1038/s41596-019-0176-0
- 129 5. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS-H, Murthy M, et al. Fast animal pose estimation
 130 using deep neural networks. *Nat Methods.* 2019;16: 117–125. doi:10.1038/s41592-018-0234-5
- 131 6. Bala PC, Eisenreich BR, Yoo SBM, Hayden BY, Park HS, Zimmermann J. Automated markerless pose
 132 estimation in freely moving macaques with OpenMonkeyStudio. *Nat Commun.* 2020;11: 4560.
 133 doi:10.1038/s41467-020-18441-5
- 134 7. Labuguen R, Matsumoto J, Negrete SB, Nishimaru H, Nishijo H, Takada M, et al. MacaquePose: A Novel “In
 135 the Wild” Macaque Monkey Pose Dataset for Markerless Motion Capture. *Frontiers in Behavioral
 136 Neuroscience.* 2021;14. Available: <https://www.frontiersin.org/articles/10.3389/fnbeh.2020.581154>
- 137 8. Desai N, Bala P, Richardson R, Raper J, Zimmermann J, Hayden B. OpenApePose: a database of annotated
 138 ape photographs for pose estimation. arXiv; 2022. doi:10.48550/arXiv.2212.00741
- 139 9. Wiltshire C, Lewis-Cheetham J, Komedová V, Matsuzawa T, Graham KE, Hobaiter C. DeepWild: Application
 140 of the pose estimation tool DeepLabCut for behaviour tracking in wild chimpanzees and bonobos. *Journal of
 141 Animal Ecology.* 2023;92: 1560–1574. doi:10.1111/1365-2656.13932
- 142 10. Bain M, Nagrani A, Schofield D, Berdugo S, Bessa J, Owen J, et al. Automated audiovisual behavior
 143 recognition in wild primates. *Science Advances.* 2021;7: eabi4883. doi:10.1126/sciadv.abi4883
- 144 11. Sakib F, Burghardt T. Visual Recognition of Great Ape Behaviours in the Wild. arXiv; 2020.
 145 doi:10.48550/arXiv.2011.10759
- 146 12. Sande Y, Pouw W, Southern L. Automated Recognition of Grooming Behavior in Wild Chimpanzees.
 147 Proceedings of the annual meeting of the Cognitive Science Society 2024. Rotterdam; 2024.
- 148 13. Turvey MT, Fonseca ST. The medium of haptic perception: A tensegrity hypothesis. *Journal of Motor
 149 Behavior.* 2014;46: 143–187. doi:10.1080/00222895.2013.798252
- 150 14. Ingber DW. Tensegrity and mechanotransduction. *Journal of Bodywork and Movement Therapies.* 2008;12:
 151 198–200. doi:10.1016/j.jbmt.2008.04.038
- 152 15. Perlman M, Salmi R. Gorillas may use their laryngeal air sacs for whinny-type vocalizations and male display.
 153 *Journal of Language Evolution.* 2017;2: 126–140. doi:10.1093/jole/lzx012
- 154 16. Krakauer AH, Tyrrell M, Lehmann K, Losin N, Goller F, Patricelli GL. Vocal and anatomical evidence for
 155 two-voiced sound production in the greater sage-grouse *Centrocercus urophasianus*. *Journal of Experimental
 156 Biology.* 2009;212: 3719–3727. doi:10.1242/jeb.033076
- 157 17. Dunn JC. Sexual selection and the loss of laryngeal air sacs during the evolution of speech. *Anthropological
 158 Science.* 2018;126: 29–34. doi:10.1537/ase.180309
- 159 18. Maciej P, Fischer J, Hammerschmidt K. Transmission Characteristics of Primate Vocalizations: Implications
 160 for Acoustic Analyses. *PLOS ONE.* 2011;6: e23015. doi:10.1371/journal.pone.0023015
- 161 19. Starnberger I, Preininger D, Hödl W. The anuran vocal sac: a tool for multimodal signalling. *Anim Behav.*
 162 2014;97: 281–288. doi:10.1016/j.anbehav.2014.07.027
- 163 20. Brussee BE, Coates PS, O’Neil ST, Casazza ML, Espinosa SP, Boone JD, et al. Invasion of annual grasses
 164 following wildfire corresponds to maladaptive habitat selection by a sagebrush ecosystem indicator species.
 165 *Global Ecology and Conservation.* 2022;37: e02147. doi:10.1016/j.gecco.2022.e02147
- 166 21. Owerkowicz T, Farmer CG, Hicks JW, Brainerd EL. Contribution of gular pumping to lung ventilation in
 167 monitor lizards. *Science.* 1999; 1661–1663.
- 168 22. Riede T, Tokuda IT, Munger JB, Thomson SL. Mammalian laryngeal air sacs add variability to the vocal tract
 169 impedance: physical and computational modeling. *J Acoust Soc Am.* 2008;124: 634–647.
 170 doi:10.1121/1.2924125
- 171 23. Fitch T, Hauser MD. Acoustic communication. In: Simmons A, Fay RR, Popper AN, editors. *Unpacking
 172 “honesty”: Vertebrate vocal production and the evolution of acoustic signals.* New York: Springer; 2002.
- 173 24. Nishimura T. Primate Vocal Anatomy and Physiology: Similarities and Differences Between Humans and
 174 Nonhuman Primates. In: Masataka N, editor. *The Origins of Language Revisited: Differentiation from Music
 175 and the Emergence of Neurodiversity and Autism.* Singapore: Springer; 2020. pp. 25–53. doi:10.1007/978-981-

- 176 15-4250-3_2
- 177 25. Reichard U, Hirohisha H, Barelli C. Evolution of Gibbons and Siamang: Phylogeny, Morphology, and
- 178 Cognition. 2016. Available: <https://link.springer.com/book/10.1007/978-1-4939-5614-2>
- 179 26. D'Agostino J, Spehar S, Abdullah A, Clink DJ. Evidence for Vocal Flexibility in Wild Siamang
- 180 (*Symphalangus syndactylus*) Ululating Scream Phrases. *Int J Primatol*. 2023 [cited 20 Sep 2023].
- 181 doi:10.1007/s10764-023-00384-5
- 182 27. Koda H. Gibbon Songs: Understanding the Evolution and Development of This Unique Form of Vocal
- 183 Communication. In: Reichard UH, Hirai H, Barelli C, editors. *Evolution of Gibbons and Siamang: Phylogeny,*
- 184 *Morphology, and Cognition*. New York, NY: Springer; 2016. pp. 349–359. doi:10.1007/978-1-4939-5614-
- 185 2_15
- 186 28. Clarke E, Reichard UH, Zuberbühler K. The Syntax and Meaning of Wild Gibbon Songs. *PLOS ONE*. 2006;1:
- 187 e73. doi:10.1371/journal.pone.0000073
- 188 29. Geissmann T. *Gibbon songs and human music from an evolutionary perspective. The origins of music.*
- 189 Cambridge, MA, US: The MIT Press; 2000. pp. 103–123.
- 190 30. Marshall JT, Marshall ER. Gibbons and their territorial songs. *Science*. 1976;193: 235–237.
- 191 doi:10.1126/science.193.4249.235
- 192 31. De Gregorio C, Valente D, Ferrario V, Carugati F, Cristiano W, Raimondi T, et al. Who you live with and what
- 193 you duet for: a review of the function of primate duets in relation to their social organization. *J Comp Physiol*
- 194 *A*. 2024 [cited 12 Mar 2024]. doi:10.1007/s00359-023-01689-9
- 195 32. McAngus Todd NP, Merker B. Siamang gibbons exceed the saccular threshold: Intensity of the song of
- 196 *Hylobates syndactylus*. *The Journal of the Acoustical Society of America*. 2004;115: 3077–3080.
- 197 doi:10.1121/1.1736273
- 198 33. Mott F. A Study by Serial Sections of the Structure of the Larynx of *Hylobates syndactylus* (Siamang Gibbon).
- 199 *Proceedings of the Zoological Society of London*. 1924;94: 1161–1170. doi:10.1111/j.1096-
- 200 3642.1924.tb03336.x
- 201 34. de Boer B. Acoustic analysis of primate air sacs and their effect on vocalization. *J Acoust Soc Am*. 2009;126:
- 202 3329–3343. doi:10.1121/1.3257544
- 203 35. Osman Hill WC, Booth AH. Voice and larynx in african and asiatic colobidae. *Journal of the Bombay Natural*
- 204 *History Society*. 54: 309–321.
- 205 36. Lieberman D. *The Evolution of the Human Head*. Harvard University Press; 2011.
- 206 37. Hastings BE. The veterinary management of a laryngeal air sac infection in a free-ranging mountain gorilla.
- 207 *Journal of Medical Primatology*. 1991;20: 361–364. doi:10.1111/j.1600-0684.1991.tb00551.x
- 208 38. Lowenstine LJ, Osborn KG. Respiratory System Diseases of Nonhuman Primates. *Nonhuman Primates in*
- 209 *Biomedical Research*. 2012; 413–481. doi:10.1016/B978-0-12-381366-4.00009-2
- 210 39. Kumar S, Fox B, Owston M, Hubbard GB, Dick EJ. Pathology of spontaneous air sacculitis in 37 baboons and
- 211 seven chimpanzees and a brief review of the literature. *J Med Primatol*. 2012;41: 266–277. doi:10.1111/j.1600-
- 212 0684.2012.00547.x
- 213 40. Hayama S. The origin of the completely closed glottis. Why does not the monkey fall from a tree? *Primate*
- 214 *Research*. 1996;12: 179–206. doi:10.2354/psj.12.179
- 215 41. Hewitt G, MacLarnon A, Jones KE. The functions of laryngeal air sacs in primates: A new hypothesis. *FPR*.
- 216 2002;73: 70–94. doi:10.1159/000064786
- 217 42. Harrison DFN. *The anatomy and physiology of the mamalian larynx*. Cambridge: Cambridge University Press;
- 218 1995.
- 219 43. Partan SR, Marler P. Communication Goes Multimodal. *Science*. 1999;283: 1272–1273.
- 220 doi:10.1126/science.283.5406.1272
- 221 44. Lameira AR, Wich SA. Orangutan Long Call Degradation and Individuality Over Distance: A Playback
- 222 Approach. *Int J Primatol*. 2008;29: 615–625. doi:10.1007/s10764-008-9253-x
- 223 45. Owren MJ, Seyfarth RM, Cheney DL. The acoustic features of vowel-like grunt calls in chacma baboons
- 224 (*Papio cyncephalus ursinus*): implications for production processes and functions. *J Acoust Soc Am*. 1997;101:
- 225 2951–2963. doi:10.1121/1.418523
- 226 46. Fitch WT. *The Biology and Evolution of Speech: A Comparative Analysis*. *Annual Review of Linguistics*.
- 227 2018;4: 255–279. doi:10.1146/annurev-linguistics-011817-045748
- 228 47. Alemseged Z, Spoor F, Kimbel WH, Bobe R, Geraads D, Reed D, et al. A juvenile early hominin skeleton
- 229 from Dikika, Ethiopia. *Nature*. 2006;443: 296–301. doi:10.1038/nature05047
- 230 48. Martinez J, Hossain R, Romero J, Little JJ. A Simple Yet Effective Baseline for 3d Human Pose Estimation.
- 231 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE; 2017. pp. 2659–2668.
- 232 doi:10.1109/ICCV.2017.288
- 233 49. Giovannello J, Grieco RV, Bartone NF. Laryngocele. *American Journal of Roentgenology*. 1970;108: 825–
- 234 829. doi:10.2214/ajr.108.4.825
- 235 50. Landau UM. Estimation of a circular arc center and its radius. *Computer Vision, Graphics, and Image*
- 236 *Processing*. 1987;38: 317–326. doi:10.1016/0734-189X(87)90116-2

- 237 51. Chernov N, Lesort C. Least squares fitting of circles and lines. arXiv; 2003. doi:10.48550/arXiv.cs/0301001
 238 52. Sumith YD. Fast Circle fitting using Landau method. MATLAB; 2024. Available: Sumith YD (2024). Fast
 239 Circle fitting using Landau method ([https://www.mathworks.com/matlabcentral/fileexchange/44219-fast-](https://www.mathworks.com/matlabcentral/fileexchange/44219-fast-circle-fitting-using-landau-method)
 240 circle-fitting-using-landau-method), MATLAB Central File Exchange. Abgerufen 22. Mai 2024.
 241 53. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for
 242 Statistical Computing; 2020.
 243 54. Anikin A. Soundgen: An open-source tool for synthesizing nonverbal vocalizations. Behav Res. 2019;51: 778–
 244 792. doi:10.3758/s13428-018-1095-7
 245 55. Anikin A. soundgen: Sound Synthesis and Acoustic Analysis. 2023. Available: [https://cran.r-](https://cran.r-project.org/web/packages/soundgen/index.html)
 246 project.org/web/packages/soundgen/index.html
 247 56. Koda H, Nishimura T, Tokuda IT, Oyakawa C, Nihonmatsu T, Masataka N. Soprano singing in gibbons.
 248 American Journal of Physical Anthropology. 2012;149: 347–355. doi:10.1002/ajpa.22124
 249 57. Ye S, Filippova A, Lauer J, Vidal M, Schneider S, Qiu T, et al. SuperAnimal models pretrained for plug-and-
 250 play analysis of animal behavior. arXiv; 2023. doi:10.48550/arXiv.2203.07436
 251 58. Lauer J, Zhou M, Ye S, Menegas W, Schneider S, Nath T, et al. Multi-animal pose estimation, identification
 252 and tracking with DeepLabCut. Nat Methods. 2022;19: 496–504. doi:10.1038/s41592-022-01443-0
 253 59. Karashchuk P, Rupp KL, Dickinson ES, Walling-Bell S, Sanders E, Azim E, et al. Anipose: A toolkit for
 254 robust markerless 3D pose estimation. Cell Reports. 2021;36: 109730. doi:10.1016/j.celrep.2021.109730
 255

256
 257

Acknowledgments

Non-author contributions: We would like to thank the student assistant Mouaz Morad for his annotation of the siamang air sacs. We would like to thank Daniel Anthes for his comments on the toolkit code.

Funding: The Comparative Bioacoustics Group is supported by Max Planck Independent Research Group Leader funding to AR. The Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF117). AR is funded by the European Union (ERC, TOHR, 101041885). WP is funded by a VENI grant (VI.Veni 0.201G.047: PI Wim Pouw), Language in Interaction, and was further supported by the Donders Postdoctoral Development fund.

CRedit Author contributions:

Conceptualization: WP

Methodology: WP, LSB

Investigation: WP, MK, LSB

Visualization: WP, LSB

Supervision: WP

Resources: WP, AR

Funding Acquisition: WP, AR

Software: LSB, YS, WP

Validation: LSB, WP

Writing—original draft: WP, LSB

Writing—review & editing: WP, LSB, MG, AR

Competing interests: All other authors declare they have no competing interests.

Data and materials availability: The computational tools can be found in our Github repository (<https://github.com/WimPouw/AirSacTracker/>). The code and data for reproducing the kinematic-acoustic analyses can also be found on our github page (<https://github.com/WimPouw/AirSacTracker/tree/main/Project>). The open dataset can be found on the Donders repository (<https://doi.org/10.34973/6apg-q804>).