PREPRINT

Title

- A computer vision toolkit for the dynamic study of air sacs in Siamang with a general application to the study of elastic kinematics in other animals

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

Authors

Lara S. Burchardt (l.s.burchardt@gmx.de)[1,2]*, Yana van de Sande (yana.vandesande@ru.nl)[1], Mounia Kehy[3], Marco Gamba[3], 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. Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Torino, 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 elements like bones and cartilage, and elastic elements like muscles and membranes. Computer vision advances have enabled automatic tracking of moving animal skeletal poses. Such developments put us on the verge of gaining insights in complex dynamics otherwise studied in more static terms (e.g., images). However, the elastic soft-tissues of organisms, like the nose of Elephant seals, or the buccal sac of frogs, have been poorly studied and no computer vision methods have been proposed. This leaves major gaps in different areas in biology. In the area of primatology, most critically, the function of air sacs is widely debated and many questions exist about their role in communication and human language evolution. Moving towards the dynamic study of soft-tissue elastic 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 Python's 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 revitalizes the study of non-skeletal morphological structures in a wide range of animals.

MAIN TEXT

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*)). The importance of these developments for research with non-human animals cannot be overstated, as it means that non-invasive research can be performed, which can further be used for automatic classification of behavioral patterns (*10*). However, the tracking of *elastic* biological structures has received little attention. This is a pity, because 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) (*11*). Together these elements form coherent interconnected systems that define how animals move and communicate (*11*, *12*). These elastic structures include vocal sacs in frogs, tissue inflation like 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).

Elastic structures are involved in a wide variety of behaviors and different ecological functions are suggested. In many species the expandable, oftentimes semi-circular structures are involved in communication, both acoustically and visually (*13*, *14*). Such expandable structure may serve a function in mate attraction (*14–17*). In birds they might also be involved in thermoregulation and serve important respiration adaptations (*18*). Monitor lizards and other reptiles have an inflatable gular cavity that aids in respiration (*19*) and may serve social functions (e.g., mate attraction). Many of these suggested functions have not been empirically tested and there are no detailed studies of dynamic shape and size variation of these inflatable biological structures. In this report we show that focusing on elastic structures can help fill knowledge gaps and make a first step towards a completely new layer to morphometry studies that so far mostly focus on skeletal, bony structures.

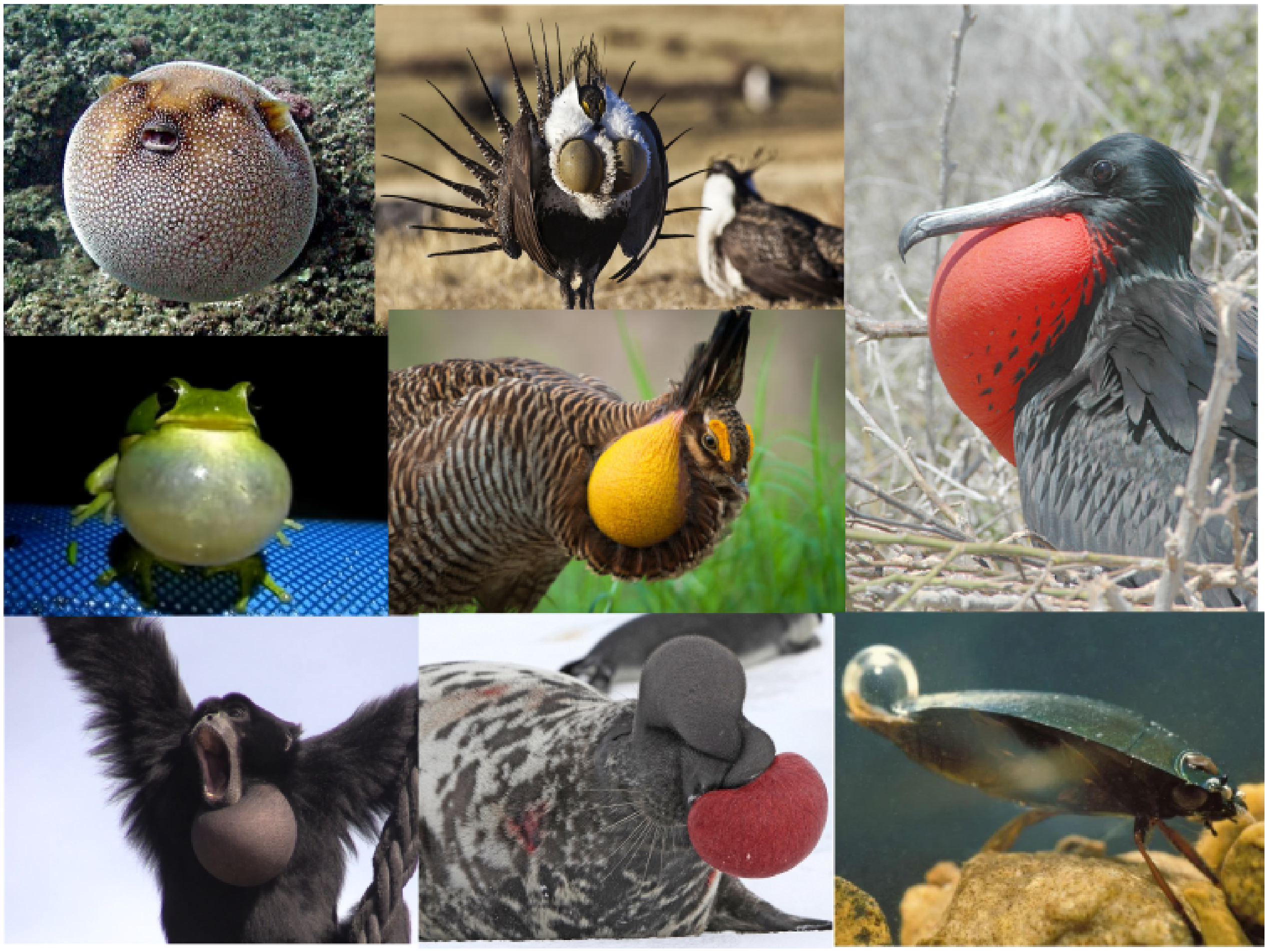


Figure 1. Examples of different species with elastic endogenous and exogenous biological structures that are trackable as demi circles with the approaches described in this report. From top-left to top right, guineafowl puffer (*Arothron meleagris*), greater sage-grouse (*Centrocercus urophasianus*),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

Especially, in the case of laryngeal air sacs, there is a lack of empirical study with an added problem of the lack of data that help develop tools or falsify hypotheses, and this is despite a widely shared conviction of the theoretical importance of air sacs for bioacoustics and evolution of communication and language (*13*, *15*, *20–22*). In this paper we resolve this gap and provide a toolkit for tracking of elastic kinematics with I) a data archive for audiovisual recordings of siamang air sac during their singing, II) a computer vision approach for analyzing spherical biological structures and air sacs, and III) a proof of concept analyses linking air sac inflation with vocal acoustics. Our toolkit can enable the study of elastic biological structures in many more systems and give mechanistic clues to behavioral observation and therefore completely non-invasive approaches.

Small Asian Apes, The Siamang, and Laryngeal Air Sacs

Small Asian Apes or gibbons (*Hylobatidae*) are genetically closely related to humans and other great apes (*23*). Like humans, they are highly vocal. Gibbons produce daily duetting songs to maintain and advertise pair bonds within an area, regulating their socially monogamous and territorial lifestyles. This singing is a loud affair and shows more diversity than is typically appreciated (*24*). The siamang, a remarkable gibbon species, have been observed to sing louder than 120 Decibels (*25*), thereby exceeding the vocalization ranges of most humans in terms of loudness, which is astonishing given human's much larger body size. Important for our current purposes, the siamang have one of the largest and most visible laryngeal air sacs in extant primates relative to body size (*22*, *23*, *26*). This is a soft-walled cavity connected to the vocal tract just above the vocal folds and below the false vocal folds (*26*). The soft-wall cavity forms a membrane under tension and thereby can resonate and radiate sounds. The air sacs are inflated during, and possibly preparatory, to producing certain types of calls during singing, suggesting their supportive role in vocal production (*21*).

Laryngeal air sacs often get infected and it is not uncommon for an animal to die from that (*27*). Air sacs evolved even though the risk and cost of having them is high (*28*). Why did they evolve? There are many hypotheses about the function of (siamang) air sacs, but little empirical work to test them (*15*, *21*, *26*, *28*–*31*). It is especially from biomechanical and acoustic *modeling* research that we expect the air sac volume to be associated with a decrease of energy at the higher formant frequencies relative to the dominant frequency and an increase in amplitude. 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 (*20*, *31*). 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 (*16*, *21*). In cluttered environments where formant structure can be lost over 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 (nex to the lungs).

Models and simulations need to be empirically verified, of course. Further, many alternative or complementary adaptive benefits for air sacs will need further verification, e.g., their role in oxygenation management (*28*), their potential contribution to generating a glottal shock (*26*), or their role in thorax stabilization during brachiation and singing (*30*). Unfortunately, not much empirical work has followed mechanical modeling studies (*31*) that dynamically verify the relation between air sac inflation, vocal acoustics, and relation with articulatory states. The state of our knowledge is best characterized by Riede (*20*), who confesses it is not even known whether the air sac can be inflated with the mouth open, which would suggest that false vocal cords have a mechanism to close off so that exhaled air may enter the air sac.

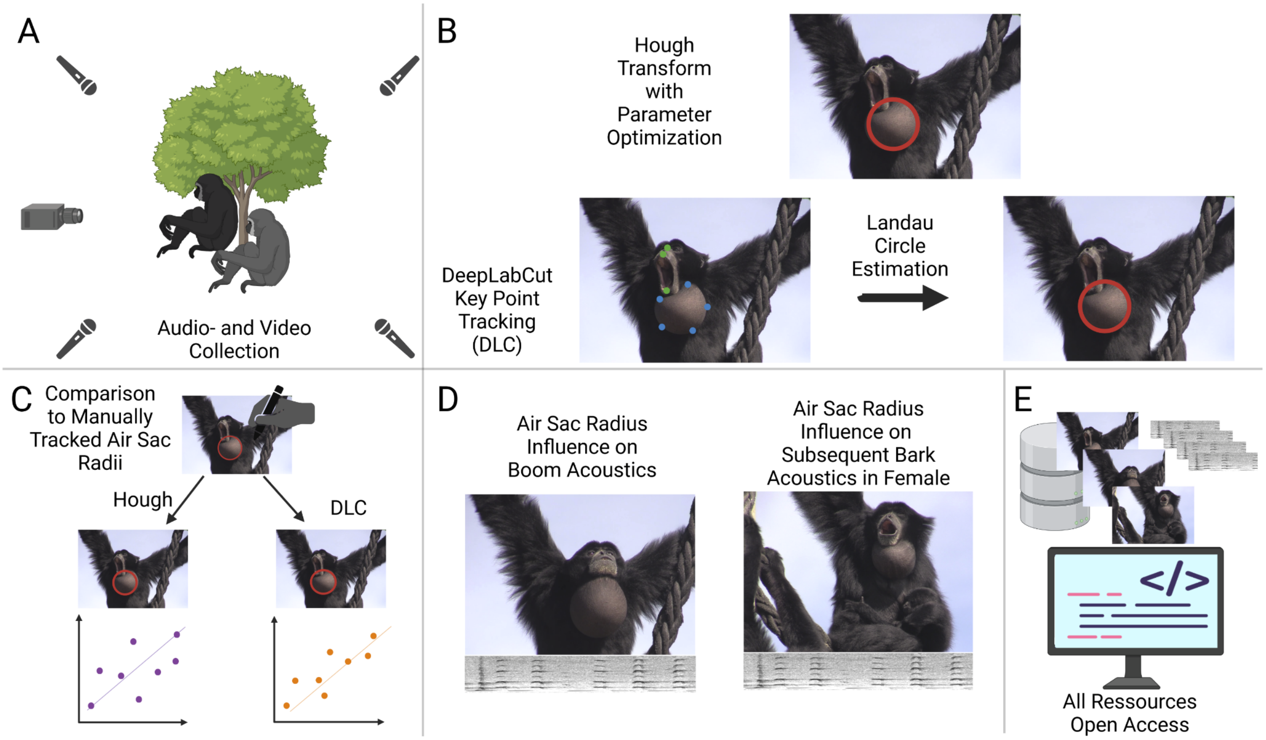
In sum, there are a lot of primary unknowns about laryngeal air sac mechanics, articulatory states, and vocalization acoustics. Studying the dynamic variation of air sacs together with articulation and acoustics will provide novel insights of their possible adaptive functions that relate to acoustic modulation, but possibly also visual signaling (*32*). 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 (*31*, *33*, *34*). Indeed it was suggested that understanding the adaptive functions of air sacs is key to understanding the evolution of vocal-articulatory communication in hominins (*35*). It is currently unknown why the laryngeal air sacs seem to have been lost in *Neanderthalis*, *Heidelbergensis*, and humans, but not in *Australopethicus* (*36*, *37*). 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 (*38*). Fitch concludes on these open issues that

“Understanding why we lost air sacs requires a clear understanding of their function” (*35*) (p. 266).

Currently we lack this functional understanding of air sacs. To change this, computer vision methods need to be optimized for elastic tracking. Furthermore, data resources need to be available that captures high-quality close-ups of siamang singing and concurrent air sac inflations. Multimodal signal processing methods can 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 to make progress on solving the mystery of the evolution of laryngeal air sacs in apes. The large contribution outside of this domain is providing a methodological approach to enable the widespread study of elastic kinematics in animals.

Summary Current toolkit

The current toolkit includes a data archive, computer vision tools, and bioacoustic analysis. We first introduce I) an open dataset of 7+ hours that allows for the detailed study of siamang air sacs (with tracking data). Then, we 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). Below follows a summary of the toolkit (see methods and results for extended information).



A
Audio- and Video Collection
B
Hough Transform with Parameter Optimization
DeepLabCut Key Point Tracking (DLC)
Landau Circle Estimation
C
Comparison to Manually Tracked Air Sac Radii
Hough
DLC
D
Air Sac Radius Influence on Boom Acoustics
Air Sac Radius Influence on Subsequent Bark Acoustics in Female
E
All Ressources Open Access

Figure 2: Overview Experimental Design. A) Audio and video data was collected in the Jaderpark. B) Air sacs were automatically tracked with two approaches: Hough Transformation (see sample here) and DeepLabCut tracking (see here sample) with Landau Circle Estimation (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 analysed and set into relation with air sac inflation. E) All Data and Codes are shared open access. The figure was created with BioRender.

Dataset (I)

In the summer of 2022 we collected close-up video recordings of a family of captive Siamang residents at Jaderpark Tier- und Freizeitpark an der Nordsee, Germany. Over a combined period of about a month, we opportunistically recorded singing events, usually occurring each morning of the day. We also recorded audio using multiple sources around the facility. This results in a dataset containing over 7+ hours of siamang singing, including analyzable data of 5 individuals (adult male and female, two subadults females, and one infantile male), consisting of 600GB of audiovisual materials in total, all openly published on the Donders repository with the following associated DOI (10.34973/6apg-q804). We have applied motion tracking to all the video data. The resulting time series data (and tracking videos) are also stored on the repository. Together this open dataset allows for easy study of air sac dynamics, but also articulatory rhythms as the kinematic dataset also contains tracking of labial kinematics.

Computer vision tools to estimate air sac inflation and other spherical objects (II)

In our investigation, we found two broad approaches that each had benefits and drawbacks. Firstly, we settled on a method that in principle could work immediately on any data that contains spherical structures: the Hough Transform. However, this method was not always stable across different scene parameters (contrast, blur, background etc), and we, therefore, also investigated a supervised approach, utilizing Deeplabcut (*3*) with additional computations for estimating spherical radii from point-based tracking.

Unsupervised Computer Vision: Hough Transform. The Hough Transform is a feature extraction technique dating back to the early sixties (*39*). It is used to find imperfect instances of shapes in an image. First, it was only implemented for straight lines. In a voting procedure, the most likely instance of the imperfect line is found in an image. It was later extended also to find other shapes, most prominently circles and ellipses (*40*). In this paper, we use this purely analytical approach to detect imperfect circles, namely the air sacs of the siamangs that can be approximated as circles in 2D space (Figure 3). The current Hough transform procedure has the advantage of being unsupervised so it can be applied without the need for pre-labeled training data. Further, our implementation works easily on local desktops or laptops with commonly available CPUs. The disadvantage is that the parameter settings are very sensitive to differences in lighting and color arrangements. This makes the Hough transform implementation sometimes tricky to stabilize for tracking large badges of videos (see performance quantification below). This is why we started to explore other supervised computer vision options. A jupyter notebook for tracking custom-picked videos with the Hough transform is provided on our https://github.com/WimPouw/AirSacTracker/tree/main.

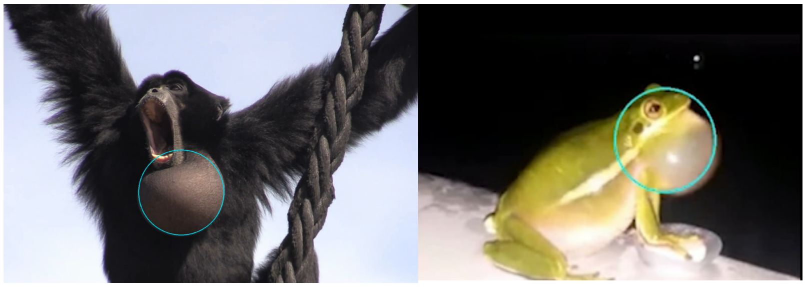


Figure 3: Example Hough transform approach. Examples: current Hough Transform tool applied to video of a siamang (see here) and a green tree frog (*Hyla cinerea*; see sample here) .

Supervised Computer Vision: DeepLabCut (DLC+). We used DeepLabCut (version 2) to train a ResNet-101 pre-trained network to detect 9 keypoints, 5 of which would be used for air sac inflation estimation, and four key points on the face (lower lip, upper lip, nosetip, eyebridge). 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*1080 = 2.073.600 pixel images). The trained ResNet-101 model and a jupyter notebook for tracking custom-picked videos with DLC, with radius estimation (DLC+) are provided on our GitHub.

Key points of the air sac with a likelihood above 0.6 are used to estimate circles. We use the ‘Landau geometric circle fit’ method by (*41*), an ordinary least square estimation method that allows for estimating circles from at least three points. For an example see of DLC+ see here (colored dots represent DLC trained position estimates, and a circle represents the landau circle fit on those points).

Automatic Circle Tracking Compared to manual 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 correlated the automatically tracked radii. The results of both methods raw and smoothed are depicted in Table 1. The smoothed results are depicted in Figure 4.

Table 1: 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 Houch transform, tracking works equally well, if not better, without smoothing using DLC+ tracking.

Circle estimation method	Best Mean 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, windowlength = 3 | ** between videos

Automatic trackings of DLC+ were of sufficient quality r > .80 and showed the highest correlation to manually tracked data (this method will also be used for the subsequent analysis in part III).

It is to be noted that these are the correlations for the whole subset of data. The Hough Transform works very well for particular examples; DLC works better on average and, across the board, has high performance.

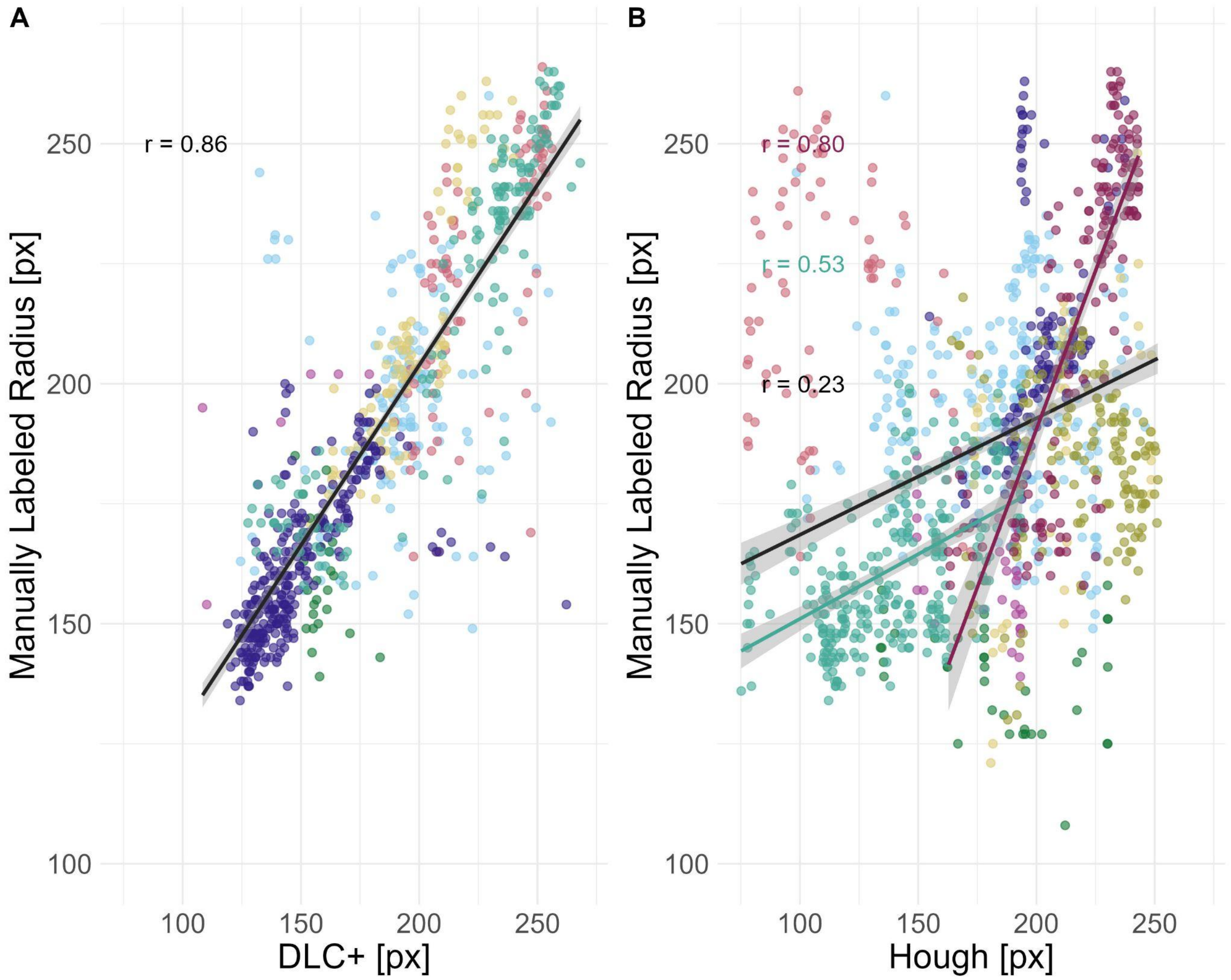


Figure 4: Automatically tracked radii predicting ground truth (manually labeled radius). For the correlations, automatically tracked radii were filtered to be a maximum of 270px as this was the maximum tracked manually and the maximum trackable radius in the Hough transform algorithm. Radii below 100 px were not regarded for any of the datasets. A) Comparison of DLC+ trackings and manual tracking of air sac radii. Radii match very well, *r* =0.86. B) Comparison of Hough transform tracking and manual tracking. The best average correlation coefficient (r) for the nine test videos was 0.23. Parameters need and can be optimized, and when set adequately for individual videos, we see correlations close to the one for DLC+ trackings. As a trendline, in red, we see the second-best correlation for one video with *r* = 0.53; in turquoise, the best correlation with r = 0.8 the correlations, automatically tracked radii were filtered to be a maximum of 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. A) Comparison of DLC+ trackings and manual tracking of air sac radii. Radii match very well, *r* =0.86. B) Comparison of Hough transform tracking and manual tracking. The best average correlation coefficient (r) for the nine test videos was 0.23. Parameters need and can be optimized, and when properly set for individual videos we do see correlations close to the one for DLC+ trackings. As a trendline in red, we see the second-best correlation for one

video with *r* = 0.53 and in turquoise the best correlation with r = 0.8. Colors denote the different video scenes of which frames were hand and machine labeled.

Analyses Relating Acoustic Parameters to Air Sac Inflation (III)

To study the influence of the air sac inflation status on the acoustic parameters of accompanied vocalizations, we analyzed 47 acoustic parameters 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 model in adults

We correlated air sac inflation of adult individuals with a set of acoustic parameters (see the documentation for the *analyze* function in the `soundgen` R-package (*42*) for the complete list of parameters) by correlating the radii as estimated from the points tracked with DLC+ to acoustic parameters analyzed in the corresponding sounds. Two adult individuals, one male and one female, were recorded. A total of 25 call sequences were analyzed with 176 adequate frames (video 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 5E, correlation plot adults). We are showing the four parameters amplitude, pitch (f0), entropy and spectral Centroid, in more detail (Figure 4, top panel, left). As predicted by theoretical models, we observe an increase in amplitude with an increase in radius (r = 0.45, p < 0.0001***) (*26*). The fundamental frequency also strongly correlates with an increasing air sac inflation status (r = 0.82, p < 0.0001***). Entropy and spectral Centroid of the boom call are negatively correlated with radius inflation, though (entropy: r = -0.31, p = 0.002**; spectral Centroid: r = -0.55, p < 0.0001***).

If we separate the results by sex of the calling individual (one male, one female), we see that the pattern differs between female and male and the overall correlations seem to bedriven by the adult female (Figure 5, bottom panel, left).

In addition to the adults, two male individuals of younger age classes were analyzed: Baju, a subadult (7 years, eight months) and Fajar, a juvenile (4 years, 11 months). Their results stand in contrast to the results we found in adults and to what theory predicts. No clear pattern emerges when we pool both age classes (Figure 5, top panel, right). If we divide further by age class (juvenile and subadult), we can even see patterns opposite to what was expected and shown in adults (Figure 5, bottom panel, right). This suggests an influence of ontogeny on the relationship between acoustic parameters and air sac inflation that should be considered in further analyses.

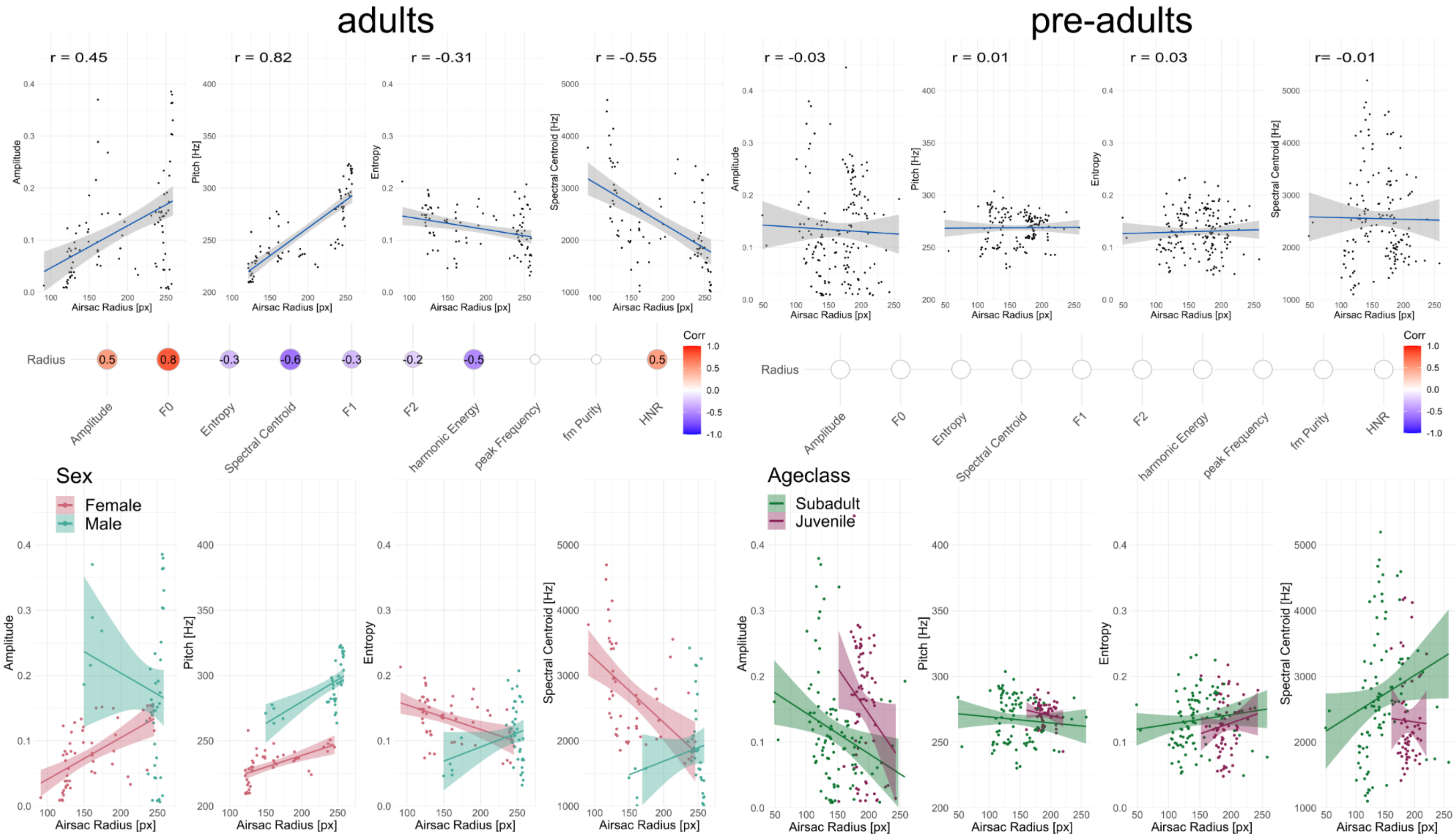


Figure 5: Significant correlations between air sac inflation (as air sac radius in pixel) and acoustic parameters. The figure is divided into results of adult siamangs (left) and pre-adult siamangs (right). The top panel shows the pooled data. Air sac radius in pixel on the x-axis is shown against four different acoustic parameters (from left to right): sound amplitude, pitch or fundamental frequency in Hertz, Wiener entropy and spectral Centroid in Hertz. For adults, we find clear correlations: The more the air sac is inflated, the higher the sound amplitude produced. This is in line with model predictions (see (*31*)). The pitch or fundamental frequency is also positively correlated with the air sac inflation. Mean Entropy is negatively correlated with air sac inflation, meaning, the more inflated the air sac is, the more tonal the produced sound. The Spectral Centroid is negatively correlated as well, indicating the higher the inflation, the more energy in the lower frequencies. We do not see those relationships in pre-adult not fully grown individuals. In the middle panel significant correlation coefficients for all tested acoustic parameters are shown. Notice, that none of the acoustic parameters showed a significant relation in pre-adults (indicated by white circles). The bottom panel divides the pooled data. For adults we divide by sex into male and female, for the pre-adults we divide by ageclass into subadult and juvenile.

Influence of Air Sac Inflation on subsequent calls

In a second analysis, we studied how air sac inflation of a ball 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 boom-bark alternations, see here). We analyzed 16 boom-bark pairs from a total of 8 great call sequences. We found that the air sac inflation influences the average spectral centroid of a subsequent bark (Figure 6B). The more inflated the air sac was for the boom call (last trackable inflation radius of 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, the lowering of the spectral centroid is in line with the theory that air sacs serve to attenuate higher formant energy to increase the acoustic appearance of body size and sound radiation (*31*). None of the other relations found for boom calls (i.e. entropy and fundamental frequency) could be observed in the data (Figure 6C). No clear correlation to the sound amplitude of the bark call could be found (r = -0.04, p = 0.14, Figure 6A), 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 (*26*).

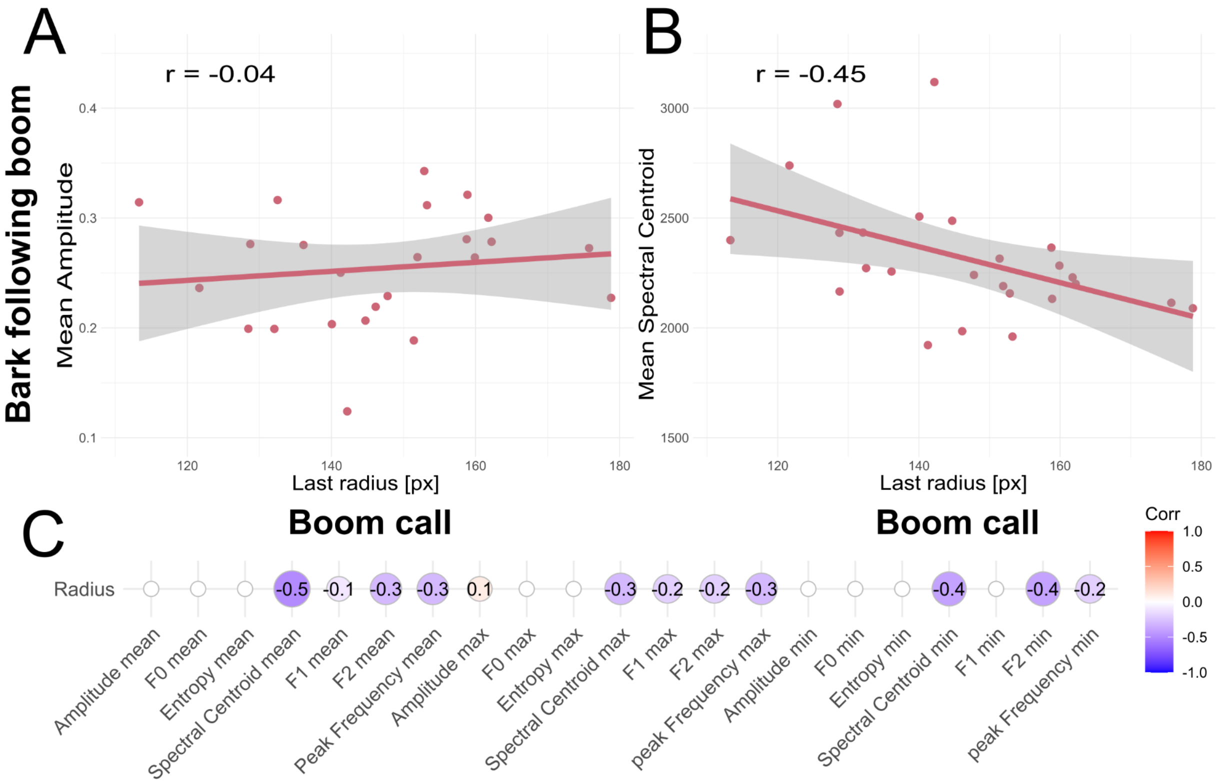


Figure 6: Relation between air sac inflation for a boom call, with acoustics of bark immediately following the boom. The top panel shows the relationship between the last observed radius of the air sac during a boom call, which is plotted against the acoustic parameters of the subsequent bark. Panel A) shows the results for the acoustic parameter mean sound amplitude, which does not relate to the inflation of the air sac in this case. Panel B) shows the relation of the mean spectral centroid (given in Hertz) of the bark relative to the inflation of the preceding boom. In panel C) correlation estimates are shown for the suite of acoustic parameters included in this exploratory analyses only statistically significant correlation coefficients are

shown. Mean values as well as minimum and maximum values of the acoustic parameters in the bark were tested against the radius of the last boom.

Discussion

The current report breaks terrain in the tracking of elastic kinematics in animals; it uses unsupervised and supervised computer vision methods, focusing on demi-circular soft-tissue structures. In our main use case we focused on the laryngeal air sacs in the Siamang (*Symphalangus syndactylus*). We thereby also provide a data archive of 7+ hours of Siamang singing with closeup video-data ideal for the study of articulatory and air sac states, with multi-source audio data ideal for acoustic analyses. Sharing of data in primatology is relatively rare, and we are convinced that this open dataset will prove valuable for future researchers and is easy to use. All data is already tracked with our DLC+ pipeline. DLC+ was the bestmethod as assessed with ground truth data. Our toolkit does however provide both unsupervised (Hough transform) and supervised computer vision tools (DLC+) to track elastic circular biological structures. Finally, we provide a kinematic-acoustic analysis of air sac inflation and its relation to acoustics of calls, summarized below. Together this report breaks new terrain to solicit research that incorporates a more complete morphometric analysis of animal behavior, which involves bones, but also elastic and expandable soft-tissue.

Our kinematic-acoustic analyses of air sac dynamics in Siamang singing confirmed that the lower frequencies are modulated while the higher frequencies are attenuated, captured by a lower spectral centroid. We find this for adult boom calls and barks following booms in the female boom-bark sequences that occur at early phases of the so-called “great call”. We also see an increased amplitude of the boom calls for higher inflated air sacs. These findings obtained from dynamic real-world data align with modeling research where air sacs were initialized at different air sac volumes (*31*). We further obtain that adults and younger Siamang individuals show different relationships between air sac inflation and the different acoustic parameters assessed, indicating a role of ontogeny. This is not surprising, as physiologically the cavities of smaller air sacs will have different resonant properties due to morphological differences. Furthermore, we observed other interesting patterns in acoustics and air sac inflation, where for example boom f0 increases with air sac inflation. The current exploratory analyses thus shows that with the current toolkit in hand, new ground can be broken to understand the role of laryngeal air sacs in primate communication.

The current toolkit also provides a boon for the collection of new audiovisual data on elastic circular biological structures in the wild, as there are now tested approaches to study said structures. The pipelines provided in our toolkit are fully reproducible and await large-scale application in the diversity of elastic circular biological structures on this planet (Figure 1). Especially the combinations of point estimates using DLC and Landau circle approximation, as implemented in our DLC+, provide the best performance we would recommend for further research endeavors. We encourage researchers to further optimize tools like the Hough transform implementation we found suboptimal at the moment. Perhaps with some adjustments in the pre-processing of images, future studies can improve the Hough transform approach to the point where it becomes more efficient for large-scale use. This would be ideal as in such a case the need for training a neural network as in DLC+ becomes obsolete, sparing labor. Our work incites one more development: the training of a pose model that is based on a range of elastic circular biological structures. This trained model could then serve as an out-of-the-box approach for tracking these circular objects, similar to models already developed (see DLC SuperAnimal, (*43*)). A drawback of the current approach is that we have focused on single-animal tracking, and it invites further development of applications for multi-animal tracking; this is relatively straightforward as a multi-animal DLC approach is already developed (*44*).

This approach, developing morphometry studies to the next level by progressing from bony structures to elastic structures, can help in answering a wide range of questions in species ranging from birds and primates to pinnipeds and frogs. The gular sacs of the greater sage grouse (*Centrocercus urophasianus,* Figure 1) serve as a great example here. This species, inhabiting wide areas of Northamerica, is an important indicator species for the sagebrush steppe ecosystem (*18*). Further study on gular sac dynamics can also answer species-specific ecological questions, for example on their influence on respiration or thermoregulation. Tracking of demi-circular biological objects could potentially also be used to assess the health and development of the species through pure observation, serving as an indicator of ecosystem health. The role of vocal sacs in anuran vocalizations is well documented, but additional functions have been proposed related to respiration, buoyancy control, chemical signaling or even thermoregulation (*17*). The ability to study vocal sac dynamics in detail will help shed light on this multi-functionality (*17*).

The current open-source dataset, open-source computer vision tools, and benchmarking and proof-of-concept analysis provide a way forward in studying diverse biological structures (see Figure 1). In the current highlighted case of Siamang air sacs it can help understand the adaptive functions of these extreme biological modifications. We thereby invite the community to study the dynamic modulation of elastic structures in animals.

Materials and Methods

Overview

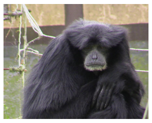
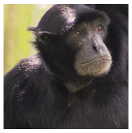
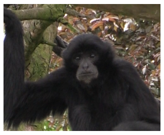
I) Audiovisual recordings were made of six Siamang (Table 2) in captivity over a combined period of about a month in the summer of 2022 (see Audiovisual Dataset). **II)** We investigated two approaches to track the recorded air sac inflations automatically. We used 1) circle tracking through mathematical transformations with the Hough Transformation, an analytic approach, finding circular shapes in images geometrically and 2) circle tracking through point tracking using a trained DeepLabCut model (version 2) with subsequent circle estimation using the Landau algorithm (38), which we refer to as DLC+. **III)** To show the toolkit in action, We then ran two different proof-of-concept analyses on the corresponding acoustic and newly obtained kinematic data. The experimental design is summarized in Figure 1.

I) Audiovisual Dataset

Audiovisual recordings were made 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 + random 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 5 p.m. The random sampling strategy consisted of 30 minutes of recording for a random time slot within the visiting hours of the zoo (but this data is not included in this report). For the opportunistic 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 a particular individual chosen at random. 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 (though 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).

Table 2. Information on individuals

	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 resolution). A Sennheiser ME64-12 was fed as an audio input to the camera (using a DXA-2T audio adapter). We also collected data with 4 go-pros but these materials are outside of the scope of this report (but see (*45*)).

Audio recording. For this study, we used two audio sources. 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 connected a combined multisource audio stream from four KE400 Sennheiser microphones with windjammers sampling at 48 kHz to the go-pros at four locations. We combined all channels into a single mono-channel source by synchronizing audio waveforms using Adobe Premier Pro 2023. This multisource audio stream is more suitable for estimating acoustic measurements sensitive to the sound producer's relative distance or sound radiation direction.

Enriched data. In the repository, there is information about the weather (humidity, temperature, cloudiness, etc.) in each recording session, next to the start and end times of the recordings. Additionally, closeup data is provided with the individual names (which can be linked to the data about age in the repository). Finally, we have tracked all videos in the repository with our DLC+ model using a GPU-based machine, also present 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 (using ImagJ) of images of Siamang closeups. We created a subset of 1612 frames from 9 different scenes and three different days (3 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 centre were exported. If no air sac was visible or deemed to not be trackable because of position or else, a very small circle, clearly smaller than any tracked circle, was drawn at the edge of the frame. That way, those frames could be included in testing the automatic tracking. Before comparison, we transformed diameters into radii.

Unsupervised Computer Vision: Hough Transformation. We used the feature extraction technique Hough Transform to detect imperfect circles (or 'demi circles') in individual video frames (37). Frames undergo a series of preprocessing steps first, to increase the detection success of circles. First, frames are converted to grayscale and optimized for brightness and contrast. We then apply a median blur, and as a next step, we apply 'Canny edge detection' and dilate the found edges. The Canny edge detection is another feature detection method, reducing an image to its edges. 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 (for details, see script xy). After dilating, the image is blurred again with a median blur.

After image preprocessing the Hough Transform can be applied to the image, in which circles with a minimal radius of 5 pixels and a maximum of 270 pixels can be found (depending on your preferences). 270 pixels as a maximum for detection was found to be optimal for our dataset by trial and error and visual inspection. This needs to be adjusted on a case-by-case basis, depending on circle size in the videos. After circle detection in the full video, we apply a Kolmogorov-Zurbenko smoothing algorithm of detected centroids and radii of circles. The Analysis was run with a custom written Python Jupyter Notebook utilizing the OpenCV module (link).

Our Hough detection and the preprocessing process is depicted in Figure 7, where circles detected with the Hough Transform are indicated for all processing steps to illustrate their necessity. The preprocessing parameters were optimized with another custom Python Jupyter Notebook, changing brightness and contrast parameters, median blur, edge detection parameters, and dilation strength (Table 3). This optimization is advised for a different dataset as well. Notice that depending on variability in the dataset, some parameter combinations might work great for some and bad for other images, while another parameter combination might work well (not great, not bad) for all. Another step for potential optimization of circle detection with the Hough transform is the correct choice of the image section. It can increase tracking success if only the particular region of interest is used in the analysis. Thus, we advise cropping the video so that the circular structure is maximally in frame.

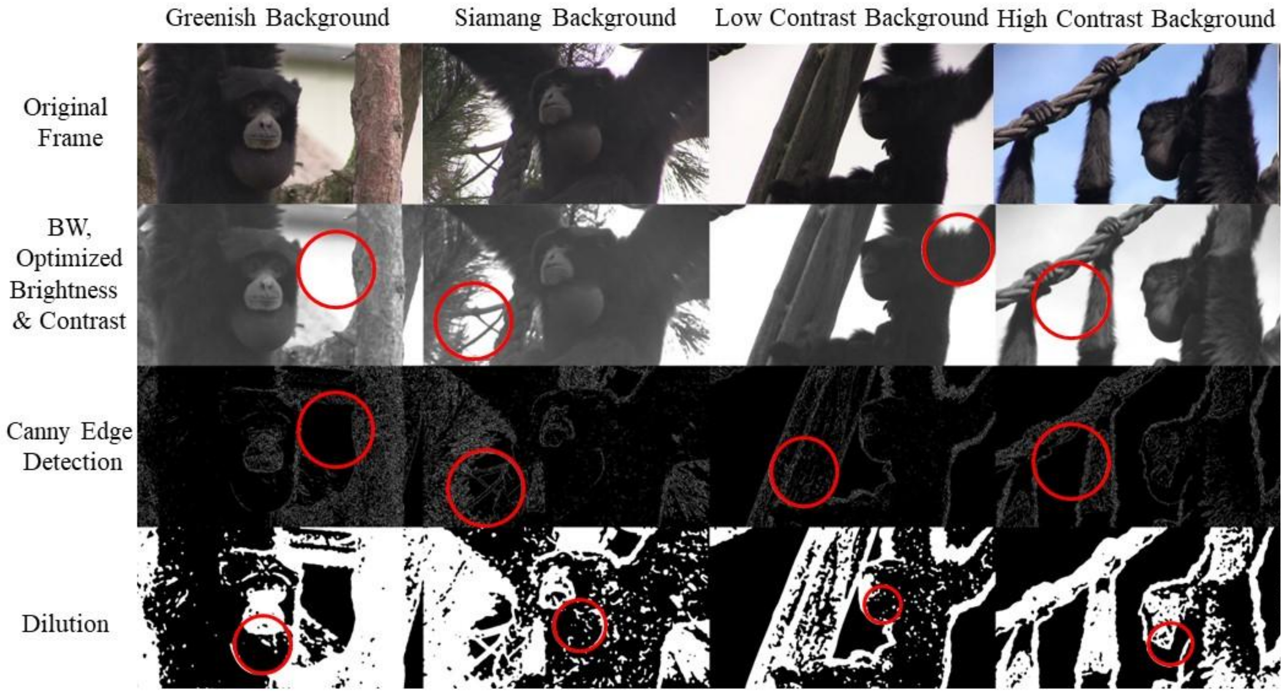


Figure 7: Image Processing to increase Hough Transform circle detection success.

Table 3: 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. Best performance was found by correlating the found radius results to the manually tracked radii and choosing the parameter combination showing 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

Supervised Computer Vision: DeepLabCut + Landau (DLC+)

DLC model info. As mentioned we trained a Resnet-101 convolutional neural network using Deeplabcut 2.0. We first trained a more shallow network, Resnet-50, but this yielded very poor tracking of the air sacs 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 the model metadata can be found here.

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. For the estimation of air sac inflation the key points containing "outline" are used. These key points provide a robust estimate for air sac inflation because of several reasons. Firstly, they were easily formalizable into geometrically defined rules (see figure 8) that can be applied for air sac tracking in multiple angles (see Figure 8 for an explanation). Secondly, though only three points are needed to estimate circles, having redundant points allows for better estimates and the tracking becomes more error-robust because keypoints can sometimes not be tracked at all (the likelihood of all points dropping out a the frame is low, while one dropping out is higher).

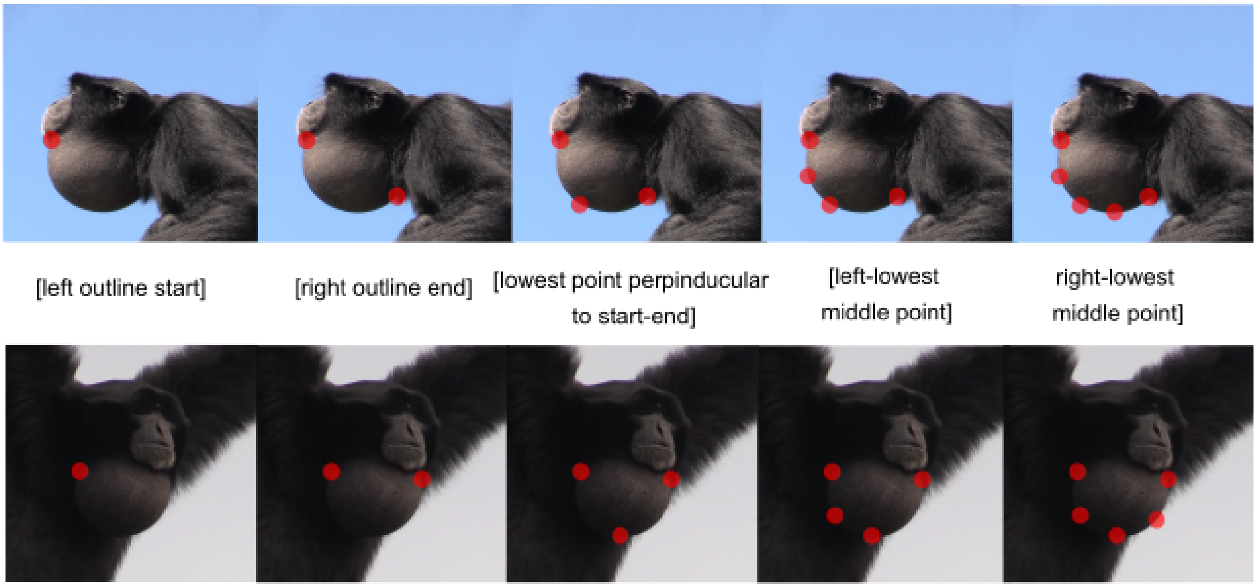

[left outline start]
[right outline end]
[lowest point perpinducular to start-end]
[left-lowest middle point]
right-lowest middle point]

Figure 8. DLC labeling approach for DLC+. 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 demi circle, then the right end of the outline. Then the lowest point is defined perpendicular to the line between start left and end right. Then the middle points along the circle are defined for the left and right side. Note, that this labeling method allows for a well-formed description of each point, regardless of perspective of the siamang.

DLC training set. The total number of hand-annotated images was 390, which was used as a training(90%)/test(10%) set for training of the DLC model. The performances are reported in the main text.

DLC+: Radius estimate using Landau. We use the geometric circle fit method by Landau 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 and was described as very robust and is widely used. The estimation 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 Euclidean distance (geometric) from data points to the fitting curve is used (*41*). We estimate the radius of a circle as well as the x and y coordinates of its centroid. In our analysis only points tracked by DLC are used if they have a likelihood value > 0.6. For frames, where less than 3 points fulfill that criterion, we do not get a radius estimate and the frame is excluded from further analysis. DLC tracking is transformed into radii, x, y, data with a custom made Python Script (link). We also implemented the same routine in an R-script (link).

Kinematic-Acoustic Analysis

Data snippets for the kinematic-acoustic analysis were sampled opportunistically from the full dataset. To be included in the analysis snippets for both analyzes had to fulfill several criteria: camera angle and zoom was not allowed to change within the snippet to get reliable radii estimations to be matched to acoustic parameters, only a single siamang was allowed to vocalize during a snippet and background noise was to be minimized. For the individual boom snippets data was sampled from all four vocalizing siamangs. As great call boom-bark sequences are only8 or primarily produced by females, sample for this analysis were only taken from the only female Pelangi. Booms produced by Pelangi in the analyzed great call sequences were not analyzed as individual booms.

All acoustic analyses are run in `R` 4.2.3. The window length for the acoustic analysis was chosen to match the duration of one videoframe, to be able to directly compare air sac inflation status to the resulting acoustic properties. Videoframe data consistently had a fps of 25. 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 (*42*) was used to perform an extensive analysis, reporting on 47 standard acoustic parameters, such as amplitude, fundamental frequency (called pitch in the analysis), entropy or spectral Centroids (see Documentation in soundgen package and codes for comprehensive list of parameters).

Acoustic analyzes were then conducted on two types of calls: a) boom calls and b) bark calls. For the boom call the acoustics 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 and compared it to the average, minimum and maximum of the acoustic parameters across the subsequent bark. The same window length was used, as for the individual boom call analysis. Acoustics were not analyzed for the boom parts of the great call sequences.

References

1. Z. Cao, T. Simon, S.-E. Wei, Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields" in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Honolulu, HI, 2017; http://ieeexplore.ieee.org/document/8099626/), pp. 1302–1310.
2. C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, M. Grundmann, MediaPipe: A Framework for Building Perception Pipelines (2019), , doi:10.48550/arXiv.1906.08172.
3. A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, M. Bethge, DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*. **21**, 1281–1289 (2018).
4. T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, M. W. Mathis, Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc*. **14**, 2152–2176 (2019).
5. T. D. Pereira, N. Tabris, A. Matsliah, D. M. Turner, J. Li, S. Ravindranath, E. S. Papadoyannis, E. Normand, D. S. Deutsch, Z. Y. Wang, G. C. McKenzie-Smith, C. C.doi Mitelut, M. D. Castro, J. D’Uva, M. Kislin, D. H. Sanes, S. D. Kocher, S. S.-H. Wang, A. L. Falkner, J. W. Shaevitz, M. Murthy, SLEAP: A deep learning system for multi-animal pose tracking. *Nat Methods*, 1–10 (2022).
6. P. C. Bala, B. R. Eisenreich, S. B. M. Yoo, B. Y. Hayden, H. S. Park, J. Zimmermann, Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nat Commun*. **11**, 4560 (2020).
7. R. Labuguen, J. Matsumoto, S. B. Negrete, H. Nishimaru, H. Nishijo, M. Takada, Y. Go, K. Inoue, T. Shibata, MacaquePose: A Novel “In the Wild” Macaque Monkey Pose Dataset for Markerless Motion Capture. *Frontiers in Behavioral Neuroscience*. **14** (2021) (available at https://www.frontiersin.org/articles/10.3389/fnbeh.2020.581154).
8. N. Desai, P. Bala, R. Richardson, J. Raper, J. Zimmermann, B. Hayden, OpenApePose: a database of annotated

ape photographs for pose estimation (2022), , doi:10.48550/arXiv.2212.00741.

9. C. Wiltshire, J. Lewis-Cheetham, V. Komedová, T. Matsuzawa, K. E. Graham, C. Hobaiter, DeepWild: Application of the pose estimation tool DeepLabCut for behaviour tracking in wild chimpanzees and bonobos. *Journal of Animal Ecology*. **92**, 1560–1574 (2023).
10. M. Bain, A. Nagrani, D. Schofield, S. Berdugo, J. Bessa, J. Owen, K. J. Hockings, T. Matsuzawa, M. Hayashi, D. Biro, S. Carvalho, A. Zisserman, Automated audiovisual behavior recognition in wild primates. *Science Advances*. **7**, eabi4883.
11. M. T. Turvey, S. T. Fonseca, The medium of haptic perception: A tensegrity hypothesis. *Journal of Motor Behavior*. **46**, 143–187 (2014).
12. D. E. Ingber, Tensegrity I. Cell structure and hierarchical systems biology. *J Cell Sci*. **116**, 1157–1173 (2003).
13. M. Perlman, R. Salmi, Gorillas may use their laryngeal air sacs for whinny-type vocalizations and male display. *Journal of Language Evolution*. **2**, 126–140 (2017).
14. A. H. Krakauer, M. Tyrrell, K. Lehmann, N. Losin, F. Goller, G. L. Patricelli, Vocal and anatomical evidence for two-voiced sound production in the greater sage-grouse Centrocercus urophasianus. *Journal of Experimental Biology*. **212**, 3719–3727 (2009).
15. J. C. Dunn, Sexual selection and the loss of laryngeal air sacs during the evolution of speech. *Anthropological Science*. **126**, 29–34 (2018).
16. P. Maciej, J. Fischer, K. Hammerschmidt, Transmission Characteristics of Primate Vocalizations: Implications for Acoustic Analyses. *PLOS ONE*. **6**, e23015 (2011).
17. I. Starnberger, D. Preininger, W. Hödl, The anuran vocal sac: a tool for multimodal signalling. *Anim Behav*. **97**, 281–288 (2014).
18. B. E. Brussee, P. S. Coates, S. T. O'Neil, M. L. Casazza, S. P. Espinosa, J. D. Boone, E. M. Ammon, S. C. Gardner, D. J. Delehanty, Invasion of annual grasses following wildfire corresponds to maladaptive habitat selection by a sagebrush ecosystem indicator species. *Global Ecology and Conservation*. **37**, e02147 (2022).
19. T. Owerkowicz, C. G. Farmer, J. W. Hicks, E. L. Brainerd, Contribution of gular pumping to lung ventilation in monitor lizards. *Science*, 1661–1663 (1999).
20. T. Riede, I. T. Tokuda, J. B. Munger, S. L. Thomson, Mammalian laryngseal air sacs add variability to the vocal tract impedance: physical and computational modeling. *J Acoust Soc Am*. **124**, 634–647 (2008).
21. T. Fitch, M. D. Hauser, "Acoustic communication" in *Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals*, A. Simmons, R. R. Fay, A. N. Popper, Eds. (Springer, New York, 2002), *Springer Handbook of Auditory Research*.
22. T. Nishimura, "Primate Vocal Anatomy and Physiology: Similarities and Differences Between Humans and Nonhuman Primates" in *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism*, N. Masataka, Ed. (Springer, Singapore, 2020; https://doi.org/10.1007/978-981-15-4250-3_2), pp. 25–53.
23. U. Reichard, H. Hirohisha, C. Barelli, *Evolution of Gibbons and Siamang: Phylogeny, Morphology, and Cognition* (2016; https://link.springer.com/book/10.1007/978-1-4939-5614-2), *Developments in Primatology: Progress and Prospects*.
24. J. D'Agostino, S. Spehar, A. Abdullah, D. J. Clink, Evidence for Vocal Flexibility in Wild Siamang (Symphalangus syndactylus) Ululating Scream Phrases. *Int J Primatol* (2023), doi:10.1007/s10764-023-00384-5.
25. N. P. McAngus Todd, B. Merker, Siamang gibbons exceed the saccular threshold: Intensity of the song of *Hylobates syndactylus*. *The Journal of the Acoustical Society of America*. **115**, 3077–3080 (2004).
26. F. Mott, A Study by Serial Sections of the Structure of the Larynx of Hylobates syndactylus (Siamang Gibbon). *Proceedings of the Zoological Society of London*. **94**, 1161–1170 (1924).
27. B. E. Hastings, The veterinary management of a laryngeal air sac infection in a free-ranging mountain gorilla. *Journal of Medical Primatology*. **20**, 361–364 (1991).
28. G. Hewitt, A. MacLarnon, K. E. Jones, The functions of laryngeal air sacs in primates: A new hypothesis. *FPR*. **73**, 70–94 (2002).
29. D. F. N. Harrison, *The Anatomy and Physiology of the Mammalian Larynx* (Cambridge University Press, Cambridge, 1995; https://www.cambridge.org/core/books/anatomy-and-physiology-of-the-mammalian-larynx/374FE10734305D20EFFF480DA818F535).
30. S. Hayama, The origin of the completely closed glottis. Why does not the monkey fall from a tree? *Primate Research*. **12**, 179–206 (1996).
31. B. de Boer, Acoustic analysis of primate air sacs and their effect on vocalization. *J Acoust Soc Am*. **126**, 3329–3343 (2009).
32. S. R. Partan, P. Marler, Communication Goes Multimodal. *Science*. **283**, 1272–1273 (1999).
33. T. Nishimura, A. Mikami, J. Suzuki, T. Matsuzawa, Development of the Laryngeal Air Sac in Chimpanzees. *International Journal of Primatology*. **28**, 483–492 (2007).

34. W. T. Fitch, B. de Boer, N. Mathur, A. A. Ghazanfar, Monkey vocal tracts are speech-ready. *Science Advances*. **2**, e1600723 (2016).
35. W. T. Fitch, The Biology and Evolution of Speech: A Comparative Analysis. *Annual Review of Linguistics*. **4**, 255–279 (2018).
36. Z. Alemseged, F. Spoor, W. H. Kimbel, R. Bobe, D. Geraads, D. Reed, J. G. Wynn, A juvenile early hominin skeleton from Dikika, Ethiopia. *Nature*. **443**, 296–301 (2006).
37. I. Martínez, J. L. Arsuaga, R. Quam, J. M. Carretero, A. Gracia, L. Rodríguez, Human hyoid bones from the middle Pleistocene site of the Sima de los Huesos (Sierra de Atapuerca, Spain). *Journal of Human Evolution*. **54**, 118–124 (2008).
38. J. Giovannello, R. V. Grieco, N. F. Bartone, Laryngocele. *American Journal of Roentgenology*. **108**, 825–829 (1970).
39. D. H. Ballard, Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*. **13**, 111–122 (1981).
40. R. O. Duda, P. E. Hart, Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM*. **15**, 11–15 (1972).
41. N. Chernov, C. Lesort, Least squares fitting of circles and lines (2003), , doi:10.48550/arXiv.cs/0301001.
42. A. Anikin, soundgen: Sound Synthesis and Acoustic Analysis (2023), (available at https://cran.r-project.org/web/packages/soundgen/index.html).
43. S. Ye, A. Filippova, J. Lauer, M. Vidal, S. Schneider, T. Qiu, A. Mathis, M. W. Mathis, SuperAnimal models pretrained for plug-and-play analysis of animal behavior (2023), , doi:10.48550/arXiv.2203.07436.
44. J. Lauer, M. Zhou, S. Ye, W. Menegas, S. Schneider, T. Nath, M. M. Rahman, V. Di Santo, D. Soberanes, G. Feng, V. N. Murthy, G. Lauder, C. Dulac, M. W. Mathis, A. Mathis, Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat Methods*. **19**, 496–504 (2022).
45. W. Pouw, M. Kehy, M. Gamba, A. Ravignani, Cross-modal constraints in multimodal vocalizations in Siamang (Symphalangus syndactylus) (2023) (available at https://ecoevorxiv.org/repository/view/5688/).

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 A.R. Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF117). 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.

Author contributions:

Conceptualization: WP
Methodology: WP, LSB
Investigation: WP, MK, LSB
Visualization: WP, LSB
Supervision: WP, LSB
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/tree/main). 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).