Postprint: This article is accepted for publication at *International Journal of Comparative Psychology*

Amplitude Increases of Vocalizations are Associated with Body Accelerations in Siamang (*Symphalangus syndactylus*)

Wim Pouw[2], Mounia Kehy[3], Marco Gamba[4], Andrea Ravignani[5, 6, 7]

1. Tilburg University, Department of Computational Cognitive Science, Tilburg University

2. Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, Netherlands

3. Equipe de Neuro-Ethologie Sensorielle, Université Jean Monnet, France

4. Dipartimento di Scienze della Vita e Biologia dei Sistemi, Università di Torino, Torino, Italy

5. Comparative Bioacoustics Research Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

6. Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music, Aarhus, Denmark

7. Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

Author note: Correspondences can be addressed to Wim Pouw (wim.pouw@donders.ru.nl). We would like to thank the 'Jaderpark Tier- und Freizeitpark an der Nordsee' for allowing us to record audiovisual data of the Siamang family at their facility. This work has been supported by the Max Planck Institute (MPI) for Psycholinguistics Nijmegen, and the Donders Institute for Cognition, Brain, and Behavior. We would like to thank Jeroen Geerts of the MPI, for support of organizing the audiovisual equipment, and Maarten Snellen with his help setting up a dedicated server to run our dashboard application. We would like to thank Diandra Düngen for her support in making the recordings possible. WP is funded by a VENI grant (VI.Veni 0.201G.047: PI Wim Pouw) and was further supported by a Donders Postdoctoral Development fund. 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). AR is funded by the European Union (ERC, TOHR, 101041885). **Open data:** All data and code supporting this manuscript can be found on https://github.com/WimPouw/siamang_physical_constraints_code_repo

Abstract

Siamangs (*Symphalangus syndactylus*), one of the few singing apes, vocalize loudly, often while they move. We hypothesize that movement and vocalization coordinate, possibly due to vigorous thorax-loading movements such as brachiation affecting vocal-respiratory dynamics. To assess this vocal-motor coordination we recorded more than a hundred stereotypical vocalizations combined with movement from two captive Siamang (isolated from 7 hours of singing). We observed that stereotypical calls coincided with a movement display and were performed by juvenile individuals during solo singing (which allowed for isolation of the calls). Investigating these vocal-motor events, we found that body acceleration estimated using computer vision was statistically associated with the nearest peak in the amplitude envelope of the call, and that body acceleration timeseries contained mutual information about the amplitude envelope timeseries during these events. By confirming via quantitative methods that singing and movement are coordinated, the current report invites further mechanistic investigation on vocal-locomotor coupling in siamang.

Keywords: Siamang, Locomotor-vocal coupling, Locomotion, Respiration, Vocalization, Multimodal Communication

Introduction

Human and non-human animals often coordinate whole-body movement with vocalizations. Multiple bat species (e.g., *Phyllostomus hastatus*) flexibly synchronize their echo-locating pulses with their wingbeats while flying, often in 1:1 or polyrhythmic fashion [1–4]. Twelve species of North- and South-American birds show an allometrically coupled wingbeat duration with vocal unit durations [5]. Gerbils (*Meriones unguiculatus*) locomote in a saltating way, when they hop and hit the ground with their forelimbs they synchronously emit a vocalization [6]. The male brown-headed cowbird (*Molothrus ater*) uses vigorous wingbeats in courtship displays, and such activity affects respiratory-vocal activity [7]. Finally, humans also show synchronization of impulses of their limb movements with their vocalization (Pearson and Pouw, 2022; Pouw et al., 2020; Pouw et al., 2023; Serré et al., 2022; for an overview see Pouw and Fuchs, 2022).

Surprisingly, it is virtually unknown whether non-human primates synchronize their whole-body movements with vocalization. One taxon seems the perfect model to study vocal-motor interactions, building on anatomical and neural circuitry shared across primates, including us: the Gibbons and Siamang (*Hylobatidae)*. They are highly vocal species and they load their entire body weight on their pectoral system during their primary mode of locomotion – *brachiation*. Here we focus on the heavier-weight Siamang (*Syndactylus symphalangus*).

Siamang and Gibbons diverge in several ways from great apes. They are a highly vocal species, performing extremely loud (sometimes > 100 Decibels [13,14], rhythmically coordinated songs [15], supporting family-bonding and territory-marking, and on occasion alarming. While usually understood as containing stereotyped vocalizations, Siamang song contain considerable variability for certain call phrases [16]. Siamang and Gibbons apes also move at extremely fast 45km/h speed using hand-over-hand grasps, also known as brachiation. These Small Asian Apes also move *while* they sing (e.g., see here). Interestingly, Haimoff (Haimoff, 1981), *p*. 135) observes a temporal coordination of locomotion and vocalization in the wild Siamang they studied. These and other widespread *qualitative* observations [18,19] suggest that singing in Siamang and Gibbons may at times be a combined display[1], such that two

[1] Note that some Gibbon species (but not Siamang) are known to also dance outside of the context of singing [20], but that refers to a different sort of behavior.

behaviors (vocalization & movement) that in principle can operate alone, structurally operate together, much like other animals that move and vocalize in coordinated ways [1,5,7,21]. However, no quantitative evidence exists to support the idea of a combined display. Perhaps in part because it is difficult to isolate calls often produced in group singing, and in part because movements of these apes are difficult to track – at present it seems virtually impossible to track movements of the Siamang in the wild as these apes move through the canopies with high speeds [22].

One interesting possible reason for vocal-motor coupling in Siamang is biomechanics. A range of animals that include their pectoral limbs for locomotion (e.g., bats, dogs, horses, rhinoceros) synchronize their locomotor cycles at increasing gait speeds with respiratory cycles [23]. This synchronization is held to occur because of a piston-like effect where the visceral organs push forward on the lungs during accelerations and decelerations of the body during locomotion strides (Daley et al., 2013; Lafortuna et al., 1996; for an overview see Pouw and Fuchs, 2022), or because muscle activation for locomotion driving movement accelerations simultaneously compress the rib-cage [1]. Interestingly, it has been generally acknowledged that brachiation in primates likely affects respiratory control and thus vocalization [26–29]. Thus from biomechanics of brachiation alone we would hypothesize some respiratory interactions, suggesting some potential for vocal-motor coupling. Other non-mutually exclusive hypotheses for coordination of voice and whole-body movements hold that combining visual and auditory signals increases the likelihood of communicative success [30,31].

In this study we opportunistically observed captive Siamang engaged in solo singing to assess vocal-motor coupling, whereby we audio-visually recorded singing for a period of 21 days. We noticed characteristic vocalizations combined with pulse-like movements (e.g., for examples see here). These movement-accompanied calls contained ululating screams as main units [32] and were produced by juveniles who engaged in solo singing after the duetting singing performed by the entire group was completed or was winding down. To our surprise these specific stereotypical calls were *always* produced with a pulse-like movement. These pulse-like movements were seemingly synchronized with the calls and we will assess whether these movements associate with the loudness of these solo-songs in two analyses. We know from biomechanics in humans and other animals that the physical impact of a movement on the

musculo-skeletal system is during acceleration or deceleration (as forces are a function by mass [a constant] and acceleration) [6,9,24,25]. Therefore, we test here whether thorax accelerations during vocalizations statistically relate to the amplitude of concurrent vocalization in Siamang.

Materials and methods

Data recording

Audiovisual recordings of a family of Siamang (6 members; female adult, male adult, two male juveniles, one infant, and a newborn) were collected in the June and August of 2022 over two visits at the Jaderpark Zoo in Lower Saxony, Germany. This yielded over 7 hours of recorded singing, collected by the first and second authors. The Siamang sang primarily in the morning around 9-10am, or after their fruit and vegetable lunchtime, around 1pm, and occasionally around 5pm. Only two juvenile/young adult individuals performed the stereotypical movement-accompanied vocalizations we consider here, and these events were ideal for acoustic analysis as there was no overlap compared to the typical collective singing of Siamang. Due to the adults and juveniles singing together, there is almost constant overlap in calls, which made us focus on the solo singing of the juveniles in this first investigation of siamang singing and movement. The two individuals were Baju (7 years and 8m) and Fajar (4 years and 11m). Baju and Fajar were both born in the Jaderpark zoo. During data collection, Baju was separated from the family due to risk of injury after a fight where all family members attacked Baju. This also means we could not collect *more* data from Baju during our second testing period as he was transferred to a smaller facility and stopped singing during our visit when the transfer was made.

Audio-visual recording

Four GOPRO Hero9 were installed, set at 1080 quality, sampling at 59.74fps, with linear lens settings. We then cropped frames and recompressed the video to 50fps. The camera positions were positioned as orthogonally from each other as the site allowed. Figure S1, panel a, shows a sketched map with geometrically estimated distances based on a laser-pointer measurement device. We further use in this study four audio sources from Sennheiser microphones with windjammers (MKE400), plugged into the GO-PRO ensuring audio-visual synchronization, sampling at 48kHz.

Recording was set at similar gains across microphones, and we checked for clipping during pilot recordings. The four acoustic waveforms were combined using an 'Adobe Premiere Pro 2019 CC' waveform alignment, which uses a cross-correlation approach to find the optimal lag to synchronize peaks in the audio. After synchronizing the waveforms of the four sources, we recompressed the audio as a single-channel audio source which thereby contains the combined time-aligned sources (48kHz). Therefore, we always estimate amplitude using four combined audio sources that collected from multiple locations to increase our measurement accuracy to track the sound's amplitude. Specifically, we placed the audiovisual recorders at four different angles; this strategy minimized problems with differences in sound radiation and differences in sender-recording distances across different events, which may otherwise affect measurement of amplitude peaks. However, the amplitude measurements will be flawed in open environments such as these. For this reason, we also report a second time series analysis unaffected by differences in amplitude measurements across events (see lagged mutual information).

Identification of movement-accompanied vocalizations and related features

The first and the second author identified opportunistically as many movement-accompanied vocalizations by going through all the recorded songs and annotating these events in ELAN [33]. These vocalizations were easy to identify because they all consisted of a ululating scream as a main unit, and any variability in the call structure was stereotypical within each of the two individuals. Interestingly, these stereotypical vocalizations always co-occurred with a pulse-like movement, though with variable intensity and different types of locomotion. The annotators (second and first author) drew a boundary around the movement-vocalization event, such that the movement and the vocalization sequence was contained. We will refer to these annotated events as movement-accompanied vocalizations throughout.

Importantly, we did not observe any stereotypical vocalizations that occurred with no observable movement at all (though some contained small amplitude movements, and these are part of the variability in our dataset; e.g., see supplemental table S1, Example 1). Please also see a longer segment of a juvenile's singing in which it becomes clear that the stereotypical movement-vocal calls that comprise our data are embedded in other types of calls such as sequences of barks: see supplemental table S1, Example 2).

Also note that we focused on Juveniles because they would sing on their own. The adult female and male produced movement and vocalizations too. However, their movement-accompanied vocalizations almost always occurred when all individuals were singing. This lead to frequent, multiple overlaps. Analyses of overlaps that would require another set of acoustic post-processing steps that would greatly complicate the analysis relative to the current approach.

The degree of physical interactions expected are likely dependent on the type of action performed during vocalizations; therefore, we attempted to apply a standardized description to locomotor actions. We used Hunt's typology of locomotion types [34] to characterize the action that occurred during the vocalization (Figure 1). In some instances, we slightly deviated from Hunt's [35] categories to accommodate for a particular locomotion action (e.g., 'drop fore limb swing': the individual sits on top of horizontal structure, and then scoots backwards or forward to drop and then swing forward with two extended arms). A common mode of locomotion for Siamang, ricochetal brachiation (e.g., see supplemental table S1, Example 3), is absent in our dataset, possibly due to the facility having more ropes than rigid and connected supports, and thus being more tailored towards swinging movements rather than ricochetal brachiation. Some movement-accompanied vocalizations remained undefined as they did not fall into a clear category (i.e., mixed locomotion modes). We will make a crude binary distinction between locomotion types which load the entire weight on the thorax via the shoulder girdle(s) (forelimb only[2]), or those that involve distribution of weight or support via the lower limbs (other). In the case of forelimb only loads, one would particularly expect accelerations to constrain respiratory-vocal interactions.

Video preprocessing and post-processing

Each event can potentially be recorded via four camera angles. We first checked all these camera angles to see whether the individual was visible. If the individual was not visible, it was excluded as a potential camera angle submitted for analysis. Video processing was performed in Python, the specific steps discussed below. Further processing to prepare the dataset for statistical analyses was performed in R and R-studio.

[2] Strictly speaking, since siamang are primarily bipedal, we could have also referred to the pectoral limbs as upper limbs (rather than forelimbs). But we decided to follow Hunt's categories here.

Cutting scenes and performing initial motion detection with OpenCv2. Firstly, a custom Python script automatically cut the videos based on the ELAN annotations begin and end times using moviepy, ffmpeg and pydub (see supplemental table S1, Script 1).

Secondly, we determined regions of interest for each scene. The installed cameras had a field of view to opportunistically capture behavior at different locations. This makes tracking with supervised computer vision more computationally costly as there may be many individuals moving in a complex structured site. As a pre-processing step we therefore created a Python pipeline (see supplemental table S1, Script 2) using OpenCv2 that determined pixel deviations from the median to ascertain a key area of movement per frame. After smoothing and obtaining maxima, this key area was used to determine a static bounding box that would further serve to crop the video to primarily contain the movement of the siamang and exclude the rest of the complex scene (for an example see Figure S1 panel b).

Kinematics

Tracking DeepLabCut. A convolutional neural network (Resnet-50) was trained using DeepLabCut (version 2)[36] with 250 hand-labeled frames (see Supplemental table S1, Resource 1). Two key points were used for the training set. The first key point effectively tracked the upper thorax region targeted at the most posterior region (i.e., which was used for acceleration). The second key point was targeted at the sacrum of the individual (which was used for the body normalization of the acceleration magnitudes).

We trained the model with half a million iterations, reaching an average error rate of 1 pixel (for keypoints with 60% confidence rates) in the training set, and 25 pixels in the test set. If we normalize these errors by the original frame-sizes (1500x1080), then we get an error of 0,0015%.

When using the trained model to extract position traces for the video recordings, we applied the model to all events with DLC's native filtering option to remove noise-related jitter, yielding x,y position traces for the two key points (see Supplemental table S1, Example 4) and likelihoods. Since derivatives increase power of noise-related jitter relative to slower frequencies [37], we also applied extra smoothing of the resultant position traces with a 9th order Kolmogorov Zurbenko filter with a span of 110ms (`R`-package `kza`).

Likelihoods were further used for data quality curation. Specifically, if a camera angle had tracking that dropped for more than 5% below a threshold of .80, than we did not submit that particular camera for further analyses; the remaining 5% of the data, that had low confidences, were linearly interpolated (`na.approx` function using `R`-package `zoo`) using the surrounding high-confidence tracking samples. Note that the DLC team has recommended a likelihood of at least .60 for good tracking, and our pipeline is slightly more conservative than that.

Normalization. The thorax accelerations were calculated by differentiating speed measured in pixels over time. However, different camera positions, and different locations of the Siamang, make pixel-units problematic: a pixel as a unit of space will differ per distance of the camera to the individual. Therefore, we normalized the kinematics to a dimensionless quantity by scaling the position traces by the mean body size of the Siamang detected (for all frames that had a DeepLabCut confidence estimate of 100%). This means that all kinematics in this report are always first normalized by body size units.

Approximating kinematics from multiple camera angles. Depending on the location of the individual, we had one to four camera angles that recorded the movement-accompanied vocalizations. The many objects on the site, the distances of the cameras, the lack of access to the site, combined did not allow us to perform stereoscopic reconstruction of the camera angles to estimate 3D postures using a Charuco board [38,39]; this means the cameras could not be spatially calibrated for angle intrinsics and extrinsics. Note further that it would not be inappropriate to simply combine the 2D accelerations determined for each camera (by taking the Euclidean norm of all the 2D accelerations recorded) as this would yield an overestimation given that camera angles have correlated information (for example, because they all capture vertical acceleration of the individual). Therefore, to combine the information we need to find the non-correlated information in each of the camera views.

To still make use of multiple cameras (in the case when more than one camera had sufficient-quality data) that were impossible in the zoo to calibrate, we devised another solution to utilize multi-angle camera information using Principal Component Analyses (see e.g., [40], which formed the basis for our approximation of the 3D acceleration of the Siamang. We combined position traces of n cameras (c) available $\mathbf{M} = [x_{c=1}, y_{c=1}, x_{c=2}, y_{c=2.} \ldots x_{c=n}, y_{c=n}]$ to a principal component analysis, to calculate the eigenvectors (v), where we extract the three

highest loading principal components, $\mathbf{PC}_{1\text{-}3} = [\mathbf{M}v(:,1), \mathbf{M}v(:2), \mathbf{M}v(:3)]$, which will approximate positional information about orthogonal planes. Subsequently, we take the Euclidean norm of the derivatives of the first three principal components to get an approximation of 3D speed vector s, such that $\vec{s} = \sqrt{(\Delta PC_1^2+\Delta PC_2^2+\Delta PC_3^2)}$. This norm is then differentiated once more, and absolutized, to yield an approximated 3D acceleration magnitude vector, ($\vec{a} = |\Delta\vec{s}|$). Note after each differentiation we smooth the data with a 5th order Kolmogorov Zurbenko filter with a span of 50ms.

Acoustics

Smoothed amplitude envelope. We extracted a smoothed amplitude envelope of the waveform from the combined audio sources using a common approach, by first applying a Hilbert transformation [41], and then taking the complex modulus. This resulted in a one-dimensional time series, which was downsampled to 100Hz and further smoothed with a 12 Hz Hanning window.

Aggregation of multiple data streams

We synchronized the motion tracking data and the acoustic data by first aligning the data samples in time, and upsampling the motion tracking to 100Hz to preserve the high-sampling rate of the acoustics. We upsampled the motion tracking data by linearly interpolating the data along a time vector using `R`-package `zoo` (function `na.approx`; Zeileis and Grothendieck, 2005) so as to have a regular sampling rate at 100Hz that exactly matches the sampling times of the amplitude envelope. This yields a combined time series with acoustics and motion tracking that could be further processed for analysis. `R`-code performing the above-mentioned processing steps is available online (see supplemental table S1, Script 3).

Final processed dataset

After having to exclude 29 events that had low-confidence tracking (trackings that dropped more than 5% below a threshold of 80% likelihood determined by DeepLabCut) the final dataset consisted of 83 events (of which; 26 Baju and 57 Fajar).

Analysis 1: Peak analyses

For each movement-accompanied vocal event we determined the global maximum in absolutized acceleration (max acc), i.e., the largest magnitude of either acceleration or

deceleration. Then we obtained the local maxima in the smoothed amplitude envelope, with a `findpeaks` function using `R`-package `pracma`. This allowed us to select the magnitude of the local peak in the envelope nearest to max acceleration. Since the acoustic and acceleration peak data were long-tailed distributed, we log transformed the variables (which also improved model fits), and z-normalized. As additional information, we also report change in amplitude in terms of Decibels by scaling the raw envelope signal by dB = 10*log10(x). This way, we can estimate how body acceleration tends to increase the amplitude by a certain amount of dB.

Analysis 2: Lagged Mutual Information Analyses

We used the function `cmi` from `R`-package `mpmi`[43], to compute for each vocal-motor event the continuous mutual information (cmi) at different lags, where cmi at some lag is defined as $I(X;Y)\tau$ in eq. 1:

$$I(X;Y)\tau = \int\int p(x_{t-\tau}, y_t) \cdot \log\left(\frac{p(x_{t-\tau}, y_t)}{p(x_{t-\tau}) \cdot p(y_t)}\right) dx_{t-\tau}\, dy_t \qquad (1)$$

where X_t is the z-normalized timeseries of body acceleration, Yt is the z-normalized timeseries of amplitude envelope, and τ is a predefined lag capturing the time delay between the two variables. Note, that the *p(x, y)* refers to the joint probability density function of X and Y, and *p(x)* and *p(y)* the marginal probabilities.

Mutual information thus provides an estimate of the information gained in Y given X (at a particular time lag). This analysis helps uncover linear and non-linear covarying information and temporal dependencies between body acceleration and amplitude envelope. By z-normalizing the time series within each event, this analysis is not dependent on the relative differences in absolute amplitude of the sound and movement between events, but rather is a test of whether movement and sound within an event are mutually coupled.

False random pair. A strong comparison condition was constructed to determine whether mutual information was higher in the observed time series relative to some baseline. Lagged mutual information was produced by comparing the amplitude envelope with a randomly simulated time series, together forming a *false random pair*. For each event comparison, the simulated body acceleration time series had the same mean, standard deviation, and

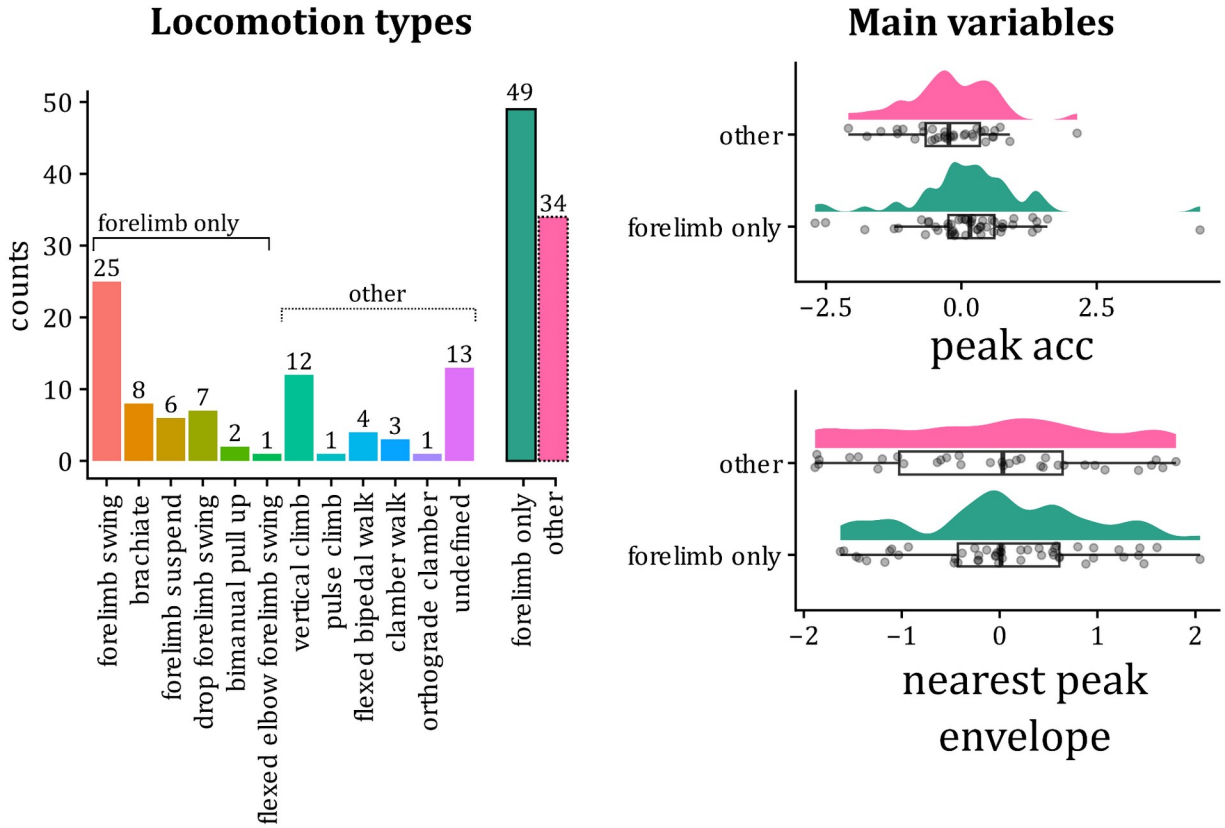
autocorrelation as the original event's body acceleration time series (using an autoregressive integrated moving average, or ARIMA, procedure). We would expect that mutual information is lower in false random pairs relative to real paired movement-vocalization events.

Main measures. To summarize the data, we obtained the maximum observed mutual information between movement and sound for each event over the different lags (max mutual information). We also collected the lag at which the maximum mutual information was observed to understand whether movement is more predictive of sound or vice versa.

Descriptive statistics

An R markdown-script supporting the statistical analyses can be found online (see Supplemental table S1, Script 4). Figure 1 and Table S2 provide information about the main variables, as well as the time windows of the annotations (approximate length of the multimodal events). The temporal inter-peak distance between the global maximum in acceleration and the nearest local maximum in amplitude was on average 120 milliseconds (*SD* = 160). This low temporal distance provides confidence that the two point-estimates for kinematics and acoustics occurred sufficiently close in time to be possibly coupled. Based on 95% confidence intervals, the older juvenile Baju seemed to generate higher peaks in the amplitude envelope (nearest to peak acceleration) as compared to his younger brother Fajar, while being comparable in their body accelerations peaks (see Table S2). Further, forelimb only locomotion tends to have higher accelerations than other locomotion types. The main variables did not dramatically differ either by individual or locomotion type.

Figure 1. Frequency distributions for the locomotion types, peak acceleration, and limbs. The left panel shows the number of different locomotion types that we categorized for each vocal-motor event. Since there are too many categories, with few instances, we created super categories that indicate locomotion actions ($N = 49$ forelimb only, $N = 34$ for other types of locomotion) that only included the fore/upper limbs (forelimb only), versus those that included another limb (other). We use this super category to check whether our continuous kinematic-acoustic analyses may give different results when we also consider what type of locomotor type was performed. On the right panel, the distributions are shown for the magnitude (z-scaled units per individual) in the global peak of acceleration, per binary locomotion category, showing for example higher mean acceleration for forelimb only locomotor actions as opposed to other actions that include the hindlimbs.



Results and Discussion

To assess whether body accelerations constrain vocalizations amplitude, we assess whether both parameters reliably scale in magnitude. Figure 2 provides a summary of the procedure and the main results.

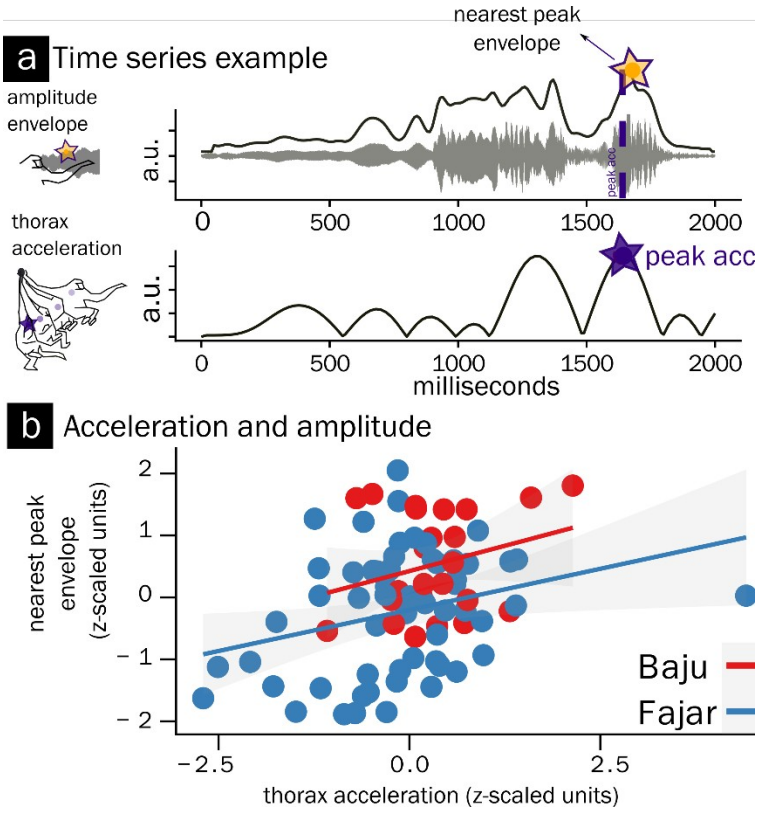
A mixed linear regression model was performed associating peak acceleration with nearest peak of the call amplitude envelope (using maximum likelihood from R-package `nlme`; Pinheiro et al., 2019). Individual (Fajar, Baju) was set as random intercept, but a model with

random slopes for individual did not converge. By setting Individual as random intercept we assess a relationship within individual, and we therefore do not simply lump together the data in establishing an association between body acceleration and amplitude envelope. The model statistically predicting peak envelope from peak acceleration reliably outperformed a base model predicting the overall mean in amplitude; change in χ^2 (1) = 7.68, p = .006. The model coefficients indicated that a higher magnitude peak in acceleration reliably associated with higher magnitude envelope peaks, b = 0.29, b 95%CI = [0.08, 0.49], t(80) 2.82, p = .006, intercept b = 0.09, t (80) = 0.41, p = 0.682. Since the models are performed on a log-log scale, we can interpret the coefficients as indicating that for a unit of increase in acceleration there is a 30% increase in the magnitude of the nearest peak envelope (i.e., ⅓ power relationship). If we rescale the amplitude envelope to dB, and perform the same model we yield that per body scaled change in acceleration, there is an increase in the peak envelope of about 1.4dB, b = 1.43, b 95%CI = [0.043, 1.435], t(80) 2.82, p = .006 (see supplemental online figure). A simple regression analysis (which lumps the data together) yields a similar conclusion for the main effect of body acceleration, r = .33, t(81)= 3.18, p = .002. Also, note that if we remove the possible outlier shown in Figure 2 panel b (see supplemental online figure without outlier) the conclusions remain unchanged, r = .38, t(80) = 3.1829, p < .0001.

We explored possible confounds (e.g., number of camera angles available) or moderators (e.g., locomotion type) of this kinematic-acoustic coupling via interactions, but such interaction models were not reliably outperforming our main model, χ^2 (1) < 8.44, p's > .21 (see supplemental online figure). These checks for confounds or conditional factors provide confidence that effects of acceleration and vocalization are stable among different measurement conditions, individuals, and locomotion types.

Thus, we can conclude that the magnitude of vocalization amplitude peaks that occur around moments when the thorax undergoes its maximum acceleration or deceleration is positively associated with the magnitude of that acceleration. For an inspection of the movement-accompanied vocalizations underlying figure 2 see our dynamic data dashboard (for code, see supplemental Table S1, Script 5).

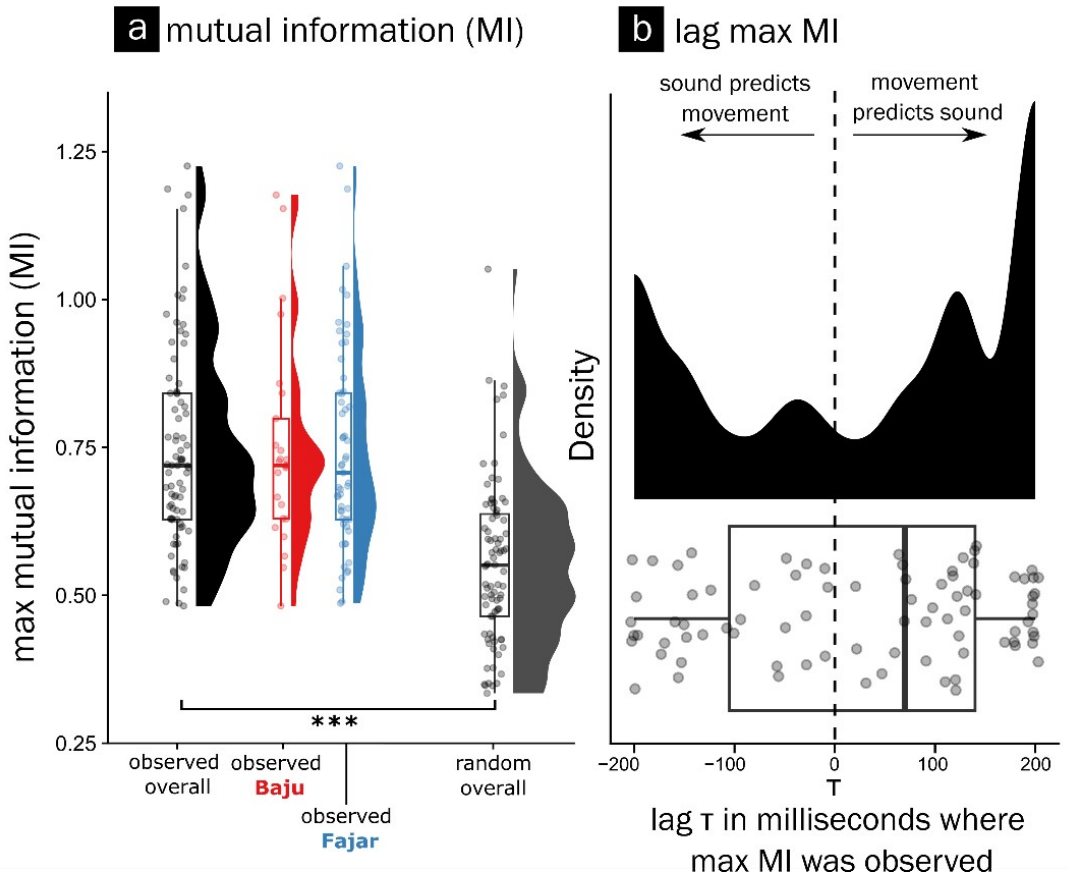
Figure 2. Nearest peak analyses and results for acceleration and amplitude envelope The time series example (a) shows data for a single event (of 83 in total, 26 Baju and 57 Fajar). The upper panel of (a) shows the smoothed amplitude envelope of the call; the lower panel of (a) shows the acceleration of the thorax, both given in arbitrary units (a.u.). The purple star shows the global maximum for acceleration. This maximum is then used to determine the nearest local maximum in the amplitude envelope (yellow star, with purple outline), which in this case also happens to be the global maximum in amplitude. The point-estimates are then submitted for linear mixed regression with random intercepts for each individual, Baju and Fajar (b). Panel (b) shows the nearest peak in body-scaled z-normalized thorax acceleration on the x-axis related to the nearest z-normalized peak in the amplitude envelope on the y-axis (this relation was found to be statistically reliable in a mixed regression analysis, $p = .006$). See the ***online dynamic dashboard* to *explore the audiovisual data*** underlying these data points which allows for exploring each underlying audiovisual event.



While the peak magnitude analysis provides a promising indication that body accelerations are associated with the loudness of the call, there is always a risk of unreliable results given noisy amplitude measurement conditions, which may vary over the different occasions the calls were produced. Therefore, we devised another test to statistically confirm an association of movement and sound based on lagged mutual information calculations that assess the continuous co-variance between two continuous variables. We assess mutual information at

different lags (-200 to +200 ms) in movement (body-scaled acceleration, z-normalized for each event) and sound (amplitude envelope, z-normalized for each event). We set the lags at a maximum of 200 milliseconds as any longer delays are unlikely to reflect a close coordination of sound and movement (but rather some sequential organization). The lagged analyses allow for possible coupling delays, indicating whether movement predicts sound in time rather than vice versa.

Figure 3. Mutual information between normalized movement and sound a) The maximum observed mutual information (in bits) in body acceleration and amplitude envelope over lags from -200 milliseconds to 200 milliseconds is shown for observed movement-accompanied vocal events (overall [N=83]; as well as split over individuals, Baju [N=26] and Fajar [N=57]) and false random pair time series (random overall [N=83]). It is clear in a) that observed time series have much higher mutual information than random pairs which was confirmed in a linear mixed regression ($p < .0001$), and it can be seen that Baju and Fajar show this high mutual information. b) shows the lag at which the maximum mutual information is found for all actually observed events as expressed in milliseconds. We obtain that relatively more often a maximum in mutual information (MI) is found for when movement predicts sound in time, but our mixed regression analyses did not find that a lag at max MI is reliably higher than 0ms ($p = .08$).



A mixed linear regression model assessed whether the observed maximum mutual information is higher than a false random pair. Individual variability (Fajar, Baju) is accounted for as a random intercept (random slopes did not converge). A model contrasting a false random pair with the observed time series outperformed a model predicting the overall mean of mutual information, $\chi 2$ (1) = 54.49, p < 0001. The mutual information for false random pairs was indeed reliably lower than the observed mutual information, b = 0.20, b 95%CI = [-0.25, -0.15], t(163) -7.982, p < .0001. We further assessed whether the lags at which we obtained the max mutual information were reliably different from 0, thereby indicating that either sound or movement was predictive of the other in time. Though movement seemed more likely to be maximally predictive of sound at a lag between 0 to 200 milliseconds (see Figure 3) rather than the other way around, the intercept of a mixed regression model (with individual as random intercept) predicting the overall mean of the lag at max mutual information was not reliably different from 0 with a two-sided test, b = +27ms, t (81) -1.80, p = .08. To conclude, we find clear additional evidence that there is mutual information in movement and sound, even if we ignore absolute measurements of sound and movement (as we z-normalize each time series per event). This report has therefore to remain inconclusive about whether movement predicts sound in time (though the data pattern toward movement predicting sound rather than vice versa).

The evolution of vocal production can be framed as a continuous stabilization of multiple interacting subsystems originally evolved for different purposes [45–49]. Many bodily systems must fall in line to produce such skilled and energetic vocal duet singing in Gibbons and Siamang. This duet singing is certainly shaped, as hypothesized till now, by several adaptations, such as complex vocal tract shaping [13] and the presence of laryngeal air sacs [50–52]. However, here we show that that co-vocal movement and vocalization are mutually coupled in time and amplitude. We firstly obtain striking stereotypical pulse-like movements that co-occur with stereotypical vocalizations in two juvenile Siamang engaged in solo singing. In fact, these stereotypical vocalizations always occurred with (some) movements (few of these vocalizations occurred with very little movement, and these are included in our data, see the left region of the ***dynamic plot*** to inspect the events with low body accelerations). Using unsupervised and

supervised computer vision and data science methods we approximated the magnitude of peaks in 3D acceleration/deceleration of the thorax, which we then related to the magnitude of the peak in the smoothed amplitude envelope nearest to the moment of peak acceleration. We observed that the more the Siamang's thorax undergoes acceleration, the higher the amplitude of the vocalization, i.e., the louder the call. We further find that body accelerations are associated in time with changes in loudness, as established through finding higher mutual information of z-normalized movement and vocalization timeseries as compared against false pairs.

We speculate that the non-overlapping and loud solo songs may serve an 'honest signaling' function towards prospective mates (see e.g., Raemaekers et al., 1984). We would suspect that if these high-amplitude solo calls are designed to inform others about one's adaptive fitness, all available bodily resources will likely be recruited. This may mean that body movements stabilize and amplify said calls through some biomechanical cooperation (i.e., a synergy), or it may simply mean that more intense movements are coordinated as a visual display – a visual honest signal – that is coordinated with vocalization (i.e., a systematic combination of two signals). The current study merely shows an association between singing and movement, and more research is needed to understand the mechanisms and possible adaptive benefits of mutual coordination of song and movement.

Our results have several implications. Firstly, vigorous movements during singing in the Siamang have generally been described as combined 'movement *displays*'. However, our results suggest that these behaviors should also be understood as a coordination, where whole-body movement and sound combined in a structural way [9,30,31]. This further has essential implications for the large field of primate gesture studies, specifically as it tailors to the need for a better understanding of how different modalities contribute to communicative signaling [54,55] - after all, it seems that Siamang signal by using their body and their vocalization in a synchronized way. This synchrony reminds us of how professional human singers couple their upper limb movements during vocalizations [8]. It may be mechanistically analogous to how other non-human animals couple their pectoral limb activity to vocalization [1,5–7,23]. Our findings further tie in with recent research suggesting that species with more arboreal locomotion repertoires also have increased vocal singing abilities [56].

Importantly however, coupling of the respiratory-vocal system with peripheral bodily movements is not necessarily something that might drive vocal flexibility over evolutionary time [5,57]. Several avian vocal learners tend to have a looser allometric scaling of their wingbeat duration with their vocalization duration, as compared to birds who are not vocal learners [5]. Mammalian vocal learners also tend to escape allometric scaling laws that relate vocalization and vocal tract size [58]. In humans, the increased flexibility in respiratory control has been in part attributed to a weakening of biomechanical constraint of locomotion with respiration (as the thorax was no longer impacted by locomotion; Bramble and Carrier, 1983). Furthermore, when mammals, including primates, need to do forceful activities that load onto the thorax, the glottis might need to be temporarily closed off to trap air to stiffen the thorax [6,9,27,59].

Thus, the possible functional interactions between locomotion and vocalization in primates is an open question. In avian species, for example, it has been shown that vocal development can be dependent on locomotor development [21,60]. Developmental research with marmoset monkeys has however shown that vocal development can occur even when locomotor development is delayed - Only an association of locomotor-postural development and vocal development was found (Gustison et al., 2019). Thus, more research is needed that not only assesses whether there is a link between locomotion and vocalization, but also how this then may play out over ontogenetic and evolutionary development.

There are several limitations to the current study. It is based on data from a single zoo obtained from captive juveniles, rather than wild adult Siamangs. It only considers certain vocalizations. These issues can only be resolved by collecting and analyzing more data and behaviors. Further, our measurements do not allow, at present, for more fine-grained 3D tracking of the animals as the cameras were impossible to calibrate at this zoo; the study setting also did not allow for perfect vocal amplitude measurements since that would require much more controlled parameters (e.g., the constant distance between source, the constant direction of radiation). We have reduced these issues through measuring from multiple directions and using signal processing techniques to obtain unique movement information from the cameras.

We should further highlight some of the hypothesized mechanisms that we put forward for vocal-motor interaction is on the level of biomechanics (kinematics) but our analyses and findings limited to kinematics, and accordingly more work is needed on the level of

biomechanics [1,22,62]. Such biomechanical work would need to evaluate how locomotion affects the surrounding respiratory system so that sub-glottal pressure increases due to some compression of the lungs. For example, activity of muscles that insert into the human rib cage (e.g., pectoralis major), measured via electromyography, predicts fluctuations in loudness in steady-state vocalizations during upper limb movements [11]. In Siamang similar but less invasive investigations could consist of measuring branch reaction forces during locomotion [22] and directly relating this to modulations in singing.

Conclusions

The potential coordination of singing with movement has fascinated gibbon researchers for a long time [17]. At the same time, current leading primatologists see the coordination of the voice and communicative movement as something that is not common in non-human primates [63]. The current report provides quantitative support that primate vocalization and whole-body movement is coordinating. Siamangs fluctuate their song amplitude together with body accelerations. They thus coordinate movement and singing in a way that is currently unknown. We hope therefore that this research further invites inquiry into the impressive movement-entangled vocalizations of Siamang and Gibbons.

References

1. Lancaster WC, Henson OW, Keating AW. Respiratory muscle activity in relation to vocalization in flying bats. J Exp Biol. 1995;198: 175–191.

2. Suthers RA, Thomas SP, Suthers BJ. Respiration, wing-beat and ultrasonic pulse emission in an echo-locating bat. J Exp Biol. 1972;56: 37–48.

3. Stidsholt L, Johnson M, Goerlitz HR, Madsen PT. Wild bats briefly decouple sound production from wingbeats to increase sensory flow during prey captures. iScience. 2021;24: 102896. doi:10.1016/j.isci.2021.102896

4. Koblitz JC, Stilz P, Schnitzler H-U. Source levels of echolocation signals vary in correlation with wingbeat cycle in landing big brown bats (Eptesicus fuscus). J Exp Biol. 2010;213: 3263–3268. doi:10.1242/jeb.045450

5. Berg KS, Delgado S, Mata-Betancourt A. Phylogenetic and kinematic constraints on avian flight signals. Proc R Soc B Biol Sci. 2019;286: 20191083. doi:10.1098/rspb.2019.1083

6. Blumberg M. Rodent ultrasonic short calls: locomotion, biomechanics, and communication. J Comp Psychol. 1992;106: 360–365. doi:10.1037/0735-7036.106.4.360

7. Cooper BG, Goller F. Multimodal signals: enhancement and constraint of song motor patterns by visual display. Science. 2004;303: 544–546. doi:10.1126/science.1091099

8. Pearson L, Pouw W. Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement. Ann N Y Acad Sci. 2022;n/a. doi:10.1111/nyas.14806

9. Pouw W, Fuchs S. Origins of vocal-entangled gesture. Neurosci Biobehav Rev. 2022;141: 104836. doi:10.1016/j.neubiorev.2022.104836

10. Pouw W, Paxton A, Harrison SJ, Dixon JA. Acoustic information about upper limb movement in voicing. Proc Natl Acad Sci. 2020 [cited 12 May 2020]. doi:10.1073/pnas.2004163117

11. Pouw W, Werner R, Burchardt L, Selen L. The human voice aligns with whole-body kinetics. bioRxiv; 2023. p. 2023.11.28.568991. doi:https://doi.org/10.1101/2023.11.28.568991

12. Serré H, Dohen M, Fuchs S, Gerber S, Rochet-Capellan A. Leg movements affect speech intensity. J Neurophysiol. 2022 [cited 23 Sep 2022]. doi:10.1152/jn.00282.2022

13. Koda H, Nishimura T, Tokuda IT, Oyakawa C, Nihonmatsu T, Masataka N. Soprano singing in gibbons. Am J Phys Anthropol. 2012;149: 347–355. doi:10.1002/ajpa.22124

14. McAngus Todd NP, Merker B. Siamang gibbons exceed the saccular threshold: Intensity of the song of *Hylobates syndactylus*. J Acoust Soc Am. 2004;115: 3077–3080. doi:10.1121/1.1736273

15. Raimondi T, Di Panfilo G, Pasquali M, Zarantonello M, Favaro L, Savini T, et al. Isochrony and rhythmic interaction in ape duetting. Proc R Soc B Biol Sci. 2023;290: 20222244. doi:10.1098/rspb.2022.2244

16. D'Agostino J, Spehar S, Abdullah A, Clink DJ. Evidence for Vocal Flexibility in Wild Siamang (Symphalangus syndactylus) Ululating Scream Phrases. Int J Primatol. 2023 [cited 20 Sep 2023]. doi:10.1007/s10764-023-00384-5

17. Haimoff EH. Video Analysis of Siamang (Hylobates syndactylus) Songs. Behaviour. 1981;76: 128–151.

18. Mitani JC. Gibbon Song Duets and Intergroup Spacing. Behaviour. 1985;92: 59–96.

19. Redmond J, Lamperez A. Leading limb preference during brachiation in the gibbon family member, Hylobates syndactylus (siamangs): A study of the effects of singing on lateralisation. Laterality. 2004;9: 381–396. doi:10.1080/13576500342000211

20. Fan P-F, Ma C-Y, Garber PA, Zhang W, Fei H-L, Xiao W. Rhythmic displays of female gibbons offer insight into the origin of dance. Sci Rep. 2016;6: 34606. doi:10.1038/srep34606

21. Berg KS, Beissinger S, Bradbury J. Factors shaping the ontogeny of vocal signals in a wild parrot. J Exp Biol. 2013;216: 338–345. doi:10.1242/jeb.073502

22. Cheyne SM. Gibbon Locomotion Research in the Field: Problems, Possibilities, and Benefits for Conservation. In: D'Août K, Vereecke EE, editors. Primate Locomotion: Linking Field and Laboratory Research. New York, NY: Springer; 2011. pp. 201–213. doi:10.1007/978-1-4419-1420-0_11

23. Boggs DF. Interactions between locomotion and ventilation in tetrapods. Comp Biochem Physiol A Mol Integr Physiol. 2002;133: 269–288. doi:10.1016/S1095-6433(02)00160-5

24. Daley MA, Bramble DM, Carrier DR. Impact loading and locomotor-respiratory coordination significantly influence breathing dynamics in running humans. PLOS ONE. 2013;8: e70752. doi:10.1371/journal.pone.0070752

25. Lafortuna CL, Reinach E, Saibene F. The effects of locomotor-respiratory coupling on the pattern of breathing in horses. J Physiol. 1996;492 (Pt 2): 587–596. doi:10.1113/jphysiol.1996.sp021331

26. Harrison DFN. The anatomy and physiology of the mamalian larynx. Cambridge: Cambridge University Press; 1995.

27. Hayama S. The origin of the completely closed glottis. Why does not the monkey fall from a tree? Primate Res. 1996;12: 179–206. doi:10.2354/psj.12.179

28. Mott F. A Study by Serial Sections of the Structure of the Larynx of Hylobates syndactylus (Siamang Gibbon). Proc Zool Soc Lond. 1924;94: 1161–1170. doi:10.1111/j.1096-3642.1924.tb03336.x

29. Negus VE. The comparative anatomy and physiology of the larynx. London: William Heinemann Medical Books; 1949.

30. Partan SR, Marler P. Communication Goes Multimodal. Science. 1999;283: 1272–1273. doi:10.1126/science.283.5406.1272

31. Halfwerk W, Varkevisser J, Simon R, Mendoza E, Scharff C, Riebel K. Toward Testing for Multimodal Perception of Mating Signals. Front Ecol Evol. 2019;7. doi:10.3389/fevo.2019.00124

32. Geissmann T. Duet Songs of the Siamang, Hylobates Syndactylus: II. Testing the Pair-Bonding Hypothesis during a Partner Exchange. Behaviour. 1999;136: 1005–1039.

33. Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H. ELAN: a Professional Framework for Multimodality Research. 2006; 4.

34. Hunt KD. Why are there apes? Evidence for the co-evolution of ape and monkey ecomorphology. J Anat. 2016;228: 630–685. doi:10.1111/joa.12454

35. Hunt KD, Cant JGH, Gebo DL, Rose MD, Walker SE, Youlatos D. Standardized descriptions of primate locomotor and postural modes. Primates. 1996;37: 363–387. doi:10.1007/BF02381373

36. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci. 2018;21: 1281–1289. doi:10.1038/s41593-018-0209-y

37. Winter DA. Biomechanics and Motor Control of Human Movement. John Wiley & Sons; 2009.

38. Karashchuk P, Rupp KL, Dickinson ES, Walling-Bell S, Sanders E, Azim E, et al. Anipose: A toolkit for robust markerless 3D pose estimation. Cell Rep. 2021;36: 109730. doi:10.1016/j.celrep.2021.109730

39. Theriault DH, Fuller NW, Jackson BE, Bluhm E, Evangelista D, Wu Z, et al. A protocol and calibration method for accurate multi-camera field videography. J Exp Biol. 2014; jeb.100529. doi:10.1242/jeb.100529

40. Mena-Chalco JP, Macêdo I, Velho L, Cesar RM. 3D face computational photography using PCA spaces. Vis Comput. 2009;25: 899–909. doi:10.1007/s00371-009-0373-x

41. He L, Dellwo V. Amplitude envelope kinematics of speech: Parameter extraction and applications. J Acoust Soc Am. 2017;141: 3582–3582. doi:10.1121/1.4987638

42. Zeileis A, Grothendieck G. zoo: S3 Infrastructure for Regular and Irregular Time Series. J Stat Softw. 2005;14: 1–27. doi:10.18637/jss.v014.i06

43. Pardy C. mpmi: Mixed-pair mutual information estimators version 0.4 from R-Forge. 2012. Available: https://rdrr.io/rforge/mpmi/

44. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Team RC. nlme: Linear and nonlinear mixed effects models. 2019.

45. Pouw W, Proksch S, Drijvers L, Gamba M, Holler J, Kello C, et al. Multilevel rhythms in multimodal communication. Philos Trans R Soc B Biol Sci. 2021. doi:10.1098/rstb.2020.0334

46. Deacon TW. The symbolic species: The co-evolution of language and the brain. W.W. Norton; 1998.

47. Deacon TW. Incomplete Nature: How Mind Emerged from Matter. W.W. Norton; 2013.

48. MacLarnon AM, Hewitt GP. The evolution of human speech: the role of enhanced breathing control. Am J Phys Anthropol. 1999;109: 341–363. doi:10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2

49. MacNeilage PF. The origin of speech. Oxford University Press; 2010.

50. Burchardt LS, Sande Y van de, Kehy M, Gamba M, Ravignani A, Pouw W. 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. 2023 [cited 15 Oct 2023]. Available: https://ecoevorxiv.org/repository/view/6080/

51. de Boer B. Acoustic analysis of primate air sacs and their effect on vocalization. J Acoust Soc Am. 2009;126: 3329–3343. doi:10.1121/1.3257544

52. Dunn JC. Sexual selection and the loss of laryngeal air sacs during the evolution of speech. Anthropol Sci. 2018;126: 29–34. doi:10.1537/ase.180309

53. Raemaekers JJ, Raemaekers PM, Haimoff EH. Loud Calls of the Gibbon (Hylobates lar): Repertoire, Organisation and Context. Behaviour. 1984;91: 146–189.

54. Liebal K, Slocombe KE, Waller BM. The language void 10 years on: multimodal primate communication research is still uncommon. Ethol Ecol Evol. 2022;0: 1–14. doi:10.1080/03949370.2021.2015453

55. Slocombe KE, Waller BM, Liebal K. The language void: the need for multimodality in primate communication research. Anim Behav. 2011;81: 919–924. doi:10.1016/j.anbehav.2011.02.002

56. Schruth DM, Templeton CN, Holman DJ, Smith EA. Evolution of primate protomusicality via locomotion. bioRxiv; 2021. p. 2020.12.29.424766. doi:10.1101/2020.12.29.424766

57. Bramble D, Carrier DR. Running and breathing in mammals. Science. 1983;219: 251–256. doi:10.1126/science.6849136

58. Garcia M, Ravignani A. Acoustic allometry and vocal learning in mammals. Biol Lett. 2020;16: 20200081. doi:10.1098/rsbl.2020.0081

59. Orlikoff RF. Voice Production during a Weightlifting and Support Task. Folia Phoniatr Logop. 2008;60: 188–194. doi:10.1159/000128277

60. Liu W-C, Landstrom M, Cealie M, MacKillop I. A juvenile locomotor program promotes vocal learning in zebra finches. Commun Biol. 2022;5: 573. doi:10.1038/s42003-022-03533-3

61. Gustison ML, Borjon JI, Takahashi DY, Ghazanfar AA. Vocal and locomotor coordination develops in association with the autonomic nervous system. Tchernichovski O, Calabrese RL, Goller F, editors. eLife. 2019;8: e41853. doi:10.7554/eLife.41853

62. Bertram JEA. New perspectives on brachiation mechanics. Am J Phys Anthropol. 2004;39: 100–117. doi:10.1002/ajpa.20156

63. Fröhlich M, Sievers C, Townsend SW, Gruber T, Schaik CP. Multimodal communication and language origins: integrating gestures and vocalizations. Biol Rev. 2019;94: 1809–1829. doi:10.1111/brv.12535

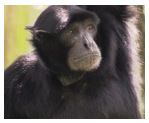
Supplemental materials

Table S1. Links to online materials

Reference	Information	Hyperlink
Example 1	Example of data of small amplitude movements	link
Example 2	Example of how stereotypical calls analyzed in this research are embedded in longer singing sequences	link
Example 3	Example ricochetal brachiation	link
Example 4	Example of video tracking DeepLaBcut	link
Script 1	Python code for snipping videos from annotations	link
Script 2	Python OpenCv2 automatic bounding box pre-processing script	link
Script 3	R Processing script	link
Script 4	R Statistical analysis	link
Script 5	Python reproducible code for data dashboard	link
Resource 1	Deeplabcut trained model + info	link

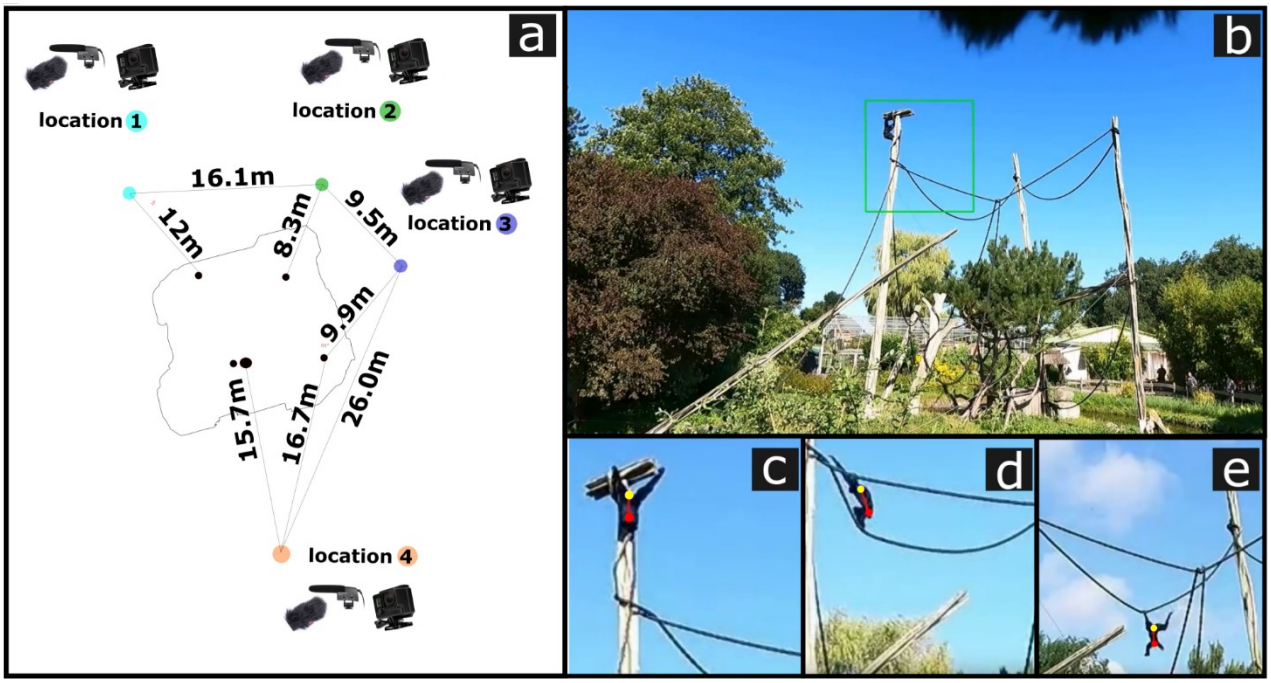
Table S2. Descriptives stats main variables

	Overall	**Baju**	**Fajar**	**Forelimb only**	**Other**
	M (*SD*) 95%CI[]	*M* (*SD*) 95%CI[]	*M* (*SD*) 95%CI[]	*M* (*SD*) 95%CI[]	*M* (*SD*) 95%CI[]

					
Time annotation	2851 ms (491) [2744, 2959]	2495 (268) [2386, 2603]	3014 (485) [2886, 3143]	2815 (485) [2675, 2954]	2905 (504) [2729,3080]
Nearest peak envelope (*z*)		0.52 (0.82) [0.19, 0.85]	-0.24 (.99) [-0.50, 0.03]	0.07 (0.91) [-0.19, 0.33]	-0.10 (1.12) [-0.49, 0.29]
Peak acceleration (*z*)		0.29 (0.69) [0.01, 0.57]	-0.13 (1.09) [-0.42, 0.16]	0.15 (1.09) [-0.21, 0.46]	-0.21 (0.83) [-0.50, 0.08]
Inter-peak distance	121 ms (160) [86, 156]	157ms (165) [90, 224]	104 (156) [62, 146]	107 (154) [62, 151]	141 (169) [83, 200]

Note. Max nearest envelope is the z-scaled magnitude of log peak smoothed amplitude envelope. Note that there are no overall descriptives for variables that have been z-scaled (amounting to $M = 0$, sd = 1).

Figure S1. Keypoints tracked for different frames



Note. Panel a) shows a sketch of the location with four different recording sites with camera and microphones and estimated distances obtained via laser-based distance estimation. b) Shows a frame from location 4 and an automatic selection of the frame denoted by the green rectangle. c-d) These subframes containing movement in the frame were then used for training DLC model to detect the two keypoints, shown in yellow (upper thorax, at T1) and red (sacrum). The Euclidean distance of the upper thorax to the sacrum was used to normalize kinematics expressed in pixel units to units relative to body size.