Università degli Studi di Torino

Tesi di Dottorato di Ricerca in *Scienze Biologiche e Biotecnologie Applicate* – XXXV ciclo

PhD Thesis in *Biological Sciences and Applied Biotechnologies* – XXXV cycle

Rhythmic categories in primate vocalizations

Teresa Raimondi

Tutor: Prof. Marco Gamba

At liquidas avium voces imitarier ore
ante fuit multo quam levia carmina cantu
concelebrare homines possent aurisque iuvare.

To imitate the liquid notes of birds
was earlier far amongst men than power to make,
by measured song, melodious verse and give delight to ears.

Lucrezio
De rerum natura,
book V, verses 1379-1381

CONTENTS

ABSTRACT

In the reconstruction of the evolutionary history of musicality, two complementary approaches can come to our aid: on the one hand, the identification of musical universals, i.e., those traits that are shared by all human cultures; on the other hand, the search for such universals in other species, from a comparative perspective. One of the universals of human musicality is represented by rhythmic categories, i.e., the fact that rhythmic patterns show temporal intervals that are not randomly distributed but are clustered around specific relationships between durations. In particular, human music exhibits ratios between temporal intervals that mathematically simple ratios can describe, the so-called *small-integer ratios*. The simplest form of relationship is *isochrony*, the rhythmic structure characterized by intervals of equal duration (1:1). Isochrony is the primary rhythmic grid on which all musical systems worldwide are based. In humans, the presence of such regularity in music is part of a compound of multifaceted rhythmic capacities that are known to derive from several parallel but interconnected functional processes: the production of oscillatory patterns, the perception and production of the beat, the synchronization to that beat, called entrainment, the generalization of patterns into a hierarchical system, or metre. Those processes rely on composite, partially co-opted from other pathways' neural substrates. Whether similar processes and pathways are shared with other animal species is debated.

This study aims at investigating the presence of categorical rhythms, and in particular of small-integer-ratios, in the vocalizations of six primate species, linked by a decreasing gradient of phylogenetic proximity to humans: the Bornean orangutan (*Pongo pygmaeus wurmbii*), the lar gibbon (*Hylobates lar*), the yellow-cheeked gibbon (*Nomascus gabriellae*), the white-cheeked gibbon (*Nomascus leucogenys*), the southern white-cheeked gibbon (*Nomascus siki*), and the indri (*Indri indri*). These species share certain aspects of signal characteristics and functions but, at the same time, differ in crucial eco-ethological features. The identification of possible rhythmic categories in these species may allow the reconstruction of the evolutionary history of this trait and selectively identify the selective pressures that made this trait adaptive.

Evidence from this work highlights how all the analysed species show isochrony. Furthermore, we highlight how exogenous stimuli, such as synchrony and vocal interaction, affect a signal's rhythmic structure. This is crucial as the interplay and integration of

endogenous and exogenous stimuli in determining rhythmic behaviours have been thought to be a human prerogative for a long time. At the same time, we show how vocal interaction is not necessary for the evolution of isochrony, as vocalizations emitted in solo behavioural contexts also show isochrony. Interestingly, the only species to show further small-integer-ratios besides isochrony is the Indri, which is the furthest phylogenetically from humans among those analysed.

The results are discussed from a perspective that takes into account the selective pressures acting on each vocal type and species analysed and then attempts to deduce overall evolutionary trends and pressures that are generalizable to the primate clade, including humans. The reconstruction of this trait from a comparative perspective allows us to advance evolutionary hypotheses on rhythm, shedding light on how rhythmic capacities, a feature thought to be exclusively human, actually show a gradient of declinations within the primate clade.

GENERAL INTRODUCTION

The exchange of information among individuals responds to a number of vital needs, such as reproduction, feeding, resource defence, and escape from danger (Hebets et al., 2016). This process is referred to, in a very general definition, as *communication*. Communication among conspecifics or heterospecifics takes place over many components of the same (unimodal) or different (multimodal) sensory modalities (Hebets et al., 2016): acoustic, visual, olfactory and tactile. Each communication *signal*, a specially evolved message that contains information meant to affect the behaviour and fitness of one or both involved individuals (Alcock & Rubenstein, 2019), is the result of many coexisting selective pressures. A communication signal is, for example, the result of the interaction between its propagation and the environment in which the signal evolves, in a process known as *sensory drive* (Alcock & Rubenstein, 2019). Also, the function of a signal, its urgency, and the trade-off between energy expense and informativeness of the signal itself are some of the pressures shaping the structure of a signal.

The production of *sounds*, defined as radiating disturbances in the pressure of a medium, is widespread in the animal kingdom (Bradbury & Vehrencamp, 2011). An acoustic signal results, first of all, from the production of the sound source through the vibration of a designated structure (e.g., vocal folds in mammals; Fant, 1971). This sound source then undergoes filtering to maximise the resonance (e.g., the vocal tract in humans comprising the pharynx, nasal and buccal cavities; Fant, 1971). An acoustic signal can be either vocal, when produced through the respiratory system of the emitter, or mechanical when air vibration is induced by the movements of a body organ or structure (e.g., stridulation in Orthoptera). Temporal and frequency features of a sound are the evolutionary result of the interaction of the acoustic signal with the environment it evolved in, i.e., the most efficient the propagation, the higher the adaptiveness of the signal (Alcock & Rubenstein, 2019). Temporal and frequency features of the sound can also convey information such as sex, age, individuality, and physical condition of the emitter (Alcock & Rubenstein, 2019). In this sense, since acoustic signals are often constrained by the body size, by the anatomy of the vocal tract or the vibrational organ that modulates the sound, and are expensive to produce, they are considered an honest signal of the sender's quality (Alcock & Rubenstein, 2019; Reby & Mccomb, 2003).

Of the various modes of communication, the acoustic mode is particularly adaptive in contexts where it is crucial to propagate the signal rapidly, in both water and air, even over great distances (Bradbury & Vehrencamp, 2011). Acoustic communication allows the fast conveyance of urgent information and the energy savings from the possibility of communicating over great distances without moving (Bradbury & Vehrencamp, 2011). Furthermore, acoustic communication is successful in environments with limited visibility (e.g., due to the density of vegetation or poor lighting; Bradbury & Vehrencamp, 2011). For all these reasons, acoustic communication is extended to an enormous number of taxa with stunningly diverse sound production and modulation characteristics (Alcock & Rubenstein, 2019; Bradbury & Vehrencamp, 2011).

What is rhythm?

All communicative signals are articulated on space and time and decline through different, specialized, and not-mutually-exclusive sensory channels (*e.g.,* visual, olfactory, acoustic; Pouw et al., 2020). All communicative signals thus share their organization on time as an unavoidable dimension. Vocal signals, for instance, take shape on two main features, both subjected to selective pressures and modification during evolution (Patel, 2010): frequency and time. The second determines the temporal structure of the signal's organization, which we call *rhythm*. Rhythm, whether periodic (i.e., if it has a specific pattern that repeats itself at a regular interval of time) or aperiodic, is defined, in the most inclusive sense possible, as a *serial pattern of durations marked by a series of events* (Mcauley, 2010).

When talking about rhythm, many authors refer to its periodicity as its antonomastic form. Neural oscillation, heartbeat, and locomotion, are common-to-all, physiological mechanisms that show a regular meter, like the beat of a metronome: when time intervals are equal to each other within a display, we speak of *isochrony*. Physiological periodicity is thus widespread in other species, and it is not a human prerogative, but, as claimed by Fitch (Fitch, 2012), a paradox, therefore, seems to be evident: if periodicity appears somehow as a primary mechanism, why the human capacity to produce and perceive rhythm, and to engage the musical motor output to entrain to others, like in collective, coordinated dancing and singing, is an exception in the animal kingdom? The synthesis of this paradox must be sought in a bio-musical approach, and in particular in the representation of musicality (in our species, intended as the *set of capacities and proclivities that allows us to generate and* *enjoy music in all of its diverse forms*: music itself is only one of the behavioural outputs of musicality; W. T. Fitch, 2015) in the light of the following considerations:

- musicality is composite in terms of general functional processes, meaning that rhythm cognition depends on separate, compound functional sub-mechanisms. Rhythm is the result of at least three independent processes (**Figure 1**; Tecumseh Fitch, 2012):

    1) *periodic motor pattern generation*, which allows one individual to produce a regular rhythmic pattern

    2) *beat* extraction from complex rhythmic patterns, that is, the ability to generalize a temporal pattern and predict it

    3) *entrainment*, or the capacity of one individual to coordinate with another, based on the inferred beat (Fitch, 1988).

    4) *meter perception*, namely the operation of arranging recursive patterns into hierarchical structures (similar to what happens in speech; Kotz et al., 2018)

- musicality is composite in terms of neural substrates that make rhythm possible (Kotz et al., 2018): multiple modules overlap with each other and allow other functions besides musicality. The pathways that make rhythm possible are not monolithic and are often co-opted by other functions. The areas deputed to rhythm perception in the human encephalon are known to be a complex network of cortico-subcortical areas (Kotz et al., 2018). This network largely corresponds to the structure deputed to the perception and production of the temporal structure of auditory signals in general (Grahn, 2012).

Rhythm is thus the result of a gradual miscellany of anatomical and functional changes throughout evolution. From this perspective, we can consider rhythm in the light of the specific selective pressures acting on a given behavioural trait or structure from a composite and systemic perspective rather than in terms of the presence or absence of striking characteristics that we observe in our species (Merchant & Honing, 2014).

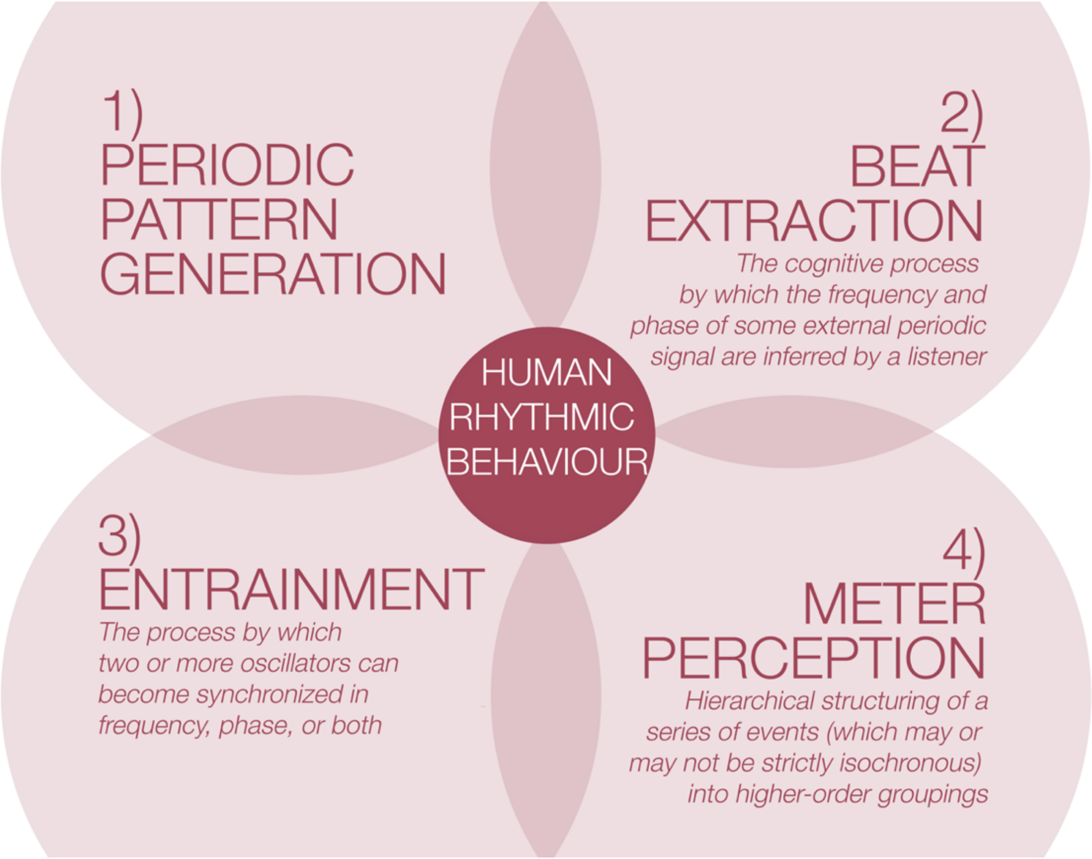


Figure 1 - Schematization of the view (Kotz et al., 2018) that human rhythmic behavior consists of at least four main, independent but interrelated mechanisms: 1) the rhythmic pattern production, 2) the extraction of the beat, or the generalization of the recognized rhythmic pattern, 3) the entrainment to the extracted beat, 4) the meter perception, or the capacity to abstractly represent the hierarchical structure of the signal (strong vs weak elements), not necessarily thought time.

Categorical rhythms and small integer ratios

Periodic rhythms are characterized by repeating a specific pattern of events through time. The relationships that exist among two temporal intervals (as inter-onset-interval, or t_k, namely the duration that separates two successive intervals' onsets; **Figure 2;** Ravignani & Madison, 2017) of a signal can be described by the ratio between them. The relationship linking them can be random, but when it's not, ratios are not continuously but categorically distributed: in other words, some ratios happen more frequently than others and determine the presence of *categorical rhythms*. What is observed in all musical systems around the world (Savage et al., 2015), is the production of *small-integer ratios*, specific categorical rhythms in which very simple mathematical ratios describe the relationships between temporal intervals (e.g., 1:2, 2:1, 1:3, 3:1, and so on). The most straightforward relationship possible between two intervals is the 1:1 ratio, or *isochrony*: all intervals have equal durations, like the tickling of a clock. In humans, the study of the role of time in musicality is strongly tied to a view whereby time itself is both a physical and objective dimension and a subjective structuring of events (Michon & Jackson, 1985). Humans tend to partition rhythmic patterns into small-integer-ratios, a process subject to factors such as the metric context in which the pattern is perceived (Desain & Honing, 2003).

To sum up, two paramount rhythmic universals, that is, features that are found in all human cultures, of music are (Savage et al., 2015):

- the presence of an isochronous grid as a reference
- a variation of the isochronous grid in terms of small integer ratios, specific rhythmic categories in which time intervals are linked by very simple mathematical relationships

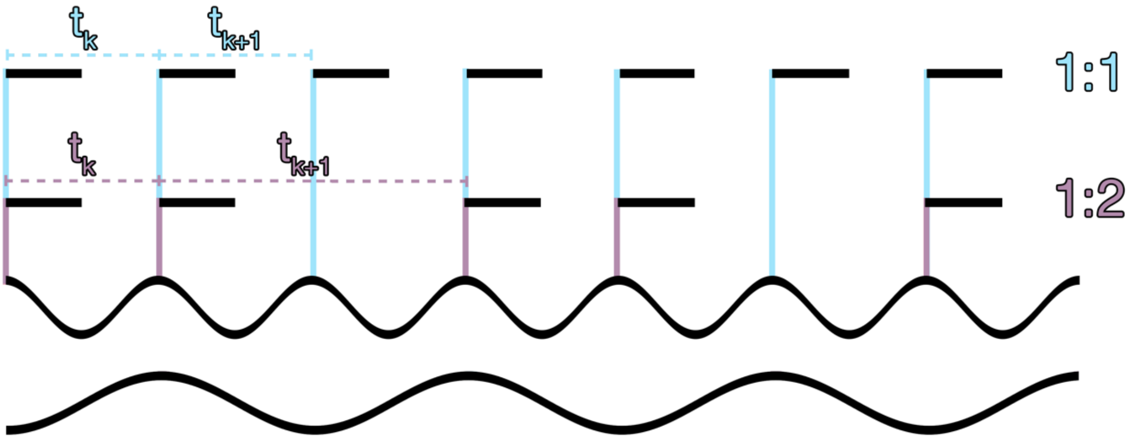


Figure 2 – Schematic representation of an isochronous pattern (up) and of a pattern corresponding to a 1:2-2:1 ratio (bottom). The first two inter-onset intervals of the two patterns (t_k and t_{k+1}) are schematized and determine, in particular, a 1:1 ratio (up) and a 1:2 ratio (bottom). The onsets of each pattern are indicated with vertical solid lines (light blue for the 1:1 pattern and purple for the 1:2-2:1 pattern). The two sine waves on the bottom display the oscillatory processes might generate the rhythmic categories, 1:1 and 1:2 ratios.

Rhythm in human communication: speech and music

Music and speech are two domains, present in all human cultures, that are greatly articulated on the temporal dimension (Kohler, 2009). All spoken languages are intrinsically built on a rhythmic system. Traditionally, authors reported the different elements of a language to recur at regular intervals, establishing the temporal organization; Ramus et al., 1999): spoken languages can be syllable-timed (e.g., Italian, more isochronous), mora-timed (e.g., Japanese), stress-timed (e.g., English, les isochronous). However, some authors have proposed that speech fails to show a consistent isochrony when strictly and quantitatively studied using musical metrics (Kotz et al., 2018). Rhythmic regularity perception in speech depends on the subjectivity of the listener, on the speaker's identity and utterances variability (Brown & Weishaar, 2010): language would be more of a gradient of intermittent quasi-periodic patterns located into an overall, heterometric system. The role of rhythm in speech is thus not strictly linked to the regularity of the pattern but is involved in facilitating speech segmentation and comprehension (Burchardt et al., 2019; Kotz et al., 2018): phonated and non-phonated events alternation, mouth movements (in the domain of a multi-modal communication), have been proposed to be ancestrally crucial for spoken communication, as they facilitate processing and speech comprehension, coupling to neurocognitive oscillators frequency (Port, 2003).

In parallel to language, all human cultures have developed a musical domain. Although in different forms and also on cultural influence, as Peretz expressed (2006), *music appears to transcend time, place, and culture*. Unlike language, however, the study of music from a biological perspective is rather recent, since its function in the range of human ethology is less obvious (Peretz, 2006; Hauser & McDermott, 2003). Unlike speech, regular rhythm is a crucial component of all human music (with few exceptions, see Kotz et al., 2018). All musical systems are built on a ground of a regular temporal pattern with variations corresponding to straightforward mathematical relationships related to temporal intervals (the so-called small integer ratios).

Music and language evolution: a shared biological root?

Music and language, in their specific form, are human prerogatives (Patel, 2010). A temporal dimension is a mesh on which both domains are articulated, with the rhythmic connotations of the two domains that are remarkably different: if music is articulated on an isochronous system, speech is quasi-periodic (Kotz et al., 2018). The shared features mainly concern general processing mechanisms (the capacity to form learned sound categories, extract statistical regularities from melodic and rhythmic sequences, and organise incoming elements into syntactic structures - Patel, 2010). The organisational aspect closely links music and language as they hierarchically group elements in a temporal grid (Fitch & Martins, 2014). Also, evidence on metrical perception suggests that listeners perceive and entrain to accented events through similar mechanisms in music and speech (Kotz et al., 2018).

Beginning with what Darwin called *musical protolanguage* in 1871, many theories link these two expressions of human communication, on the strength of the many shared characteristics. Those who theorize that music and language share homologous origins (Darwin, 1872; Brown, 2001) hypothesize that the current manifestation of language and music is the result of an evolutionary process that, at least in part, was based on a common substrate. Those modules would have partly maintained and shared the same areas for the same functions in the two systems, and on the other hand integrated their functioning and specialized additional parallel circuits divergently in the two systems.

Whether the areas activated by both systems have a common origin or were co-opted by one of the two systems at a later time, in any case, what we can observe today in the human encephalon is a strong overlap in terms of activations of the various encephalic regions in the two domains, both at the cortical and subcortical levels (Patel, 2010). The neural substrates for vocalizing (Perry et al., 1999; Nakada et al., 1998), reading and comprehension of language and music (Nakada et al., 1998), and music execution and listening (Patel, 2003), show a high degree of overlap, for instance in the inferior frontal gyrus and the ventral premotor cortex (Fadiga et al., 2009). This connection is corroborated by the fact that musical expertise facilitates, among others, language learning (Francois & Schön, 2011). This is related to the observation that music and language rely on similar syntactic organisation beyond the primary features shared between the two domains (e.g., affective prosody, vocalization - Brown, 2001), music and language rely on similar syntactic organization.

During the past decades, music and speech processing has often been described as dichotomous: music on the right hemisphere, speech on the left. This view is largely outdated, giving way to a modular representation of the organization of music cognition (Altenmüller, 2001). Different aspects of music processing involve different areas of both hemispheres that partially overlap with each other and with those of language. Moreover, these modules are not shaped in a fixed manner but flexibly transformed based on the individual's life learning experience (Altenmüller, 2001). The modules involved in rhythm processing and production are cortico-subcortical and form a network that broadly corresponds to the structure deputed to the perception and production of the temporal structure of auditory signals in general (Schwartze & Kotz, 2013). The modules needed for synchronization to music correspond to the brain regions traditionally associated with motor output and production (Kotz et al., 2018). A human characteristic in the process of entrainment between body movement and an exogenous rhythmic stimulus is that humans are more inclined to entrain to auditory than visual stimuli (Gámez et al., 2018; Repp & Penel, 2004). Research on metre aspects of rhythm perception is in its infancy, but interestingly evidence suggests that musical training might affect where and at which degree musical rhythms are processed and hierarchically categorized at an abstract level.

Rhythm in non-human animals

As humans, many species show physiological, neurological, and locomotion outcomes that follow an isochronous pattern (Ravignani & Madison, 2017): endogenous mechanisms that allow the production of such a rhythmic ability are possibly widespread among the animal kingdom (Ravignani & Madison, 2017). Spontaneous rhythmic behaviours seem widespread in non-human species in general. Even in our closest relatives, the great apes, the production of regular rhythmic patterns is common (e.g., buttress drumming in chimpanzees, Arcadi et al., 1998; speech-like rhythm in *clicks* and *faux-speech* of orangutans, Lameira et al., 2015), but their regularity has rarely been quantified. Recently, spontaneous production of isochronous rhythms has been proven in several, not-closely-related to humans species: a bat, *Saccopteryx bilineata* (Burchardt et al., 2019), rock hyraxes, *Procavia capensis*, (Demartsev et al., 2022), Australian pied butcherbird, *Cracticus nigrogularis*, (Xing et al., 2022), and thrush nightingales, *Luscinia luscinia* (Roeske et al., 2020).

Beat perception and entrainment, intended as the ability to perceive a beat in music and synchronize bodily movement with it have been observed in some species other than humans Geissmann, 1999). The entrainment of an individual to artificial stimuli appears scarcely widespread in non-human species, even with trained animals (Schachner et al., 2009), but the capacity of animals to entrain to conspecifics' vocalizations seems to be sporadic but more common in vertebrates and invertebrates; Merker et al., 2009). Mammals such as dogs, horses or non-human primates fail to efficiently entrain to a stimulus, with or without training. Non-human primates (e.g., macaques) performance, in particular, is comparable to humans in simple tasks (e.g., interval reproduction) but shows differences in more complex tasks such as rhythmic entrainment and synchronization (Merchant & Honing, 2014). Another crucial difference is that humans easily synchronize in the audiomotor domain, while other primates (e.g., macaques) can synchronize in the visual domain, but show less sensitivity in the auditory one (Merchant & Honing, 2014). Crucially, vocal mimics, such as parrots or elephants, display a form of entrainment (Schachner et al., 2009).

The interplay between rhythm and song organization is critical to understand whether meter perception is present in non-human animals. Non-human animals produce regular rhythmic patterns and show hierarchical structures in their signals (Zuberbühler, 2018; Liao et al., 2022), but how these two key features enhance each other? The approach of combining rhythm production and meter perception is rare in literature, but slight evidence for rhythm and syntax interaction exists, at least in birds (Xing et al., 2022).

The evolution of rhythm

It has been proposed that human musical rhythm incorporates primitive primate biological elements to typically-human features that evolved more recently, shaping musical behaviour in our species (Fitch, 2012; Kotz et al., 2018). Circuitry engaged in human communication, when looking at the general cytoarchitecture and shared non-linguistic functional properties, shows homologues regions in non-human primate brains (e.g., Broca's and Wernicke's area in macaque's brain; Rilling, 2014; Corballis, 2009; Wilson & Petkov, 2011). Primates share the basic brain cytoarchitecture involved in communication, which, in humans, seem to have evolved specific musical and linguistic modules that gradually diverged from other species.

Turning to the observation of rhythmic skills from a comparative perspective, from one side, we have seen that there is fragmented evidence that non-human animals spontaneously produce isochronous patterns and that great apes' facial displays happen at frequency ranges that are compatible with human speech. We also know that beat extraction and entrainment are rare, and mainly confined to the visual domain, even in trained mammals (dogs and horses). On the other hand, entrainment is shown by distantly-related-to-humans taxa, as birds, arguably for convergent evolution. It's been proposed that non-human primates cannot rhythmically entrain in the way humans do because of the absence of a strong coupling between the auditory and motor system of non-human primates, which would have appeared in the hominoid evolutionary branch at a later evolutionary stage (Gámez et al., 2018; Merchant & Honing, 2014; Repp & Penel, 2004). It was also hypothesized that the typically human entrainment mechanism is the result of convergent evolution that has led vocal learners to evolve this capacity: the neural regions that support musical beat perception in humans (basal ganglia) would in fact be the same as those that are associated with vocal mimicry in birds (Doupe et al., 2005).

On the one hand, we observe the sharing of some composite neural modules, deputed to the perception and production of rhythm, that are potentially homologous; on the other hand, we detect that the identification of rhythmic skills in the different taxa more or less related to humans go in the direction of the affirmation of at least four separate systems in rhythm processing that are shaped independently through evolution: periodic pattern generation, beat perception, entrainment and meter perception. Unfortunately, evidence from a comparative perspective on these four aspects is currently patchy, not comparable in methodological terms, and rarely observed in light of the selective pressures that may have acted differentially in the various taxa.

Are music and rhythm adaptive?

The biological history and adaptive function of music and rhythm are unclear: many explanations have been theorised over time (Hauser & McDermott, 2003). Hagen and Bryant suggest that music and dance evolved for group-oriented purposes in the advertisement and coordination of coalitions (Hagen & Bryant, 2003). Coordination is made possible by a predictable and generalizable pattern (beat evolution), so that multiple individuals can coordinate with each other (entrainment). Many suggest that isochrony originates precisely from this need to coordinate not only one's vocal displays with motor displays (singing and dancing, present in all human cultures), but also one individual with others (Bowling et al., 2013; Merker et al., 2009). This need may have served our hominoid ancestors to signal the presence of resources in a given place, crucially to migrating females for mating (Merker et al., 2009). The ability to predict a rhythmic pattern and coordinate with it, in fact, allows group members to sum up the signal so that it is transmitted more effectively in the environment. Savage et al. (2015) also suggest that music has group coordination and cohesion functions and recognize the universal tendency to sing, play percussion instruments, and dance to simple, repetitive music in groups as unequivocal support for this hypothesis. To this, Fitch (2012) adds that isochrony is one of the ways, the simplest one, to make a signal predictable and allows beat perception and entrainment, but that this does not mean that the intrinsic biological basis of rhythm is to be found exclusively in isochrony, or that at the very least isochrony is not biologically necessary, per se, for the evolution of the various modules that, as we have seen, constitute the system of rhythmic capacities that denote human and, plausibly, animal behaviour in general.

Darwin first argued that music would be primarily a sexually selected system that originally served to broadcast the emitter's quality for mate attraction (Darwin, 1872). Since the inability to produce and perceive rhythmic regularity has been shown to underlie, at times, a developmental or genetic impairment that potentially impacts the fitness of the subject (Van Den Broek & Todd, 2003), rhythm may signal the quality of the emitter. Indeed, there are several species in which signal rhythmic regularity has been found to mediate the conveyance of transmitter quality and is subject to sexual selection (Van Den Broek & Todd, 2009). There are established cases of human cultures in which musical performance is also subject to sexual selection (Smith & Bird, 2000), but other studies failed to prove a connection between music abilities and mating success (Mosing et al., 2015).

Still, others, including Pinker (Pinker, 1997), argue that music would be a fortuitous by-product of perceptual and cognitive mechanisms deputed to other functions: music would not, therefore, have an adaptive function per se. This assessment, however, is only partially convincing, as the maintenance of energetically costly displays such as dancing and singing is difficult to justify without the presence of an advantage (McAndrew, 2019) that is even a boost to social cooperation and bonding, alliances maintenance and advertisement or personal features displaying.

Coming to the adaptive potential intrinsic to an isochronous signal, isochrony represents the simplest possible rhythmic structure in a signal and the one with the lowest entropy. At the same time, isochrony is intertwined with redundancy, making the risk of information loss during the communicative process low: information repetition is particularly adaptive in noisy environments and long-distance communication (Favaro et al., 2020). There are numerous cases, even in animal communication, of adaptation to typically human linguistic laws that enhance redundancy and lower energy costs (Favaro et al., 2020; Gustison et al., 2016; Heesen et al., 2019; Valente et al., 2021).

The study of animal rhythms

The study of animal rhythms revolves around three main approaches (Ravignani, 2019): the study of spontaneous individual production of rhythmic patterns, interactive and collective rhythm production in groups, and coordination among individuals producing rhythmic patterns (**Figure 3**). We can imagine tackling these aspects from a single perceptual channel or using a multimodal approach.

These three approaches underlie the abovementioned view that different rhythmic behaviours can be expected to depend on parallel, multi-component, stand-alone mechanisms (Fitch, 2015; Kotz et al., 2018). Producing rhythmic patterns individually or collectively does not require coordination or entrainment. For example, the ability of two individuals to coordinate with each other does not require beat perception or entrainment (this can often be a sudden response, in terms of "when you emit, I immediately emit"). In this sense, the study of animal rhythm, in the first instance and as it is in its infancy, may be limited to attesting which animal groups exhibit which of these mechanisms before inferring which mechanisms and functions are underlying.

Each rhythmic trait can be investigated through the declination of Tinbergen's four questions. First of all, as a mechanism: what is the mechanism, or neural and biological predisposition underlying rhythm and coordinated behaviour? Also, from an ontogenetic (i.e., when does a particular rhythmic ability appear?), phylogenetic and functional perspectives (i.e., when does a particular rhythmic ability appear through evolution and why might it be adaptive?).

A comparative approach, which considers the multi-component structure of mechanisms making rhythm possible, effectively complements the phylogenetic approach, but not only in looking at a particular trait. It allows us to reconstruct which features of human rhythmic behaviour are present or similar in other species and why (Kotz et al., 2018). Identifying the production of rhythmic patterns in solitary or in groups, and the presence of coordination among individuals, is the first step in inferring the evolutionary history of rhythm in the other clades (Hauser & McDermott, 2003). Such a process, Fitch (2015) suggests, should be *broad* (from the closest to human to the most distant) and cross-cutting in the homology and analogy aspects of rhythmic outputs: if homologies allow inferences to be made about inherited by descent from a common ancestor, homoplasy has complementary functions, such as the potential to identify what selective pressures, regardless of evolutionary history, cause a particular trait to appear by convergence.

Finally, in studying musicality from a biological perspective, it is crucial to combine a comparative approach with the identification of musical universals (Fitch, 2015). Musical universals correspond to the features that are common to all human cultures around the world. A particular ability or behavioural output in all human cultures may underlie a pivotal biological mechanism that shapes musicality. The identification of universals is controversial and has largely been debated since some universal capacities may not be expressed in some specific cultures or musical genres, referring back to long-standing considerations of strict nature and nurture separation from which it is problematic to draw a synthesis. However, a quantitative empirical approach (Savage et al., 2015) has made it possible to partially overcome this obstacle. Savage and colleagues (2015) identified some universals that, although not absolute, are consistent across all cultures and ascribable as generalizable. They found 18 universals, including an isochronous beat, a metrical hierarchy, singing and dancing and group performance.

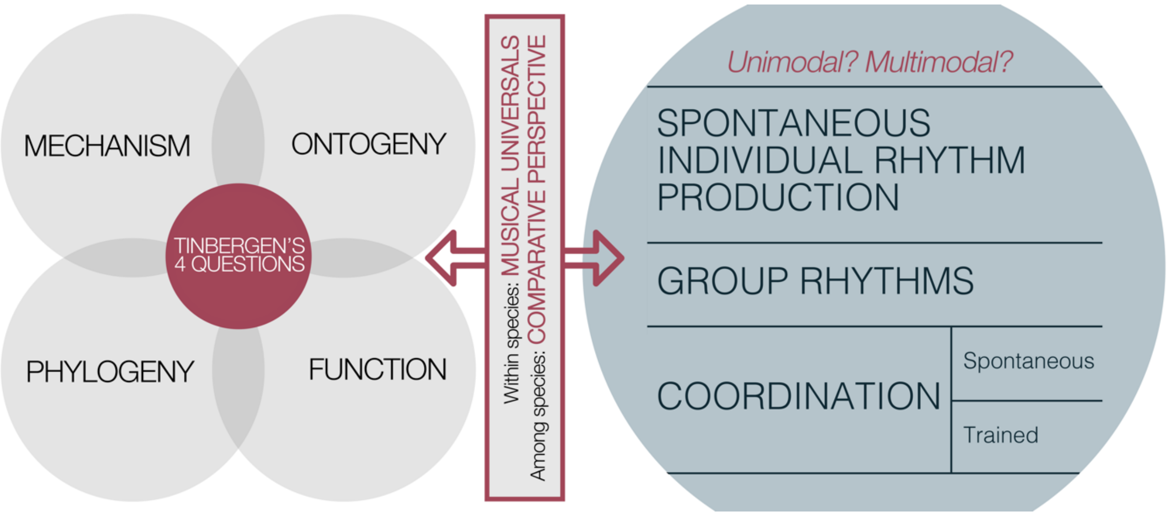


Figure 3 - Schematization of the main approaches for studying rhythm and its evolutionary history (Fitch 2015; Ravignani, 2019; Pouw 2020). We can approach the observation of unimodal or multimodal rhythmic behaviours such as the spontaneous production of individual rhythm, collective displays (e.g., duets, choruses) and coordinated displays (in spontaneous or trained contexts) by examining the four Tinbergen questions as the investigation keystones. Two possible means to articulate the questions are identifying musical universals (within-species approach) and a comparative perspective (among-species approach).

Aims of the thesis

This study aims to complement a comparative view through the identification of rhythmic categories, and in particular small integer ratios, in the vocalizations of primates (**Figure 4A**). The presence of isochrony or other small-integer ratios has never been verified in the vocalizations of this clade. It is unclear whether the production of regular rhythmic patterns was present in the vocalizations of our primate ancestor: assessing the presence of this trait in other primate species may allow reconstructing its evolutionary history.

To this end, six primate species were identified within three different families (**Figure 4B-C**): starting with the most closely-related to humans, Hominidae, with the genus Pongo (divergence from humans 7 Ma), Hylobatidae, with the genera Hylobates and Nomascus (divergence from humans 12 Ma), and Indriidae, with the genus Indri (divergence from humans ~74 Ma). The study of different evolutionary branches allows the backwards reconstruction of the evolutionary history that determined the selection, or not, of rhythmic categories in primates. A comparative perspective enables to identify, through a kind of *differential diagnosis*, the selective pressures that may have determined the emergence or maintenance of categorical rhythms in the vocalizations of each species. These species, in fact, share the essential presence of vocal activity in their communication, but show radically different eco-ethological characteristics (**Table 1**).

In the course of this investigation, I paid special attention to aspects such as the effect of structure and hierarchical organization of the signal, the potential for an effect of sexual selection on the signal, the adaptiveness of redundancy for the specific vocal type depending on its function, the presence of physiological constraints that may determine the rhythmic structure of the signal, and the effect of social context and vocal interaction in determining the rhythmic structure of the signal.

The thesis is organized into five chapters in a descending gradient of phylogenetic proximity to humans. The first two chapters (**CHAPTER 1-2**) investigate the presence of rhythmic categories in the long calls of a great ape, the male Bornean orangutan (*Pongo pygmaeus wurmbii*). The first chapter considers long calls as a whole, while the second delves into the presence of nested rhythmic structures typical of human musicality. The third chapter (**CHAPTER 3**) deals with the presence of rhythmic categories in the duet and solos songs of a smaller ape, the lar gibbon (*Hylobates lar*), with a focus on the flexibility of the song's rhythmic structure based on the context of emission and vocal interaction. The role of physiological constraints in determining the temporal structure of the signal is also assessed. The fourth chapter (**CHAPTER 4**) focuses on the study of rhythmic categories in the choruses of three gibbon species of the genus *Nomascus:* the yellow-cheeked gibbon (*Nomascus gabriellae*), the white-cheeked gibbon (*Nomascus leucogenys*), and the southern white-cheeked gibbon (*Nomascus siki*). Finally, the fifth chapter (**CHAPTER 5**) deals with the presence of rhythmic categories in the only species of singing lemur, the indri (*Indri indri*), with a focus on sexual dimorphism in the temporal structure of the song. Indri allows us to investigate the presence of categorical rhythms in a species that shares an ancestor with humans dating back to a basal branch of primate evolution. An **OVERALL DISCUSSION** aims to outline the general patterns emerging from these works, with possible insights into evolutionary hypotheses that may help answer the question: what is the biological origin of rhythm?

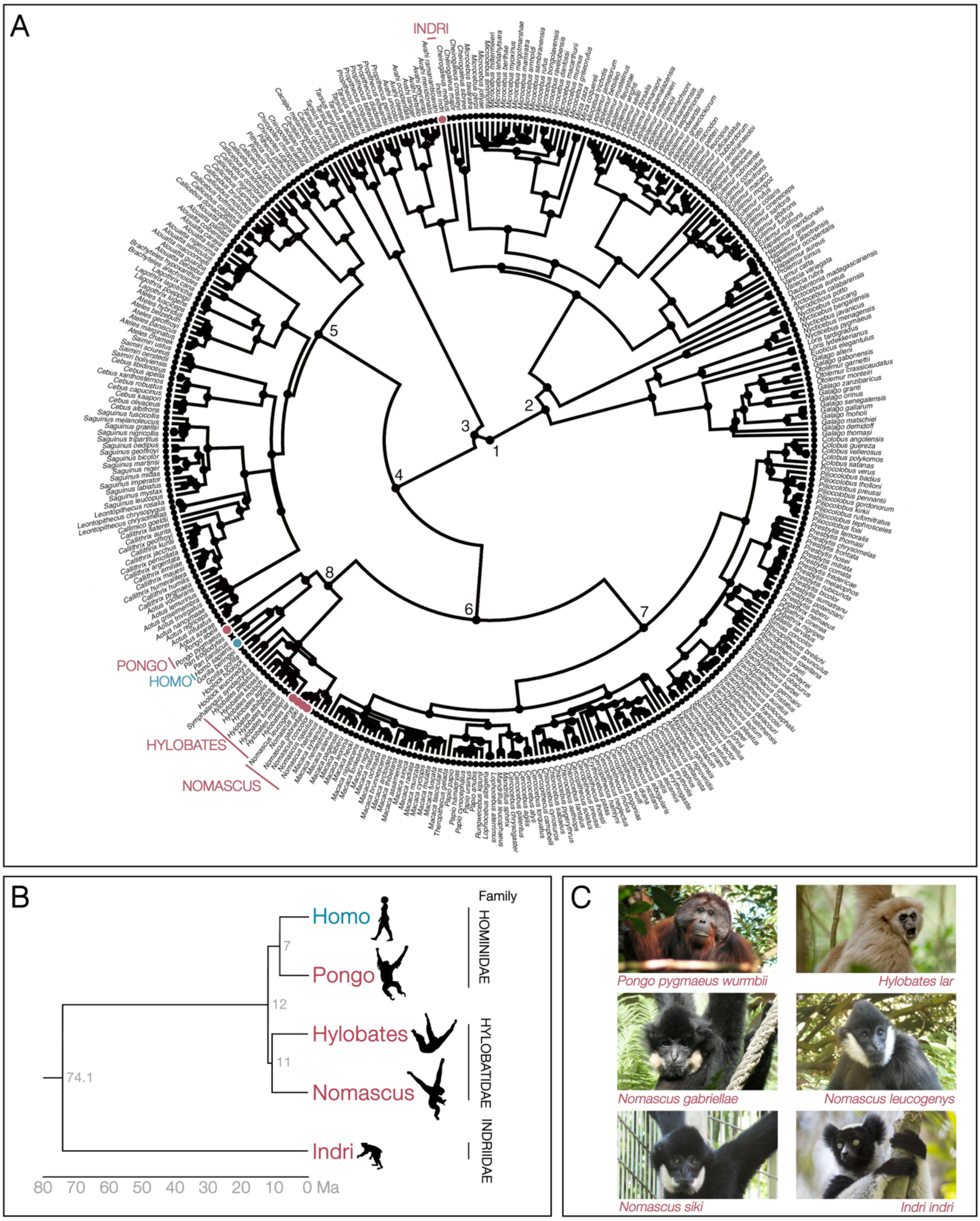


Figure 4 – **A.** Mod. from Kappeler and Pozzi 2019. Primate phylogeny (roots: 1: primate; 2: Strepsirhini; 3: Haplorhini; 4: Anthropoidea; 5: Platyrrhini; 6: Catarrhini; 7: Cercopithecoidea; 8: Hominoidea) with colored red dots showing the studied species (*Pongo pygmaeus wurmbii*, *Hylobates lar*, *Nomascus gabriellae*, *Nomascus leucogenys*, *Nomascus siki*, *Indri indri*) and the blue dot *Homo sapiens*. Genera covered by this research are highlighted on the phylogenetic tree with colored lines that include the set of species belonging to that genus. **B.** Details of the phylogenetic tree representing only the relationships between the genus Homo and the genera studied in this thesis. The grey numbers on the phylogenetic tree indicate, in millions of years, the time of the divergence node estimated by the most up-to-date literature (Pozzi et al., 2014), while the horizontal axis on the bottom represents the reference time scale (0 Ma, today - 80 Ma). **C.** Pictures of the 6 studied species. Photo credits – *P.p.w.*: Madeleine Hardus and Adriano Lameira; *H.l.*: Giovanni Di Panfilo; *N.*spp: Camilla Mancassola; *I.I.*: Filippo Carugati.

Table 1 - Table summarising the main eco-ethological characteristics of the analysed species and vocal type studied in this thesis. At the bottom of the table, the bibliographical references for each specific trait can be found. As can be seen, the six species display characteristic traits that are partly shared (by homology or analogy), which make it possible to implement a 'differential analysis' to identify the selective pressures that may have acted in shaping the vocal type rhythmic structure.

TAXONOMY		ECO-ETHOLOGY				VOCALIZATIONS					
Family	Species	Singing primate	Social system	Territorial	Analyzed vocal type	Long distance signal	>1 individuals participating to the display	Rhythmic locomotory display associated to vocalization	Hierarchical structure of the signal	Signal subjected to sexual selection	Advertisement function
HOMINIDAE	*Pongo pygmaeus wurmbii*	N De Gregorio et al. 2022	Semi-solitary Spillmann et al. 2015	N Kunz et al. 2022	Male long calls	Y Lameira & Wich, 2008	N Delgado et al. 2009	N	-	- Mitani, 1985	Y Mitani 1985
HYLOBATIDAE	*Hylobates lar*	Y De Gregorio et al. 2022	Pair-living / Polyandrous Reichard 1995 Reichard et al. 2012	Y Bartlett 2009	Song (duets and male solos)	Y Raemarkers & Raemarkers 1985	Y *Raemarkers et al. 1984*	Y Geissmann 2000	Y Geissmann 2002	Y Cowlishaw 1992	Y Cowlishaw 1992
	Nomascus gabriellae	Y De Gregorio et al. 2022	Pair-living / Multifemale Kenyon et al. 2011 Barca et al. 2016	Y Kenyon et al. 2011	Song (duets and choruses)	Y Fan et al. 2009 (on *N. concolor*)	Y *Merker & Cox 1999*	Y Fan et al. 2016 (on *N. nasutus*)	Y Geissmann 2002	Y Cowlishaw 1992	Y Cowlishaw 1992
	Nomascus leucogenys	Y De Gregorio et al. 2022	Multifemale Harding 2012	-	Song (duets and choruses)	Y Fan et al. 2009 (on *N. concolor*)	Y *Konrad & Geissmann 2006*	Y Fan et al. 2016 (on *N. nasutus*)	Y Geissmann 2002	Y Cowlishaw 1992	Y Cowlishaw 1992
	Nomascus siki	Y De Gregorio et al. 2022	-	-	Song (duets and choruses)	Y Fan et al. 2009 (on *N. concolor*)	Y *Geissmann et al. 2007*	Y Fan et al. 2016 (on *N. nasutus*)	Y Geissmann 2002	Y Cowlishaw 1992	Y Cowlishaw 1992
INDRIDAE	*Indri indri*	Y De Gregorio et al. 2022	Pair-living Bonadonna et al. 2019	Y Bonadonna et al. 2017	Advertisement songs (duets and choruses)	Y Bonadonna et al. 2017	Y Gamba et al. 2016	N Torti et al. 2013	Y Thalmann et al., 1993	Y Bonadonna et al. 2014 Gamba et al. 2016	Y Torti et al. 2013

LEGEND

Y = yes

N = no

- = no information / lack of evidence

REFERENCES

Bonadonna G, Torti V, De Gregorio C, Valente D, Randrianarison RM, Pozzi L, Gamba M, Giacoma C. 2019. Evidence of genetic monogamy in the lemur Indri (Indri indri). Am J Primatol. 81:e22993.

Bonadonna, G., Torti, V., Randrianarison, R. M., Martinet, N., Gamba, M., & Giacoma, C. (2014). Behavioral correlates of extra-pair copulation in Indri indri. Primates, 55, 119-123.

Bonadonna, G., Torti, V., Sorrentino, V., Randrianarison, R. M., Zaccagno, M., Gamba, M., ... & Giacoma, C. (2017). Territory exclusivity and intergroup encounters in the indris (Mammalia: Primates: Indridae: Indri indri) upon methodological tuning. The European Zoological Journal, 84(1), 238-251.Cowlishaw, G. U. Y. (1992). Song function in gibbons. Behaviour, 121(1-2), 131-153.

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., ... & Giacoma, C. (2022). Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. Ethology Ecology & Evolution, 34(3), 205-219.Reichard, U. (1995). Extra-pair copulations in a monogamous gibbon (Hylobates lar). Ethology, 100(2), 99-112.Reichard et al. 2012

Delgado, R. A., Lameira, A., Davila Ross, M., Husson, S. J., Morrogh-Bernard, H. C., & Wich, S. A. (2009). Geographical variation in orangutan long calls. Orangutans: Geographic variation in behavioral ecology and conservation, 215-224.

Fan, P. F., Ma, C. Y., Garber, P. A., Zhang, W., Fei, H. L., & Xiao, W. (2016). Rhythmic displays of female gibbons offer insight into the origin of dance. Scientific reports, 6(1), 1-8.

Fan, P. F., Xiao, W. E. N., Huo, S., & Jiang, X. L. (2009). Singing behavior and singing functions of black-crested gibbons (Nomascus concolor jingdongensis) at Mt. Wuliang, Central Yunnan, China. American Journal of Primatology: Official Journal of the American Society of Primatologists, 71(7), 539-547.

Gamba M, Torti V, Estienne V, Randrianarison RM, Valente D, Rovara P, Bonadonna G, Friard O, Giacoma C. 2016. The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. Front Neurosci. 10:249. doi:10.3389/fnins.2016.00249

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. The origins of music, 103-123.

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. Biological Reviews, 77(1), 57-76.

Harding, L. E. (2012). Nomascus leucogenys (Primates: Hylobatidae). Mammalian Species, 44(890), 1-15.Bartlett 2009

Kenyon, M., Roos, C., Binh, V. T., & Chivers, D. (2011). Extrapair paternity in golden-cheeked gibbons (Nomascus gabriellae) in the secondary lowland forest of Cat Tien National Park, Vietnam. Folia Primatologica, 82(3), 154-164.Barca et al. 2016

Kunz, J. A., van Noordwijk, M. A., van Schaik, C. P., & Shackelford, T. (2022). Orangutan Sexual Behavior.

Mitani, J. C. (1985). Sexual selection and adult male orangutan long calls. Animal Behaviour, 33(1), 272-283.

Raemaekers, J. J., & Raemaekers, P. M. (1985). Long-range vocal interactions between groups of gibbons (Hylobates lar). Behaviour, 95(1-2), 26-44.

Spillmann, B., van Noordwijk, M. A., Willems, E. P., Mitra Setia, T., Wipfli, U., & van Schaik, C. P. (2015). Validation of an acoustic location system to monitor Bornean orangutan (Pongo pygmaeus wurmbii) long calls. American Journal of Primatology, 77(7), 767-776.

Thalmann, U., Geissmann, T., Simona, A., & Mutschler, T. (1993). The indris of Anjanaharibe-sud, northeastern Madagascar. International Journal of Primatology, 14, 357-381.

Torti, V., Gamba, M., Rabemananjara, Z. H., & Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. Italian Journal of Zoology, 80(4), 596-607.

CHAPTER 1

Isochrony in the long calls of a great ape

Image generated on DALL.E, an Artificial Intelligence capable of generating images from textual descriptions through synography. "Digital art of an orangutan playing the drums"

Isochrony in the long calls of a great ape

Teresa Raimondi[1], Madeleine E. Hardus[^], Marco Gamba[1], Adriano R. Lameira[2]

[1]Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy

[2]Department of Psychology, University of Warwick, Coventry, UK

[ì]Aarhus/Aalborg, Denmark

[^]Independent researcher

ABSTRACT

Humans, in all the world's cultures, are the only species to have evolved a proper musical system: but what is the biological origin of such exceptional behaviour? The study of rhythm from a comparative perspective can provide information regarding the biological origin of musicality. We, therefore, investigated the presence of isochrony, i.e., a rhythmic structure characterised by temporal intervals of equal duration, like the beat of a metronome, in the male long calls of a great ape, the Bornean orangutan (*Pongo pygmaeus wurmbii*). This species shares a recent ancestor with humans (~7 Ma), and hence is an ideal animal model for reconstructing the phylogenetic origin of periodic pattern generation in the primate clade. We found that isochronous pattern production is present in the Bornean orangutan vocalizations. The detection of such a trait shared with humans in this species allows us to advance hypotheses concerning the adaptivity of isochrony production. The various hypotheses in this regard are discussed from the point of view of the ethology and ecology of the species under consideration and generalised to mechanisms that may also have acted in the evolution of musicality in our species.

INTRODUCTION

The role of rhythm in animal communication is subject to growing interest since it can provide indirect information on the evolution of human communication and its correlation to musicality (Ravignani & Madison, 2017). In particular, the function of isochrony (the property of a signal that shows equal duration temporal intervals, like the beat of a metronome; Ravignani & Madison, 2017) in primate communication has recently been deepened by two studies that found isochrony in the song of two singing species: indris (*Indri indri*; De Gregorio et al., 2021) and lar gibbons (*Hylobates lar;* Raimondi et al., 2023). Since a signal showing rhythmic regularity is predictable, the presence of isochrony in humans, and potentially in singing primates, has been hypothesized to play a role in orchestrating vocal interaction (Merker et al., 2009). Primates emit songs in solos, duets or choruses (De Gregorio et al., 2022) and are coordinated, flexible in their temporal structure, and interactive (de Reus et al., 2020; Gamba et al., 2016; Haimoff, 1986; Terleph et al., 2018). In lar gibbons, for instance, isochrony is closely related to synchrony (Raimondi et al., 2023), in line with the evolutionary hypothesis that traces the origin of rhythm to the need for multiple individuals to coordinate with each-other in multimodal, collective displays (Merker et al., 2009; Savage et al., 2021). In our species, these displays are universal, meaning that, regardless of the specific culture, they are present all around the world (Savage et al., 2015): the strong link between rhythm and individuals' coordination and interaction is widespread and probably adaptive in our species in general.

Primates are not the only taxon displaying isochrony in spontaneous vocal production, as three singing bird species also show rhythmic regularity in their vocalizations: male thrush nightingales (*Luscinia Luscinia*; Roeske et al., 2020), male zebra finches (*Taeniopygia guttata;* Roeske et al., 2020), and Australian pied butcherbird (*Cracticus nigrogularis;* Xing et al., 2022). Male rock hyraxes' (*Procavia capensis*) songs also show isochrony and the rhythmic regularity is positively correlated with the fitness of the emitter in terms of surviving offspring, in line with the idea that rhythmic regularity can reflect the genetic and developmental quality of the singer (Demartsev et al., 2022).

To sum up, the presence of isochrony seems to be the result of manifold, non-mutually exclusive selective pressures that make a regular beat adaptive: in species in which the male emits solo songs (as in many birdsongs or rock hyraxes), it is likely that rhythmic regularity advertises the quality of the emitter, and can be subjected to sexual selection, while we can assume that the level of required coordination and rhythmic interaction required for duetting and chorusing is higher (such as Australian pied butcherbirds or the singing primates). With the current knowledge, it appears however hazardous to infer general evolutionary trends across *taxa*, given that isochrony has been quantitatively searched and found in the vocalizations of only six singing, not-closely-related, species. The vocalizations of non-singing species have not yet been investigated, and whether rhythmic regularity evolved for convergence is still unknown.

In this study, we tested for the presence of isochrony in male long calls of a great ape, the Bornean orangutan (*Pongo pygmaeus wurmbii*; **Figure 1**). This species represents an interesting animal model for the study of rhythm evolution, since it embodies the crasis of unique evolutionary, social, eco-ethological characteristics, thus allowing to selectively isolate the pressures acting on the rhythmic features of the signal. It is the only great ape that can be found outside Africa, it is highly dimorphic, adult males show bimaturism (unflanged males can maintain their morph for up to 20 years in the wild), they have a huge body mass, but an arboreal lifestyle (Delgado Jr. & Van Schaik, 2000). Flanged males are vocally active, emitting long calls, and semi-solitary (Delgado Jr. & Van Schaik, 2000). Orangutans are also closely related to humans and share a recent ancestor (around 17MY; Pozzi et al., 2014). Their communication shares pivotal features with the human one: to give a few examples, *clicks* (voiceless calls) and *faux-speech* (voiced calls) show a speech-like rhythm, compatible with a potential articulatory common evolutionary origin between great ape calls and human consonants and vowels (Lameira et al., 2015); orangutans flexibly and multimodally adjust tactical communication decisions depending on the different degrees of understanding of human caregivers (Cartmill & Byrne, 2007; Poss et al., 2006); they show a form of vocal learning and spontaneously and plastically acquire and modulate human whistles (Lameira et al., 2013; Wich et al., 2009); they show tool-use, potentially as a cultural display, to lower the maximum frequency of their kiss squeaks (M. Hardus et al., 2009).

Flanged male individuals typically emit long calls, which travel long distances (Lameira & Wich, 2008) and can last several minutes (Spillmann et al., 2010) and they often begin with a series of *grumbles*, followed by further *pulses* organized in clusters (Hardus et al., 2008). The call usually ends with a series of *bubbles* (Hardus et al., 2008). Long calls serve for intra-sexual spacing (Mitani, 1985a), but their function is however still debated, as Bornean orangutans seem to show variable-among-sites social organization, and significantly differ from Sumatran orangutans (Setia et al., 2009). The role of long calls in mate attraction, for instance, is still nebulous (Mitani, 1985a): for the time being, there is no conclusive, quantitative evidence to support or reject it. Consequently, it is unclear whether long calls are subjected to sexual selection and whether they convey information on the emitter's quality.

Even though long calls are emitted by solitary males that don't need to coordinate with each-other, we expect that orangutans may anyway show isochrony in their long calls, as a potential advertisement of the emitter quality for intra- and inter-sex communication. Also, we claim that isochrony, if present, may result from parallel constraints (*e.g.,* physiological) or serve unexplored additional functions (*e.g.,* energy saving and redundancy; Ravignani & Madison, 2017; Ravignani, 2021).

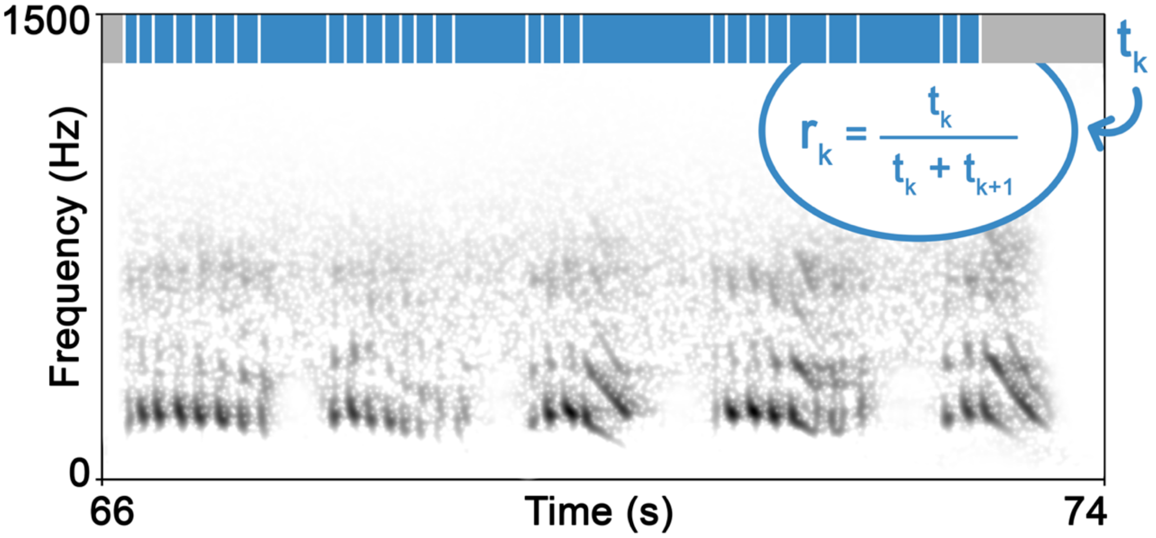


Figure 1 – Spectrogram of short section of a male Bornean orangutan's long call, generated in Praat (wl = 0.06 s). The colored bar on the top of the spectrograms displays the inter-onset-intervals (t_k) with onsets in white. The rhyrhm ratios (r_k) formula is shown in the ellipse.

MATERIAL AND METHODS

Animals and recordings. We conducted our research at the Tuanan Research Station (2°09′S; 114°26′E), Central Kalimantan, Indonesia. Long calls were recorded from 11 identified flanged males (*Pongo pygmaeus wurmbii*) through a Marantz Analogue Recorder PMD222 in combination with a Sennheiser Microphone ME 64 or a Sony Digital Recorder TCD-D100 in combination with a Sony Microphone ECM-M907.

Acoustic analyses. We transferred audio recordings to a computer with a sampling rate of 44.1 kHz. Pulses durations were manually extracted directly from the spectrogram window (window type: Hann; 3 dB filter bandwidth: 124 Hz; grid frequency resolution: 2.69

Hz; grid time resolution: 256 samples) using Raven interactive sound analysis software (version 1.5, K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, 2022.). We then extracted durations on a *CSV* file from which we calculated the temporal intervals between the onset of a pulse and the following one (inter-onset interval, or t_k). We focused on $0.025 \leq t_k \leq 5$ sec (t_k n = 7086) since these values are hypothesized to be the lower and upper limit for meter perception and performance in macaques and humans and the thresholds on orangutans are unknown (Kuhl & Padden, 1983; London, 2004). The rhythm ratio (r_k) was then calculated between a t_k and the next one, t_{k+1}, as $t_k/(t_k+t_{k+1})$ (r_k n = 6538).

Statistical analyses. All statistical analyses were performed in R (version 4.2.2, R Core Team, 2022). We plotted the probability density functions of t_k and r_k values and identified the number and position of the peak values on the curves. We tested the significance of the r_k peaks on the density distribution falling in the vicinity of isochrony (corresponding to r_k=0.5) through the methodology displayed by Roeske and colleagues (2020). We identified the boundaries referable to on-isochrony sectors between 0.440-0.555 (centred around 0.5 r_k value) and the external, symmetric off-isochrony boundaries on 0.400-0.440 and 0.555-0.600 r_k values. We then counted, per individual contribution, the number of r_k instances falling into the off-isochrony vs. on-isochrony sectors of the curve. To test the significance of the peak around isochrony, we used a Generalized Linear Mixed Model (GLMM; *glmmTMB* R package; Brooks et al., 2017): the count of r_k was the response variable, which followed a negative-binomial distribution (we specified the family distribution), in line with a count variable. We used the bin (off or on-isochrony) as a predictor and added an offset weighting the r_k count based on the width of the bin (off-isochrony bins: 0.044; on-isochrony bins: 0.056). The individual contribution code nested in the individual were used as random intercepts. We tested the significance of the *Full* model comparing it with a *Null* model, the latter comprising only the offset and the random factors, with a likelihood ratio test (Anova with “Chisq” argument; Dobson & Barnett, 2018). We obtained p-values for predictors using the R *summary* function. We checked normality, homogeneity (via function provided by R. Mundry), and number of the residuals. We checked for overdispersion with the *performance* R package (Lüdecke et al., 2020).

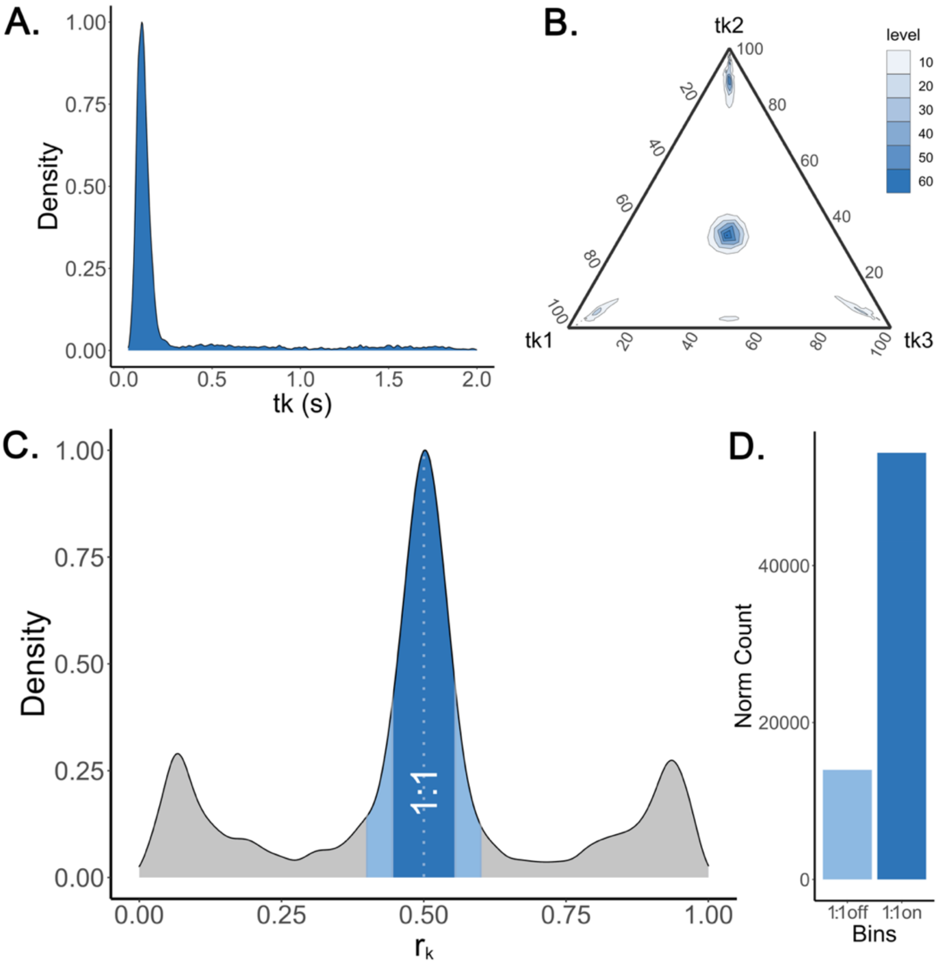


Figure 2 – **(a)** Probability density function showing the distributions of t_k. **(b)** Ternary density plot displaying the duration of three successive t_k. The three axes are normalized on a scale of 0-100. **(c)** Probability density function shoeing the distributions of rhythm ratios (r_k), suggesting the prominence of r_k distribution around isochrony (r_k =0.5). The dark-blue section of the curve indicates on-isochrony r_k ranges and the light-blue sections indicate off-isochrony r_k ranges. **(d)** Barplot of the r_k adjusted count (r_k count / bin width) for values falling within on-isochrony vs off-isochrony ranges of the density; as displayed in **Table 1**, r_k values falling on-isochrony are significantly more than r_k values falling off-isochrony.

RESULTS

T_k raw values have an average duration of 0.329±0.567 (mean±sd) and show a single peak of 0.098 s on the density distribution curve (**Figure 2A**). When plotting the t_k values on a ternary density plot (**Figure 2B**), a prominent central cluster is visible, corresponding to three successive t_k intervals with roughly equal duration (i.e. isochrony). The probability density function of r_k distribution confirms that, and shows a conspicuous peak around isochrony (**Figure 2C**; peak value=0.503). The apparent presence of isochrony at both two-successive and three-successive t_k scales of investigation is validated by the GLMM testing for the effect of the on-off-isochrony sector of the bin (**Table 1**; Full vs Null: df=1, Chisq= 103.206, p<0.001) on the number of observed r_k. In other words, on-isochrony sectors of the curve contain significantly more r_k values than off-isochrony ones (**Table 1**; bin on-isochrony vs bin off-isochrony – Estimate= 1.249, z.value=16.700, p<0.001): orangutans vocalize at a highly regular pulse in which the relationship between successive t_k is consistent through time.

Table 1 – Summary of the GLMM testing for the difference between the number of on-isochrony t_k versus the number of off-isochrony t_k. Full vs Null: AIC_Null= 1063.028, AIC_Full 959.822, Chisq = 103.206, df=1, p<0.001.

Generalized Linear Mixed Model
Family = Negative binomial
R_k count ~ Isochrony bin (off/on-isochrony) + offset (bin width) + (1 I Individual / Individual Contribution)

CONDITIONAL MODEL

	Estimate	Std. Error	Z value	Pr(>lzl)
(Intercept)	4.299	0.157	[a]	[a]
Bin (on) [b,c]	1.249	0.075	16.700	<0.001

[a] Not shown as not having a meaningful interpretation

[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.

[c] Reference category: “Bin (off)”

DISCUSSION

We provide the first evidence that another great ape, besides humans, shows isochronous rhythms in their vocalizations. We show that vocal interaction is not a necessary condition to develop a regular pace in vocalizations. Since isochrony is present in the long calls of orangutans, we can postulate that this results from additional selective pressures that make rhythmic regularity adaptive. Long calls are prolonged, loud, expensive displays: this aspect makes them potentially subjected to sexual selection and physiological constraints, and particularly favoured by an energy rationalization.

Many animal physiological and neural processes in animals rely on local rhythmic regularity (e.g., the oscillation of neuron groups, heartbeat, respiration, and locomotion;

Ravignani & Madison, 2017). Neural oscillators have been suggested to have a role in orchestrating many rhythmic displays, and a rhythmically impaired display can be a sign of an individual's dysfunctionality, and consequently of its lower fitness. In contrast, rhythmic regularity can represent an honest signal of the emitter's quality (Van Den Broek & Todd, 2009). In rock hyraxes, isochrony reflects the male quality and is correlated to male reproductive success (Demartsev et al., 2022). In the Sumatran orangutan (*Pongo abelii*), long calls also serve for female attraction (T. M. Setia & van Schaik, 2007; sexually active females slightly approach the emitter of spontaneous long calls). Male Bornean orangutan long calls serve as a driver for advertising the intended travel trajectory, to convey individual identity, and they reflect the context of emission (females with offspring ignore long calls caused by a disturbance but move away if the long call is spontaneous; Askew & Morrogh-Bernard, 2016; Spillmann et al., 2010). It is then not clear whether, in Bornean orangutans, long calls are subjected to sexual selection and whether they serve for mate attraction, and there is no conclusive evidence to support this hypothesis or reject it (Mitani, 1985a; Spillmann et al., 2010). Future insights correlating the rhythmic regularity of long calls with the attraction of sexually active females may shed light on the male's quality advertisement function of this vocalization.

Isochrony can also be the by-product of physiological and respiratory constraints. The link between movement, vocalization and breathing is evident (Ravignani & Kotz, 2020) and implies an advantage in emitting a signal, and thus breathing, at regular intervals. At the same time, isochrony may be the result of a minimum possible interval duration below which an animal cannot vocalize: since a high call rate has the potential to be adaptive and often represents an honest signal of the emitter's quality (Catchpole & Slater, 2003), the individual calls as fast as possible, but can't indefinitely reduce its inter-onset intervals. Even though this is not the case for pied butcherbirds, in which rhythmic song structure is probably not a biomechanical by-product of respiration constraints (Xing et al., 2022), the presence of a minimum threshold in the inter-onset-intervals duration could not be rejected in lar gibbons, which show a positive correlation between call rate and isochrony (the faster the signal, the more regular the intervals; Raimondi et al., 2023). Here we show short temporal intervals (peak value of 0.098) that may be compatible with the presence of a respiratory constraint (the shortest inter-onset-interval possible is emitted) that determines isochrony. Coming to our studied species, Askew and Morrogh-Bernard (2016) suggest that the dominant male in their sample showed the shortest units and highest pulse rate, especially in the first 30 s of the long calls, compared to non-dominant individuals. Further insights unequivocally linking isochrony to call rate will be needed in the future to support this hypothesis.

From the point of view of information theory, isochrony is the simplest existing structure, is purely deterministic and predictable and guarantees the lowest entropy possible (Ravignani & Madison, 2017). A regular pattern is then adaptive since it minimizes entropy and enhances redundancy, avoiding the risk of corrupting the conveyed information caused by transmission error. However, this generally comes at the expense of signal expressivity (i.e. the amount of contained information). However, in the face of the fixity of the rhythmic organization of the signal, however, expressiveness can be mediated otherwise. In Bornean orangutans, in addition to temporal parameters, also the frequency and general descriptive parameters (call rate, call duration, number of pulses and of pulses type) have an effect in determining, for instance, the individuality or context-specific characteristics of the long calls (Spillmann et al., 2010): it is possible to maintain consistent inter-onset-intervals over time through an adjustable trade-off between phonation and non-phonation intervals and to convey additional information through frequency parameters. Also, the combinatorics of different vocal types has been shown to determine emergent semantic properties in many species (Crockford & Boesch, 2005; Ouattara et al., 2009), and is potentially compatible with an isochronous rhythmic pattern. Furthermore, isochrony does not necessarily decline according to one level of organization: nested isochrony (isochrony within-isochrony) may be present in the long calls, but hidden when looking at the rhythmic structure overall, as we did in the current study. Hierarchical, combinatorial structures (i.e. combining elements into increasingly larger size constructions) are not only decisive in human communication but also widespread in animal communication (Zuberbühler, 2018; e.g., primate songs - Zanoli et al., 2020 - bird songs - Suzuki et al., 2016 – cetaceans vocalizations (Kello et al., 2017). An approach that further investigates the hierarchical structure of orangutans' long calls would be in line with the grounded idea that rhythm is strictly intertwined with the signal syntax (i.e., the sentence structure, and how it interplays with other dimensions of linguistic information; Xing et al., 2022). The quantitative description of a signal's rhythmic structure, hierarchical organization and complexity confronts a methodological challenge (Janney et al., 2016; Sasahara et al., 2015; Xing et al., 2022), but their synthesis probably represents a paramount keystone to infer the proximate and remote mechanisms that determined the evolution of rhythm, and isochrony in particular, in communication.

The temporal structure of a communication signal is the outcome of the interplay of numerous, coexisting selective pressures. We suggest that their identification and correlation with the presence of rhythmic categories, from a comparative perspective within the primate clade, may be the key to reconstructing the evolutionary steps that led to the evolution of musicality in our species.

REFERENCES

Askew, J. A., & Morrogh-Bernard, H. C. (2016). Acoustic Characteristics of Long Calls Produced by Male Orang-Utans (*Pongo pygmaeus wurmbii*): Advertising Individual Identity, Context, and Travel Direction. *Folia Primatologica*, *87*(5), 305–319. https://doi.org/10.1159/000452304

Brooks, M., E., Kristensen, K., Benthem, K., J. ,van, Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H., J., Mächler, M., & Bolker, B., M. (2017). *GlmmTMB* Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378. https://doi.org/10.32614/RJ-2017-066

Cartmill, E. A., & Byrne, R. W. (2007). Orangutans Modify Their Gestural Signaling According to Their Audience's Comprehension. *Current Biology*, *17*(15), 1345–1348. https://doi.org/10.1016/j.cub.2007.06.069

Catchpole, C. K., & Slater, P. J. B. (2003). *Bird Song: Biological Themes and Variations*. Cambridge University Press.

Crockford, C., & Boesch, C. (2005). Call combinations in wild chimpanzees. *Behaviour*, 397-421.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. Proceedings of the National Academy of Sciences, 106(51), 22026-22031.De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Gamba, M., & Giacoma, C. (2022). Notes on a tree: Reframing the relevance of primate choruses, duets, and solo songs. *Ethology Ecology & Evolution*, 1–15. https://doi.org/10.1080/03949370.2021.2015451

De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A., & Gamba, M. (2021). Categorical rhythms in a singing primate. *Current Biology*, *31*(20), R1379–R1380. https://doi.org/10.1016/j.cub.2021.09.032

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M. D., Bruno, J. H., Trainor, L., & Ravignani, A. (2020). *Rhythm in dyadic interactions* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/9yrkv

Delgado Jr., R. A., & Van Schaik, C. P. (2000). The behavioral ecology and conservation of the orangutan (*Pongo pygmaeus*): A tale of two islands. *Evolutionary Anthropology: Issues, News, and Reviews*, *9*(5), 201–218. https://doi.org/10.1002/1520-6505(2000)9:5<201::AID-EVAN2>3.0.CO;2-Y

Demartsev, V., Haddas-Sasson, M., Ilany, A., Koren, L., & Geffen, E. (2022). Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *Journal of Animal Ecology*, 1365-2656.13801. https://doi.org/10.1111/1365-2656.13801

Dobson, A. J., & Barnett, A. G. (2018). *An Introduction to Generalized Linear Models* (4th ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315182780

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., Bonadonna, G., Friard, O., & Giacoma, C. (2016). The Indris Have Got Rhythm! Timing and Pitch Variation of a Primate Song Examined between Sexes and Age Classes. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00249

Haimoff, E. H. (1986). Convergence in the duetting of monogamous Old World primates. *Journal of Human Evolution*, *15*(1), 51–59. https://doi.org/10.1016/S0047-2484(86)80065-3

Hardus, M. E., Lameira, A. R., Singleton, I., Morrogh-Bernard, H. C., Knott, C. D., Ancrenaz, M., Utami Atmoko, S. S., & Wich, S. A. (2008). A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In S. A. Wich, S. S. Utami Atmoko, T. M. Setia, & C. P. van Schaik (Eds.), *Orangutans: Geographic Variation in Behavioral Ecology and Conservation* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199213276.003.0004

Hardus, M., Lameira, A., Schaik, C., & Wich, S. (2009). Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation? *Proceedings. Biological Sciences / The Royal Society*, *276*, 3689–3694. https://doi.org/10.1098/rspb.2009.1027

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (n.d.). *Raven Pro: Interactive Sound Analysis Software* (1.5). Ithaca, NY: The Cornell Lab of Ornithology. https://ravensoundsoftware.com/

Kello, C. T., Dalla Bella, S., & Médé, B. (2017). *Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes*.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *The Journal of the Acoustical Society of America*, *73*(3), 1003–1010. https://doi.org/10.1121/1.389148

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE*, *10*(1), e116136. https://doi.org/10.1371/journal.pone.0116136

Lameira, A. R., Hardus, M. E., Kowalsky, B., de Vries, H., Spruijt, B. M., Sterck, E. H. M., Shumaker, R. W., & Wich, S. A. (2013). Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. *The Journal of the Acoustical Society of America*, *134*(3), 2326–2335. https://doi.org/10.1121/1.4817929

Lameira, A. R., & Wich, S. A. (2008). Orangutan Long Call Degradation and Individuality Over Distance: A Playback Approach. *International Journal of Primatology*, *29*(3), 615–625. https://doi.org/10.1007/s10764-008-9253-x

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). Performance*: Assessment of regression models performance. R package version 0.4.*

Merker, B. H., Madison, G. S., & Eckerdal Gamla Kyrkv, P. (2009). *On the role and origin of isochrony in human rhythmic entrainment. 45(1)*, 4–17. https://doi.org/10.1016/j.cortex.2008.06.011

Mitani, J. C. (1985). Sexual selection and adult male orangutan long calls. *Animal Behaviour*, *33*(1), 272–283. https://doi.org/10.1016/S0003-3472(85)80141-X

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. Proceedings of the National Academy of Sciences, 106(51), 22026-22031.

Poss, S. R., Kuhar, C., Stoinski, T. S., & Hopkins, W. D. (2006). Differential use of attentional and visual communicative signaling by orangutans (*Pongo pygmaeus*) and gorillas (*Gorilla gorilla*) in response to the attentional status of a human. *American Journal of Primatology*, *68*(10), 978–992. https://doi.org/10.1002/ajp.20304

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *75*, 165–183. https://doi.org/10.1016/j.ympev.2014.02.023

R Core Team. (2022). *R: A language and environment for statistical computing.* (4.2.2). R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raimondi, T., Di Panfilo, G., Pasquali, M., Zarantonello, M., Favaro, L., Savini, T., Gamba, M., & Ravignani, A. (2023). Isochrony and rhythmic interaction in ape duetting. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1990), 20222244. https://doi.org/10.1098/rspb.2022.2244

Ravignani, A. (2021). Isochrony, vocal learning, and the acquisition of rhythm and melody. *Behavioral and Brain Sciences*, *44*, e88. https://doi.org/10.1017/S0140525X20001478

Ravignani, A., & Kotz, S. A. (2020). Breathing, voice, and synchronized movement. *Proceedings of the National Academy of Sciences*, *117*(38), 23223–23224. https://doi.org/10.1073/pnas.2011402117

Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology*, *8*(NOV), 1–13. https://doi.org/10.3389/fpsyg.2017.01820

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology*, *30*(18), 3544-3555.e6. https://doi.org/10.1016/j.cub.2020.06.072

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Setia, T., Delgado, R., Atmoko, S., Singleton, I., & Schaik, C. (2009). Social organization and male-female relationships. *Orangutans: Geographic Variation in Behavioral Ecology and Conservation*. https://doi.org/10.5167/uzh-31343

Setia, T. M., & van Schaik, C. P. (2007). The Response of Adult Orang-Utans to Flanged Male Long Calls: Inferences about Their Function. *Folia Primatologica*, *78*(4), 215–226. https://doi.org/10.1159/000102317

Spillmann, B., Dunkel, L. P., Noordwijk, M. A. V., Amda, R. N. A., Lameira, A. R., Wich, S. A., & Schaik, C. P. V. (2010). Acoustic Properties of Long Calls Given by Flanged Male Orang-Utans (*Pongo pygmaeus wurmbii*) Reflect Both Individual Identity and Context. *Ethology*, *116*(5), 385–395. https://doi.org/10.1111/j.1439-0310.2010.01744.x

Terleph, T. A., Malaivijitnond, S., & Reichard, U. H. (2018). Male white-handed gibbons flexibly time duet contributions. *Behavioral Ecology and Sociobiology*, *72*(1), 1–13. https://doi.org/10.1007/s00265-017-2432-z

Van Den Broek, E. M. F., & Todd, P. M. (2009). Evolution of rhythm as an indicator of mate quality. *Musicae Scientiae*, *13*(2_suppl), 369–386. https://doi.org/10.1177/1029864909013002161

Wich, S. A., Swartz, K. B., Hardus, M. E., Lameira, A. R., Stromberg, E., & Shumaker, R. W. (2009). A case of spontaneous acquisition of a human sound by an orangutan. *Primates*, *50*(1), 56–64. https://doi.org/10.1007/s10329-008-0117-y

Xing, J., Sainburg, T., Taylor, H., & Gentner, T. Q. (2022). Syntactic modulation of rhythm in Australian pied butcherbird song. *Royal Society Open Science*, *9*(9), 220704. https://doi.org/10.1098/rsos.220704

Zuberbühler, K. (2018). Combinatorial capacities in primates. *Current Opinion in Behavioral Sciences*, *21*, 161–169. https://doi.org/10.1016/j.cobeha.2018.03.015

CHAPTER 2

Recursive self-embedded vocal motifs in wild orangutans

Accepted - preprint in revision at eLife (April 2023)

Image generated on DALL.E, an Artificial Intelligence capable of generating images from textual descriptions through synography. "Digital art of an orangutan playing the saxophone"

Accepted - Preprint *eLife* (April 2023)

https://doi.org/10.1101/2023.04.14.536836

Recursive self-embedded vocal motifs in wild orangutans

Adriano R. Lameira[1*], Madeleine E. Hardus[^], Andrea Ravignani[2,3], **Teresa Raimondi**[4], Marco Gamba[4]

[1]Department of Psychology, University of Warwick, Coventry, UK

[2]Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

[3]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Denmark

[4]Department of Life Sciences and Systems Biology, University of Turin, Torino, Italy

^Independent researcher

*corresponding author: adriano.lameira@warwick.ac.uk

ABSTRACT

Recursive procedures that allow placing a vocal signal inside another of similar kind provide a neuro-computational blueprint for syntax and phonology in spoken language and human song. There are, however, no known vocal patterns among nonhuman primates arranged in self-embedded combinations that evince vocal recursion or potential insipient forms or neuro-procedures thereof, suggesting a neuro-cognitive transformation exclusive to humans. Here, we uncover that wild flanged male orangutan long calls show two hierarchical strata, wherein rhythmically isochronous call sequences are nested within self-similar isochronous call sequences. Remarkably, three unrelated recursive motifs occurred simultaneously in long calls, refuting that motifs resulted from three parallel linear procedures or that motifs were simple anatomical artifacts of bodily constrains. Findings represent a case of recursive hominid vocal production in the absence syntax, semantincs, phonology or music. Second-order combinatorics, ‘sequences within sequences’, involving hierarchically organized and cyclically structured vocal sounds in ancient hominids may have preluded the evolution of recursion in modern language-able humans.

INTRODUCTION

Among the many definitions of recursion (M. Martins, 2012), the view that it represents the capacity to iterate a signal within a self-similar signal has crossed centuries and disciplines, from von Humboldt (von Humboldt, 1836) and Hockett (Hockett, 1960), to Mandelbrot (Mandelbrot, 1980) and Chomsky (Chomsky, 2010); from fractals in mathematics (Mandelbrot, 1980) to generative grammars in linguistics (Chomsky, 2010). Across varying terminologies, the common denominator across fields is that to re-curse (i.e., to 're-*invoke*') or re-iterate is an algorithmic "hack" to produce infinite signal states from a finite signal set. Although classically associated with syntax (Chomsky, 2010; Idsardi et al., 2018), recursive signal production and the resulting self-embedded structures have also been recognised in phonology (Bennett, 2018; Elfner, 2015; Kabak & Revithiadou, 2009; Nasukawa, 2015, 2020; Vogel, 2012) and (verbal and non-verbal) music (Jackendoff, 2009; Koelsch et al., 2013; M. D. Martins et al., 2017; Sharma & Chimalakonda, 2018). Given that language and music are uniquely human, this hints at a neuro-cognitive or neuro-computational transformation in the human brain that enabled the emergence of multiple open-ended communication systems in the human lineage, but seemingly, none other (Hauser et al., 2002). The absence of vocal structures in (nonhuman) primates that could help inform insipient or transitional recursive states, namely within the hominid family, has led some scholars to question altogether whether recursion was the result of natural selection (Berwick & Chomsky, 2019; Bolhuis & Wynne, 2009) and to rebutte the role of shared ancestry (Jackendoff & Pinker, 2005) and evolution as an incremental path-dependent process (de Boer et al., 2020; Kershenbaum et al., 2014; P. T. Martins & Boeckx, 2019), favouring instead sudden "hopeful monster" mutant scenarios.

Decades-long debates on the evolution of recursion have ensued, carved around the successes and limitations of empirical comparative animal research. For example, perception and processing of syntax-like vocal combinatorics has been identified in some bird (Engesser et al., 2016, 2019; Gentner et al., 2006; Liao et al., 2022b; T. Suzuki et al., 2016; T. N. Suzuki et al., 2017) and primate species (Jiang et al., 2018; Wang et al., 2015; Watson et al., 2020) but results' interpretation has received various criticisms (Bolhuis et al., 2018a; Bowling & Fitch, 2015; Corballis & Corballis, 2014; Rawski et al., 2021). For example, animal studies have, thus far, almost exclusively focused on *perception* (cf. Ferrigno et al., 2020): animals recognizing or responding to recursive signal structures; however, processing operations of recursive signals are not necessarily recursive themselves (Corballis & Corballis, 2014; Miyagawa, 2021). Critically, most research has focused on laboratory animals in artificial test settings who were presented with synthetic stimuli after dedicated human training (Gentner et al., 2006; Jiang et al., 2018; Liao et al., 2022b; Watson et al., 2020). This opens comparative research to two critical drawbacks. First, results are mute about evolutionary precursors and processes. Settings and stimuli have not been those to which species adapted over evolutionary time, with most species tested thus far being too distantly related to humans to allow inferences about recursion emergence among hominids in the first place. Second, results have been slanted by the dominant classic theory that, counter-intuitively, disfavours a gradual evolutionary scenario for recursion and instead conceives it as a punctuated event (Berwick & Chomsky, 2019; Bolhuis et al., 2015; Bolhuis & Wynne, 2009). Namely, experimental stimuli have consisted of recursive signal sequences organized along a single temporal scale (though not structurally linear), similarly with how *Merge* and syntax operate. However, recursive signal structures can also unfold in other manners, such as across nested temporal scales and in the absence of semantics (T. W. Fitch, 2017a), as in music, but these remain unprobed and untested.

An approach to the production of recursive signals, complementary to perceptual studies, is thus, desirable. Data from wild primates in particular may help infer signal patterns that were recursive in some degree or kind in an extinct past and subsequently moulded into the recursive structures observed today in modern humans. By virtue of their own primitive nature, proto-recursive structures did not likely fall within modern-day classifications, and thus, will often fail to be predicted based on assumptions guided by full-fledged language (Kershenbaum et al., 2014; Miyagawa, 2021). To this end, a structural approach is particularly advantageous. First, no prior assumptions are required about species' cognitive capacities. These are directly inferred from how signal sequences are organised. For example, Chomsky's definition of recursion (Chomsky, 2010) can generate non-self-embedded signal structures, but these would be operationally undetectable amongst other signal combinations. Second, no prior assumptions are required about signal meaning. There are no certain parallels with semantic content and word meaning in animals, but analyses of signal patterning allow to identify similarities between non-semantic (nonhuman) and semantic (human) combinatoric systems (Lipkind et al., 2013; Sainburg et al., 2019). Therefore, the search for recursion can be made in the absence of meaning-base operations, such as *Merge*. Third, no prior assumptions are required about signal function. Under punctuated or gradual evolutionary hypotheses, ancestral signal function (whether cooperative, competitive or otherwise) is expected to have derived or been leveraged by its proto-recursive structure, otherwise, once present, recursion would not have been fixated among human ancestral populations. A structural approach opens, therefore, the field to untapped signal diversity in nature and yet unrecognised bona fide combinatoric states within the human clade.

Here, undertaking an explorative structural approach to recursion, we provide evidence for recursive self-embedded vocal patterning in a (nonhuman) great ape, namely, in the long calls of flanged orangutan males in the wild. We conducted precise rhythm analyses (De Gregorio et al., 2021b; Roeske et al., 2020b) of 66 long call audio recordings produced by 10 orangutans (*Pongo pygmaeus wurmbii*) across approximately 2510 observation hours at Tuanan, Central Kalimantan, Indonesian Borneo. We identified 5 different element types that comprise the structural building blocks of long calls in the wild (Hardus et al., 2009; Lameira & Wich, 2008), of which the primary type are full pulses (Fig. 1A). Full pulses do not, however, always exhibit uninterrupted vocal production throughout a long call (as during a long call's climax; Spillmann et al., 2010) but can break-up into 4 different "sub-pulse" element types: (*i*) grumble sub-pulses (quick succession of staccato calls that typically constitute the first build-up pulses of long calls (Hardus et al., 2009), (*ii*) sub-pulse transitory elements and (*iii*) pulse bodies (typically constituting pulses before and/or after climax pulses) and (*iv*) bubble sub-pulses (quick succession of staccato calls that typically constitute the last tail-off pulses of long calls; **Fig. 1A**). We characterised long calls' full- and sub-pulses' rhythmicity to determine if orangutan long calls present a re-iterated structure across different hierarchical strata. We extracted inter-onset-intervals (IOIs; i.e. time difference between the start of a vocal element and the preceding one - t_k) from 8993 vocal long call elements (**Fig. 1A**): 1930 full pulses, 757 grumble sub-pulses, 1068 sub-pulse transitory elements, x pulse bodies and 4422 bubble sub-pulses. From the extracted IOIs, we calculated their rhythmic ratio by dividing each IOI for its duration plus the duration of the following interval. We then computed the distribution of these ratios to ascertain whether the rhythm of long call full and sub-pulses presented natural categories, following published protocols (De Gregorio et al., 2021b; Roeske et al., 2020b; **Fig. 1B, C, D**).

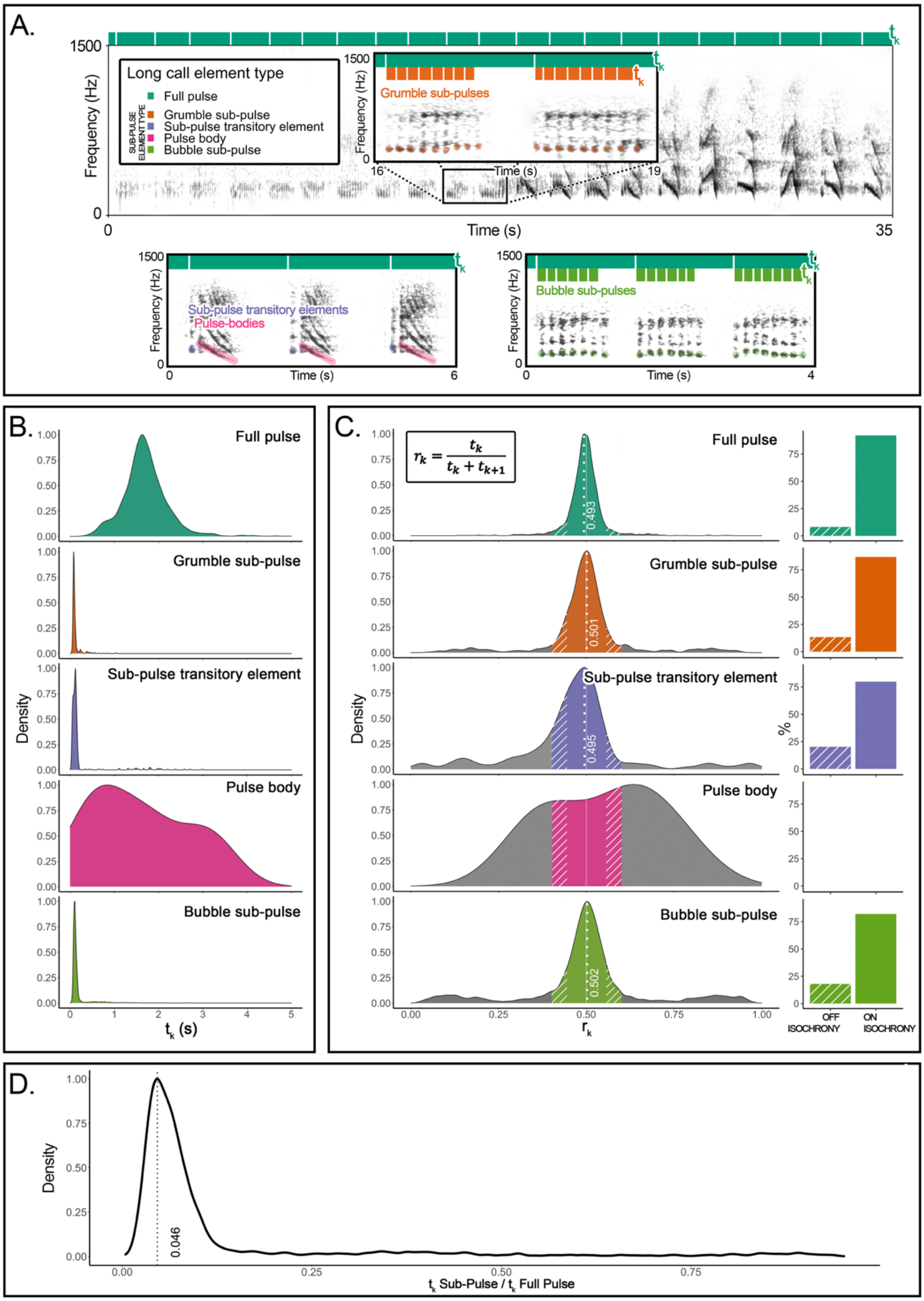


Figure 1 – Organization and rhythmic features of orangutans' long calls. (A) On top, the spectrogram of a *Full Pulse* and its organization in Sub-Pulses (e.g., *Grumble sub-pulses*). Below are the spectrograms of the three other sub-element types: *Sub-pulse transitory* elements, *Pulse bodies* and *Bubble sub-pulses*. Bars on the top of each spectrogram schematically quantify durations of inter-onset intervals (t_k): dark green denotes

the higher-level of organization (*Full pulse*). Orange (in the inset) and light green (bottom right) denote the lower-level organization (sub-pulse element types). **(B)** Probability density function showing the distributions of the inter-onset-intervals (t_k) for each of the long call element types. **(C)** The distributions on the left show rhythm ratios (r_k) per element type as calculated on 12 flanged males for a total of 1915 Full-pulses and 5309 sub-pulses. Solid sections of the curves indicate on-isochrony r_k values; striped sections indicate off-isochrony r_k values. A solid white line indicates the 0.5 r_k value corresponding to isochrony. White dotted lines denote the on-isochrony peak value extracted from the probability density function. On the right, a bar plot per each element type shows the percentage of observations (r_k) falling into the on-isochrony boundaries (solid bars) or on off-isochrony boundaries (striped bars). The number of on-isochrony r_k is significantly larger (GLMM, *Full* vs *Null*: Chisq=2717.543, $p<0.001$) than the number of off-isochrony r_k for all long call element types (*Full pulse*: t-ratio=-25.164, $p<0.001$; *Bubble sub-pulse*: t-ratio= -30.694, $p<0.001$; *Grumble sub-pulse*: t-ratio=-14.526, $p<0.001$; *Sub-pulse transitory element*: t-ratio=-3.148, $p<0.001$). *Pulse body* showed no r_k values falling within the on-off-isochrony boundaries. **(D)** Distribution of a variable calculated as the ratio between the t_k of a sub-pulse and the t_k of the corresponding higher level of organization, the *Full Pulse*. We report the peak value of the curve (0.046) and tested the significance of the extent of the central quartiles, which was significantly smaller than peripheral quartiles (Wilcoxon signed-rank test: W=2272, $p<0.001$). Spectrograms are generated in Praat (wl = 0.04-0.06).

RESULTS

The density probability function of orangutan full pulses showed one peak (r_k=0.493) in close vicinity to a theoretically pure isochronic rhythm, that is, full pulses were regularly paced at 1:1 ratio, following a constant tempo along the long call (**Fig. 1C**). Our model (GLMM, full model vs null model: Chisq=298.2876, df=7, $p<0.001$; see **Table 1**) showed that pulse type, range of the curve (on-off-isochrony), and their interaction, had a significant effect on the count of rk values. In particular, full pulses' isochronous peak tested significant (t.ratio=-15.957, $p<0.0001$), that is, the number of rk values falling inside on-isochrony range was significantly higher than the number of rks falling inside the off-isochrony range (**Fig. 1C**). Critically, three (of the four) orangutan sub-pulse element types – grumble sub-pulses, sub-pulse transitory elements and bubble sub-pulses – also showed significant peaks (**Table 1**; grumble sub-pulses: t.ratio = -5.940, $p<.0001$; sub-pulse transitory elements: t.ratio=-4.048, p=0.0001; bubble sub-pulses: t.ratio= - 10.640, $p<.0001$) around pure isochrony (peak rk: grumble sub-pulses = 0.501; sub-pulse transitory elements=0.495; bubble sub-pulses=0.502; **Fig. 1C**). That is, sub-pulses were regularly paced within regularly paced full pulses, denoting isochrony within isochrony (**Fig. 1C**) at different average tempi (mean tk (sd): full pulses=1.696 (0.508); grumble sub-pulses=0.118 (0.111); sub-pulse transitory elements=0.239 (0.468); bubble sub-pulses= 0.186 (0.292); **Fig. 1B**). Overall, sub-pulses' t_k was equivalent to 0.046 of their comprising full-pulses (**Fig. 1D**), which put sub-pulses at an approximate ratio of 1:20 relative to that of full-pulses, the smallest categorical temporal rhythmic interval registered thus far in a vertebrate (De Gregorio et al., 2021b; Roeske et al., 2020b). Permuted discriminant function analyses (Mundry & Sommer, 2007; crossed, in order to control for individual variation) in R (Team, 2013) based on seven acoustic measures extracted from grumble, transitory and bubble sub-pulses confirmed that these represented indeed distinct sub-pulse categories, where the percentage of correctly classified selected cases (62.7%) was significantly higher (p=0.001) than expected (37%).

Table 1 - Summary and details of the GLMM and pairwise *post-hoc* testing for the effect of the pulse level (BSP, FP, GSP, SBTE) in interaction with Type (ON- or OFF-isochrony range) on the count of r_k observations.

Conditional model

Count of obs ~ Pulse level *Type + offset + (1 | ID Contribution)

Pulse level: BSP, FP, GSP, SBTE

Type: ON or OFF isochrony

Full (AIC=2114.608) vs Null (AIC=2398.896): df=7, Chisq=298.288, p<0.001

	Estimate	Std Error	zvalue	Pr(>|z|)
(Intercept)				
Type (ON) [a]	1.2959	0.1445	8.97	<0.0001
Pulse level (FP)[a]	-1.3532	0.1670	-8.10	<0.001
Pulse level (GSP) [a]	-0.8826	0.2449	-3.60	<0.001
Pulse level (SBTE) [a]	-2.3268	0.2966	-7.85	<0.001
Pulse level (FP) : type (ON) [a]	1.0314	0.2150	4.80	<0.001
Pulse level (GSP) : type (ON) [a]	0.1615	0.3241	0.50	0.618
Pulse level (SBTE) : type (ON)[a]	-0.1903	0.3626	-0.52	0.600

[a] *Pulse level (BSP) and Type (OFF) being the reference categories*

***Post-hoc* test**

Contrast	Estimate	SE	df	t.ratio	p.value
FP: OFF - ON	-2.57	0.161	319	-15.957	<.001
BSP: OFF - ON	-1.54	0.144	319	-10.640	<.001
GSP: OFF - ON	-1.70	0.286	319	-5.940	<.001
SBTE: OFF - ON	-1.35	0.333	319	-4.048	<.001

DISCUSSION

Orangutan long call nested rhythmic patterns reveal self-similar embedded isochrony in the vocal production of a wild great ape, notably, with two discernible structural strata – full- and sub-pulse level – and three non-exclusive rhythmic arrangements in the form of [isochrony^{A} [isochrony^{a,b,c}]].

Human and nonhuman great apes have similar auditory capacities (Quam et al., 2015). There are no identified skeletal differences in inner ear anatomy that could suggest significantly distinct sound sensitivity, resolution or activation thresholds in the time domain (Quam et al., 2015). Humans perceive an acoustic pulse as a continuous pitch, instead of a rhythm, at rates higher than 30 Hz (i.e., 30 beats per second). Long call sub-pulses exhibited average rhythms at ~9.263 (3.994) Hz (i.e., t_k=0.184 (0.303) s). Therefore, hominid ear anatomy offers strong confidence that orangutans, like humans (and other great apes) perceive sub-pulse rhythmic motifs as such, i.e., as a train of signals, not as one uninterrupted signal. Assuming otherwise would imply that auditory time-resolution differ by more than one order of magnitude between humans and other great apes in the absence of obvious anatomical culprits.

The simultaneous occurrence of non-exclusive recursive patterns excludes the likelihood that orangutans concatenate long calls and their subunits in linear structure without any recursive processes. To generate the observed vocal motifs as linear structure, three independent neuro-computational procedures would need to run in parallel to generate distinct isochronic rhythms at the sub-pulse level, whilst being indistinguishable, transposable and/or interchangeable at the pulse level without interference, which would be unlikely, if theoretically possible at all. Non-exclusive simultaneous recursive patterns also help exclude the probability that recursion was the primary by-product of anatomic constrains, such as breath length, heart beat and other physiological rhythmic processes or movements (Pouw et al., 2020). Such processes could generate various simultaneous rhythmic patterns; however, these would be expected to be in the form of harmonics of the same base "carrier" rhythm, but the three observed rhythmic arrangements at the sub-pulse level were not related to the pulse level by any small integer ratios. Together, this strongly suggests that the observed recursive self-embedded motifs are most likely generated by a recursive neurological procedure or recursion algorithm.

Recursive self-embedded vocal motifs in orangutans indicate that vocal recursion among hominids is not exclusive to human vocal and cognitive systems. This is not to suggest that they exhibit *all* properties that recursion exhibits in modern language-able humans. Such expectation would be unreasonable as it would imply that no evolution occurred in >10 million years since the split between the orangutan and human phylogenetic lineages. Thus, any apparent differences with recursion in today's syntax, phonology or music do not invalidate the probability that the reported vocal motifs represent an ancient, or potential ancestral, state for the ensuing evolution of recursion along the human clade.

Recursion and fractal phenomena are prevalent across the universe. From celestial and planetary movement to the splitting of tree branches and the morphology of bacteria colonies, patterns within self-similar patterns are the norm, not the exception. This makes the seeming singularity of human recursion amongst vertebrate signals only the more enigmatic. Our findings indicate that ancient vocal patterns organized across nested structural strata were likely present in ancestral hominids. Recursive vocal production likely predated the evolutionary emergence of (spoken) language within the Hominid family and human lineage. This data-driven possibility poses new evolutionary trajectories and timelines distinct from the classic theoretical notions that the advent of recursion was a saltational all-or-nothing mutational event that took place only recently in modern humans (Berwick & Chomsky, 2019). Gradual cumulative evolutionary scenarios for the emergence and evolution of recursion (de Boer et al., 2020; P. T. Martins & Boeckx, 2019) can and should be considered and investigated, with the advantage that living primate models in general, and natural hominid behaviour in particular, offer superior empirical and comparative validity beyond purely theoretical considerations and dissertations (Lameira et al., 2021). A combination of both approaches will likely pay the highest heuristic dividends. For example, future research may implement playback experiments designed to clarify whether or how great apes cognitively represent natural recursive self-embedded motifs as informed by previous designs and theoretical considerations (Corballis & Corballis, 2014; Engesser et al., 2016; Watson et al., 2020).

Implications for the evolution of recursion and cognition

The presence of recursive self-embedded vocal motifs in orangutans carries four major implications for the evolution of recursion and cognition. First, much ink has been laid on the topic, yet, the possibility of self-embedded isochrony, or alternatively, non-exclusive self-embedded patterns occurring within the same signal sequence, has on no account been formulated or conjectured as a possible state of recursive signalling, be it in vertebrates, mammals, primates or otherwise, extant or extinct. This suggests that controversy may have thus far been underscored by data-poor circumstances, namely, lack of rigorous and detailed knowledge about call combinations in wild primates in general, and great apes in particular. The presence of recursive vocal patterns in a wild great ape in the absence of syntax, phonology or music helps prevent prior assumptions that recursion precursors ought, or should be expected to, operate as modern-day recursion. Orangutan recursive vocal patterns provide the first data point and open a new charter for possible insipient or transitional states within the hominid family that, whilst inherently distinct from modern-day recursion, were nonetheless homologous to it and fully functional in their own right. The open discussion about what kind of properties would or could make a structure proto-recursive vs. non-proto-recursive will be essential to move the state-of-knowledge past antithetical, jointly exhaustive or mutually exclusive options of what recursion is and how it emerged and evolved.

Second, primate loud calls are functionally analogous to, for instance, bird and whale song. Accordingly, sophisticated signalling in far-related species does not undermine the value of great apes as living models for the study of language and its underpinning neuro-motoric operations, as claimed (Berwick & Chomsky, 2019; Bolhuis et al., 2018b; Bolhuis & Wynne, 2009; Lattenkamp & Vernes, 2018; T. N. Suzuki et al., 2018; Vernes et al., 2021) on the false assumption that great ape vocal research has been at a standstill for the last half-century (Belyk & Brown, 2017; Bianchi et al., 2016; Crockford et al., 2004; Hopkins et al., 2007; Lameira, 2017; Lameira et al., 2013, 2015a, 2016; Lameira & Shumaker, 2019; Pereira et al., 2020; Russell et al., 2013; Staes et al., 2017; Taglialatela et al., 2012; Watson et al., 2015; S. Wich et al., 2009; S. A. Wich et al., 2012). Our findings invite results across lineages to be tested with primates in order to explicitly assess their evolutionary relevance within the human clade. This will help avoid prematurely interpreting absence of evidence for evidence of absence in primates.

Third, our findings suggest that, despite criticism, recursive perceptual capacities identified in primates (Watson et al., 2020) may be factual and likely evolved to subserve vocal signals in the species' natural repertoire (though potentially yet unrecognized). This invites for a renewed interest and re-analysis of primate signalling behaviour in the wild (Gabrić, 2021). However, findings also show that it may be too hasty to discuss whether perceptual capacities in primates or birds are equivalent to those engaged in syntax (Watson et al., 2020) or phonology (Rawski et al., 2021). Such classifications may be putting the proverbial cart before the horse; they are based on untested assumptions (e.g., that syntax and phonology evolved as separate "modules", that one attained modern form before the other, etc) that may not have applied to proto-recursive ancestors (Kershenbaum et al., 2014; Miyagawa, 2021).

Forth, given that isochrony universally governs music and that recursion is a feature of music, findings suggest an evolutionary link between great ape loud calls and vocal music. Loud calling is an archetypal trait in primates (S. A. Wich & Nunn, 2002) but is absent in modern humans. Our findings suggest may not be coincidental. Great ape loud calling may have preceded and subsequently transmuted into modern recursive vocal structures in humans. Given their conspicuousness, loud calls represent one of the most studied aspects of primate vocal behaviour (S. A. Wich & Nunn, 2002), but their rhythmic patterns have seldom been characterized with precision (Clink et al., 2020; De Gregorio et al., 2021b; Gamba et al., 2016). Besides our analyses, there are remarkably few confirmed cases of isochrony great apes, but the behaviours that have been rhythmically measured with accuracy have been implicated in the evolution of percussion (Fuhrmann et al., 2015) and musical expression (Dufour et al., 2015; Hattori & Tomonaga, 2020), such as social entrainment in chimpanzees in connection with the origin of dance (Lameira et al., 2019; a capacity once falsely assumed to be neurologically absent in great apes; T. W. Fitch, 2017b; Patel, 2014). This opens the possibility that recursive vocal production and supporting neural procedures were first and foremost a feature of the proto-musical expression in human ancestors, later recruited and “re-engineered” for the generation of linguistic combinatorics.

Future studies outlining the distribution of isochrony across primate (vocal) behaviour offer promising new paths to empirically assess the evolution of recursive signal structures in music and language and will help move the needle forward on one the most tantalizing riddles in the evolution of language and cognition. These crucial data and insights will materialise if, as stewards of our planetary co-habitants, humankind secures the survival of these species and the preservation of their natural wild habitat (Estrada et al., 2017, 2022; Laurance, 2013; Laurance et al., 2012).

METHODS AND MATERIALS

Study site

We conducted our research at the Tuanan Research Station (2°09′S; 114°26′E), Central Kalimantan, Indonesia. Long calls were opportunistically recorded from identified flanged males (*Pongo pygmaeus wurmbii*) using a Marantz Analogue Recorder PMD222 in combination with a Sennheiser Microphone ME 64 or a Sony Digital Recorder TCD-D100 in combination with a Sony Microphone ECM-M907.

Acoustic data extraction

Audio recordings were transferred to a computer with a sampling rate of 44.1 kHz. Seven acoustic measures were extracted directly from the spectrogram window (window type: Hann; 3 dB filter bandwidth: 124 Hz; grid frequency resolution: 2.69 Hz; grid time resolution: 256 samples) by manually drawing a selection encompassing the complete long call (sub)pulse from onset to offset, using Raven interactive sound analysis software (version 1.5, Cornell Lab of Ornithology). These parameters were duration(s), peak frequency (Hz), peak time, peak frequency contour average slope (Hz), peak frequency contour maximum slope (Hz), average entropy (Hz), signal-to-noise ratio (NIST quick method). Please see software's documentation for full description of parameters (https://ravensoundsoftware.com/knowledge-base/pitch-tracking-frequency-contour-measurements/). Acoustic data extraction complemented the classification of long calls elements, both at the pulse and sub-pulse levels, based on close visual and auditory inspection of spectrograms, both based on elements' distinctiveness between each other as well as in relation to the remaining catalogued orangutan call repertoire (Hardus et al., 2009; see also supplementary audio files). Of these parameters, duration and peak frequency in particular have been shown to be resilient across recording settings(Lameira et al., 2013) and to adequately represent variation in the time and frequency axes (Lameira et al., 2017).

Rhythm data analyses

Inter-onset-intervals (IOI's = t_k) were only calculated from the begin time (s) of each full- and sub-pulse long call elements using Raven interactive sound analysis software, as above explained. t_k was calculated only from subsequent (full/sub) pulse elements of the same type. Ratio values (r_k) were calculated as $t_k/(t_k+t_{k+1})$. Following the methodology of Roeske et al., 2020 and De Gregorio et al. 2021, to assess the significance of the peaks around isochrony (corresponding to the 0.5 r_k value), we counted the number of r_ks falling inside on-isochrony ranges ($0.440 < r_k < 0.555$) and off-isochrony ranges ($0.400 < r_k < 0.440$ and $0.555 < r_k < 0.600$), symmetrically falling at the right and left sides of 1:1 ratios (0.5 r_k value). We tested the count of on-isochrony r_ks versus the count of off-isochrony r_ks, per pulse type, with a GLMM for negative-binomial family distributions, using *glmmTMB* R library. In particular, we built a *full* model with the count of r_k values as the response variable, the pulse type in interaction with the range the observation fell in (on- or off- isochrony) as predictors. We added an offset weighting the r_k count based on the width of the bin. The individual contribution was set as random factor. We built a *null* model comprising only the offset and the random intercepts. We checked the number of residuals of the full and *null* models, and compared the two models with a likelihood ratio test (Anova with “Chisq” argument). We calculated p-values for each predictor using the R *summary* function and performed pairwise comparisons for each level of the explanatory variables with *emmeans* R package, adjusting all p-values with Bonferroni correction. We checked normality, homogeneity (via function provided by R. Mundry), and number of the residuals. We checked for overdispersion with *performance* R package [90]. Graphic visualization was prepared using R (Team, 2013) packages *ggplot2* (Wickham, 2009) and *ggridges* (Wilke, 2022). Data reshape and organization were managed with *dplyr* and *tidyr* R packages.

Acoustic data analyses

Permutated discriminant function analysis with cross classification was performed using R and a function provided by Roger Mundry (Mundry & Sommer, 2007). The script was: pdfa.res=pDFA.crossed (test.fac="Sub-pulse-type", contr.fac="Individual.ID", variables=c("Delta.Time", "Peak.Freq", "Peak.Time", "PFC.Avg.Slope", "PFC.Max.Slope", "Avg.Entropy", "SNR.NIST.Quick"), n.to.sel=NULL, n.sel=100, n.perm=1000, pdfa.data=xdata). These analyses assured that long call elements, at the pulse and sub-pulse level, indeed represented biologically distinct categories.

REFERENCES

Belyk M, Brown S. The origins of the vocal brain in humans. Neurosci Biobehav Rev. 2017;77: 177–193. doi:10.1016/j.neubiorev.2017.03.014

Bennett R. Recursive prosodic words in Kaqchikel (Mayan). Glossa: a journal of general linguistics. 2018;3: 67. doi:10.5334/gjgl.550

Berwick RC, Chomsky N. All or nothing: No half-Merge and the evolution of syntax. PLoS Biol. 2019;17: e3000539. doi:10.1371/journal.pbio.3000539

Bianchi S, Reyes LD, Hopkins WD, Taglialatela JP, Sherwood CC. Neocortical grey matter distribution underlying voluntary, flexible vocalizations in chimpanzees. Scientific reports. 2016;6: 34733. doi:10.1038/srep34733

Bolhuis JJ, Beckers GJ, Huybregts MA, Berwick RC, Everaert MB. Meaningful syntactic structure in songbird vocalizations? PLoS biology. 2018;16: e2005157. doi:10.1371/journal.pbio.2005157

Bolhuis JJ, Beckers GJL, Huybregts MAC, Berwick RC, Everaert MBH. The slings and arrows of comparative linguistics. PLoS Biol. 2018;16: e3000019. doi:10.1371/journal.pbio.3000019

Bolhuis JJ, Tattersall I, Chomsky N, Berwick RC. Language: UG or Not to Be, That Is the Question. PLoS Biol. 2015;13: e1002063. doi:10.1371/journal.pbio.1002063

Bolhuis JJ, Wynne CD. Can evolution explain how minds work? Nature. 2009;458: 832–833. doi:10.1038/458832a

Bowling DL, Fitch TW. Do Animal Communication Systems Have Phonemes? Trends in Cognitive Sciences. 2015;19: 555–557. doi:10.1016/j.tics.2015.08.011

Chomsky N. Some simple evo devo theses: how true might they be for language? In: Larson RK, Deprez V, Yamakido H, editors. The Evolution of Human Language. Cambridge: Cambridge University Press; 2010. pp. 45–62. doi:10.1017/CBO9780511817755.003

Clink DJ, Tasirin JS, Klinck H. Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. Hare J, editor. Current Zoology. 2020;66: 173–186. doi:10.1093/cz/zoz035

Corballis MC, Corballis MC. The Recursive Mind: The Origins of Human Language, Thought, and Civilization - Updated Edition. Princeton University Press; 2014. doi:10.1515/9781400851492

Crockford C, Herbinger I, Vigilant L, Boesch C. Wild Chimpanzees Produce Group-Specific Calls: a Case for Vocal Learning? Ethology. 2004;110: 221–243. doi:10.1111/j.1439-0310.2004.00968.x

de Boer B, Thompson B, Ravignani A, Boeckx C. Evolutionary Dynamics Do Not Motivate a Single-Mutant Theory of Human Language. Sci Rep. 2020;10: 451. doi:10.1038/s41598-019-57235-8

De Gregorio C, Valente D, Raimondi T, Torti V, Miaretsoa L, Friard O, et al. Categorical rhythms in a singing primate. Current Biology. 2021;31: R1379–R1380. doi:10.1016/j.cub.2021.09.032

Dufour V, Poulin N, Curé C, Sterck EH. Chimpanzee drumming: a spontaneous performance with characteristics of human musical drumming. Scientific reports. 2015;5: 11320. doi:10.1038/srep11320

Elfner E. Recursion in prosodic phrasing: evidence from Connemara Irish. Nat Lang Linguist Theory. 2015;33: 1169–1208. doi:10.1007/s11049-014-9281-5

Engesser S, Holub JL, O'Neill LG, Russell AF, Townsend SW. Chestnut-crowned babbler calls are composed of meaningless shared building blocks. PNAS. 2019; 201819513. doi:10.1073/pnas.1819513116

Engesser S, Ridley AR, Townsend SW. Meaningful call combinations and compositional processing in the southern pied babbler. Proceedings of the National Academy of Sciences of the United States of America. 2016; 201600970. doi:10.1073/pnas.1600970113

Estrada A, Garber PA, Gouveia S, Fernández-Llamazares Á, Ascensão F, Fuentes A, et al. Global importance of Indigenous Peoples, their lands, and knowledge systems for saving the world's primates from extinction. Sci Adv. 2022;8: eabn2927. doi:10.1126/sciadv.abn2927

Estrada A, Garber PA, Rylands AB, Roos C, Eduardo F-D, Fiore A, et al. Impending extinction crisis of the world's primates: Why primates matter. 2017; e1600946. doi:10.1126/sciadv.1600946

Ferrigno S, Cheyette SJ, Piantadosi ST, Cantlon JF. Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. Sci Adv. 2020;6: eaaz1002. doi:10.1126/sciadv.aaz1002

Fitch TW. Dendrophilia and the Evolution of Syntax. In: Boë L-J, Joël F, Pascal P, Jean-Luc S, editors. Origins of human language: Continuities and discontinuities with nonhuman primates. Berlin: Peter Lang; 2017. pp. 305–328.

Fitch TW. Empirical approaches to the study of language evolution. Psychonomic Bulletin & Review. 2017;24: 1–31. doi:10.3758/s13423-017-1236-5

Fuhrmann D, Ravignani A, Sarah M-P, Whiten A. Synchrony and motor mimicking in chimpanzee observational learning. Sci Rep-uk. 2015;4: srep05283. doi:10.1038/srep05283

Gabrić P. Overlooked evidence for semantic compositionality and signal reduction in wild chimpanzees (Pan troglodytes). Anim Cogn. 2021 [cited 9 Dec 2021]. doi:10.1007/s10071-021-01584-3

Gamba M, Torti V, Estienne V, Randrianarison R, Valente D, Rovara P, et al. The Indris Have Got Rhythm! Timing and Pitch Variation Of A Primate Song Examined Between Sexes And Age Classes. Frontiers in neuroscience. 2016;10: 249. doi:10.3389/fnins.2016.00249

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. Recursive syntactic pattern learning by songbirds. Nature. 2006;440: 1204–1207. doi:10.1038/nature04675

Hardus ME, Lameira AR, Singleton I, C M-B Helen, Knott CD, Ancrenaz M, et al. A description of the orangutan's vocal and sound repertoire, with a focus on geographic variation. In: Wich S, Setia MT, Utami SS, Schaik C, editors. Orangutans. New York: Oxford University Press; 2009. pp. 49–60.

Hattori Y, Tomonaga M. Rhythmic swaying induced by sound in chimpanzees (Pan troglodytes). Proc Natl Acad Sci USA. 2020;117: 936–942. doi:10.1073/pnas.1910318116

Hauser MD, Chomsky N, Fitch TW. The faculty of language: what is it, who has it, and how did it evolve? Science (New York, NY). 2002;298: 1569–1579. doi:10.1126/science.298.5598.1569

Hockett CF. The origin of speech. Scientific American. 1960;203: 89–96.

Hopkins WD, Taglialatela JP, Leavens DA. Chimpanzees differentially produce novel vocalizations to capture the attention of a human. Animal Behaviour. 2007;73: 281–286.

Idsardi WJ, Gallego ÁJ, Martin R. Why Is Phonology Different? No Recursion. In: Gallego ÁJ, Martin R, editors. Language, syntax, and the natural sciences. Cambridge University Press; 2018. pp. 212–223. Available: https://doi.org/10.1017/9781316591529

Jackendoff R, Pinker S. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). Cognition. 2005;97: 211–225. doi:10.1016/j.cognition.2005.04.006

Jackendoff R. Parallels and Nonparallels between Language and Music. Music Perception. 2009;26: 195–204. doi:10.1525/mp.2009.26.3.195

Jiang X, Long T, Cao W, Li J, Dehaene S, Wang L. Production of Supra-regular Spatial Sequences by Macaque Monkeys. Current Biology. 2018;0: 1851-1859.e4. doi:10.1016/j.cub.2018.04.047

Kabak B, Revithiadou A. An interface approach to prosodic word recursion. In: Grijzenhout J, Kabak B, editors. Phonological Domains. Berlin, New York: Mouton de Gruyter; 2009. pp. 105–134. doi:10.1515/9783110219234.2.105

Kershenbaum A, Bowles AE, Freeberg TM, Jin DZ, Lameira AR, Bohn K. Animal vocal sequences: not the Markov chains we thought they were. Proceedings Biological sciences / The Royal Society. 2014;281: 20141370. doi:10.1098/rspb.2014.1370

Koelsch S, Rohrmeier M, Torrecuso R, Jentschke S. Processing of hierarchical syntactic structure in music. Proceedings of the National Academy of Sciences of the United States of America. 2013;110: 15443–15448. doi:10.1073/pnas.1300272110

Lameira AR, Alexandre A, Gamba M, Nowak MG, Vicente R, Wich S. Orangutan information broadcast via consonant-like and vowel-like calls breaches mathematical models of linguistic evolution. Biol Lett. 2021;17: 20210302. doi:10.1098/rsbl.2021.0302

Lameira AR, Eerola T, Ravignani A. Coupled whole-body rhythmic entrainment between two chimpanzees. Sci Rep. 2019;9: 18914. doi:10.1038/s41598-019-55360-y

Lameira AR, Hardus ME, Bartlett AM, Shumaker RW, Wich SA, Menken SB. Speech-like rhythm in a voiced and voiceless orangutan call. PloS one. 2015;10: e116136. doi:10.1371/journal.pone.0116136

Lameira AR, Hardus ME, Kowalsky B, de Vries H, Spruijt BM, Sterck E, et al. Orangutan (Pongo spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension. Journal of the Acoustical Society of America. 2013;134: 1–11. doi:10.1121/1.4817929

Lameira AR, Hardus ME, Mielke A, Wich SA, Shumaker RW. Vocal fold control beyond the species-specific repertoire in an orang-utan. Scientific reports. 2016;6: 30315. doi:10.1038/srep30315

Lameira AR, Shumaker RW. Orangutans show active voicing through a membranophone. Sci Rep. 2019;9: 12289. doi:10.1038/s41598-019-48760-7

Lameira AR, Vicente R, Alexandre A, Gail C-S, Knott C, Wich S, et al. Proto-consonants were information-dense via identical bioacoustic tags to proto-vowels. Nature Human Behaviour. 2017;1: 0044. doi:10.1038/s41562-017-0044

Lameira AR, Wich S. Orangutan Long Call Degradation and Individuality Over Distance: A Playback Approach. International Journal of Primatology. 2008;29: 615–625. doi:10.1007/s10764-008-9253-x

Lameira AR. Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. Neuroscience & Biobehavioral Reviews. 2017;83: 429–439. doi:10.1016/j.neubiorev.2017.09.021

Lattenkamp EZ, Vernes SC. Vocal learning: a language-relevant trait in need of a broad cross-species approach. Curr Opin Behav Sci. 2018. doi:10.1016/j.cobeha.2018.04.007

Laurance WF, Useche CD, Rendeiro J, Kalka M, Bradshaw CJ, Sloan SP, et al. Averting biodiversity collapse in tropical forest protected areas. Nature. 2012;advance on. doi:10.1038/nature11318

Laurance WF. Does research help to safeguard protected areas? Trends in Ecology {& Evolution. 2013;28: 261–266. doi:10.1016/j.tree.2013.01.017

Liao DA, Brecht KF, Johnston M, Nieder A. Recursive sequence generation in crows. Sci Adv. 2022;8: eabq3356. doi:10.1126/sciadv.abq3356

Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, et al. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. Nature. 2013;498: 104–108. doi:10.1038/nature12173

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2020). Performance: Assessment of regression models performance. R package version 0.4.

Mandelbrot BB. FRACTAL ASPECTS OF THE ITERATION OF z →Λz(1- z) FOR COMPLEX Λ AND z. Annals of the New York Academy of Sciences. 1980;357: 249–259. doi:10.1111/j.1749-6632.1980.tb29690.x

Martins M. Distinctive signatures of recursion. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367: 2055–2064. doi:10.1098/rstb.2012.0097

Martins MD, Gingras B, Puig-Waldmueller E, Fitch WT. Cognitive representation of “musical fractals”: Processing hierarchy and recursion in the auditory domain. Cognition. 2017;161: 31–45. doi:10.1016/j.cognition.2017.01.001

Martins PT, Boeckx C. Language evolution and complexity considerations: The no half-Merge fallacy. PLoS Biol. 2019;17: e3000389. doi:10.1371/journal.pbio.3000389

Miyagawa S. Revisiting Fitch and Hauser’s Observation That Tamarin Monkeys Can Learn Combinations Based on Finite-State Grammar. Front Psychol. 2021;12: 772291. doi:10.3389/fpsyg.2021.772291

Mundry R, Sommer C. Discriminant function analysis with nonindependent data: consequences and an alternative. Animal Behaviour. 2007;74: 965–976. doi:10.1016/j.anbehav.2006.12.028

Nasukawa K, editor. Morpheme-internal recursion in phonology. Berlin ; Boston: De Gruyter Mouton; 2020.

Nasukawa K. Recursion in the lexical structure of morphemes. In: Oostendorp M van, Riemsdijk H van, editors. Representing Structure in Phonology and Syntax. DE GRUYTER; 2015. pp. 211–238. doi:10.1515/9781501502224-009

Patel AD. The evolutionary biology of musical rhythm: was Darwin wrong? PLoS biology. 2014;12: e1001821. doi:10.1371/journal.pbio.1001821

Pereira AS, Kavanagh E, Hobaiter C, Slocombe KE, Lameira AR. Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. Biol Lett. 2020;16: 20200232. doi:10.1098/rsbl.2020.0232

Pouw W, Paxton A, Harrison SJ, Dixon JA. Acoustic information about upper limb movement in voicing. Proc Natl Acad Sci USA. 2020; 202004163. doi:10.1073/pnas.2004163117

Quam R, Martínez I, Rosa M, Bonmatí A, Lorenzo C, de Ruiter DJ, et al. Early hominin auditory capacities {\textbar} Science Advances. Sci Adv. 2015;1: e1500355. doi:10.1126/sciadv.1500355

Rawski J, Idsardi W, Heinz J. Comment on “Nonadjacent dependency processing in monkeys, apes, and humans.” Sci Adv. 2021;7: eabg0455. doi:10.1126/sciadv.abg0455

Roeske TC, Tchernichovski O, Poeppel D, Jacoby N. Categorical Rhythms Are Shared between Songbirds and Humans. Current Biology. 2020;30: 3544-3555.e6. doi:10.1016/j.cub.2020.06.072

Russell JL, Joseph M, Hopkins WD, Taglialatela JP. Vocal learning of a communicative signal in captive chimpanzees, Pan troglodytes. Brain and Language. 2013;127: 520–525. doi:10.1016/j.bandl.2013.09.009

Sainburg T, Theilman B, Thielk M, Gentner TQ. Parallels in the sequential organization of birdsong and human speech. Nat Commun. 2019;10: 1–11. doi:10.1038/s41467-019-11605-y

Sharma N, Chimalakonda S. Learning Recursion from Music and Music from Recursion. 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT). Mumbai: IEEE; 2018. pp. 257–261. doi:10.1109/ICALT.2018.00066

Spillmann B, Dunkel LP, van Noordwijk MA, Amda RN, Lameira AR, Wich SA, et al. Acoustic properties of long calls given by flanged male orang-utans (Pongo pygmaeus wurmbii) reflect both individual identity and context. Ethology. 2010;116: 385–395.

Staes N, Sherwood CC, Wright K, de Manuel M, Guevara EE, Tomas M-B, et al. FOXP2 variation in great ape populations offers insight into the evolution of communication skills. Scientific reports. 2017;7: 258. doi:10.1038/s41598-017-16844-x

Suzuki T, Wheatcroft D, communications G-M. Experimental evidence for compositional syntax in bird calls. Nature communications. 2016.

Suzuki TN, Wheatcroft D, Griesser M. Call combinations in birds and the evolution of compositional syntax. PLoS Biol. 2018;16: e2006532. doi:10.1371/journal.pbio.2006532

Suzuki TN, Wheatcroft D, Griesser M. Wild Birds Use an Ordering Rule to Decode Novel Call Sequences. Current Biology. 2017;27: 2331-2336.e3. doi:10.1016/j.cub.2017.06.031

Taglialatela JP, Reamer L, Schapiro SJ, Hopkins WD. Social learning of a communicative signal in captive chimpanzees. Biology letters. 2012;8: 498–501. doi:10.1098/rsbl.2012.0113

Team R. R: A language and environment for statistical computing. 2013.

Vernes SC, Janik VM, Fitch WT, Slater PJB. Vocal learning in animals and humans. Phil Trans R Soc B. 2021;376: 20200234. doi:10.1098/rstb.2020.0234

Vogel I. Recursion in phonology? In: Botma B, Noske R, editors. Phonological Explorations. Berlin, Boston: DE GRUYTER; 2012. doi:10.1515/9783110295177.41

von Humboldt W. Über die Verschiedenheit des Menschlichen Sprachbaues und ihren Einfluss auf die geristige Entwickelung des Menschengeschlechts. Berlin: Königlichen Akademie der Wissenschaften; 1836.

Wang L, Uhrig L, Jarraya B, Dehaene S. Representation of Numerical and Sequential Patterns in Macaque and Human Brains. Current Biology. 2015;25: 1966–1974. doi:10.1016/j.cub.2015.06.035

Watson SK, Burkart JM, Schapiro SJ, Lambeth SP, Mueller JL, Townsend SW. Nonadjacent dependency processing in monkeys, apes, and humans. Sci Adv. 2020;6: eabb0725. doi:10.1126/sciadv.abb0725

Watson SK, Townsend SW, Schel AM, Wilke C, Wallace EK, Cheng L, et al. Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees. Current Biology. 2015;25: 495–499. doi:10.1016/j.cub.2014.12.032

Wich S, Swartz K, Hardus M, Lameira A, Stromberg E, Shumaker R. A case of spontaneous acquisition of a human sound by an orangutan. Primates. 2009;50: 56–64. doi:10.1007/s10329-008-0117-y

Wich SA, KRU T MICHAEL, Lameira AR, ALEXANDER N, NATASHA A, Bastian ML, et al. Call cultures in orang-utans? PloS one. 2012;7: e36180. doi:10.1371/journal.pone.0036180

Wich SA, Nunn CL. Do male "long-distance calls" function in mate defense? A comparative study of long-distance calls in primates. Behavioral Ecology and Sociobiology. 2002;52: 474–484. doi:10.1007/s00265-002-0541-8

Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2009.

Wilke C. ggridges. 2022. Available: https://CRAN.R-project.org/package=ggridges

ACKNOWLEDGEMENTS

We thank the Indonesian Ministry of Research and Technology, the Indonesian Ministry of Environment and Forestry, the Indonesian Ministry of Home Affairs, the Directorate General of Natural Resources and Ecosystem Conservation and the former Directorate General of Forest Protection and Nature Conservation for authorization to carry out research in Indonesia; the Universitas National for supporting the project and acting as sponsors and counter-partners; the Bornean Orangutan Survival Foundation and the MAWAS Programme in Palangkaraya for their support and permission to stay and work in the MAWAS Reserve. A.R.L. was supported by the UK Research & Innovation, Future Leaders Fellowship grant agreement number MR/T04229X/1.

AUTHOR CONTRIBUTIONS

A.R.L. conceived and designed the study. A.R.L. and M.E.H. collected data. A.R.L., A.R., T.R. and M.G. analysed data. A.R.L., M.E.H., A.R., T.R. and M.G. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to Adriano R. Lameira.

CHAPTER 3

Isochrony and rhythmic interaction in ape duetting

Proceedings of the Royal Society B (2023), *290*(1990), 20222244.

Image generated on DALL.E, an Artificial Intelligence capable of generating images from textual descriptions through synography. "Digital art of a lar gibbon playing the guitar"

Proceedings of the Royal Society B, *290*(1990), 20222244.

https://doi.org/10.1098/rspb.2022.2244

Isochrony and rhythmic interaction in ape duetting

Teresa Raimondi[1*¶], Giovanni Di Panfilo[1¶], Matteo Pasquali[1], Martina Zarantonello[1], Livio Favaro[1], Tommaso Savini[2#], Marco Gamba[1#], Andrea Ravignani[3,4#*]

[1] *Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy*

[2] *Conservation Ecology Program, King Mongkut University of Technology Thonburi, School of Bioresources & Technology, Bangkok, Thailand*

[3] *Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

[4] *Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Denmark*

[¶] *Co-first authors.*

[#] *Co-senior authors.*

[*] *Correspondence:*
Teresa Raimondi, *Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy; teresa.raimondi@unito.it*

Andrea Ravignani, *Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands; andrea.ravignani@mpi.nl*

ABSTRACT

How did rhythm originate in humans, and other species? One cross-cultural universal, frequently found in human music, is isochrony: when note onsets repeat regularly like the ticking of a clock. Another universal consists in synchrony, e.g., when individuals coordinate their notes so that they are sung at the same time. An approach to biomusicology focuses on similarities and differences across species, trying to build phylogenies of musical traits. Here we test for the presence of, and a link between, isochrony and synchrony in a non-human animal. We focus on the songs of one of the few singing primates, the lar gibbon (*Hylobates lar*), extracting temporal features from their solo songs and duets. We show that another ape exhibits one rhythmic feature at the core of human musicality: isochrony. We show that an enhanced call rate overall boosts isochrony, suggesting that respiratory physiological constraints play a role in determining the song's rhythmic structure. However, call rate alone cannot explain the flexible isochrony we witness. Isochrony is plastic and modulated depending on the context of emission: gibbons are more isochronous when duetting than singing solo. We present evidence for rhythmic interaction: we find statistical causality between one individual's note onsets and the cosinger's onsets, and a higher than chance degree of synchrony in the duets. Finally, we find a sex-specific trade-off between individual isochrony and synchrony. Gibbon's plasticity for isochrony and rhythmic overlap may suggest a potential shared selective pressure for interactive vocal displays in singing primates. This pressure may have convergently shaped human and gibbon musicality while acting on a common neural primate substrate. Beyond humans, singing primates are promising models to understand how music and, specifically, a sense of rhythm originated in the primate phylogeny.

INTRODUCTION

Rhythm permeates human life across physiological, behavioural, and social domains: heartbeats, neural oscillations, spoken language, and music are all built on precise rhythmic patterns, which act as building blocks of physiological and communicative processes[1]. This is true for humans, but how widespread is endogenous rhythm production in the animal kingdom[2]? The simplest, lowest-entropy rhythmic structure is *isochrony*, namely a pattern where time intervals between successive onsets of a signal all have *roughly equal durations*[2]. Isochronous rhythm can be produced at an individual level: the heartbeat, walking, and singing are some examples. Rhythm also plays a crucial role in interactive processes, as in human music and animal coordinated displays. Inter-individual interaction has an adaptive value in animal evolution[3] rhythm, as a tool, may help individuals to coordinate by predicting the interlocutor's turn and preparing and adapting their own. This can be mediated through two opposite strategies[3,4]. The first is synchrony, where the overlap between individuals' phonation is maximized; this happens, for example, in human joint singing. The second is turn-taking, where the overlap is minimized on the advantage of rapid exchange of short turns, as, for example, in speech.

The spontaneous production of isochronous vocal patterns seems relatively rare in other vertebrates; it has only been reported in two bird species[5,6], a bat[7], rock hyraxes[8], and one primate[9]. The empirical connection between isochrony and synchrony is even less clear because the two are often studied in isolation[10]. This lack of data strikingly contrasts with the hypothesis that a link between isochrony and synchrony, potentially already present in our hominoid ancestors' loud-calls, may be the mechanism that shaped our coordinated communication[4]. In fact, every human culture shows collective, pulse-based rhythmic singing and dancing displays, in which individuals entrain to an isochronous, predictable pulse, allowing group coordination[11]. This purported role of cooperative vocal interaction in enhancing rhythmic regularity has been detected not only in collective dance and music but also in vocalizations[12]: in humans, synchrony boosts isochrony.

Are humans the only apes capable of flexibly orchestrating isochrony and synchrony in their behaviour? To date, the joint presence of, and connection between, isochrony and synchrony in other mammals has never been found. Here we show the first evidence for isochrony, rhythmic flexibility mediated by co-singing, and synchrony in the vocalizations of an ape, the lar gibbon (*Hylobates lar* – **Fig1a**). Gibbons are closely related to humans and share with us an unusual form of vocal communication: the song (**Fig1b-c**). Primate song plays a crucial role in various social and sexual dynamics (*e.g.,* territorial defence, hierarchies and partnership assessment, courtship, social bonding, emotions sharing). Singing primates, in general, may thus represent a convenient animal model to unravel the origins and mechanisms shaping the evolution of speech and musicality[13,14]. Gibbons, and lar gibbons specifically, are ordinarily found in *ex-situ* contexts, making controlled collection of acoustic data possible. Similarities exist between gibbon songs and human music, for instance the collective context of emission of the song[15] and the association of ritualized locomotor displays to the song[16,17]. Because of this, gibbons may be good models for unravelling the biological origin of musicality in our *taxon*[16].

We searched for isochronous patterns in lar gibbons' songs, specifically in *female* and *male contributions to duet* (**Fig1b**) and *male solos* (**Fig1c**). We hypothesized that gibbons may flexibly deploy and modulate rhythmic isochrony based on social context[18,4,12]. We extracted the inter-onset-intervals (t_k), namely the duration of the interval separating the onset of a song element from the next one. Then, we calculated the ratios between two adjacent intervals ($r_k=t_k/(t_k+t_{k+1})$), analysing their distribution to detect the existence of categorical rhythms (density peaks on specific, small-integer ratios; e.g., a 1:1 peak represents isochrony)[5,9]. We then probed the biological substrates of isochrony in this species, testing the effect of biomechanical constraints, in terms of call rate, in explaining rhythmic patterns of gibbons' duet: calling fast is energetically expensive and a potential signal of signaller's quality[19]. Because duetting seems a coordinated display[18], we also investigated the role of interaction in shaping the songs' rhythmic structure. To do so, we first assessed the causality of one individual's phonation on the one of the cosinger. Then, as synchrony may underpin coordinated displays[4,14], we quantified the overlapped phonation between duetting females and males, predicting that shared advertisement purposes enhance cosinger's synchrony. Finally, we tested for a relationship between isochrony and synchrony in the two sexes. We predicted that isochrony and synchrony may be strictly linked, similarly to humans. We also expected that the sexual dimorphism in form and function of gibbons' song may translate into a sex-specific trade-off between isochrony and synchrony.

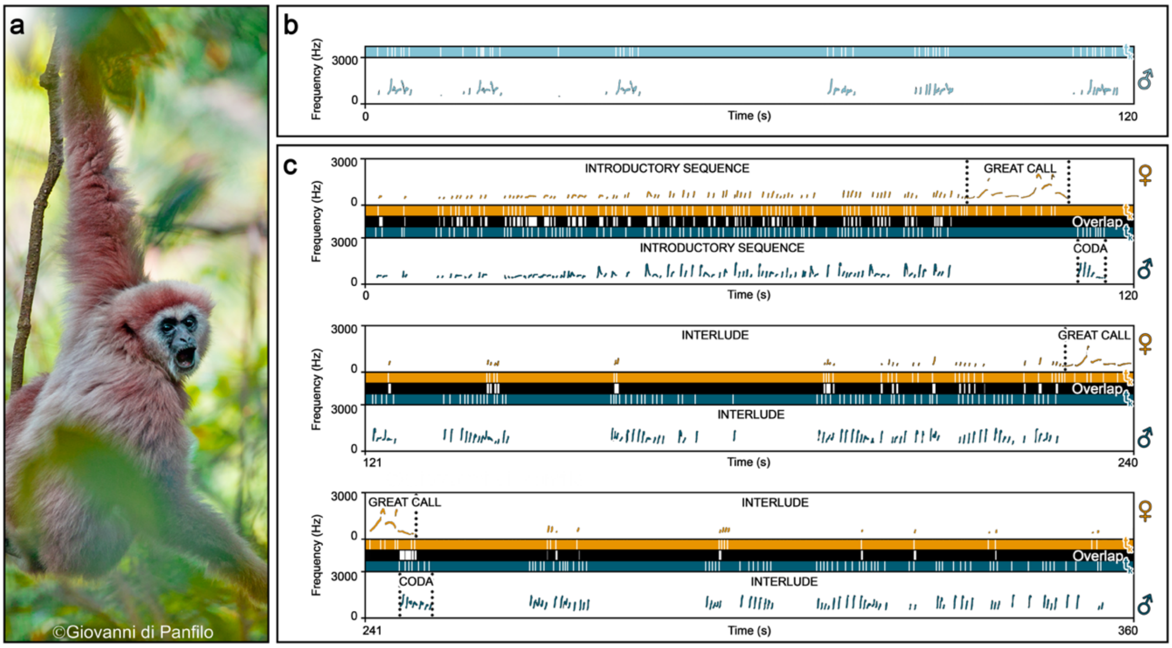


Figure 1 – Organization of lar gibbon's songs. (a) Male lar gibbon singing in the Huai Kha Khaeng Wildlife Sanctuary (Thailand). **(b)** Spectrogram and inter-onset-interval graphs of the *male solo*. The fundamental frequency is highlighted in light blue on the spectrogram. The coloured bar indicates inter-onset-intervals (t_k) of the solo singing male, where solid white lines on the bar represent the onsets. **(c)** Spectrograms and inter-onset-interval graphs of the reproductive couple's whole duet. The fundamental frequency of individuals' contributions is highlighted on the spectrogram in dark blue for the *male contribution to duet* and dark yellow for the *female contribution to duet*. The sections of the song are labelled in the upper part of the spectrogram and separated with dotted lines. Coloured bars indicate inter-onset intervals (t_k) of the contributions of each individual with white lines again corresponding to the onsets. Black bars turn white when the cosingers overlap. Notice how rhythmicity unfolds heterogeneously throughout the duet, alternating periods of higher and lower overlap. Note clusters onsets of the duetting gibbons influence each-other (see also **Fig4**), with introductory sequences and interludes showing higher levels of synchrony, while great calls and codas partly overlapped.

MATERIAL AND METHODS

Ethics statement

We complied with all the good practice protocols for field primatology of the International Primatological Society (2014) to work with wild species of primates. NRCT (National Research Council of Thailand) provided us with research permits for collecting non-invasive acoustic recordings of wild gibbons, while according to Italian law, no permits were required on captive groups. We never used playback stimuli to avoid any modification of gibbons' behaviour. All procedures follow the guidelines of the Association for the Study of Animal Behaviour for the care and use of animals for research activities (2021).

Animals and recordings

Six habituated gibbon familiar groups were followed, with a total of 12 individuals and 215 songs, and specifically 157 *female contributions to duet*, 157 *male contributions to duet* and 58 *male solos*. Four groups were inhabiting the forests of Huai Kha Khaeng Wildlife Sanctuary (Thailand), and two other *ex-situ* groups were living at the Cappeller faunistic park (Corvigliano, Italy) and the Falconara zoo park (Ancona, Italy). The contribution of each group and gibbon to the final dataset is provided in the Supplementary Information, **table S1**.

We recorded the animals between 6:00AM and 12:00AM using a a Sennheiser ME67 microphone connected to a solid-state digital audio recorder Tascam DR-100MKII (44.1KHz sampling rate). All vocalizations were recorded at 5-50 meters distance from the animals, aiming the microphone toward the individual vocalizing to maximize recording quality.

Acoustic analyses

We edited the songs using Praat 6.0.14 and saved them as WAV audio files [20]. Field notes and video recordings allowed us to recognize and separate individual contributions to each song; each annotated contribution was subsequently saved as a single Praat *TextGrid*, an object featuring onset and offset of each note. A computing cluster (OCCAM[21]) processed all 196768 vocal units via a custom Praat script and exported all onsets of song units from separate TextGrids into one .csv datasheet. We calculated the temporal interval between an onset and the next one, which defines an inter-onset interval (t_k). We focused on all $t_k \leq 5$ sec, as this value is typically hypothesized as upper limit for meter perception and performance in humans and there is no quantitative evidence about an upper threshold on other apes[22]. We calculated the tempo frequency as the inverse of peak values of t_k per song type (Hz). The ratio (r_k) was then calculated between a t_k and the next one, t_{k+1}, as $t_k/(t_k+t_{k+1})$.

Statistical analyses

Testing isochrony per song type. To test the significance of the peaks of the r_k density distribution falling in the vicinity of isochrony (corresponding to r_k=0.5) we followed the methodology in Roeske and colleagues[5] and De Gregorio and colleagues[9]. On-isochrony ratio ranges were centred on a 1:1 ratio, i.e. r_k=0.500, while the off-isochrony ones correspond to the peripheral range at the left and right sides of the on-isochrony range. Specifically, we took on/off-isochrony boundaries at 0.400, 0.440, 0.555, 0.600 (r_k values) and counted, per individual contribution and per song type (*female contribution to duet*, *male contribution to duet*, *male solo*), the number of r_k instances falling into the off-isochrony vs on-isochrony sectors of the curve. We calculated an *Isochrony Rate,* as the ratio between on- and off-isochrony observations per contribution.

Using a Generalized Linear Mixed Model (GLMM; *glmmTMB*[23] R package), we tested whether r_k observation counts differed by song type in interaction with the type of interval (*on* or *off-isochrony* interval). The response variable was the observations count (r_k), which followed a Poisson distribution, and the individual contribution code was used as a random factor (table **S3 Supplementary Information)**. To test the significance of the *full* model, we built a *null* model comprising only the random factors and compared the *full* and the *null* with a likelihood ratio test (Anova with “Chisq” argument[24]). We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons for each level of the explanatory variables with *emmeans*[25] package (p-values adjustment with Tukey method).

We then tested the effect of song types (fixed factor) on *Isochrony Rate* (response variable), which is the ratio between the number of on-isochrony and the number of off-isochrony observations per contribution, using a GLMM (*lme4* package[26]). The song code was entered as random factor. *Isochrony Rate* was log-transformed and followed a normal distribution. *Full* vs *null* models were compared with a likelihood ratio test (Anova with “Chisq” argument[24]). We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons for each level of the explanatory variables with *emmeans*[25] package (p-values adjustment with Tukey method).

***Call Rate* calculation and assessment of its effect on isochrony.** We calculated the *Call Rate* (number of onsets/10s) and recalculated *Isochrony Rate* on contributions chunks of 10 seconds. A GLMM tested whether song type (predictor) affected *Call Rate* values (response variable), with individual contribution code as random factor (table **S4 Supplementary Information**). The response variable, once transformed as 1/(*Call Rate* +1), fit a Gaussian distribution. We built (*lme4*[26] package) a *full* model and a *null* model comprising only the random factors. We compared them with a likelihood ratio test (Anova with “Chisq” argument[24]). We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons for each level of the explanatory variables with *emmeans*[25] package (p-values adjustment with Tukey method).

We tested the effect of *Call Rate* in interaction with song type (fixed factors) on *Isochrony Rate* (response variable) with a GLMM (*lme4*[26] package), where the individual contribution code was a random factor (table **S5 Supplementary Information**). The response variable, once log-transformed, followed a Gaussian distribution. We built and compared *full* and *null* models with a likelihood ratio test[24]. We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons and p-adjustment (Tukey method) with *emmeans*[25] and slope estimation with *lsmeans*[27] packages.

We wanted to test whether high vs low call rate affected isochrony. To test this hypothesis, we partitioned the *Call Rate* values between *low* and *high Call Rate,* depending on whether these were *below* or *above* the median of the specific song type. We counted how many r_k observations (response variable) for each song type (predictor) fell *on* or *off* peaks (predictor). We then built a GLMM model testing whether r_k counts could be explained not only by song type and *on* or *off* peaks – similarly to previous models – but also whether this purported relationship was modulated by an interaction with *low* vs. *high Call Rate* (*glmmTMB*[23] package; see table **S6 Supplementary Information**). We compared *full* and *null* model with a likelihood ratio test[24]. We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons and p-adjustment (Tukey method) with *emmeans*[25].

Assessing causality between male and female's contributions. We converted each individual contribution into binary series. We binned contributions into intervals of 10ms. We mapped the interval to a new time series containing a 1 when phonated, 0 otherwise. We performed a Granger causality test (GrangerTest function in R – lag = 5, order = 500), which assesses whether a time series is efficient in predicting another, between the female and male binary time series of the same song, in both directions (F→M, M→F). We then counted the number of occurrences where $p<0.001$, for F→M and M→F causality directions, to infer whether one of the two sexes affects more than the other the partner's timing of vocalizations. We limited the causality analysis to songs recorded in the wild to minimise the effect of potential external disturbances that often occur in captivity (e.g., other species' singing, visitors' presence).

Is duets' overlap different from chance? For every contribution recorded in the wild, we split songs' TextGrids into chunks of 3 minutes and calculated the overlapped phonation, summing the TextGrids' tiers with a custom Python[28] script. We then normalized it in two ways (figure **S7 Supplementary Information)**. First, we measured the rate (from 0, where no overlap is present, to 1, where all phonation is overlapped) of overlapped phonation on the total phonation of an individual contribution. Second, we also calculated the *Normalized Overlap*, by dividing the overlap duration in seconds by the total duration of phonation of the two cosingers (duration of phonation of the male + phonation of the female). We simulated random duets permuting each chunk against the others. For biological plausibility, the randomized permutation was carried out only between different sexes. Random duets were randomly paired songs either from real couples or from female-male non-coupled individuals. In simulations, we calculated the overlap and the *Normalized Overlap* as for real duets. We then tested whether the real overlap (REAL) was different from chance (SIMULATED permutations) with a GLMM (*lme4*[26] package*;* table **S8 Supplementary Information**). The response variable, the *Normalized Overlap*, was normally distributed. We used the group code of the two singers as the random factor. We compared *full* and *null* model with a likelihood ratio test[24]. We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons and p-adjustment (Tukey method) with *emmeans*[25].

Correlation between overlap and isochrony. We investigated the effect of overlap on the *Isochrony Rate* with a GLMM (table **S9 Supplementary Information**). Both *Isochrony Rate* and *Normalized Overlap* were calculated on the whole individual contribution. *Isochrony Rate* was the response variable, while *Normalized Overlap* and sex, in interaction, were fixed factors. We tested for the interaction between overlap and sex, expecting potentially dimorphic effects of those variables on isochrony. We used the song code as a random factor. The *Isochrony Rate* followed a beta distribution, so we used the *glmmTMB* package to build the *full* and *null* models and compared them with a likelihood ratio test[24]. We obtained p-values for each predictor using the R *summary* function and performed pairwise comparisons and p-adjustment (Tukey method) with *emmeans*[25]. We performed slope estimation using the *lsmeans*[27] package.

For every model, we assessed normality, homogeneity (via function provided by R. Mundry), and number of the residuals; the *performance*[29] package tested for collinearity among fixed factors. Effect plots were produced with *sjPlot*[68] package.

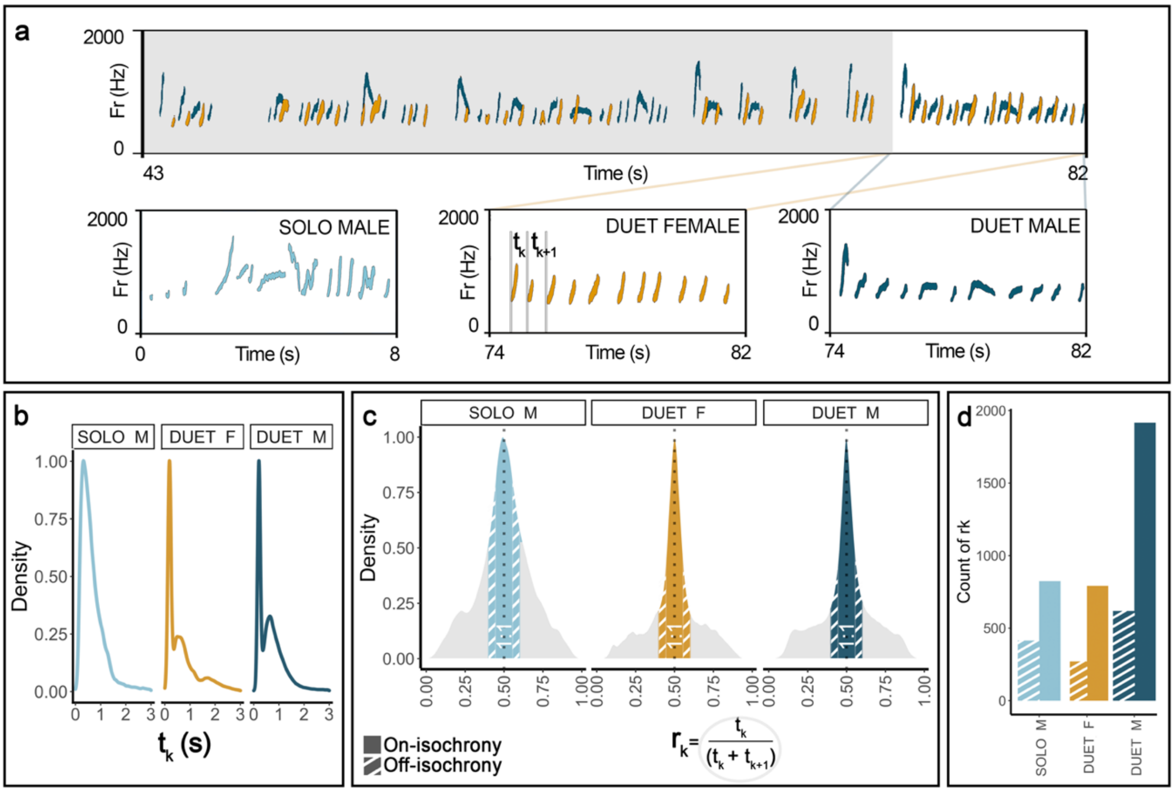


Figure 2 - Rhythmic features of the lar gibbon's songs. (a) Zoomed-in portions of the spectrograms of both duet contributions and male solo. The fundamental frequency of individual contributions is highlighted in light-blue for the *male solo*, in dark-blue for the *male contribution to duet* and dark-yellow for the *female contribution to duet* (both detailed in specific portions indicated by the coloured lines). A representation of the onsets (grey lines) and the relative inter-onset intervals (t_k) is reported on the spectrogram of the female contribution. **(b)** Probability density function representing the distributions of t_k per song type. t_k values are calculated on 12 adult individuals for a total of 372 individual contributions to songs. **(c)** Probability density function representing the distributions of rhythm ratios (r_k) per song type, suggesting a difference between *male solos* and *male* and *female contributions to duets*. Solid coloured sections of the curves indicate on-isochrony r_k ranges, striped sections indicate off-isochrony r_k ranges. **(d)** Histogram of the counts for r_k values falling within on-isochrony vs off-isochrony ranges of the density function (depicted in panel c), per song type. For all song types, on-isochrony observations (solid bars) are significantly more numerous than off-isochrony ones (striped).

RESULTS

Lar gibbons deploy isochrony in both duets and solos

The distribution of raw inter-onset intervals (t_k - **Fig2a**; figure **S2 Supplementary Information**) values shows two peaks for female and male duet contributions (**Fig2b**), with maxima corresponding to t_k values of 0.181s (5.525 Hz) and 0.503s (1.988 Hz) for the female duet, at 0.204s (4.902 Hz) and 0.637s (1.570 Hz) for the male duet. Male solos show only one peak at 0.322s (3.105 Hz). Density plots of r_k values (**Fig2c**) show a clear peak on the small integer-ratio of 1:1, namely isochrony. At least visually, there seems to be strong rhythmic regularity. We quantified this intuition by computing and comparing the count of r_k values falling into on/off-isochrony boundaries[9] (**Fig1d**). Through our model (*full* vs *null*: df=5, Chisq=24989.590, p<0.001; table **S3 Supplementary Information**), we found that the r_k peaks around isochrony were statistically significant in the three song types (*female contribution to duet*, off- vs.on-isochrony r_k count: estimate=-1.168, t-value=-95.220, p<0.001; *male contribution to duet*, off- vs.on-isochrony r_k count: estimate=-1.067, t-value=-106.792, p-value<0.001; *male solo* off- vs.on-isochrony r_k count: estimate=-0.580, t-value=-34.293, p<0.001), meaning that the number of r_k values falling into on-isochrony range of the curve were significantly higher than those falling in the off-isochrony range.

Call rate, isochrony and their relationship are context-dependent

We obtained average *Call Rate* values (±sd) of 0.768±0.360 Hz for the *female contribution to duet*, 0.774±0.315 Hz for the *male contribution to duet* and 0.943±0.468 Hz for the *male solo*. Our model testing the effect of song type on *Call Rate* values (*full* vs *null*: Chisq=65.469, df=2, p<0.001; table **S4 Supplementary Information**) showed no rate difference between female and male contributions to the duet, but significantly higher values of *Call Rate* for the male solo (*post-hoc* comparisons; *female contribution to duet* vs *male contribution to duet*: estimate= 0.007, z.ratio=1.495, p=0.293; *female contribution to duet* vs *male solo*: estimate= 0.046, z.ratio= 8.056, p<0.001; *male contribution to duet* vs *male solo*: estimate: 0.039, z ratio=7.028, p<0.001 – figure **S2** and table **S4 Supplementary Information** - *Call Rate* values showed in **Fig3a**). Our model (*full* vs *null*; Chisq=1569.052, df=5, p<0.001 – table **S5 Supplementary Information**) testing the effect of *Call Rate* in interaction with the song type on the *Isochrony Rate* (**Fig3f**) showed no differences between female and male contributions in a duet (*post-hoc* comparisons; fe*male contribution to duet* vs *male contribution to duet*: estimate:0.075, z.ratio=2.363, p-value=0.048; fe*male contribution to duet* vs *male solo*: estimate=0.383, z.ratio=9.442, p-value<0.001 – table **S5**

Supplementary Information). Conversely, we found significantly lower isochrony in the *male solo* compared to both the female and male duets (*post-hoc* comparisons; *male contribution to duet* vs *male solo*: estimate=0.309, z.ratio=7.742, p-value<0.001 – figure **S2** and table **S5 Supplementary Information**). Isochronous regularity differs between sexes in a duet - females are more isochronous -, and either sex has more isochrony in duets than males have in solos. Male solos have higher call rates but lower isochrony. These two results showed a context-dependent level of isochrony in the male, with the male singing more regularly when duetting than when singing alone; they also suggest that higher levels of isochrony *cannot* exclusively derive from higher call rates. Our model shows that call rate has a significant positive effect on isochrony; however, song type modulates this effect (i.e., the solo shows a significantly less steep slope, and thus significantly weaker effect, than the duet contributions of male and female - **Fig3c**, table **S4 Supplementary Information**). To summarize, male solos show higher call rates, lower isochrony and weaker correlation between these two variables than the other song types.

We tested whether high vs low levels of *Call Rate* modulated isochrony. All the isochronous peaks were still significant at both high and low values of *Call Rate* (table **S5 Supplementary Information**), confirming that *Call Rate* alone cannot explain isochrony levels.

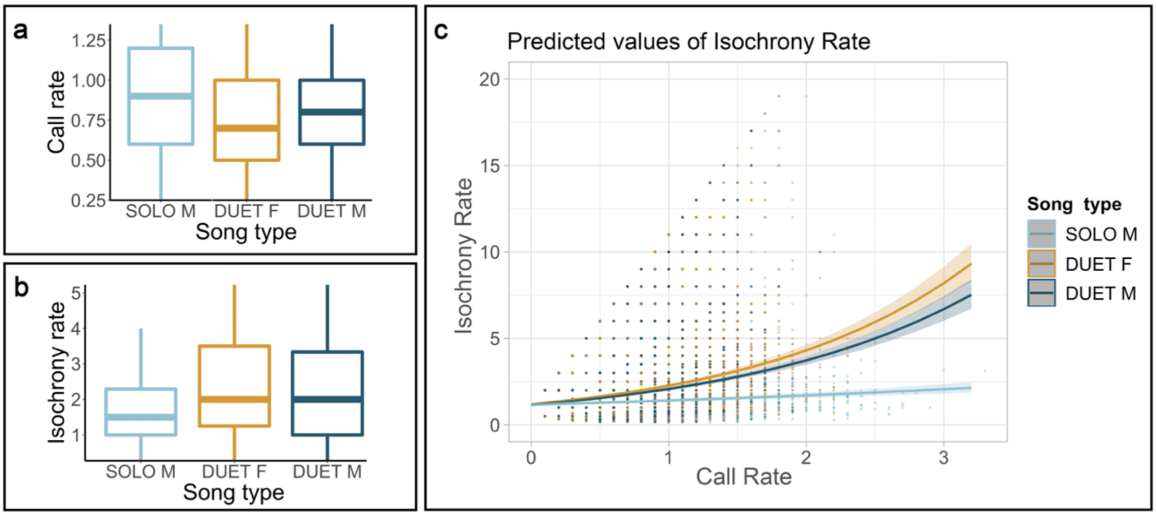


Figure 3 - *Call Rate* and isochrony covariation in lar gibbon's song. (a) Boxplots representing the *Call Rate* per song type, calculated on chunks of 10s for every individual contribution. The solo of the male shows significantly higher values of *Call Rate* than the male and female duets. Outliers corresponding to the 10% of higher and lower values of the variable are excluded from the plot. **(b)** Boxplots depicting the *Isochrony Rate* (on-isochrony counts / off-isochrony counts) per song type, calculated on chunks of 10s for every individual contribution. The *Isochrony Rate* does not differ between male and female duet contributions, while the

Isochrony Rate of the male singing solo is significantly lower than the one of the male and female in the duet. Outliers corresponding to the 10% of higher and lower values of the variable are excluded from the plot but included in the statistical computations. **(c)** Effect plot showing the predicted values derived from a GLMM looking at the effect of *Call Rate* on *Isochrony Rate*. The *female contribution to duet* and the *male contribution to duet* have a positive effect on *Isochrony Rate*, but both show a significantly steeper slope to the one of the *male solo*.

Male and female influence each other's phonation onsets

Since *Call Rate* was not sufficient to explain context-dependent isochrony levels, we investigated the presence of rhythmic interaction, by testing whether an individual's vocal rhythm affects its partner's. For this we used Granger Causality, a test probing whether future vocal onsets of one individual can be better predicted by considering past onsets of the cosinger (as opposed to past vocal onsets of the first individual). We found that 95% of the individual contributions to duets affect (i.e. Granger-caused at p<0.001– **Fig4**) the partner's phonation. Moreover, this effect emerged bidirectionally, both for the female on male's contribution (F→M in **Fig4**) and the male on female's contribution (M→F in **Fig4**). At the song level, 91.39% of the songs showed a reciprocal highly significant causality (p<0.001) between male and female contributions, 6.89% showed at least one significant contribution, 1.72% showed no significant contribution. This provided compelling evidence that the two sexes influence each other's onsets during the duet.

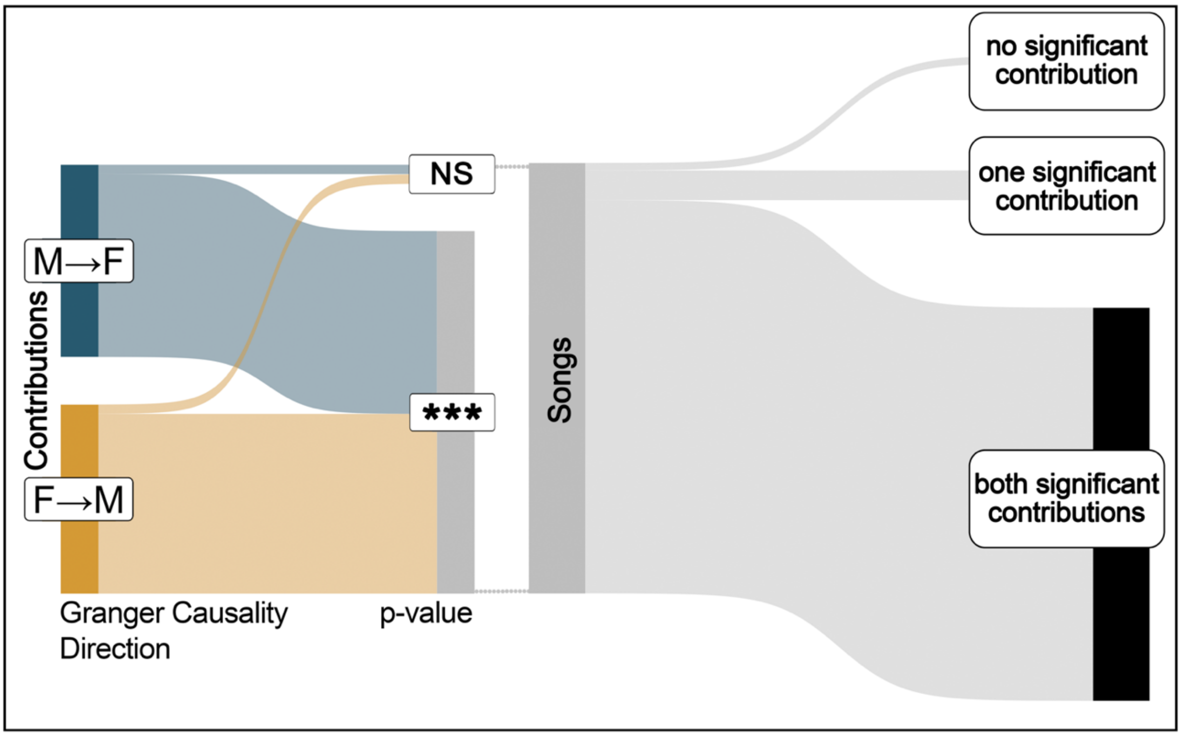


Figure 4 - Granger causality in lar gibbon's duet. Aggregated Sankey plots showing: on the *left*, the *proportion of contributions* having a significant causality (***=p<0.001) on the onsets of the cosinger, for both directions (male on female and female on male) and non-significant ones (NS); on the *right*, the *proportion of songs* showing a reciprocal causality (both male on female and female on male), having significant causality in only one-direction (male on female or female on male) or no significant causality at all .

Synchrony: duet overlap is higher than chance

Since we found a potential for rhythmic interaction, we then quantified overlap in the duets through two different metrics. The first metric, normalized on the individual total phonation, showed that females presented a rate (from 0 to 1) of overlap (mean±sd) of 0.177±0.145 while males of 0.158±0.131. The second metric, the *Normalized Overlap* (normalized on the sum of phonations of both cosingers – a rate from 0 to 0.5), in actual duets (*REAL;* mean±sd=0.079±0.037) assessed the degree of synchrony between actual singing partners (table **S8 Supplementary Information**). To have baselines and quantify randomness, the *Normalized Overlap* was also calculated on the randomly permuted opposite-sex individuals' contributions, for both existing couples (SIMULATED: real couple; mean±sd=0.060±0.037; table **S8 Supplementary Information**) and randomly paired individuals (SIMULATED: random couple; mean±sd=0.062±0.035; table **S8 Supplementary Information**). When testing the difference in overlap (response variable) between real duets and randomly simulated ones (**Fig5a**; GLMM, *full* vs *null*; Chisq=28.711, p<0.001; table **S9 Supplementary Information**), we found that real duets showed significantly higher overlap than both types of simulated ones (SIMULATED random couple – REAL estimate=-0.018; z-value=-2.416; p=0.037; SIMULATED real couple – REAL: estimate=-0.016; z-value= -5.347; p<0.001). The two types of simulated duets showed no difference in their rate of *Normalized Overlap*. These results indicate that the overlap, and thus synchrony, heard in actual duets is not perfect but significantly higher than chance.

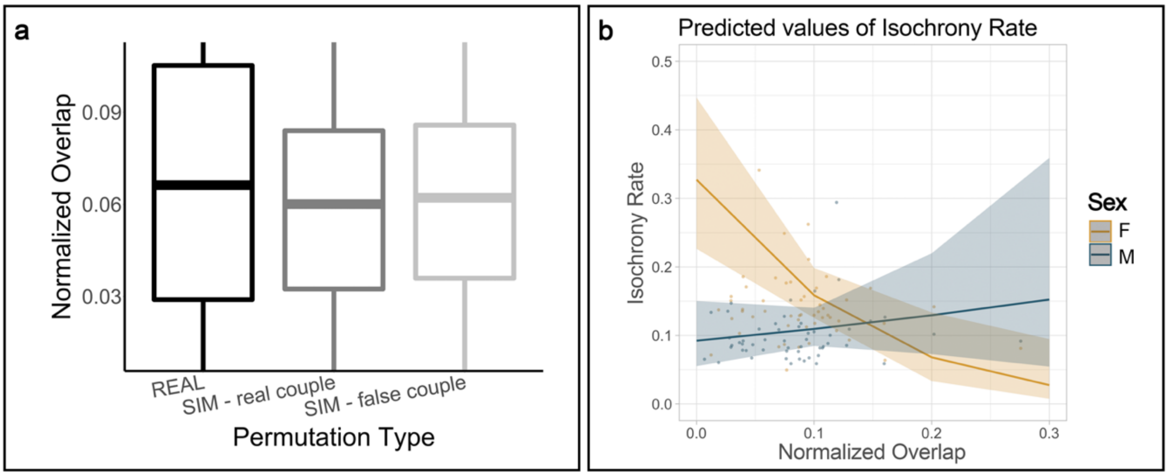


Figure 5 – Synchrony and isochrony covariation in lar gibbon's song. (a) Boxplots showing the amount of *Normalized Overlap* (sum of all durations of overlapping female-male phonation / sum of all durations of female-male phonation), measuring the degree of synchrony. *Normalized overlap* is a rate going from 0 (no overlap between the cosingers) to 0.5 (cosingers perfectly overlapped). The amount of overlap is calculated for the real duets (REAL) and simulated ones (SIM - two types of simulations: simulated duets of real couples and simulated duets with random couples composed of opposite-sex individuals from the sample). Real duets show significantly higher overlap than simulated ones, displaying that the animals are statistically more synchronous than expected by chance. **(b)** Effect plot showing the predicted values taken from the GLMM looking at the effect of *Normalized Overlap* on *Isochrony Rate*. The two regression lines show the sex-specific trends that link the two variables.

The effect of synchrony on isochrony is sex-specific

Our model looking at the effect of synchrony (*Normalized Overlap*) and sex on isochrony (*Isochrony Rate - full* vs *null*; Chisq=30.526, df=3, p<0.001; table **S10 Supplementary Information**) showed a significant effect of synchrony on the *Isochrony Rate* (estimate=-9.484, t-value=-2.416, p=0.001; table **S10 Supplementary Information**). In particular, we found a significant interaction between overlap and sex (estimate=11.383, t-value=-5.347, p<0.001; table **S10 Supplementary Information**). There was a significant negative effect for females (confidence interval: lower cl. = -15.26; upper cl. = -3.71; table **S10 Supplementary Information**) - which means higher isochrony corresponding to lower synchrony -, and a non-significant positive trend for males (confidence interval: lower cl. = -3.37; upper cl. = 7.17 – table **S10 Supplementary Information**, **Fig5b**): the effect of synchrony on isochrony is overall significant but sex-specific.

DISCUSSION

We provide compelling evidence for isochrony and synchrony in the song of an ape and elucidate the existing link between the two. Our results show how lar gibbon's songs are organized into highly isochronous, though plastic, rhythmic patterns; this isochrony is only partially explained by physiological constraints linked to call rate. Male inter-onset intervals become less variable and more isochronous from solo to duet, suggesting that the interaction between cosingers may shape the song's rhythmic structure. Furthermore, the timing of phonation of each individual in a duet is predicted by the timing of the cosinger's vocalizations, suggesting a potential rhythmic interaction. The exact coincidence of events in time can be quantified as the overlap between two individuals' vocalizations, namely their synchrony, which we showed is higher than chance across singers when compared with artificial dyads. Finally, we searched and found a link between synchrony and isochrony, specifically a significant statistical interaction indicating different trends across sexes. Females show a significant negative relationship between rhythm regularity and overlap, suggesting a potential trade-off between individual and group rhythms.

Isochrony and its role in coordinated displays

The production of isochronous rhythm in spontaneous animal vocalizations is rare and poorly studied[2,30]. Isochrony may have important biological functions: for instance, rock hyraxes producing more isochronous rhythmic patterns have higher fitness, in terms of reproductive success[8]. Here we provide evidence for above-chance occurrence of isochrony in an ape species. This occurs in both male and female songs. t_k peak values in lar gibbons are dimorphic and bimodal (similarly to another primate, *Indri indri*[9]), and differ between song types. Gibbons' isochrony transcends absolute temporal intervals: raw t_k values may vary but nonetheless lead to the same isochronous 1:1 ratio across sexes and song types. That is to say, as in human music, isochrony emerges even when tempo or note duration vary[11]. At the same time, t_k values are not randomly distributed, but mainly clustered around short intervals that produce the isochronous patterns: gibbon songs show two dominant tempi ($1/t_k$ peak values) in the duet contributions which are tentatively close to two human behaviors: music and speech. The slower tempo - at 1.988 Hz for the female and 1.570 Hz for the male - is closer to Western music, where 2Hz is the preferred tempo[31] and human locomotion[32]. The faster tempo we found - at 5.525 Hz for the female and 4.902 Hz for the male – falls in the range of speech, which relies on the association of vocal output and facial posture in the 3-7Hz range [33]. Similarly, macaques' "lip smacks"[34] and orangutans' "clicks/faux-speech"[35] have tempi close to speech. Speech-like rhythm in nonhuman primate vocalizations, often associated with facial expressions, may underpin – according to some - an ancestral original audiovisual rhythmic feature, still detectable in some branches of the primate phylogeny[34,35]. Gibbon tempi are numerically bounded by those of human music and speech. Beyond tempo, music and speech differ in rhythm regularity: music - like gibbon song - has it, while speech does not[36]. Human music and speech may have branched, says one hypothesis, from an ancestral "musical proto-language"[37,38]; our finding of human-compatible song tempi in another ape provides indirect support to this hypothesis. Looking ahead, the sound-facial expressions link in gibbons is an open topic for future research, and so is the similarity between their songs' tempo and human music or speech.

The use of isochrony to engage in group coordination, as for choruses and dance, was deemed a uniquely human trait[4]. Some features of gibbons' song are considered inherited and developmentally fixed[39,16]. Nonetheless, lar gibbons can instantly and flexibly adjust their vocal contribution; they can start and stop singing depending on subtle temporal and spectral variations in the co-singer utterances[18]. This may involve a duet coordination mechanism that potentially requires mutual learning and fine-tuned adaptation. Supporting this hypothesis, established pairs show higher organization and coordination when compared to new ones[40]. Our data also support this learning and adaptation hypothesis. We found a reciprocal causal link between the timing of the vocal emission of duetting individuals: the onset of one individual determines the onset of the cosinger. Moreover, males show higher isochrony in duets than in solos. We thus show that isochrony is deployed differently depending on social context, potentially fulfilling a need for tuning rhythmic structure via tempo regularity. Speculatively, if reciprocal rhythmic adaptation were a flexible and expensive learning process, it might represent an explicit energy commitment; this rhythm-advertised commitment may enhance the strength of the pair-bond and lower the risk of partner desertion[41].

Human experiments show that both music and speech display temporal regularity when people mutually share a coordination purpose[12]: regular rhythm may support cooperative interaction. Our results showing enhanced isochrony in gibbons duetting might imply that they share with us some signalling aims, such as cooperatively communicating and then being subjected to similar selective pressures. If so, they may also rely on homolog or analogue neural mechanisms that allow rhythm and coordination in humans.

Finally, physiological constraints in vocal emission may also partly explain our results. Solo songs are, on average, longer[15] and show a higher call rate than duets. It is likely that a longer vocal display and a high call rate demand higher energy investment and depend on breathing constraints[19]: vocal fatigue may determine less regular rhythmic patterns in prolonged bouts[8,43]. At the same time, a costly display like the song is an honest signal of the emitter quality[42], making solo songs highly subjected to sexual selection[44]. The flexibility of rhythmic structure may thus not only serve vocal interaction between cosingers, but also be the outcome of a trade-off between quality signaling and physiological constraints.

Overlap vs turn-taking: parallels with human music and speech

Across species, synchronous coordination stems from two primary purposes: either minimization of the overlap in favour of turn-taking or maximization of the overlap towards signal amplification[3,4]. We found a higher than chance rate of overlap in the songs. This finding dovetails with studies suggesting that overlap in the songs may mediate mechanisms such as mate attraction, anti-predatory purposes (e.g., gibbons[43], anurans[44]), and signalling the cohesion of an alliance in mammals (e.g., lemurs[45], dolphins[46]). In bottlenose dolphins, cooperative context enhances motor and vocal synchrony: a shared function may enhance synchrony and coordination[46]. Similarly, male and female gibbons jointly advertise their presence in the forest, and signal amplification through synchrony enhances transmission over long distances[14,47] which Merker and colleagues[4] hypothesized crucial in our ancestors' long-distance calls. Long-range signal transmission is needed to localize conspecifics and reduce the costs of territorial conflicts[48,49]. The adaptiveness of group signalling more than individual broadcasting, and the preponderance of synchrony rather than turn-taking, may be the result of a monogamous mating system[50]. The avoidance of overlap seems to be preeminent in morphologically dimorphic species subjected to higher levels of sexual selection[51]. Our results showing a higher than chance level of overlap in a monogamous, non-morphologically dimorphic species corroborate the idea that species subjected to weaker sexual selection[52] benefit from synchrony more than from turn-taking.

Human speech and music are notably different in the amount of overlap vs turn-taking occurring during interactions[53]. Both speech and music rely on rhythm. In conversations, speakers tend to minimize silence and avoid overlap: turn-taking is enhanced through spectro-temporal clues like prosody. Cross-cultural work showed that overlap avoidance and silence minimization, quantified via response latency in conversation, are the norm across languages[54]. Conversely, ensemble music often shows high levels of overlap: in many musical cultures (near-)synchrony is enhanced. Both mechanisms involve flexibility, adjustment and anticipation that allow the two domains to serve their different, specific adaptive functions. Our gibbons exhibit an average overlap rate of 16-18%, with higher than chance level of synchrony. It is difficult to compare our overlap rates (obtained by diving duration overlap by duration of phonation of the singer) to those of human conversations (often expressed as duration of silent gaps in seconds) or music. However, qualitatively, we would argue that gibbon overlap rates are well above what observed in human spoken conversations, and below those characteristics of several musical genres[54].

Sex-specific relationship between isochrony and synchrony

Our results highlighted a complex relationship linking synchrony and isochrony. We found an overall negative effect of synchrony on isochrony on the whole duet, and dimorphic effects of synchrony on isochrony. Synchrony did not enhance isochrony in males, but significantly decreased isochrony in females. This supports the idea that songs serve multiple functions that vary depending on dimorphic selection pressures acting on the duet contributions[55,56,13,57]. We propose that, for both sexes, isochrony may function to increase the redundancy of the signal during communication[2] and to mediate the overlap between the co-singers aiming at signal summation[4]. A trade-off mechanism may explain how females' isochrony decreases with higher synchronization: besides the interest in advertising the couple's mated status to neighbouring groups, females may broadcast their quality to higher quality males and female competitors. Previous work suggests that females' great calls are both an index of their physical condition[58] and the least overlapped part of the song[18]. Hence, we propose that female advertisement can be achieved by minimizing synchrony, hence changing the rhythmic structure of a bout, thus making it less predictable for their mate. In other words, the female would be the most rhythmically flexible[59] of the two in a race to the arms with its partner[60].

Some hypothesized human music originated via an ancestral form of loud call, shared with primates closely related to us[4,16]. Numerous extant ape species show vocal displays, e.g., loud calls and songs, that share features with human music and may also derive from a proto-musical loud call of our last common primate ancestor. Such features include loudness, for long-distance communication, tonal notes organized into higher-level structures, inked locomotor displays coordinated to vocal utterances and an inherited component[16]. After the divergence of the human branch, our species coordinated vocal displays specialized and acquired unique characteristics[61,62,16]: more flexibility in song structure improvisation and new conventions, a crucial role of learning shaping all those processes, the presence of a stable beat structuring rhythmic music structure. Our results offer indirect support to this hypothesis.

Isochrony is a "statistical universal" feature of human music, meaning it is found more often than not across diverse musical cultures[63,64]. Some argue it may be rare in other species, or even a human prerogative[61]. Others maintain that isochrony's function for group coordination may be the one feature exclusive to our species[4]. Indeed, whether endogenous isochrony can be driven by exogenous (e.g., environmental, social) factors, in other species, is unknown[2], even in the only non-human mammals (*Indri indri*) shown to date to have rhythmic categories[9]. Our results not only support the idea that a primate closely-related to us can spontaneously produce isochronous rhythms, but also that this rhythmic pattern can be shaped by exogenous *stimuli* linked to interaction. Endogenous and exogenous mechanisms coexist and are not mutually exclusive in determining rhythm[65]; consequently, rhythm should be investigated in terms of both underlying neural circuitry[66] and species-specific social and communicative adaptations[9]. From a comparative perspective, rhythmic capacities result from a mosaic of anatomic and functional changes throughout evolution. This perspective is in line with the *gradual audiomotor evolution* hypothesis, suggesting a coupling between the auditory and motor system in non-vocal-learners primate species[7]; the closer a primate is to humans, the more developed their rhythmic skills should be in terms of coordination and entrainment skills. Rhythmic capacities in gibbons support for the *gradual audiomotor evolution hypothesis*: our similar neural substrate, and potentially convergent pressures for singing, may grant similar rhythmic capacities. The rhythmic features we found are only a first step towards understanding potential rhythmic entrainment in gibbons.

Our results all together support the idea that evolution, by convergence or shared ancestral traits, may have selected isochrony as an adaptive trait for collective, coordinated vocal displays. Still, the role of isochrony in mediating synchronization in animal species remains widely unexplored. Now that individual building blocks of animal rhythms are increasingly studied, one can start probing their interaction. We suggest that elucidating the link between synchrony and isochrony may represent a substantial step in reconstructing musicality evolution and its meaning for our and other species.

DATA AVAILABILITY

Acoustic data used for this study have been deposited in *GitHub* at the following link:
https://github.com/sciabola/largib/blob/main/TOT_GIBBONI_DURATIONS_01c_TR_ratiotype_tk_tk%2B1_binomD%26S.csv.zip

ACKNOWLEDGMENTS

Our thanks go to NRCT (National Research Council of Thailand) for the research permits and DNP (Department of National Parks, Wildlife and Plants Conservation) people of Khao Nang Rum Research Station and to the CEG (Conservation Ecology Group) of the King's Mongkut University of Bangkok. We would also like to thank Parco Faunistico Cappeller (Cartigliano, Vicenza, Italy) and Parco Zoo Falconara (Falconara Marittima, Ancona, Italy) for giving us the opportunity to study their animals. We are grateful to Olivier Friard for his help during the scripting and performing of the quantitative statistical analysis using OCCAM.

AUTHOR CONTRIBUTIONS

MG, AR, TR, LF and GDP contributed to the conceptualization of the paper. Data collection was carried out by GDP, MP and MZ. Methodology was conceptualized by MG, AR, TR and LF. GDP, MP, MZ, and MG processed the sound files and the corresponding *TextGrids*. TR and MG performed the analyses on the data and produced data visualization. TR, AR and MG wrote the manuscript. All authors reviewed the manuscript before the submission. All steps were supervised by MG and AR.

COMPETING INTERESTS

The authors declare no competing interests.

MATERIALS & CORRESPONDENCE

Code and custom-written scripts are available from the corresponding author, T.R., upon request.

REFERENCES

1. Patel, A. D. Rhythm in Language and Music. *Annals of the New York Academy of Sciences* **999**, 140–143 (2003).
2. Ravignani, A. & Madison, G. The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology* **8**, 1–13 (2017).
3. de Reus, K. *et al.* Rhythm in dyadic interactions. *Phil. Trans. R. Soc. B* **376**, 20200337 (2021).
4. Merker, B. H., Madison, G. S. & Eckerdal Gamla Kyrkv, P. On the role and origin of isochrony in human rhythmic entrainment. **45(1)**, 4–17 (2009).
5. Roeske, T. C., Tchernichovski, O., Poeppel, D. & Jacoby, N. Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology* **30**, 3544-3555.e6 (2020).
6. Norton, P. & Scharff, C. “Bird Song Metronomics”: Isochronous Organization of Zebra Finch Song Rhythm. *Frontiers in Neuroscience* **10**, 309–309 (2016).
7. Burchardt, L. S., Norton, P., Behr, O., Scharff, C. & Knörnschild, M. General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*. *R. Soc. open sci.* **6**, 181076 (2019).
8. Demartsev, V., Haddas-Sasson, M., Ilany, A., Koren, L. & Geffen, E. Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *Journal of Animal Ecology* 1365-2656.13801 (2022) doi:10.1111/1365-2656.13801.
9. De Gregorio, C. *et al.* Categorical rhythms in a singing primate. *Current Biology* **31**, R1379–R1380 (2021).
10. Greenfield, M. D., Honing, H., Kotz, S. A. & Ravignani, A. Synchrony and rhythm interaction: from the brain to behavioural ecology. *Phil. Trans. R. Soc. B* **376**, 20200324 (2021).
11. Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci USA* **112**, 8987–8992 (2015).
12. Bowling, D. L., Herbst, C. T. & Fitch, W. T. Social Origins of Rhythm? Synchrony and Temporal Regularity in Human Vocalization. *PLoS ONE* **8**, e80402 (2013).
13. Clarke, E. & Reichard, U. H. The Syntax and Meaning of Wild Gibbon Songs. *PLoS ONE* 10 (2006).
14. Gamba, M. *et al.* The Indris Have Got Rhythm! Timing and Pitch Variation of a Primate Song Examined between Sexes and Age Classes. *Front. Neurosci.* **10**, (2016).
15. Raemaekers, J. J., Raemaekers, P. M. & Haimoff, E. H. Loud Calls of the Gibbon (*Hylobates lar*): Repertoire, Organisation and Context. *Behaviour* **91**, 146–189 (1984).
16. Geissmann, T. Gibbon Songs and Human Music from an Evolutionary Perspective. in *The Origins of Music* (eds. Brown, S., Merker, B. & Wallin, C.) (The MIT Press, 1999). doi:10.7551/mitpress/5190.003.0011.

17. Fan, P.-F. *et al.* Rhythmic displays of female gibbons offer insight into the origin of dance. *Scientific Reports* **6**, 34606 (2016).

18. Terleph, T. A., Malaivijitnond, S. & Reichard, U. H. Male white-handed gibbons flexibly time duet contributions. *Behavioral Ecology and Sociobiology* **72**, 1–13 (2018).

19. Catchpole, C. K. & Slater, P. J. B. *Bird Song: Biological Themes and Variations*. (Cambridge University Press, 2003).

20. Boersma, P. Praat, a system for doing phonetics by computer. *Glot International* **5:9/10**, 341–345 (2001).

21. Aldinucci, M. *et al.* OCCAM: a flexible, multi- purpose and extendable HPC cluster. *Journal of Physics: Conference Series* **898**, 082039 (2017).

22. London, J. *Hearing in time: psychological aspects of musical meter*. (Oxford University Press, 2004).

23. Brooks, M., E. *et al. glmmTMB* Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* **9**, 378 (2017).

24. Dobson, A. J. & Barnett, A. G. *An Introduction to Generalized Linear Models*. (Chapman and Hall/CRC, 2018). doi:10.1201/9781315182780.

25. Lenth, R. *emmeans*: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.0. (2020).

26. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [stat]* (2014).

27. Lenth, R. V. Least-Squares Means: The *R* Package **lsmeans**. *J. Stat. Soft.* **69**, (2016).

28. Van Rossum, G. & Drake, F. L. *Python/C Api Manual - Python 3*. (CreateSpace, 2009).

29. Lüdecke, D., Makowski, D., Waggoner, P. & Patil, I. *Performance*: Assessment of regression models performance. R package version 0.4. (2020).

30. Bouwer, F. L., Nityananda, V., Rouse, A. A. & ten Cate, C. Rhythmic abilities in humans and non-human animals: a review and recommendations from a methodological perspective. *Phil. Trans. R. Soc. B* **376**, 20200335 (2021).

31. Stevens, C. *ICMPC7: 7th International Conference on Music Perception & Cognition: conference proceedings*. (Causal Productions [for] AMPS, 2002).

32. G. Hamish MacDougall & Steven T. Moore. Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. **99**, 1164–1173 (2005).

33. Chandrasekaran, C. & Trubanova, A. The Natural Statistics of Audiovisual Speech. *PLoS Computational Biology* **5**, 18 (2009).

34. Ghazanfar, A. A. & Takahashi, D. Y. Facial Expressions and the Evolution of the Speech Rhythm. **26**, 12.

35. Lameira, A. R. *et al.* Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE* **10**, e116136 (2015).

36. Brown, S. & Weishaar, K. Speech is “heterometric”: The changing rhythms of speech. 4.

37. Ravignani, A. & Boer, B. de. Joint origins of speech and music: testing evolutionary hypotheses on modern

humans. *Semiotica* **2021**, 169–176 (2021).

38. Brown, S. A Joint Prosodic Origin of Language and Music. *Front. Psychol.* **8**, 1894 (2017).

39. Geissmann, T. Inheritance of Song Parameters in the Gibbon Song, Analysed in 2 Hybrid Gibbons (*Hylobates pileatus* X *H. lar*). 20 (1984).

40. Haimoff, E. H. Acoustic and organizational features of gibbon songs. *The Lesse Apes* (1984).

41. Wickler, W. Vocal dueting and the pair bond: I. coyness and partner commitment. A hypothesis. *Zeitschrift für Tierpsychologie* **52(2)**, 201–209 (1980).

42. Barelli, C., Mundry, R., Heistermann, M. & Hammerschmidt, K. Cues to Androgens and Quality in Male Gibbon Songs. *PLoS ONE* **8**, e82748 (2013).

43. Brockelman, W. Y., Reichard, U., Treesucon, U. & Raemaekers, J. J. Dispersal, pair formation and social structure in gibbons (*Hylobates lar*). *Behavioral Ecology and Sociobiology* **42**, 329–339 (1998).

44. Legett, H. D., Aihara, I. & Bernal, X. E. The dual benefits of synchronized mating signals in a Japanese treefrog: attracting mates and manipulating predators. *Phil. Trans. R. Soc. B* **376**, 20200340 (2021).

45. Torti, V., Gamba, M., Rabemananjara, Z. H. & Giacoma, C. The songs of the indris (Mammalia: Primates: Indridae): contextual variation in the long-distance calls of a lemur. *Italian Journal of Zoology* **80**, 596–607 (2013).

46. Moore, B. L., Connor, R. C., Allen, S. J., Krützen, M. & King, S. L. Acoustic coordination by allied male dolphins in a cooperative context. *Proc. R. Soc. B.* **287**, 20192944 (2020).

47. Rehberg-Besler, N., Doucet, S. M. & Mennill, D. J. Overlapping vocalizations produce far-reaching choruses: a test of the signal enhancement hypothesis. *BEHECO* arw176 (2016) doi:10.1093/beheco/arw176.

48. Bonadonna, G. *et al.* Behavioral correlates of extra-pair copulation in *Indri indri*. *Primates* **55**, 119–123 (2014).

49. Reichard, U. Extra-pair Copulations in a Monogamous Gibbon (*Hylobates lar*). *Ethology* **100**, 99–112 (1995).

50. Haimoff, E. H. Convergence in the duetting of monogamous Old World primates. *Journal of Human Evolution* **15**, 51–59 (1986).

51. Deputte, B. L. Duetting in male and female songs of the white-cheeked gibbon (*Hylobates concolor leucogenys*). in (Primate Communication).

52. Leigh, S. R. & Shea, B. T. Ontogeny and the evolution of adult body size dimorphism in apes. *Am. J. Primatol.* **36**, 37–60 (1995).

53. de Reus, K. *et al. Rhythm in dyadic interactions*. https://osf.io/9yrkv (2020) doi:10.31234/osf.io/9yrkv.

54. Stivers, T. *et al.* Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10587–10592 (2009).

55. Cowlishaw, G. Song Function in Gibbons. *Behav* **121**, 131–153 (1992).

56. Mitani, J. C. Male Gibbon (*Hylobates agilis*) Singing Behavior: Natural History, Song Variations and Function. *Ethology* **79**, 177–194 (1988).

57. Geissmann, T. Duet-splitting and the evolution of gibbon songs. *Biological Reviews of the Cambridge*

Philosophical Society **77**, 57–76 (2002).

58. Terleph, T. A., Malaivijitnond, S. & Reichard, U. H. Age related decline in female lar gibbon great call performance suggests that call features correlate with physical condition. *BMC Evolutionary Biology* **16**, 1–13 (2016).

59. De Gregorio, C. *et al.* Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Current Zoology* **65**, 89–97 (2019).

60. Greenfield, M. D. & Rolzen, I. Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. 3 (1993).

61. Fitch, W. T. The Evolution of Music in Comparative Perspective. *Annals of the New York Academy of Sciences* **1060**, 29–49 (2005).

62. Hauser, M. D. & McDermott, J. The evolution of the music faculty: a comparative perspective. *Nat Neurosci* **6**, 663–668 (2003).

63. Savage, P. E., Brown, S., Sakai, E. & Currie, T. E. Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci USA* **112**, 8987–8992 (2015).

64. Jacoby, N. & McDermott, J. H. Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology* **27**, 359–370 (2017).

65. Merker, B. Synchronous chorusing and the origins of music. *Musicae Scientiae* **3**, 59–73 (1999).

66. Honing, H., Bouwer, F. L., Prado, L. & Merchant, H. Rhesus Monkeys (*Macaca mulatta*) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis. *Frontiers in Neuroscience* **12**, (2018).

67. Merchant, H. & Honing, H. Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* **7**, (2014).

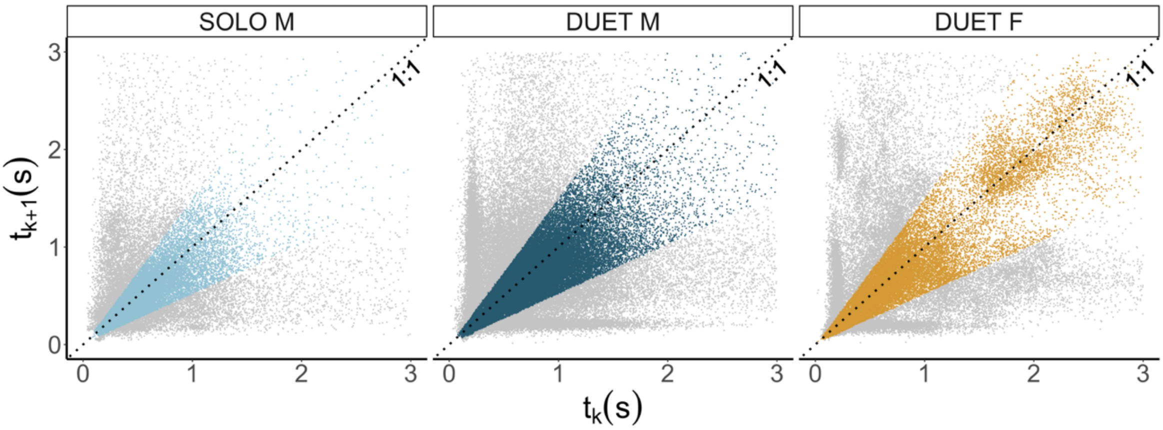
68. Lüdecke D. *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.11, https://CRAN.R-project.org/package=sjPlot (2022).

SUPPLEMENTARY INFORMATION

S1 - Group site, composition, and their contribution to the sample.

Group	Group members	N.Duet Contributions	N.Solo Contributions	Site
A	Adult Female	21		Huai Kha Khaeng Wildlife Sanctuary (Thailand)
	Adult Male	21	13	
	Subadult Male			
	Juvenile			
	Infant			
B	Adult Female	29		Huai Kha Khaeng Wildlife Sanctuary (Thailand)
	Adult Male	29	18	
	Subadult Male			
	Juvenile			
	Infant			
D	Adult Female	11		Huai Kha Khaeng Wildlife Sanctuary (Thailand)
	Adult Male	11	10	
	Subadult Male			
	Infant			
L	Adult Female	5		Huai Kha Khaeng Wildlife Sanctuary (Thailand)
	Adult Male	5	6	
	Juvenile			
	Infant			
P	Adult Female	66		Adult Falconara zoo park (Ancona, Italy)
	Adult Male	66		
	Juvenile			
	Infant			
Z	Adult Female	25		Cappeller Faunistic park (Corvigliano, Italy)
	Adult Male	25	11	
	Juvenile			
	Infant			
6 **Tot. groups**	12 **Tot. reproductive individuals**	314 **Tot. (male/female) duet contributions**	58 **Tot. (male) solo contributions**	372 **Tot. contributions**

S2 – Scatter-plots showing the values of one inter-onset interval (t_k - x axis) vs the successive interval (t_{k+1} - y axis) per song type (*male solo*, *male contribution to duet*, *female contribution to duet*). Coloured points are the observations falling inside the off/on boundaries of the isochronous ratio (1:1 – boundaries at 0.400 and 0.660, as in Roeske et al. 2020), grey ones are the observations falling outside isochrony boundaries. Dotted black lines represent the bisector, the 1:1 ratio ($r_k = 0.5$). Notice how the *female contribution to duet* shows clusters of observations around isochrony even at higher values of t_k. Conversely, the *male solo* visually displays shorter t_k values (higher call rate – tested in **S4**). Overall, as shown by **Fig2b** and **S3**, shorter t_k are the most frequent and fall around isochrony in the three song types. Call rate has an overall positive effect on isochrony, potentially depending on respiration constraints, but the relationship between the two variables is weaker in the *male solo* (**S5**) than in duet contributions. For this reason, call rate alone is not sufficient to explain the presence of isochrony across the three song types.



S3 - Summary and details of the GLMM and pairwise *post-hoc* test searching for the effect of r_k interval type (on/off-isochrony) and song type on the observation count.

a Summary of the *full* model - Influence of the fixed factors (Song type: *FEMALE DUET- MALE DUET - MALE SOLO*; R_k interval: ON-ISOCHRONY - OFF-ISOCHRONY) on the count of the observations. The identification code of the contribution is the random factor (contribution ID).

b Results for the pairwise *post-hoc* test on the interaction Song type * r_k interval.

---

a

Generalized Linear Mixed Model
Family = Poisson
Count(r_k obs) ~ Song type * r_k interval + (contribution ID)
Full vs Null (Chisq: 24989.590; df=5, p<0.001)

Variable	**Estimate**	**SE**	**z value**	**p-value (z)**
(Intercept)	3.660	0.091	[a]	[a]
Song Type (DUET M) [b,c]	0.363	0.128	2.83	0.005
Song Type (MALE SOLO) [b,c]	0.212	0.176	1.21	0.227
R_k interval (ON) [b,c]	1.168	0.012	95.22	<0.001
Song Type (DUET M) : R_k interval (ON) [b,c]	-0.101	0.016	-6.41	<0.001

Song Type (MALE SOLO) : R_k interval (ON) [b,c]	-0.588	0.021	-28.15	<0.001

[a] Not shown as not having a meaningful interpretation
[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.
[c] Reference categories: “Song Type (FEMALE DUET)”, “R_k interval (OFF)”

b

Post-hoc comparisons

Emmeans

Song Type	R_k int	emmean	SE	df	lower.CL	upper.CL
FEMALE DUET	**OFF**	3.66	0.091	695	3.48	3.84
DUET M	**OFF**	4.02	0.090	695	3.85	4.20
MALE SOLO	**OFF**	3.87	0.151	695	3.58	4.17
FEMALE DUET	**ON**	4.83	0.090	695	4.65	5.01
DUET M	**ON**	5.09	0.090	695	4.91	5.27
MALE SOLO	**ON**	4.45	0.151	695	4.16	4.75

Contrasts

Contrast	Estimate	SE	df	t.ratio	p-value
FEMALE DUET OFF - MALE DUETOFF	-0.363	0.128	695	-2.833	0.053
FEMALE DUET OFF - MALE SOLO OFF	-0.213	0.176	695	-1.208	0.833
FEMALE DUET OFF - FEMALE DUET ON	-1.168	0.012	695	-95.220	<0.001
FEMALE DUET OFF - MALE DUETON	-1.429	0.128	695	-11.183	<0.001
FEMALE DUET OFF - MALE SOLO ON	-0.793	0.176	695	-4.509	<0.001
MALE DUETOFF - MALE SOLO OFF	0.150	0.176	695	0.853	0.957
MALE DUETOFF - FEMALE DUET ON	-0.806	0.128	695	-6.308	<0.001
MALE DUETOFF - MALE DUETON	-1.067	0.010	695	-106.792	<0.001
MALE DUETOFF - MALE SOLO ON	-0.430	0.176	695	-2.447	0.142
MALE SOLO OFF - FEMALE DUET ON	-0.956	0.177	695	-5.436	<0.001
MALE SOLO OFF - MALE DUETON	-1.217	0.176	695	-6.924	<0.001
MALE SOLO OFF - MALE SOLO ON	-0.580	0.017	695	-34.293	<0.001
FEMALE DUET ON - MALE DUETON	-0.261	0.127	695	-2.049	0.316

FEMALE DUET ON - MALE SOLO ON	0.376	0.175	695	2.140	0.268
MALE DUETON - MALE SOLO ON	0.637	0.175	695	3.629	0.004

S4 - Summary and details of the GLMM and pairwise *post-hoc* test searching for the effect of song type on the call rate (number of onsets/10s).

a Summary of the *full* model - Influence of the fixed factors (Song type: *FEMALE DUET- MALE DUET - MALE SOLO*) on call rate. The call rate was transformed as 1/(call rate+1), so the estimate values correspond to the inverse of call rate values. The identification code of the individual contribution is the random factor (contribution ID).

b Results for the pairwise *post-hoc* test on the Song type fixed effect.

---

a

Generalized Linear Mixed Model
Family = Normal
Call rate ~ Song type + (contribution ID)
Full vs Null (df=2; Chisq: 65.469; p<0.001)

Variable	Estimate	SE	t value	p-value (>ltl)
(Intercept)	0.594	0.003	[a]	[a]
Song Type (DUET M) [b,c]	-0.007	0.004	548.783	0.136
Song Type (MALE SOLO) [b,c]	-0.046	0.006	600.543	<0.001

[a] Not shown as not having a meaningful interpretation
[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.
[c] Reference categories: “Song Type (FEMALE DUET)”

b

***Post-hoc* comparisons**

Contrasts

Contrast	Estimate	SE	df	z.ratio	p-value
FEMALE DUET - MALE DUET	0.007	0.004	2	1.495	0.293
FEMALE DUET - MALE SOLO	0.046	0.005	2	8.056	<0.001
MALE DUET - MALE SOLO	0.039	0.006	2	7.028	<0.001

S5 – Summary and details of the GLMM and pairwise *post-hoc* test searching for the effect of Call rate in interaction with Song type on *Isochrony rate* (number of on-isochrony r_ks/number of off-isochrony r_ks per contribution).

a Summary of the *full* model - Influence of the fixed factors (Call rate: numeric; Song type: *FEMALE DUET- MALE DUET - MALE SOLO*) on the isochrony rate. The identification code of the individual contribution is the random factor (contribution ID).

b Results for the pairwise *post-hoc* test on the interaction Song type * Call rate.

a

Generalized Linear Mixed Model
Family = Normal
Isochrony rate ~ Call rate * Song type + (contribution ID)
Full vs Null (df=5; Chisq: 1569.052; p<0.001)

Variable	Estimate	SE	t value	p-value (>ltl)
(Intercept)	0.173	0.029	[a]	[a]
Call rate	0.643	0.643	28.235	<0.001
Song Type (DUET M) [b,c]	-0.029	-0.029	-0.720	0.472
Song Type (MALE SOLO) [b,c]	-0.022	-0.022	-0.435	0.663
Call rate : Song Type (DUET M) [b,c]	-0.058	-0.058	-1.832	0.067
Call rate : Song Type (MALE SOLO) [b,c]	-0.453	-0.453	-12.288	<0.001

[a] Not shown as not having a meaningful interpretation
[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.
[c] Reference categories: “Song Type (FEMALE DUET)”

b

***Post-hoc* comparisons**

Contrasts (Song type)

Song type	Estimate	SE	df	z.ratio	p-value
FEMALE DUET - DUET M	0.075	0.032	-	2.363	0.048
FEMALE DUET – MALE SOLO	0.383	0.041	-	9.442	<0.001
MALE DUET– MALE SOLO	0.309	0.309	-	7.742	<0.001

Slope estimates (interaction Song type * Call rate)

Song type * call rate	Call rate trend	SE	df	Asymp.LCL	Asym.UCL
FEMALE DUET * call rate	0.643	0.023	-	0.599	0.688
MALE DUET* call rate	0.585	0.0220	-	0.542	0.628
MALE SOLO * call rate	0.190	0.0290	-	0.134	0.247

Contrasts (interaction Song type * Call rate)

Contrast	Estimate	SE	df	z.ratio	p-value
FEMALE DUET * Call rate - MALE DUET* Call rate	0.058	0.0317	-	1.832	0.1591
FEMALE DUET * Call rate - MALE SOLO * Call rate	0.453	0.0368	-	12.288	<0.001
MALE DUET* Call rate - MALE SOLO * Call rate	0.395	0.0364	-	10.861	<0.001

S6 – Summary and details of the GLMM and pairwise *post-hoc* test searching for the effect of r_k interval type (on/off-isochrony), song type and Call rate code (high/low) on the observation count. **a** Summary of the *full* model - Influence of the fixed factors (Song type: *FEMALE DUET- MALE DUET* - *MALE SOLO*; R_k interval: ON-ISOCHRONY - OFF-ISOCHRONY; Call rate code: HIGH - LOW) on the count of the r_k observations. The identification code of the individual contribution is the random factor (contribution ID). **b** Results for the pairwise *post-hoc* test on the interaction Song type * r_k interval * Call rate code.

---

a

Generalized Linear Mixed Model
Family = Poisson
Count(r_k obs) ~ Song type * r_k interval * Call rate code (high/low) + (contribution ID)
Full vs Null (df=11; Chisq: 42652.980; p<0.001)

Variable	Estimate	SE	z.value	Pr(>lzl)
(Intercept)	0.446	0.020	[a]	[a]
Song Type (DUET M) [b,c]	-0.056	0.027	-2.05	0.040
Song Type (MALE SOLO) [b,c]	0.339	0.340	9.95	<0.001
Type (ON) [b,c]	1.196	0.014	83.51	<0.001
Call rate code (low)	-0.885	0.022	-39.48	<0.001
Song type (DUET M) * Type (ON) [b,c]	-0.141	0.019	-7.40	<0.001
Song type (MALE SOLO) * Type (ON) [b,c]	-0.621	0.024	-25.56	<0.001
Song type (DUET M) *Call rate code (low)	0.203	0.028	7.23	<0.001
Song type (MALE SOLO) *Call rate code (low)	-0.004	0.036	-0.10	0.919
Type (ON) * Call rate code(low)	-0.159	0.026	-6.21	<0.001
Song type (DUET M) * Type (ON) * Call rate code (low)	0.125	0.322	3.88	<0.001
Song type (MALE SOLO) * Type (ON) * Call rate code(low)	0.151	0.043	3.49	<0.001

[a] Not shown as not having a meaningful interpretation
[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.
[c] Reference categories: "Song Type (FEMALE DUET)", "R_k interval (OFF)", "Call rate (high)"

---

b

***Post-hoc* comparisons**

Emmeans

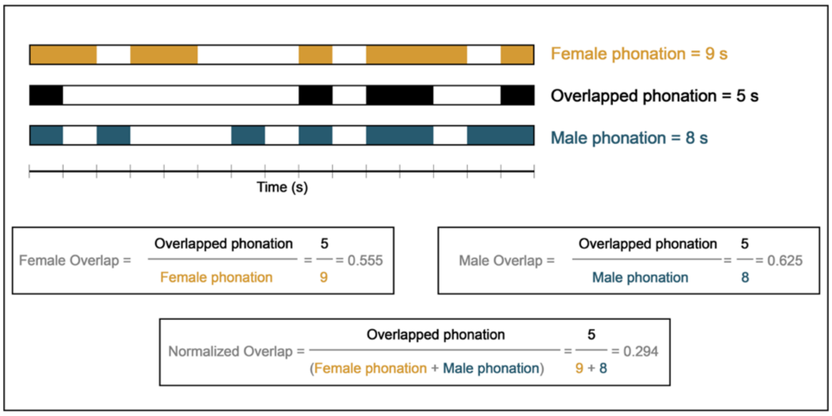
Song type	Rk int	Call rate code	emmean	SE	df	lower.CL	upper.CL

FEMALE DUET	**OFF**	**High**	0.446	0.020	49261	0.407	0.485
DUET M	**OFF**	**High**	0.390	0.019	49261	0.354	0.423
MALE SOLO	**OFF**	**High**	0.785	0.028	49261	0.731	0.839
FEMALE DUET	**ON**	**High**	1.642	0.017	49261	1.609	1.675
DUET M	**ON**	**High**	1.445	0.016	49261	1.413	1.477
MALE SOLO	**ON**	**High**	1.360	0.026	49261	1.310	1.410
FEMALE DUET	**OFF**	**Low**	-0.439	0.024	49261	-0.486	-0.392
DUET M	**OFF**	**Low**	-0.291	0.020	49261	-0.330	-0.253
MALE SOLO	**OFF**	**Low**	-0.104	0.032	49261	-0.167	-0.041
FEMALE DUET	**ON**	**Low**	0.598	0.019	49261	0.560	0.634
DUET M	**ON**	**Low**	0.729	0.017	49261	0.696	0.762
MALE SOLO	**ON**	**Low**	0.463	0.028	49261	0.408	0.518

Contrasts
(only contrasts regarding peaks-test are shown)

Contrast	**Estimate**	**SE**	**df**	**z.ratio**	**p-value**
FEMALE DUET OFF high - FEMALE DUET ON high	-1.196	0.014	49261	-83.511	<.0001
MALE DUETOFF high - MALE DUETON high	-1.054	0.013	49261	-83.482	<.0001
MALE SOLO OFF high - MALE SOLO ON low	0.322	0.024	49261	13.374	<.0001

S7 – Graphical representation of the calculation of the two overlap metrics. The first is normalized on the phonation of the singer, while the second, the *Normalized Overlap*, on the sum of the phonations of the cosingers.



S8 – Table summarizing the rate of *Normalized Overlap* (duration of the overlap/summed duration of the phonation of the two individuals' contributions) per REAL DUETS and SIMULATED ones (real couples but simulated duets or random couples simulated duets). Notice that the overlap rate goes from 0, where no overlap is observed, to 0.5, where the two cosingers completely overlap.

Type	Min	1st Qu	Median	Mean	3rd Qu	Max	Sd
REAL DUETS	0.001	0.029	0.066	0.079	0.106	0.290	0.037
SIMULATED - Real couples	0.000	0.032	0.060	0.060	0.084	0.293	0.037
SIMULATED - Random couples	0.000	0.036	0.062	0.062	0.086	0.265	0.035

S9 - Summary and details of the GLMM and pairwise *post-hoc* test testing the effect of the type of coupled contribution (real duets – simulated duets) on Synchrony.

a Summary of the *full* model - Influence of the fixed factors (Permutation Type: REAL DUETS - SIMULATED: real couple - SIMULATED: random couple) on Synchrony (*Normalized Overlap*). The identification code of the group is the random factor (Group ID).

b Results for the pairwise *post-hoc* test on the fixed factor Permutation Type.

a

Generalized Linear Mixed Model
Family = Normal
Overlap ~ Permutation Type + (Group ID)
Full vs Null (Chisq: 28.711, p<0.001)

Variable	Estimate	SE	Df	t value	p-value
(Intercept)	8.253e-02	6.055e-03	1.415e+01	[a]	[a]
Perm Type (SIM: random couple) [b,c]	-1.792e-02	7.413e-03	1.145e+01	-2.416	0.033
Perm Type (SIM: real couple)[b,c]	-1.652e-02	3.090e-03	1.812e+04	-5.347	<0.001

[a] Not shown as not having a meaningful interpretation
[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.
[c] Reference category: "Permutation Type (REAL)"

b

***Post-hoc* comparisons**

Emmeans

Permutation type	emmean	SE	df	Lower.CL	Upper.CL
REAL	0.082	0.006	14.11	0.070	0.095
SIM: false couple	0.065	0.005	7.91	0.055	0.074
SIM: real couple	0.066	0.005	8.15	0.054	0.078

Contrasts

Contrast	Estimate	SE	z.value	p-value

SIM: random couple – REAL	-0.018	0.007	-2.416	0.037
SIM: real couple – REAL	-0.016	0.003	-5.347	<0.001
SIM: real couple – SIM: random couple	0.001	0.007	0.205	0.975

S10 - Summary and details of the GLMM and pairwise *post-hoc* test searching for the effect of Synchrony, quantified as the *Normalized Overlap* between individual contributions, and Sex of the emitter, on *Isochrony Rate*.

a Summary of the *full* model - Influence of the fixed factors (Synchrony=*Normalized Overlap*; sex: F - M) on *Isochrony Rate*. The identification code of the song is the random factor (Song ID).

b Results for the pairwise *post-hoc* test on the interaction between Synchrony and Sex.

a

Generalized Linear Mixed Model
Family = Beta
Isochrony rate ~ Synchrony*sex + (Song ID)
Full vs Null (Chisq: 30.526, p<0.001)

Variable	Estimate	SE	Df	t value	p-value
(Intercept)	-0.720	0.256	-2.813	[a]	[a]
Overlap	-9.484	2.913	-3.256	-2.416	0.001
Sex (Male)[b,c]	-1.565	0.307	-5.094	-5.347	<0.001
Overlap*Sex	11.383	3.275	3.475	-5.347	<0.001

[a] Not shown as not having a meaningful interpretation

[b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor.

[c] Reference category: “Sex (Female)”

b

Slope Estimates and Confidence Intervals per Sex

Sex	Estimate	SE	df	Confidence interval
Female	-9.48	2.91	102	-15.26 - -3.71
Male	1.90	2.66	102	-3.37 - 7.17

CHAPTER 4

Isochronous singing in three crested gibbons' species (*Nomascus* spp.)

In publication at *Current Zoology* (July 2023)

Image generated on DALL.E, an Artificial Intelligence capable of generating images from textual descriptions through synography. "Digital art of a *Nomascus* gibbon playing the violin"

In publication at *Current Zoology,* July 2023

Isochronous singing in three crested gibbons' species (*Nomascus* spp.)

Chiara De Gregorio[1,$], **Teresa Raimondi**[1,$,*], Valeria Bevilacqua[2,$], Chiara Pertosa[1,*], Daria Valente[1], Filippo Carugati[1], Francesca Bandoli[3], Livio Favaro[1], Brice Lefaux[4], Andrea Ravignani[5,6,7,,§], Marco Gamba[1,§]

$ Co-first

§ Co-last

[1]University of Turin, Department of Life Sciences and Systems Biology, via Accademia Albertina 13, 10125, Torino, Italy

[2]University of Padua, Via VIII Febbraio, 2, 35122 Padova, Italy

[3]Giardino Zoologico di Pistoia, Via Pieve a Celle, 160/A, 5110, Pistoia, Italy

[4]Botanical and Zoological Garden of Mulhouse, 11 Av. de la 1ère Division Blindée, 68100 Mulhouse, France

[5]Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

[6]Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Building 1710, Universitetsbyen 38000, Denmark

[7]Department of Human Neuroscience, Sapienza University of Rome, Viale Regina Elena 334, 00185 Rome, Italy

*Address correspondence to: **Teresa Raimondi**. Email: teresa.raimondi@unito.it and Chiara Pertosa. Email: chiara.pertosa64@edu.unito.it

ABSTRACT

The search for common characteristics between the musical abilities of humans and other animal species is still taking its first steps. One of the most promising aspects from a comparative point of view is the analysis of rhythmic components, which are crucial features of human communicative performance, but also well-identifiable patterns in the vocal displays of other species. Therefore, the study of rhythm is becoming essential to understanding the mechanisms of singing behaviour and the evolution of human communication. Recent findings provided evidence that particular rhythmic structures occur in human music and some singing animal species, such as birds and rock hyraxes, but only two species of non-human primates have been investigated so far (*Indri indri* and *Hylobates lar*). Therefore, our study aims to consistently broaden the list of species studied regarding the presence of rhythmic categories. We investigated the temporal organisation in the singing of three species of crested gibbons (*Nomascus gabriellae, N. leucogenys, N. siki*) and found that the most prominent rhythmic category was isochrony. Moreover, we found slight variation in songs' tempo among species, with *N. gabriellae* and *N. siki* singing with a temporal pattern deriving from a gradually increasing tempo (a musical *accelerando*), and *N. leucogenys* with a more regular pattern. Here we show how the prominence of a peak at the isochrony establishes itself as a shared characteristic in the small apes considered so far.

KEY WORDS:

singing primates, music, tempo, *accelerando*, song, isochrony, rhythm

INTRODUCTION

Can you imagine a nightingale whistling Beethoven's Ninth Symphony, a woodpecker singing The Rolling Stones' 'Can't You Hear Me Knocking', or a chimpanzee drumming at the rhythm of Queen's 'We Will Rock You'? Probably not. In fact, what we generally consider as “music” is a social and cultural construct (Honing, 2018), which is however based on perceptual, motor and vocal capacities, that have been identified as musicality (Honing et al., 2015; Honing, 2018). Once identified, those characteristics and biological and cognitive substrates underlying human musical abilities can be used for cross-species investigation (Fitch, 2015).

The last few years have revealed that several species may share the building blocks of musicality with humans. For example, chickadees can discriminate pitch ratios presented at different absolute frequencies (Hoeschele et al. 2012), zebra finches can generalize across different timbres (Spierings & ten Cate 2014), and thrush nightingale songs share an isochronous rhythmic pattern with human music (Roeske et al. 2020). Singing primates, a small circle of primate species that produce modulated songs composed from tens to thousands of vocal units (Haimoff, 1986; De Gregorio et al., 2022a), have been indicated as among the most promising species to investigate features shared with our musical displays (Ravignani et al., 2014; Levinson, 2016). Singing primates tend to live in small social groups, often defending a territory and engaging in loud choruses that propagate for kilometres in the forest (De Gregorio et al., 2022a). Singing primates are, in some cases, phylogenetically distant (e.g., 80 mya) and in others within the Hominoidea clade (e.g., 20 mya). This phylogenetic distance suggests the possibility of convergent evolution for singing behaviour in this taxon (Geissmann, 2002). Nevertheless, the lack of knowledge about specific traits often prevents the formulation of further hypotheses.

Past research on two nonhuman primate species is particularly relevant here. Previous studies showed that the indri (*Indri indri*) is currently the only known species that perform duet and choruses (De Gregorio et al., 2019) that displays multiple rhythmic categories, as seen in human music (De Gregorio et al., 2021a). Rhythmic categories occur when temporal intervals between note onsets assume precise values, rather than being uniformly distributed, and rhythms with small integer ratio between these values are common in human music (Jacoby & McDermott, 2017; Savage et al., 2015). Besides indris, bioacoustics work on gibbons has shown interactive and individual rhythmic features in their songs. In particular, white-handed gibbons (*Hylobates lar*) showed a single prominent rhythmic category corresponding to small integer ratio, isochrony (Raimondi et al., 2023).

The last common ancestor between the white-handed gibbon (*Hylobates lar*) and modern humans is dated 20 mya (Glazko & Nei, 2003), while indris have an evolutionary history separated from humans by about 75 mya (Kumar et al., 2017), making the presence of multiple rhythmic categories in this lemur species even more intriguing. In fact, despite being phylogenetically distant from humans, indri lemurs are currently the non-human primates showing multiple rhythmic categories shared with human music, while small apes like the white-handed gibbons show only one.

As the mosaic concerning the occurrence of rhythmic categories within the vocal displays of animals is still largely incomplete, this study aims to extend the currently available knowledge by analysing the rhythmic categories of three species of crested gibbons. Gibbons are one of the few primate families featuring singing species. Hence, mapping rhythmic capacities across gibbon species is crucial to understanding whether what was observed in white-handed gibbons generalises to singing apes.

By studying the singing behaviour of three species of crested gibbons (*Nomascus gabriellae*, *N. leucogenys*, and *N. siki*), we aimed to understand whether we could find similarities across the rhythmic categories of closely related species. In fact, crested gibbons' common ancestor lived around 1.3 mya, while white-handed gibbons' and crested gibbons' common ancestor lived around 8.5 mya (Kumar et al. 2017). Gibbons belonging to the *Nomascus* genus are virtually unstudied in the wild and produce species-specific songs, although songs given by *N. leucogenys*, *N. siki* and *N. gabriellae* are very similar and only minor differences were observed (Geissmann, 2002; Thinh et al., 2011). Finding that a crested gibbon species would produce non-random metrical structures would be relevant not just for our understanding of the species' biology but also for providing a critical piece of information for understanding whether and how conformity of animal rhythms to small-integer ratio is widespread in the animal kingdom. Furthermore, by examining which primate species, besides humans, have evolved the capacity for musical rhythm, we can create the conditions to test functional hypotheses about why this capacity is adaptive (ten Cate & Healy, 2017).

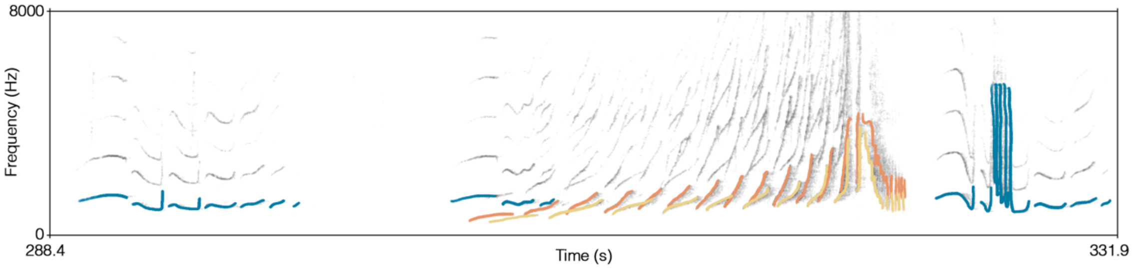


Figure 1 - Spectrogram of a song by a group of *N. gabriellae*.

The fundamental frequency (f0) of each individual contribution is highlighted with a different colour: the male in blue and the two females in orange and yellow respectively. of *N. gabriellae*, highlighting the contributions of the three animals to the chorus with different colours.

MATERIAL AND METHODS

1. Observations and recordings

We recorded songs (**Figure 1**) emitted by adult captive gibbons in two zoological parks (**Table SM1**). At the Pistoia Zoo (Italy), we studied a family group of *N. gabriellae* (N= 3, two males and one female) from October to December 2021. At the Zoological and Botanical Park of Mulhouse (France) from April to June 2022, we recorded a group of *Nomascus gabriellae* (N = 3, two females and one male), a group of *Nomascus leucogenys* (N = 3, two females and one male) and a group of *Nomascus siki* (N = 3, two females and one male).

We recorded songs using both a solid-state recorder (Zoom F1 equipped with a shotgun microphone) and a passive recorder (Audiomoth). The solid-state recorder was employed for the three species of *Nomascus* at the Zoological and Botanical Park of Mulhouse. The animals were observed from 8 am to 4 pm, each day focusing on two species, one in the morning and the other in the afternoon, rotating groups each day. The recordist pointed the microphone towards a particular individual and attributed each vocalisation to the signaler via the focal animal sampling technique. We also recorded songs through passive recorders (Aumdiomoth) monitoring the groups at Pistoia Zoological Park, with a recording schedule from 9 am to 2 pm (low gain, sampling rate: 48 kHz), placing it outside the exhibit of *N. gabriellae* (about 10 metres from the animals). Whenever the gibbons sang in the presence of a recordist, we used a digital camera Panasonic Lumix DMC-LZ7 to ensure the correct association of the emitter to each vocalisation.

2. Acoustic analyses

We edited and saved the portions of our recordings containing gibbons' songs as single mono audio files (WAV format). Video recordings and focal animal sampling allowed assigning each contribution to a singing individual. An individual contribution comprises all the units emitted by a single individual. We obtained a total of 334 individual contributions; 147 from *N. gabriellae* at the Pistoia Zoological garden, while at Mulhouse Zoological and Botanical Garden we recorded 79 contributions from *N. gabriellae*, 28 from *N. siki* and 81 from *N. leucogeny*s (Tab. SM1). Using the TextGrid tool in Praat (version 6.2.05; Boersma & Weenink, 2022) we annotated the onsets and offsets of all the units emitted in each individual contribution to a song. Through a visual inspection of spectrograms, we identified all vocal and non-vocal intervals and measured starting point and duration of each interval (Gamba & Giacoma, 2007). Next, using the software RStudio (RStudio Team 2020), we calculated the duration between the onsets of each couple of adjacent notes (inter-onset interval or IOI; hereafter $t_{k,}$), to evaluate the rhythmic structure of contributions (Sasahara et al., 2015; Gamba et al., 2016; De Gregorio et al., 2021b). We then calculated rhythmic ratios (r_k) by dividing each interval t_k for itself plus the following one: $r_k = t_k / (t_k + t_{k+1})$ (see Roeske et al. (2020) and De Gregorio et al. (2021a)). Ratios were calculated excluding t_k <0.025 s and >5 s, as these values are hypothesised being the lower and upper limit for metre perception and performance in macaques and humans and the thresholds on other ape species are unknown (Kuhl & Padden, 1983; London, 2012).

3. Statistical analyses

We performed all statistical analyses below using RStudio (RStudio Team, 2020).

3.1 t_k distribution: To evaluate how subsequent t_k and, consequently, the rhythmic categories were sequentially distributed among gibbons' species, we created a ternary plot (*ggtern* packages; Hamilton & Ferry, 2018) for each species. Ternary plots are graphical representations of any three components, typically expressed as percentage and whose sum must be 100%. In the case of human music, these plots have been called "time clumping maps", chronotopic maps usually employed in the human music field to capture, in principle, all possible performances of all rhythms of n+1 onset (Desain & Honing, 2003). The ternary plots depict a three-dimensional space: each axis represents one t_k, and each area within the graph represents a rhythm resulting from three consecutive t_k. Moreover, following the time-clumping map described by Desain & Honing (2003), all points in proximity to a empirical area belong to the same rhythmic category.

3.2 Rhythmic categories: r_k distribution. To evaluate the empirical occurrence of small-integer ratios, following the methodology shown in Roeske et al. (2020) and De Gregorio et al. (2021a), we divided the ratio distribution into on-integer and off-integer ratio ranges, centering the on-integer ratio range around 1:1 (or 0.50), 1:2 (or 0.33), 1:3 (0.25), 2:1 (0.66), 3:1 (0.75). The 0.50 value matches isochrony, i.e. two consecutive intervals have equal duration. According to Roeske et al. (2020), we set the on-off-peak boundaries for the 1:3 ratio at 0.222, 0.235, 0.267, 0.286; for 1:2 ratio at 0.286, 0.308, 0.364. 0.400; for 1:1 at 0.400, 0.444, 0.556, 0.600; for 2:1 at 0.600, 0.636, 0.692, 0.714; for 3:1 at 0.714, 0.733, 0.765, 0.778. We then counted all ratios that fell in each on- and off-integer ratio ranges for each individual contribution and adjusted the count by bin size.

3.3 r_k occurrence. To test whether the values' amount of t_k ratios (r_k) fell statistically more into the on-integer ratio range rather than the off-integer ratio range we used three Generalized Linear Mixed Models (GLMM, *glmmTMB* package; Brooks, 2017), one for each of the three *Nomascus* species. Before creating the models, we checked the distribution of the response variable (beta) via the package *fitdistrplus* (Delignette-Muller & Dutang, 2015) as a suitable theoretical distribution. Next, we entered the rk adjusted count as a response variable, the rk bin type (factor levels: OFF1:1, ON1:1, OFF1:2, ON1:2, OFF1:3, ON1:3, OFF2:1, ON2:1, OFF3:1, ON3:1) as a fixed factor, and individual ID and individual's contribution ID to a song as nested random factors. We then used a post-hoc test to perform all pairwise comparisons for all levels of the rk bin type (*emmeans* package; Lenth, 2019). Specifically, this test runs multiple comparisons between the observations that fall on integer and off integer for each integer ratio, with a P value adjustment based on the Tukey method. For all models, we verified the assumption of normality and homogeneity of residuals by visually inspecting the *qqplot* and the residuals' distribution (a function provided by R. Mundry). To test for the significance of our full models, we compared them against null models containing only the random factors (Forstmeier, 2011), with a likelihood ratio test (ANOVA with argument "Chisq"; Dobson, 2002).

RESULTS

1. t_k distribution

For *N. gabriellae,* t_k showed a bimodal distribution with a higher peak at 0.870 s, and a second peak for shorter t_k at 0.174 s (**Fig. 2B**). Differently, *N. leucogenys* and *N. siki* had a single peak at 0.392 s (**Fig. 2G**) and 0.481 s (**Fig. 2L**) respectively. Visual inspection of the ternary plots (**Fig. 2C, 2H, 2M**) indicated a shared pattern in the succession of t_k and, consequently, rhythmic categories. All species shared a higher density of values in the middle of the graphs. Following the time clumping map (Desain & Honing, 2003), rhythmic sequences were marked by three-time intervals with a reciprocal ratio of 1:1:1, indicating isochrony.

2. Rhythmic categories: r_k distribution

Visual inspection of r_k types' occurrence indicated a cluster in correspondence of isochronous ratios (1:1 or 0.5) for all three gibbons' species (**Fig. 2D, I, N**). The r_k probability density functions confirmed the results from ternary plots (**Fig.2C, H, M**), namely that subsequent t_k tend to have the same values, either when focusing on two or three following intervals of the same individual contribution.

3. r_k occurrence and species differences

The three models showed, for each species, significant differences in the r_k count depending on the r_k bin types (*N. gabriellae*: Null vs Full - Chisq=1400.514, df=9, $p<0.001$; *N. leucogenys*: Null vs Full - Chisq=829.813, df=9, $p<0.001$; *N. siki*: Null vs Full - Chisq=215.380, df=9, $p<0.001$). The post-hoc comparisons showed that the songs of all species were characterised by a prominent isochronous component ($p < 0.001$, Table 1), but not for 1:2, 2:1. Note that, *N. gabriellae* and *N. leucogenys* displayed significantly more ratios falling in the 2:1 *off-integer* bin than in the 2:1 *on-integer* one ($p < 0.001$ for both species, Table 1). The local maxima for the isochrony peak for each species was at 0.503 for *N. gabriellae*, for *N. leucogenys* at 0.501 and at 0.505 for *N. siki.*

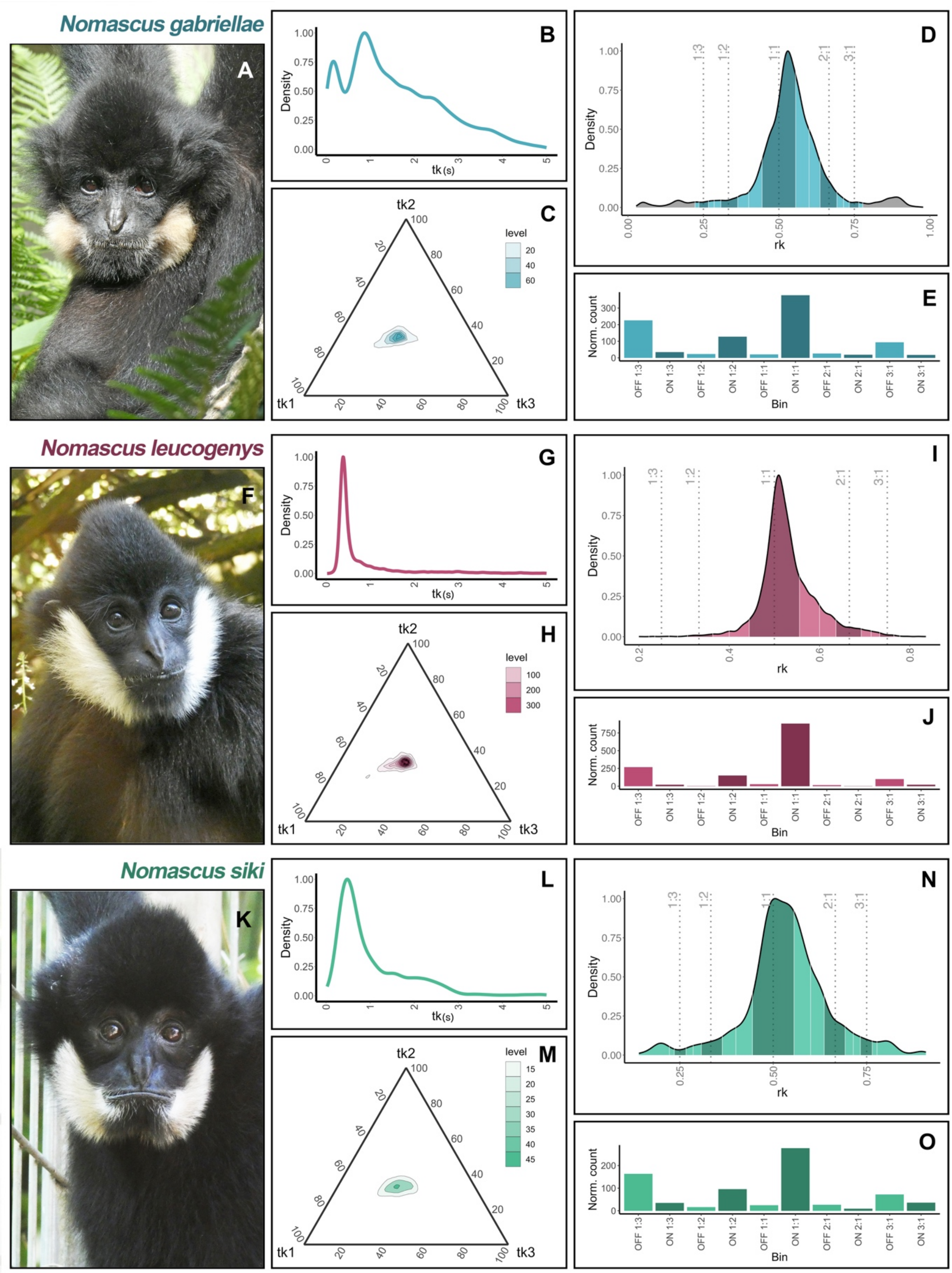


Figure 2 - T_k and r_k distribution for the three *Nomascus* spp.

A, F, K) *Nomascus* spp. **B, G, L)** Probability density functions of t_k for each species. **C, H, M)** Ternary plots for the three species, each axis represents one t_k, and each area within the graph represents a rhythm resulting from three consecutive tk. **D, I, N)** Probability density functions of rhythm ratios (r_k), a shift of central peak to the right of the 1:1 ratio may be diagnostic of an accelerando. **E, J, O)** Barplot showing the average adjusted

r_k occurrence for on-integer (dark colours) and off-integer (light colours) ratio ranges. Photo credits: C. Mancassola.

Table 1 - Post-hoc peak-wise comparisons of the GLMMs testing for the effect of r_k bin type (OFF1:1, ON1:1, OFF1:2, ON1:2, OFF1:3, ON1:3, OFF2:1, ON2:1, OFF3:1, ON3:1) on the r_k adjusted count in the three *Nomascus* spp. Tukey p-value adjustment for a 10 estimates comparison

Nomascus gabriellae

Adjusted r_k count ~ bin type + (1| ID / IDcontribution)

contrast	estimate	SE	df	t.ratio	p.value
OFF1:1-ON1:1	**-0.522**	**0.072**	**2237**	**-7.269**	**<.0001**
OFF1:2-ON1:2	0.207	0.095	2237	2.185	0.468
OFF1:3-ON1:3	0.017	0.096	2237	0.178	1.000
OFF2:1-ON2:1	0.407	0.088	2237	4.649	<.0001
OFF3:1-ON3:1	0.085	0.096	2237	0.889	0.997

Nomascus leucogenys

Adjusted r_k count ~ bin type + (1| ID / IDcontribution)

contrast	estimate	SE	df	t.ratio	p.value
OFF1:1-ON1:1	**-159.778**	**0.089**	**797**	**-17.991**	**<0.001**
OFF1:2-ON1:2	0.135	0.156	797	0.863	0.9974
OFF1:3-ON1:3	0.049	0.157	797	0.311	1.000
OFF2:1-ON2:1	0.552	0.130	797	4.262	<0.001
OFF3:1-ON3:1	0.253	0.156	797	1.618	0.839

Nomascus siki

Adjusted r_k count ~ bin type + (1| ID / IDcontribution)

contrast	estimate	SE	df	t.ratio	p.value
OFF1:1-ON1:1	**-0.964**	**0.206**	**267**	**-4.670**	**<0.001**
OFF1:2-ON1:2	0.095	0.267	267	0.354	1.000
OFF1:3-ON1:3	0.141	0.271	267	0.521	1.000
OFF2:1-ON2:1	0.342	0.249	267	1.374	0.934
OFF3:1-ON3:1	0.171	0.273	267	0.626	1.000

DISCUSSION

Our work aimed to broaden our knowledge of the occurrence of rhythmic categories in singing primates, particularly investigating the vocal behaviour of three species of crested gibbons (*Nomascus* spp.). We found that the songs of *N. gabriellae, N. siki and N. leucogenys* are characterised by a prominent isochronous component, in agreement with a previous study on lar gibbons (Raimondi et al., 2023). The consistency of distribution of rhythmic ratios among species is even more striking when considering the differences in the distribution of raw intervals, which are not randomly distributed but aggregate around one value for *N. leucogenys* and *N. siki*, and two values for *N. gabriellae*. Although is know that gibbons' songs are strongly genetically determined (Geismann 1984), the concordance in rhythmic structure among different crested gibbons species confirms the strong link between phylogeny, genetic features and singing behaviour in this genus (Geissmann, 2002; Konrad & Geissmann, 2006; Thinh et al., 2011). Nevertheless, we also found some degree of temporal variability between species, in particular concerning songs' tempo: the slight shift of the isochrony peaks over 0.5 in the three *Nomascus* spp. seems to indicate that *N. gabriellae* and *N. siki* sing with a temporal pattern imputable to a gradually increasing tempo (a musical *accelerando*), while *N. leucogenys'* singing shows a more regular pattern. However, we cannot infer if these slight differences in tempo could be perceived by the animals as it is known for humans, where the tempo of a song can influence listeners' emotions (van der Zwaag et al., 2011).

Here, we also show how the prominence of a peak at the isochrony establishes itself as a shared characteristic in the small apes considered so far. As Raimondi et al. (2023) observed in lar gibbons (*Hylobates lar*), we found that crested gibbons showed isochrony as the sole prominent rhythmic category. However, it remains an open question whether isochrony in crested gibbons' songs results from the anatomical constraints influencing call rate and/or is based on a neuronal substrate in the so-called oscillators (Large & Kelso, 2002; Desain & Honing, 2003). Results on lar gibbons (Raimondi et al., 2023) would suggest that these two aspects might be intertwined, as, in this species, the presence of a strong isochronous rhythm was only partially explained by physiological constraints linked to call rate.

Moreover, even if *Nomascus* spp., as *Hylobates* spp., are phylogenetically closer to humans than indris, all lesser apes investigated so far share less rhythmic categories with human music than lemurs do, as indris produces songs with an additional small integer ratio (1:2 - De Gregorio et al., 2021a). The study of rhythmic categories in animals, and in particular in the vocal signals of animals, is becoming increasingly important. For example, studies on male hyraxes have shown that males that vocalise with a more regular isochronous pattern and for longer periods have greater reproductive success and generate offspring that tend to survive longer (Demartsev et al. 2022).Our results complement previous data on lar gibbons, confirming that small apes show only one peak in the distribution of rhythmic ratios, corresponding to isochrony. Taking the multiple ratios characterising popular music (Jacoby and McDermott 2017), we should expect that a small ape, whose evolutionary history diverged later from humans compared to the indris, should feature a higher rhythmic complexity. Two rhythmic categories in the indris correspond significantly with the small integer ratios 1:2 and 1:1 (De Gregorio et al. 2021a).

Thus, two questions remain: Which features in the indris singing behaviour differ from the crested gibbons? What factors might have favoured the evolution of a song showing two rhythmic categories in the indris? The first aspect could stem from the extent to which singing results from collective behaviour. Evidence in lar and crested gibbons shows that solitary singing is common in males but sometimes in females (De Gregorio et al. 2022a). On the other hand, studies on indri have shown that singing is always a collective behaviour, rarely emitted by a single individual and potentially as an unanswered duet attempt (De Gregorio et al. 2022a). This aspect seems to be an essential element in distinguishing indri from other singing primates. A second aspect is related to the organisation of the chorus and the overlapping of individuals. Non-reproductive gibbons within a group (i.e., offspring of the pair-bonded adults) tend to overlap with the pair (Merker and Cox 1999). Female indris singing template guides the offspring's participation in the song (De Gregorio et al. 2019, De Gregorio et al. 2022b). Subadults also appear to avoid actively overlapping with their parents' singing (Gamba et al. 2016, De Gregorio et al. 2022b). We can speculate that this regular alternation of singers in the song and that they take turns but overlap non-randomly (Gamba et al. 2016) may have played a role in the evolution of complex rhythmic abilities.

Our findings corroborate the hypothesis that isochrony is a fundamental temporal organisation in the singing behaviour of primates. The mosaic regarding the rhythmic capabilities of primates is still in its infancy, and future investigations on other singing species and natural populations would be a valuable contribution to its construction.

FUNDING

This research was supported by the University of Torino.

AUTHORS' CONTRIBUTIONS

CDG and TR: conceptualization. CDG and TR: methodology. VB, CP, FC: data collection. CDG, TR, VB: investigation. CDG and MG: writing–original draft. DV, FC, AR, FB, LF, BL: writing–review and editing. TR: visualization. All authors contributed to the article and approved the submitted version.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the staff at the Zoological institutions where data were collected for their help and support. We also thank the editor and the reviewers for their helpful suggestions during the revision process.

REFERENCES

Boersma P & Weenink S, 2022. Praat: doing phonetics by computer [Computer program]. Version 6.2.05, retrieved 5 January 2022 from http://www.praat.org/

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Maechler M, Bolker BM, 2017. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J* **9**:378–400.

De Gregorio C, Zanoli A, Valente D, Torti V, Bonadonna G et al., 2019. Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. *Curr Zool* **65**:89–97.

De Gregorio C, Valente D, Raimondi T, Torti V, Miaretsoa L, Friard O, Giacoma C, Ravignani A, Gamba M, 2021a. Categorical rhythms in a singing primate. *Curr Biol* **31**(20):R1379-R1380.

De Gregorio C, Carugati F, Estienne V, Valente D, Raimondi T, Torti V, et al. 2021b. Born to sing! Song development in a singing primate. *Curr Zool* **67**:585–596.

De Gregorio C, Carugati F, Valente D, Raimondi T, Torti V, Miaretsoa L, et al., Giacoma C, 2022a. Notes on a tree: reframing the relevance of primate choruses, duets, and solo songs. *Ethol Ecol Evol* **34**(3):205-219.

De Gregorio C, Zanoli A, Carugati F, Raimondi T, Valente D, Torti V, Miaretsoa L, Andry R, Gamba M, Giacoma C. 2022b. Parent-offspring turn-taking dynamics influence parents' song structure and elaboration in a singing primate.*Front Ecol Evol.* 2022:01-14.

Delignette-Muller ML & Dutang C, 2015. fitdistrplus: An R Package for Fitting Distributions. *J Stat Softw* **64**:1–34.

Demartsev V, Haddas-Sasson M, Ilany A, Koren L, Geffen E. Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *J Anim Ecol*, 00:1– 12.

Desain P & Honing H, 2003. The formation of rhythmic categories and metric priming. *Perception* **32**(3):341-365.

Dobson AJ, 2002. An Introduction to Generalized Linear Models, 2nd Edn. Boca Raton, FL: Chapman and Hall/CRC Press.

Fitch WT, 2015. Four principles of bio-musicology. *Philos Trans R Soc B* **370**(1664):20140091.

Forstmeier W & Schielzeth H, 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav Ecol Sociobiol* **65**:47–55.

Gamba M & Giacoma C, 2007. Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. *Ethol Ecol Evol* **19**(4):323-343.

Gamba M, Torti V, Estienne V, Randrianarison RM, Valente D, Rovara P, Bonadonna G, Friard O, and Giacoma C, 2016. The indris have got rhythm! Timing and pitch variation a primate song examined between sexes and age classes. *Front Neurosci* **10**:249.

Geissmann T, 2002. Duet-splitting and the evolution of gibbon songs. *Biol Rev* **77**(1):57-76.

Geissmann, T 1984. Inheritance of Song Parameters in the Gibbon Song, Analysed in 2 Hybrid Gibbons (*Hylobates pileatus* × *H. lar*), *Folia Primatol*, 42(3-4), 216-235

Glazko GV & Nei M, 2003. Estimation of divergence times for major lineages of primate species. *Mol Biol Evol* **20**(3):424-434.

Gray PM, Krause B, Atema J, Payne R, Krumhansl C, Baptista L, 2001. The music of nature and the nature of music. *Science* **291**(5501): 52-54.

Haimoff EH, 1986. Convergence in the duetting of monogamous Old World primates. *J Hum Evol* **15**:51-59.

Hamilton NE & Ferry M, 2018. ggtern: Ternary diagrams using ggplot2. *J Stat Softw* **87**:1-17.

Hoeschele M, Guillette LM, Sturdy CB. 2012 Biological relevance of acoustic signal affects discrimination performance in a songbird. *Anim. Cogn.*15, 677–688

Honing H, 2012. Without it no music: beat induction as a fundamental musical trait. *Ann N.Y Acad Sci* **1252**:85–91.

Honing H, ten Cate C, Peretz I, Trehub SE, 2015. Without it no music: cognition, biology and evolution of musicality. *Philos Trans R Soc B* **370**(1664):20140088.

Honing H, 2018. On the biological basis of musicality. *Ann N.Y Acad Sci* **1423**(1):51-56.

Jacoby N & McDermott JH, 2017. Integer ratio priors on musical rhythm revealed cross- culturally by iterated reproduction. *Curr Biol* **27**:359–370.

Konrad R, Geissmann T, 2006. Vocal Diversity and Taxonomy of Nomascus in Cambodia. *Int J Primatol* **27**:713–745.

Kumar S, Stecher G, Suleski M, Hedges SB, 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 1;**34**(7):1812-1819.

Kuhl PK, Padden DM, 1983. Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *J Acoust Soc Am* **73**(3):1003-10.

Large EW, Fink P & Kelso SJ, 2002. Tracking simple and complex sequences. *Psychol Res* **66**:3–17.

Lenth R, Singmann H, Love J, Buerkner P, Herve M, 2019. Package 'emmeans'.

Levinson SC, 2016. Turn-taking in human communication–origins and implications for language processing. *Trends C Fogn Sci* **20**:6–14.

London J, 2012. Hearing in time: psychological aspects of musical meter. Oxford: Oxford University Press.

Patel AD, 2006. Musical rhythm, linguistic rhythm, and human evolution. *Music Percept* **24**:99–104.

Raimondi T, Di Panfilo G, Pasquali M, Zarantonello M, Favaro L, Savini T, Gamba M, Ravignani A, 2023. Isochrony and rhythmic interaction in ape duetting. *Proc R Soc B* 11;**290**(1990):20222244.

Ravignani A, Bowling DL, Fitch WT, 2014. Chorusing, synchrony, and the evolutionary functions of rhythm. *Front Psychol* **5**:1118.

Roeske TC, Tchernichovski O, Poeppel D, Jacoby N, 2020. Categorical rhythms are shared between songbirds and humans. *Curr Biol* **30**(18):3544-3555.

RStudio Team, 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Sasahara K, Tchernichovski O, Takahasi M, Suzuki K, and Okanoya K, 2015. A rhythm landscape approach to the developmental dynamics of birdsong. *J R Soc Interf* **12**:20150802.

Savage PE, Brown S, Sakai E, Currie TE, 2015. Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci*, **112**(29):8987-8992.

Spierings MJ, ten Cate C. 2014 Zebra finches are sensitive to prosodic features of human speech. *Proc. R. Soc. B*2, 81, 20140480.

ten Cate C, Healy S, 2017. *Avian Cognition*. Cambridge: Cambridge University Press.

Thinh VN, Hallam C, Roos C, Hammerschmidt K, 2011. Concordance between vocal and genetic diversity in crested gibbons. *BMC Evol Biol* **11**(1):1-9.

van der Zwaag MD, Westerink JH, van den Broek EL, 2011. Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Music Sc*. 15, 250–269

SUPPLEMENTAL MATERIAL

Table SM1. Name, sex, age and rearing conditions for all the individuals analysed, with the numbers of individual contributions obtained for each animal.

Species	Individual	Sex	Age	Zoo	Rearing environment	N. contributions
Nomascus gabriellae	Chloe	F	32	Mulhouse	born in captivity (1990)	27
	Dan	M	31	Mulhouse	unknown location (1991)	25
	Firmine	F	12	Mulhouse	born in captivity (2010)	27
	Lii	M	35	Pistoia	born in the wild (1985-1986)	46
	Siam	M	7	Pistoia	born in captivity (2015)	55
	Trilli	F	18	Pistoia	born in captivity (2003)	45
Nomascus leucogenys	Chukhao	F	15	Mulhouse	born in captivity (2007)	27
	Connie	F	32	Mulhouse	born in captivity (1990)	27
	LaiCao	M	7	Mulhouse	born in captivity (2015)	27
Nomascus siki	Anoie	F	17	Mulhouse	born in captivity (2005)	10
	Dorian	M	33	Mulhouse	born in captivity (1989)	8
	FengShui	F	12	Mulhouse	born in captivity (2010)	10

CHAPTER 5

Categorical rhythms in a singing primate

Current Biology (2021), *31*(20), R1379-R1380

Image generated on DALL.E, an Artificial Intelligence capable of generating images from textual descriptions through synography. "Digital art of an *Indri indri p*laying the trumpet"

Current Biology, *31*(20), R1379-R1380

https://doi.org/10.1016/j.cub.2021.09.032

Categorical rhythms in a singing primate

Chiara De Gregorio[1], Daria Valente[1,2], **Teresa Raimondi**[1], Valeria Torti[1], Longondraza Miaretsoa[1,3], Olivier Friard[1], Cristina Giacoma[1,5], Andrea Ravignani[4,5,6,*], and Marco Gamba[1,5]

[1] Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy.

[2] Equipe de Neuro-Ethologie Sensorielle (ENES), Centre de Recherche en Neurosciences de Lyon (CRNL), CNRS, INSERM, University of Lyon/Saint-Etienne, Saint-Etienne, France.

[3] Groupe d'Etude et de Recherche sur les Primates de Madagascar (GERP), Antananarivo, Madagascar.

[4] Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

[5] Co-senior authors.

[6] Lead contact. *E-mail: andrea.ravignani@mpi.nl

SUMMARY

What are the origins of musical rhythm? One approach to the biology and evolution of music consists in finding common musical traits across species. These similarities allow biomusicologists to infer when and how musical traits appeared in our species. A parallel approach to the biology and evolution of music focuses on finding statistical universals in human music. These include rhythmic features that appear above chance across musical cultures. One such universal is the production of categorical rhythms, defined as those where temporal intervals between note onsets are distributed categorically rather than uniformly. Prominent rhythm categories include those with intervals related by small integer ratios, such as 1:1 (isochrony) and 1:2, which translates as some notes being twice as long as their adjacent ones. In humans, universals are often defined in relation to the beat, a top-down cognitive process of inferring a temporal regularity from a complex musical scene. Without assuming the presence of the beat in other animals, one can still investigate its downstream products, namely rhythmic categories with small integer ratios detected in recorded signals. Here we combine the comparative and statistical universals approaches, testing the hypothesis that rhythmic categories and small integer ratios should appear in species showing coordinated group singing. We find that a lemur species displays, in its coordinated songs, the isochronous and 1:2 rhythm categories seen in human music, showing that such categories are not, among mammals, unique to humans.

MAIN TEXT

What are the origins of musical rhythm? One approach to the biology and evolution of music consists in finding common musical traits across species. These similarities allow biomusicologists to infer when and how musical traits appeared in our species[1]. A parallel approach to the biology and evolution of music focuses on finding statistical universals in human music[2]. These include rhythmic features that appear above chance across musical cultures. One such universal is the production of categorical rhythms[3], defined as those where temporal intervals between note onsets are distributed categorically rather than uniformly[2,4,5]. Prominent rhythm categories include those with intervals related by small integer ratios, such as 1:1 (isochrony) and 1:2, which translates as some notes being twice as long as their adjacent ones. In humans, universals are often defined in relation to the beat, a topdown cognitive process of inferring a temporal regularity from a complex musical scene[1]. Without assuming the presence of the beat in other animals, one can still investigate its downstream products, namely rhythmic categories with small integer ratios detected in recorded signals. Here we combine the comparative and statistical universals approaches, testing the hypothesis that rhythmic categories and small integer ratios should appear in species showing coordinated group singing[3]. We find that a lemur species displays, in its coordinated songs, the isochronous and 1:2 rhythm categories seen in human music, showing that such categories are not, among mammals, unique to humans[3].

Beyond melodic features (**Figure 1A**), individual animal vocalizations have onsets (blue lines in **Figure 1C**). Two onsets delimit an inter-onset interval (t_k), i.e. the time between the onset of a note and the next one[5,6]. Ratios between these intervals are calculated by dividing an interval by itself plus its adjacent one[3]. Patterns in these ratios may emerge: for instance, two identical intervals generate a 1:1 ratio, and an interval followed by another twice its duration generates a 1:2 ratio. This approach has successfully highlighted rhythmic similarities between birdsong and human music[3]. In particular, for the first time in a non-human species, it was recently demonstrated that songs of thrush nightingales show a significant isochronous rhythmic category (1:1 ratio) and a potential bias towards 1:2 categories[3]. Apart from songbirds, singing mammals are promising species to look for musical universals, offering the advantage of phylogenetic proximity to humans.

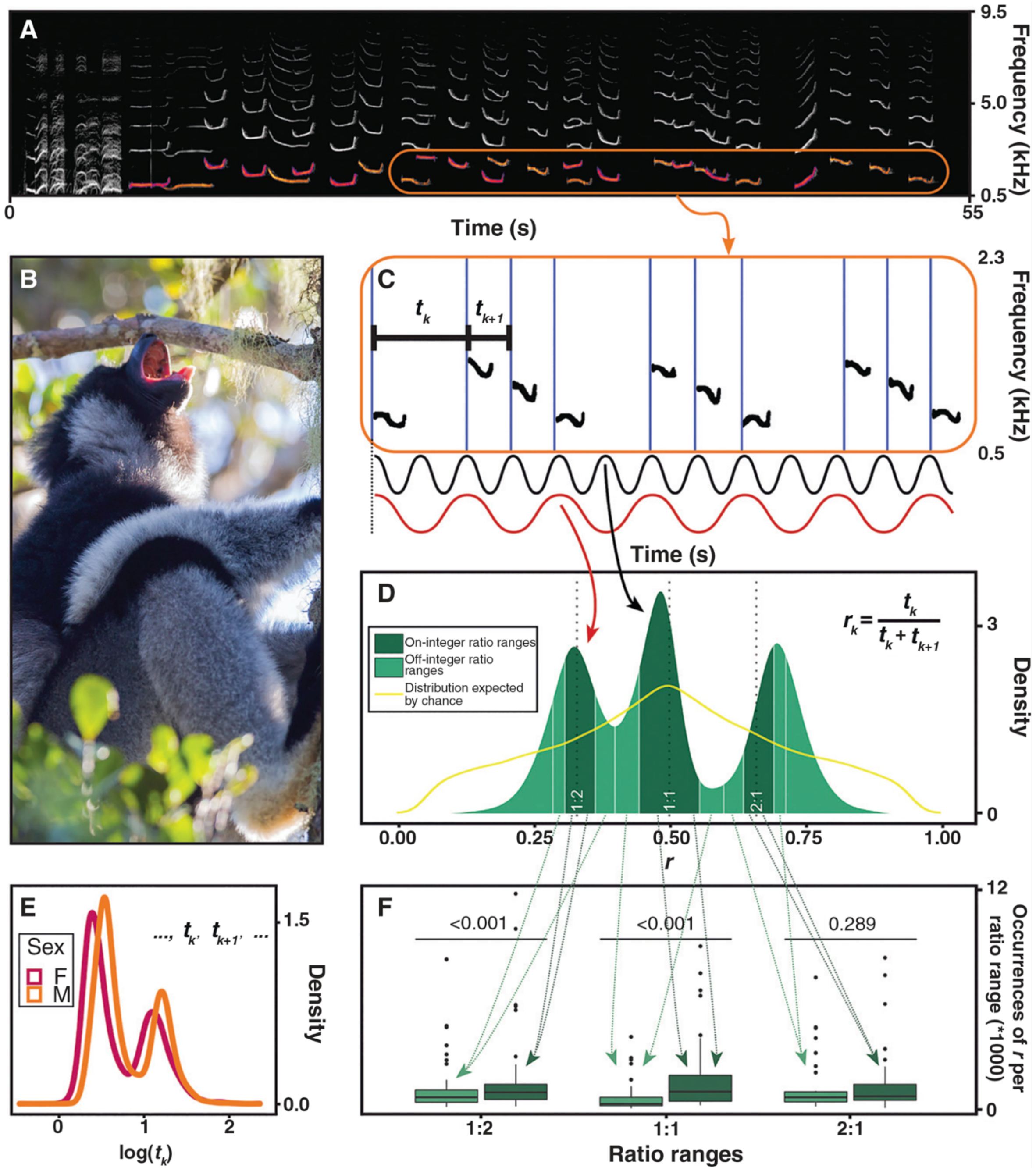


Figure 1. Indris' songs: Extracting individual inter-onset intervals and rhythm ratios from group choruses.

(A) Spectrogram of the indris' song, highlighting the fundamental frequencies of male (orange) and female (pink) notes (iZotope RX). The orange inset highlights the male song portion detailed in C. **(B)** Adult male indri singing in the Maromizaha New Protected Area, Madagascar (credit: Filippo Carugati). **(C)** Schematic representation of onsets (solid blue lines) in a male indri song. Pairs of onsets define inter-onset intervals t_k, marked by solid black lines. Sine waves (bottom) exemplify which purported oscillatory processes might generate the rhythmic categories found in the data: 1:1 ratio in black, 1:2 ratio in red. **(D)** Probability density function of rhythm ratios (r_k), which we calculated across 39 adult indris and 636 individual contributions to songs. On-integer (dark green) and off-integer (light green) ratio ranges are highlighted. The yellow line depicts a null distribution, showing how the ratios would be distributed in the absence of rhythmic categories, i.e. if the underlying intervals were uniformly distributed (see Supplemental information). The empirical ratio distribution significantly differs from this null distribution (2-sample Kolmogorov-Smirnov tests: $0.109 < D < 0.127$, $p < 0.001$). The central peak being slightly to the left of the 1:1 ratio may be diagnostic of a ‘ritardando’[3]. **(E)** Probability density function of the logarithm of inter-onset intervals (t_k), which significantly differs from a uniform distribution with the same boundaries (2-sample Kolmogorov-Smirnov test: $D = 0.656$, $p < 0.001$) and shows sexual dimorphism (i.e. a statistically significant sex difference; see Supplemental information). **(F)** Boxplots of adjusted r_k occurrence for on-integer (dark green) and off-integer (light green) ratio ranges. The ratio counts for individual indris constitute the data points for the analysis, and counts are normalized by bin size (see Sup- plemental information). Indris significantly produce more on-integer than off-integer ratios for the 1:2 (Wilcoxon signed-rank test, $V = 81$, p-values in figure) and the isochronous 1:1 categories ($V = 0$) but not for the 2:1 category ($V = 313$).

We focused on one of the few singing primates, the lemur *Indri indri* (**Figure 1B**), sampling approximately 1% of all living individuals from this critically endangered species (for details, see Supplemental information, published with this article online). All members of a family group sing in temporally coordinated duets and choruses (see: https://youtu.be/wxVYsAodZ6U)[7,8]. Over 12 years, we recorded songs from 20 indri groups (39 individuals) living in their natural habitat, the rainforest of Madagascar. We extracted temporal features of indris’ songs, analyzing inter-onset intervals (**Figure 1E**) and their ratios (Figure 1D).

First, we find that intervals between note onsets are not uniformly distributed (**Figure 1E**), and their ratios appear to form three clusters (green density function in **Figure 1D**). In other words, the intervals between notes are not sampled with the same probability among all possible values (yellow line in **Figure 1D**), similarly to the discretely sampled note durations found in human music[3,4]. Note that finding these clusters does not imply that they match specific integer ratios[5].

Second, we find that ratios match two rhythmic categories: 1:1, isochronous, similar to the pace of a metronome, and 1:2, a fundamentally small integer ratio. We ask whether ratios produced by each individual fall more frequently on-integer, i.e. in the vicinity of a small integer ratio, than off-integer, i.e. in the vicinity of its adjacent non-integer ratio (**Figure 1D** and **Supplement information**)[3]. By pairing the number of on-integer to off- integer ratios produced by each indri, we find that the empirical rhythmic ratios from songs fall statistically more often on small integer ratios than off integer ratios (**Figure 1F**). In particular, and similarly to songbirds[3], indri songs have a strong, above chance, isochronous component with some tempo flexibility; indris sing at a slightly decreasing isochronous tempo, similar to musical ‘ritardando’ (**Figure 1D**). While songbirds do not produce 1:2 categories above chance[3], indris' empirical ratios do statistically match the theoretical 1:2 category (**Figure 1F**). This provides direct evidence for one musical universal, categorical rhythms[2,5]. The small integer ratios, falling at 1:1 and 1:2, are exactly those expected for beat production and binary, metrical subdivisions[5]; both of these are aspects of human musicality rare in other species[1]. The 1:1 and 1:2 integer ratios we witness are not direct evidence of beat or meter, though they may hint at similar nested periodic processes (bottom of **Figure 1C**).

Third, we find that in males and females the absolute interval lengths are different (**Figures 1E** and **S1F**) but the 1:1 ratios are the same (see Supplemental information). This may be relevant to selective hypotheses for music, hypothesizing a role of sexual selection, social bonding, etc. on rhythm origins[9]. Productive inference on how rhythm evolved requires multicomponent thinking[1,6,9]: rhythm as a whole is a mystery but some of its defining features may be present in other species, as we show here. Male and female indris may produce different singing tempi and interval durations, but 1:1 ratios are the same between sexes. This suggests that sexual selection should not affect the evolution of isochrony in indris; it might, however, play a role in single interval timing and in non-isochronous rhythmic categories[1].

Why should another primate, apart from humans, produce categorical rhythms? As the last common ancestor between humans and indris lived 77.5 MYA, common ancestry of categorical rhythms is unlikely; instead, this ability may have convergently evolved among singing species, such as songbirds, indris, and humans[10]. As in songbirds, isochrony and rhythmic categories in indris may facilitate song coordination, processing, and potentially learning[1,3,6,7]. The mechanisms supporting rhythmic categories in indri are unknown; cognitive or not, they might still constitute a foundational phenomenon for beat-related behaviors in our and other species. We encourage comparative work on indri and other endangered species to gain more data before it is too late to witness their breath-taking singing displays.

Supplemental information

Supplemental information contains one figure, one table, experimental procedures, results and can be found with this article online at https://doi.org/10.1016/j.cub.2021.09.032.

REFERENCES

1. Kotz, S.A., Ravignani, A., and Fitch, W.T. (2018). The evolution of rhythm processing. Trends Cogn. Sci. 22, 896–910.

2. Savage, P.E., Brown, S., Sakai, E., and Currie, T.E. (2015). Statistical universals reveal the structures and functions of human music. Proc. Natl. Acad. Sci. USA 112, 8987–8992.

3. Roeske, T.C., Tchernichovski, O., Poeppel, D., and Jacoby, N. (2020). Categorical rhythms are shared between songbirds and humans. Curr. Biol. 30, 3544–3555.

4. Jacoby, N., and McDermott, J.H. (2017). Integer ratio priors on musical rhythm revealed cross- culturally by iterated reproduction. Curr. Biol. 27, 359–370.

5. Ravignani, A., Delgado, T., and Kirby, S. (2017). Musical evolution in the lab exhibits rhythmic universals. Nat. Hum. Behav. 1, 0007.

6. Ravignani, A. (2021). Isochrony, vocal learning and the acquisition of rhythm and melody. Behav. Brain Sci. 44, e88.

7. De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Ratsimbazafy, J., Gamba, M., and Giacoma, C. (2021). Born to sing! Song development in a singing primate. Curr. Zool., https://doi. org/10.1093/cz/zoab018.

8. Gamba, M., Torti, V., Estienne, V., Randrianarison, R.M., Valente, D., Rovara, P., Bonadonna, G., Friard, O., and Giacoma, C. (2016). The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. Front. Neurosci. 10, 249.

9. Savage, P.E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., and Fitch, W.T. (2021). Music as a coevolved system for social bonding. Behav. Brain Sci. 44, E59.

10. Haimoff, E.H. (1986). Convergence in the duetting of monogamous Old World primates. J. Hum. Evol. 15, 51–59.

SUPPLEMENTAL INFORMATION

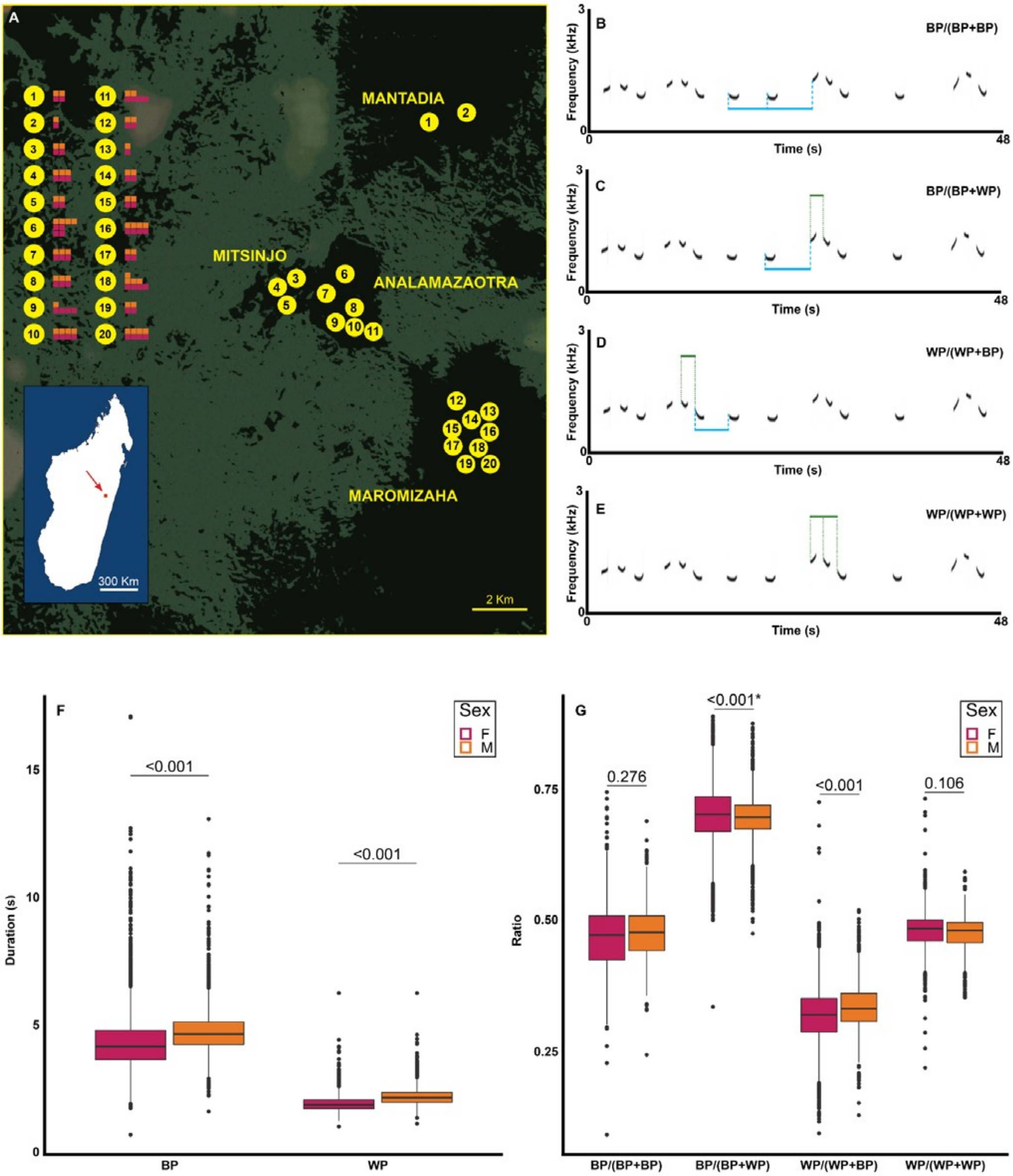


Figure S1. Song recording locations, with schematic representation of durations and ratios types and their values. (A) Map showing the four sites where indris were recorded. Yellow numbered circles represent the ID of the familiar group sampled. Colored lines next to the circles indicate that a reproductive male (orange) or female (pink) belonged to that group; the number of squares in each bar denotes the number of years that individual was recorded. **(B)** Schematic representation of a “BP/(BP+BP)” r_k type; the light blue line indicates two BP intervals: a t_k between two isolated notes, and a t_k between an isolated note and the first note of a phrase. **(C)** Schematic representation of a “BP/(BP+WP)” r_k type; the light blue line indicates a BP interval, and the green line indicates a WP interval: a t_k between an isolated note and the first note of a phrase, and a t_k between the first two notes of the same phrase. **(D)** Schematic representation of a “WP/(WP+BP)” r_k type; the light blue line indicates the BP interval, the green line indicates the WP interval: a t_k between two notes of the same phrase, and a t_k between a note belonging to a phrase and an isolated note. **(E)** Schematic representation of a "WP/(WP+WP)" r_k type; the green line indicates two WP intervals: in this case a t_k between the first and second note of the same phrase, and a t_k between the second and third note of the same phrase. **(F)** Boxplots of t_k duration by type. Notice that the WP boxplots here correspond to the first and second peaks in Figure 1E, and the BP boxplots here correspond to the third and fourth peaks in Figure 1E. **(G)** Boxplots of r_k values by type. The * denotes that, although the Tukey test is statistically significant, the difference is negligible because of the small effect size of the comparison.

Table S1. Summary and details of models and Tukey tests. **(A)** Influence of the fixed factors on t_k *duration*. Full model vs. Null model: $\chi 2 = 24468.48$, df = 3, $p < 0.001$. **(B)** Influence of the fixed factors on r_k *type*. Full model vs. Null model: $\chi 2 = 30155$, df = 11, $p < 0.001$. **(C)** Results of the Tukey test for r_k *type*. **(D)** Results of the Tukey test for the interaction between t_k *type* and *sex*. **(E)** Results of the Tukey test for the interaction between r_k *type* and *sex*. [a] Not shown as not having a meaningful interpretation. [b] Estimate ± SE refer to the difference of the response between the reported level of this categorical predictor and the reference category of the same predictor. [c] This predictor was dummy coded, with "SexF," being the reference category. Statistically significant values are indicated in bold; M: males; F: females.

A	**[LMM] Response Δ**		**Fixed Factor**	**Random Factors**	**Package**
	t_k duration		Sex * t_k type	ID Individual, ID contribution	lme4
Predictors	*Estimate*	*SE*	*df*	*t*	*p*
(Intercept)	1.423	0.028	[a]	[a]	[a]
WP[b]	-0.080	0.038	13770	-207.956	**<0.001**
M[b,c]	0.163	0.033	40.130	5.443	**<0.001**
WP*M[b,c]	0.037	0.006	13760	5.875	**<0.001**

B	**[GLMM] Response Δ**	**Fixed Factor**	**Random Factors**	**Package**
r_k		Sex * r_k type	ID Individual, ID contribution	glmmTMB
Predictors	*Estimate*	*SE*	*z*	*p*
(Intercept)	-0.130	0.008	[a]	[a]
BP/(BP+WP)[b]	0.982	0.009	105.460	**<0.001**
WP/(BP+WP)[b]	-0.632	0.009	-68.020	**<0.001**
WP/(WP+WP)[b]	0.046	0.009	4.860	**<0.001**
M[b,c]	0.034	0.015	2.300	**0.021**
BP/(BP+WP)*M[b,c]	-0.069	0.016	-4.170	**<0.001**
WP/(BP+WP)*M[b,c]	0.040	0.016	2.430	**0.015**
WP/WP+WP*M[b,c]	-0.054	0.016	-3.270	**0.001**

C		**[Tukey Test] r_k type**		**Package:** multcomp
	Estimate	*SE*	*z*	*p*
BP/(BP+WP) - BP/(BP+BP)	0.982	0.009	105.457	**<0.001**
WP/(WP+BP) - BP/(BP+BP)	-0.632	0.009	-68.021	**<0.001**
WP/(WP+WP) - BP/(BP+BP)	0.045	0.009	4.858	**<0.001**

WP/(WP+BP) - BP/(BP+WP)	-1.614	0.006	-259.709	**<0.001**
WP/(WP+WP) - BP/(BP+WP)	-0.936	0.006	-144.786	**<0.001**
WP/(WP+WP) - WP/(WP+BP)	0.677	0.006	105.373	**<0.001**

D		**[Tukey Test] t_k type * Sex**		**Package:** multcomp
	Estimate	*SE*	*z*	*p*
(F:F) WP-BP	-0.796	-0.796	-207.956	**<0.001**
(M:F) BP-BP	0.163	0.038	5.443	**<0.001**
(M:F) WP-BP	-0.596	0.030	-19.994	**<0.001**
(M:F) BP-WP	0.958	0.030	32.090	**<0.001**
(M:F) WP-WP	0.199	0.030	6.697	**<0.001**
(M:M) WP-BP	-0.759	0.005	-153.539	**<0.001**

E		**[Tukey Test] r_k type * Sex**		**Package**: multcomp
	Estimate	*SE*	*z*	*p*
(M:F) BP/(BP+BP)-BP/(BP+BP)	-0.796	-0.796	-207.956	0.276
(M:F) BP/(BP+WP)-BP/(BP+WP)	0.163	0.038	5.443	**<0.001**
(M:F) WP/(WP+BP)-WP/(WP+BP)	-0.596	0.030	-19.994	**<0.001**
(M:F)WP/(WP+WP)-WP/(WP+WP)	0.958	0.030	32.090	0.106

Supplemental Experimental Procedures

1. Observations and recordings

Indris (*Indri indri*) are lemurs whose ancestor diverged from ours approximately 77.5 MYA[S1]; phylogenetically, they are one of the farthest primate species from humans. Indris are classified as 'critically endangered' by the IUCN Red List of Threatened Species[S2]. While their population size is currently unknown, it has been estimated that only around 1000 - 10000 individuals remain in the wild[S3]. Indris have never been successfully bred in captivity[S4].

We recorded songs produced by 20 indri groups (39 individuals) living in four different rainforest patches in Madagascar (Figure S1A): six groups in the Analamazaotra Reserve (Andasibe-Mantadia National Park, 18°56' S, 48° 25' E), two groups in Mantadia (Andasibe-Mantadia National Park), three groups in the Mitsinjo Station Forestière (18°56'S, 48°4' E), and nine groups in the Maromizaha Forest (18°56'49" S, 48°27'53"E). Genetic analyses on seven groups inhabiting Maromizaha Forest confirmed that indris live in groups composed of an unrelated mating pair plus one to three individuals, usually their offspring[S5].

We collected data in the field from 2005 to 2016 for a total of 41 months. We observed one group per day from 06:00 a.m. to 1:00 p.m., using natural marks to identify each individual. Our analyses focused on adult reproductive individuals because temporal features in the vocalizations of adult reproductive indris may differ from those found in young or non-reproductive individuals[S6,S7].

We recorded songs using solid-state recorders (SoundDevices 702, Olympus S100 and LS05, and Tascam DR-100, DR-40, and DR-05) equipped with Sennheiser (ME 66 and ME 67) or AKG (CK 98) shotgun microphones. We set the recorders at a sampling rate of 44.1 kHz and an amplitude resolution of 16 bits during all the recording sessions. We recorded the animals at a distance ranging from 2 to 20 meters, with the microphone aimed at the focal singing animals. The recordist pointed the microphone towards a particular individual and attributed each vocalization to the signaler via the focal animal sampling technique[S8]. Songs can take the form of duets when two indris (usually the reproductive pair) sing together, or choruses, when one or more non-reproductive individuals join the pair in singing. An illustrative video can be found online at https://youtu.be/4w04ohzB40g. In the indris' songs, notes can be either organized in phrases or produced as isolated units. Phrases are characterized by adjacent units with a descending frequency pattern and separated by silent intervals. Isolated units are single notes between two phrases (Figure 1C, Figure S1B-E). A new phrase starts when the starting fundamental frequency of a note is higher than the ending fundamental frequency of the preceding one.

2. Acoustic analyses

From group recordings of 346 duets and choruses we extracted a total of 636 individual contributions uttered by 39 reproductive indris, 20 females and 19 males. Using Praat 5.3.46[S9], we edited and saved the recorded portion containing the indris' song as a single mono audio file (WAV format). Using field notes and video recordings, we then identified, annotated, and saved the onsets and offsets of each note for each individual as a Praat textgrid. We identified units and silences via visual inspection of the spectrograms. We labeled the notes, differentiating if they were organized in a sequence or produced as isolated units. Our analysis does not include ‘roars’, which are harsh and chaotic emissions that usually introduce the song[S10].

We then labeled the silent portions of the individual contributions, differentiating silences depending on whether they occurred between two notes of the *same phrase*, vs. between two *different phrases* or two different *isolated notes*. For each of the above units and intervals, we extracted the timing of note onsets and duration using Praat. We imported them in R (R Core Team 2017; version 3.4.3) and calculated the inter-onset intervals (t_k, Figure 1C) based on onsets to evaluate the rhythmic structure of contributions[S11]. Based on previous work[S7,S12], we used the information on notes and silences type (belonging to a phrase or an isolated note) to calculate two types of t_k. The within-phrase t_k (WP: Within-Phrase inter-onset intervals) are all and only the intervals between notes *within* a phrase (roughly mapping to the two leftmost peaks in Figure 1E). The between-phrase t_k (BP: Between-Phrase inter-onset intervals) are all and only the intervals *between* two different phrases or between two isolated notes (roughly mapping to the two rightmost peaks in Figure 1E). Note that a sequence of *n* notes will produce *n-1* inter-onset intervals, i.e. $t_1,...,t_{n-1}$.

We then calculated rhythmic ratios r_k following the methodology of Roeske et al.[S13], namely by dividing each t_k for its duration plus the duration of the following interval: $r_k = t_k /(t_k+t_{k+1})$. Since each t_k can either be a WP or a BP, we obtained four ratio types: WP/(WP+WP), WP/(WP+BP), BP/(WP+BP) and BP/(BP+BP), depending on the type of two adjacent t_k. These four ratio types and their corresponding intervals in indris' songs are depicted in Figure S1B-E (and their pooled distribution in Figure 1D). Further analyses were performed both on the t_k values and their ratios r_k.

3. Statistical analysis

Four main analyses were performed. All statistical analyses in 3.1-3.3 below were performed in R (R Core Team 2017; version 3.4.3). Analyses in 3.1 aimed at statistically substantiating the intuition in Figure 1E, namely that the distributions of t_k durations have 4 distinct peaks, potentially different by sex and t_k type. Analyses in 3.2 aimed at testing this same hypothesis on the r_k data. Analyses in 3.3 correspond to, and provide details for, the main result reported in the manuscript and Figure 1F. Compared to 3.3, analyses in 3.2 afford the possibility of zooming in on the overall pattern of ratios and analyzing them by sex and ratio type. In addition, comparison between results in 3.1 and 3.2 allow inference on potential sex differences in durations and/or ratios. Finally, the simulation described in 3.4 aimed at providing a plausible baseline scenario of how empirical ratios would look if indris had no rhythmic categories. While density functions (Figure 1D and 1E) were calculated for illustrative purposes, all statistical analyses were based on the original data points, not the smoothed, continuous density functions.

3.1. t_k durations

We tested whether sex and the type of t_k statistically predicted the inter-onset interval duration, using a Linear Mixed Model (LMM, lmer function of lme4 package[S14]). Before fitting the models, t_k duration was log-transformed ($\log_e$) because it was not normally distributed. We entered the values of t_k as response variable, and sex of the emitter, t_k type (WP and BP) and their interaction as fixed factors. We included the individual's identity and the specific song contribution from which we extracted the t_k as nested random factors. We used the Tukey test[S15] to perform all the pairwise comparisons for all the interaction levels between sex and t_k type (multiple contrast package *multcomp* in R). Finally, we used a two-sample Kolmogorov-Smirnov test in order to test if the two-peak distribution of our observed t_k durations (Fig. 1E) would differ from a random, uniformly distributed one, generated based on the upper and lower limit of the real distribution and containing the same number of observations (N= 14192).

3.2. t_k ratios (r_k)

To test whether the values of t_k ratios (r_k) were statistically influenced by the sex of the singer and the r_k type (WP/(WP+WP), WP/(WP+BP), BP/(WP+BP), and BP/(BP+BP)), we used a Generalised Linear Mixed Model (GLMM, *glmmTMB* package[S16]) fitting a beta distribution, continuous between 0 and 1. Beta was chosen via the package *fitdistrplus*[S17] as a suitable theoretical distribution. The values of r_k were entered in the model as response variable, and sex of the emitter, r_k type, and their interaction as fixed factors. We included the individual identity and the specific song contribution from which we extracted the ratios r_k as nested random factors. We used the Tukey test (within the multiple contrast package *multcomp* in R) to perform all pairwise comparisons for all levels of the interaction between sex and r_k type and for all levels of r_k type[S15].

For both models, we verified the assumptions of normality and homogeneity of residuals by visually inspecting the *qqplot* and the residuals' distribution (a function provided by R. Mundry). We also excluded the presence of collinearity among predictors based on variance inflation factors (*vif* package[S18]). To test for significance of our full models[S19] we compared them against null models containing only the random factors, with a likelihood ratio test (Anova with argument test “Chisq”[S20]). We report estimates, standard error (S.E.), z- and p-values for the Tukey tests (Table S1C-E).

Finally, we used Cohen’s d test to compute the magnitude of the effect size for the Tukey tests’ comparisons between different types of durations and ratios between the two sexes.

3.3. Rhythmic categories: Ratios distribution and their peaks

To evaluate the empirical occurrence of small integer ratios, following the methodology of Roeske et al.[S13], we divided the ratio distribution into on-integer and off-integer ratio ranges, centering the on-integer ratio ranges around 1:2 (or 0.333; a fundamentally small integer ratio), 1:1 (or 0.500; corresponding to isochrony), and 2:1 (or 0.666; a fundamentally small integer ratio). While the 1:1 ratio corresponds to two intervals of equal duration, the 1:2 and 2:1 ratios correspond to the second interval being, respectively, double and half the duration of the first.

Conversely, again following the methodology of Roeske et al.[S13], off-integer ratio ranges were centered around 1:3.5 (or 0.285), 1:2.5 (or 0.400), 1-1:2.5 (or 0.600), and 1-1:3.5 (or 0.710); the boundaries of all on- and off integer ratio ranges were 1:3.25 (or 0.307), 1:2.75 (or 0.363), 1:2.25 (or 0.444), 1-1:2.25 (or 0.555), 1-1:2.75 (or 0.637), and 1-1:3.25 (or 0.693). All these ratios were those used by Roeske et al.[S13]. We then counted all occurrences of ratio values that fell in each on- and off-integer ratio range for each individual, and we normalized these counts according to the size of their range on the x-axis[S13]. Notice, for instance, how the on-integer ratio range of 1:2 in Figure 1D is narrower than the 1:1 range; normalization allowed to correct and account for this and other inequalities. The Shapiro-Wilk tests confirmed that our count data did not follow a normal distribution, so we compared on-integer and off-integer ratio ranges using three (paired) Wilcoxon signed-rank tests.

3.4. Simulated ratio distribution

For reference, we simulated the null ratio distribution N expected by chance. The yellow line in Figure 1D is based on 100,000 simulated ratios produced using a custom script in Python 2.7.10. Based on the upper (681 msec) and lower (17089 msec) bound of inter-onset intervals naturally produced by indri[S12], we simulated what would happen if no rhythmic categories existed by sampling inter-onset intervals from two uniform distributions U and V with bounds at 681 and 17089 msec and calculating each ratio as in the formula in Figure 1D. In other words, this sampling simulated the distribution N of the ratio between one random variable U and its sum with another random variable V, both random variables being uniform.

Finally, to understand if the simulated ratio distribution (yellow line in Figure 1D) significantly differed from the empirical one, we resampled 17100 observations (the number of observations in our dataset) from the 100,000 simulated ones. This was done 1000 times, and every resampled distribution was compared with the empirical one using Kolmogorov-Smirnov tests in R.

Supplemental Results

1. t_k durations

The average duration of t_k was 2.114 ± 0.298s for WP and 4.651 ± 1.008s for BP. This difference is also visible in Figure 1E, where the probability density function clearly shows the presence of two clusters in each distribution of the t_k durations. Males showed longer t_k than females (Figure S1F, Table S1A), both overall (M = 3.429 ± 1.372s, F = 2.919 ± 1.403s; $p < 0.001$) and by t_k type (WP: M = 2.339 ± 0.340s, F = 1.899 ± 0.258s; $p < 0.001$; BP: M = 5.002 ± 0.928s, F = 4.318 ± 1.083s; $p < 0.001$; Table S1D; Figure S1F). WPs had a shorter duration than BPs ($p < 0.001$; Table S1A). When comparing the two sexes, Cohen's d was 1.477 for WP and 0.679 for BP, confirming that the significant sexual differences were non-negligible. In brief, there is both a sexual dimorphism in durations and a significant difference between the t_k types BP and WP, exemplified as the four peaks in Figure 1E. Finally, the twosample Kolmogorov-Smirnov test confirmed that the distribution of observed t_k durations significantly differed from a null uniform one with the same boundary values ($D = 0.656$, $p < 0.001$)

2. t_k ratios (r_k)

The average ratio r_k was 0.499 ± 0.161. Our model showed significant differences among r_k types (Table S1B), and the Tukey test confirmed that all four r_k types differed significantly ($p < 0.001$, for every comparison; Table S1C). The average WP/(WP+BP) was 0.324 ± 0.054, WP/(WP+WP) was 0.477 ± 0.033, BP/(BP+WP) was 0.698 ± 0.054, and BP/(BP+BP) was 0.470 ± 0.068. Notice how both values linked to isochrony (BP/BP+BP and WP/WP+WP) are slightly smaller than 0.5, suggesting increasing duration of adjacent intervals, i.e. ritardando.

Males presented overall higher ratio values than females ($p = 0.021$; Table S1B). However, the Tukey test indicated no significant sex differences for the ratio types linked to isochrony (BP/(BP+BP), $p = 0.276$; WP/(WP+WP), $p = 0.106$; Table S1E) emerged. Instead, we did find sexual dimorphism in r_k type WP/(WP+BP), where males showed lower values than females ($p < 0.001$), and in BP/(BP+WP), where males were the ones showing higher values ($p < 0.001$). When comparing between sexes, Cohen's d was 0.152 for WP/(WP+WP), 0.059 for BP/(BP+BP), 0.278 for WP/(WP+BP). Instead, it was 0.111 for

BP/(BP+WP), suggesting that this difference between males and females is negligible, even if it is statistically significant (Figure S1G). Therefore, the only actual (significant and non-negligible) difference in r_k types concerns WP/(WP+BP) values between males and females. WP/(WP+BP) corresponds to the first peak in Figure 1D and the first pair of boxplots in Figure 1F. The two significant peaks in Figure 1D characterize the 1:1 and the 1:2 ratio; the latter is sexually dimorphic and with higher values in males (third pair of boxplots in Figure S1G). This necessarily implies that the last inter-onset interval within a phrase is longer in males. In brief, there is little sexual dimorphism in ratios and categories and no dimorphism at all for significant rhythmic categories but the 1:2 ratio. On the one hand, this speaks *against* sexual selection hypotheses for isochrony; on the other hand, sexual selection might be responsible for longer intervals in males.

Summary and details of both models and Tukey tests appear in Table S1.

3. Rhythmic categories: Ratios distribution and their peaks

Visual inspection of the occurrence of different r_k types indicated a possible presence of three clusters (green density function in Figure 1D). Statistical comparison between data points from this density function confirmed and refined this intuition. The dependent 2-group Wilcoxon tests between on-integer and off-integer ratio ranges (Figure 1F) confirmed that indris produce songs characterized by at least two rhythmic categories: isochrony, corresponding to 1:1 ratio ($p < 0.001$, $V = 0$) and a 1:2 ratio ($p < 0.001$, $V = 81$). A V value of 0 for the isochronous 1:1 category means that each of the 39 indris produced more on-integer than off-integer ratios, not only statistically but also numerically at an individual level. However, we did not find significant differences between on-integer and off integer ratio ranges for 2:1 ($p = 0.289$, $V = 313$). In other words, the first and second (but not the third) peaks in Figure 1D result from indris producing ratios falling on small on-integer ratio neighborhoods rather than equallysized off-integer neighborhoods.

4. Simulated ratio distribution

The results of the Kolmogorov-Smirnov two-sample test showed that the 1000 resampled ratio distributions were significantly different from the empirical real ratio distribution: all $0.109 < D < 0.127$, all $p < 0.001$. In other words, statistical comparison between data points from the empirical distribution and data points from the distribution expected by chance (yellow line in Figure 1D) suggested that our empirical ratios differed from those expected by chance.

ACKNOWLEDGEMENTS

This research was supported by the University of Torino and the Parco Natura Viva – Centro Tutela Specie Minacciate, with the financial assistance of the European Union, through the Project BIRD (ACP SandT Program, Contract FED/2009/217077). The authors are grateful to the local field guides and field assistants who helped out during the data collection. We would also like to thank the GERP (Groupe d'Étude et des Recherche sur les Primates de Madagascar) for their support during the field activities, and to Dr. Cesare Avesani Zaborra and Dr. Caterina Spiezio for the financial and technical support. We are grateful to Andrea Schiavio, Diandra Düngen, Ine Alvarez van Tussenbroek, Maria Goncharova, Taylor Hersh, and two anonymous reviewers for providing critical feedback on the manuscript. We have received permits for this research, each year, from "Direction des Eaux et Forêts" and "Madagascar National Parks" (formerly ANGAP): 2005 [N°197/MINENV.EF/SG/DGEF/DPB/SCBLF/RECH], 2006 [N°172/06/MINENV.EF/SG/DGEF/DPB/SCBLF], 2007 [N°0220/07/MINENV.EF/SG/DGEF/DPSAP/SSE], 2008 [N° 258/08/MEFT/SG/DGEF/DSAP/SSE], 2009 [N°243/09/MEF/SG/DGF/DCB.SAP/SLRSE], 2010 [N°118/10/MEF/SG/DGF/ DCB.SAP/SCBSE and N°293/10/MEF/SG/DGF/DCB.SAP/SCB], 2011 [N°274/11/MEF/SG/DGF/DCB.SAP/SCB], 2012 [N°245/12/MEF/SG/DGF/DCB.SAP/SCB], 2014 [N°066/14/MEF/SG/DGF/DCB.SAP/SCB], 2015 [N°180/15/MEEMF/SG/ DGF/DAPT/SCBT], 2016 [N°98/16/MEEMF/SG/DGF/DAPT/SCB.Re and N°217/16/MEEMF/SG/DGF/DSAP/SCB.Re]. Data collection did not require a permit for 2013 because it has been performed by Malagasy citizens only.

AUTHOR CONTRIBUTIONS

Conceptualization, C.D.G., D.V., T.R., A.R. and M.G.; Methodology, C.D.G., D.V., T.R., A.R., O.F. and M.G.; Investigation, C.D.G., D.V., V.T., T.R. and L.M.; Writing – Original Draft, A.R. and C.D.G.; Writing – Review & Editing, M.G., V.T., T.R., D.V. and L.M.; Visualization, C.D.G., D.V., T.R., A.R. and M.G.; Supervision, M.G. and C.G.

DATA AND CODE AVAILABILITY

Data, code, and custom-written scripts are available from the corresponding author upon request.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL REFERENCES

S1. Masters, J.C., Silvestro, D., Génin, F., and DelPero, M. (2013). Seeing the wood through the trees: the current state of higher systematics in the Strepsirhini. Folia Primatol. *84*, 201-19.

S2. King, T., Dolch, R., Randriahaingo, H.N.T., Randrianarimanana, L., and Ravaloharimanitra, M. (2020). *Indri indri*. The IUCN Red List of Threatened Species 2020: e.T10826A115565566. https://dx.doi.org/10.2305/IUCN.UK.2020-

2.RLTS.T10826A115565566.en

S3. Andriaholinirina, N., Baden, A., Blanco, M., Chikhi, L., Cooke, A., Davies, N., Dolch, R., Donati, G., Ganzhorn, J., Golden,

C., Groeneveld, L.F., Hapke, A., Irwin, M., Johnson, S., Kappeler, P., King, T., Lewis, R., Louis, E.E., Markolf, M., Mass, V., Mittermeier, R.A., Nichols, R., Patel, E., Rabarivola, C.J., Raharivololona, B., Rajaobelina, S., Rakotoarisoa, G., Rakotomanga, B., Rakotonanahary, J., Rakotondrainibe, H., Rakotondratsimba, G., Rakotondratsimba, M., Rakotonirina, L., Ralainasolo, F.B., Ralison, J., Ramahaleo, T., Ranaivoarisoa, J.F., Randrianahaleo, S.I., Randrianambinina, B., Randrianarimanana, L., Randrianasolo, H., Randriatahina, G., Rasamimananana, H., Rasolofoharivelo, T., Rasoloharijaona,

S., Ratelolahy, F., Ratsimbazafy, J., Ratsimbazafy, N., Razafindraibe, H., Razafindramanana, J., Rowe, N., Salmona, J., Seiler,

M., Volampeno, S., Wright, P., Youssouf, J., Zaonarivelo, J., and Zaramody, A. (2020). *Indri indri*. The IUCN Red List of Threatened Species. Version 2020-3. https://www.iucnredlist.org

S4. Britt, A., Randriamandratonirina, N.J., Glasscock, K.D., and Iambana, B.R. (2002). Diet and feeding behaviour of *Indri indri* in a low-altitude rain forest. Folia Primatol. *73*, 225–239.

S5. Torti, V., Bonadonna, G., De Gregorio, C., Valente, D., Randrianarison, R.M., Friard, O., Pozzi, L., Gamba, M., and Giacoma, C. (2017). An intra-population analysis of the indris' song dissimilarity in the light of genetic distance. Sci. Rep. *7*, 10140.

S6. Gamba, M., Torti, V., Estienne, V., Randrianarison, R.M., Valente, D., Rovara, P., Bonadonna, G., Friard, O., and Giacoma, C. (2016). The indris have got rhythm! Timing and pitch variation a primate song examined between sexes and age classes. Front. Neurosci. *10*, 249.

S7. De Gregorio, C., Carugati, F., Estienne, V., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Ratsimbazafy, J., Gamba, M., and Giacoma, C. (2021). Born to sing! Song development in a singing primate. Curr. Zool., zoab018. S8. Altmann, J. (1974). Observational study of behavior: sampling methods. Behaviour *49*, 227–267.

S9. Boersma, P., and Weenink, D. (2008). Praat: Doing Phonetics by Computer (Computer Program). Available online at: http://www.praat.org

S10. Pollock, J. I. (1986). The song of the indris (*Indri indri*; Primates: Lemuroidea): natural history, form, and function. Int. J. Primatol. *7*, 225-264.

S11. Sasahara, K., Tchernichovski, O., Takahasi, M., Suzuki K, and Okanoya K. (2015). A rhythm landscape approach to the developmental dynamics of birdsong. J. R. Soc. Interface *12*, 20150802.

S12. De Gregorio, C., Zanoli, A., Valente, D., Torti, V., Bonadonna, G., Randrianarison, R.M., Giacoma, C., and Gamba, M. (2019). Female indris determine the rhythmic structure of the song and sustain a higher cost when the chorus size increases. Curr. Zool. *65*, 89-97.

S13. Roeske, T. C., Tchernichovski, O., Poeppel, D., and Jacoby, N. (2020). Categorical rhythms are shared between songbirds and humans. Curr. Biol. *30*, 3544-3555.

S14. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1–48.

S15. Bretz, F., Hothorn, T., and Westfall, P. (2010). Multiple Comparisons Using R (Chapman and Hall/CRC Press).

S16. Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Maechler, M., and Bolker, B.M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. R. J. *9*, 378–400.

S17. Delignette-Muller, M.L., and Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. J. Stat. Softw. *64*, 1– 34.

S18. Fox, J., and Weisberg, S. (2011). An R companion to applied regression (SAGE Publications, Inc).

S19. Forstmeier, W., and Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. Behav. Ecol. Sociobiol. *65*, 47–55.

S20. Dobson, A. J. (2002). An Introduction to Generalized Linear Models (Chapman and Hall/CRC Press).

OVERALL DISCUSSION

This thesis aimed to investigate the biological origins of rhythm in the vocalizations of the primate clade, specifically looking at the presence of a universal of human music, categorical rhythms, in the vocalizations of six primate species. This comparative study aims to reconstruct this trait's evolutionary history in species with different phylogenetic relationships to humans, with insights into the role of its hierarchical signal organization, function, synchrony and vocal interaction.

The evidence shows that all the analysed vocal types exhibit at least one rhythmic category. In particular, all vocal types show a significant peak around isochrony in the distribution of ratios between consecutive temporal intervals. Crucially, we show two other features in common with human rhythmicity: the flexibility of the rhythm dependent on the emission context (cosinging boosts rhythmic regularity in lar gibbons) and the recursiveness of the hierarchical structure of the vocal signal in the long calls of Bornean orangutans. At the same time, we present the first evidence in an animal species of an additional rhythmic category besides isochrony, coinciding with a small integer ratio (1:2): indris.

The production of isochronous patterns is widespread in animal communication

The presence of isochrony, found in all the vocal displays I considered, highlights how other non-human-primate species show a universal of human music (**CHAPTERS 1-5**). This finding supports the hypothesis that the production of isochronous patterns is common and potentially homologous in this clade. The production of isochronous patterns appears to be widespread in animal communication and does not per se require cognitive/neurological mediations exclusive to humans. In general, the oscillatory pattern of vocalizations and timing processes have been proposed to be the direct result of neural oscillations involving basal ganglia in the brain that seem to be shared with other primates and taxa (Ravignani & Madison, 2017). Such a mechanism has been hypothesised to underlie the production of rhythmic patterns in phylogenetically extremely distant from primate taxa, such as arthropods and amphibians. In amphibians, for instance, connections have been identified between the timing of vocalizations and neurophysiology, with an effect of the hormonal environment on the rhythmic structure of the signal (Zornik & Kelley, 2011).

The pivotal difference between the production of isochronous patterns in other taxa compared to humans has been proposed to be the fact that, in humans, isochrony is the result and interaction with exogenous factors, and not only the direct result of endogenous oscillations (Ravignani & Madison, 2017). Our evidence on lar gibbons (**CHAPTER 3**) shows flexibility in rhythm regularity depending on the context of emission, in line with the idea that the production of regular intervals is not exclusively the result of the endogenous environment but is subject to exogenous stimuli, as in humans.

The role of vocal interaction: isochrony as a coordination tool?

Male lar gibbons adapt their rhythmic structure on the base of the context of emission (**CHAPTER 3**): exogenous stimuli can shape isochrony. This is further evidence favouring a gradual view of rhythmic capabilities from a comparative perspective rather than a dichotomic view of human and animal rhythmic capacities (Honing et al., 2018; B. Merker, 1999): humans produce music, animals don't.

The susceptibility of rhythm to social factors is a relevant clue in identifying the forces that made isochrony adaptive. It may play a role in anticipating the rhythm of cosingers and coordination between individuals (Bowling et al., 2013; B. H. Merker et al., 2009). There is no evidence of a direct connection between isochrony and vocal interaction in non-human species. However, the evidence that social factors greatly influence the structure of vocal signals and that it is subject to flexible adjustment based on among-individuals-interaction (de Reus et al., 2020) is solid in primates (Clink et al., 2020; Gamba et al., 2016; Terleph et al., 2018) and amphibians (Song et al., 2020).

At the same time, regular rhythmic patterns, even on several hierarchical levels, are also present in species that do not vocalise in interactive mechanisms but individually (orangutans; **CHAPTERS 1-2**). In other words, although isochrony is boosted in the presence of cosinging, and is therefore flexible and sensitive to the context of emission, it also occurs in vocalisations emitted in solitude: vocal interaction is not necessary to develop isochrony, but it appears to boost regularity in social singing species.

Respiratory constraints and isochrony

The production of isochronous patterns may not only be related to neuronal oscillation but may also result from physiological and respiratory constraints. This may be particularly relevant in species that emit at a high call rate: the effect of call rate on isochrony has been partially highlighted in the lar gibbons' song (**CHAPTER 3**). In vocalizations emitted for advertisement purposes, emitting faster is adaptive since it allows the transmission of more information in a shorter time and represents an honest signal of the quality of the emitter, making the signal pivotal in sexual selection (Catchpole & Slater, 2003; Searcy & Andersson, 1986). However, for physiological and mechanical reasons, increasing the call rate indefinitely is impossible. The action of sexual selection on the call rate is a widespread phenomenon in nature (Catchpole & Slater, 2003; Sullivan, 1983), but also regularity itself can be subject to it (Demartsev et al., 2022; E. Van Den Broek & Todd, 2003). The lower limit of the interval length could represent the factor determining the equal length of all intervals, thus isochrony. With this work, I show how this mechanism may partially underlie isochrony but also that physiological constraints alone don't explain the rhythmic structure flexibility according to the context of emission: gibbon solo songs have higher call rates but lower rhythmic regularity. The orangutans (**CHAPTER 2**) reinforce the idea that isochrony can only be partly explained by physiological constraints since we observe a nested isochrony that, at the level of sub-elements (sub-pulses), could be determined by respiratory constraints, but this does not apply to higher-order grouping (full-pulses), where periodicity is significantly slower.

The signal's function: advertisement needs redundancy, redundancy exploits isochrony

The primate song is an outlier behaviour within this clade: it only appears in a few families, namely Tarsiidae, Pitheciidae, Hylobatidae and Indriidae (De Gregorio et al., 2022). Even though the functions of primate songs are still a topic of research, it is assumed to cover some central functions (Geissmann, 1999), such as territorial defence and conflicts mediation (Cowlishaw, 1992), inter-group spacing and communication (Mitani, 1985b), intra-group communication and cohesion (Geissmann & Orgeldinger, 2000). Songs are long, loud, temporally segmented, and frequency-varied vocal signals (De Gregorio et al., 2022). Since they convey crucial information, such as group composition and territory location, in the context of inter- and intra-group communication, it is possible to assume that the smaller the information loss, the more adaptive the signal. In this sense, the fact that all specific songs I considered (**CHAPTERS 3-5**) showed isochrony is meaningful since isochrony potentially maximizes signal redundancy, minimizes information loss and allows cosingers to manage their timing and predict the others', and thus flexibly mediate synchrony (more overlap for signal summing, less overlap for conveying information about the emitter individual). Male orangutan long calls are emitted in a rather different social and ethological context than songs, but they share with songs some potential functions and selective pressures to which the signal is subject. Long calls, which we showed to be isochronous (**CHAPTERS 1-2**), transmit the position and status (Mitani, 1985a) of the emitter and mediate spacing between males (somewhat like the song serves a territorial purpose). Like songs, long calls are potentially subject to sexual selection, and greater regularity in rhythmic patterning may reflect a well-functioning neurological system (E. M. F. Van Den Broek & Todd, 2009). Like songs, orangutans' long calls mediate the transmission of crucial information, where redundancy could be remarkably adaptive.

Correlation between a hierarchical signal organization and rhythmic categories

All studied species showed isochrony in their vocalizations, but only one species showed a multi-modal distribution of the rhythmic ratios, i.e., several categorical rhythms. Crucially, out of the three shown categorical rhythms, two of them fell on small-integer ratios (1:2, 1:1), with the third falling in the vicinity of another small-integer-ratio (2:1), with a slight trend towards a *ritardando*. The fact that the only species exhibiting a further small-integer-ratio apart from isochrony is a strepsirrhine (**CHAPTER 5**), i.e. furthermore, the most distant from humans among those analysed in this thesis, suggests that the presence of this trait is dictated by eco/ethological convergent rather than homologue factors. No species studied with this approach had ever shown additional small-integer ratios outside of isochrony (Anichini et al., 2023; Demartsev et al., 2022; Roeske et al., 2020; Xing et al., 2022). The question that emerges is, therefore: what is the selective pressure that determines the presence of a further significant small-integer ratio in the indri song but not in the other analysed vocal signals?

An isochronous pattern is intrinsically related to a non-hierarchical signal. In cold theory, multiple rhythmic categories, regardless of whether they coincide with small-integer ratios, can only result from the discrete distribution of at least two types of intervals, long or short. This is what happens in the indri song, in which the organization of aggregated units into clusters, called phrases, determines the alternation of long intervals (between two phrases) and short intervals (between units of the same phrase), resulting in three rhythmic categories, two of which surprisingly coincide with small integer rations (1:1 and 1:2). Since gibbon songs, as the other primate songs (De Gregorio et al., 2022), also typically have a hierarchical organization, why do we not detect the presence of multiple peaks? This observation may appear problematic, but it allows us to emphasize a relevant aspect of studying rhythmic categories. In indris, the organizational structure of the core section of the song shows descending phrases that consist of 1-6 units each (Gamba et al., 2016). To generalize, gibbon song clusters comprise a very high number of sub-elements (more than a dozen, Raemaekers et al., 2016). For instance, the Sub-elements of *Nomascus* gibbons have a hierarchy showing numerous sub-elements clustered into a higher-order organization (e.g., the great call of the female) separated by longer than 5 s intervals. The hierarchy of gibbon songs may thus not be mirrored by our results, as a time interval of this magnitude has been excluded from our analyses since evidence on humans and other primate species indicate that metre perception and production is very unlikely to be based on similar neural mechanisms (Kuhl & Padden, 1983; London, 2004). In this sense, we can hypothesize that the systems of signal hierarchization, be it speech, human musicality or animal communication, are based on systems that, depending on the time scale on which they are articulated, disregard metre perception and production. This is not surprising when considering that hierarchical organization and its abstraction, in the animal behavioural spectrum in general, is not relegated to communication but also involves, for instance, long-lasting sequences of goal-directed actions (lower-level goals organized into higher-level goals to fulfil a purpose). Bearing in mind that humans and other primates share basic brain architecture, with processing that has gradually changed throughout evolution, it has been proposed that such forms of hierarchization are based on similar mechanisms that are linked, for example, to working memory (Asano, 2021).

Another aspect emerging from this research is that not only the songs but also the long calls of orangutans show a hierarchical structure, i.e., a nested isochrony (a level of isochrony linked to the full pulse and a level of isochrony linked to the elements constituting the full pulses; **CHAPTER 2**), highlighting how hierarchy is not only not exclusive to the primate song but, above all, that recursion is not exclusive to humans. What relationships subsist between rhythm, recursiveness, combinatorics and syntax of a communication system, in our and other species, remains a nebulous issue.

Does tempo affect categorical rhythms?

As noted above, isochrony may partially be a by-product of a fast call rate in lar gibbons (**CHAPTER 3**). However, the relationship between a fast call rate and the presence of categorical rhythms other than isochrony remains to be clarified: could the presence of categorical rhythms in indri but not in gibbons have something to do with the slow call rate of the former and a fast call rate of the latter? What is observed in human music is that tempo has an effect on the perception and production of rhythmic patterns, and not merely in kinematic terms but in the signal organization and generalization at a neurological level: the deviations from rhythmic categories depend on the internal organization of the signal and its tempo (Repp et al., 2002), and, in particular, faster tempo patterns show larger deviations from the requested rhythm. Specifically, in patterns with marked variations between long and short intervals (with thresholds between long and short definitions changing depending on the authors), when the tempo of the performance increases and the short intervals can no longer be shortened (for kinematic reasons), then only the longer intervals are shortened, which leads to a flattening of the differences between long and short intervals (Repp et al., 2002; Roeske et al., 2020), and a deviation of the ratios linking intervals. To put it another way, the progressive similarity of durations of long and short intervals leads to the production of increasingly isochronous intervals. As a general mechanism, at fast tempi, the distinction and production of discrete intervals of three different durations (for instance, the 3:2:1 swing ratio) seem difficult to maintain: there is a convergence between the durations (and thus a loss of the 3:2:1 ratio).

Interestingly, besides the tendency toward reduction of simple ratios (known as *assimilation*) at a faster tempo, at a slower tempo contrast between long and short intervals is enhanced. In other words, perceptual reorganization, or regrouping, in humans causes that identical rhythmic patterns can be grouped in different ways at different tempi, which makes it difficult to recognize and reproduce the same rhythm at different tempi (Clarke, 1982; Repp et al., 2002). For example, signals like the sub-pulses of orangutan long calls or gibbon songs show shorter intervals (i.e., higher call rates; **CHAPTERS 2-5**) than indri's song. Suppose similar mechanisms of reorganization of interval ratios, based on time, also apply to other primate species. In that case, we might expect the presence of additional rhythmic categories besides isochrony in indri to be partly related to the evolution of a slower tempo signal.

Homologies and analogies in primate rhythm production

The study of musicality from a comparative perspective represents a fundamental building block in understanding the mechanisms that have made music adaptive in our species (Kotz et al., 2018; Ravignani & Madison, 2017). Since we are the only species to have evolved a musical system, musical features in non-human animals can be read as a side effect of more general features of perception and cognition that may serve communicative functions in general (Hauser & McDermott, 2003). This process goes toward understanding the composite nature of communicative systems and musicality in particular (Fitch, 2012; Kotz et al., 2018). Musicality is not a monolithic trait and cannot be merely looked at in terms of the presence or absence of very general features in a given species. With this in mind, we can interpret rhythmic categories as the result of a long series of separate neural and cognitive processes, even co-opted from systems with different functions, depending on the selective pressures species are subjected to in that specific evolutionary moment. This premise is intended to reiterate that it is implausible to consider a trait that undoubtedly relies on many different substrates and neural systems, which fulfils many diverse functions, as homologous or analogous as a whole: we can expect that what enables the production of isochrony or small integer ratios in non-human animals may be the set of both analogous and homologous sub-systems, concerning our and other non-human species. It is therefore proposed that this approach should also be applied to the observation of the isochronous category independently of the different identified rhythmic categories: the evolutionary history of isochrony could disregard, at least in part, that of other rhythmic categories.

The evidence provided in this thesis shows how all primate species studied to date in terms of the presence of rhythmic categories show isochrony. From a certain point of view, by Occam's razor principle, one might infer that the substrates necessary for the production of isochrony were present at the primate root rather than evolving multiple times, by convergence, in the primate clade. This view agrees with the evidence mentioned above showing that the production of isochronous patterns is widespread in communication in the animal kingdom (Ravignani & Madison, 2017). At the same time, the only species that shows further rhythmic categories besides isochrony, and small integer ratios in particular, is indri, a Strepsirrhine resulting from a rather basal primate evolutionary branch. Small-integer ratios like the 1:2 ratio may have been lost from most primates throughout evolution if homologous between humans and indri. Alternatively, it may have emerged many times (at least two, according to the current evidence), convergently in the primate clade, resulting from all those composite, non-mutually exclusive processes discussed above.

Rhythmic abilities result from a compound system in terms of substrates and deriving capacities, potentially causing isochrony and other rhythmic categories to evolve independently. If isochrony is at least partially mediated by neuronal oscillations and mechanical/physiological mechanisms, different small integer ratios could be the emergent result (albeit dependent on an isochronous oscillation) of a process of hierarchization and tempo slowing of the signal. The role of metre perception, in terms of mediation of only partially, identified cognitive systems and ecological and social pressures on the emergence of this trait, have not yet been elucidated. It is suggested that the study of spontaneous vocal communication from a comparative perspective is imperative in this investigation process.

REFERENCES

Alcock, J., & Rubenstein, D. R. (2019). *Animal behavior* (Eleventh edition). Oxford University Press.

Altenmüller, E. O. (2001). How Many Music Centers Are in the Brain? *Annals of the New York Academy of Sciences*, *930*(1), 273–280. https://doi.org/10.1111/j.1749-6632.2001.tb05738.x

Anichini, M., de Reus, K., Hersh, T. A., Valente, D., Salazar-Casals, A., Berry, C., Keller, P. E., & Ravignani, A. (2023). Measuring rhythms of vocal interactions: A proof of principle in harbour seal pups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1875), 20210477. https://doi.org/10.1098/rstb.2021.0477

Arcadi, A. C., Robert, D., & Boesch, C. (1998). Buttress drumming by wild chimpanzees: Temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates*, *39*(4), 505–518. https://doi.org/10.1007/BF02557572

Asano, R. (2021). The evolution of hierarchical structure building capacity for language and music: A bottom-up perspective. *Primates*, *63*. https://doi.org/10.1007/s10329-021-00905-x

Bowling, D. L., Herbst, C. T., & Fitch, W. T. (2013). Social Origins of Rhythm? Synchrony and Temporal Regularity in Human Vocalization. *PLoS ONE*, *8*(11), e80402. https://doi.org/10.1371/journal.pone.0080402

Bradbury, J. W., & Vehrencamp, S. L. (2011). *Principles of animal communication* (2nd ed). Sinauer Associates.

Brown, S. (2001). Are Music and Language Homologues? *Annals of the New York Academy of Sciences*, *930*(1), 372–374. https://doi.org/10.1111/j.1749-6632.2001.tb05745.x

Brown, S., & Weishaar, K. (2010). Speech is “heterometric”: The changing rhythms of speech. *Speech Prosody 2010-Fifth International Conference*, 4.

Burchardt, L. S., Norton, P., Behr, O., Scharff, C., & Knörnschild, M. (2019). General isochronous rhythm in echolocation calls and social vocalizations of the bat *Saccopteryx bilineata*. *Royal Society Open Science*, *6*(1), 181076. https://doi.org/10.1098/rsos.181076

Catchpole, C. K., & Slater, P. J. B. (2003). *Bird Song: Biological Themes and Variations*. Cambridge University Press.

Clarke, E. F. (1982). Timing in the performance of Erik Satie’s ‘Vexations.’ *Acta Psychologica*, *50*(1), 1–19. https://doi.org/10.1016/0001-6918(82)90047-6

Clink, D. J., Tasirin, J. S., & Klinck, H. (2020). Vocal individuality and rhythm in male and female duet contributions of a nonhuman primate. *Current Zoology*, *66*(2), 173–186. https://doi.org/10.1093/cz/zoz035

Corballis, M. C. (2009). The Evolution of Language. *Annals of the New York Academy of Sciences*, *1156*(1), 19–43. https://doi.org/10.1111/j.1749-6632.2009.04423.x

Cowlishaw, G. (1992). Song Function in Gibbons. *Behaviour*, *121*(1–2), 131–153. https://doi.org/10.1163/156853992X00471

Darwin, C. (1872). *The Descent of Man: Selection in Relation to Sex*. D. Appleton.

De Gregorio, C., Carugati, F., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Gamba, M., & Giacoma, C. (2022). Notes on a tree: Reframing the relevance of primate choruses, duets, and solo songs. *Ethology Ecology & Evolution*, 1–15. https://doi.org/10.1080/03949370.2021.2015451

De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A., & Gamba, M. (2021). Categorical rhythms in a singing primate. *Current Biology*, *31*(20), R1379–R1380. https://doi.org/10.1016/j.cub.2021.09.032

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M. D., Bruno, J. H., Trainor, L., & Ravignani, A. (2020). *Rhythm in dyadic interactions* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/9yrkv

Demartsev, V., Haddas-Sasson, M., Ilany, A., Koren, L., & Geffen, E. (2022). Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *Journal of Animal Ecology*, 1365-2656.13801. https://doi.org/10.1111/1365-2656.13801

Desain, P., & Honing, H. (2003). The Formation of Rhythmic Categories and Metric Priming. *Perception*, *32*(3), 341–365. https://doi.org/10.1068/p3370

Doupe, A. J., Perkel, D. J., Reiner, A., & Stern, E. A. (2005). Birdbrains could teach basal ganglia research a new song. *Trends in Neurosciences*, *28*(7), 353–363. https://doi.org/10.1016/j.tins.2005.05.005

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's Area in Language, Action, and Music. *Annals of the New York Academy of Sciences*, *1169*(1), 448–458. https://doi.org/10.1111/j.1749-6632.2009.04582.x

Fant, G. (1971). *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*. Walter de Gruyter.

Favaro, L., Gamba, M., Cresta, E., Fumagalli, E., Bandoli, F., Pilenga, C., Isaja, V., Mathevon, N., & Reby, D. (2020). Do penguins' vocal sequences conform to linguistic laws? *Biology Letters*, *16*(2), 20190589. https://doi.org/10.1098/rsbl.2019.0589

Fitch, W. T. (1988). *Paradox of Rhythm*. 1–17.

Fitch, W. T. (2012). The biology and evolution of rhythm: Unravelling a paradox. In *Language and Music as Cognitive Systems*. https://doi.org/10.1093/acprof:oso/9780199553426.003.0009

Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140091. https://doi.org/10.1098/rstb.2014.0091

Fitch, W. T., & Martins, M. D. (2014). Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, *1316*(1), 87–104. https://doi.org/10.1111/nyas.12406

Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex (New York, N.Y.: 1991)*, *21*(10), 2357–2365. https://doi.org/10.1093/cercor/bhr022

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., Bonadonna, G., Friard, O., & Giacoma, C. (2016). The Indris Have Got Rhythm! Timing and Pitch Variation of a Primate Song Examined between Sexes and Age Classes. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00249

Gámez, J., Yc, K., Ayala, Y. A., Dotov, D., Prado, L., & Merchant, H. (2018). Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Annals of the New York Academy of Sciences*, *1423*(1), 396–414. https://doi.org/10.1111/nyas.13671

Geissmann, T. (1999). Gibbon Songs and Human Music from an Evolutionary Perspective. In S. Brown, B. Merker, & C. Wallin (Eds.), *The Origins of Music*. The MIT Press. https://doi.org/10.7551/mitpress/5190.003.0011

Geissmann, T., & Orgeldinger, M. (2000). The relationship between duet songs and pair bonds in siamangs, Hylobates syndactylus. *Animal Behaviour*, *60*(6), 805–809. https://doi.org/10.1006/anbe.2000.1540

Grabe, E., & Low, E. L. (2022). Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in Laboratory Phonology*, *7*(1982), 515–546.

Grahn, J. A. (2012). Neural Mechanisms of Rhythm Perception: Current Findings and Future Perspectives. *Topics in Cognitive Science*, *4*(4), 585–606. https://doi.org/10.1111/j.1756-8765.2012.01213.x

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, *113*(19). https://doi.org/10.1073/pnas.1522072113

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, *6*(7), 663–668. https://doi.org/10.1038/nn1080

Hebets, E. A., Barron, A. B., Balakrishnan, C. N., Hauber, M. E., Mason, P. H., & Hoke, K. L. (2016). A systems approach to animal communication. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826), 20152889. https://doi.org/10.1098/rspb.2015.2889

Heesen, R., Hobaiter, C., Ferrer-i-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896), 20182900. https://doi.org/10.1098/rspb.2018.2900

Honing, H., Bouwer, F. L., Prado, L., & Merchant, H. (2018). Rhesus Monkeys (*Macaca mulatta*) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis. *Frontiers in Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnins.2018.00475

Kello, C. T., Dalla Bella, S., & Médé, B. (2017). *Hierarchical temporal structure in music, speech and animal vocalizations: Jazz is like a conversation, humpbacks sing like hermit thrushes*.

Kohler, K. J. (2009). Rhythm in Speech and Language. *Phonetica*, *66*(1–2), 29–45. https://doi.org/10.1159/000208929

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018a). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018b). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *The Journal of the Acoustical Society of America*, *73*(3), 1003–1010. https://doi.org/10.1121/1.389148

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. J. (2015). Speech-

Like Rhythm in a Voiced and Voiceless Orangutan Call. *PLoS ONE*, *10*(1), e116136. https://doi.org/10.1371/journal.pone.0116136

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*(4), 2142–2152. https://doi.org/10.1016/j.neuroimage.2003.08.016

Liao, D. A., Brecht, K. F., Johnston, M., & Nieder, A. (2022). Recursive sequence generation in crows. *Science Advances*, *8*(44), eabq3356. https://doi.org/10.1126/sciadv.abq3356

London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

McAndrew, F. T. (2019). Costly Signaling Theory. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 1–8). Springer International Publishing. https://doi.org/10.1007/978-3-319-16999-6_3483-1

Mcauley, J. (2010). Tempo and Rhythm. In *Music Perception* (pp. 165–199). https://doi.org/10.1007/978-1-4419-6114-3_6

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00274

Merker, B. (1999). Synchronous chorusing and the origins of music. *Musicae Scientiae*, *3*(1_suppl), 59–73. https://doi.org/10.1177/10298649000030S105

Merker, B. H., Madison, G. S., & Eckerdal Gamla Kyrkv, P. (2009). *On the role and origin of isochrony in human rhythmic entrainment*. *45(1)*, 4–17. https://doi.org/10.1016/j.cortex.2008.06.011

Michon, J. A., & Jackson, J. L. (1985). *Time, Mind, and Behavior*. Springer Science & Business Media.

Mitani, J. C. (1985a). Sexual selection and adult male orangutan long calls. *Animal Behaviour*, *33*(1), 272–283. https://doi.org/10.1016/S0003-3472(85)80141-X

Mitani, J. C. (1985b). Gibbon Song Duets and Intergroup Spacing. *Behaviour*, *92*(1/2), 59–96.

Mosing, M. A., Verweij, K. J. H., Madison, G., Pedersen, N. L., Zietsch, B. P., & Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior*, *36*(5), 359–366. https://doi.org/10.1016/j.evolhumbehav.2015.02.004

Nakada, T., Fujii, Y., Suzuki, K., & Kwee, I. L. (1998). ‘Musical brain’ revealed by high-field (3 Tesla) functional MRI. *NeuroReport*, *9*(17), 3853.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), Article 7. https://doi.org/10.1038/nn1082

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press, USA.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal. *Current Biology*, *19*(10), 827–830. https://doi.org/10.1016/j.cub.2009.03.038

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*(1), 1–32.

https://doi.org/10.1016/j.cognition.2005.11.004

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, *10*(18), 3979.

Pinker, S. (1997). *How the mind works*. NY Norton.

Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, *31*(3), 599–611. https://doi.org/10.1016/j.wocn.2003.08.001

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R. S., & Wiggins, G. (2020). *Multilevel rhythms in multimodal communication* [Preprint]. Open Science Framework. https://doi.org/10.31219/osf.io/psmhn

Pozzi, L., Hodgson, J. A., Burrell, A. S., Sterner, K. N., Raaum, R. L., & Disotell, T. R. (2014). Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *75*, 165–183. https://doi.org/10.1016/j.ympev.2014.02.023

Raemaekers, J. J., Raemaekers, P. M., & Haimoff, E. H. (2016). *Loud Calls of the Gibbon (Hylobates lar): Repertoire , Organisation and Context Stable URL : http://www.jstor.org/stable/4534384 REFERENCES Linked references are available on JSTOR for this article: You may need to log in to JSTOR to access the linked*. *91*(1), 146–189.

Raimondi, T., Di Panfilo, G., Pasquali, M., Zarantonello, M., Favaro, L., Savini, T., Gamba, M., & Ravignani, A. (2023). Isochrony and rhythmic interaction in ape duetting. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1990), 20222244. https://doi.org/10.1098/rspb.2022.2244

Ramus, F., Nespor, M., & Mehler, J. (1999). *Correlates of linguistic rhythm in the speech signalq*.

Ravignani, A. (2019). Rhythm and synchrony in animal movement and communication. *Current Zoology*, *65*(1), 77–81. https://doi.org/10.1093/cz/zoy087

Ravignani, A., & Madison, G. (2017). The paradox of isochrony in the evolution of human rhythm. *Frontiers in Psychology*, *8*(NOV), 1–13. https://doi.org/10.3389/fpsyg.2017.01820

Ravignani, A., & Norton, P. (2017). Measuring rhythmic complexity: A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution*, *2*(1), 4–19. https://doi.org/10.1093/jole/lzx002

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal behaviour*, *65*(3), 519-530.

Repp, B. H., Luke Windsor, W., & Desain, P. (2002). Effects of Tempo on the Timing of Simple Musical Rhythms. *Music Perception*, *19*(4), 565–593. https://doi.org/10.1525/mp.2002.19.4.565

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research*, *68*(4), 252–270. https://doi.org/10.1007/s00426-003-0143-8

Rilling, J. K. (2014). Comparative primate neurobiology and the evolution of brain language systems. *Current Opinion in Neurobiology*, *28*, 10–14. https://doi.org/10.1016/j.conb.2014.04.002

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology*, *30*(18), 3544-3555.e6. https://doi.org/10.1016/j.cub.2020.06.072

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology*, *19*(10), 831–836. https://doi.org/10.1016/j.cub.2009.03.061

Schwartze, M., & Kotz, S. A. (2013). A dual-pathway neural architecture for specific temporal prediction. *Neuroscience & Biobehavioral Reviews*, *37*(10, Part 2), 2587–2596. https://doi.org/10.1016/j.neubiorev.2013.08.005

Searcy, W. A., & Andersson, M. (1986). Sexual Selection and the Evolution of Song. *Annual Review of Ecology and Systematics*, *17*, 507–533.

Smith, E. A., & Bird, R. L. B. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and Human Behavior*.

Song, J., Sun, R., Fang, K., Zhang, B., & Fang, G. (2020). Flexibility as a Strategy for Avoiding Call Overlap in Male Anhui Treefrogs. *Asian Herpetological Research*, *11*, 230–239. https://doi.org/10.16373/j.cnki.ahr.1900331

Sullivan, B. K. (1983). Sexual selection in Woodhouse's toad (Bufo woodhousei). II. Female choice. *Animal Behaviour*, *31*(4), 1011–1017. https://doi.org/10.1016/S0003-3472(83)80006-2

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2020). The syntax–semantics interface in animal vocal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180405. https://doi.org/10.1098/rstb.2018.0405

Terleph, T. A., Malaivijitnond, S., & Reichard, U. H. (2018). Male white-handed gibbons flexibly time duet contributions. *Behavioral Ecology and Sociobiology*, *72*(1), 1–13. https://doi.org/10.1007/s00265-017-2432-z

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, *4*(1). https://doi.org/10.1515/lp-2013-0005

Valente, D., De Gregorio, C., Favaro, L., Friard, O., Miaretsoa, L., Raimondi, T., Ratsimbazafy, J., Torti, V., Zanoli, A., Giacoma, C., & Gamba, M. (2021). Linguistic laws of brevity: Conformity in Indri indri. *Animal Cognition*. https://doi.org/10.1007/s10071-021-01495-3

Van Den Broek, E. M. F., & Todd, P. M. (2009). Evolution of rhythm as an indicator of mate quality. *Musicae Scientiae*, *13*(2_suppl), 369–386. https://doi.org/10.1177/1029864909013002161

Van Den Broek, E., & Todd, P. M. (2003). Piep Piep Piep – Ich Hab' Dich Lieb: Rhythm as an Indicator of Mate Quality. In W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, & J. T. Kim (Eds.), *Advances in Artificial Life* (Vol. 2801, pp. 425–433). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-39432-7_45

Wilson, B., & Petkov, C. I. (2011). Communication and the Primate Brain: Insights from Neuroimaging Studies in Humans, Chimpanzees and Macaques. *Human Biology*, *83*(2), 175–189.

Xing, J., Sainburg, T., Taylor, H., & Gentner, T. Q. (2022). Syntactic modulation of rhythm in Australian pied butcherbird song. *Royal Society Open Science*, *9*(9), 220704. https://doi.org/10.1098/rsos.220704

Zornik, E., & Kelley, D. B. (2011). A neuroendocrine basis for the hierarchical control of frog courtship vocalizations. *Frontiers in Neuroendocrinology*, *32*(3), 353–366. https://doi.org/10.1016/j.yfrne.2010.12.006

Zuberbühler, K. (2018). Combinatorial capacities in primates. *Current Opinion in Behavioral Sciences*, *21*, 161–169. https://doi.org/10.1016/j.cobeha.2018.03.015

AKNOWLEDGMENTS

A special thanks to Prof. Marco Gamba, my tutor and mentor, whose support has been decisive in every phase of my PhD. Working with him has been a constant and inspiring learning process. Over the years, Marco has forged my idea of a scientific and personal role model, to which I will aspire in the years to come. It is impossible to express how much his daily presence along this path, with wisely modulated toughness and gentleness, has determined more than anything else the personal and professional success that this PhD has represented for me.

Thanks to the whole DBIOS Ethology Lab: in you, I have found formidable colleagues and friends with whom I've been sharing such precious moments. It has been amazing, over the course of time, to bring together the characteristics of each of us, through the experience that comes from sharing entire days, at work and outside. Getting to know each of you was a real pleasure. It was good to see this Lab become not only the sum of our (INNUMEROUS) complementary qualities, but something more.

Thanks to Prof. Andrea Ravignani and Prof. Adriano Lameira, for sharing their knowledge and time, for allowing me to grow scientifically, with their guidance, for making me feel valued, for transmitting their enthusiasm for their job, involving me in thrilling projects. They have been a special reference point that I never took for granted.

Thanks to Prof. Cristina Giacoma, who set up a massive and exceptional project that led to the constitution of the research centre in Maromizaha (Madagascar). Thanks to her, many people, including myself, were able to live the unforgettable experience of living in a primary forest and working first-hand for its conservation.

Thanks to the extraordinary research guides, the GERP staff (Groupe d’Etude et de Recherche sur les Primates), the inhabitants of Anevoka, and all those people who make fieldwork and conservation in the Maromizaha Protected Area possible.

Thanks to my mum, my dad, my brother, and Léo, who, with their love, have given me the happiest substrate and balance, allowing me to invest my energies serenely in my studies. I hope my love for you is one of those certainties you will never doubt.

Thanks to my friends and loved ones. Thanks to the ones who stayed at my side despite the geographical distance, despite the crazy working hours, who decided to share with me the burden and joys of the last years. Thanks for showing how many different ways there are to love each other and to share pieces of our lives.