Animal brains can follow the beat : the neuro-evolution of rhythmic acoustic processing

Piette, Théophane

How to cite

PIETTE, Théophane. Animal brains can follow the beat : the neuro-evolution of rhythmic acoustic processing. Doctoral Thesis, 2025. doi: 10.13097/archive-ouverte/unige:186384

This publication URL: https://archive-ouverte.unige.ch/unige:186384
Publication DOI: 10.13097/archive-ouverte/unige:186384

UNIVERSITÉ
DE GENÈVE
FACULTY OF MEDICINE

nccr evolving language
NATIONAL CENTRE OF COMPETENCE IN RESEARCH

LEMANIC NEURO SCIENCE
DOCTORAL SCHOOL

DOCTORAT EN NEUROSCIENCES
des Universités de Genève
et de Lausanne

UNIVERSITÉ DE GENÈVE
UNIL | Université de Lausanne

UNIVERSITÉ DE GENÈVE FACULTÉ DE MÉDECINE

Professeure Anne-Lise Giraud, directrice de thèse
Professeur Didier Grandjean, co-directeur de thèse
Dr Éloïse Déaux, co-directrice de thèse

TITRE DE LA THESE

ANIMAL BRAINS CAN FOLLOW THE BEAT :

THE NEURO-EVOLUTION OF RHYTHMIC ACOUSTIC PROCESSING

THESE
Présentée à la Faculé de Médecine
de l'Université de Genève
pour obtenir le grade de
Docteur en Neurosciences

par

Théophane PIETTE

de Rang-du-Fliers (France)

Thèse N° **439**

Genève

Editeur ou imprimeur : Université de Genève

2025

TABLE OF CONTENTS

REMERCIEMENTS

Que serait une thèse sans des remerciements, pour toutes ces personnes qui m’ont accompagné tout au long de cette aventure de quatre années incroyables, parfois longues, mais toujours passionnantes !

Évidemment, je tiens à commencer par remercier la Prof. Anne-Lise Giraud, superviseuse d’un style inimitable, sans qui rien de tout cela n’aurait été possible. C’est elle qui m’a offert l’opportunité de me lancer dans cette aventure de thèse, en me permettant de combiner ma passion pour le comportement animal et mon émerveillement pour le fonctionnement du cerveau. Je voudrais la remercier pour sa confiance sans faille, son enthousiasme à toute épreuve, et son ouverture aux idées, souvent un peu folles, qui m’ont emmené bien loin.

Comment ne pas remercier aussi ma co-superviseuse, la Dr. Éloïse Déaux, pour ses conseils avisés lors de la rédaction des articles, la mise en place des protocoles, et ses questions toujours pertinentes qui m’ont poussé à réfléchir davantage, à penser aux détails. Mais aussi pour nos sessions EEG sur Thalya dans les montagnes savoyardes, cette jeune chienne fougueuse qui a bien supporté nos petites (et grandes) bêtises humaines.

Un grand merci au Prof. Didier Grandjean et au Dr. Adrien Meguerditchian, qui me soutiennent depuis longtemps, bien avant que cette thèse ne commence. Grâce à eux, j’ai pu découvrir que la primatologie et les neurosciences peuvent coexister, et j’ai eu la chance de suivre mes baby baboons (Noé forever) pendant ces quatre années. Au-delà de l’aspect scientifique, je tiens aussi à remercier tous mes superviseurs pour leur bonne humeur inébranlable et nos soirées inoubliables qui m’ont rappelé qu’une thèse, ce n’est pas que du travail.

Je remercie également tout le personnel de primatologie à Marseille et du Wolf Science Center à Vienne, sans qui tout cela n’aurait pas été possible.

Un énorme merci à tous mes collègues. Comme je l'ai dit dans une interview pour l'UNIGE, ce sont bien plus que des collègues de travail, ce sont de véritables amis qui m'ont soutenu, supporté (franchement, respect !), suivi dans mes projets fous, et à qui je sais que je ne dis pas adieu en rendant cette thèse.

Un grand merci à Sarah, pour son énergie, sa bonne humeur, ses photos, et nos débats sur les plus beaux designs de présentation. À Émilie, celle avec qui tout a commencé, nous deux seules dans les bureaux alors que le COVID terrorisait la terre entière. À Naïma pour ses idées folles, ses conseils et ses remèdes de grand-mère. À Coralie, dont je ne cesserai jamais de vouloir prendre la place. À Mathieu, qui s'est retrouvé embarqué dans une histoire d'EEG sur babouins et un voyage en Afrique du Sud alors qu'il n'avait rien demandé. À Sam pour m'avoir supporté pendant les six mois les plus intenses de ma vie. À Irene (pour avoir sauvé mes plantes), Nosrat et Kinkini (pour avoir accepté ma haine de l'ail), Mamady (ou peut-être Thomas), Tim (le doopleganger), Celia, Mael, Soraya, Martina, Alexandra (Marmie girl), Arthur…

Je tiens aussi à remercier ceux que je connais depuis bien plus longtemps et qui ont toujours été là pour moi : Will et Laurène, qui, je le sais, viendront me voir présenter ce travail, et dont je n'oublierai jamais ces années de licence passées ensemble. Johanna, que la vie a bien fait de mettre dans le même appartement que moi à Grenoble, et qui a su révéler l'artiste en moi. Et bien évidemment, ma famille, qui me soutient depuis 29 ans maintenant, ne sachant pas toujours où je vais, mais m'aidant à chaque étape, et sans qui je ne serais pas là aujourd'hui.

Et finalement, je voudrais aussi remercier tous les animaux, participants parfois malgré eux de mes petites expériences, mais qui ne m'en ont jamais tenu rigueur. Merci à vous, babouins (Fidji, Fabienne, Feline, Formule, Ozone, Ota, Pepita, Geraldine, Chet et Gomme), chienne (Thalya), loup (Amarok), lapin (Balthazar), pigeon et cochon.

PUBLISHED ARTICLES

First Author

T. Piette[†], C. Cathcart[†], C. Barbieri, K. M. Ming’ D. Grandjean, B. Bickel, E. C. Déaux*, A-L. Giraud[*] (2024). A universal rhythm in animal acoustic communication. BioRxiv https://doi.org/10.1101/2024.10.07.616955. Under Review in Science Advances.

[†] Joint first authors

* Joint last authors

T. Piette, C. Lacaux, M. Scheltienne, V. Sterpenich, M.Isnardon, V.Moulin, D. Grandjean, A. Meguerditchian, E.C Déaux, A-L.Giraud (2025). Light propofol anaesthesia for non-invasive auditory EEG recording in unrestrained non-human primates. BioRxiv https://doi.org/10.1101/2025.03.24.644890. Submitted in Hearing.

Author

Déaux EC, Piette T, Gaunet F, Legou T, Arnal L, A-L.Giraud (2024) Dog–human vocal interactions match dogs’ sensory-motor tuning. PLOS Biology https://doi.org/10.1371/journal.pbio.3002789

ABSTRACT (English)

Acoustic communication is a fundamental aspect of animal behavior, shaping interactions across a wide range of species. While human speech relies heavily on rhythmic structures for comprehension, the role of rhythm in animal vocalizations and its evolutionary origins remain largely unexplored. This thesis therefore investigates rhythm in animal communication and its potential neural and evolutionary underpinnings. By analyzing vocalizations from 98 species, spanning mammals, birds, amphibians, insects, reptiles, and fish, we uncover a conserved rhythmic structure in animal acoustic communication around 2.9 Hz, within the delta frequency range of brain oscillations. This challenges the notion that rhythm is primarily influenced by biomechanical constraints and suggests an ancient evolutionary pressure shaping communication rhythms across species, likely related to the neural processing of sound. To explore the neural basis of rhythm perception, we then developed a non-invasive auditory passive listening EEG protocol using light propofol anesthesia in non-human primates. Our findings reveal that olive baboons, like humans, exhibit neural entrainment to rhythmic vocalizations, supporting the idea that brain oscillations play a crucial role in structuring acoustic communication. Moreover, we provide evidence that as humans do for speech, baboon auditory processing operates on three timescales, delta (1–4 Hz), theta (4–8 Hz), and low-gamma (25–55 Hz), despite baboon rarely using theta oscillations in conspecific vocalizations. This therefore suggests that the capacity for theta-based processing was already present but not yet fully integrated into their natural vocal communication system, implying that the emergence of a three-timescale hierarchical organization, essential for speech processing, might be an evolutionary development of the primate lineage. Finally, we investigated the conservation of roughness, characterized by fast amplitude modulations between 30 and 150 Hz, particularly in distress calls of bird offspring. We reported a wide conservation of the use of roughness is these calls, suggesting that roughness is an ancestral mechanism for grabbing attention, playing a key role in high-arousal contexts. Taken together, our results suggest that the processing of rhythmic structures in vocal communication is a deeply conserved trait, predating the

emergence of complex language in humans. By linking universal rhythms in animal vocalizations to neural entrainment mechanisms and attention-grabbing features like roughness, this work provides new insights into the neurobiological foundations of acoustic communication and the evolutionary roots of speech, suggesting that the rhythmic characteristics of speech were inherited from ancient and conserved acoustic processing mechanisms.

ABSTRACT (French)

La communication acoustique est un élément clé du comportement animal, influençant les interactions entre de nombreuses espèces. Alors que la compréhension de parole humaine repose largement sur des structures rythmiques, le rôle du rythme dans les vocalisations animales et ses origines évolutives restent encore incertaines. C'est pourquoi cette thèse explore le rythme dans la communication animale et ses potentielles bases évolutives. En étudiant les vocalisations de 98 espèces, allant des mammifères aux poissons, nous avons identifié un rythme universel de la communication animale autour de 2,9 Hz, dans la plage de fréquence des oscillations cérébrales delta. Ces résultats remettent en question l'idée selon laquelle le rythme serait influencé par des contraintes biomécaniques et suggèrent qu'une pression évolutive ancienne pourrait avoir façonné ces rythmes de communication, probablement liée au traitement du son par le cerveau. Afin de mieux comprendre les bases neuronales de la perception du rythme, nous avons mis au point un protocole EEG non invasif d'écoute passive sous légère sédation au propofol chez des primates non humains. Nos résultats montrent que les babouins olive, comme les humains, présentent une synchronisation neuronale avec les signaux rythmiques, ce qui soutient l'idée que les oscillations cérébrales jouent un rôle important dans le traitement de la communication acoustique. De plus, nous avons observé que comme les humains, les babouins traitent les vocalizations sur trois échelles temporelles, delta (1-4 Hz), thêta (4-8 Hz) et gamma (25-55 Hz), bien qu'ils n'utilisent pas ou très peu de rythme thêta dans leurs vocalisations. Cela suggère donc que la capacité à traiter des rythmes thêta était déjà présente chez les babouins, ce qui laisse penser que l'apparition d'une organisation à trois échelles temporelles pourrait être un développement évolutif propre aux primates. Enfin, nous avons exploré la conservation de la rugosité, caractérisée par des modulations rapides de l'amplitude entre 30 et 150 Hz, en particulier dans les appels de détresse des oisillons. Nous avons constaté que cette rugosité est largement utilisée dans ces appels, suggérant qu'elle pourrait être un mécanisme ancestral de capture de l'attention. En résumé, nos résultats suggèrent que

le traitement des structures acoustiques rythmiques par les oscillations cérébrales est un trait largement conservé, bien avant l'apparition du langage chez les humains. En mettant en lien la découverte d'un rythme universel de la communication acoustique animale avec les mécanismes neuronaux d'entraînement et des éléments tels que la rugosité, cette recherche propose une nouvelle perspective sur les bases neurobiologiques de la communication sonore. Elle suggère que les caractéristiques rythmiques de la parole pourraient être héritées de mécanismes acoustiques anciens et communs à la plupart des espèces animales.

INTRODUCTION

THESIS FRAMEWORK

Acoustic communication is a fundamental behavior in the animal kingdom, used by a wide variety of species to convey information crucial for survival. Through sounds, animals can communicate with conspecifics, even at considerable distances, without the need for direct visual or tactile interaction. While rhythm is known to play a key role in human speech comprehension, its significance in animal communication and the selective forces driving its evolution remain poorly understood. Rhythmic patterns appear in the acoustic displays of many animal species, contributing to emotional communication, urgency, or mating rituals. Certain vocal-learning species, such as parrots, pinnipeds, and some primates, have demonstrated the ability to synchronize body movements with external acoustic rhythms which provides direct evidence of their capacity to process rhythm. Partly due to this ability, the study of rhythm in animal communication has largely focused on primates and songbirds, leaving gaps in our understanding of how rhythm evolved. However, many other species use rhythm in their vocalizations without displaying such motor entrainment, suggesting that rhythm perception may be a widespread and ancient ability.

The evolutionary significance of rhythm processing may lie in the role of brain oscillations, cyclical patterns of neural activity that structure how the brain detect and interpret salient stimuli in the environment. In humans, these oscillations are deeply integrated into speech perception, with slow oscillations underlying syllables and word processing, while fast oscillations are believed to be involved in the processing of phonemic information. The interaction between brain oscillations and acoustic stimuli has been underexplored in animals, partly due to the assumption that complex, hierarchical vocal structures are uniquely human. Quite interestingly, brain oscillations, particularly slow rhythms, are conserved across animal species and might therefore represent the evolutionary backbone of auditory processing. Investigating neural oscillations in the context of rhythm processing provides a valuable framework for tracing the roots of acoustic communication and understanding the shared evolutionary pathways that may connect human speech with other forms of animal communication.

By integrating perspectives from neurobiology, evolutionary theory, and cognitive science, this thesis provides a multidimensional understanding of rhythm as a crucial element of communication. It explores how rhythmic structures may have influenced the evolution of language and auditory processing, revealing shared principles that transcend species boundaries. Ultimately, these findings offer a novel perspective on the foundational aspects of animal communication, providing critical insights into the neural and evolutionary origins of rhythm. This work not only sheds light on the ancient roots of speech and language but also contributes to a broader understanding of how rhythmicity has shaped the development of complex communication systems throughout the animal kingdom.

COMMUNICATION

2.1 An Overview of Communication

If one were to explore the literature on animal communication, she would quickly realise that biologists have debated its definition for decades. In his book « The Evolution of Communication », Marc Hauser tells us that communication is defined as « the transfer of information from a signaler to a receiver » (*1*) . As simplistic as this definition might seem, it does encompass the three major aspects of communication that most biologists agree upon.

First, there has to be a sender, and at least one receiver for communication to happen. Krebs and Dawkins (1984) went as far as describing communication as a game of « manipulation » from the sender, which is trying to make the receiver perform an action that would grant him benefit, and «mind-reading» from the receiver that is trying to use communication to predict the behaviour of the sender (*2*). This view of communication directly relates to the second point of agreement, that, communication involves a signal carrying information. Robert Seyfart (2010) stated that « the concept of information plays a central role in studies of animal communication ». Animals' responses to different types of signals suggest their ability to extract information and use them to enhance their fitness, putting the receiver on the same level of importance as the sender in communicative interactions (*3*). Finally, these signals must have been shaped for communication purposes by natural selection (*4*, *5*). Let's take the example of a pack of wolves finding deer's tracks in the snow. These tracks would give the wolves vital information on their prey's presence, their number, and the direction they took. Still, no one would argue that deers were communicating this information to wolves. These tracks in the snow are cues, elements that are by-products of an animal's behavior, carrying information, but unlike signals, have not been selected for communication purposes.

By its definition, communication is universal and present in every aspect of animals' daily lives. It allows for the exchange of relevant information between individuals and groups and is the core of every social organization.

2.2 Modalities in Animal Communication

Communication being ubiquitous, with research even claiming its existence in plants and fungi (*6–8*), it can take many forms depending on the species, situation and information that needs to be carried. Animals communicate using an amazing diversity of signals, that can be tactile (*9–11*), chemical (*12*, *13*), visual (*14*, *15*), acoustic (*16*, *17*) or even electrical (*18*).

While all modalities of communication are equally important and finely tuned to the ecological and social needs of each species, this thesis will primarily focus on acoustic communication. Acoustic signals are widespread in our daily environment and present in the communication systems of most vertebrates (*19–21*). By using sound, an undulation that propagates in every environment, it allows for the instantaneous exchange of information between individuals who cannot see, smell or touch each other, even at considerable distances. Sounds can also be quickly modulated, allowing for dynamic variation of information, making it a perfect tool for communication and adaptive response. In humans, spoken language plays a central role in shaping social interactions, culture, and cognition. Similarly, many vertebrates rely heavily on acoustics for essential and vital functions. This universality makes acoustic communication a valuable lens through which to explore broader questions about the evolution and function of communicative systems.

2.3 Chapter Summary

Communication is the foundation of interaction in the animal world, shaping relationships between organisms through the exchange of information. By focusing on the versatility and ubiquity of acoustic signals, we could gain insight into how communication systems evolve to meet the ecological and social needs of species.

ACOUSTIC COMMUNICATION

3.1 Sound Production

Sound being produced by almost every tetrapod, from fishes to mammals and birds, as well as some insects, there is an incredible diversity of production mechanisms. The signals produced by these mechanisms are often placed in two main categories, as vocal and non-vocal.

Vocal signals are generated by the movement of air from the lungs through the vocal tract. In mammals and amphibians, this airflow passes through the larynx, causing the vocal cords to vibrate and produce sound waves (22). In birds, sound is produced in the syrinx, a complex, muscle-controlled cartilaginous structure located at the base of the trachea (23). As air flows through the syrinx, the muscles adjust their tension and shape, creating sound waves (Fig 1a). These waves travel through the vocal tract, where they are modified by the anatomy of the species, resulting in a vast range of vocal signals. Human speech, whale songs, bird chirps, and amphibian courtship calls are just a few examples of these. Reptiles, on the other hand, typically have simpler vocal tracts and often rely on rapid air expulsion, resulting in hissing, or growling sounds (24). Some fishes also produce sound by vibrating their swim bladders using specialized drumming muscles (Fig 2c), which are believed to share an evolutionary origin with the lungs and breathing muscles of the tetrapods (25) .

Non-vocal sounds encompass all acoustic signals not generated by the passage of air through the vocal tract. These signals are categorized based on their production methods into three types. Stridulation, for sounds produced by rubbing one body part, known as the scrapper, against a series of ridges or files, as seen in crickets' chirps. Percussion, by striking a body part against another object, like woodpeckers drumming or gorillas' chest-beating display, and tymballic, specific to some insect species like cicadas, who produce sounds through a specialised organ called a tymbal.

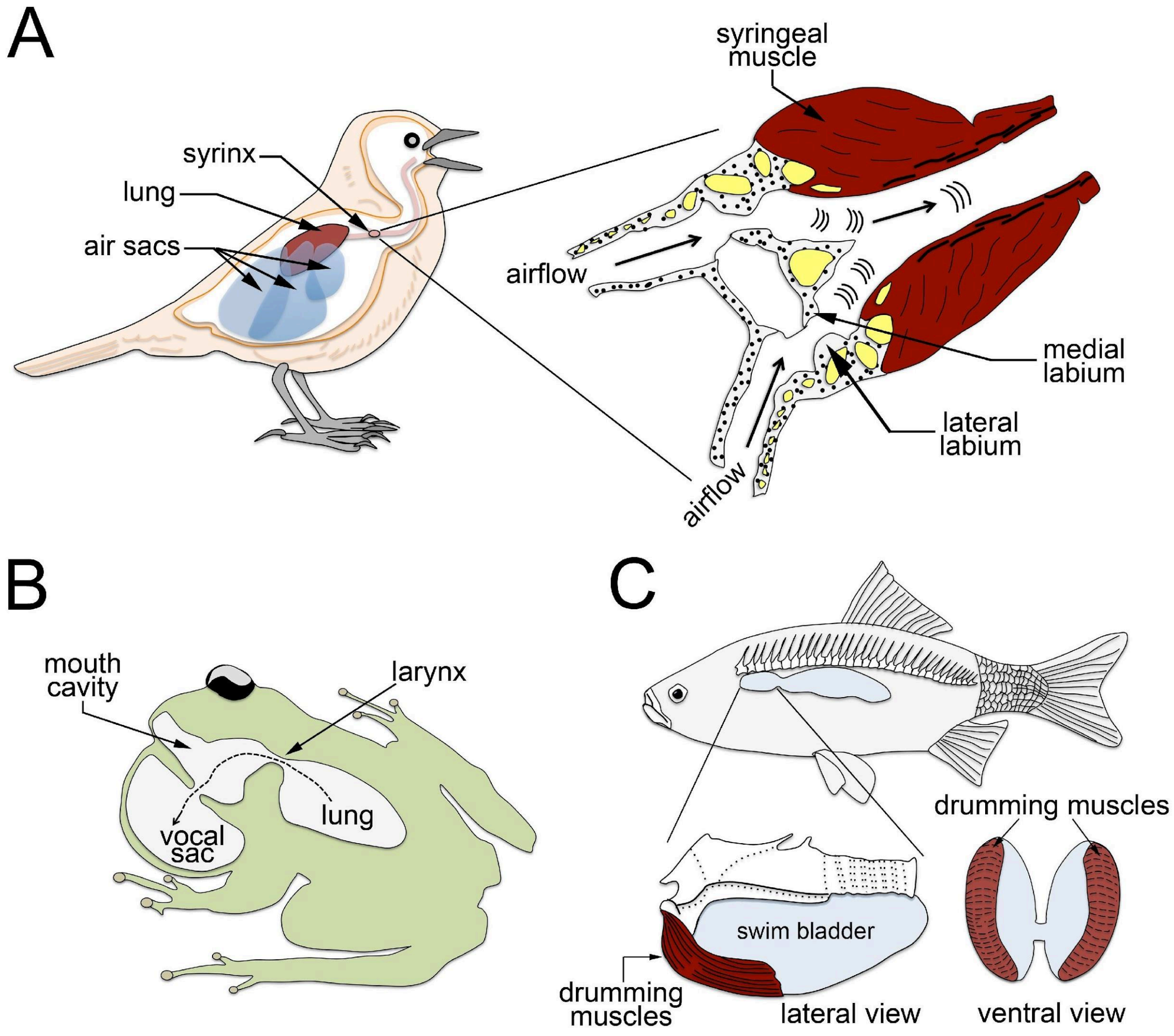


Figure 1: Comparative anatomy of the sound production apparatus in birds, amphibians, and fish. The swim bladder in fish, thought to be the precursor of terrestrial vertebrate lungs, is highlighted, alongside the complex muscular organization of the bird syrinx and the simpler vocal tract of amphibians. Adapted from (*19*).

3.2 Frequency Modulation in Animal Acoustic Communication

To be functional, signals must convey adaptive information that helps animals adjust to their environment or interact in ways that enhance their fitness. Acoustic signals, being sound waves, often rely on frequency modulation to encode and transmit important information. While non-vocal signals typically consist of simple, pure tones with a single fundamental frequency, vocal signals are generally more complex.

The source-filter theory tells us that the sound produced by the larynx or syrinx (the source) is modified as it travels through the vocal tract. This tract can alter the sound based on the anatomical properties of vocal parts such as the teeth, lips, or tongue (the filters). The placement of these organs changes the originally produced frequency, adding what are known as formants to the sounds. By introducing these additional frequency components, animals produce a richer and wider repertoire of sounds (*26*, *27*). In human speech, for example, the first and second formants help distinguish one phoneme from another (*28*) . Similarly, formants in meerkat (*Suricata suricatta)* barks convey information about the caller's identity and the presence of predators (*29*). In non-human primates, such as guinea baboons *(Papio papio*), grunts' formants serve to encode individual identity or emotional state.

Formants can also provide information about the size and sex of vocalizing animals due to a consistent allometric relationship between formant frequencies and body weight, observed both within and across species. This relationship. shaped by anatomical constraints is mainly linked to the length and structure of the vocal tract. (*30–33*). Larger animals have longer vocal tract, allowing them to produce sounds with lower formants, and vice versa. While formants are the best index of body size due to their direct link to vocal tract length, they tend to be difficult to measure, and are even absent in species using tonal sound (made of one single frequency) to communicate. Therefore, Fundamental frequency (F0), the original frequency produced by the vocal folds, and dominant frequency (dF), the frequency of highest amplitude in a call, are often used in comparative study, and show inter-species allometric relationship with weight (*34*, *35*) (Fig 2). Despite this relationship being less reliable, these frequencies remain valuable in animal communication. For example, Australian frogs (*Uperoleia rugosa*) use the dominant frequency of other males' calls to assess their fighting ability and decide whether to enter their territory (*36*). Similarly, male scops owls (*37*) and other bird species (*38*) use these frequencies to honestly communicate their size and fighting ability to defend or invade territory. In some species, the fundamental frequency plays a role in female mating choices. For instance, female red deer (*Cervus elaphus*) tend to prefer males with higher pitched roars (*39*) , and male scop owls (*Otus scops*) with lower hoots have higher reproductive success (*40*).

Variations in the fundamental frequency of calls in a single individual can also give information on his emotional state. In mammals, increases in Fo have been correlated with the caller's increased arousal , a characteristic particularly present in human speech, in which alerting screams tend to present higher fundamental frequency than neutral speech (*41*).

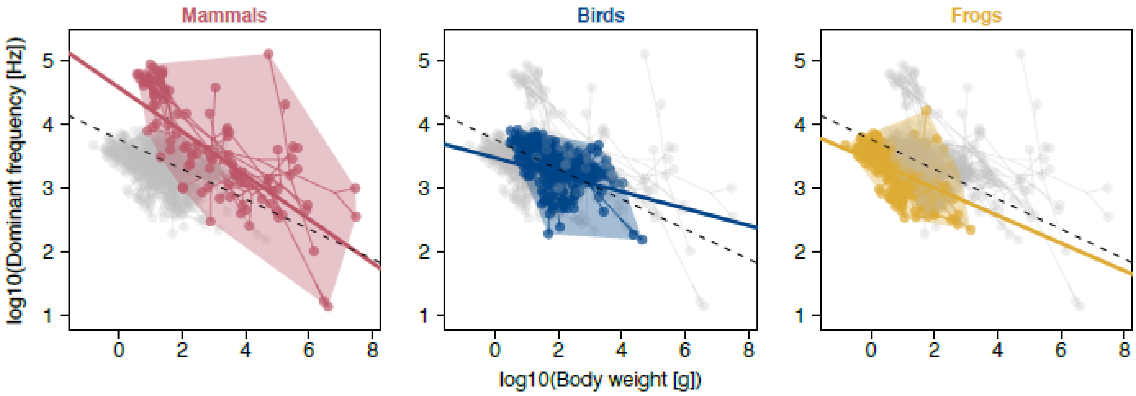


Figure 2 : Phylomorphospace showing the covariation between the body weight and dominant frequency of mammals (red, left), birds (blue, centre) and frogs (yellow, right). The estimated PGLS line for each clade is depicted as a solid-coloured line. For reference, the general tetrapod acoustic is depicted as a dashed black line in each panel. Adapted from (*34*)

3.3 Rhythm in Animal Acoustic Communication

Acoustic signals are not only spectrally but also temporally structured, with rhythm having communicative functions at least as important as spectral features. This is particularly well exemplified by human speech, whereby speech rate plays a key role in comprehension. In most languages, syllable production rate ranges between 4 to 9 syllables per second (*42*). Going above or below this range negatively impacts speech intelligibility by reducing the amount and quality of information extracted from the speech stream (*43*).

In animal calls, temporal features are no less important; for instance, rhesus monkeys *(Macaca mulatta)* often produce complex vocalizations consisting of short calls separated by silent pauses, known as inter-pulse intervals. Altering these intervals can disrupt the normal responses of other animals, eliminating the usual right ear orientation bias, and highlighting the importance of temporal patterns in vocal recognition (*44*). In certain species, temporal cues are also crucial for mating. Female canaries (*Serinus canaria*) show a strong preference for males with a faster syllabic rate (*45*). Similarly, female zebra finches (*Taeniopygia guttata*) remember the timing of male songs and may reject a male if there is a timing error during song production (*46*).

Like fundamental frequency, rhythm is closely linked to the emotional state of the signaler. When frightened, humans often speak more rapidly, a pattern observed across most mammals, in which increased calling rates are associated with higher arousal (*47*). In primates, individuals under attack may increase their call rate in an attempt to attract attention from others. This increase in call rate is also a common strategy for predator avoidance; in marmots and meerkats, the faster the rhythm of alarm calls, the closer the predator is (*48, 49*). In addition to regular rhythm, nonlinear temporal phenomena can enhance the urgency of alarm calls. Among these, roughness, rapid amplitude modulation between 30 to 150Hz, have been described in human screams, enhancing salience and making it more likely to attract attention and trigger a rapid response (*50, 51*) .

3.4 Combination, Compositionality and Human Speech

While frequency and rhythm play essential roles in acoustic communication, some animals also organize their calls into sequences, creating the potential for more complex patterns of information, enhancing the communicative richness of their vocal systems.

Some species can combine meaningless units into larger meaningful ones. This ability known as combinatoriality is the base for more complex auditory communication systems. Some animals, such as birds or whales, combine individual syllables or notes into unique songs and sequences (52, 53). Combination is also at the heart of human language. The smallest acoustic units and building blocks of spoken language are phonemes, which allow for the differentiation of meaning. For instance, phonemes like /d/ /ɒ/ and /ɡ/ can be combined to form the word "dog", and the difference between the words « beau » and « peau » in French only comes from one phoneme.

Even more complex, some species can build up sequences of meaningful units, in which the meaning of the sequence is derived from the meaning of its part. This ability, known as compositionality, is also at the core of human speech. Often set apart for its complexity and compositional nature, the hierarchical structure of language, which allows for the generation of an infinite variety of meanings from a finite set of elements, is a prime example of compositionality. When combined together, phonemes form syllables that typically contain a vowel and one or several surrounding consonants. Syllables are then combined into words, which will then be organized in sentences, following grammatical rules specific to each language. This hierarchical structure of language allows for incredible versatility. By rearranging and composing basic units together, speakers can generate an endless array of new meaningful ones, each conveying unique meanings (fig 3).

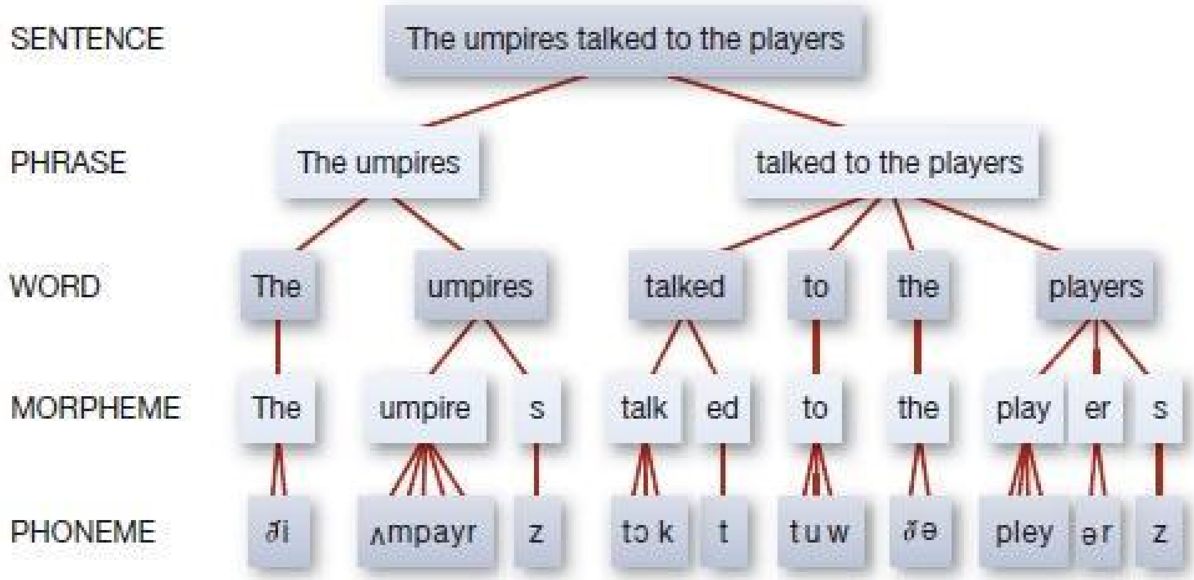


Figure 3 : Hierarchical structure showing how an english sentence is built by associating smaller subparts into bigger ones.

While compositionality was once thought to be unique to human language, recent studies have revealed its occurrence in primates and certain bird species. A well-known example is seen in male Campbell's monkeys (*Cercopithecus campbelli*), which produce distinct alarm calls such as 'krak' for leopards and 'hok' for eagles. These calls can be modified by adding an '-oo' suffix, resulting in 'krak-oo,' which indicates a broader range of disturbances, and 'hok-oo,' which signals non-ground-based threats (54). Similarly, in birds, Pied babblers (*Turdoides bicolor*) combine their typical alert and recruitment calls into a specific vocalization when faced with a terrestrial threat that requires mobbing (55).

3.5 Chapter Summary

Sound production across the animal kingdom, while varied in mechanisms, is unified by shared principles that enhance communication. Whether through vocal or non-vocal means, the modulation of frequency, rhythm, and compositional structures reflects common strategies animals use to convey critical information.

EVOLUTION OF ACOUSTIC COMMUNICATION

« There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved. »

Charles Darwin, On the Origin of Species.

4.1 Brief History of Evolution Theory

At the origin of evolution lies a fundamental question that philosophers have debated for centuries and continue to do so today: 'Where do humans come from?'. Despite being present quite early in the history of science, the evolutionary idea always faced strong resistance, opposed to the religious belief of mankind's godly creation. Influential philosophers from ancient Greece like Plato or Aristotle defended this idea of humans' divine origin and their special role in the cosmos, distinguishing them from other animals. These ideas were also strongly put forward by the rise of monotheist religion in the Middle Ages, often linked to the concept of vitalism or « breath of life».

It was only in the 18th century, during the Enlightenment, that they started to be challenged. The classification of species, done by Carl von Linnaeus in his book *« Systema Naturae » (1735),* as well as his binomial nomenclature of species, would be the start of a revolution in the century to come (56). Even though Linée was considered a fixist, who believed that species were created as such and could not be modified, his classification, surprisingly still widely accurate three centuries later, and his work on plant hybridization inspired the mind of his contemporary.

A few decades later, at the beginning of the 19th century, Jean-Baptiste Lamarck proposed the first theory of evolution, known as transformism or Lamarckism. Lamarck believed in the inheritance of acquired characteristics, meaning that characteristics acquired by an individual during his lifetime could be passed on to the next generation, pushed by an inherent drive to become more complex or perfect. He exemplified his theory using the giraffe's neck, which would have become longer over generations due to ancestors stretching their necks to reach higher leaves (Fig 4). Despite not endorsing the idea of a common lineage between species, Lamarck's work was highly influential in developing the modern view of evolution.

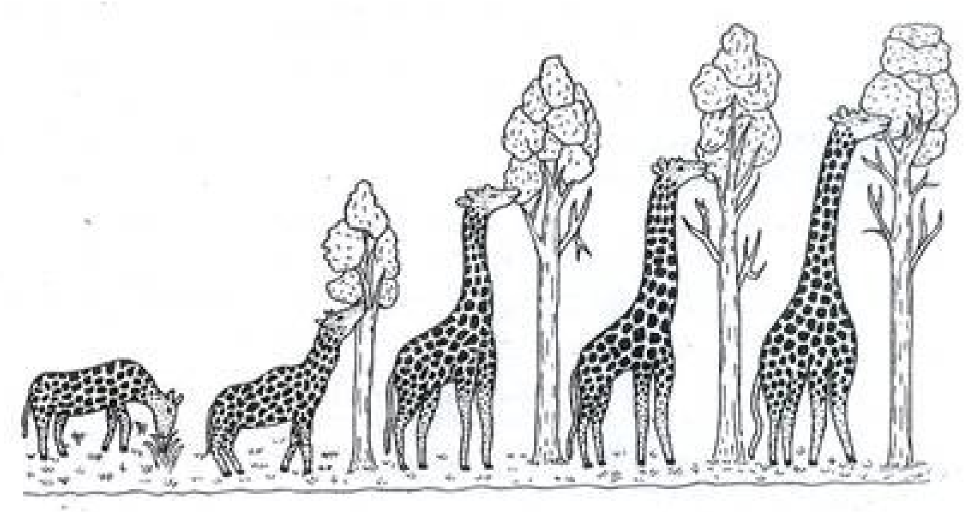


Figure 4 : Drawing of Lamarck Giraffe's theory, suggesting that giraffes gain long necks by stretching them trying to get to higher leaves. Adapted from (*57*)

Finally, in the middle of the 19th century the development of evolution theory reached its biggest milestones, carried by Alfred Wallace and Charles Darwin. Both travelers and in close contact during their entire careers, Darwin co-jointly presented their theory at the Linnean Society in London in 1858. Later, with the first publication of his famous book «On the Origin of Species » in 1959 Darwin established his now well-accepted theory of natural selection. For Darwin, the environment exerts selective pressures on species, favoring the survival and reproduction of individuals with advantageous traits. Building on Lamarck's idea, Darwin's theory suggests that giraffes born with longer necks would have better access to food, enhancing their fitness and reproductive success. Over time, this would result in a new generation of giraffes with even longer necks. These longer-necked individuals would breed with one another, progressively leading to giraffes with increasingly longer necks. Darwin postulates for a common lineage to all species, pictured through the « Tree of Life » in which each branch represents a lineage that splits into new species. In doing so, he implies the existence of a common ancestor for all species (*58*).

The discovery of Deoxyribonucleic Acid (DNA) by James Watson and Francis Crick a century later supported Darwin's theory of evolution by providing a concrete mechanism for how traits are inherited and passed on through generations (*59*). While Darwin's theory explained how natural selection favors individuals with beneficial traits, DNA revealed the genetic basis of these traits and their transmission from parents to offspring (*60*). This genetic understanding reinforced Darwin's ideas and allowed for the updating of Lineaus' classification, by comparing species' DNA to establish their lineage through the science of genetics instead of empirical observation and arbitrary decisions, giving birth to the phylogenetic classification of living organisms (Fig 5).

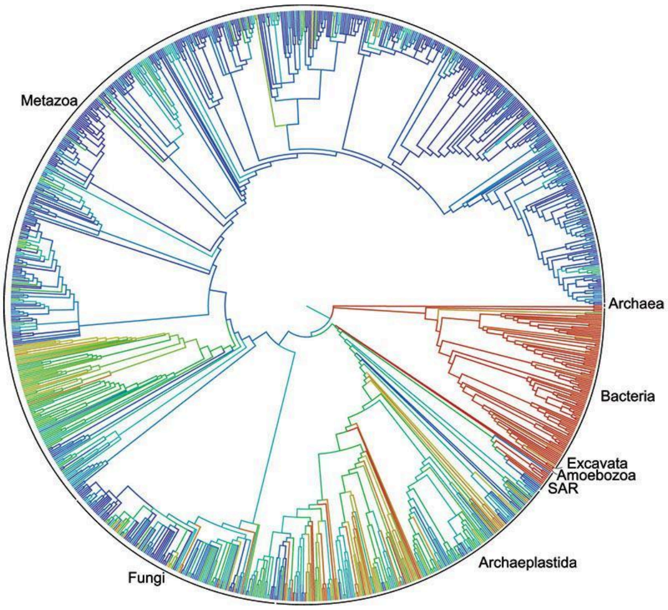


Figure 5 : Phylogenetic tree of life representing all living organisms and their last universal common ancestor (LUCA). Adapted from (*61*)

4.2 Phylogeny

In the phylogenetic classification of life, all species are organized into three domains: Archaea, Bacteria, and Eukarya, each of which is further divided into kingdoms (Fig 5). In biology, animals are defined as multicellular eukaryotic organisms belonging to the Animalia kingdom within the Eukarya domain. The Animalia kingdom is subdivided into 36 phyla (Arthropoda, Chordata ...), which are further divided into approximately 110 classes (Mammalia, Reptilia, ...). These classes are divided into around 1,200 orders (Carnivora, Anura, Squamata ...) and then into 5,000 to 6,000 families (Canidae, Alytidae, Gekkonidae ...), encompassing a total of about 1.5 to 2 million discovered species. To represent the relationships between species, we use tree-like representation, where each species is connected by branches. The length of these branches represents the phylogenetic distance between the species, and the nodes where branches meet represent the common ancestors of two or more species.

Phylogeny can be used to study the origins of specific traits. Let's consider echolocation, the ability to navigate the environment using sound, found in both dolphins and bats. There are two possible evolutionary scenarios for this trait. First, echolocation may have appeared once in the common ancestor of both dolphins and bats. Alternatively, echolocation could have evolved independently twice, once in the ancestor of all dolphins and once in the ancestor of all bats. By examining the presence or absence of echolocation across species that are positioned between dolphins and bats on the phylogenetic tree, researchers can determine the most likely scenario. If most of these intermediate species possess echolocation, it likely appeared in the common ancestor of dolphins and bats. On the other hand, if most of these species lack echolocation, as is the case here (*62*), it suggests that echolocation evolved independently in both lineages.

4.3 Phylogenetic Models

In cases of continuous traits, more complex phylogenetic models can help understand how these traits evolved over time, revealing patterns of conservation, diversification, or drift. Conservation indicates that a trait has remained relatively unchanged through evolutionary history, suggesting its fundamental importance. Diversification shows how a trait has varied, possibly adapting to different environments or ecological niches, highlighting the dynamic nature of evolutionary processes. Drift, on the other hand, refers to random changes in trait frequency, which can lead to traits that are neither strictly conserved nor diversified, but fluctuate unpredictably over generations.

To represent these evolutionary processes, mathematical models are often employed to simulate the evolution of traits under various assumptions. One of the most widely used and foundational models is based on Brownian motion (BM) processes. This model assumes that the evolutionary changes in a trait over time are driven by a series of small, independent, and normally distributed random steps, without any long-term directional trends or tendencies to revert to a specific state (59). As such, the BM model is particularly well-suited for representing processes like diversification or genetic drift. For example, the evolution of body mass in some primates lineage has been shown to follow a Brownian motion process, suggesting that changes in body size occur randomly across species without a clear, consistent directional trend (*63*)

However, a limitation of the BM process is that it is not well-suited to model conservation scenarios in which the evolutionary pressure persists over time and affects different branches similarly. To overcome this issue, Butler et al. (2004) introduced the Ornstein-Uhlenbeck (OU) process. Unlike BM, the OU model assumes that evolution is not only random, but also stabilized by an optimal value to which the trait tends to return (60). This makes the OU process particularly useful for modeling scenarios where traits tend to return to an evolutionary optimum due to stabilizing selection or other consistent environmental pressures. For example, The evolution of size and tooth morphology in canines can be modeled by an Ornstein-Uhlenbeck (OU) process, with phenotypic evolution being directed toward adaptively optimal morphologies, such as larger body size and broader molars (*64*).

4.4 Evolution of Acoustic Communication and Language

Acoustic communication being an essential trait for animal fitness and survival, it is no surprise to find it widespread across the phylogeny. It is present in insects, but also in every class of vertebrates, and believed to have evolved separately in these two branches.

In vertebrates, vocal communication was believed to have evolved separately in the different classes. However recent developments, notably establishing the use of vocalizations in most turtles and caecilians, have traced back its origin to the common ancestor of all choanate vertebrates 405 Million Years Ago (*65*).

While acoustic communication is widespread among species and exhibits remarkable diversity and complexity, it serves as a foundation upon which more intricate systems like human language may have evolved. As previously mentioned, the impressive combinatorial mechanism of language has led to debates about its origin and whether it is unique to humans. Historically, philosopher René Descartes stated that "language distinguishes us from other animals that cannot arrange their thoughts by which to make their thoughts known."

This view of language's uniqueness, strongly supported by Noam Chomsky's work, is still defended today but faces challenges from researchers like Tecumseh Fitch and others. Chomsky theorized that the ability to learn language is innate to humans, and humans alone, proposing a universal grammar, present from birth, that facilitates the acquisition of languages. He argues for a cognitive gap between humans and non-human animals, enabling the human brain to process complex and organized communication, while non-human animals cannot (*66*).

Fitch and colleagues, on the other hand, propose a gradual model of language evolution, where language traits would have emerged over time across various species (*67*). Vocal learning, as observed in species like California sea lions, is one such trait, where individuals modify their vocalizations based on auditory experience (*68*). This ability is considered a precursor to more advanced language skills, as language is not an innate communication but acquired through learning. Likewise, combinatoriality, seen in birds and primates, involves combining smaller vocal units, pointing to an incremental development toward language. For example, baboons produce syllable-like sounds, which could indicate early stages of speech-like production (*69*).

Behavioral and genetic studies, including the discovery of the FoxP2 gene and experiments attempting to teach language to chimpanzees and parrots, show that non-human animals cannot learn and use language with its complex rules, reinforcing the uniqueness of human language. However, acoustical and neuroscientific studies reveal a continuum of communication perception and production between humans and non-human animals. Evidence of brain asymmetry, such as the hypertrophy of the left planum temporale and the depth of the central sulcus, has been found in primates, suggesting shared neurological features (70–72) . Volitional control over vocalizations in crows (73), referential alarm calls in vervet monkeys (74), and the cultural transmission of songs in humpback whales (75) further strengthen the idea of this continuum. Instead of viewing these theories as opposing, recent research proposes a synthesis of both perspectives, incorporating elements of each.

In this context, rhythm being a key feature of speech comprehension in speech, its study in animal communication becomes crucial for understanding language evolution. While most research focuses on vocal learning species, those able to detect and synchronize motor actions to external rhythms, rhythmic abilities are also observed in non-vocal learning species like primates (76), marmots (77), and even non-vocal animals. This ability to perceive and produce rhythm may have been a key precursor to the development of more complex language systems in humans, further supporting the idea of a gradual evolution of language across species. Studying the evolution of rhythmic capabilities could therefore provide valuable insights into how language may have emerged in humans.

4.5 Chapter Summary

While the origin of acoustic communication remains unclear, the origin of vocal communication, crucial for survival and fitness, can be traced back to the common ancestor of all choanate vertebrates and later diversified across species. Although the gradual evolution of language from these vocalizations is still a theory, studying the development of rhythmic abilities and vocal learning in animals could provide valuable insights into how complex language may have gradually emerged in humans.

NEUROANATOMY OF ACOUSTIC COMMUNICATION

5.1 Anatomy of Ascending Auditory System Pathway

In mammals, despite some specialization in species using echolocation, the anatomy of the auditory system is quite conserved. The journey of sound starts after the cochlea, in which sounds will be transformed into an electrical signal that will travel through the auditory nerve. The auditory nerve will carry this signal toward the brain stem, in the Superior Olivary Complex (SOC) or the Cochlear Nuclear Complex (CNC) which will perform spatial analysis of sound sources. Afference from both CNC and SOC will then converge to the Inferior Colliculus (IC) in the midbrain, in which information carried by the sound will be further processed. From there, the signal will travel through the Medial Geniculate Nucleus (MGN) of the thalamus, which will serve as a relay station toward the Primary Auditory Cortex (A1) that will analyze the sound in terms of pitch, volume and rhythm (Fig 6). Finally, information can be sent to higher parts of the auditory cortex and associative areas for integration of more complex sensory information and cognitive functions (*78*).

The anatomy of the bird and reptiles' auditory pathway is quite similar. The auditory nerves travel through the same brainstem and midbrain nucleus toward the thalamus. From there, information will be sent to Field L, the bird equivalent of the primary auditory cortex (A1), which is further divided into five subregions, or in the auditory pallium in reptiles and amphibians. Finally, as it is the case in mammals, information can be sent to higher auditory and associative areas, that coordinate more complex functions such as vocal learning, song memory or species-specific call recognition (*79–81*).

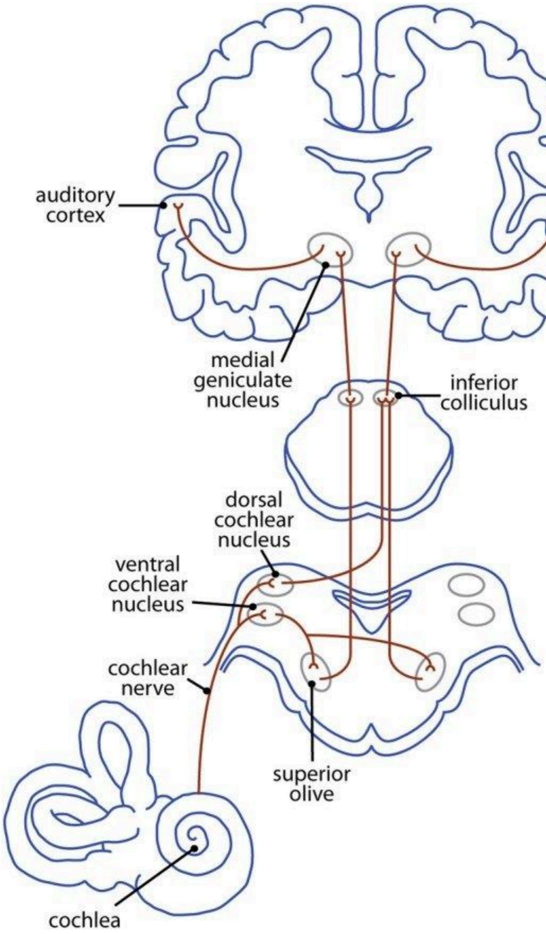


Figure 6 : Human ascending auditory pathway from the cochlea to the primary auditory cortex. Adapted from (*82*).

Due to the vast diversity of insect species, describing a single auditory pathway is challenging. However, the sound processing steps in insects are generally similar to those in reptiles and amphibians. The key difference is that while large nuclei of neurons handle these computations in other species, in insects, these tasks are often performed by one or small groups of neurons, likely due to size limitations (*83*) .

5.2 Brain Anatomy of Speech Processing

Historical research on the neural pathways underlying speech processing has emphasized the lateralization of language functions, particularly through the Wada test. In this procedure, researchers temporarily anesthetize one side of the brain to observe the specific contributions of each hemisphere to language processing. In most patients, anesthetizing the left hemisphere results in a loss of language production, a condition known as aphasia. In 1861, through studies of aphasic patients, French neuroscientist Paul Broca proposed the existence of a brain region responsible for speech production, located in the left Brodmann areas 44 and 45, which he named Broca's area (*84*). A few years later, in 1874, German neurologist Carl Wernicke, through studies of patients with impaired speech comprehension, identified a speech comprehension center in the left parietal lobe, later named Wernicke's area (*85*). Over the years, numerous functional and anatomical asymmetries related to language function have been described, such as the leftward asymmetry of the planum temporale (*86*) and the central sulcus (*87*).

While research on the left asymmetry of language and the roles of Broca's and Wernicke's areas dominated the field, studies like Robin et al.'s 1990 paper, which showed impairment of spectral information processing in patients with right auditory cortex lesions (*88*), or Belin et al.'s work showing bilateral activation during slow-changing acoustic transitions (*89*), contributed to the proposal of a dual-stream model of speech processing by Hickok and Poeppel. In this widely accepted model, speech is processed through two pathways. The dorsal pathway, primarily left-lateralized, encompasses key regions involved in transforming acoustic signals into articulatory representations for speech production. This pathway includes an articulatory network, such as the posterior inferior frontal gyrus and premotor cortex, and a sensorimotor interface, notably in planum temporale, which facilitates the mapping of auditory input onto motor plans. In contrast, the ventral pathway, organized bilaterally, supports speech comprehension. It includes a combinatorial network in the anterior temporal lobe (ATL), which integrates syntactic and semantic elements, and a lexical interface in the superior temporal sulcus (STS), linking sound to meaning. A core component of the ventral pathway is the superior temporal gyrus (STG), which processes incoming auditory information and plays a critical role in extracting phonetic and semantic content from speech (Fig. 8). Together, these pathways form a dynamic system for producing and understanding speech (*90*).

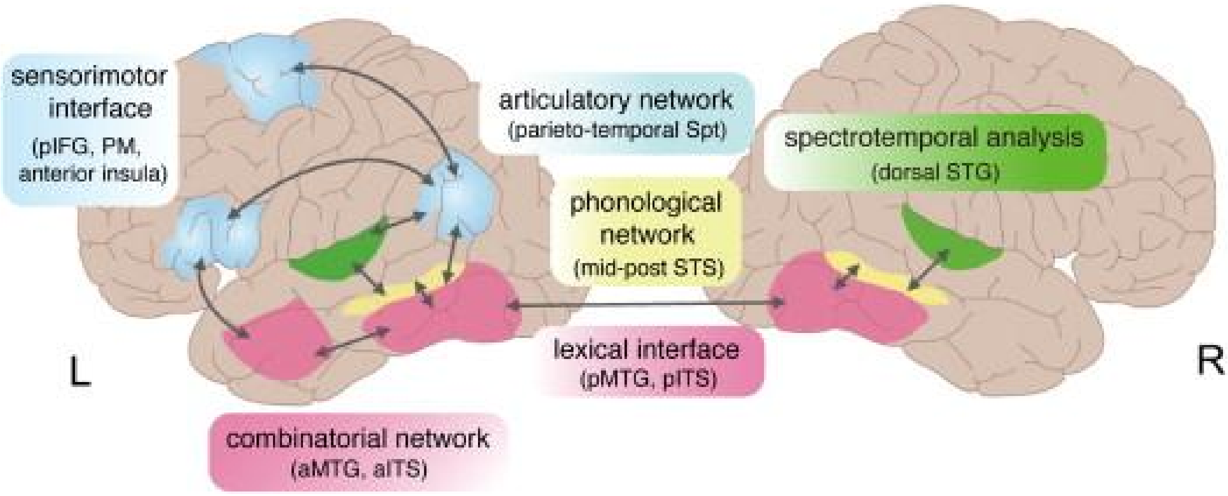


Figure 8 : (A) The dual-stream model of Hickok and Poeppel in schematic form and the neural correlates of each node (B). Blue areas represent the dorsal stream, while pink areas the ventral one. Adapted from (*91*)

5.3 Brain Anatomy of Vocal Processing in Animals

Studies on vocal processing in animals have primarily focused on non-human primates (NHP) because of their close phylogenetic relationship to humans, and on songbirds due to the intricate complexity of their vocal displays.

Anatomical studies in non-human primates have revealed conserved brain asymmetry related to conspecific vocal processing, shared between NHP and humans. Notably, leftward asymmetry of the planum temporale has been observed in chimpanzees, olive baboons, and rhesus monkeys, echoing the leftward asymmetry of Wernicke's area in humans (*72*). Similarly, research on sulcal depth in olive baboons has demonstrated a conserved leftward asymmetry in the depth of the superior temporal sulcus (STS) (*70*).

Functional brain research in NHP, particularly in rhesus monkeys, has highlighted similarities in species-specific sound processing between human and NHP brains. When exposed to conspecific vocalizations, macaque brains exhibit increased activity in the anterior superior temporal gyrus (STG) and superior temporal sulcus (STS), analogous to the human temporal voice area (*92*) . Additionally, conserved mechanisms for processing vocal sequences have been identified. For instance, several frontal cortex regions, including the ventral frontal and opercular cortex (vFOC), insula, and Brodmann areas 44/45, show heightened activity in response to sequences that violate previously learned grammatical rules in both humans and macaques (*93*).

Interestingly, these similarities are not limited to primates but extend across a diverse range of species, highlighting a potential evolutionary convergence in neural mechanisms for processing vocal communication. In mammals, mustached bats (Pteronotus parnellii) and mice (Mus musculus) demonstrate pronounced left-hemisphere dominance when responding to conspecific vocalizations (*94*, *95*), suggesting that temporal and spectral features of these calls are preferentially processed by specialized neural circuits in that hemisphere.

Similarly, in birds, species like zebra finches (Taeniopygia guttata) exhibit a striking left-hemisphere preference for vocal processing in the Field L region, in a structure analogous to Broca's area in humans (*96*). Interestingly, this left-hemisphere dominance for vocal processing extends beyond zebra finches to other bird species, such as canaries (Serinus canaria) and European starlings (Sturnus vulgaris), both of which also show a lateralized response to song processing in the left hemisphere (*97*, *98*).

5.4 Chapter Summary :

Comparative neuroanatomy of acoustic communication reveals conserved neural mechanisms across species, such as the left hemisphere’s dominance in vocal processing in both mammals and birds, suggesting an evolutionary convergence in how complex vocalizations are processed. These findings support the idea that intricate vocal communication in both animals and humans relies on specialized, lateralized brain circuits for producing, learning, and understanding sounds.

OSCILLATIONS AND BRAIN FUNCTION

6.1 Brain, Neurons, and Synapses

To understand brain oscillations, one should first talk about the brain. In animals, information within the brain is encoded through electrical and chemical signals. These signals are produced by neurons, the fundamental units of the nervous system. Even though neurons exist in various forms, they share similar basic structures. Each neuron is made of a cell body, called the soma, dendrites, spike-shaped extensions of the soma receiving information from other neurons, and an axon which transmits the electrical signals to other cells. Inside the brain, axons from neurons make contact with the dendrites of other neurons, creating a zone of information exchange called the synapse (Fig 6). A synapse is made of the end of an axon, called the pre-synaptic component, and the beginning of one or several dendrites, called the post-synaptic component.

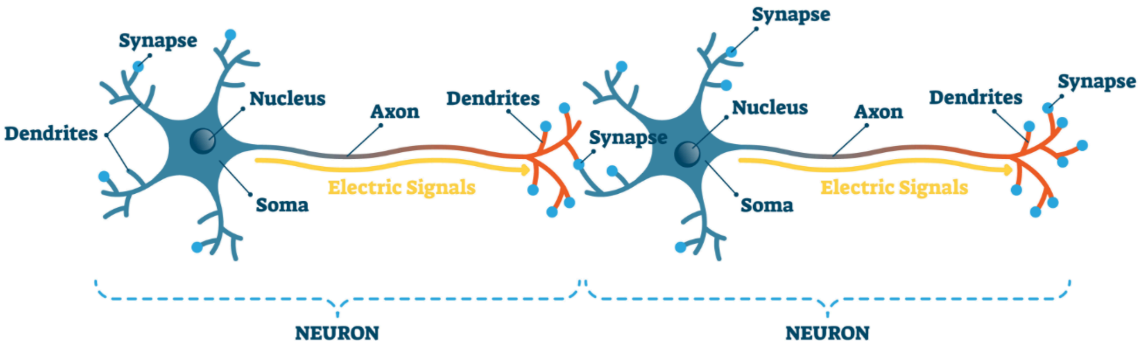


Figure 6 : Schematic representing two human neurons connecting with each others through electrical signals.

Fundamentally, neurons encode information through their spiking activity, characterized by the rate and pattern of firing. When an action potential occurs in the pre-synaptic neuron, it prompts the release of neurotransmitters into the synapse. These chemical compounds bind to specific receptors on the surface of the dendrites of the post-synaptic neuron. Depending on the type of neurotransmitter, this binding will either depolarize the post-synaptic membrane, creating an Excitatory Post-Synaptic Potential (EPSP), or hyperpolarize it, resulting in an Inhibitory Post-Synaptic Potential (IPSP). EPSPs increase neuronal excitability, whereas IPSPs decrease it. When action potentials fire rhythmically and synchronously, they induce synchronous EPSPs or IPSPs in the receiving neurons. This synchronization leads to rhythmic alternations between periods of high and low excitability in the receiving cells . When happening at a neuron population level, these alternations create electrical waves called brain oscillations (*99*). These oscillations can be measured in single neuron activity and small groups of neurons using invasive local field potential techniques, or at a more general scale, using non-invasive techniques such as scalp electroencephalography (EEG) or magnetoencephalography (MEG) (*100*).

6.2 Brain Oscillations

Brain oscillations have first been described in the human brain by Hans Berger, the inventor of the electroencephalography (EEG) technique in 1925. By placing electrodes on the scalp of a male subject, Berger was able to measure spontaneous oscillatory activity in the brain when the eyes were closed, as well as their suppression when the eyes were open. In his paper at the time, Berger concludes « Indeed, I believe that I have discovered the electroencephalogram of man and that I have published it here for the first time ». While Hans Berger first described alpha oscillations as endogenous, arising spontaneously during states of rest, Edgar Douglas Adrian later demonstrated that oscillations could also be stimulus-driven, occurring in direct response to sensory input. This distinction between intrinsic, resting-state activity and externally evoked responses not only reinforced Berger's findings but also solidified EEG as a reliable and versatile brain imaging technique.

Later studies using frequency decomposition techniques identified spontaneous and stimulus-driven oscillatory activity in the human brain, across four primary frequency ranges: delta (1-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-30 Hz), and low gamma

(25-55 Hz). As oscillation bands reflect synchronous neuronal activities, their functional roles are closely tied to their region of origin in the brain, and can represent the physical and molecular constraints that shape brain function.

Delta oscillations or slow oscillations are preeminent during the deep sleep stage. Delta oscillations are linked to basic homeostatic and motivational processes, including hunger, sexual arousal, and attention (*101*). In the motor cortex, delta oscillations interact dynamically with beta rhythms and contribute to movement planning and execution (*102*). Moreover, delta oscillations are actively involved in sensory input processing and integration (*103*).

Theta oscillations are present in a consortium of mechanisms. Preeminent in the hippocampus, they have been linked to visual memory (*104*), auditory processing (*105*) ,sustained attention (*106*), spatial coding (*107*), anxiety behavior (*108*) and decision-making (*109*). Theta oscillations in the human brain are also related to language and music processing (*110*).

Gamma oscillations are involved in higher cognitive functions, including attention and consciousness (*111*). They are crucial for the binding of sensory inputs, enabling the brain to create a cohesive perception of the environment (*112*). Gamma rhythms are also associated with the processing of complex stimuli, such as tasks requiring high levels of concentration, playing a role in memory formation, learning, and the synchronization of activity across different brain regions (*113, 114*) .

6.3 Evolution of Brain Oscillations in Animals

Even though oscillations quickly became a human-oriented subject of interest, mainly due to how easy and non-invasive it is to record them, they were first described in animal brains, and are ubiquitous across the animal realm. In mammals, oscillations have been extensively studied. Data from humans, chimpanzees, cats, dogs, mice, guinea pigs, and many others show clear conservation of rhythmic patterns across most mammals (*115*). Among the four primary bands, only theta oscillations, recorded in the hippocampus, vary between species, slowing down as brain weight increases (Fig 7).

Research on oscillations in birds are less common. Beta (13-19 Hz) and gamma (33-48 Hz) oscillations have been identified in crows' brains. Gamma oscillations (36Hz) were also induced in the hippocampal formation of domestic chicks through cholinergic activation.

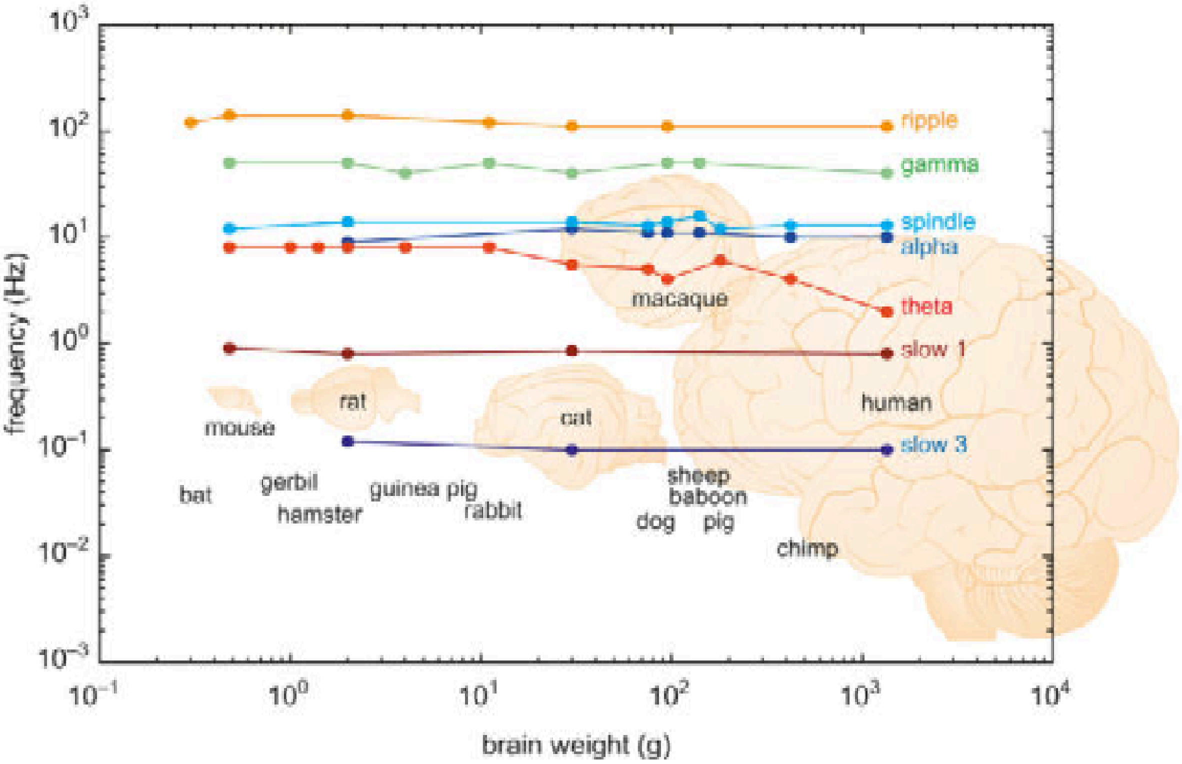


Figure 7 : Relationship between brain weight and frequency of the various rhythm classes on a log-log scale. Note the small variation of frequency changes despite increases in brain weight of several orders of magnitude. Adapted from (*115*).

In reptiles, delta oscillations (0.5-4 Hz) dominate brain activity and are primarily associated with ancient, evolutionarily conserved processes like selective attention and the detection of salient stimuli in the environment (*101*). In turtles, visual stimuli have been shown to induce oscillatory activity in the alpha (10 Hz) and beta (20 Hz) bands (*116*).

In insects, most research on oscillations has focused on Drosophila. Delta oscillations (0.5-2Hz) have been observed during sleep (*117*), stimulus-dependent gamma oscillations can be recorded from the optic lobes, and beta oscillations (20-30 Hz) in the central complex are linked to visual attention, specifically locking onto the temporal features of attended objects (*118*).

6.4 Neural Substrates of Speech Processing

While fMRI studies have highlighted the lateralization of language function, research has also emphasized the role of neural oscillations in speech decoding. This focus stemmed from the discovery of a correlation between syllabic rate (2-8 syllables/s) and the frequency of theta oscillations. Temporal variation at different timescales is a specific feature of speech that directly relates to its hierarchical structure. Indeed, as units combine together to form the higher one, phonemic information occurs in the low gamma range (25-35Hz), syllabic in the theta range (4-8Hz) and word in the delta range (1-4Hz).

There is evidence that in human listeners, theta oscillations entrain to the syllabic rhythm by resetting their phase at syllabic boundaries. As speech, which is hierarchically organized, theta oscillations also work in close relation with slow-gamma oscillations that serve to process segmental cues. Theta oscillations regulate the amplitude of slow-gamma oscillations through a mechanism of phase-amplitude coupling. While the phase of theta oscillations goes up, the amplitude of gamma oscillations goes down and vice-versa (Fig 9). This creates an alignment of neuronal excitability with speech acoustic structure, supporting speech decoding through syllables parsing and identification (*119*). Quite importantly, this processing of speech by brain oscillations has been shown to rely mostly upon rhythm, and not unit length, as intelligibility of three-time compress speech can be regained by enlarging the silence gap between the compressed syllables (*120*).

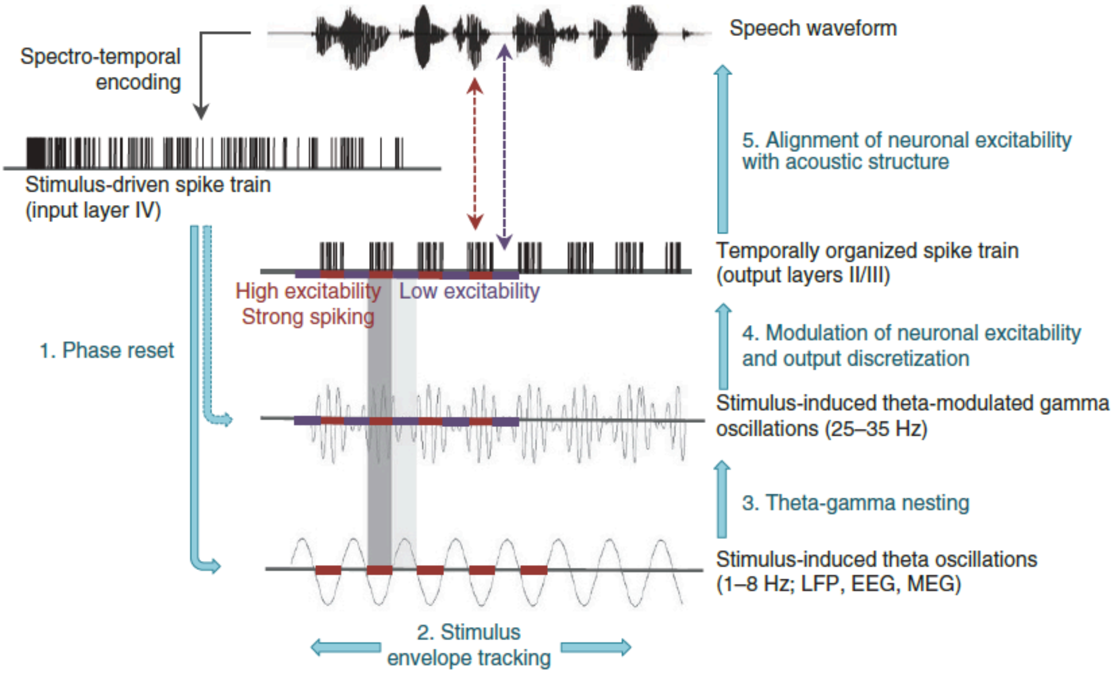


Figure 9 : Giraud and Poeppel model of speech processing by the brain through theta and gamma oscillations. The model involves five key operations: resetting theta oscillations with speech onset, tracking the speech envelope, coupling theta and gamma oscillations, using gamma power to regulate neuronal excitability, and aligning neuronal activity with speech modulations to optimize signal processing. Adapted from (*119*).

6.5 Neural Substrates of Animal Acoustic Communication Processing

The model of speech processing through brain oscillations was initially developed on the work of Peter Lakatos and Charles Schroeder, who demonstrated that oscillations in the auditory cortex of rhesus monkeys are hierarchically organized (*121*). Yet, research on rhythmic stimulus processing in non-human animal brains remains limited. Recent studies from Sonja Kotz's lab have shown that slow oscillations in the rhesus monkey brain entrain to rhythmic environmental sounds (*122*), suggesting that the fundamental mechanisms for neural entrainment were likely present in our common ancestors. Additionally, the discovery of neuronal entrainment to syllables in canary songs (*123*) implies that these mechanisms may have originated even earlier, in the common ancestor of birds and mammals.

Despite the similarities in sequence production and processing between humans and animals, there is still a significant gap in our understanding of how vocal sequences relate to brain oscillations in general and how brain oscillations respond to rhythmic sounds across different species.

6.6 Chapter summary

Brain oscillations, especially delta, theta and gamma rhythms, play a crucial role in speech processing by synchronizing neuronal activity to decode syllabic structures and acoustic patterns. While their role in animal acoustic processing remains unclear, the conservation of these oscillatory mechanisms across species, alongside evidence of their similar organisation, suggests an evolutionary continuity of acoustic processing in animal species.

THESIS OBJECTIVES

Despite the importance of rhythm in animal vocalization, and its tight link to brain oscillation, especially in human speech, the mechanism underlying rhythm perception, as well as the implication of brain oscillation in its processing in animals remain largely unknown.

To address this gap, this thesis employs a comprehensive approach. First, acoustic rhythms were quantified across a diverse range of animal species from various taxa, investigating the factors that may have influenced the evolution of rhythm and constructing models of its most plausible evolutionary scenarios. This research aims to reconstruct the origins of rhythm, provide a foundational understanding of its emergence and offer insights into the potential neural mechanisms underlying rhythm processing.

Next, to investigate auditory processing in non-human animals, we developed a sedation protocol for non-human primates. This protocol allows for EEG recordings on sedated subjects while preserving auditory processing in the brain. Using this protocol, we examined the relationship between brain oscillations and vocalization rhythm in Olive baboons (*Papio anubis*) and humans. By comparing how rhythm is processed in conspecific vocalizations between the two species, we aimed to uncover the fundamental auditory processing mechanisms that support the complex hierarchical structure of speech in the human brain, as well as enhancing our understanding of animal communication.

Finally, to investigate the prevalence and functional significance of non-linear rhythmic patterns in animal vocalizations, we analyzed the calls of nesting offspring from 33 bird species across 11 orders, focusing on the presence or absence of roughness in their vocalizations.

This work, by uncovering a conserved rhythm in animal vocalizations, points to ancient mechanisms underlying acoustic communication. By exploring the neural and evolutionary roots of rhythm processing, this thesis reveals connections between animal vocalizations and the origins of human speech. Highlighting the potential adaptive roles of nonlinear phenomena, it raises questions about how acoustic signals evolved to meet diverse environmental needs. Overall, these findings suggest a deeper, shared framework for communication that transcends species, offering new insights into the relationship between rhythm, brain function, and the evolution of language.

EXPERIMENTAL PART

CHAPTER 1

A universal rhythm in animal acoustic communication

T. Piette†, C. Cathcart†, C. Barbieri, K. M. Ming, D. Grandjean, B. Bickel, E. C. Déaux*, A-L. Giraud*

† Joint first authors

* Joint last authors

Summary Of Results

Although rhythm is recognized as a critical component of speech comprehension, little is known about its role in animal acoustic communication, as well as its origin and evolutionary history. Consequently, this chapter delves into its evolution, from its production to neural processing, and examines how these factors may have influenced its use across time and phylogeny. First, we quantified rhythm in acoustic sequences from 98 species including birds, mammals, amphibians, insects, reptiles, and fish. We investigated whether physiological and morphological factors, such as breathing rate, heart rate, biomechanical constraints, environmental influences, and social complexity, constrain communication rhythms across species. Phylogenetic modeling did not support an influence of these species-specific selective forces, instead, suggesting a phylogeny-wide evolutionary pressure towards an optimum of 2.9 Hz, in the neuronal delta range (1-4Hz).

CHAPTER 2

Light propofol anaesthesia for non-invasive auditory EEG recording in unrestrained non-human primates.

T. Piette, C. Lacaux, M. Scheltienne, V. Sterpenich, Marie (vétérinaire), Valérie (assistante véto), D. Grandjean, A. Meguerditchian, E.C Déaux, A-L.Giraud.

Summary Of Results

While studying rhythm through a bioacoustic lens has provided valuable insights into its evolution, confirming the link between brain oscillations and acoustic rhythm in animals requires recording brain signals across diverse species. Developing a simple, adaptable passive listening experiment could enable tracing the origins and evolution of rhythm processing in the brain. However, conducting EEG research in awake animals presents significant challenges. EEG is highly sensitive to motion artifacts, and the overall fragility of the recording equipment further complicates data acquisition in active and moving animals. While some studies rely on invasive intracranial, or contention protocols, these set-ups greatly limit the scope and generalizability of the experiments, and raise serious ethical questions. We, therefore, explored the use of light propofol anesthesia for auditory-related tasks in non-human animals. We show that light propofol anesthesia injected at 0.1 mg/kg/h induces light sleep in female olive baboons *(Papio anubis)* while conserving basic auditory processing, as shown by the presence of classical N100 and P200 auditory ERP components. These findings thus validate the use of light propofol anesthesia for auditory-related tasks in non-human animals.

CHAPTER 3

Hierarchical coupling of slow and low-gamma oscillations for vocal processing in humans and non-human primates.

T. Piette,D. Grandjean, A. Meguerditchian, E.C Déaux, A-L.Giraud.

Summary Of Results

Theta/gamma oscillation coupling is a hallmark of speech processing that reflects the hierarchical encoding of phoneme-level acoustic cues by the low-gamma rhythm within syllable boundaries flexibly tracked by theta oscillations. While theta and gamma oscillations are relatively conserved across species, whether their coupling is also a fixed conserved trait or a functional adaptation linked to human speech is unknown. In this study, we compared EEG responses to speech, grunt calls and matched rhythmic noises in humans and baboons. The stimuli were selected to examine the neural processing of different types of acoustic signals, ranging from those that are biologically relevant (conspecific calls) to more complex human-generated sounds (speech). The results indicate tracking of grunt via delta and theta oscillations (2-6Hz) in the baboon's brain during listening, as well as theta-low gamma phase-amplitude coupling while listening to conspecific vocal sequences. Interestingly, we also observed tracking of syllables by theta oscillations (4-8 Hz) in the baboons' brain when listening to human speech, and phase-amplitude coupling between theta and low gamma oscillations when exposed to familiar words.

DISCUSSION

THE EVOLUTION OF ACOUSTIC PROCESSING

1.1 A Universal Rhythm in Animal Acoustic Communication

Our finding of an animal-wide optimum rhythm around 2.9 Hz for acoustic communication challenges the current main idea that rhythm is shaped by biomechanical constraints, such as masticatory movements (jaws, tongue, teeth) in mammals or beak morphology in birds. Indeed, since body weight is a reliable proxy for organ size and mastication speed, the latter would predict a strong negative allometric relationship between rhythm and weight, at least in masticating species. Instead, the weak effect of weight on acoustic rhythms observed in masticatory species, and the absence of effect of beak morphology (depth, length and width) in birds both indicates that biomechanical constraints only exert a marginal influence on acoustic rhythm in animals.

This stands in clear contrast to the evolution of dominant frequency (DF). Unlike rhythm, DF is strongly influenced by the size of the sound-producing organs, showing a well-documented negative allometric relationship with body weight. Rhythm, on the other hand, exhibits a distinct and independent pattern, further highlighting its unique evolutionary trajectory.

In addition to showing that the emitter’s anatomical characteristics only have a marginal effect, our analyses also demonstrated that rhythm evolution cannot be explained by neither environmental pressures nor social complexity. Likewise, given the diversity of hearing structures and production modes, it seems unlikely that this common rhythm would be constrained by similarities in receivers’ anatomical traits. In fact, the early emergence of this conserved rhythm, as indicated by its conservation among birds, mammals and more remote species (insect, reptiles, fish), indicates a common, ancient and endogenous constraint on rhythm that predates the divergence of these lineages.

1.2 The Challenges Of Acoustic Processing

When investigating acoustic communication, it is crucial to understand the fundamental principles underlying basic acoustic processing in the animal brain. While the origin of acoustic communication remains uncertain and may have evolved independently across different evolutionary branches, the origin of hearing in vertebrates can be traced back to their common ancestor (*65*). Early vertebrates possessed a lateral line system with sensory hair cells adapted for detecting waterborne vibrations (*124*). These hair cells provided the mechanosensory foundation for the evolution of the inner ear, enabling the detection of more complex acoustic signals. This shared evolutionary origin suggests that vertebrates likely rely on conserved mechanisms for basic acoustic processing, which may have significantly influenced the development and diversification of acoustic communication systems over time.

Interestingly basic neural mechanisms are remarkably conserved, and could therefore well explain a widespread optimum rhythm. Vertebrates share a common location for motor neurons controlling vocalizations in the caudal hindbrain, where central pattern generators (CPGs) produce rhythmic outputs (*125*). Some of these CPGs regulate vital motor activities, while others are believed to control vocal production, influencing call duration and timing through motor neurons projecting to vocal muscles, supporting conserved mechanisms of acoustic rhythmic production. In terms of auditory reception, this conserved rhythm around 2.9 Hz (85% interval 1.1Hz-4.2Hz) best matches delta brain oscillations (1–4 Hz), which have been observed across species, including mammals, reptiles, and insects (*101*, *115*, *117*).

When developing the ability to hear, our common ancestor most likely faced two major challenges. First, sounds need to be detected as quickly as possible. Any delay in detection could reduce the time available to respond to a potential danger, significantly decreasing the individual's fitness. Second, it is equally important to distinguish and identify one sound from another. An individual should not respond the same way to the approach of a conspecific, a predator, or a natural danger. While discrimination, identifying that sound A is different from sound B, does not necessarily require fine-grain frequency resolution, identification, on the other hand, involves associating a sound with a specific meaning or source. This requires the resolution of detailed frequency information by the auditory system, enabling the accurate recognition of specific sound features. These two challenges, detectability and identification, most likely played a key role in shaping how we process natural sounds in our environment.

Importantly, one role of the vertebrate inner ear is to process acoustic information through frequency decomposition (*126*), sharing similar limitations to the well-known mathematical Fourier transform used in signal processing. Using a short time window would result in good temporal resolution (i.e., good detectability) but poor frequency resolution (i.e., poor discrimination). Conversely, using a long time window would result in poor temporal resolution (i.e., poor detectability) but good frequency resolution (i.e., good discrimination) (*127*), a limitation known as acoustical uncertainty principle

1.3 A Two Timescale Processing Of Sound

To resolve this trade-off, one possible strategy would be to analyse sound simultaneously at two different time-scales, ensuring both good detectability and discrimination. As such, brain oscillations offer an effective tool providing a rhythmic framework that can align with the temporal structure of natural acoustic signals (*119*).

Interestingly, delta oscillations have been shown to play a key role in structuring sniffing behaviors in mice (*128*) and eye-gaze patterns in humans (*129*), underlining their importance in active sensing. These slow oscillations have also been found to track word and prosodic-level information during speech, with their rate aligning with the conserved animal rhythm. This suggests that slow oscillations integrate sensory information and act as a primary gate for external stimuli, making them particularly effective for processing slow-varying acoustic cues, such as those essential for vocal recognition (*103*). This rhythmic mechanism likely aids in the discrimination of acoustic signals, ensuring efficient communication. In line with this, recent research indicates that despite millennia of domestication, dogs still process speech at a slow rhythm between 1 and 4 Hz (*130*), such that humans naturally adjust their speech rate in noisy environments to enhance intelligibility (*131*), reinforcing the notion of a conserved and effective slow communications rhythm across species.

On the other hand, analyzing acoustic signals at faster rhythms, if occurring simultaneously, could help maintain good detectability. Interestingly our preliminary results show tracking rate or grunt in the delta range as well as phase-amplitude coupling between slow and low gamma oscillations during the processing of conspecific vocalizations in baboons, supporting the idea of multi-time scale processing in non-human primates. Similarly, phase-amplitude coupling between slow oscillations (2-8 Hz) and low gamma oscillations (25-55 Hz) has also been observed in the auditory cortex of bats (*132*). These hint at conserved neural mechanisms for processing sound across species, with slow and fast oscillatory rhythms working together to process auditory information at different timescales. However, further research is needed to confirm these findings, and to extend them beyond the mammalian branch to better understand the evolutionary origin and significance of these mechanisms.

As we consider the brain's rhythmic processing of auditory information, it's important to recognize that these mechanisms likely originate from circuits that have been positively selected for through evolution. Given the constant exposure to auditory stimuli, the brain should rely on efficient processing mechanisms. Slower oscillations, requiring less energy to generate, strike an optimal balance between energy efficiency and the capacity to continuously process auditory input. In this context, the conserved rhythm around 2.9 Hz might reflect the default firing rate of these neurons, providing an energy-efficient range at which these neural circuits function. This would ensure that the brain can effectively screen and respond to the auditory environment while minimizing energetic costs.

1.4 The Evolution Of Speech Three Time Scale Processing

While this thesis proposes a two-time scale processing of acoustic communication in non-human animals, human speech processing involves a more complex mechanism, relying on three distinct timescales, delta, theta, and gamma rhythms, respectively corresponding to phonemic, syllabic, and prosodic rates (*119*). This raises intriguing questions about the origin and evolution of the third theta timescale in speech processing, particularly in relation to how it might have developed to support the finer temporal aspects of human language.

Interestingly, while our previous study showed that dogs track language in the delta range, our preliminary analysis of speech processing in female baboons reveals that their brains can track faster rhythms in the theta range, rather aligning with the syllabic rate of speech. Furthermore, we also observed that theta oscillations directly influence gamma oscillations during speech and conspecific vocal processing. as it would in the human brain. Additionally, research on macaques has demonstrated a three-level hierarchical organization of brain oscillations in their auditory cortex, where delta oscillations influence theta oscillations, which in turn influence low-gamma oscillations (*121*). The ability to track theta rhythms that we found in baboons, and the presence of theta low-gamma PAC, coupled with the three-tiered organization of brain oscillations observed in macaques, suggests that the intermediary theta level of acoustic processing is already latently present in non-human primates, though it may not yet play a functional role in their acoustic communication.

A plausible explanation for the development of such intermediary processing level in speech lies in the unique complexity of human social interactions. Unlike most animals, which limit their social exchanges to a relatively small number of individuals, humans engage in intricate social networks involving more than 100 individuals (*133*, *134*). As these interactions grew increasingly complex over time, the volume of information needed to be transferred between individuals likely surpassed the capacity of slower oscillatory processing. This pressure could have driven the use of an already existing faster intermediary timescale, allowing for the transmission of more detailed information. Although such faster mechanisms may be less reliable in term of frequency resolution than slower oscillations, which could explain why other primates may not rely on them, this limitation might be overcome by shifting from passive acoustic signal processing to a predictive framework. Recent studies on the role of predictive coding in human speech processing support this idea, highlighting how prediction enhances the efficiency and accuracy of processing faster temporal information in complex communication (*135–137*). This mechanism, while likely present in other animal communication systems (*138*), may have become particularly crucial for speech processing in humans. enhancing the efficiency and accuracy of processing faster temporal information in complex communication.

1.5 Low Gamma High Gamma And The Influence Of Roughness

While it is intuitive to see how delta oscillations, as the slowest available rhythm, became the primary gate for auditory processing, the role of low gamma as the faster timescale for ensuring detectability remains less clear. Faster rhythms would offer smaller time windows, enabling better temporal resolution to detect rapid acoustic changes. This raises a compelling question: why would low gamma (25-35hz), rather than higher gamma rhythms (>40Hz), be selected for communication processing?

An intriguing perspective on this emerges from studies on human screams, which hold a unique place in our vocal repertoire, being rarely used in routine communication. Research using the modulation power spectrum has shown that screams amplify amplitude modulations within a specific perceptual range, between 30-150 Hz, known as the roughness range (*50*). When present, these fast amplitude modulations are perceived as salient and aversive, creating a acoustic niche majoritarily used to communicate danger and distress (*51*).

Interestingly, our results reveal that these amplitude modulations are also present in the vocalizations of most bird offspring calls, primarily used to attract caregiver attention, and have been conserved since the common ancestor of all birds. Additionally, studies have demonstrated the presence of roughness in the vocal signals of animals with more limited vocal repertoires, such as frogs, crickets, and drumfish .

This suggests that roughness may represent a conserved mechanism for capturing attention in acoustic communication, inherited from ancient auditory processing strategies. Notably, rough acoustic characteristics are also found in sounds emitted by natural dangers, such as thunder or crackling fire. During the evolution of hearing, natural selection may have favored individuals capable of quickly identifying environmental dangers based on these auditory cues. As acoustic communication evolved, this roughness niche may have been conserved as an effective strategy to grab attention and signal immediate danger. With the increasing complexity of communication systems, low gamma oscillations (25-35 Hz) were likely selected as the fastest remaining rhythm for processing other forms of communication. This frequency range, crucially, avoids overlap with the roughness spectrum. By operating at this lower gamma range, auditory processing ensures sufficient temporal resolution for detecting non-threatening sounds, while preserving the roughness characteristic for high-arousal, danger-related cues.

LIMITATIONS

While offering valuable insights into basic acoustic processing mechanisms and proposing a comprehensive hypothesis on the evolution of acoustic communication through brain oscillations, driven by novel comparative studies, this thesis, like any, is not without its flaws.

First, constructing a comparative dataset of animal vocal sequences required selecting recordings from various public and private databases, as well as seeking collaboration from researchers worldwide to share their recordings. In doing so, we often faced challenges accessing comprehensive metadata, particularly regarding call types and production contexts, which are known to impact rhythm in animal vocal communication (47). The same applies to our analysis of roughness in bird offspring, for which we primarily gathered data from the public Xenocanto database.

Although our EEG experiment with female baboons yielded promising results, the small sample size limited the ability to generalize these findings. A larger cohort would be crucial to enhance their generalisability and reproducibility. Furthermore, while these preliminary results offer valuable insights into the evolution of acoustic communication in primates, they are not applicable beyond this group, and research on other species remains sparse. Expanding these experiments to include more distantly related species is therefore vital and could significantly deepen our understanding of fundamental acoustic processing mechanisms, particularly in relation to the evolution of the three-time scale processing of speech in the human brain.

Furthermore, while our anesthetic protocol appears to preserve basic acoustic processing in non-human primates, we cannot completely rule out an effect of anesthesia on higher cognitive processing of auditory signals, potentially limiting our comprehension of complex auditory signal processing by the brain.

Another limitation arises from EEG itself. While offering impressive temporal resolution, ideal for studying the brain's processing of rapidly changing rhythmic information, it suffers from a poor spatial resolution. Although this limitation can be addressed in human experiments using standardized caps with an extensive number of electrodes, up to over 200, this approach cannot be applied to most animal species, limiting the potential for source localization.

Finally, while this thesis advocates for an ancient two-timescale processing of sound by the animal brain, it does not present direct proof of a conserved fast processing rhythm for acoustic signals across species. While we observe compelling evidence of a two-timescale mechanism in the acoustic processing of various species, the data primarily supports the hypothesis rather than confirming it through direct experimental proof. Similarly, although our findings on the coupling between delta and low gamma oscillations in the processing of conspecific vocalizations in baboons suggest the existence of such a mechanism, the lack of direct proof leaves room for alternative interpretations.

One such alternative interpretation is the possibility that what we have considered distinct timescales of auditory processing might actually represent a single, slower timescale combining both delta and theta oscillations. This could also explain why most animals do not engage the higher frequencies typically associated with more complex auditory processing seen in humans. A shift to these higher frequencies could come at the cost of efficiency, with faster mechanisms potentially being more prone to errors or requiring greater energy expenditure. Therefore, further investigation into the interplay between delta and theta oscillations across species is crucial for evaluating these alternative hypotheses and understanding the evolutionary trade-offs involved in auditory processing.

PERSPECTIVES

While this thesis offers precious advances in the comparative study of acoustic rhythm processing by the animal brain, further experiments are still needed to fully understand the mechanisms of acoustic processing, and how they evolved through time and species. First, to further support the claim of a preferred rhythm in acoustic production and processing, an adapted auditory steady-state response experiment could be conducted, incorporating rhythms both within and outside the conserved communication rhythm. If performed across various species from different classes, such an experiment could help confirm the existence of a preferred acoustic rhythm in the brain conserved throughout evolution.

Similarly, directly investigating the existence of a conserved fast processing rate for acoustic signals would be crucial to confirm the presence of a two-timescale processing strategy across species. To achieve this, an experiment could be designed to examine brain activity in response to acoustic stimuli at varying time scales, including both slow and fast rhythms. Using techniques such as EEG or magnetoencephalography (MEG), researchers could analyze phase-amplitude coupling or oscillatory entrainment across different species, including those with limited vocal repertoires, to assess whether the fast processing rate observed in humans is similarly conserved in animals.

Finally, future studies using EEG with non-human animals (NHAs) should focus on developing new protocols for investigating brain mechanisms in awake animals. Although minimal propofol was used to control the depth of anesthesia, we cannot entirely rule out the potential influence of anesthesia on higher cognitive functions. This limitation highlights the need for more advanced techniques to study brain activity in awake animals, particularly when examining the rapid brain rhythms involved in acoustic processing. Additionally, recent advancements in magnetoencephalography (MEG) techniques, particularly the development of portable MEG systems, may one day overcome the spatial resolution limitations of EEG. These innovations could provide a more accurate understanding of the brain's acoustic processing mechanisms, particularly those underlying acoustic communication.

CONCLUSIONS

Through extensive cross-species comparative work, this thesis offers a novel perspective on the evolution of acoustic communication, notably identifying a conserved slow rhythm that transcends biomechanical constraints, environmental, and social pressures. By highlighting the likely constraint of ancient and conserved neural mechanisms, it presents a strong case for a two-timescale processing framework in acoustic communication, based on the simultaneous processing of sound by delta and low gamma oscillations. This dual system, balancing detection and discrimination, appears to be a foundational feature of vertebrate auditory systems, shaping the evolution of acoustic communication across taxa. Furthermore, this framework sets the stage for the appearance of a third, intermediate theta timescale, within the primate lineage, where it likely plays a critical role in processing complex acoustic signals, such as speech. While limitations in data availability, sample size, and the use of anesthesia necessitate further investigation, these findings open promising avenues for comparative research into the neural and evolutionary bases of sound processing, putting forward non-invasive brain recording methods, and hopefully promoting an evolutionary approach to the investigation of rhythm processing.

BIBLIOGRAPHY

---

1. M. D. Hauser, *The Evolution of Communication* (MIT Press, Cambridge, Mass., 1. paperback ed., 4. print., 2000)*A Bradford book*.
2. J. R. Krebs, N. B. Davies, Eds., *Behavioural Ecology: An Evolutionary Approach* (Blackwell Science, Oxford, 3. ed., reprinted., 1995).
3. R. M. Seyfarth, D. L. Cheney, T. Bergman, J. Fischer, K. Zuberbühler, K. Hammerschmidt, The central importance of information in studies of animal communication. *Animal Behaviour* **80**, 3–8 (2010).
4. R. M. Seyfarth, D. L. Cheney, Meaning and Emotion in Animal Vocalizations. *Annals of the New York Academy of Sciences* **1000**, 32–55 (2006).
5. D. Otte, Effects and Functions in the Evolution of Signaling Systems. *Annu. Rev. Ecol. Syst.* **5**, 385–417 (1974).
6. F. Cottier, F. A. Mühlschlegel, Communication in Fungi. *International Journal of Microbiology* **2012**, 1–9 (2012).
7. R. Karban, Plant Communication. *Annu. Rev. Ecol. Evol. Syst.* **52**, 1–24 (2021).
8. K. Yoneya, J. Takabayashi, Plant–plant communication mediated by airborne signals: ecological and plant physiological perspectives. *Plant Biotechnology* **31**, 409–416 (2014).
9. G. Carter, L. Leffer, Social Grooming in Bats: Are Vampire Bats Exceptional? *PLoS ONE* **10**, e0138430 (2015).
10. C. D. Saunders, G. Hausfater, The Functional Significance of Baboon Grooming Behaviora. *Annals of the New York Academy of Sciences* **525**, 430–432 (1988).
11. L. Lefebvre, The organization of grooming in budgerigars. *Behavioural Processes* **7**, 93–106 (1982).
12. E. O. Wilson, Chemical communication among workers of the fire ant Solenopsis saevissima (Fr. Smith) 1. The Organization of Mass-Foraging. *Animal Behaviour* **10**, 134–147 (1962).
13. T. D. Wyatt, *Pheromones and Animal Behaviour: Communication by Smell and Taste* (Cambridge University Press, ed. 1, 2003; https://www.cambridge.org/core/product/identifier/9780511615061/type/book).
14. M. Brecht, W. A. Freiwald, The many facets of facial interactions in mammals. *Current Opinion in Neurobiology* **22**, 259–266 (2012).
15. E. H. Meseth, The Dance of the Laysan Albatross, Diomedea Immutabilis. *Behav* **54**, 217–257 (1975).
16. R. Zelick, D. A. Mann, A. N. Popper, "Acoustic Communication in Fishes and Frogs" in *Comparative Hearing: Fish and Amphibians*, R. R. Fay, A. N. Popper, Eds. (Springer New York, New York, NY, 1999; http://link.springer.com/10.1007/978-1-4612-0533-3_9)vol. 11 of *Springer Handbook of Auditory Research*, pp. 363–411.
17. S. K. Watson, S. W. Townsend, F. Range, Wolf howls encode both sender- and context-specific information. *Animal Behaviour* **145**, 59–66 (2018).
18. C. D. Hopkins, The neuroethology of electric communication. *Trends in Neurosciences* **4**, 4–6 (1981).
19. F. Ladich, Fish bioacoustics. *Current Opinion in Neurobiology* **28**, 121–127 (2014).
20. K. C. Colafrancesco, M. Gridi-Papp, "Vocal Sound Production and Acoustic Communication in Amphibians and Reptiles" in *Vertebrate Sound Production and Acoustic Communication*, R. A. Suthers, W. T. Fitch, R. R. Fay, A. N. Popper, Eds. (Springer International Publishing, Cham, 2016; http://link.springer.com/10.1007/978-3-319-27721-9_3)vol. 53 of *Springer Handbook of Auditory Research*, pp. 51–82.

21. M. Naguib, Ed., *Advances in the Study of Behavior* (Elsevier/AP, Amsterdam, 2009)*Advances in the study of behavior*.
22. T. Fitch, "Production of Vocalizations in Mammals" in *Encyclopedia of Language & Linguistics* (Elsevier, 2006; https://linkinghub.elsevier.com/retrieve/pii/B008044854200821X), pp. 115–121.
23. A. S. Gaunt, M. K. Wells, Models of Syringeal Mechanisms. *Am Zool* **13**, 1227–1247 (1973).
24. A. P. Russell, A. M. Bauer, Vocalization by extant nonavian reptiles: A synthetic overview of phonation and the vocal apparatus. *The Anatomical Record* **304**, 1478–1528 (2021).
25. D. A. Mann, Sound production by fishes. *The Journal of the Acoustical Society of America* **104**, 1826–1826 (1998).
26. A. M. Taylor, B. D. Charlton, D. Reby, "Vocal Production by Terrestrial Mammals: Source, Filter, and Function" in *Vertebrate Sound Production and Acoustic Communication*, R. A. Suthers, W. T. Fitch, R. R. Fay, A. N. Popper, Eds. (Springer International Publishing, Cham, 2016; http://link.springer.com/10.1007/978-3-319-27721-9_8)vol. 53 of *Springer Handbook of Auditory Research*, pp. 229–259.
27. A. M. Taylor, D. Reby, The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology* **280**, 221–236 (2010).
28. J. N. Holmes, W. J. Holmes, P. N. Garner, "Using formant frequencies in speech recognition" in *5th European Conference on Speech Communication and Technology (Eurospeech 1997)* (ISCA, 1997; https://www.isca-archive.org/eurospeech_1997/holmes97_eurospeech.html), pp. 2083–2086.
29. S. W. Townsend, B. D. Charlton, M. B. Manser, Acoustic cues to identity and predator context in meerkat barks. *Animal Behaviour* **94**, 143–149 (2014).
30. S. A. Reber, J. Janisch, K. Torregrosa, J. Darlington, K. A. Vliet, W. T. Fitch, Formants provide honest acoustic cues to body size in American alligators. *Sci Rep* **7**, 1816 (2017).
31. D. Reby, K. McComb, Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour* **65**, 519–530 (2003).
32. T. Riede, T. Fitch, Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). *Journal of Experimental Biology* **202**, 2859–2867 (1999).
33. W. T. Fitch, Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America* **102**, 1213–1222 (1997).
34. M. I. Muñoz, M. Marsot, J. Ellers, W. Halfwerk, Tetrapod vocal evolution: higher frequencies and faster rates of evolution in mammalian vocalizations. [Preprint] (2023). https://doi.org/10.1101/2023.08.09.552622.
35. D. L. Bowling, M. Garcia, J. C. Dunn, R. Ruprecht, A. Stewart, K.-H. Frommolt, W. T. Fitch, Body size and vocalization in primates and carnivores. *Sci Rep* **7**, 41070 (2017).
36. J. G. M. Robertson, Male territoriality, fighting and assessment of fighting ability in the Australian frog Uperoleia rugosa. *Animal Behaviour* **34**, 763–772 (1986).
37. L. A. Hardouin, D. Reby, C. Bavoux, G. Burneleau, V. Bretagnolle, Communication of Male Quality in Owl Hoots. *The American Naturalist* **169**, 552–562 (2007).
38. P. Galeotti, N. Saino, R. Sacchi, A. P. Møller, Song correlates with social context, testosterone and body condition in male barn swallows. *Animal Behaviour* **53**, 687–700 (1997).
39. D. Reby, B. D. Charlton, Y. Locatelli, K. McComb, Oestrous red deer hinds prefer male roars with higher fundamental frequencies. *Proc. R. Soc. B.* **277**, 2747–2753 (2010).
40. L. A. Hardouin, V. Bretagnolle, P. Tabel, C. Bavoux, G. Burneleau, D. Reby, Acoustic cues to reproductive success in male owl hoots. *Animal Behaviour* **78**, 907–913 (2009).
41. E. F. Briefer, Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology* **288**, 1–20 (2012).
42. C. Coupé, Y. M. Oh, D. Dediu, F. Pellegrino, Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Sci. Adv.* **5**, eaaw2594 (2019).
43. S. Fujii, The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Frontiers*

in Human Neuroscience, 15.

44. M. D. Hauser, B. Agnetta, C. Perez, Orienting asymmetries in rhesus monkeys: the effect of time-domain changes on acoustic perception. *Animal Behaviour* **56**, 41–47 (1998).
45. T. I. Drăgănoiu, L. Nagle, M. Kreutzer, Directional female preference for an exaggerated male trait in canary (*Serinusanaria*) song. *Proc. R. Soc. Lond. B* **269**, 2525–2531 (2002).
46. K. Riebel, “Chapter 6 Song and Female Mate Choice in Zebra Finches: A Review” in *Advances in the Study of Behavior* (Elsevier, 2009; https://linkinghub.elsevier.com/retrieve/pii/S0065345409400068)vol. 40, pp. 197–238.
47. E. F. Briefer, Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool* **288**, 1–20 (2012).
48. M. B. Manser, Response of foraging group members to sentinel calls in suricates, Suricata suricatta. *Proc. R. Soc. Lond. B* **266**, 1013–1019 (1999).
49. D. T. Blumstein, K. B. Armitage, Alarm calling in yellow-bellied marmots: I. The meaning of situationally variable alarm calls. *Animal Behaviour* **53**, 143–171 (1997).
50. L. H. Arnal, A. Flinker, A. Kleinschmidt, A.-L. Giraud, D. Poeppel, Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology* **25**, 2051–2056 (2015).
51. L. H. Arnal, A. Kleinschmidt, L. Spinelli, A.-L. Giraud, P. Mégevand, The rough sound of salience enhances aversion through neural synchronisation. *Nat Commun* **10**, 3671 (2019).
52. R. S. Payne, S. McVay, Songs of Humpback Whales: Humpbacks emit sounds in long, predictable patterns ranging over frequencies audible to humans. *Science* **173**, 585–597 (1971).
53. J. Cynx, Experimental determination of a unit of song production in the zebra finch (Taeniopygia guttata). *Journal of Comparative Psychology* **104**, 3–10 (1990).
54. K. Ouattara, A. Lemasson, K. Zuberbühler, Campbell’s Monkeys Use Affixation to Alter Call Meaning. *PLoS ONE* **4**, e7808 (2009).
55. S. Engesser, A. R. Ridley, S. W. Townsend, Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5976–5981 (2016).
56. C. V. Linné, J. F. Gmelin, G. Emanuel. Beer, *Systema Naturae per Regna Tria Naturae : Secundum Classes, Ordines, Genera, Species, Cum Characteribus, Differentiis, Synonymis, Locis* (Impensis Georg. Emanuel. Beer, Lipsiae [Leipzig], 1788; https://www.biodiversitylibrary.org/bibliography/545).
57. P. Corsi, *Lamarck: Genèse et enjeux du transformisme (1770-1830)* (CNRS Éditions, 2001; http://books.openedition.org/editionscnrs/35613).
58. C. Darwin, J. Murray, William Clowes and Sons, Bradbury & Evans., *On the Origin of Species by Means of Natural Selection, or, The Preservation of Favoured Races in the Struggle for Life* (John Murray, Albemarle Street, London, 1859; https://www.biodiversitylibrary.org/bibliography/82303).
59. J. D. Watson, F. H. C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
60. S. T. Kalinowski, M. J. Leonard, T. M. Andrews, Nothing in Evolution Makes Sense Except in the Light of DNA. *LSE* **9**, 87–97 (2010).
61. C. E. Hinchliff, S. A. Smith, J. F. Allman, J. G. Burleigh, R. Chaudhary, L. M. Coghill, K. A. Crandall, J. Deng, B. T. Drew, R. Gazis, K. Gude, D. S. Hibbett, L. A. Katz, H. D. Laughinghouse, E. J. McTavish, P. E. Midford, C. L. Owen, R. H. Ree, J. A. Rees, D. E. Soltis, T. Williams, K. A. Cranston, Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12764–12769 (2015).
62. J. Parker, G. Tsagkogeorga, J. A. Cotton, Y. Liu, P. Provero, E. Stupka, S. J. Rossiter, Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228–231 (2013).
63. M. Grabowski, W. L. Jungers, Evidence of a chimpanzee-sized ancestor of humans but a gibbon-sized ancestor of apes. *Nat Commun* **8**, 880 (2017).
64. G. J. Slater, Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4897–4902 (2015).
65. G. Jorgewich-Cohen, S. W. Townsend, L. R. Padovese, N. Klein, P. Praschag, C. R. Ferrara, S.

Ettmar, S. Menezes, A. P. Varani, J. Serano, M. R. Sánchez-Villagra, Common evolutionary origin of acoustic communication in choanate vertebrates. *Nat Commun* **13**, 6089 (2022).

66. N. Chomsky, *Language and Mind* (Cambridge University Press, Cambridge, UK New York, 3rd ed., 2006).
67. W. T. Fitch, *The Evolution of Language* (Cambridge University Press, Cambridge, 2010).
68. C. Reichmuth, C. Casey, Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology* **28**, 66–71 (2014).
69. L.-J. Boë, F. Berthommier, T. Legou, G. Captier, C. Kemp, T. R. Sawallis, Y. Becker, A. Rey, J. Fagot, Evidence of a Vocalic Proto-System in the Baboon (Papio papio) Suggests Pre-Hominin Speech Precursors. *PLOS ONE*, 15 (2017).
70. W. D. Hopkins, Neuroanatomical asymmetries in nonhuman primates in the homologs to Broca's and Wernicke's areas: a mini-review. *Emerging Topics in Life Sciences* **6**, 271–284 (2022).
71. Y. Becker, J. Sein, L. Velly, L. Giacomino, L. Renaud, R. Lacoste, J.-L. Anton, B. Nazarian, C. Berne, A. Meguerditchian, Early Left-Planum Temporale Asymmetry in newborn monkeys (Papio anubis): A longitudinal structural MRI study at two stages of development. *NeuroImage* **227**, 117575 (2021).
72. D. Marie, M. Roth, R. Lacoste, B. Nazarian, A. Bertello, J.-L. Anton, W. D. Hopkins, K. Margiotoudi, S. A. Love, A. Meguerditchian, Left Brain Asymmetry of the Planum Temporale in a Nonhominid Primate: Redefining the Origin of Brain Specialization for Language. *Cerebral Cortex* **28**, 1808–1815 (2018).
73. D. A. Liao, K. F. Brecht, L. Veit, A. Nieder, Crows "count" the number of self-generated vocalizations. *Science* **384**, 874–877 (2024).
74. R. M. Seyfarth, D. L. Cheney, P. Marler, Monkey Responses to Three Different Alarm Calls: Evidence of Predator Classification and Semantic Communication. *Science* **210**, 801–803 (1980).
75. E. C. Garland, A. W. Goldizen, M. L. Rekdahl, R. Constantine, C. Garrigue, N. D. Hauser, M. M. Poole, J. Robbins, M. J. Noad, Dynamic Horizontal Cultural Transmission of Humpback Whale Song at the Ocean Basin Scale. *Current Biology* **21**, 687–691 (2011).
76. M. Fukushima, A. M. Doyle, M. P. Mullarkey, M. Mishkin, B. B. Averbeck, Distributed acoustic cues for caller identity in macaque vocalization. *R. Soc. open sci.* **2**, 150432 (2015).
77. D. T. Blumstein, K. B. Armitage, Alarm calling in yellow-bellied marmots: I. The meaning of situationally variable alarm calls. *Animal Behaviour* **53**, 143–171 (1997).
78. D. B. Webster, A. N. Popper, R. R. Fay, Eds., *The Mammalian Auditory Pathway: Neuroanatomy* (Springer New York, New York, NY, 1992; http://link.springer.com/10.1007/978-1-4612-4416-5)vol. 1 of *Springer Handbook of Auditory Research*.
79. R. L. Boord, THE ANATOMY OF THE AVIAN AUDITORY SYSTEM. *Annals of the New York Academy of Sciences* **167**, 186–198 (1969).
80. C. Gerhardt, Auditory selectivity for the acoustic properties of conspecific mate-attracting signals in lower vertebrates and songbirds. *OAAP*, 73 (2015).
81. C. E. Carr, R. A. Code, "The Central Auditory System of Reptiles and Birds" in *Comparative Hearing: Birds and Reptiles*, R. J. Dooling, R. R. Fay, A. N. Popper, Eds. (Springer New York, New York, NY, 2000; http://link.springer.com/10.1007/978-1-4612-1182-2_5)vol. 13 of *Springer Handbook of Auditory Research*, pp. 197–248.
82. B. E. Butler, S. G. Lomber, Functional and structural changes throughout the auditory system following congenital and early-onset deafness: implications for hearing restoration. *Front. Syst. Neurosci.* **7** (2013).
83. M. C. Göpfert, R. M. Hennig, Hearing in Insects. *Annu. Rev. Entomol.* **61**, 257–276 (2016).
84. M. P. Broca, REMARQUES SUR LE SIÉGE DE LA FACULTÉ DU LANGAGE ARTICULÉ, SUIVIES D'UNE OBSERVATION D'APHÉMIE (PERTE DE LA PAROLE).
85. C. Wernicke, "Der aphasische Symptomenkomplex" in *Der aphasische Symptomenkomplex* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1974; http://link.springer.com/10.1007/978-3-642-65950-8_1), pp. 1–70.

86. N. Geschwind, W. Levitsky, Human Brain: Left-Right Asymmetries in Temporal Speech Region. *Science* **161**, 186–187 (1968).
87. K. Amunts, G. Schlaug, A. Schleicher, H. Steinmetz, A. Dabringhaus, P. E. Roland, K. Zilles, Asymmetry in the Human Motor Cortex and Handedness. *NeuroImage* **4**, 216–222 (1996).
88. D. A. Robin, D. Tranel, H. Damasio, Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language* **39**, 539–555 (1990).
89. P. Belin, M. Zilbovicius, S. Crozier, L. Thivard, A. A. Fontaine, M.-C. Masure, Y. Samson, Lateralization of Speech and Auditory Temporal Processing. *Journal of Cognitive Neuroscience* **10**, 536–540 (1998).
90. G. Hickok, D. Poeppel, The cortical organization of speech processing. *Nat Rev Neurosci* **8**, 393–402 (2007).
91. Y. Sahara, T. Kobayashi, H. Toya, T. Suzuki, Applications of functional magnetic resonance imaging for analysis of oral functions. *Journal of Oral Biosciences* **54**, 101–106 (2012).
92. C. I. Petkov, C. Kayser, T. Steudel, K. Whittingstall, M. Augath, N. K. Logothetis, A voice region in the monkey brain. *Nat Neurosci* **11**, 367–374 (2008).
93. B. Wilson, Y. Kikuchi, L. Sun, D. Hunter, F. Dick, K. Smith, A. Thiele, T. D. Griffiths, W. D. Marslen-Wilson, C. I. Petkov, Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat Commun* **6**, 8901 (2015).
94. G. Ehret, Left hemisphere advantage in the mouse brain for recognizing ultrasonic communication calls. *Nature* **325**, 249–251 (1987).
95. J. S. Kanwal, Right–left asymmetry in the cortical processing of sounds for social communication vs. navigation in mustached bats. *Eur J of Neuroscience* **35**, 257–270 (2012).
96. S. Moorman, S. M. H. Gobes, M. Kuijpers, A. Kerkhofs, M. A. Zandbergen, J. J. Bolhuis, Human-like brain hemispheric dominance in birdsong learning. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12782–12787 (2012).
97. I. George, “Hemispheric Asymmetry of Songbirds” in *The Two Halves of the Brain*, K. Hugdahl, R. Westerhausen, Eds. (The MIT Press, 2010; https://academic.oup.com/mit-press-scholarship-online/book/19046/chapter/177441402), pp. 91–120.
98. I. George, H. Cousillas, J.-P. Richard, M. Hausberger, Song perception in the European starling: hemispheric specialisation and individual variations. *Comptes Rendus. Biologies* **325**, 197–204 (2002).
99. G. Buzsáki, C. A. Anastassiou, C. Koch, The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* **13**, 407–420 (2012).
100. M. X. Cohen, Where Does EEG Come From and What Does It Mean? *Trends in Neurosciences* **40**, 208–218 (2017).
101. G. G. Knyazev, EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience & Biobehavioral Reviews* **36**, 677–695 (2012).
102. A. Hamel-Thibault, F. Thénault, K. Whittingstall, P.-M. Bernier, Delta-Band Oscillations in Motor Regions Predict Hand Selection for Reaching. *Cereb. Cortex*, cercor;bhw392v1 (2016).
103. B. Morillon, L. H. Arnal, C. E. Schroeder, A. Keitel, Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neurosci Biobehav Rev* **107**, 136–142 (2019).
104. H. Lee, G. V. Simpson, N. K. Logothetis, G. Rainer, Phase Locking of Single Neuron Activity to Theta Oscillations during Working Memory in Monkey Extrastriate Visual Cortex. *Neuron* **45**, 147–156 (2005).
105. V. Kolev, O. A. Rosso, J. Yordanova, A transient dominance of theta ERP component characterizes passive auditory processing: evidence from a developmental study: *Neuroreport* **12**, 2791–2796 (2001).
106. W. Xie, B. M. Mallin, J. E. Richards, Development of infant sustained attention and its relation to EEG oscillations: an EEG and cortical source analysis study. *Developmental Science* **21**, e12562 (2018).

107. M. E. Hasselmo, C. E. Stern, Theta rhythm and the encoding and retrieval of space and time. *NeuroImage* **85**, 656–666 (2014).
108. T. Korotkova, A. Ponomarenko, C. K. Monaghan, S. L. Poulter, F. Cacucci, T. Wills, M. E. Hasselmo, C. Lever, Reconciling the different faces of hippocampal theta: The role of theta oscillations in cognitive, emotional and innate behaviors. *Neuroscience & Biobehavioral Reviews* **85**, 65–80 (2018).
109. J. Jacobs, G. Hwang, T. Curran, M. J. Kahana, EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage* **32**, 978–987 (2006).
110. L. Meyer, The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur J of Neuroscience* **48**, 2609–2621 (2018).
111. A. Guan, S. Wang, A. Huang, C. Qiu, Y. Li, X. Li, J. Wang, Q. Wang, B. Deng, The role of gamma oscillations in central nervous system diseases: Mechanism and treatment. *Front. Cell. Neurosci.* **16**, 962957 (2022).
112. C. Tallon-Baudry, Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences* **3**, 151–162 (1999).
113. M. J. Jutras, P. Fries, E. A. Buffalo, Gamma-Band Synchronization in the Macaque Hippocampus and Memory Formation. *J. Neurosci.* **29**, 12521–12531 (2009).
114. U. Friese, J. Daume, F. Göschl, P. König, P. Wang, A. K. Engel, Oscillatory brain activity during multisensory attention reflects activation, disinhibition, and cognitive control. *Sci Rep* **6**, 32775 (2016).
115. G. Buzsáki, N. Logothetis, W. Singer, Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms. *Neuron* **80**, 751–764 (2013).
116. M. S. Hoseini, J. Pobst, N. Wright, W. Clawson, W. Shew, R. Wessel, Induced cortical oscillations in turtle cortex are coherent at the mesoscale of population activity, but not at the microscale of the membrane potential of neurons. *J Neurophysiol* **118**, 2579–2591 (2017).
117. D. Raccuglia, S. Huang, A. Ender, M.-M. Heim, D. Laber, R. Suárez-Grimalt, A. Liotta, S. J. Sigrist, J. R. P. Geiger, D. Owald, Network-Specific Synchronization of Electrical Slow-Wave Oscillations Regulates Sleep Drive in Drosophila. *Current Biology* **29**, 3611-3621.e3 (2019).
118. M. J. Grabowska, R. Jeans, J. Steeves, B. van Swinderen, Oscillations in the central brain of Drosophila are phase locked to attended visual features. *Proc Natl Acad Sci U S A* **117**, 29925–29936 (2020).
119. A.-L. Giraud, D. Poeppel, Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**, 511–517 (2012).
120. O. Ghitza, S. Greenberg, On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica* **66**, 113–126 (2009).
121. P. Lakatos, A. S. Shah, K. H. Knuth, I. Ulbert, G. Karmos, C. E. Schroeder, An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology* **94**, 1904–1911 (2005).
122. A. Criscuolo, M. Schwartze, L. Prado, Y. Ayala, H. Merchant, S. A. Kotz, Macaque monkeys and humans sample temporal regularities in the acoustic environment. *Progress in Neurobiology* **229**, 102502 (2023).
123. S. Boari, G. B. Mindlin, A. Amador, Neural oscillations are locked to birdsong rhythms in canaries. *Eur J of Neuroscience* **55**, 549–565 (2022).
124. C. V. H. Baker, M. S. Modrell, J. A. Gillis, The evolution and development of vertebrate lateral line electroreceptors. *J Exp Biol* **216**, 2515–2522 (2013).
125. A. H. Bass, B. P. Chagnaud, Shared developmental and evolutionary origins for neural basis of vocal–acoustic and pectoral–gestural signaling. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10677–10684 (2012).
126. E. Thoret, S. Ystad, R. Kronland-Martinet, Hearing as adaptive cascaded envelope interpolation. *Commun Biol* **6**, 671 (2023).
127. K. Gröchenig, “Time-Frequency Analysis and the Uncertainty Principle” in *Foundations of*

Time-Frequency Analysis (Birkhäuser Boston, Boston, MA, 2001; http://link.springer.com/10.1007/978-1-4612-0003-1_3)*Applied and Numerical Harmonic Analysis*, pp. 21–36.

128. J. Ito, S. Roy, Y. Liu, Y. Cao, M. Fletcher, L. Lu, J. D. Boughter, S. Grün, D. H. Heck, Whisker barrel cortex delta oscillations and gamma power in the awake mouse are linked to respiration. *Nat Commun* **5**, 3572 (2014).
129. H. Gunasekaran, L. Azizi, V. Van Wassenhove, S. K. Herbst, Characterizing endogenous delta oscillations in human MEG. *Sci Rep* **13**, 11031 (2023).
130. E. C. Déaux, T. Piette, F. Gaunet, T. Legou, L. Arnal, A.-L. Giraud, Dog–human vocal interactions match dogs' sensory-motor tuning. *PLoS Biol* **22**, e3002789 (2024).
131. T. D. Hanley, M. D. Steer, Effect of Level of Distracting Noise upon Speaking Rate, Duration and Intensity. *J Speech Hear Disord* **14**, 363–368 (1949).
132. F. García-Rosales, L. López-Jury, E. González-Palomares, Y. Cabral-Calderín, M. Kössl, J. C. Hechavarria, Phase-amplitude coupling profiles differ in frontal and auditory cortices of bats. *Eur J of Neuroscience* **55**, 3483–3501 (2022).
133. R. I. M. Dunbar, *Grooming, Gossip, and the Evolution of Language* (Harvard University Press, Cambridge, Mass, 1996).
134. M. Tomasello, The ultra-social animal. *Euro J Social Psych* **44**, 187–194 (2014).
135. A. G. Lewis, J.-M. Schoffelen, H. Schriefers, M. Bastiaansen, A Predictive Coding Perspective on Beta Oscillations during Sentence-Level Language Comprehension. *Front. Hum. Neurosci.* **10** (2016).
136. H. Park, R. A. A. Ince, P. G. Schyns, G. Thut, J. Gross, Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Current Biology* **25**, 1649–1653 (2015).
137. S. Hovsepyan, I. Olasagasti, A.-L. Giraud, Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nat Commun* **11**, 3117 (2020).
138. Z. C. Chao, K. Takaura, L. Wang, N. Fujii, S. Dehaene, Large-Scale Cortical Networks for Hierarchical Prediction and Prediction Error in the Primate Brain. *Neuron* **100**, 1252-1266.e3 (2018).

PUBLICATIONS

A universal rhythm in animal acoustic communication

T. Piette[†], C. Cathcart[†], C. Barbieri, K. M. Ming, D. Grandjean, B. Bickel, E. C. Déaux*, A-L. Giraud*

[†] Joint first authors

* Joint last authors

(Under review in Science Advances)

Short Title

Acoustic signals share an optimum rhythm.

Abstract

Acoustic communication is crucial for survival across the animal kingdom, with acoustic signals being shaped by the interaction of producer and receiver selective pressures. While spectral features' variation reflects species-specific selection, the evolutionary history of acoustic rhythm remains unknown. Using data from 98 species spanning mammals, birds, amphibians, reptiles, fishes, and insects, we investigate whether rhythm is shaped by the producer's physical characteristics, environmental constraints, or social complexity. Phylogenetic modeling did not support an influence of these species-specific selective forces, instead, showing that most species' rhythms are conserved around an evolutionary optimum of 2.9 Hz that falls within the neural delta range (1-4 Hz) and predates mammalian divergence. Given the known conserved brain oscillations across species and delta involvement in active sensing, we propose that, unlike spectral features, acoustic rhythm is governed by a universal neural

mechanism facilitating effective intra and interspecific communication through a shared channel that has persisted through evolutionary times.

Teaser

Animal sounds share a common rhythm around 2.9 Hz, suggesting a common neural basis for communication that spans millions of years across species.

Introduction

Acoustic signals allow for effective, instantaneous communication between individuals even at considerable distances. To be functional these signals' structure must carry adaptive information with the importance of spectral features in doing so being well established (*1–5*). Yet, acoustic signals are not only spectrally but also temporally structured, with rhythm having equally important communicative functions. This is particularly well exemplified by human speech, where speech rate is sufficient for comprehension (*6*, *7*). In most languages, syllable production rate ranges between 4 to 9 syllables per second (*8*). This frequency range corresponds to theta neural oscillations, which flexibly adapt to this rhythm during perception, such that modifying this flow impedes this process and decreases speech intelligibility. Rhythmic patterns allow the identification of syllables, words, and sentences and can help convey meaning, such as emphasis, intonation, and emotional state (*9*, *10*). In animal calls, temporal features are no less important, for instance in vocal recognition (*11*), mating behavior (*12*), and predator avoidance (*13*).

Thus, the temporal patterning of acoustic sequences bears significant communicative function(s) and is not speech-specific. Yet questions remain as to what influences the evolution of rhythm across the animal kingdom. In this study we quantify rhythm across animal clades, test the most prominent hypotheses of signal structure evolution and build models of the most plausible evolutionary scenario. Four major selective forces may drive rhythm evolution. Firstly, the presence of theta rhythm in vocal productions and mouth movements of non-human primates suggests that this rhythm originates from the natural oscillatory movements of the articulators that are directly inherited from mastication (*14–16*). Second, in animals that vocalize, morphological and physiological characteristics, such as breathing rate, heart rate, or metabolism could constrain rhythm range in an analogous manner to spectral features (*17*). Thirdly, to be effective, acoustic signals must reach the receiver despite constraints from the living environment, which could thus influence not only spectral but also temporal features (*18*). Additionally, the sociality-complexity hypothesis suggests a positive relationship between the complexity of the social environment and the complexity of a species' vocal repertoire (*19*). As rhythm would bear a direct relationship with the amount of information that can be transmitted in a given time unit, social complexity may also influence acoustic rate. Finally, it may be that none of these species-specific selective forces have influenced rhythm and that instead it has evolved through phylogenetic mechanisms of conservation and diversification that are either shared or diverse across lineages.

To test these different hypotheses, we quantified rhythm in acoustic sequences from birds, mammals, amphibians, insects, reptiles, and fish. Using Bayesian multilevel models for phylogenetic regression (20), we controlled for phylogenetic relationships and evaluated, in accordance with the previously stated hypotheses, whether weight (as a proxy for breathing rate, heart rate, and metabolism), mastication status, sociality level or ecological characteristics could account for differences in rhythm. Finally, we compared phylogenetic models to assess the most plausible evolutionary scenario of rhythm across animals.

Main Text

Rhythm computation

We analyzed acoustic sequences from 98 species (58 birds, 28 mammals, 4 amphibians, 4 insects, 1 reptile, and 1 fish) to study the evolution of rhythm in animals. We calculated rhythm by analyzing variation in the signal amplitude, allowing broad applicability across species (Fig 1a, 1b, 1c, 1d). Validation against conventional methods showed consistent results (Fig 1e: F2.171=0.33, p=0.72) confirming the robustness of our method. Control of the impact of signal-to-noise ratio (SNR) and sequence length revealed no significant relationship with rhythm (Supp Fig 1a; Supp Fig 1b). Additional investigation of the allometric relationship between weight and dominant frequency

(Supp Fig 2), and of the effect of context on rhythm (Supp fig 4a) further confirmed the validity of our vocal database.

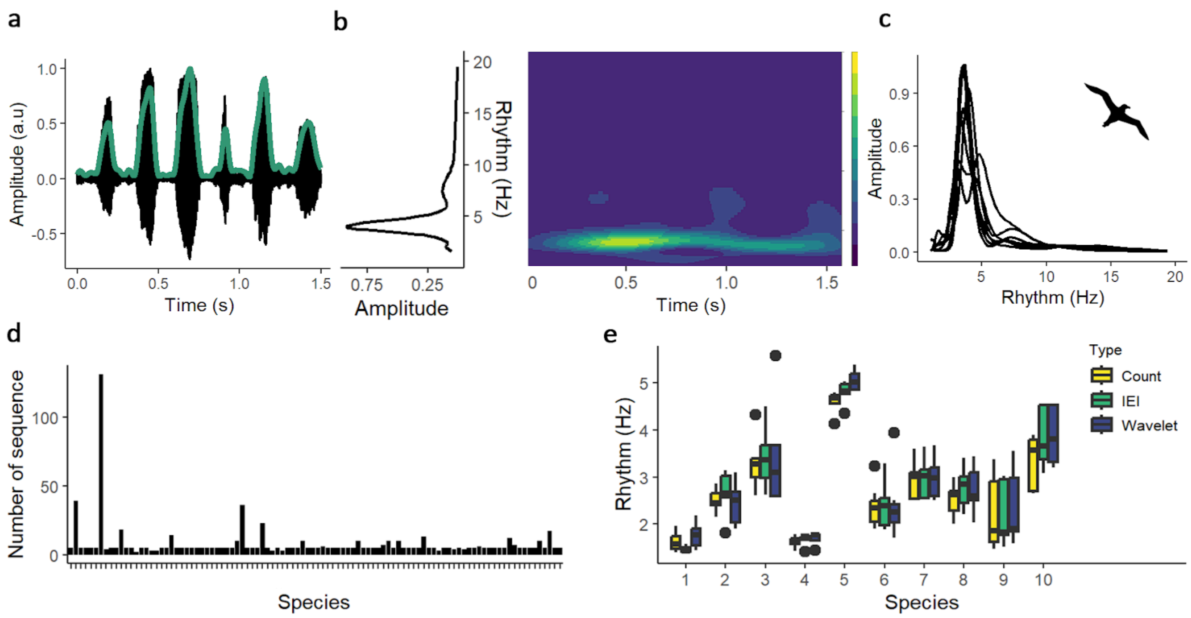


Figure 1. Methodology a) Oscillogram of one acoustic sequence of polar skua call. b) Power spectrum and time frequency representation of the envelope of the previous sequence. c) Power spectra of the envelopes of all polar skua acoustic sequences b) Number of sequences per species (ordered by alphabetical order) c) Rhythm computation using sequences of 10 randomly selected species, employing, from left to right, number of elements per second (Count), inter-element intervals (IEI), and wavelet method (Wavelet).

Species Specific Selective Pressure

To investigate which factors influence the evolution of rhythm, we fitted two phylogenetic regressions in a Bayesian multilevel framework A full model investigating the impact of weight, mastication status, and living environment while controlling for phylogenetic relatedness, and a null model controlling for phylogenetic relatedness only. We compared models via their leave-one-out expected log pointwise density (ELPD) and stacking weight. Due to heteroskedasticity, distributional (scale-location) models better fit the data, leveraging over 90% of the stacking weights (Sup Fig 3). Including the predictors did not improve predictive performance (Fig 2a) and the null model leveraged the highest stacking weight (Fig 2b). The posterior distribution of the regression coefficients of the full model revealed that none of our predictors had any decisive effect on rhythm, given that their 95% credible intervals (CI) all contain zero. We observed only very weak evidence for a negative effect of weight on rhythm in masticating species as zero was outside the 85% CI of their interaction coefficient (Fig 2c, 2d).

Taken together, these results indicate that inclusion of these predictors does not add decisive explanatory value to phylogenetic history. As vocal complexity, our indicator of social complexity, could not be included in the phylogenetic model due to unavailability for all species, we additionally regressed rhythm on vocal complexity (Sup Fig 4d: t=-0.75, p=0.46), revealing no relationship between the two.

Phylogenetic History of Rhythm

Visual inspection of the median rhythm of our tested species shows that in every class, acoustic rhythm spans mainly the lower rates, in what is commonly called the delta range (Fig 2e). To understand if rhythm has randomly evolved in this range, favoring species-specific rhythm, or has been maintained around an optimum value, we fitted models of the evolution of vocal rhythm under Brownian Motion (BM) and Ornstein-Uhlenbeck (OU) processes (*21*), representing each evolutionary scenario respectively.

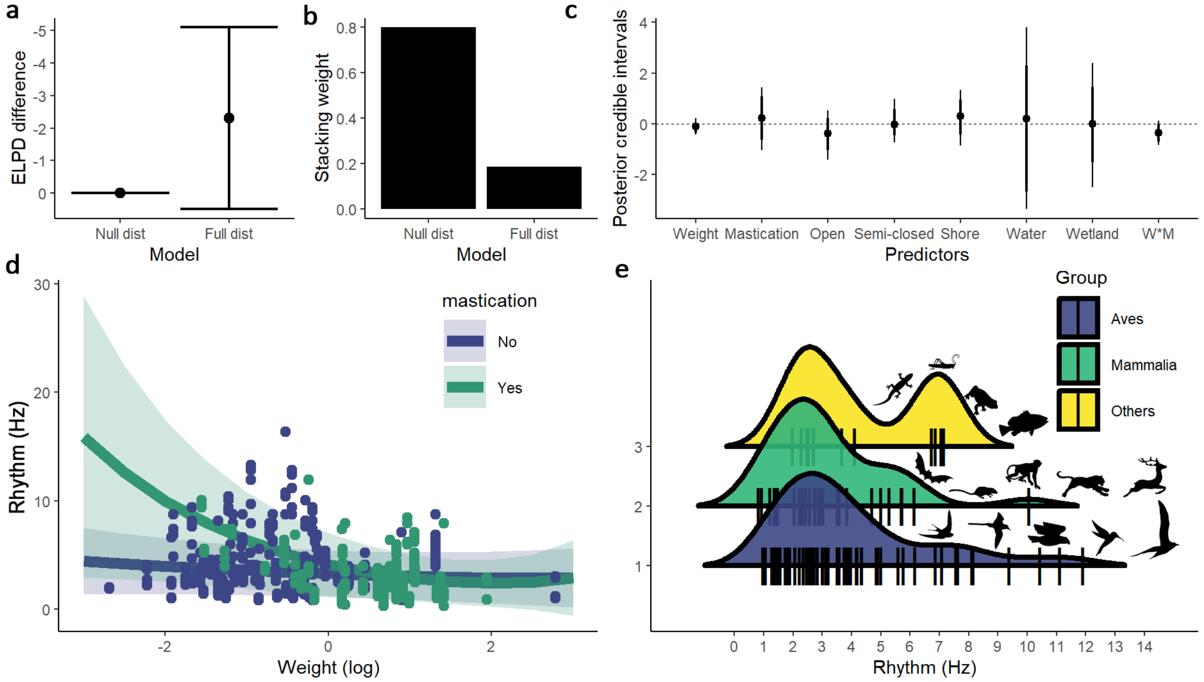


Figure 2. Bayesian multilevel model a) Leave-one-out expected log pointwise density difference (ELPD) between the null and the full distributional (scale-location) models. b) Stacking weight of the models. c) Posterior credible intervals (95% and 85%) of the full distributional model (W*M = Weigh*Mastication) d) Rhythm plotted as a function of log-transformed weight with predicted slopes from the full distributional model and their standard error. e) Distribution of raw median rhythm in Hz in the different clades.

We fitted models of the evolution of vocal rhythm under BM and OU processes. Comparison of ELPD values and model stacking show that the OU model best fit the data (Fig 3a). The result is validated by models (Sup results) fitted separately for mammals and birds alone, with the OU model having the highest stacking weight (birds w_{OU} = 1, mammals w_{OU} = 0.925). The median posterior estimate of rhythm is 2.9Hz, with 95% CI of [.5, 5.1] Hz and 85% credible interval of [1.1Hz, 4.2] Hz (Fig 3b). This represents both the optimum to which the OU process reverts over time and the likely state at the root of the phylogeny. The posterior distribution of sigma represents the stochastic volatility (Fig 3c) and has a median of 0.69 and a 95% CI of (0.29 1.90). The posterior distribution of alpha represent the strength of attraction (Fig 3d) and has a median of 2.79 and a 95% CI of (0.38 6.76). The proportion of posterior half-life estimates lower than 1 (height of the tree) is 97% (Fig 3e), supporting strong selection with fast reversals to the optimum. Consistent with this, rhythm values close to the optimum are also reconstructed for most interior nodes of the phylogeny (Fig 3f). In summary, the model suggests that there is a phylogeny-wide evolutionary pressure towards an optimal rhythm, to which species that deviate quickly revert to.

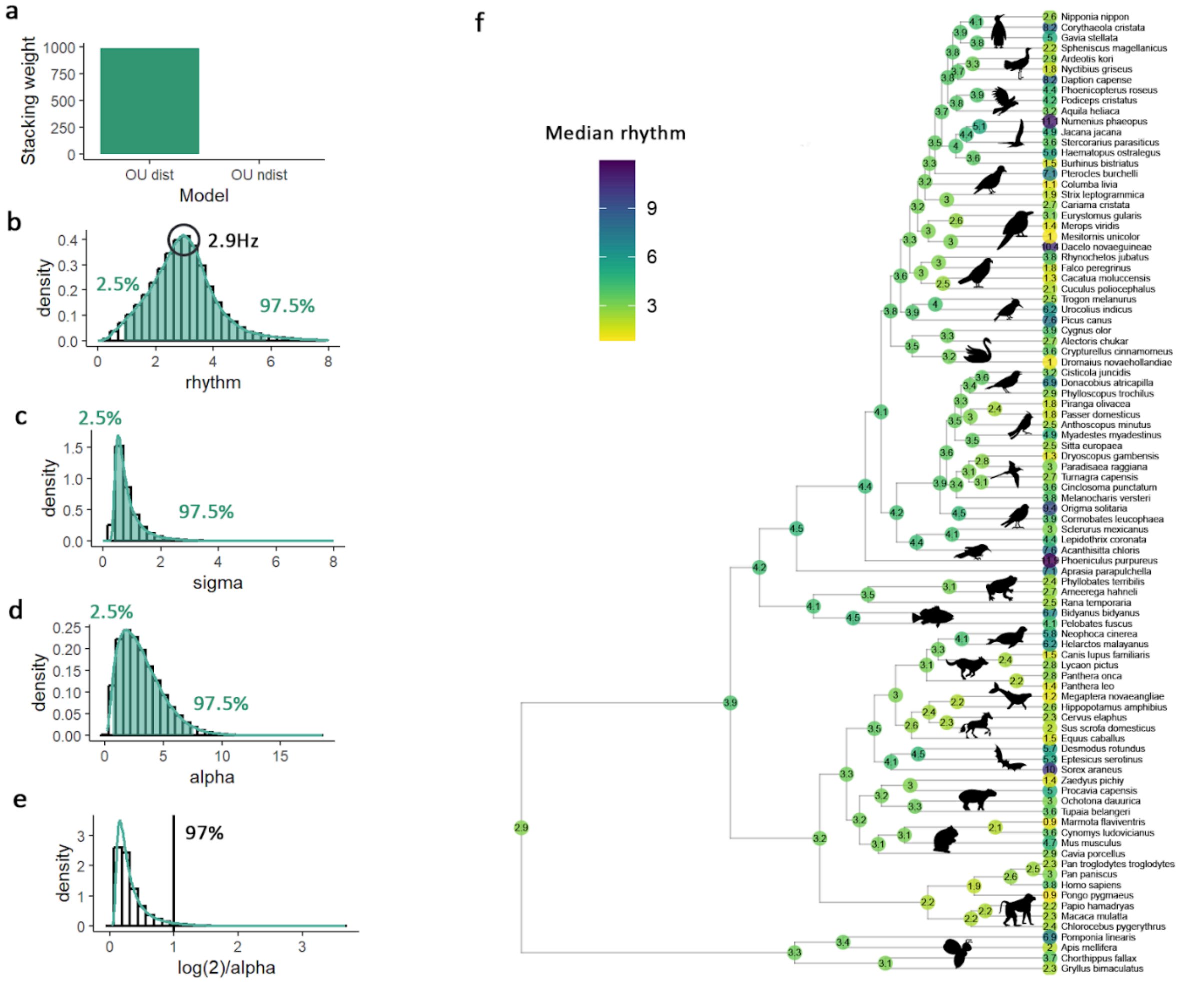


Figure 3. Phylogenetic modeling a) Stacking weight of the OU and BM distributional models. b) Posterior distribution of rhythm in the OU distributional model. c) Posterior distribution of sigma, the scale of the drift process. d) Posterior distribution of alpha, the strength of selection. e) Posterior distribution of the phylogenetic half-life. f) Visualization of reconstructed median posterior values using a maximum clade credibility (MCC) tree.

Discussion

Our finding of an animal-wide optimum rhythm for acoustic communication challenges the idea that it is shaped by biomechanical constraints. The latter would predict distinct rhythms between masticating and non-masticating species, along with a strong negative allometric relationship between rhythm and weight in masticating species. Instead, the weak effect of weight on acoustic rhythms observed in masticatory species indicates that biomechanical constraints only exert a marginal influence. By contrast, our analyses of dominant frequency (Supp fig 2) reconfirmed that spectral characteristics primarily depend on the specifics of the production organ, which chiefly vary with weight (*35*).

With regards to the emergence of this optimum acoustic rhythm, its widespread evidence among birds and mammals indicates its presence at least in their last common ancestor around 340 million years ago. Further, its existence in more remote species in our sample (insects, amphibians, fishes) speaks to even older roots (Fig. 3). This begs the question as to what factors contributed to the emergence and persistence of this rhythm throughout evolutionary times. In addition to showing that producteur physical characteristics have only a peripheral effect, our analyses demonstrated that, unlike spectral features, rhythm evolution cannot be explained by either environmental pressures nor social complexity (Fig 2c, Sup Fig 4d). Likewise, given the diversity of hearing structures (cochlea, otic vesicle, tympanal organs etc.) and production modes (vocalization, stridulation, percussion etc.), it seems unlikely that this common rhythm would be constrained by similarities in receivers' anatomical traits (*23*).

By contrast, basic neural mechanisms are remarkably conserved, and could therefore better explain a widespread optimum rhythm. On the production side, vertebrates share a common location for motoneurons responsible for vocalizations in the caudal hindbrain, believed to be involved in call duration and timing, supporting the claim of conserved mechanisms for acoustic rhythmic production (*24*). In terms of auditory reception, this conserved rhythm around 2.9 Hz (85% interval 1.1Hz-4.2Hz) best matches delta brain oscillations (1–4 Hz), which have been observed across species, including mammals, reptiles, and insects (*25–27*). Interestingly, delta oscillations are linked to active sensing, where organisms' perception occurs via the cyclic sampling of their environment (*71*). Slow oscillations help integrate sensory information and, being the primary gate for external stimuli perception, are particularly suited for integrating slow-varying acoustic cues important for signal detectability and discriminability, as in vocal recognition (*29*). Recent findings even show that despite millenia of domestication, dogs process speech at a slower rhythm between 1Hz to 3 Hz (*30*), further strengthening this notion of a conserved and effective communication rate across species. In sum, this production rhythm best matches a neural rhythm that is important for assigning significance to acoustic signals, which would thus result in an effective communicative design.

Yet, speech perception research has revealed that auditory processing relies on (at least) a two-timescale processing (*10*). Similarly, our control analyses on context show that different call types operate within this conserved rhythm (Sup Fig 4), requiring a faster analysis window to be resolved. Interestingly, calling rate is typically linked to arousal (*31*) with high vocal rhythms eliciting rapid responses, possibly by targeting salience-related brain networks (*32*, *33*). As such, it is possible to envision a dual receptive strategy relying on slow analyses for signal identification, and fast analyses for within-repertoire call discrimination (*34*–*37*). Though this report does not demonstrate the fast timescale's evolution, an interesting avenue would be to probe for an optimum that combines fast and slow scales during acoustic communication.

Finally, these results suggest another interesting aspect. The maintenance throughout evolution of a slow rhythm for acoustic signal production and perception de facto results in a common communication channel across coexisting species. The common low-range (delta scale) could thus permit cross-species identification and offer possibilities for interspecies signaling (e.g. a common danger) and/or eavesdropping, conferring evolutionary advantages.

In summary, spectral and temporal features, both key to effective communication, evolve following different trajectories. While spectral traits diversify along with the hearing organ anatomy, rhythm may be primarily shaped by basic neural factors, responsible for both production and perception processes and which lead to a conserved optimal rhythm.

Materials and Methods

Vocal sequences

To perform an extensive phylogenetic comparison on a balanced phylogeny, we collected acoustic and biological data for at least one species per infra-order of tetrapods, when data were available, as well as a few species of insects and fishes, to obtain a good representation of rhythm throughout the phylogeny. Acoustic sequences were gathered from public and private databases (Fonozoo, Cisro, Berlin Museum fur Naturkunde and Xeno-canto), online videos platform (Youtube, Dailymotion), and from different research groups that kindly shared audio files.

Weight

Weight presents an allometric relationship with morphological features and physiological processes involved in acoustic productions across various species. Thus we collected biological data on the mean weight for each species, to serve as a proxy for heart rate, breathing rate and metabolism. This involved calculating the average weight by considering both the minimum and maximum weights recorded for each species, irrespective of sex. These data were primarily obtained from the handbook of mammals of the world (38), and the handbook of birds of the world (39). When data were unavailable from these sources, we looked for reference articles.

Beak Size

Just as masticatory abilities may have influenced rhythm in mammals, a similar proposition could be made regarding beak morphology in birds. Unlike other animals, the morphological traits of a bird's beak do not consistently adhere to an allometric relationship with its weight (40). Thus we collected information on beak length, width and depth of our species. As these measures were not available for all species, we could not include them in the phylogenetic model. We therefore built an additional linear mixed model investigating the variation of rhythm, including group and order as random effects, and beak length, width, depth, and their interactions as fixed effects. Data are available in the study github.

Living environment

As environmental conditions can impact vocal communication (*18*), we also collected data on the typical habitat of each species. We used a five level categorical classification, with habitats being either classified as closed (defined as habitats with heavy tree coverage), semi-closed (defined as habitats with light three coverage or human cities), open (defined as fully open habitat with no three coverage or obstacle), shore (for species living near a significant amount of water such as lakes, rivers or seas) or water (for species living below the water surface). These data were primarily obtained from the handbook of mammals of the world, and the handbook of birds of the world. When data were unavailable from these sources, we looked for reference articles. Data and linked references are available in the study github.

Mastication status

As some have proposed that rhythmic communication in vocalizing animals may be linked to mastication regime (*14–16*), we also classified each species according to their mastication status (yes or no).

Social complexity

As social complexity increases, individuals may need to communicate more information in a given time, and therefore speed up their communication. As communication signals have been linked to social complexity (*19*), we also gathered species vocal repertoire complexity (number of distinct calls in the species vocal repertoire) when this information was available. As this measure was not available for all species, we could not include it in the phylogenetic model. We therefore built an additional linear mixed model investigating the variation of rhythm, including group and order as random effects, and vocal complexity as fixed effect. These data were primarily obtained from the handbook of birds of the world, and reference articles. Data and linked references are available in the study github.

Acoustic sequence selection and pre-processing

Based on a cross-species literature search (*41–44*), we defined sequences as recordings of acoustic displays emitted by a single individual, containing more than two calls separated by less than two seconds of silence. For computational purposes, we selected only recordings lasting more than one second. We included in our analysis species for which we had at least five different sequences from five different individuals. Species with fewer sequences were included if they were the only available representatives of their infra-order, resulting in a database of 98 species, including 58 birds, 28 mammals, 4 amphibians, 4 insects, 1 reptile, and 1 fish.

Rhythm analyses

To quantify rhythm in these acoustic sequences, we decided to adapt the method developed by Tilsen et al to compute rhythm in human speech production (45). This method uses the signal amplitude to automatically compute the rhythmic component of a sequence, without making any assumption on the components' size, and is thus widely applicable across all species regardless of variations in unit size or spectral characteristics. First, we denoised the sequences using a first-order Butterworth filter, with a bandpass filter between a minimal frequency (minF), defined as 200 Hz below the minimum frequency of the animal call, and a maximum frequency (maxF), defined as 200 Hz above the maximum frequency of the animal acoustic signal obtained from reference articles. When this information was not available, we applied a large range filter with a 100 Hz minF and 10000 Hz maxF. We then computed the normalized envelope of the denoised sequences using the Hilbert Transform. Next, we low pass-filtered this envelope with a fourth-order Butterworth filter with a 20 Hz cut-off frequency to obtain the slow changes in acoustic energy. Before further analysis, we downsampled the resulting signal at 150 Hz for computational purposes, and then applied a continuous wavelet transform using the Morlet wavelet to obtain a time-frequency representation of the amplitude envelope. We replaced the Fourier transform with a wavelet transform, to allow for more flexibility with regards to the variation in sequence length present in our dataset. We finally analyzed that representation's power spectrum to extract the five frequency peaks of highest amplitude in the power spectrum and used the time-frequency representation to select the main rhythmic component conserved across the entire sequence.

To assess the validity of the proposed methodology, we conducted a comparative analysis between the calculated vocal rates of a subset comprising 10% of our database and those derived from two widely accepted conventional approaches: 1) by counting the number of elements per second (Count) and 2) by computing the inter-element interval (IEI). The three methods gave sensibly similar results (Figure 1d: $F_{2.171}=0.33$, $p=0.72$), hence validating our rhythm quantification method.

Signal to noise ratio (SNR) and length

As further control analyses, we quantified recording durations in seconds and signal-to-noise ratio in decibels. To control for the effect of both factors and their interaction on rhythm, we build a linear mixed model investigating the variation of rhythm including group and order as random effects, and SNR, length and their interaction as fixed effects.

Dominant frequency analyses

To determine dominant frequencies, which unlike the fundamental frequency are measurable in all types of communicative signals (*22*), we isolated the first acoustic unit in each denoised sequence. We then applied a single discrete Fourier transform to compute the power spectrum of these units, and extract the peak of highest amplitude. The obtained results were also visually controlled in Praat, to make sure that the extracted dominant frequency matched the acoustic energy present in the unit. If the first unit had poor signal-to-noise ratio leading to inaccurate computation of the dominant frequency, we selected the next unit in the sequence.

Context effect control

While the existing literature highlights the importance of context and its correlated arousal levels on acoustic signal rate (*1*), for most species we were not able to obtain these data. Nevertheless, whenever possible we selected recordings of different call types for each species. Further, we performed separated analysis of variance (ANOVA) to control for the effect of call type on rhythm in three species: one avian and two mammalian, from which we could obtain different call types, including contact calls, alarm calls, songs, agonistic and antagonistic vocal displays.

Phylogenetic tree sample reconstructed from genetic sequences

To test our hypotheses of interest we used phylogenetic comparative methods (PCM), a broad family of methodological tools for characterizing and controlling for the evolutionary dynamics thought to give rise to the data under study. PCMs require a representation of the relatedness of the taxa under study in the form of a phylogenetic tree sample. To represent the tree topology of the species we performed a phylogenetic analysis based on comparable genetic sequences, using a Bayesian framework to infer a posterior tree sample. We first matched each species in the sample with their closest genetic proxies in GenBank (*46*), and extracted mitochondrial DNA for the corresponding species. For 54 of the 98 species, matching mitochondrial genomes were available from literature and deposited in GenBank. For the 44 remaining species we chose proxies from another closely related species. We took a species within the same genus when possible; if this was not available, we chose species within the family of the target species, after confirming that no more than one species per family was included in the original list. Only for four target species - *Correlophus ciliates* (Squamata), *Galbula ruficauda* (Aves), *Leptosomus discolor* (Aves), *Phaethon rubricauda* (Aves) - we did not find proxies within the family and we had to find a proxy within the order. To choose the best mtDNA proxy with those deposited in GenBank, we considered completeness of the available mitochondrial sequences and comparable average size, weight and environment of the target species. Maximum missing data is 100 base pairs, for an average size of 16706 base pairs. MtDNA genomes were aligned with MAFFT software (*47*) and standard settings. The alignment was manually screened in BioEdit (version 7.2, https://thalljiscience.github.io/) for spotting irregularities and potential outlier sequences. Sequences were then cut to keep only the coding region, which is more conserved across species, using the *Homo sapiens* sequence as a reference. The final alignment consisted of 21860 base pairs, which include large INDELs sections to accommodate alignment between the most divergent species (e.g. *Apis mellifera*).

We used BEAST2 to generate the trees, running 10'000'000 iterations of Markov chain Monte Carlo (MCMC) with a thinning interval of 1000. We used the following settings to approximate the broad evolutionary range of the species considered: assuming an HKY substitution model, a strict clock (Uniform rates across branches), and a

Birth-Death tree prior with a Yule birth rate. This resulted in 10'000 trees, of which we use 50 for phylogenetic comparative analyses.

Bayesian multilevel models for phylogenetic regression

To assess the impact of several predictors of interest on different properties of variation in vocal rhythm and dominant frequency, we used phylogenetic regression modeling, a comparative method that assesses the effect of predictors on a response while controlling for the phylogenetic relatedness of the taxa. Due to heterogeneity in the number of datapoints and individuals in each species, we employed both non-distributional regression models, which model the mean of the response variable as a function of predictors, and distributional (scale-location) regression models, which model both the mean and standard deviation of the response variable as a function of predictors. We control for species-level idiosyncrasies in both the median and (in some cases) standard deviation of rhythm via phylogenetic random intercepts and slopes (phylogenetic random effects are similar to the standard random effects used in hierarchical regression modeling, but are generated by a Gaussian Process with a covariance kernel that is a function of the phylogenetic patristic distances between species under study rather than independently and identically distributed with diagonal variance).

We fitted four phylogenetic regression models using brms (*20*) for each response variable (vocal rhythm, dominant frequency), resulting in eight models. These predictors consisted of average species weight in kilograms, which was log-transformed, centered around zero, and standardized (*48*), mastication status, and living environment of the species. We also consider the interaction between mastication status and weight.

Our distributional models have the following basic generative process (below, $\alpha^{\mu} + \beta X_i$ is shorthand for all model predictors, including fixed and random effects):

$$y_i \sim LogNormal(\mu_i, \sigma_i)$$

$$\mu_i = \alpha^{\mu} + \beta^{\mu} X_i$$

$$\sigma_i = exp(\alpha^\sigma + \beta^\sigma X_i)$$

Non-distributional models have the following structure:

$$y_i \sim LogNormal(\mu_i, \sigma)$$

$$\mu_i = \alpha^\mu + \beta^\mu X_i$$

We employ the default priors of brms.

The first of the four models was a full distributional one that modeled both the expected median and variance of the response variable as a function of these predictors, while controlling for species-level idiosyncrasies in both the median and variance of rhythm via phylogenetic random intercepts and slopes. The second was a full non-distributional model that treated only the expected median rhythm as a function of the predictor variables as well as phylogenetic random intercepts and slopes. The third of these was a null distributional model that included only phylogenetic random intercepts and slopes for the expected median and variance. The final model was a null non-distributional that included only phylogenetic random intercepts and slopes for mean rhythm. We ran each of these models for 4000 iterations of the no U-turn sampler over 4 chains with a log-normal link function and discarded the first half of samples, aggregating posterior samples across the retained sampled trees.

Models in brms have the following formulae:

Full, distributional

```
bf(frequency ~ weight*mastication + environment + (1 +
weight*mastication + environment | gr(taxon, cov =
phylo.cov))

sigma ~ weight*mastication + environment + (1 +
weight*mastication + environment | gr(taxon, cov =
phylo.cov)))
```

Full, non-distributional

```
bf(frequency ~ weight*mastication + environment + (1 +
weight*mastication + environment | gr(taxon, cov =
phylo.cov)))
```

Null, distributional

```
bf(frequency ~ (1 | gr(taxon, cov = phylo.cov)), sigma ~
(1 | gr(taxon, cov = phylo.cov)))
```

Null non-distributional

```
bf(frequency ~ (1 | gr(taxon, cov = phylo.cov))
```

We compared fitted models via their leave-one-out expected log pointwise density (ELPD) values (*49*) and stacking (50), which average predictive distributions of different models to generate weights representing their relative predictive power. We used the function loo_compare to measure differences in ELPD across models. Finally, We inspected posterior distributions of regression coefficients of the full distributional model to assess the effects of predictors of interest.

Evolutionary dynamics of vocal rhythm

We further investigate the properties of rhythm across species using two Gaussian Process models of continuous trait evolution, asking specifically whether the evolution of vocal rhythm is characterized by a random process of drift (characterized by Brownian motion) or whether selective forces draw rhythm values toward an optimal value over time (a mean-reverting scenario characterized by an Ornstein-Uhlenbeck process). Under Brownian motion, the displacement of a continuous trait at time s has a variance proportional to the amount of time elapsed over the course of displacement (below denoted as t), where σ represents the scale of the drift process:

$$X(s) \sim Normal(X(s - t), \sigma t)$$

Under an OU process, the displacement of a character has the following formula:

$$dX(s) = \alpha(\theta - X(t))dt + \sigma dW(t)$$

In the first component of the sum, α represents the strength of selection to the optimal value θ. The second component represents a process of Brownian motion, with σ representing the scale of drift. Thus, the OU process allows for both selective and random forces in character evolution.

An standard way to interpret α is to transform it to the phylogenetic half-life, ln 2/α (50). This is interpreted as the average time for a trait to evolve halfway from an ancestral state toward a new optimum, indicating how long it will take before adaptation to a new regime is more influential than constraints from the ancestral state. If half-life values are greater than the height of the phylogeny (1 in our case, as the tree length is scaled to unit height), the process increasingly resembles Brownian motion and involves a slower adaptation speed.

As above, we employ a distributional approach, allowing species-level mean rhythm values and species-level standard deviations of rhythm values to evolve over the phylogeny according to BM or OU processes.

The distributional BM process has the following generative process:

$y_i \sim LogNormal(\mu[species_i], \sigma[species_i])$
$\mu \sim MVNormal(\theta_0^\mu, \tau^\mu \Psi)$
$\sigma \sim MVNormal(\theta_0^\sigma, \tau^\sigma \Psi)$

θ_0 represents the trait value at the root of the tree, while Ψ is a matrix of the shared history (the time between the root age of the tree and the most recent common ancestor) of each pair of nodes in the tree and τ is the positive scale of drift. Conventions are as above.

The distributional OU process has the following generative process:

$y_i \sim LogNormal(\mu[species_i], \sigma[species_i])$
$\mu \sim MVNormal(\theta_0^\mu, \tau^\mu exp(-\alpha^\mu \Delta))$
$\sigma \sim MVNormal(\theta_0^\sigma, \tau^\sigma exp(-\alpha^\sigma \Delta))$

θ_0 represents the trait value at the root of the tree, τ is the positive scale of drift, α is the positive strength of selection, and Δ is a matrix of pairwise cophenetic distances between species in the phylogeny, scaled to a maximum distance of 1. Conventions are as above. We place *Normal(0,1)* priors over unconstrained parameters and *Gamma*(1,1) priors over positive parameters. In addition to running these models on all species in our sample, we validate results by running models on bird and mammal species alone.

Phylogenetic reconstruction of rhythm values

Rhythm values were reconstructed to internal nodes of the maximum clade credibility (MCC) tree of the phylogeny, using ggtree package (*51*) by drawing ten draws from each of the posterior distributions inferred from the 50 different trees in the tree sample and sampling values at internal nodes of the tree from the normal distribution parameterized by the OU process, conditioned at the expected tip values (fig 3f).

Software:

All analyses and visualization were done using Stan and R version 4.1.2 (2021-11-01) with the following packages Seewave (*52*), Soundgen (*53*) , DoBy (*54*) , Lme4 (*55*), MuMYn (*56*), brms (*20*), ggplot2 (*57*).

Acknowledgments

The NCCR Evolving Language, Swiss National Science Foundation Agreement Nr. #51NF40_180888, funded this work. The authors are grateful to the CSIRO Australian National Wildlife Collection, (https://ror.org/059mabc80), Fonoteca Zoologica (https://www.fonozoo.com/), the Museum für Naturkunde of Berlin (https://www.museumfuernaturkunde.berlin) and the Xeno-Canto Foundation for Nature Sounds (https://xeno-canto.org) for their assistance in undertaking this research. The authors thank all the researchers from the NCCR Evolving Language as well as Pascal Belin and Steffen R.Hage for their insights on the project. The authors would like to thank Aaron Bauer, Andrew Spencer, Alex Kwet, Alex Rohtla, Camila Ferrara, Daniel Blumstein, Élodie Briefer, Émilie Genty, Frank Lambert, Isabelle

Charrier, Gerald Carter, Hans Schneider, Marc D. Hauser, Nikola Falk, Peter Boesman and Robert Seyfarth for providing us with recordings for this study. ALG is funded by the Fondation pour l'Audition (GA-FPA IDA11).

Competing Interest

All authors declare that they have no conflicts of interest.

Contributions

T.P, E.D, D.G, and A-L.G conceptualized the project. T.P conducted the acoustic data collection, while K.K.M was responsible for the genetic data collection. The acoustic analysis methodology was developed by T.P and E.D, with T.P conducting the acoustic data analysis. C.B constructed the phylogenetic tree. Phylogenetic analyses were designed by C.C and B.B and carried out by C.C. T.P wrote the initial draft, with C.C providing detailed methodological input on the phylogenetic analysis. All authors contributed to the review and editing of the manuscript. E.D and A-L.G provided supervision. A-L.G provided resources, and funding for the project.

Data and materials availability

All data and code used for the analysis in this study are available in a public GitHub repository. This repository contains the raw data, scripts for data processing, analysis, and visualization, as well as detailed instructions for reproducing the results. The repository can be accessed at https://github.com/chundrac/phylo-acoustic-rhythm.

Supplementary Text

Phylogenetic regression of Dominant Frequency (DF)

As our acoustic data were gathered from both supervised and unsupervised dataset, we decided, as a validation procedure, to check for the presence of the well established allometric relationship between weight and dominant frequency in our dataset. To do so, we fitted two phylogenetic regressions using Bayesian multilevel models. A full model investigating the impact of weight, mastication status, and living environment while controlling for phylogenetic relatedness, and a null model controlling for phylogenetic relatedness only. Due to heterogeneity in the number of datapoints and individuals in each species, we employed both non-distributional and distributional regression models. The full distributional model with predictors included has the highest ELPD value (Sup Figure 2a). In terms of stacking weight, we see that the full distributional models with predictors has the majority of the stacking weight, but the null distributional model still has a substantial amount of stacking weight (Sup Figure 2b). We inspect the credible intervals (CIs) of posterior distributions of model coefficients of the the full distributional model to determine which predictors have an effect on dominant frequency, the response value. 95% and 85% CIs for model coefficients are found below, along with the proportion of posterior samples for which the coefficient value is greater or less than zero. We take coefficients whose 95% CIs exclude zero to represent decisive evidence for an effect, and coefficients whose 85% CIs exclude zero to represent strong evidence for an effect. As is clear, there is a decisive negative effect of weight on dominant frequency, and strong evidence for a negative interactive effect between mastication and weight on dominant frequency (Sup Figure 2c, 2d), thus validating the accuracy of our database.

Phylogenetic History of Rhythm in birds and mammals

Independent analyses of the evolution of rhythm in birds and mammals confirm the results on the entire dataset, with the OU model having the highest stacking weight (birds = 1, mammals = 0.925). This indicates that evolutionary history of rhythm is consistent across clades, showing conservation of rhythm and its constraints across the different groups.

Context

As the existing literature highlights the importance of context and its correlated arousal levels on acoustic signal rate, we investigated the effect of context on acoustic rhythm in three distinct species. Analysis of variance (ANOVA) of the effect of call type on rhythm in baboons (*Papio anubis*) revealed significant effect of call type on rhythm

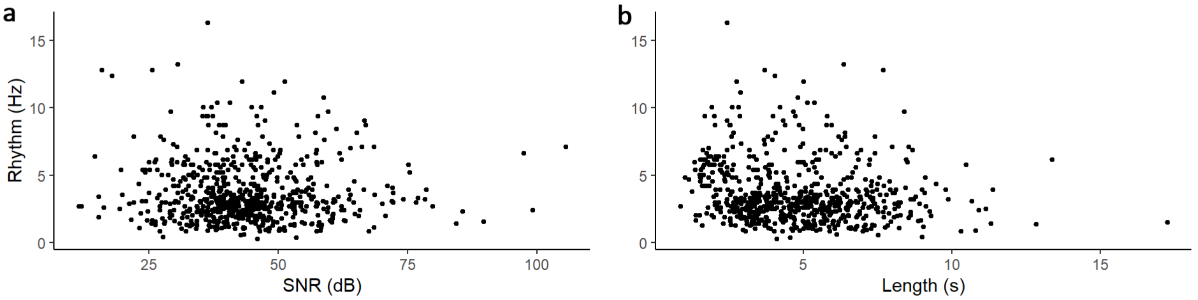
($F_{2,27}$ = 14.99, p=0.0004) with post-hoc tukey tests showing significant differences in rhythm between scream and affiliative grunt, as well as between threat grunt and affiliative grunt (Sup fig 4a). Similar analysis in European stone-curlews (*Burhinus oedicnemus)* and dogs *(Canis lupus familiaris)* showed comparable results ($F_{2,21}$=4.546, p=0.02; $F_{4,136}$=3.632, p=0.007) , with call type having a significant impact on rhythm. However, the median rhythm for each call types are contained between ± 1Hz around the median rhythm of the species (Sup fig 4a,4b,4c). This illustrates that, even though context bears a significant influence on vocal rate, this influence is limited to a narrow and constrained range around the median vocal rate of the species.

Vocal complexity

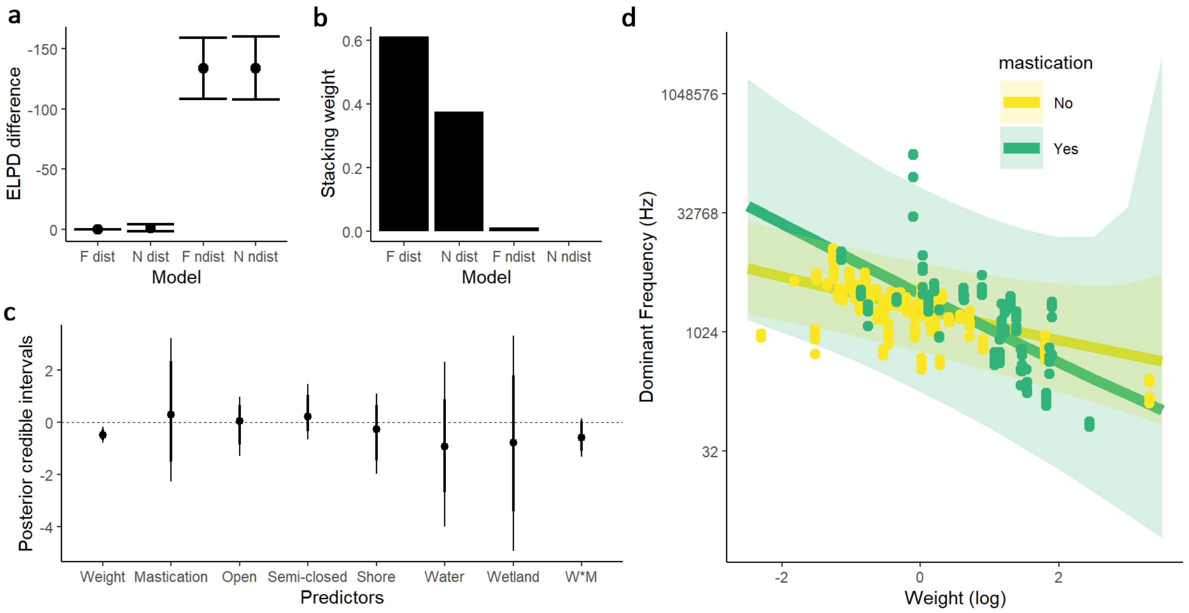
As vocal complexity is linked to social complexity, we build a linear regression model between rhythm and vocal complexity, to investigate the effect of social complexity on rhythm in animals. Vocal complexity did not show any significant relationship with rhythm (t=-0.748, p=0.46). The model failed to account for a significant proportion of variance in rhythmic behavior (R^2=-0.01, p=0.46), suggesting that other factors beyond social complexity may influence rhythmic behavior (Sup Figure 4d).

Beak size

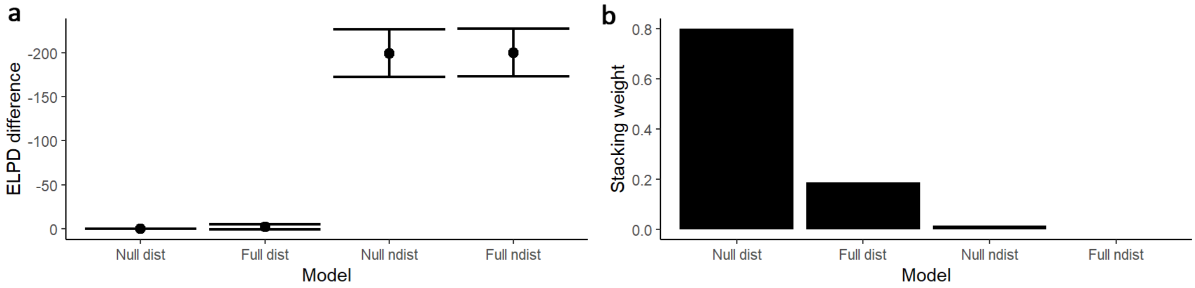
To investigate the effect of beak morphology on rhythm in birds, we build a linear regression model examining the relationship between rhythm, beak length, width, depth, and their interactions. Beak length (t=-1.4, p= 0.17), beak width (t=-0.98, p= 0.33), and beak depth (t=0.35, p=0.73) did not show any significant relationship with rhythm (Sup Figure 4e,4f,4g). Similarly, interactions between these variables were not significant predictors of rhythmic behavior (t=-0.41, p=0.68). The model failed to account for a significant proportion of variance in rhythmic behavior (R^2=-0.01, p=0.5), suggesting that other factors beyond morphological traits may influence bird rhythmic behavior.



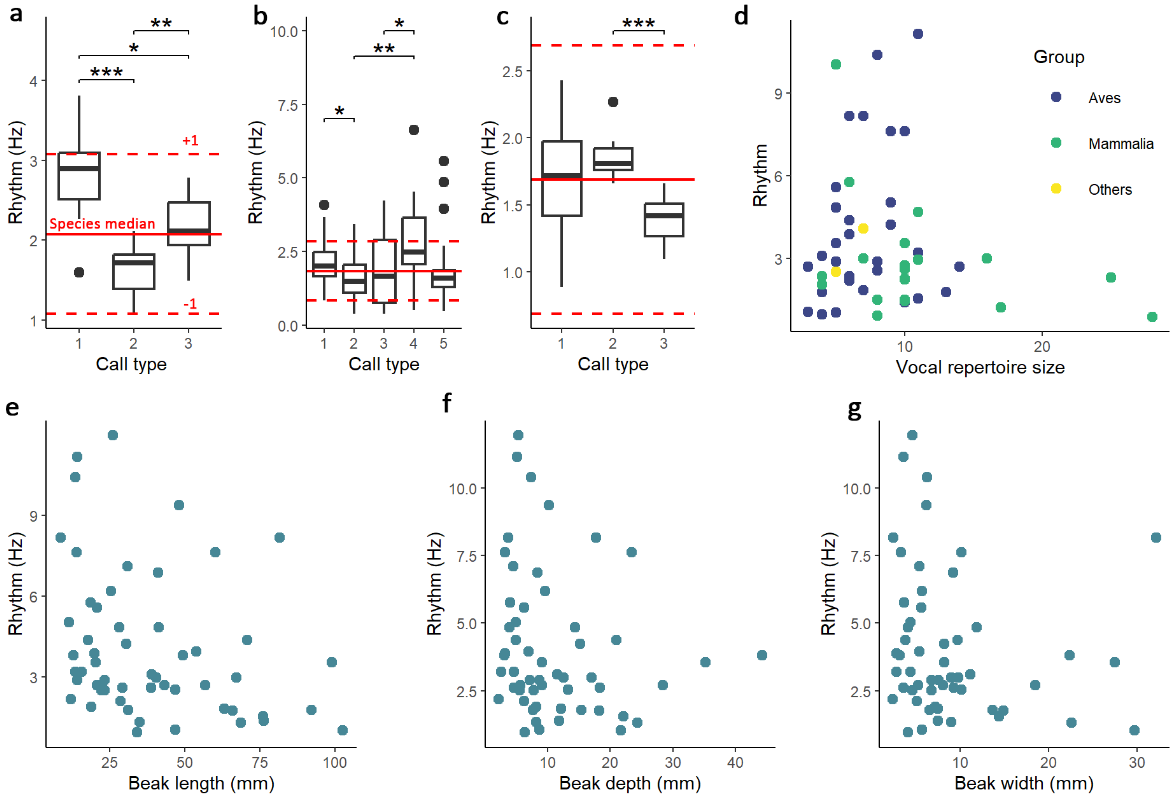
Supplementary Figure 1 : Signal to Noise Ratio and sequence length effect on rhythm a) Rhythm (Hz) as a function of signal-to-noise ratio (SNR) showing the absence of relationship between SNR and Rhythm (t=-0.21, p=0.84, R^2=0.02) b) Rhythm (Hz) as a function of recording length showing the absence of relationship between sequence length and rhythm (t=-1.45, p=0.15, R^2=0.02).



Supplementary Figure 2. Dominant Frequency Phylogenetic Regression a) Leave-one-out expected log pointwise density difference (ELPD) between the full and the null distributional models. b) Stacking weight of the models. c) Posterior credible interval (95% and 85%) of the full model d) Dominant frequency plotted on a logarithmic scale as a function of log-transformed weight with predicted slopes from the full distributional model and their standard error.



Supplementary Figure 3. Rhythm Phylogenetic Regression a) Leave-one-out expected log pointwise density difference (ELPD) between the null distributional ("dist") model and the others ("ndist" for "non-distributional, modeling only the mean") . b) Stacking weight of the models



Supplementary figure 4 Additional Analysis on Rhythm a) Rhythm in sequences of different call types (1=affiliative grunt, 2=scream, 3=threat grunt) in olive baboons *(Papio anubis)*. b) Rhythm in sequences of different call types (1=bark, 2=growl, 3=howl, 4=snarl, 5= whine) in dogs *(Canis lupus familiaris)*. c) Rhythm in sequences of different call types (1=alarm call, 2=flight call, 3=song) in Eurasian stone-curlews *(Burhinus oedicnemus)* d) Rhythm plotted as a function of vocal repertoire size. e) Rhythm plotted as a function of beak length in birds. f) Rhythm plotted as a function of beak depth in birds. g) Rhythm plotted as a function of beak width in birds.

References

1. E. F. Briefer, Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool* **288**, 1–20 (2012).
2. N. H. Fletcher, A simple frequency-scaling rule for animal communication. *The Journal of the Acoustical Society of America* **115**, 2334–2338 (2004).
3. M. Fukushima, A. M. Doyle, M. P. Mullarkey, M. Mishkin, B. B. Averbeck, Distributed acoustic cues for caller identity in macaque vocalization. *R. Soc. open sci.* **2**, 150432 (2015).
4. H. J. Kriesell, T. Aubin, V. Planas-Bielsa, M. Benoiste, F. Bonadonna, H. Gachot-Neveu, Y. Le Maho, Q. Schull, B. Vallas, S. Zahn, C. Le Bohec, Sex identification in King Penguins *Aptenodytes patagonicus* through morphological and acoustic cues. *Ibis* **160**, 755–768 (2018).
5. D. A. Puts, A. K. Hill, D. H. Bailey, R. S. Walker, D. Rendall, J. R. Wheatley, L. L. M. Welling, K. Dawood, R. Cárdenas, R. P. Burriss, N. G. Jablonski, M. D. Shriver, D. Weiss, A. R. Lameira, C. L. Apicella, M. J. Owren, C. Barelli, M. E. Glenn, G. Ramos-Fernandez, Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proc. R. Soc. B.* **283**, 20152830 (2016).
6. O. Ghitza, S. Greenberg, On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica* **66**, 113–126 (2009).
7. A. Nakamura, N. Seiyama, R. Ikezawa, T. Takagi, E. Miyasaka, “Real time speech rate converting system for elderly people” in *Proceedings of ICASSP ’94. IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, Adelaide, SA, Australia; http://ieeexplore.ieee.org/document/389678/)vol. ii, p. II/225-II/228.
8. C. Coupé, Y. M. Oh, D. Dediu, F. Pellegrino, Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Sci. Adv.* **5**, eaaw2594 (2019).
9. R. Banse, K. R. Scherer, Acoustic Profiles in Vocal Emotion Expression. (1996).
10. A.-L. Giraud, D. Poeppel, Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**, 511–517 (2012).
11. A. A. Ghazanfar, M. D. Hauser, The auditory behaviour of primates: a neuroethological perspective. *Current Opinion in Neurobiology* **11**, 712–720 (2001).
12. T. I. Drăgănoiu, L. Nagle, M. Kreutzer, Directional female preference for an exaggerated male trait in canary (*Serinusanaria*) song. *Proc. R. Soc. Lond. B* **269**, 2525–2531 (2002).
13. D. T. Blumstein, K. B. Armitage, Alarm calling in yellow-bellied marmots: I. The meaning of situationally variable alarm calls. *Animal Behaviour* **53**, 143–171 (1997).
14. T. J. Bergman, J. C. Beehner, M. C. Painter, M. L. Gustison, The speech-like properties of nonhuman primate vocalizations. *Animal Behaviour* **151**, 229–237 (2019).
15. A. S. Pereira, E. Kavanagh, C. Hobaiter, K. E. Slocombe, A. R. Lameira, Chimpanzee lip-smacks confirm primate continuity for speech-rhythm evolution. *Biol. Lett.* **16**, 20200232 (2020).
16. C. Risueno-Segovia, S. R. Hage, Theta Synchronization of Phonatory and Articulatory Systems in Marmoset Monkey Vocal Production. *Current Biology* **30**, 4276-4283.e3 (2020).
17. A. M. Taylor, D. Reby, The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology* **280**, 221–236 (2010).
18. T. Guilford, M. S. Dawkins, Receiver psychology and the evolution of animal signals. *Animal Behaviour* **42**, 1–14 (1991).
19. I. Krams, T. Krama, T. M. Freeberg, C. Kullberg, J. R. Lucas, Linking social complexity and vocal complexity: a parid perspective. *Phil. Trans. R. Soc. B* **367**, 1879–1891 (2012).
20. P.-C. Bürkner, **brms** : An *R* Package for Bayesian Multilevel Models Using *Stan*. *J. Stat. Soft.* **80** (2017).
21. M. A. Butler, A. A. King, Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *The American Naturalist* **164**, 683–695 (2004).
22. D. L. Bowling, M. Garcia, J. C. Dunn, R. Ruprecht, A. Stewart, K.-H. Frommolt, W. T. Fitch, Body

size and vocalization in primates and carnivores. *Sci Rep* **7**, 41070 (2017).
23. B. Warren, M. Nowotny, Bridging the Gap Between Mammal and Insect Ears – A Comparative and Evolutionary View of Sound-Reception. *Front. Ecol. Evol.* **9**, 667218 (2021).
24. A. H. Bass, B. P. Chagnaud, Shared developmental and evolutionary origins for neural basis of vocal–acoustic and pectoral–gestural signaling. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10677–10684 (2012).
25. G. Buzsáki, N. Logothetis, W. Singer, Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms. *Neuron* **80**, 751–764 (2013).
26. G. G. Knyazev, EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neuroscience & Biobehavioral Reviews* **36**, 677–695 (2012).
27. D. Raccuglia, S. Huang, A. Ender, M.-M. Heim, D. Laber, R. Suárez-Grimalt, A. Liotta, S. J. Sigrist, J. R. P. Geiger, D. Owald, Network-Specific Synchronization of Electrical Slow-Wave Oscillations Regulates Sleep Drive in Drosophila. *Current Biology* **29**, 3611-3621.e3 (2019).
28. B. Morillon, L. H. Arnal, C. E. Schroeder, A. Keitel, Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neurosci Biobehav Rev* **107**, 136–142 (2019).
29. M. D. Hauser, B. Agnetta, C. Perez, Orienting asymmetries in rhesus monkeys: the effect of time-domain changes on acoustic perception. *Animal Behaviour* **56**, 41–47 (1998).
30. E. C. Déaux, T. Piette, F. Gaunet, T. Legou, L. Arnal, A.-L. Giraud, Dog–human vocal interactions match dogs' sensory-motor tuning. *PLoS Biol* **22**, e3002789 (2024).
31. A. Ravignani, S. Dalla Bella, S. Falk, C. T. Kello, F. Noriega, S. A. Kotz, Rhythm in speech and animal vocalizations: a cross-species perspective. *Ann. N.Y. Acad. Sci.* **1453**, 79–98 (2019).
32. L. H. Arnal, A. Flinker, A. Kleinschmidt, A.-L. Giraud, D. Poeppel, Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology* **25**, 2051–2056 (2015).
33. L. H. Arnal, A. Kleinschmidt, L. Spinelli, A.-L. Giraud, P. Mégevand, The rough sound of salience enhances aversion through neural synchronisation. *Nat Commun* **10**, 3671 (2019).
34. P. Lakatos, A. S. Shah, K. H. Knuth, I. Ulbert, G. Karmos, C. E. Schroeder, An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology* **94**, 1904–1911 (2005).
35. R. Narayan, G. Graña, K. Sen, Distinct Time Scales in Cortical Discrimination of Natural Sounds in Songbirds. *Journal of Neurophysiology* **96**, 252–258 (2006).
36. I. Olasagasti, A.-L. Giraud, Integrating prediction errors at two time scales permits rapid recalibration of speech sound categories. *eLife* **9** (2020).
37. A. F. Osman, C. M. Lee, M. A. Escabí, H. L. Read, A Hierarchy of Time Scales for Discriminating and Classifying the Temporal Shape of Sound in Three Auditory Cortical Fields. *J. Neurosci.* **38**, 6967–6982 (2018).
38. D. E. Wilson, R. A. Mittermeier, P. Cavallini, Eds., *Handbook of the Mammals of the World* (Lynx Edicions : Conservation International : IUCN, Barcelona, 2009).
39. J. del Hoyo, A. Elliott, J. Sargatal, J. Cabot, Eds., *Handbook of the Birds of the World* (Lynx Edicions, Barcelona, 1992).
40. L. M. K. Rombaut, E. J. R. Capp, C. R. Cooney, E. C. Hughes, Z. K. Varley, G. H. Thomas, Allometric conservatism in the evolution of bird beaks. *Evolution Letters* **6**, 83–91 (2022).
41. C. Boesch, C. Crockford, Call combinations in wild chimpanzees. *Behav* **142**, 397–421 (2005).
42. K. Collier, S. W. Townsend, M. B. Manser, Call concatenation in wild meerkats. *Animal Behaviour* **134**, 257–269 (2017).
43. C. Coye, K. Ouattara, K. Zuberbühler, A. Lemasson, Suffixation influences receivers' behaviour in non-human primates. *Proc. R. Soc. B.* **282**, 20150265 (2015).
44. S. Engesser, A. R. Ridley, S. W. Townsend, Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5976–5981 (2016).
45. S. Tilsen, K. Johnson, Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America* **124**, EL34–EL39 (2008).
46. R. B. J. Benson, P. D. Mannion, Multi-variate models are essential for understanding vertebrate

diversification in deep time. *Biol. Lett.* **8**, 127–130 (2012).

47. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).
48. A. Gelman, J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, ed. 1, 2006; https://www.cambridge.org/core/product/identifier/9780511790942/type/book).
49. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* **27**, 1413–1432 (2017).
50. M. Grabowski, B. T. Kopperud, M. Tsuboi, T. F. Hansen, Both Diet and Sociality Affect Primate Brain-Size Evolution. *Systematic Biology* **72**, 404–418 (2023).
51. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).
52. J. Sueur, T. Aubin, C. Simonis, SEEWAVE, A FREE MODULAR TOOL FOR SOUND ANALYSIS AND SYNTHESIS. *Bioacoustics* **18**, 213–226 (2008).
53. A. Anikin, Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behav Res* **51**, 778–792 (2019).
54. Søren Højsgaard, Ulrich Halekoh, doBy: Groupwise Statistics, LSmeans, Linear Estimates, Utilities, (2006); https://doi.org/10.32614/CRAN.package.doBy.
55. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using **lme4**. *J. Stat. Soft.* **67** (2015).
56. Kamil Bartoń, MuMIn: Multi-Model Inference, (2010); https://doi.org/10.32614/CRAN.package.MuMIn.
57. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer, Switzerland, Second edition., 2016)*Use R!*

Light propofol anaesthesia for non-invasive auditory EEG recording in unrestrained non-human primates.

Authors: T. Piette[1], C. Lacaux[1], M. Scheltienne[2], V. Sterpenich[1], M.Isnardon[3], V.Moulin[4], A.Cermolacce[4] , D. Grandjean[5], A. Meguerditchian[4,6], E.C Déaux[1,7], A-L.Giraud[1,8].

[1] Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2] Fondation Campus Biotech Geneva, Geneva, Switzerland

[3] Uniivo, Paris, France.

[4] Station de Primatologie UARS846, CNRS, Rousset-sur-Arc, France

[5] Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

[6] Centre de Recherche en Psychologie et Neurosciences UMR7077, CNRS, Univ Aix-Marseille, Marseille, France

[7] ARGOS, Chevagnes, France

[8] Université Paris Cité, Institut Pasteur, AP-HP, Inserm, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France

Abstract

Non-invasive electroencephalographic (EEG) experiments have been instrumental in advancing our understanding of the brain mechanisms involved in the production and perception of sounds and human speech. Performing similar experiments in non-human primates (NHPs) would help further deepen our knowledge by allowing us to investigate the evolutionary roots of these processes. However, performing EEG on NHPs is a challenge, given its sensitivity to motion artefacts, device cost and durability, and animal training requirements. For these reasons, most attempts have used invasive intracranial recordings, which led us to develop an alternative that minimises stress and prioritises animal welfare. By using mild propofol sedation, neurophysiological experimentation can easily be integrated into the routine sanitary checks of captive animals and allows the optimisation of both EEG quality and animal welfare. To assess the influence of propofol on brain activity in NHPs, we sedated three olive baboons (*Papio anubis*), scored their sleep stages under different doses, and recorded auditory event-related potentials (ERP)in response to grunts. Analyses of the EEG recordings with regards to sleep stage and ERP components indicate that at low dose (< 0.1mg/kg/h), propofol induces a light sleep state conducive to recording stimulus-elicited auditory activity. Overall, this experiment confirms the use of propofol sedation as an appropriate technique to study auditory processes through unrestrained, non-invasive EEG in NHPs.

Introduction

Over the past decades, neuroimaging technologies have significantly advanced language research, providing critical insights into acoustic and speech processing. Among these technologies, electroencephalography (EEG) has played a central role in uncovering key mechanisms underlying human communication, such as the organisational role of brain oscillations, and has been instrumental in language disorder research and hearing aid development [1–3]. EEG can also be used to study the evolution of language and acoustic communication, in the context of comparative studies, when equipped on non-human primates (NHPs)[4–8]. However, it presents significant challenges; EEG is highly sensitive to motion artifacts[9], and the overall fragility of the recording equipment further complicates data acquisition in moving animals.

To circumvent these limitations, past studies have often relied on invasive intracranial recordings or used extensive training of a few individuals over several years [4,10,11]. These studies frequently rely on primate chairs, a contention set-up that places the animals in stressful working conditions[12]. While such set-ups minimise motion artefacts and prevent damage to the devices, they greatly limit the scope and generalizability of the experiments and raise serious ethical questions.

An alternative approach for experiments that do not explicitly require animal cooperation involves the use of light anaesthesia to inhibit movement in test subjects, enabling high-quality data acquisition. Importantly, this method eliminates the need for physical restraints, reducing stress and fear in the animals. Additionally, by removing the requirement for animal training and ensuring device protection, this approach could lower costs, allowing for larger sample sizes and ultimately contributing to more robust scientific results. Yet, the major drawback to using anaesthetic drugs is the risk of inducing deep sleep stages and impairing auditory processing. Indeed, common veterinary anaesthetic procedures include injecting intramuscular ketamine combined with medetomidine or inhaled sevoflurane. However, these drugs have been shown to reduce auditory evoked potential responses and diminish oddball mismatch negativity, making them suboptimal candidates for neuroacoustic studies [13–16].

However, while veterinary practices have favoured these anaesthetics for their reliability in various settings, other agents may be better suited for sensory processing experiments. Among them, propofol, an intravenous drug from the alkylphenol family, shows promising potential. Unlike ketamine and sevoflurane, which significantly inhibit early brainstem auditory response in humans, propofol, when administered at low dosage, has minimal impact on auditory response latency and negligible effects on response amplitude [17,18]. Furthermore, while propofol induces widespread slow oscillations (0.5-1H) and increased alpha oscillations (8-12Hz) in the frontal cortex[19,20], no direct effect on other oscillations their coupling in the auditory cortex has been reported. Information on propofol's effect on late auditory response, and especially ERP components N100 and P200 is however lacking. To date, only one study used propofol in a near-infrared spectroscopy experiment, and reported similarities in the processing of conspecific emotional vocalizations between humans and sedated baboons, highlighting propofol’s potential[21]. Importantly, propofol’s rapid anaesthetic response allows for fine-tuning anaesthesia at very low dosages[22], enabling the induction of light sleep stages (NREM1–NREM2), where auditory processing is maintained, while avoiding deeper stages (NREM3-REM) that would impair auditory processing[23,24].

In this study, we investigated the impact of propofol anaesthesia on auditory processing using non-invasive EEG recordings in olive baboons (*Papio anubis*). Initial pilot tests were conducted on two subjects anaesthetised with propofol at 0.2 mg/kg/h. These tests revealed conflicting results in sleep depth and auditory ERP responses between the two subjects, suggesting significant individual variability in the effects of propofol. To address these discrepancies, we performed a more detailed experiment on a third subject. Using continuous EEG monitoring, we assessed sleep levels under varying doses of propofol and compared auditory ERP responses during sevoflurane anaesthesia and the lowest tested dose of propofol (0.1 mg/kg/h). Our findings demonstrated that at low dosage, propofol (0.1 mg/kg/h) successfully induced a light sleep stage while preserving basic auditory processing capabilities, as evidenced by the presence of classical ERP components such as N100 and P200. These results confirm the potential of using low-dose propofol anaesthesia as a valuable tool for non-invasive brain recording of auditory processing in non-human primates.

Method

Subjects

The study included three healthy female baboons of respectively 12, 16, and 6 years old (Table 1). Male baboons were excluded due to their larger, thicker masticatory muscles over the temporal cortex, which could interfere with the collection of neural activity. The three individuals are part of the Primatology Station (Marseille, France). The baboons were born in captivity and housed in social groups in enclosures with access to outdoor and indoor areas. Their enclosures include climbing structures of wood and metal, along with substrate favouring natural foraging behaviours. Water is provided *ad libitum*, and the baboons receive daily feeds of monkey pellets, seeds, fresh fruits, and vegetables. For each of the three females, health assessments and daily behavioural monitoring by veterinary and animal welfare staff confirmed that they had normal hearing and no structural neurological impairments. This was further validated by T1-weighted anatomical brain imaging acquired under anaesthesia using a 3-Tesla MRI, performed as part of another ongoing research[25]. All procedures received approval from the "C2EA-71 Ethical Committee of Neurosciences" (INT Marseille) and complied with French law, CNRS guidelines, and European Union regulations.

Stimuli

ERP

Auditory stimuli for the ERP experiment consisted of a sequence featuring a 100 ms grunt from an olive baboon unknown from our subjects, repeated either 50 or 100 times. Each grunt was followed by a randomly timed silence interval lasting between 1 and 2 seconds.

Rhythm

Auditory stimuli for the rhythm experiment (used in another study) consisted of a recording made of 5-second long sequences of human speech, baboon grunts, and white noise, repeated either 50 or 100 times each. Each sequence was followed by a randomly timed silence interval lasting between 1 and 2 seconds.

General anaesthesia protocol

For all three subjects, we started with the general anaesthesia protocol routinely as part of general health assessments. First, subjects were separated from their social group and anaesthetised with an intramuscular injection of ketamine (4.5 mg/kg, Ketamine 1000) combined with medetomidine (45 µg/kg, Domitor). Each baboon was positioned in ventral decubitus with the head stabilised using foam supports, and a routine health check was performed. Before recordings, animals were intubated and sevoflurane (3–5%, Sevotek) was administered. Subjects were then prepared for the experiment, and sevoflurane was discontinued after installation of the EEG recording set-up. The following anaesthesia procedure depended on the subject, as described below.

Subject-specific protocol

Pilot tests (Fidji and Chet)

Following common veterinary guidelines, propofol (Propovet®) was administered intravenously at 0.2 mg/kg/h following sevoflurane termination (Table 1), throughout the rest of the experiment. Ten minutes after propofol administration, we conducted a 2-minute auditory ERP experiment and a 17-minute auditory rhythmic discrimination task (not included in this manuscript).

Follow-up experiment (Ozone)

Because results (see below) from the previous two subjects suggested possible subject-dependent dosage effects, we aimed to better characterise this phenomenon. Thus, for the third subject, propofol was initiated at the lower dose of 0.15 mg/kg/h immediately following sevoflurane termination. After maintaining this rate for 5 minutes, the dose was reduced to 0.10 mg/kg/h for approximately 1 hour, during which an auditory ERP experiment and an auditory rhythmic discrimination task were conducted. Following this period, the propofol infusion rate was increased to 0.20 mg/kg/h for 5 minutes and finally elevated to 0.25 mg/kg/h for the last 5 minutes (Table 1).

Table 1. Summary information for each female baboon. Name, age, weight, anaesthetic type, and dosage for each subject, associated with details (type and duration) of the experimental protocol.

Subject name	Age (years)	Weight (kg)	Anaesthetic	Dosage	Experiment type	Recording time (s)
Fidji	**12**	**16.75**	**Propofol**	**0.20 mg/kg/h**	**ERP**	**120**
Chet	**16**	**18.50**	**Propofol**	**0.20 mg/kg/h**	**ERP**	**120**
Ozone	**6**	**13.35**	**Sevoflurane**	**3%**	**ERP**	**240**
Ozone	**6**	**13.35**	**Propofol**	**0.15 mg/kg/h**	**Rest**	**400**
Ozone	**6**	**13.35**	**Propofol**	**0.10 mg/kg/h**	**ERP**	**240**
Ozone	**6**	**13.35**	**Propofol**	**0.10 mg/kg/h**	**Rhythm**	**1860**
Ozone	**6**	**13.35**	**Propofol**	**0.20 mg/kg/h**	**Rest**	**120**
Ozone	**6**	**13.35**	**Propofol**	**0.25 mg/kg/h**	**Rest**	**120**

EEG recording

All three subjects were similarly prepared for the EEG recording. First, the temporal areas were shaved. The scalps were then washed with soap and a scrubbing skin gel (Neurprep, Spes Medica), and then dried with paper towels. Eight gold cup electrodes were positioned on the scalp at locations Cz, Fz, Pz, Oz, C3, T3, C4, and T4, along with a reference electrode at FCz and a ground electrode on the nose (Supp fig. 1). The electrodes were secured using conductive paste (SAC2, Spes Medica) and medical tape. They were then connected to a G.Nautilus amplifier (g.tec medical engineering GmbH, Austria), positioned on the operating table beside the baboons, with data wirelessly transmitted to a receiver connected to a DELL recording laptop. Electrode impedance was maintained below 30 kΩ, and data were recorded at 500 Hz sampling rate. After securing the electrodes, a wired headphone (JVC HA-S500) connected to the stimulus computer was placed on the ears of the anaesthetised baboon. Ten minutes after the start of the propofol injection, the auditory stimuli (i.e., the ERP or rhythm tasks) were broadcast to the baboon through the positioned headset.

EEG data

All EEG preprocessing steps were done in MATLAB using the Fieldtrip toolbox and custom-written scripts. EEG data were bandpass filtered between 0.1 and 150 Hz (order =3) and a DFT filter was applied at 50, 100, and 150 Hz, as well as a third-order bandstop filter between 48-52Hz to remove background electrical noise. (Table 1).

ERP analysis

EEG data were bandpass filtered between 0.5 and 30Hz. Recordings were epoched from 0.5 s pre-stimulus onset to 1s post-stimulus onset. Artefact rejection was done through an amplitude-based rejection fieldtrip function, with a cut-off value at $\pm\ 100\, mV$. A final visual inspection of all trials was used to remove any remaining noisy trial that escaped the rejection procedure. In the case of Subject Ozone, the C3 electrode was shown to contain more than 80% of the artefacts and was removed from further analysis. For all subjects, EEG data were segmented and aligned to the onset of each event using the Fieldtrip Ft_timelockanalysis function[26], then averaged across trials and electrodes, to produce an event-related potential (ERP). To normalize the data and allow comparisons across individuals, a z-score transformation was applied.

Sleep scoring

EEG signals were band-pass filtered between 0.3 and 30 Hz and segmented into 30-second epochs. Sleep scoring was performed offline by two experienced scorers (V.S. and C.L.), who were blind to the experimental conditions. To better assess variations in sleep depth in response to different anaesthetics and doses, we used an adapted version of the former 4-stage Rechtschaffen and Kales (1968) classification (instead of the current 3-stage AASM classification)[27]. In our system, stage N3 was defined as the presence of delta waves exceeding 50% of the epoch, while stage N4 was identified by the predominance of large amplitude delta waves throughout the entire 30-second epoch. The interrater agreement was moderate (Cohen's κ coefficient = 0.44), and any discrepancies between the two scorers were resolved by jointly reviewing the data and reaching a consensus. Central tendency (mode) was then computed for each subject and condition to identify the dominant sleep stage under respective anaesthetic types and doses.

Software

All analyses and visualisation were done using Matlab version 2023a and R version 4.1.2 (2021-11-01) with the packages ggplot2[28], and tydiverse [29].

Results

We first conducted ERP recordings on two anaesthetised female baboons, aged 12 years (Fidji) and 16 years (Chet), while monitoring their sleep stages. The animals were maintained under propofol anaesthesia at 0.2 mg/kg/h, aiming to induce a light sleep stage. While our first subject (Fidji) displayed a light sleep stage, with a predominance of stage 1 sleep during the experiment (fig 1a, c), our second subject (Chet), although anaesthetised with the same dose of propofol exhibited a deeper sleep stage, with a predominance of stage 3 sleep (fig 1b,c). Interestingly, and in accordance with the sleep assessment, auditory ERP responses in Fidji show a mean minimum amplitude of -1.72 at 108 ms within the 80–120 ms post-stimulus window and a mean maximum amplitude of 1.44 at 186 ms within the 150–250 ms window (fig 1d). This indicates that the classical N100 and P200 components are present under propofol anaesthesia in this subject. In contrast, ERP responses in Chet exhibit a mean minimum amplitude of -1.207 at 80 ms within the 80–120 ms window and a mean maximum amplitude of -0.21 at 0.156 ms within the 150–250 ms window (fig 1e), indicating that the classical N100 and P200 components are absent in her case.

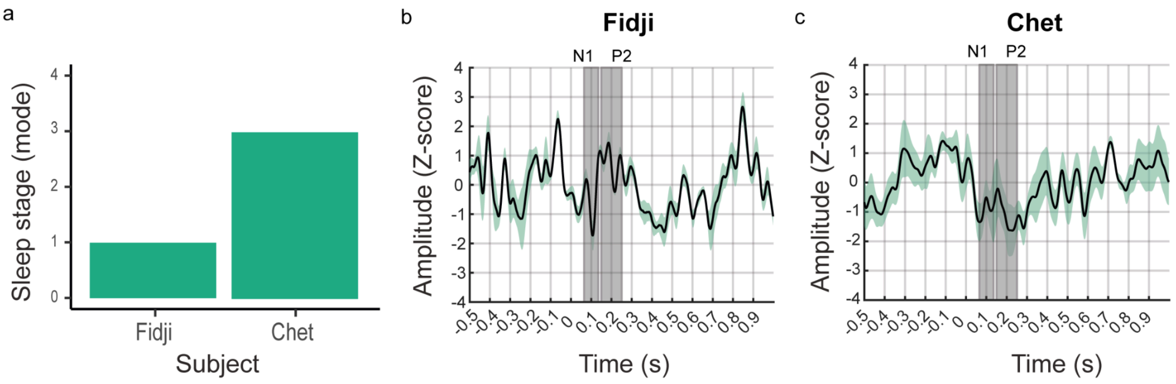


Figure 1. Impact of propofol infusion at 0.2mg/kg/h on vigilance states and auditory ERP (0.2 mg/kg/h). a) Central tendency sleep stage (mode) under Propofol anaesthesia in Fidji and Chet. b) Auditory ERP response (n=42) averaged across all electrodes recorded from the female baboon Fidji during propofol infusion, showing the presence of classical auditory components N100 and P200 c) Auditory ERP response (n=40) averaged across all electrodes recorded from the female baboon Chet during propofol infusion, showing the absence of classical auditory components N100 and P200.

To investigate the discrepancy in sleep and auditory responses between Fidji and Chet under propofol anaesthesia, we monitored sleep stages at varying propofol doses in a third 6-year-old female baboon (Ozone). Additionally, we recorded auditory ERP responses under sevoflurane and at a minimal propofol dose (0.10 mg/kg/h). At the start of the experiment, Ozone was under sevoflurane anaesthesia, exhibiting deep sleep stages (Central tendency = stage 4). Upon transitioning to propofol at 0.15 mg/kg/h, she remained in deep sleep stages (NREM stage 4). The first signs of sleep lightening appeared 6 minutes after the switch, becoming more pronounced 15 minutes post-transition, following propofol decrease at 0.10 mg/kg/h. During the auditory experiment, Ozone alternated between light and deep sleep stages, with light sleep predominating (75%). At the end of the experiment, following an increase in the propofol infusion rate, deep sleep stages reappeared 5 minutes after the adjustment, with a complete transition to deep sleep occurring 9 minutes later (t+14minutes;fig 2 a, 2b).

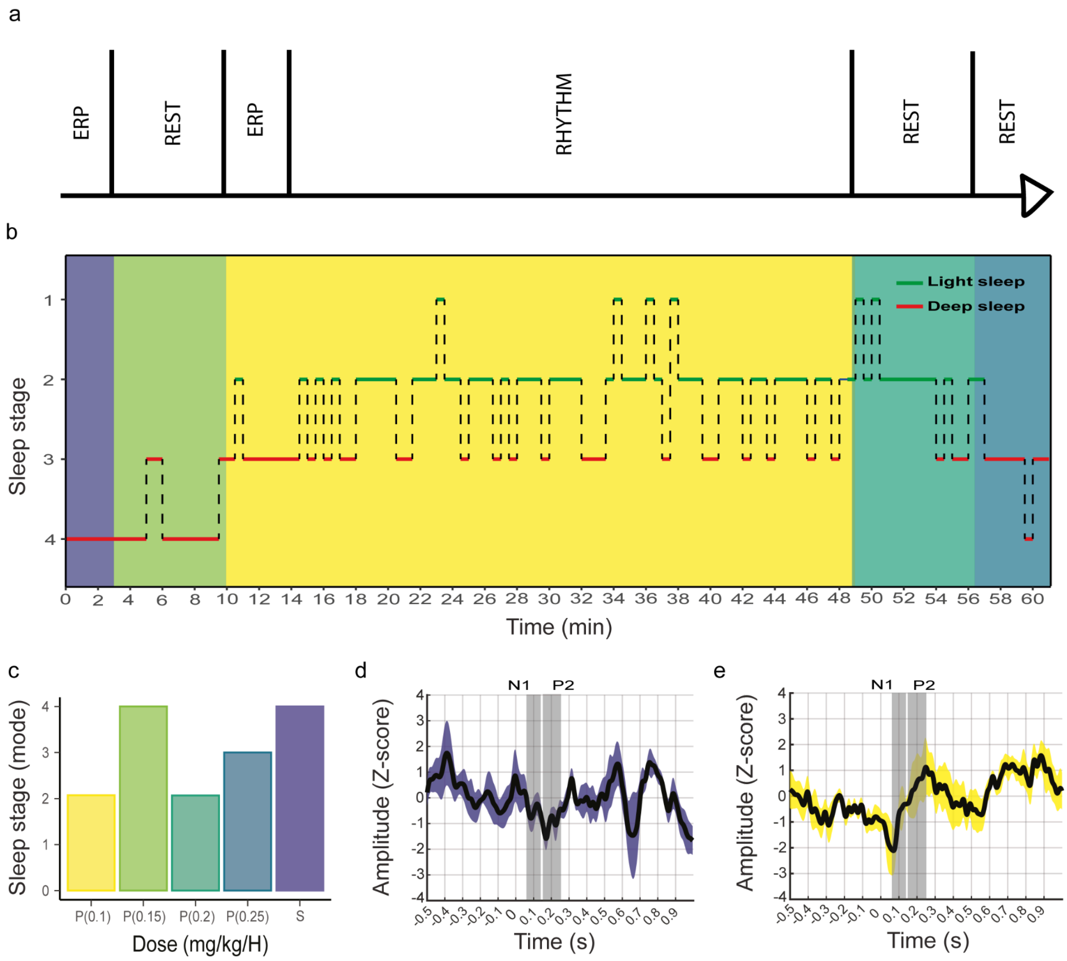


Figure 2: Impact of propofol infusion rate on vigilance states and auditory ERP a) Experiment outline. b) Hypnogram showing the evolution of sleep stages across time under different Sevoflurane and Propofol infusion rates. c) Central tendency sleep stage (mode) under Sevoflurane and Propofol at different infusion rates. d) Auditory ERP response (n=76) averaged across all electrodes recorded during Sevoflurane infusion, showing the absence of classical auditory components N100 and P200. e) Auditory ERP responses (n=76) averaged across all electrodes recorded during Propofol infusion at 0.1mg/kg/h, showing the presence of classical auditory components N100 and P200.

The ERP responses under Sevoflurane show a mean minimum amplitude of -0.80 at 86 ms within the 80–120 ms post-stimulus window and a mean maximum amplitude of -0.44 at 250 ms within the 150–250 ms window. These findings indicate that Sevoflurane anaesthesia abolished the classical N100 and P200 components. In contrast, ERP responses under light propofol anaesthesia exhibited a mean minimum amplitude of -1.87 at 80 ms within the 80–120 ms window and a mean maximum amplitude of 1.16 at 248 ms within the 150–250 ms window, confirming that light propofol anaesthesia preserves the classical N100 and P200 components.

Discussion

Performing EEG in non-human primates presents significant challenges due to motion artefacts, the fragility of recording equipment, and ethical considerations. Despite these obstacles, EEG remains crucial for advancing our understanding of brain mechanisms, particularly regarding the evolution of cross- and conspecific vocalization processing. In light of these challenges, we aimed to assess the feasibility of using propofol anaesthesia as an alternative approach to minimise motion artefacts and stress, while enabling high-quality, non-invasive EEG recordings. By examining the effects of propofol on sleep depth and auditory processing, we sought to determine its suitability for conducting passive EEG experiments in NHPs.

Discrepancies in sleep levels and auditory responses between Fidji and Chet, despite being anaesthetised with the same dose of propofol at an injection rate of 0.2 mg/kg/h, highlight significant individual variability in propofol response (fig 1). This variability is unlikely to be related to body weight, as the injection rate was adjusted accordingly. However, since propofol is lipophilic, body composition, a factor not typically measured in NHPs, could have influenced the anaesthetic response. Additionally, factors such as the animals' medical histories and the number of previous anaesthetic procedures they have undergone, which were not available to us, might also contribute to these differences. Consequently, such variability underscores the importance of starting with the lowest recommended dose of propofol anaesthesia, currently set at 0.1 mg/kg/h according to veterinary guidelines, to assess the subject's response and adjust the dosage as needed.

In the third subject Ozone, sleep levels under propofol anaesthesia exhibited a dose-dependent response, with deeper sleep observed at an infusion rate of 0.25 mg/kg/h compared to 0.1 mg/kg/h (fig. 2c). While propofol at 0.1 mg/kg/h predominantly triggers light sleep stages, an alternation between NREM2 and NREM3 still occurs. This suggests that further reducing the dose could potentially favour an alternation between lighter sleep stages, best suited for exploring auditory processing. Continuous monitoring of sleep levels throughout the experiment also revealed a residual effect of Sevoflurane anaesthesia. Sleep levels started to decrease 6 minutes after the start of propofol and stabilised around 15 minutes post-initiation, emphasising the need for a waiting period between anaesthesia transitions and the start of auditory experiments. Similarly, when the propofol infusion rate was increased from 0.1 mg/kg/h to 0.2 mg/kg/h, changes in sleep stages took about 5 minutes to manifest (fig. 2a, 2b).

Importantly, in our two subjects exhibiting predominantly light sleep stages during the auditory ERP protocol (Fidji and Ozone), brain signals revealed the presence of the classical auditory components N100 and P200, indicating that basic auditory processing is preserved under light propofol anesthesia (Fig. 1d, 2e). This finding highlights the feasibility of using light propofol anesthesia for non-invasive brain recordings in non-human animals, preserving essential neural functions while limiting the need for invasive methods.

Performed alongside routine veterinary care, light propofol anaesthesia for brain recording in non-human animals offers significant practical benefits from both ethical and experimental perspectives. The key advantages of this protocol are: 1) stress reduction compared with traditional restraint methods for both the experimenter and the animal subject, 2) the naturalness of the possible experimentations, and 3) the amount of data that can be acquired as more animals can be tested without harm. Given that propofol does not inhibit pulmonary function, it is also widely used in human medicine for anaesthesia without the need for artificial ventilation [30,31]. Therefore, an adapted propofol protocol could potentially be applied to wild, untamed animals, facilitating EEG recordings in natural and uncontrolled environments under veterinary supervision. While our protocol involved shaving the scalp to improve electrode contact, it is worth noting that similar studies in humans have achieved high-quality data with the same electrodes on unshaved subjects[32], suggesting the potential of minimising the invasiveness of the procedure in future experiments. Alternative methods, such as the use of spider electrodes or other innovative electrode types, could also be explored to minimise the procedure's invasiveness and enhance its applicability in field settings.

Overall, our findings validate the use of light propofol anaesthesia for auditory-related tasks in non-human animals, while highlighting key considerations for its use. Starting with a low dose of 0.1 mg/kg/h or even lower, and adjusting the dose based on individual responses is essential to preserve basic auditory processing in NHP's brain. A waiting period of 15 minutes should be observed between anaesthesia initiation and the start of the experiment to allow sleep to stabilise. Any increase in dose or adjustment in infusion rate should be closely monitored, as changes in sleep stages may take several minutes to appear, significantly affecting auditory processing. This approach not only provides an ethical and practical solution for conducting brain recordings but also opens the door for future research that could explore more complex auditory processing, such as mismatch negativity (MMN) in an auditory oddball paradigm, local-global paradigms[33], as well as long natural processing experiments. This would help assess whether higher-order auditory functions are also preserved under light propofol anaesthesia, paving the way for more advanced investigations into neural processes in both captive and wild non-human animals.

Ethical Statement

All animal procedures were approved by the "C2EA–71 Ethical Committee of neurosciences" (INT Marseille) under the number APAFIS##46189-2023111410456237 v5, and have been conducted at the Station de Primatologie under the number agreement G130877 for conducting experiments on vertebrate animals (Rousset-Sur-Arc, France). All methods were performed in accordance with the relevant French law, CNRS guidelines and the European Union regulations (Directive 2010/63/EU).

Acknowledgements

The NCCR Evolving Language, Swiss National Science Foundation Agreement Nr. #51NF40_180888, funded this work. We are very grateful to the Station de Primatologie CNRS, particularly the animal care staff and technicians, for their critical involvement in this project. A.M has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme grant agreement No 716931—GESTIMAGE—ERC-2016-STG, as well as French National Research Agency ANR-23-CE28-0029-01 (BABONTO), ANR-16-CONV-0002 (ILCB).

Contributions

T.P, E.D, D.G, A.M and A-L.G conceptualised the project. T.P, M.S, M.I, V.M adapted and refined the anaesthetic protocol. M,I and V.M performed animal handling and anaesthesia. T.P, M.S, A.M and D.G conducted EEG data collection. T.P performed EEG data analysis and visualisation. C.L and V.S performed sleep scoring. T.P wrote the initial draft. All authors contributed to the review and editing of the manuscript. E.D and A-L.G provided supervision. A-L.G provided resources and funding for the project.

Competing Interest

All authors declare that they have no conflicts of interest.

References

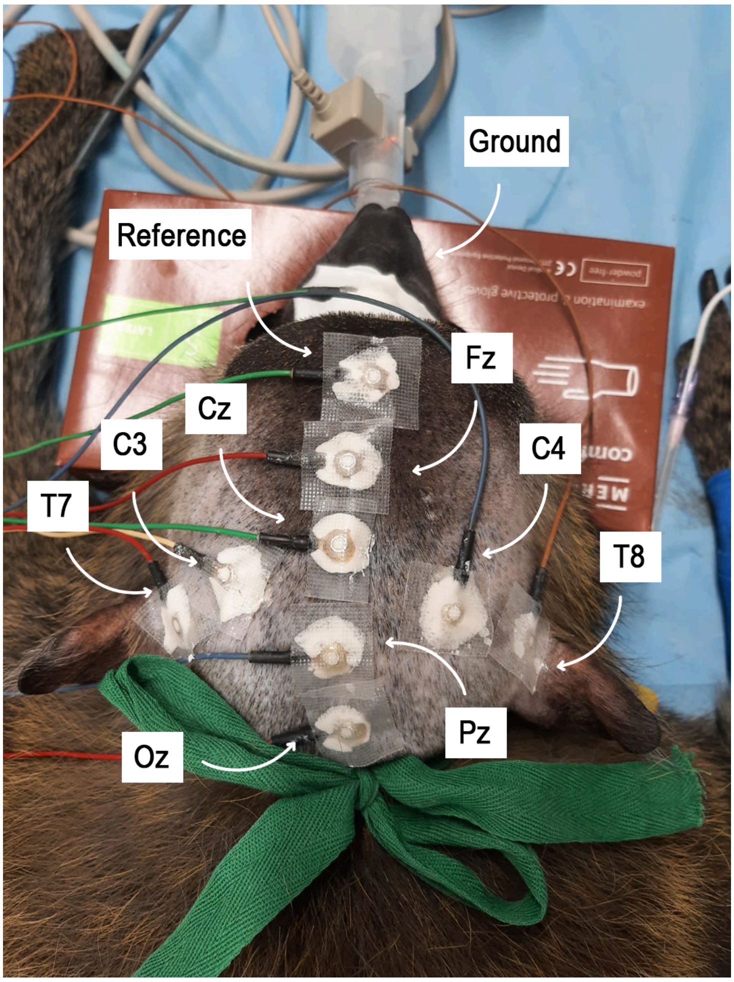
1. Giraud, A.-L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**, 511–517 (2012).
2. Ghitza, O. & Greenberg, S. On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica* **66**, 113–126 (2009).
3. Meyer, L. The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur J of Neuroscience* **48**, 2609–2621 (2018).
4. Lakatos, P. *et al.* An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. *Journal of Neurophysiology* **94**, 1904–1911 (2005).
5. Risueno-Segovia, C. & Hage, S. R. Theta Synchronization of Phonatory and Articulatory Systems in Marmoset Monkey Vocal Production. *Current Biology* **30**, 4276-4283.e3 (2020).
6. Ravignani, A. *et al.* Rhythm in speech and animal vocalizations: a cross-species perspective. *Ann. N.Y. Acad. Sci.* **1453**, 79–98 (2019).
7. Briefer, E. F. Vocal expression of emotions in mammals: mechanisms of production and evidence. *J Zool* **288**, 1–20 (2012).
8. Boesch, C. & Crockford, C. Call combinations in wild chimpanzees. *Behav* **142**, 397–421 (2005).
9. Tandle, A., Jog, N., D'cunha, P. & Chheta, M. Classification of Artefacts in EEG Signal Recordings and EOG Artefact Removal using EOG Subtraction. *CAE* **4**, 12–19 (2016).
10. Criscuolo, A. *et al.* Macaque monkeys and humans sample temporal regularities in the acoustic environment. *Progress in Neurobiology* **229**, 102502 (2023).
11. Nakamura, T. *et al.* Non-invasive electroencephalographical (EEG) recording system in

awake monkeys. *Heliyon* **6**, e04043 (2020).

12. McMillan, J. L., Bloomsmith, M. A. & Prescott, M. J. An International Survey of Approaches to Chair Restraint of Nonhuman Primates. *Comp Med* **67**, 442–451 (2017).
13. Kim, J.-W., Kim, P.-J., Kim, H.-G. & Shin, T. J. Changes of auditory stimulus processing in sevoflurane-induced sedation. *Neuroscience Letters* **698**, 121–125 (2019).
14. Felix, R. A., Kadner, A. & Berrebi, A. S. Effects of ketamine on response properties of neurons in the superior paraolivary nucleus of the mouse. *Neuroscience* **201**, 307–319 (2012).
15. Tatsumi, K., Hirai, K., Furuya, H. & Okuda, T. Effects of sevoflurane on the middle latency auditory evoked response and the electroencephalographic power spectrum. *Anesth Analg* **80**, 940–943 (1995).
16. Rosch, R. E., Auksztulewicz, R., Leung, P. D., Friston, K. J. & Baldeweg, T. Selective Prefrontal Disinhibition in a Roving Auditory Oddball Paradigm Under N-Methyl-D-Aspartate Receptor Blockade. *Biol Psychiatry Cogn Neurosci Neuroimaging* **4**, 140–150 (2019).
17. Chassard, D. *et al.* AUDITORY EVOKED POTENTIALS DURING PROPOFOL ANAESTHESIA IN MAN. *British Journal of Anaesthesia* **62**, 522–526 (1989).
18. Heinke, W. *et al.* Sequential Effects of Increasing Propofol Sedation on Frontal and Temporal Cortices as Indexed by Auditory Event-related Potentials. *Anesthesiology* **100**, 617–625 (2004).
19. Feshchenko, V. A., Veselis, R. A. & Reinsel, R. A. Propofol-Induced Alpha Rhythm. *Neuropsychobiology* **50**, 257–266 (2004).
20. Weiner, V. S. *et al.* Propofol disrupts alpha dynamics in functionally distinct thalamocortical networks during loss of consciousness. *Proc Natl Acad Sci U S A* **120**, e2207831120 (2023).
21. Debracque, C., Gruber, T., Lacoste, R., Meguerditchian, A. & Grandjean, D. Cerebral Activity in Female Baboons (Papio anubis) During the Perception of Conspecific and Heterospecific Agonistic Vocalizations: a Functional Near Infrared Spectroscopy Study. *Affec Sci* **3**, 783–791 (2022).
22. Skues, M. A. & Prys-Roberts, C. The pharmacology of propofol. *Journal of Clinical Anesthesia* **1**, 387–400 (1989).
23. Karaka, S., Bekçi, B., Çakmak, E. D., Erzengin, Ö. U. & Aydin, H. Information processing in sleep based on event-related activities of the brain. *Sleep Biol Rhythms* **5**, 28–39 (2007).
24. Jourde, H. R. & Coffey, E. B. J. Auditory processing up to cortex is maintained during sleep spindles. *PNAS Nexus* **3**, pgae479 (2024).
25. Marie, D. *et al.* Left Brain Asymmetry of the Planum Temporale in a Nonhominid Primate: Redefining the Origin of Brain Specialization for Language. *Cerebral Cortex* **28**, 1808–1815 (2018).
26. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* **2011**, 1–9 (2011).
27. Wolpert, E. A. A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. *Arch Gen Psychiatry* **20**, 246 (1969).
28. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer, Switzerland, 2016). doi:10.1007/978-3-319-24277-4.
29. Wickham, H. *et al.* Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).
30. Van Keer, L., Van Aken, H., Vandermeersch, E., Vermaut, G. & Lerut, T. Propofol does not inhibit hypoxic pulmonary vasoconstriction in humans. *Journal of Clinical Anesthesia* **1**, 284–288 (1989).
31. Salukhe, T. V. *et al.* Propofol sedation administered by cardiologists without assisted ventilation for long cardiac interventions: an assessment of 1000 consecutive patients undergoing atrial fibrillation ablation. *Europace* **14**, 325–330 (2012).
32. Moyne, M. *et al.* Brain reactivity to emotion persists in NREM sleep and is associated

with individual dream recall. *Cerebral Cortex Communications* **3**, tgac003 (2022).
33. Bekinschtein, T. A. *et al.* Neural signature of the conscious processing of auditory regularities. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1672–1677 (2009).

Supplementary Materials



Supplementary Figure 1 : Electrode placement. Placement of electrodes for EEG recording on Ozone according to the 10-20 EEG layout.

Hierarchical coupling of slow and low gamma oscillations for conspecific vocal processing in olive baboons (*Papio Anubis*).

T. Piette[1] ,D. Grandjean[2] , A. Meguerditchian[3,4], E.C Déaux[1,5] , A-L.Giraud[1 ,6].

[1] Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[2] Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

[3] Laboratoire de Psychologie Cognitive UMR7290, CNRS, Univ Aix-Marseille, Marseille,

[4] Station de Primatologie UPS846, CNRS, Rousset-sur-Arc, France

[5] ARGOS, Chevagnes, France

[6] Université Paris Cité, Institut Pasteur, AP-HP, Inserm, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France

Abstract

Brain oscillations play a fundamental role in auditory processing, particularly in speech perception in humans. The tracking of prosodic and syllabic speech rates by low-frequency oscillations, along with hierarchical coupling between theta and gamma bands, is essential for speech comprehension. However, the role of these oscillatory mechanisms in processing rhythm and vocal signals in non-human animals remains poorly understood. In this study, we employed a validated anesthetic protocol to conduct a passive listening experiment with three female olive baboons (Papio anubis), aiming to explore their neural responses to biologically relevant and irrelevant auditory stimuli. Scalp EEG recordings were obtained while the baboons were exposed to sequences of conspecific vocalizations (grunts), rhythmic noise, and human speech sentences. Our results demonstrate that delta (1-4 Hz) and theta (4-8 Hz) oscillations track conspecific vocalizations and rhythmic noise sequences, accompanied by phase-amplitude coupling (PAC) between theta and low gamma oscillations (25-45Hz). During speech listening, theta oscillations tracked syllabic rates, and PAC between theta and low gamma was observed during exposure to familiar words. This ability to

track higher rhythms in the theta range contrasts with previous studies in other mammals, which typically exhibit a two-timescale model of auditory processing. These findings suggest an evolutionary shift in auditory processing in primates, leading to the development of a third, faster processing scale within the theta range. Such an advancement likely facilitated more efficient information transfer and may have played a key role in the emergence of human language.

Methods

Subjects

The study included three healthy female baboons of respectively 13, 13 and 6 years old (Table 1). Male baboons were excluded due to their larger, thicker masticatory muscles over the temporal cortex, which could interfere with the collection of neural activity data. The three individuals are part of the Primatology Station (Marseille, France). The baboons were born in captivity and housed in social groups in enclosures with access to outdoor and indoor areas. Their enclosures include climbing structures of wood and metal, along with substrate favoring natural foraging behaviors. Water is provided *ad libitum*, and the baboons receive daily feeds of monkey pellets, seeds, fresh fruits, and vegetables. For each of the three females, health assessments and daily behavioral monitoring by veterinary and animal welfare staff confirmed that they had normal hearing and no structural neurological impairments. This was further validated by T1-weighted anatomical brain imaging acquired under anesthesia using a 3-Tesla MRI, performed as part of another ongoing research. All procedures received approval from the "C2EA-71 Ethical Committee of Neurosciences" (INT Marseille) and complied with French law, CNRS guidelines, and European Union regulations.

Table 1 Summary information for each female baboon. Name, age, weight, anesthetic type and dosage for each subject.

Subject name	Age (years)	Weight (kg)	Anesthetic	Dosage (mg/kg/h)
Feline	**13**	**17.10**	**Propofol**	**0.10**
Formule	**13**	**16.65**	**Propofol**	**0.10**
Ozone	**6**	**13.35**	**Propofol**	**0.10**

General experimental protocol

Each subject went through the same experimental protocol, starting with an anesthesia, health check, eeg set-up, oddball task to assess the depth of acoustic processing under anesthesia, passive listening acoustic paradigm and then waking up (Fig 1). Below, we describe each aspect of the session in more detail. The sessions lasted on average 3 hours.

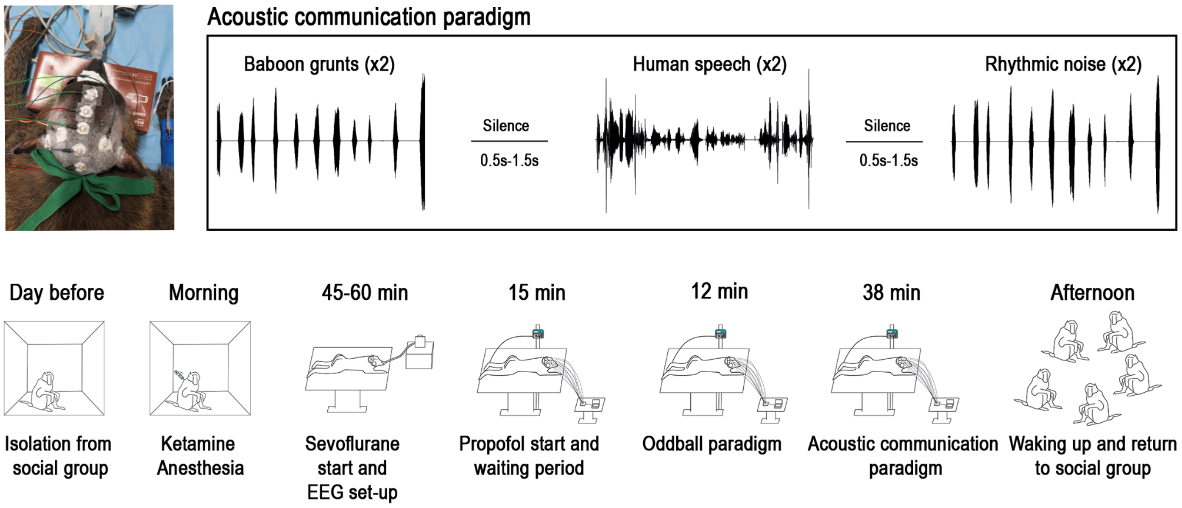


Figure 1. Overview of the experimental procedure, including electrode placement, auditory stimuli, and timeline. (Left) Example of electrode placement on a female subject under anesthesia. (Right) Acoustic communication paradigm consisting of three types of auditory stimuli: baboon grunts, human speech, and

rhythmic noise, each presented in pairs with interleaved silent intervals (0.5s–1.5s). (Bottom) Experimental timeline. The day before, subjects were isolated from their social group. On the day of the experiment, ketamine anesthesia was induced, followed by sevoflurane administration and EEG electrode placement (45–60 min). After a 15-min stabilization period with propofol at 0.1mg/kg/h, two auditory paradigms were presented: an oddball paradigm (12 min) and the acoustic communication paradigm (38 min). In the afternoon, subjects recovered and were returned to their social group.

Auditory stimulus preparation

Oddball

To validate the presence of basic auditory processing capacities under anesthesia, we used an oddball task. The auditory stimuli consisted of a sequence of standard tones (1000 Hz) presented 300 times (80%), including an initial block of 20 repetitions. Odd tones (500 Hz) were presented 75 times (20%), ensuring no consecutive repetitions. Each tone was separated by a jittered silent interval ranging from 1 to 2 seconds, resulting in a 12 minutes long recording.

Baboon grunts

Sequences of wild olive baboon grunts were sourced from the public acoustic database of the Berlin Museum für Naturkunde, ensuring that the vocalizing individuals were not known to our tested animals.

Rhythmic white noise

To obtain an irrelevant control condition, we engineered rhythmic white noise sequences from the baboon grunt sequences using Praat. The amplitude envelopes of each grunt sequence were extracted using the "To Intensity" function. White noise sequences of the same length as the baboon grunts were then generated, and both signals were merged using the "Multiply" function, resulting in white noise sequences that preserved the rhythmic structure of the baboon grunts.

Human speech

Both recordings featured French sentences, the language to which the baboons were regularly exposed. The first recording, spoken by a native French-speaking male, contained familiar words for the baboons: “Des graines dedans et si c’est bien mixé on pourrait le boire comme un smoothie, comme ça, ça ferait plein de fruits.” The second recording, extracted from a radio interview with a female French speaker, included vocabulary that was not contextually relevant to the subjects: “De mettre un genou à terre, comme le souhaiterait manifestement le ministre de l’Intérieur.”

Acoustic communication paradigm

The auditory track for the Acoustic communication experiment consisted of a recording comprising six sequences in total, two of olive baboon grunts, two of rhythmic white noise and two human speech sentences. All the sequences were first independently normalized at -1dB, then repeated 50 times (i.e. 300 sequences in total) and randomly ordered. A randomly timed silent interval of 1 to 2 seconds followed each sequence, resulting in a 34 minutes long recording.

Rhythm computation

Rhythms of each sequence were calculated following the protocol described in Piette et al, 2024[1], through wavelet analysis on the sequence amplitude envelope.

EEG recording

All three subjects were similarly prepared for the EEG recording. First, the temporal areas were shaved. The scalps were then washed with soap and a scrubbing skin gel (Neurprep, Spes Medica), then dried with paper towels. Eight gold cup electrodes were positioned on the scalp at locations Cz, Fz, Pz, Oz, C3, T3, C4, and T4, along with a reference electrode at FCz and a ground electrode on the nose (Fig 1). The electrodes were secured using conductive paste (SAC2, Spes Medica) and medical tape. They were then connected to a BrainVision.Vamp recorder (Brain Products GmbH, Germany), positioned on the operating table beside the baboons, with data wirelessly transmitted to a receiver connected to a DELL recording laptop. Electrode impedance was maintained below 30 kΩ, and data were recorded at 500 Hz sampling rate. After securing the electrodes, a wired headphone (JVC HA-S500) connected to the stimulus computer was placed on the ears of the anesthetized baboon. As suggested in Piette et al 2025, the oddball task started 15 minutes after the beginning of the propofol injection, then followed by the rhythm experiment.

EEG data processing

All EEG preprocessing was conducted in MATLAB using the FieldTrip toolbox[2] and custom scripts. The EEG data were bandpass filtered between 1 and 45 Hz, with a DFT filter applied at 50, 100, and 150 Hz. Given that the experiments took place in an open environment subject to electrical noise, a third-order bandpass filter was also applied between 48 and 52 Hz.

Oddball

Pre-processed EEG data were bandpass filtered between 1 and 30 Hz. Recordings were then epoched from 0.5 s before stimulus onset to 1 s after stimulus onset. Artifact rejection was performed using an amplitude-based rejection function from FieldTrip, with a cutoff value of 75 µV. A final visual inspection of all trials was conducted to remove any trial that had escaped the rejection procedure. EEG data were then aligned to the onset of each event using the FieldTrip ft_timelockanalysis function (100). The odd and standard tones were separated, and the resulting data were averaged across trials and electrodes to produce two event-related potentials (ERPs) for each stimulus. Finally, data were z_scored to allow comparison and averaged across subjects for each electrode.

Acoustic communication paradigm

Pre-processed EEG data were epoched from -0.5 s before stimulus onset to the end of the stimulus. Artifact rejection was performed using an amplitude-based rejection function from FieldTrip, with a cutoff value of 75 µV. A final visual inspection of all trials was conducted to remove any trial that had escaped the rejection procedure. EEG data were aligned to the onset of each event using the FieldTrip ft_timelockanalysis function, and then demeaned and baseline-corrected with a 500ms pre-stimulus window (-0.5 to 0 s) using the ft_preprocessing function. As basic auditory processing mechanisms are located in both left and right temporal area, signals from temporal electrodes T7 and T8 were then retained for further analysis.

Time frequency analysis

Time-frequency analysis was performed through wavelet decomposition. Spectral power was then computed using ft_freqanalysis, covering frequencies from 1 to 45 Hz in 0.5 Hz steps. The time window for analysis ranged from -0.5 to 4 s onset stimulus to normalise stimulus length across conditions, with a temporal resolution of 10 ms. Baseline correction was applied using ft_freqbaseline and a 500ms pre-stimulus interval (-0.5 to 0 s) with a relative baseline method. Time-frequency representations were then computed across participants using ft_freqgrandaverage, separately for each condition. Finally, spectral power values were averaged first across conditions, electrodes, and finally individuals.

Cerebro-acoustic coherence (CaCoh)

To assess cortical phase-locking to the different stimulus temporal structures, we employed the cerebro-acoustic coherence index. For each stimulus, we first computed the cross-spectral density between the neural signal and the stimulus envelope using a wavelet method, with frequencies ranging from 1 to 20 Hz in 0.1 Hz steps and time steps of 0.01 ms, from 0.6 s post-stimulus onset to 4 s. Next, we used the ft_connectivityanalysis function to compute the phase coherence between the stimulus envelope and the neural signal. Coherence values were then averaged across conditions, electrodes, and finally individuals. To evaluate how the brain of baboons tracked the acoustic signals, we compared the actual coherence to random coherence values obtained from the pairings of neural data with randomised acoustic envelopes averaged over 100 runs.

Phase-amplitude coupling (PAC)

As phase-amplitude coupling has been shown to be highly left-lateralized, signal from the T7 electrode was selected for PAC analysis. Phase frequencies were set between 2 and 8 Hz with a 0.5 Hz step size, while amplitude frequencies ranged from 20 to 45 Hz, also in 0.5 Hz steps. For each phase frequency, the signal was filtered within a ±1 Hz range, and for each amplitude frequency, the signal was filtered within a ±5 Hz range using a Butterworth bandpass filter. The modulation index (MI) was then computed to quantify the strength of phase-amplitude coupling (PAC) across trials. To generate a null distribution, surrogate data were created by randomly shifting the phase and amplitude signals relative to each other, and the MI was recalculated for each surrogate. Z-scores were obtained by comparing the observed MI values to this null distribution. Z_score were then averaged across conditions and individuals. To identify significant clusters of high Z-score coherence values, we performed a cluster analysis on data points exceeding a Z-score threshold of 1.96. A data point was included in a cluster if its Z-score was above 1.96, and adjacent points with Z-scores above 1.8 were also incorporated to account for the fine-grained frequency resolution of our analysis (0.5 Hz steps). This approach ensured the inclusion of contiguous high-coherence regions while minimizing the impact of isolated data points near the significance threshold.

Software

All analyses and visualization were done using Matlab version 2023a and R version 4.1.2 (2021-11-01) with the packages ggplot2[3], and tydiverse[4].

Results

Sequence rhythm

We first analyzed the rhythmic structure of each presented sequence. Notably, the first baboon grunt sequence exhibited a consistent delta rhythm at 3.6 Hz, whereas the second grunt sequence displayed a more complex rhythmic organization, characterized by a general delta rhythm at 2.1 Hz interspersed with localized segments exhibiting a faster theta-range rhythm at 4.85 Hz. As expected, both noise sequences exhibited corresponding rhythmic patterns. Additionally, the first human sentence, which contained familiar words, displayed a syllabic rate of 5.38 Hz, while the second sentence, composed of unfamiliar words, exhibited a slower syllabic rate of 4.22 Hz.

Oddball Paradigm

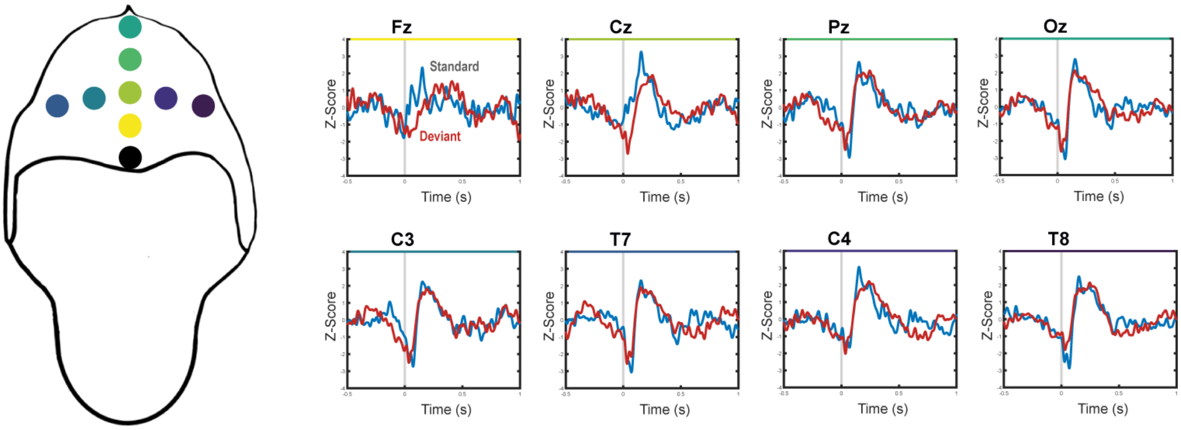


Figure 2: Oddball Paradigm (a) Schematic representation of electrode placement on the baboon skull. (b) Average brain activity of Feline and Formule in response to standard (Blue line: 1000 Hz, n = 300) and deviant (Red line: 500 Hz, n = 75) sounds at each electrode.

First, we conducted a classical auditory oddball paradigm to evaluate the effects of anesthesia on auditory processing and verify subjects' hearing capabilities. Notably, we observed robust N100 and P200 components in both Standard and Deviant conditions across electrodes Oz, Pz, C3, T7, and T8. However, no mismatch negativity (MMN) response was detected in any subject or electrode (Fig 2). These findings suggest that while our anesthetic protocol preserves basic sensory encoding and early cortical processing, it may impair higher-order auditory functions, which are crucial for more complex perceptual and cognitive processes. As basic auditory processing occurs in the temporal cortex, we selected electrodes T7 and T8 for further analysis.

Acoustic communication paradigm

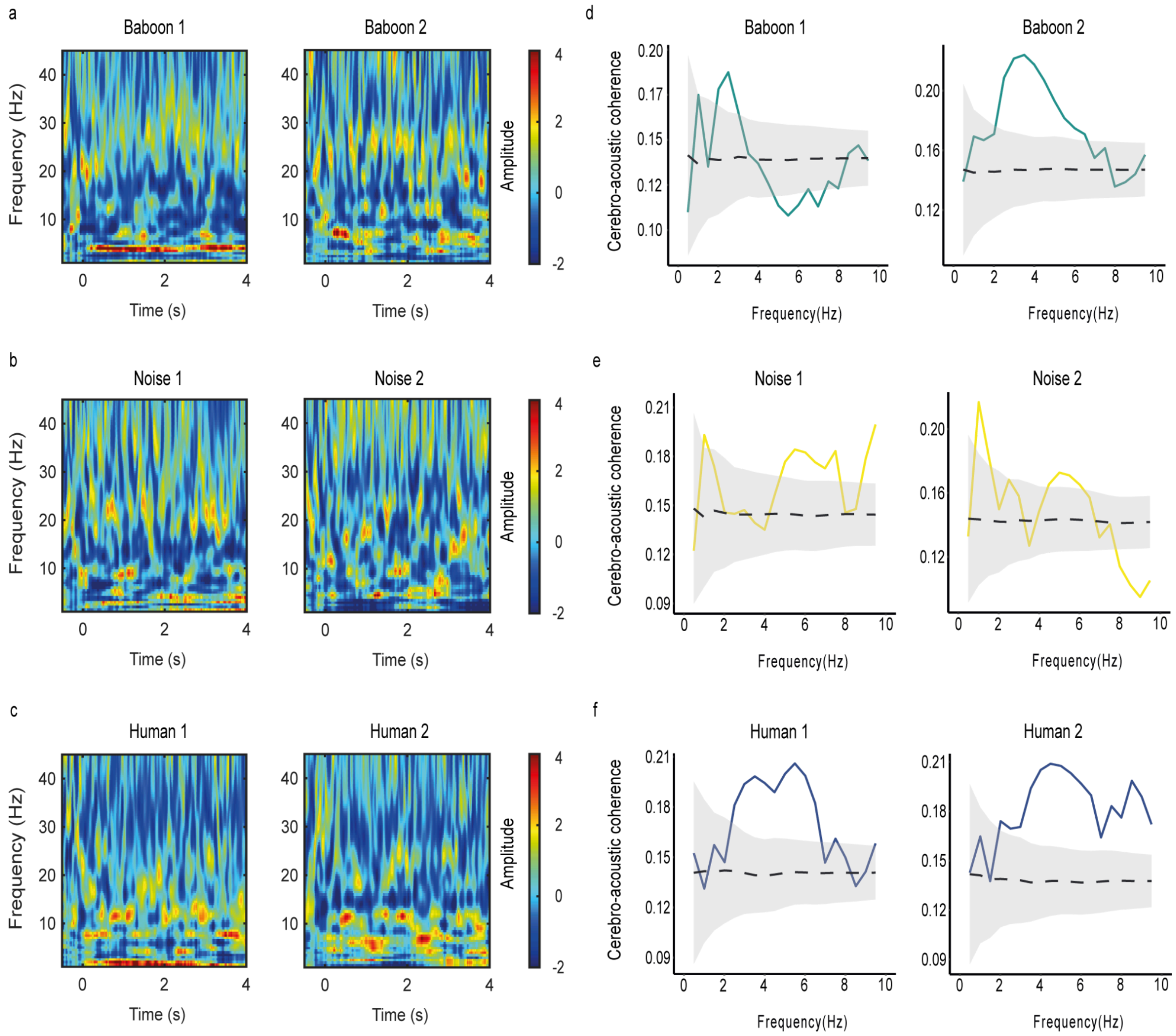


Figure 3: Brain activity and coherence in response to auditory stimuli (a,b,c) Time-frequency representation of the averaged brain activity electrodes individuals and electrodes T7/T8 while listening to baboon grunts, human speech, and white noise sequences respectively, (d,e,f) Averaged cerebro-acoustic coherence over the 0–10 Hz range across individuals and electrodes T7/T8 when exposed to baboon grunts, human speech and artificially created noise sequences respectively. The black dashed line and shaded area represent the mean (± SD) of permuted coherence.

Across subjects, neural activity exhibited a general increase in power relative to the pre-stimulus baseline, predominantly within the low-frequency range (<10 Hz; Fig. 3). However, notable differences emerged across conditions. In response to baboon vocalizations, neural activity varied between the two sequences. For the first sequence, power increased primarily in the delta range (~4 Hz), aligning with the dominant rhythmic component of the stimulus. In contrast, the second sequence elicited increased activity in the theta range (~6.5 Hz), closer to the rhythm of the localized segment within the sequence (Fig 3a). A similar pattern was observed for noise sequences. Neural activity increased in the delta range (~3 Hz) for the sequence derived from the first baboon vocalization, whereas the sequence derived from the second vocalization elicited increased activity in the theta range (~5Hz, Fig 3b). When exposed to human speech, neural activity exhibited a more complex pattern with increased power for both sequences at three distinct frequencies: delta (1.5 Hz), theta (~7.5 Hz), and alpha (~11.5 Hz) (Fig 3c).

When examining higher-frequency responses, distinct patterns emerged across conditions. Notably, a significant power increase at 26.5 Hz in the low-gamma range, specific to baboon grunts, was observed in both sequences but was absent in the other conditions (Fig 3a,b,c).

Given the presence of a stimulus-related and sustained neural response, we examined stimulus-specific neural tracking across conditions. When exposed to the first sequence of baboon grunts, cerebro-acoustic coherence was above the random mean within the delta range (2–3 Hz, Fig 3d), whereas the second sequence elicited increased coherence within the delta-theta range (2–6 Hz, Fig 3d). Exposure to artificially generated noise resulted in coherence exceeding the random mean exclusively within the theta range (4-8 Hz, Fig 3e) for the first sequences while the second sequence triggered significant coherence in both delta (1-1.5HzFig 3e) and theta range (4.5-6Hz, Fig 3e). When listening to human speech, coherence was above the random mean in the delta-theta range (2–6 Hz, Fig 3f) for both sequences.

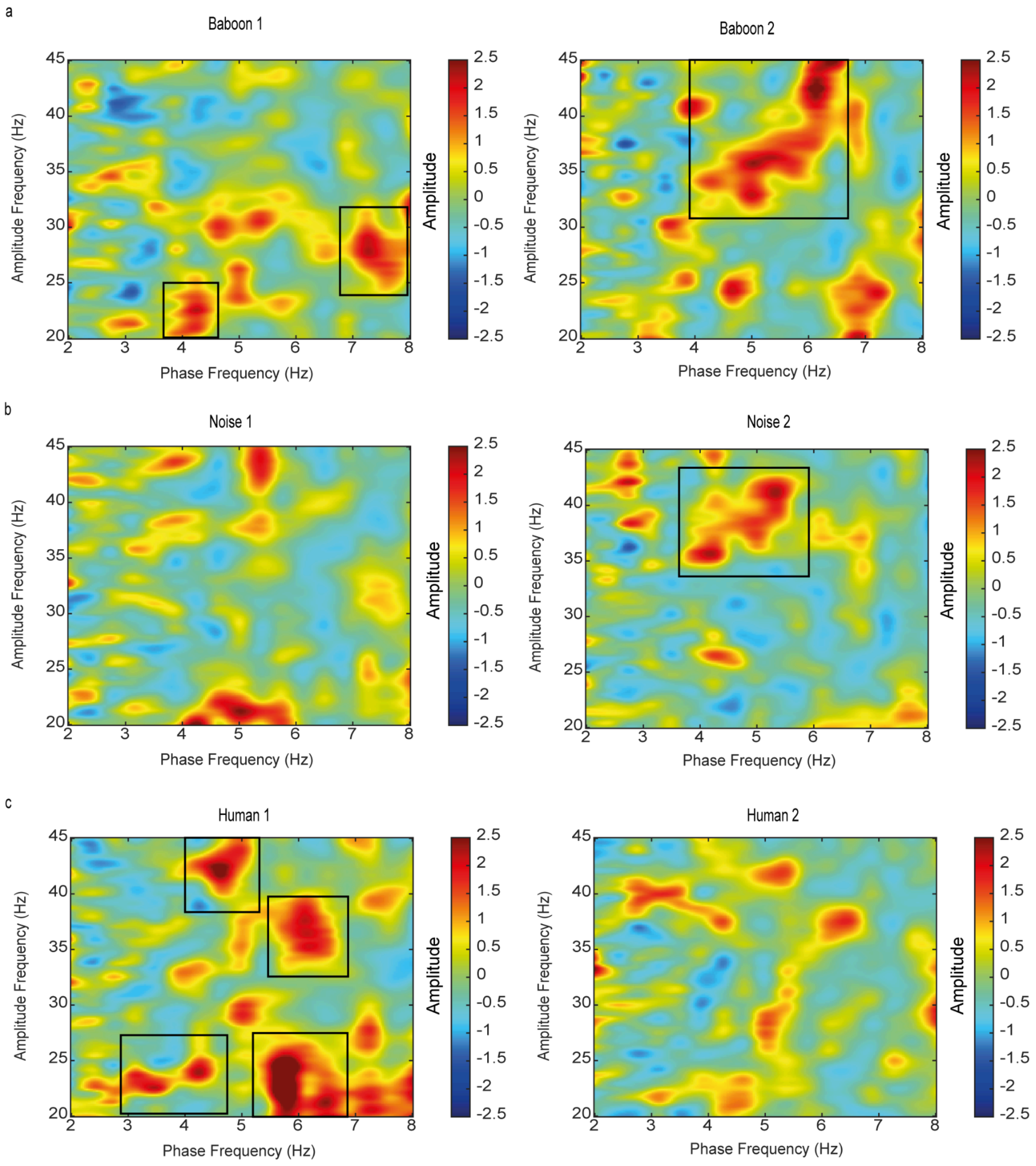


Figure 4: Phase-Amplitude Coupling in Response to Auditory Stimuli. (a, b, c) Average PAC for electrode T7 across individuals, showing coupling between slow (0–8 Hz) and fast (20–45 Hz) oscillations during exposure to baboon grunts, artificially generated noise sequences, and human speech, respectively. Black boxes highlight significant PAC clusters containing z-scores above 1.96.

We then investigated phase-amplitude coupling (PAC) between slow (1–10 Hz) and fast (15–45 Hz) oscillations.When exposed to the first sequence of baboon grunts, we identified two significant PAC clusters between theta (4–5 Hz and 7-8Hz respectively) and low-gamma (20-25 Hz and 25-32Hz respectively) oscillations. Similarly exposure to the second baboon grunt sequence primarily elicited significant PAC between theta (4–7 Hz) and low-gamma (32–45 Hz) oscillations. A similar pattern emerged during exposure to artificially generated noise sequences. The sequence derived from the first baboon vocalization did not exhibit significant PAC in our frequency range of interest , whereas the sequence derived from the second baboon vocalization showed significant PAC between theta (4-6 Hz) and low-gamma (34–43 Hz) oscillations. When listening to human speech, PAC differed between the two sequences. The sequence containing familiar words elicited significant PAC between delta (3–4,5 Hz) and low-gamma (20–25 Hz) oscillations, as well as between theta (5.5–8 Hz) and low-gamma (20–25 Hz and 33-45Hz) oscillations. However, no significant PAC was observed when participants were exposed to the unfamiliar speech sequence.

Discussion

By employing an auditory oddball paradigm, we confirmed that our anesthetic protocol preserves basic auditory processing, as demonstrated by the presence of robust N100 and P200 components (Fig 2). However, the absence of a mismatch negativity (MMN) component across all subjects suggests an impairment in higher-order auditory discrimination. This limitation should be considered when interpreting subsequent findings, particularly regarding the extent to which higher-order auditory processes remain intact.

Our stimulus presentation paradigm uncovered distinct neural responses based on the characteristics of the acoustic signal. When exposed to the first baboon grunts sequence, an increase in delta (~4 Hz) and low-gamma activity (~26.5 Hz) was found, along with enhanced cerebro acoustic coherence in the delta range. On the other hand, when exposed to the second sequences, containing chunks of faster rhythm, an increase in the theta (~6.5Hz) and low-gamma range (26.5Hz) was observed. Interestingly, responses to artificially generated noise sequences from these grunt sequences also exhibited similar power in the delta and theta range respectively.

This increase in both delta and theta power during exposure to conspecific vocalizations (baboon grunt) suggests that slow oscillations play a fundamental role in tracking slow-varying, rhythmic features of acoustic communication. This finding supports the hypothesis that slow oscillations are fundamental mechanisms for processing species-specific vocalizations and extracting supra-segmental cues from the auditory stream. This is further supported by the stronger cerebro-acoustic coherence (CaCoh) in the delta and theta range observed under the same conditions. Current research on speech processing suggests that slow oscillations emerge in response to the rhythmic properties of vocal signals[6,7]. However, the sustained slow activity observed during noise exposure suggests that this mechanism may be more general, potentially reflecting an intrinsic neural constraint on auditory segmentation. The presence of such activity in response to rhythmic noise, together with recent evidence suggesting delta rhythm is a universal feature of animal acoustic communication[1], indicates that neural processing may be inherently tuned to a slow timescale. Rather than oscillations adapting to the structure of communication signals, vocalizations and other rhythmic communicative sounds may instead be shaped to align with this pre-existing neural constraint. Fully supporting this hypothesis would however require further research, particularly on the preferred rhythmic frequency of the auditory cortex across different animal species, and the resting state hierarchy of their oscillations.

Additionally, the species-specific increase in low-gamma activity suggests a unique role for faster oscillations in processing conspecific vocalizations. This increase is also accompanied by phase-amplitude coupling (PAC) between theta and low-gamma oscillations during conspecific listening. This hierarchical organization of brain oscillations supports a multi-timescale processing framework for acoustic communication in olive baboons. In this model, slow oscillations provide a temporal reference frame, enabling supra-segmental cues to be processed through high-frequency activity coupled with the underlying slow rhythm, similar to the relationship between theta and low-gamma oscillations in speech processing.

Interestingly, delta and theta oscillations serve distinct roles in human speech processing [8,9], and the consistent PAC between theta and low gamma oscillation when exposed to both baboon grunt sequences suggest a similar separation in the baboon brain. Indeed, the presence of increased activity and tracking in the delta range when exposed to grunts, coupled with the PAC between theta and low-gamma oscillation suggest a three time scale processing of sound in baboons, happening simultaneously at the delta, theta and low-gamma scale. Furthermore, this pattern mirrors the hierarchical organization observed in the auditory cortex of humans and rhesus monkeys, hinting at a conserved mechanism for acoustic processing in the primate brain.

This is further supported by the response to human speech, where neural activity displayed a more complex spectral pattern, with increased power at 1.5 Hz, 7.5 Hz, and 11.5 Hz. Notably, the significant theta-range (4–8 Hz) response suggests that baboons can track speech rhythms at syllabic rates, consistent with prior findings in non-human primates. Further investigation into cerebro-acoustic coherence supports this idea, as speech induced coherence above the random mean in the theta range. Furthermore the presence of PAC between theta and low gamma oscillations when exposed to sentences made of familiar words, absent when they are not, mimic the effect of language proficiency on PAC in humans[10], reinforcing the idea that this three time-scales processing of acoustic stimulus is conserved in primates.

Interestingly, this theta–low gamma PAC is present not only during vocalization processing but also during noise processing, despite the absence of increased low-gamma activity. This suggests that the hierarchical organization of theta and low-gamma oscillations is not specific to communication but rather reflects a fundamental acoustic processing mechanism, with delta and theta oscillations segmenting the auditory scene. The increase in low-gamma activity during conspecific vocalization processing may therefore reflect the engagement of a specialized neural population involved in higher-order sound analysis. Alternatively, this increase could be driven by attentional mechanisms, where behaviorally relevant conspecific sounds recruit more auditory cortex neurons than environmental noise, enhancing stimulus processing and leading to greater low-gamma activity. However, this effect would likely depend on top-down interactions, which may be suppressed under the current anesthetic protocol (Fig 2).

Notably, in humans, low-gamma oscillations have been closely linked to phoneme processing, supporting the idea that high-frequency activity plays a role in segmenting the speech stream into smaller perceptually relevant units[6]. However, baboon grunts do not contain phoneme-like structures, yet still elicit a species-specific low-gamma increase. This suggests that low-gamma activity in baboons may serve a more fundamental acoustic processing function. Rather than low-gamma oscillations emerging to process phonemes specifically, it is possible that human phoneme rates evolved to align with this pre-existing neural timescale, repurposing a general auditory segmentation mechanism for linguistic processing.

Remarkably, this three-timescale hierarchy of oscillations in the primate brain contrasts with research done on other mammal taxa. Indeed, intracranial recordings in bats' auditory cortex have described a two-timescale hierarchy of their oscillations, through PAC between slow (2-8Hz) and low-gamma oscillations (25-55Hz)[11]. Similarly, a recent study on speech processing in dogs showed that they were tracking speech primarily in the delta range[12]. This suggests that a dual-timescale framework, in which slow oscillations (delta-theta) provide a temporal scaffold for parsing the auditory scene while faster oscillations (low-gamma) enhance detectability and facilitate fine-grained analysis, may represent an ancestral mammalian auditory processing strategy. However, the emergence of theta oscillations as an independent timescale in primates may have provided an additional layer of temporal organization, allowing for more efficient extraction of structured information from complex acoustic environments. This expanded processing capacity could have then ultimately been repurposed for communication. In this view, the primate auditory system would shift from a two to a three-timescale organization laying the groundwork for the complex hierarchical structure of speech.

Overall, our results provide compelling evidence that baboons process communication signals through a three-timescale mechanism, with delta and theta oscillations tracking and parsing slow-varying acoustic features and low-gamma oscillations enhancing signal detectability. This emergence of theta-range responses in baboons suggests that primates possess an intermediary processing level that, while not always functionally integrated into communication, may represent a latent precursor to the hierarchical organization observed in human speech processing. The evolutionary development of speech processing in humans may, therefore, have built upon these foundational mechanisms, ultimately giving rise to the complex hierarchical structure of oscillations observed in human speech processing. Comparative studies across animal taxa, particularly in species with complex social vocalizations such as cetaceans or songbirds, could reveal whether this three-timescale processing is strictly primate-specific or a convergent adaptation in highly communicative species.

References

1. Piette, T. *et al.* Animal acoustic communication maintains a universal optimum rhythm. Preprint at https://doi.org/10.1101/2024.10.07.616955 (2024).
2. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience* **2011**, 1–9 (2011).
3. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer, Switzerland, 2016). doi:10.1007/978-3-319-24277-4.
4. Wickham, H. *et al.* Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).
5. Piette, T. *et al.* Light propofol anaesthesia for non-invasive auditory EEG recording in unrestrained non-human primates. Preprint at https://doi.org/10.1101/2025.03.24.644890 (2025).
6. Giraud, A.-L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* **15**, 511–517 (2012).
7. Giroud, J., Trébuchon, A., Mercier, M., Davis, M. H. & Morillon, B. The human auditory cortex concurrently tracks syllabic and phonemic timescales via acoustic spectral flux. *Sci. Adv.* **10**, eado8915 (2024).
8. Poeppel, D. & Assaneo, M. F. Speech rhythms and their neural foundations. *Nat Rev Neurosci* **21**, 322–334 (2020).
9. Mai, G. & Wang, W. S. -Y. Distinct roles of delta- and theta-band neural tracking for sharpening and predictive coding of multi-level speech features during spoken language processing. *Human Brain Mapping* **44**, 6149–6172 (2023).
10. Lizarazu, M., Lallier, M. & Molinaro, N. Phase-amplitude coupling between theta and gamma oscillations adapts to speech rate. *Ann N Y Acad Sci* **1453**, 140–152 (2019).
11. García-Rosales, F. *et al.* Phase-amplitude coupling profiles differ in frontal and auditory cortices of bats. *Eur J of Neuroscience* **55**, 3483–3501 (2022).
12. Déaux, E. C. *et al.* Dog–human vocal interactions match dogs' sensory-motor tuning. *PLoS Biol* **22**, e3002789 (2024).

ANNEXES

SUPPLEMENTARY CHAPTER

Grabbing attention: The role of sound *roughness* in acoustic communication.

T. Piette, E. Esteves, E.C Déaux, A-L.Giraud, L.Arnal.

Summary Of Results

While roughness, fast amplitude modulations of sound between 30 to 150 Hz, has been identified as an acoustic niche used by humans to signal danger, its prevalence and functional role in animal communication remain largely unexplored. Given its striking salience in humans, roughness may represent an ancient and conserved mechanism for grabbing attention in critical contexts, such as distress or survival. To investigate its potential function as an attention-grabbing feature, we analyzed the vocalizations of nesting offspring from 33 bird species spanning 11 orders, assessing the presence or absence of roughness in their calls. Our findings reveal that in the majority of these species (n=30) roughness is present in offspring vocalizations. Notably, roughness was detected in at least one species within every order examined. Phylogenetic analysis further strengthens the idea of an evolutionary ancient feature showing that this characteristic rhythm is conserved across bird offspring, tracing back at least to the common ancestor of all birds.

SUPPLEMENTARY PUBLICATION

Roughness as an ancestral communication niche in animal acoustic communication.

T. Piette[1] , E. Esteves[2], M. Massenet[3] , E.-C. Déaux[4] , A-L.Giraud[1 ,2], R. Ryder[5], L.H. Arnal[2].

[1] Department of Basic Neurosciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland.

[2] Université Paris Cité, Institut Pasteur, AP-HP, Inserm, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, F-75012 Paris, France.

[3] Equipe de Neuro-Ethologie Sensorielle Bioacoustics Research Laboratory, Centre de Recherche en Neurosciences de Lyon, Université de Saint-Etienne, CNRS, Inserm, Saint-Etienne 42100, France.

[4] ARGOS, Chevagnes, France.

[5] Department of Mathematics, Imperial College London, United Kingdom.

Abstract

The ability to convey extreme emotional states and arousal through salient, non-referential vocalizations is widespread among mammals and other animals. A prime example of this can be found in juveniles, such as cries or begging calls, instinctive vocalizations which mainly serve to attract parental attention. Effective attention-grabbing vocalizations often rely on nonlinear acoustic phenomena that produce salient, temporally patterned features capable of triggering rapid behavioral responses. One such feature, roughness, rapid amplitude modulations between 30 and 150 Hz, have been identified as a key acoustic characteristic for signaling danger in humans, effectively capturing attention and eliciting swift reactions. While roughness has been well-documented in human communication and alarm sounds, its prevalence and functional significance in other animal groups remain largely unexplored. To investigate its broader evolutionary role, we analyzed its presence in the calls of

nesting juveniles from 33 bird species across 11 avian orders. Our findings reveal the presence of roughness in 30 species, distributed across all orders studied, indicating deep evolutionary roots. The presence of roughness in juvenile bird calls supports its status as an intrinsic feature of vocal expression. Furthermore, its phylogenetic conservation suggests roughness functions as a shared mechanism for capturing attention. By highlighting its widespread occurrence in both humans and birds, we propose roughness as a fundamental and conserved acoustic strategy that enhances attentional capture and adaptive responses across diverse taxa.

Methods

Vocal sequences

To perform an extensive phylogenetic comparison on a balanced phylogeny, we collected acoustic and biological data for three species per order of birds, when data were available. Acoustic sequences were gathered from the public database Xeno-canto and from different research groups that kindly shared audio files.

Call selection

To examine the role of roughness in attention-grabbing vocal displays of young birds, we focused on offspring calls, as these vocalizations primarily function to attract parental attention. We hypothesized that if roughness plays a role in this function, it should be particularly prominent at the onset of a call sequence. Accordingly, we selected 10 call sequences from 10 different individuals per species and extracted the first call from each sequence. In cases where the first call exhibited a low signal-to-noise ratio, the subsequent call in the sequence was selected instead.

Roughness

To quantify power within the roughness range in each recording, we adapted the method proposed by Arnal et al. (2015)[1], applying a two-dimensional Fourier transform to the spectrogram to obtain the modulation power spectrum (MPS). Energy in the roughness range was then measured by averaging MPS values across temporal modulations from 30 to 150 Hz, integrating over all spectral modulations. Since roughness measurements are highly influenced by stimulus duration, it was essential to conduct this analysis on the shortest possible time window to maximize species inclusion, as species with call durations shorter than the selected window would otherwise be excluded. To determine the appropriate window length, we tested roughness in human screams using time windows ranging from 100 to 1000 ms in 100 ms increments (100, 200, 300 ... 1000 ms). This analysis demonstrated that a 100 ms window was sufficient to replicate the roughness differences observed between screamed and neutral speech in human data. Consequently, we applied the MPS analysis to a 100 ms window positioned at the midpoint of each call to ensure the broadest possible species representation.

Artificial sounds

To generate artificial neutral sounds that replicated the spectral characteristics of a neutral vocalization while eliminating roughness, we used Praat[2] to extract the amplitude envelope of each vocalization and replaced its spectral content with either white noise or a tonal sound (1000 Hz). Amplitude envelopes were obtained using the "To Intensity" function. We then generated white noise and tonal sounds matching the duration of the original calls and applied the extracted envelopes using the "Multiply" function. This approach effectively removed amplitude modulations above 20 Hz, producing sounds devoid of roughness. We then tested this protocol by comparing the energy in the roughness range between human scream, neutral voice, and these artificially created sounds from human screams, showing that sounds with white noise best matched the roughness characteristic of neutral speech. Based on this finding, we used white noise to construct artificial neutral sounds for comparison in birds.

Phylogenetic model

To assess the conservation of roughness use, we examined the binary distribution of the roughness trait (use vs. non-use) across the phylogeny of our selected species. A dated phylogeny was obtained from Hedges et al. (2015) using the R package datelife[3]. We assumed that the roughness trait evolves over time in this phylogeny according to an extended Mk model[4] with distinct transition rates. This model allows for emergence of the trait once or multiple times across the phylogeny. We estimated the transition rates using maximum likelihood and calculated the marginal probability of roughness use at each internal node using phytools[5].

Software:

All analyses and visualization were done using Matlab version 2023a and R version 4.1.2 (2021-11-01) with the packages ggplot2[6], and tydiverse[7].

Results

While previous studies investigating the use of roughness in human scream could compare the latter to human neutral speech, the same approach on animal vocalizations, especially when investigating juveniles, is hardly applicable. To address this, we created artificial sounds based on animal vocalizations, that would mimic the acoustic properties of neutral vocalizations. We achieved this by extracting the envelope of each vocalization and replacing its spectral content with white noise. To validate our approach, we applied it to human screams, comparing the roughness characteristics of our artificially generated sounds to those of human screams and neutral speech. Our results show that these white-noise-based artificial sounds exhibit less energy in the roughness range than screams ($t(12) = 7.9$, $p < 0.005$) and a comparable level to neutral speech ($t(17) = -0.7$, $p = 0.49$), supporting their use as neutral controls for assessing roughness in animal vocalizations. (Fig 1a).

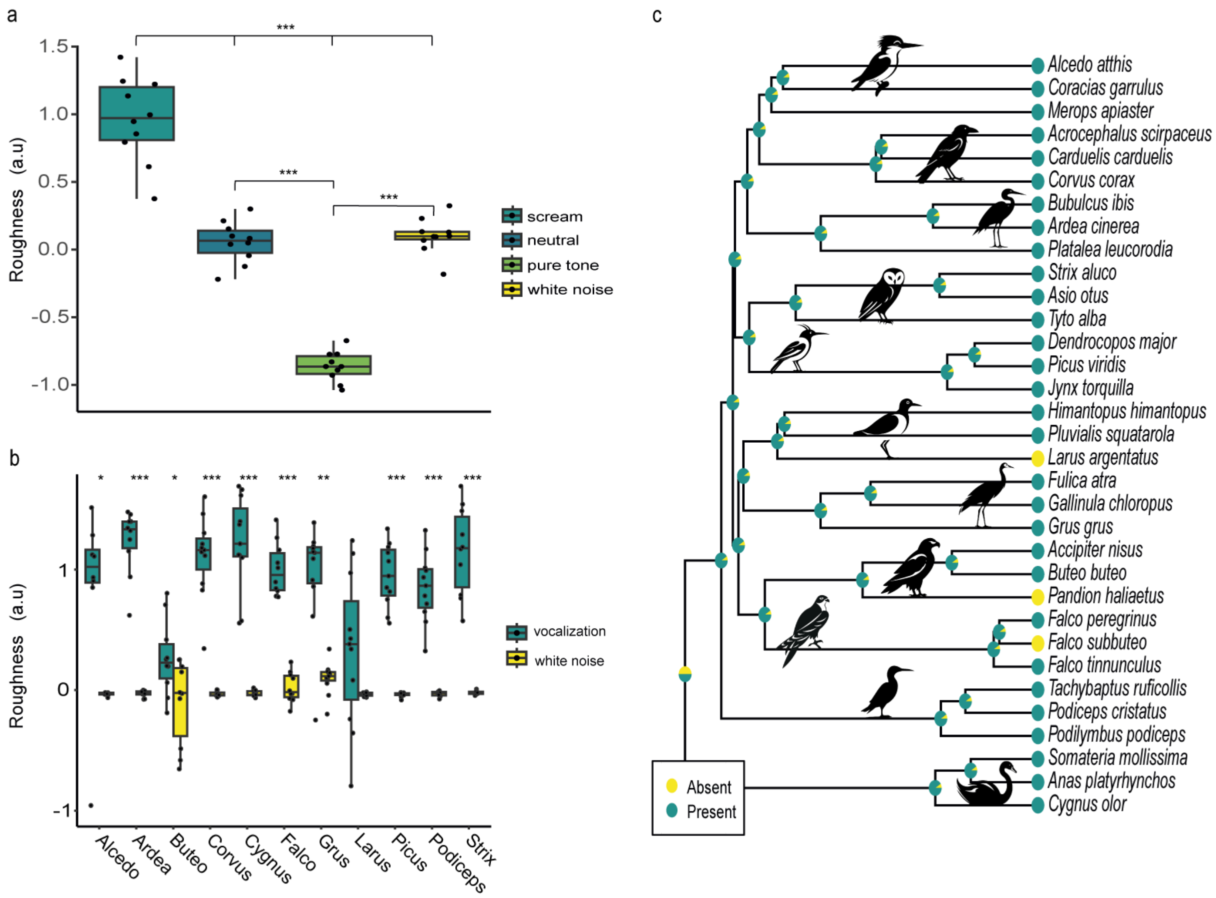


Figure 1: Conservation of roughness in nesting bird vocalizations (a) Roughness levels across human screams, human neutral voices, artificial neutral voices generated from white noise, and artificial neutral voices generated from a single tone. (b) Roughness levels in one nesting bird species per order, compared to artificial neutral vocalizations generated from white noise, showing the presence of roughness in 10 species and its absence in *Larus argentatus*. (c) Phylogenetic tree illustrating the evolutionary distribution of roughness in the vocalizations of nesting juveniles across bird species.

With our method validated, we examined the presence or absence of roughness in the vocalizations of 33 nesting bird species spanning 11 orders. Roughness was considered present when energy in the roughness range exceeded that of white-noise-based artificial sounds (Fig. 1b). Our results indicate that 30 out of 33 species exhibit significantly higher energy in the roughness range within their vocalizations. Notably, roughness was detected across all tested orders.

We next plotted the binary trait Presence/Absence of Roughness on the phylogeny of these species (Fig 1c). Phylogenetic reconstruction further suggests a high probability of roughness being present in nearly all internal nodes, consistent with its widespread occurrence across species. An exception can be observed at the root of the phylogeny, where the ancestral state is highly uncertain. Two scenarios are thus supported: either Roughness existed at the root, and survived into most modern species; or Roughness did not exist at the root, appeared independently on the two branches stemming from the root, and then survived into most modern species. The uncertainty at the root, and inability to distinguish between these scenarios, is presumably due to the uncertainty in the evolutionary rates, itself a consequence of the small number of leaves at which roughness is absent.

References

1. Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L. & Poeppel, D. Human Screams Occupy a Privileged Niche in the Communication Soundscape. *Current Biology* **25**, 2051–2056 (2015).
2. García-Rosales, F. *et al.* Phase-amplitude coupling profiles differ in frontal and auditory cortices of bats. *Eur J of Neuroscience* **55**, 3483–3501 (2022).
3. Sánchez Reyes, L. L., McTavish, E. J. & O'Meara, B. DateLife: Leveraging Databases and Analytical Tools to Reveal the Dated Tree of Life. *Systematic Biology* **73**, 470–485 (2024).
4. Lewis, P. O. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology* **50**, 913–925 (2001).
5. Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024).
6. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer, Switzerland, 2016). doi:10.1007/978-3-319-24277-4.
7. Wickham, H. *et al.* Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).