Vocal communication in harbour seal pups

Implications for language evolution

de Reus, Koen

Citation for published version (APA):
de Reus, K. (2026). *Vocal communication in harbour seal pups: Implications for language evolution*. [PhD Thesis, Vrije Universiteit Brussel, Radboud University Nijmegen]. Van Marle Drukwerk.

MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

VOCAL COMMUNICATION IN HARBOUR SEAL PUPS

Implications for language evolution

KOEN DE REUS

Vocal communication in harbour seal pups

Implications for language evolution

Koen de Reus

Funding Bodies

This research was funded by the Fonds Wetenschappelijk Onderzoek project "Interactive vocal rhythms" (G034720N) awarded to Bart de Boer, and by the Max Planck Society for the Advancement of Science (www.mpg.de/en).

International Max Planck Research School (IMPRS) for Language Sciences

The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs.

The MPI series in Psycholinguistics

Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series.

ISBN: 978-94-92910-72-1
Cover design by Stijn de Reus and Ivi Apostolatos
Printed and bound by Van Marle Drukwerk, Hengelo

The research reported in this thesis was conducted at Sealcentre Pieterburen, in Pieterburen, and the Max Planck Institute for Psycholinguistics, in Nijmegen, both in the Netherlands.

Vocal communication in harbour seal pups

Implications for language evolution

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties

en

ter verkrijging van de graad van doctor
aan de Vrije Universiteit Brussel
op gezag van de rector magnificus prof. dr. J. Danckaert

in het openbaar te verdedigen op
vrijdag 20 februari 2026
om 10.30 uur precies

door

Koen de Reus
geboren op 4 januari 1995
te Emmen

Promotoren:

Prof. dr. J.A.M. Holler

Prof. dr. B. de Boer (Vrije Universiteit Brussel, België)

Manuscriptcommissie:

Prof. dr. H.A. Özyürek

Prof. dr. E. Gonzalez Boix (Vrije Universiteit Brussel, België)

Prof. dr. P. Van Eecke (Vrije Universiteit Brussel, België)

Prof. dr. S. Nozaradan (Université Catholique de Louvain, België)

Prof. dr. S. C. Vernes (University of St Andrews, Verenigd Koninkrijk)

Vocal communication in harbour seal pups

Implications for language evolution

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board

and

to obtain the degree of doctor
from Vrije Universiteit Brussel
on the authority of the Rector Magnificus prof. dr. J. Danckaert

to be defended in public on
Friday, February 20, 2026
at 10.30 am

by

Koen de Reus
born on January 4, 1995
in Emmen (the Netherlands)

Supervisors:

Prof. dr. J.A.M. Holler

Prof. dr. B. de Boer (Vrije Universiteit Brussel, Belgium)

Manuscript Committee:

Prof. dr. H.A. Özyürek

Prof. dr. E. Gonzalez Boix (Vrije Universiteit Brussel, Belgium)

Prof. dr. P. Van Eecke (Vrije Universiteit Brussel, Belgium)

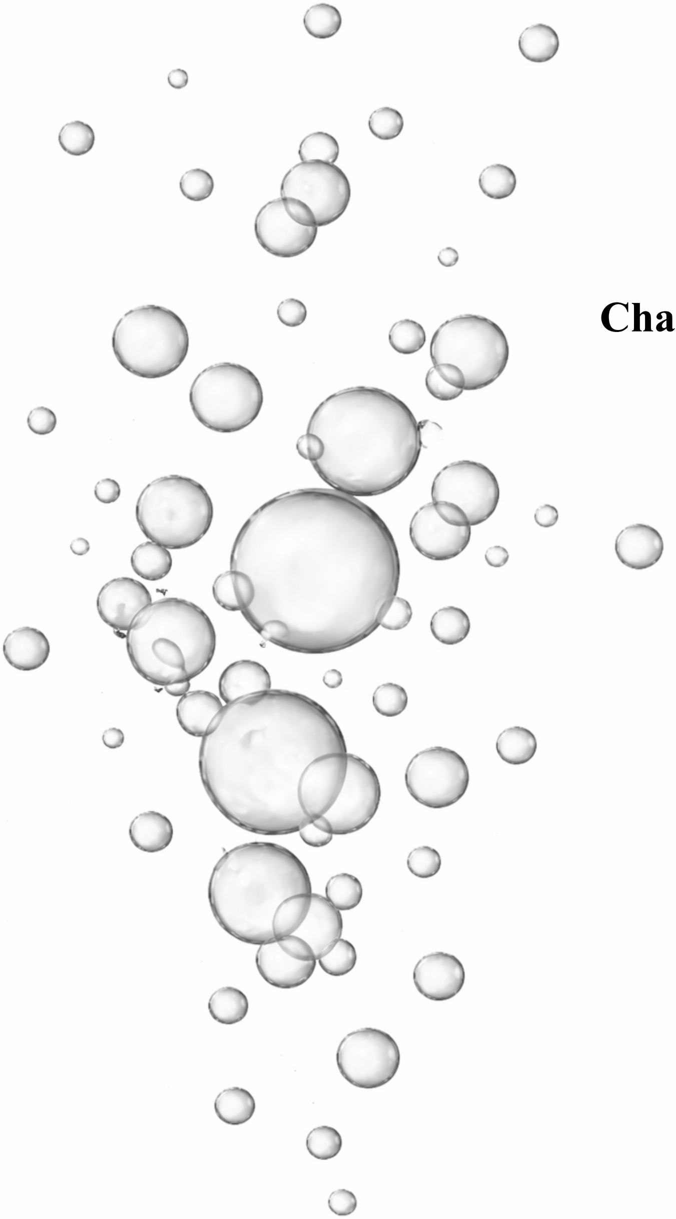
Prof. dr. S. Nozaradan (Université Catholique de Louvain, Belgium)

Prof. dr. S. C. Vernes (University of St Andrews, United Kingdom)

To my amazingly supportive grandmother Wil

Table of contents

Chapter 1



1 | General introduction

"Nothing in biology makes sense except in the light of evolution."
Theodosius G. Dobzhansky (1964)

Have you ever triggered a vocal interaction with a non-human animal, by barking at a dog or miaowing at a cat? I personally have done this many times. I make eye contact with the dog, squat down to its level to appear less threatening, and bark, mimicking a signal from its own vocal repertoire. In response, the dog locks eyes with me, starts wagging its tail, approaches, and barks too. If I continue to bark back, an extended interaction can ensue, with the dog and me alternating barks as if engaged in conversation. Although I always wondered if my barking was meaningful to the dog, one thing was obvious: my behaviour triggered a communicative interaction.

The "interaction engine" hypothesis posits that humans possess a set of specialised communicative and cognitive skills to facilitate complex social interactions (Levinson, 2006). These fundamental skills, which include intentionality, multimodal communication, sequence organisation, and turn-taking, enable us to coordinate social action (e.g., cooperative behaviours). Yet, social interaction can only take place if all individuals actively participate. In the example above, I use multiple signals across different communicative modalities—eye contact, body posture, and vocalisation—to approach the dog and clearly communicate my intent to communicate. Similarly, the dog uses several modalities, such as eye contact, tail wagging, and vocalisation, to signal its desire to engage in social interaction. By responding to my initial bark, our interaction begins with a greeting-response pattern, a fundamental unit of sequence organisation known as an adjacency pair. If the interaction continues, we alternate vocal productions, structuring the interaction by taking turns. In doing so, the dog and I coordinate our actions in real time, all the while enjoying a mutually rewarding interaction through petting. Observing similar communicative behaviour in another animal therefore suggests that it, too, possesses communicative and cognitive skills that facilitate social interaction.

Language is a remarkably powerful tool for social interaction and allows humans to communicate about virtually anything. It is often regarded as a hallmark of humanity, and is fundamental in shaping our cognition, behaviour, and culture (Fitch, 2010). However, language did not evolve in a vacuum. Its evolution is best understood from a bio-cultural perspective that acknowledges the importance of the biological structures enabling language production and perception, the role of cultural transmission as a result of repeated social interactions, and the interplay between them (Arnon et al., 2025). Language emergence thus required the convergence of several distinct abilities and cannot be attributed to a single factor, such as a specific genetic mutation, neuroanatomical specialisation, behavioural trait, or cultural innovation (e.g., Chomsky, 2014; Crow, 2000). Like other complex systems (e.g., the visual system), language emerged gradually through the modification and recombination of pre-existing biological structures and cognitive mechanisms, as well as the repurposing (i.e., exaptation) of others (Fitch, 2011; Gould & Vrba, 1982). Consequently, certain features of the language system likely have evolutionary origins that extend beyond humans. Adopting a comparative approach to identify features shared between human language and the communication systems of other animal species could provide valuable insights into the biological and social contexts that shaped their evolution.

A cross-species comparative approach to language evolution

Since language does not fossilise and its evolution can only be understood through indirect evidence (Christiansen & Kirby, 2003; Fitch, 2010; Hauser et al., 2014), studying what other animals can and cannot do is essential for understanding how these communication systems compare (Fitch, 2005). An essential first step in studying language evolution is to break language down into its constituent parts. One of the first to attempt this was Charles Hockett, who outlined 13 design features of language (later expanded to 16) in his seminal paper *The Origin of Speech*, including vocal-auditory channel, duality of patterning, traditional transmission, and displacement (Hockett, 1960, 1963). Although some features, like the vocal-auditory channel, have been reconsidered in light of work on sign languages and multimodal communication (Goldin-Meadow & Brentari, 2017; Holler & Levinson, 2019; Pleyer et al., 2025), his work provided a framework for comparing human language to other animal communication systems.

Simply put, the comparative method identifies shared and unique features of communication systems across species, allowing researchers to study language evolution using empirical methods (Fitch, 2005). Specifically, it considers the evolutionary or phylogenetic relationships between species to draw inferences (Lorenz, 1974). Homologous features, which are shared by species within the same evolutionary group, help infer characteristics of extinct common ancestors (e.g., lactation in mammals), whereas analogous features, which evolved independently in separate evolutionary groups, support inferences about the biological and social contexts driving their evolution (e.g., wings of birds, bats, and insects). By comparing communication systems across species, we can reconstruct the evolutionary steps that led to language emergence and determine which features are uniquely human and which features arose through shared ancestry or convergent evolution.

The evolution of speech

The evolution of speech offers a valuable lens for understanding language evolution. While language itself is a multimodal phenomenon, speech is widely regarded as the primary mode of communication in humans (Fitch, 2005). Our capacity for speech depends on a specialised set of anatomical, neural, and cognitive traits not found in any other species, including our primate relatives. However, fundamental mechanisms of human vocal production are largely shared with other animals, including the basic anatomy and physiology of the vocal tract. For example, the source–filter theory of sound production (Fant, 1971) and the myoelastic-aerodynamic theory of phonation (van den Berg, 1958) describe how sound is produced and modified in the vocal tract, and their core principles are broadly applicable across vertebrates, including mammals and birds (Elemans et al., 2015; Taylor & Reby, 2010). The source–filter theory explains how airflow from the lungs passes through a source (i.e., the larynx in mammals or the syrinx in birds), making the vocal folds (or membranes, in the case of the syrinx) vibrate and generating an acoustic wave that is then shaped by the filter (i.e., the oro-nasal cavities of the upper vocal tract) before being released into the environment as a vocalisation (Fant, 1971). The myoelastic-aerodynamic theory suggests that vocal fold vibration in the larynx generates sound due to the elastic properties of the vocal folds and the aerodynamic principles of airflow (van den Berg, 1958). Notably, producing a vocalisation requires precise coordination between the lungs, the larynx or syrinx, and the upper vocal tract (Fitch, 2006). For example, in mammals, the lungs generate the necessary airflow and pressure to make the vocal folds vibrate. The tension of the vocal folds, controlled by the laryngeal muscles, determines their vibration rate and consequently the perceived pitch. Finally, by adjusting the movement and shape of vocal tract articulators such as the tongue, lips, jaw, and palate, an animal can modify its upper vocal tract to selectively amplify or dampen specific frequencies, creating distinct formant patterns that help differentiate between vocalisations. In sum, producing vocalisations depends on the precise coordination of several elements, underscoring the complexity of vocal production and the value of cross-species investigations in identifying which anatomical, neural, and cognitive traits are uniquely human and which have deeper evolutionary roots.

The descended larynx: An adaptation for speech?

One defining anatomical feature of the human vocal tract is our permanently descended larynx—a trait not observed in non-human primates. Early on, scientists hypothesised that this reconfiguration played a key role in the evolution of speech. Specifically, a descended larynx lengthens the vocal tract, potentially expanding the range of possible formant patterns we can produce (Lieberman et al., 1969). Consequently, the descended larynx has been proposed as a reason for why humans can produce a large repertoire of distinguishable vowels and speech sounds. However, recent modelling work in long-tailed macaques (*Macaca fascicularis*) has shown that, even without a descended larynx, their vocal tracts are flexible enough to produce a wide range of sounds similar to those of human speech (Fitch et al., 2016). This finding demonstrates that a descended larynx alone is not sufficient for speech production and further highlights the importance of dynamic vocal tract adjustments in producing distinct speech sounds.

Additionally, comparative work has identified several mammals with a permanently descended larynx including koalas, deer, and large cats (Fitch & Reby, 2001; Frey et al., 2012, 2018; Weissengruber et al., 2002). Given the lack of speech-like vocal flexibility in these species, the evolution of a descended larynx was likely driven by other selective pressures. One leading explanation is the “size-exaggeration” hypothesis, which proposes that laryngeal descent evolved to increase vocal tract length, thereby exaggerating perceived body size. Acoustic allometry refers to the relationship between an animal’s body size and the acoustic properties of its vocalisations, with larger animals generally producing lower formants (Bowling et al., 2017). Since longer vocal tracts produce vocalisations with lower formant frequencies, laryngeal descent may have evolved to deviate from this relationship. Such an adaptation allows callers to produce disproportionately low formants, simulating a larger body size to intimidate rivals and attract mates (Charlton & Reby, 2016). The “size-exaggeration” hypothesis is further supported by the secondary descent of the larynx observed in human males during puberty (Fitch & Giedd, 1999), indicating a clear role in reproductive signalling (Pisanski et al., 2016). Although the descended larynx undoubtedly contributed to the expansion of our human vocal range, its evolution was likely driven by multiple selective pressures, and its importance in the emergence of speech may originally have been overstated (e.g., Carstairs-McCarthy, 1999; Fitch, 2000).

Neural control of speech: The “direct connections” hypothesis

Speech requires fine-grained motor control over the laryngeal muscles controlling the vocal folds and the position of vocal tract articulators (i.e., tongue, lips, palate, etc.), which together coordinate and shape vocal production. Unlike our primate relatives, humans show direct connections between neurons in the motor cortex of the brain and the primary motor neurons controlling the muscles of the larynx (Simonyan & Jürgens, 2003), allowing for more control over vocalisations. This finding suggests that enhanced voluntary control over vocal production was a necessary pre-adaptation to the evolution of speech (Jürgens, 2002; Nieder & Mooney, 2019; Simonyan & Horwitz, 2011). Support for the “direct connections” hypothesis comes from studies on a rare trait observed in some mammals and birds: vocal learning—the ability to modify, learn, and produce vocalisations based on auditory experience (Fitch, 2018; Vernes et al., 2021). Specifically, direct connections were observed between the brain and the syrinx in vocal learning parrots and songbirds, but were absent in non-vocal learning birds (Striedter, 1994; Wild, 1997). The presence of these direct neural connections in humans and other vocal learning animals therefore emphasises the critical role of fine motor control in the evolution of complex vocal behaviours such as speech.

Vocal learning: A key ability for speech acquisition

Vocal learning is critical for human speech acquisition (Janik & Slater, 1997; Jarvis, 2019), as it enables humans to learn new vocalisations and develop a rich, open-ended vocabulary necessary for effective social interaction. In contrast, although non-human primates have vocal tracts capable of producing speech-like sounds (e.g., Fitch et al., 2016), they lack the direct neural connections required for fine-grained motor control of the vocal tract (Simonyan & Jürgens, 2003). While studies of speech evolution have traditionally focused on non-human primates—our closest living relatives—their limited vocal flexibility means they offer only partial insights into the mechanisms underlying speech development. In contrast, vocal learning has evolved independently in several other groups of mammals and birds, including parrots, songbirds, hummingbirds, bats, whales, dolphins, and seals (Tyack, 2019; Vernes et al., 2021). Vocal learning in these species serves to expand their vocal repertoires and facilitate social behaviours like mate attraction, territorial defence, and individual and group recognition (Fitch, 2005), highlighting the role of vocal learning in regulating social interactions. It also enables animals to modify vocalisations and facilitate transmission of acoustic signals in different environments (Ey & Fischer, 2009). Comparative work could help determine whether vocal learning evolved through sexual selection for complex vocalisations or kin selection for information sharing, with acoustic adaptation potentially playing a role under either hypothesis (Nowicki & Searcy, 2014). Vocal learning thus plays a fundamental role in the development of advanced vocal communication systems and offers an important window through which we can better understand the evolution of speech. Specifically, studying vocal learning across a broader range of taxa can provide valuable insights into the evolutionary pressures—whether sexual, kin, or environmental—that may have shaped cognitive and neural mechanisms involved in the emergence of speech. By shifting the focus away from a primate-centric approach, we can gain a more comprehensive understanding of which evolutionary pressures and cognitive/neural mechanisms drove the development of vocal learning, shedding light on the emergence of vocal communication systems in both humans and other species.

A framework to study vocal learning

The first theoretical framework to study vocal learning classified this trait into two distinct forms: (i) vocal production learning, where individuals modify or learn new vocalisations, and (ii) vocal contextual learning, where individuals learn to use vocalisations in novel contexts and associate vocalisations with new meanings or outcomes (Janik & Slater, 1997). Despite vocal contextual learning being more widespread across species, studies on the evolution of speech have traditionally focused on species capable of vocal production learning, such as the imitation abilities of songbirds and cetaceans (i.e., whales and dolphins; Janik & Slater, 1997). This focus emerged largely because the ability to modify and learn vocalisations based on auditory experience is considered critical to human speech acquisition.

A subsequent framework conceptualised vocal production learning as a trait existing along a continuum (Petkov & Jarvis, 2012), with species ranging from "vocal non-learners" to "high vocal learners". Under this model, the most basic form of vocal production learning involves modifying the amplitude and temporal structure of innate calls, while more advanced forms involve spectral modifications, culminating in the ability to imitate acoustic signals by actively altering spectro-temporal properties of vocalisations (Arriaga & Jarvis, 2013). However, the continuum hypothesis assumes a linear evolutionary pathway for vocal production learning, which may not fully capture its actual complexity, particularly since these different forms may rely on distinct neural pathways (Tyack, 2019).

The latest theoretical developments propose to study vocal learning as a trait consisting of multiple dimensions or modules (Vernes et al., 2021; Wirthlin et al., 2019). Specifically, one influential framework proposes that vocal learning can be broken down into three modules that capture different aspects of how vocalisations are structured, modified, and used in communication, namely (1) vocal versatility, (2) vocal production variability, and (3) vocal coordination (Wirthlin et al., 2019). These modules can evolve independently under different evolutionary pressures, allowing a species to exhibit a high degree of complexity in some while being more limited in others. Below is a brief overview of the three modules:

1. **Vocal versatility** encompasses both the diversity of vocal sounds a species can produce and the degree to which these sounds can be modified through experience. This interplay between vocal tract anatomy and the neural circuits controlling vocal production is key to understanding how vocal flexibility evolved. For example, comparing expected vocalisations based on vocal tract anatomy and body size with actual vocal output can help infer the extent of a species' vocal control.

2. **Vocal production variability** encompasses how vocal output changes throughout development or in response to social and environmental influences, providing valuable insights into the role of auditory feedback in shaping vocal development. This variability allows for disentangling the relative contributions of biological factors (i.e., age, sex, or body size) and social influences on vocal output, thereby improving our understanding of vocal learning plasticity across an individual's lifetime.

3. **Vocal coordination** refers to the ability to time vocalisations during vocal exchanges, as seen in species that engage in turn-taking during communication (e.g., human conversation). This module emphasises the importance of vocal timing in structuring social interactions, offering insights into predictive timing, adjustment mechanisms, and how social dynamics influence these processes.

In this thesis, I apply this framework to investigate the connections between vocal flexibility, vocal coordination, and the biological and social influences on vocal communication, offering a robust approach that provides valuable insights into the mechanisms underlying vocal learning. Among species identified as vocal learners, seals—particularly harbour seals—could serve as promising comparative models for investigating both vocal learning and the evolution of speech.

Why seals can contribute to research on the evolution of speech

Given their unique combination of vocal and ecological diversity, phylogenetic relevance, and anatomical similarities, pinnipeds (seals, sea lions, and walruses) represent an underutilised yet valuable comparative model for exploring the evolution of complex vocal communication systems—such as language in humans. Pinnipeds are one of the rare mammalian groups which present evidence for vocal learning (Schusterman, 2008). Over three-quarters of the 33 extant species use vocalisations to modulate social interactions, exhibiting diverse repertoires in both air and underwater environments (Berta et al., 2018; Charrier, 2021; Schusterman & van Parijs, 2003). These vocalisations range from broadband calls and barks to rhythmic trills and complex underwater sequences, serving adaptive functions in various behavioural contexts such as mother–offspring recognition, mate attraction, and territorial displays (Charrier, 2020; Hanggi & Schusterman, 1994; Mathevon et al., 2017). Notably, their amphibious lifestyle (i.e., spending time both on land and underwater) makes them more accessible for controlled experiments compared to cetaceans, which are challenging to observe and study in natural settings due to their large size, extensive home ranges, and fully aquatic lifestyle. Moreover, as mammals, pinnipeds are phylogenetically closer to humans than the well-studied vocal learning songbirds and parrots (Ravignani et al., 2016). Last but not least, pinnipeds have a similar vocal tract anatomy to humans, producing vocalisations with their larynx positioned at a comparable angle in relation to airflow (Schneider, 1962, 1964). Furthermore, their vocal production follows the passive mechanisms outlined in the source–filter theory, where sounds generated by the larynx are shaped by the upper vocal tract (Taylor & Reby, 2010). Altogether, these characteristics make pinnipeds particularly interesting candidates for comparative studies on the evolution of speech.

Despite being one of the less vocal pinnipeds, the harbour seal (*Phoca vitulina*) possesses advanced vocal learning capabilities, as demonstrated by the striking case of Hoover, who, separated from its mother as a pup and raised in captivity, began imitating words and phrases presumably spoken by his human caretaker as an adult (Duengen et al., 2023; Ralls et al., 1985). To date, Hoover's case of vocal imitation remains the only well-documented evidence of vocal production learning in harbour seals, highlighting the potential value of further investigating their vocal behaviour. Fortunately, existing research on vocal communication in this species provides a solid foundation for exploring their vocal learning capabilities in more detail. For example, harbour seal pups are very vocal throughout the lactation period, producing individually distinctive mother attraction calls (MACs) that help mothers locate their pups after returning from foraging trips at sea (Perry

& Renouf, 1988; Renouf, 1984). While mothers can recognise their offspring shortly after birth (Renouf, 1985), the spectro-temporal properties of MACs vary with age, sex, and body size (Khan et al., 2006; Sauvé et al., 2015), likely as a result of developmental changes in the pups' vocal anatomy. In addition to these maturational changes, harbour seal pups exhibit vocal flexibility by adjusting the fundamental frequency of their MACs in response to background noise (Torres Borda et al., 2021), demonstrating acoustic adaptation to environmental conditions. Furthermore, harbour seal pups adjust the timing of their calls to avoid overlap with those of surrounding pups (Anichini et al., 2023; Ravignani, 2019), indicating that they can modify their vocalisation patterns to optimise their chances of reuniting with their mothers in the noisy and crowded colony environment. After weaning, however, these calls disappear entirely from their vocal repertoire, with most vocalisations ceasing aside from occasional growls during agonistic interactions (Renouf, 1984). Adult males vocalise again during the breeding seasons to defend underwater territories and attract mates (Hanggi & Schusterman, 1994; Hayes et al., 2004), with male vocalisations exhibiting site-specific variation that contributes to the development of regional vocal dialects (Bjørgesæter et al., 2004; van Parijs et al., 2003). This combination of early-life vocal activity and evidence of vocal variability, flexibility, and coordination makes the harbour seal an ideal candidate model for comparative studies on the evolution of vocal learning and speech.

Research objectives and thesis outline

While harbour seals have been identified as vocal learners, research on vocal communication in harbour seals has largely focussed on describing their vocal repertoire, developmental changes, and geographic variation. However, experimental studies directly assessing their vocal variability, flexibility, and coordination remain scarce, with many conclusions drawn indirectly. Notably, since Hoover's case of vocal imitation from the 1980s (Ralls et al., 1985), only a few studies have explored harbour seals' ability to modify vocalisations in response to environmental factors (Torres Borda et al., 2021) or their ability to adjust call timing in social contexts (Anichini et al., 2023; Ravignani, 2019). A more comprehensive understanding of these aspects provides a crucial foundation for future cross-species comparisons aimed at uncovering the evolutionary mechanisms underlying vocal learning and its role in the emergence of speech.

This thesis contributes to that foundation by applying a recent modular vocal learning framework (Wirthlin et al., 2019) to develop a more comprehensive understanding of harbour seals' vocal capabilities and their role in social interactions. It investigates key vocal learning components from the framework's modules, including vocal coordination during social interactions (Chapter 2 and 3), vocal flexibility and control (Chapter 4), and the effect of social factors on vocal production during development (Chapter 5). To study these components, this thesis integrates a combination of non-invasive experimental methodologies, including post-mortem anatomical studies, as well as acoustic recordings and behavioural playback experiments. Since harbour seals are highly vocal in their first weeks of life, with vocal activity declining after weaning, this thesis focusses exclusively on pups. Below is an overview of the thesis chapters.

Chapter 2 begins this thesis by demonstrating how the comparative method—drawing from multiple animal groups and disciplines—can offer broader insights into communicative behaviours. This chapter is a literature review that explores the role of call timing and rhythm in interactions between dyads (i.e., a group of two individuals) across various communicative modalities, behavioural contexts, and animal groups, including mammals, birds, anurans, and insects. The review includes a table of definitions for terms related to rhythm in dyadic interactions and distinguishes between rhythmic and non-rhythmic interactions, based on whether individuals adjust the timing of their signals to facilitate communication. The findings are evaluated in light of the four fundamental ethological questions: function, phylogeny, development, and mechanism (Tinbergen, 1963). The review highlights that rhythmic signalling behaviours are widespread in the animal kingdom and play an important role in regulating social interactions. However, it also underscores the need for further research on the role of signal timing development and the mechanisms that regulate it. Additionally, while the scientific literature has largely focused on interactive signalling in the acoustic and visual modalities, more work is needed on tactile and multimodal signals. Ultimately, adopting a comparative approach provides a powerful framework for understanding the evolution of communicative behaviours, bridging gaps in the literature and highlighting critical areas for future investigation that will advance the study of animal communication and language evolution.

Chapter 3 investigates turn-taking behaviour in vocal interactions between harbour seal pups. Turn-taking refers to the rapid exchange of communicative turns during social interactions, characterised by minimal silent gaps and minimal overlap between turns (Sacks et al., 1974). In human conversation, the average gap is around 200 ms (Stivers et al., 2009), suggesting that turn-taking relies on predictive timing mechanisms and mutual timing adjustments to facilitate communication. Similar turn-taking behaviour has also been observed in various non-human animals, where it serves to strengthen group cohesion, maintain social bonds, and facilitate cooperation (Badihi et al., 2024; Demartsev et al., 2018; Pika et al., 2018). Previous studies on harbour seal pups have shown that call sequences exhibit temporal organisation reminiscent of interactive communication and that pups adopt an antisynchronous calling strategy to avoid acoustic overlap (Anichini et al., 2023; Ravignani, 2019; Ravignani, Kello, et al., 2019). However, these studies did not apply the standard methodology used in conventional turn-taking research to determine whether harbour seal pups exhibit turn-taking behaviour. This chapter addresses this gap by applying response latency analyses from human behavioural research to assess whether vocal interactions exhibit the core characteristics of turn-taking, namely minimal gaps between turns, avoidance of overlap, and mutual timing adjustment. Minimal gaps between turns and avoidance of overlap are reflected in short, positive response latencies, while mutual adjustment is reflected in a negative lag-1 autocorrelation derived from time series analyses (see Pouw & Holler, 2022). Additionally, this chapter analysed results from a playback experiment, in which previously recorded harbour seal pup calls were played back to pups, to test how they would adjust their call timing when exposed to variations in rhythm, tempo, call duration, and sex of the caller. Results indicate that harbour seal pups show all the core characteristics of turn-taking behaviour and that they adapt their call timing to the tempo of the interaction and the duration of conspecific signals, demonstrating flexible vocal coordination during social interactions.

Chapter 4 describes vocal tract anatomy in harbour seals and examines its role in shaping their vocal flexibility. A key focus of this chapter is on acoustic allometry, which describes how body size is reflected in the acoustic properties of vocalisations (Bowling et al., 2017). Typically, larger animals produce lower-frequency sounds, but deviations from this pattern point to mechanisms that allow an animal to sound smaller or larger than it actually is. The chapter first examines whether harbour seal vocalisations strictly follow expected acoustic allometry. Reanalysing data from Torres Borda and colleagues (2021), it shows that seals of different sizes can produce similar fundamental frequency values under varying noise levels, suggesting that behavioural adaptations to environmental factors can override body size constraints on vocal production, allowing them to deviate from acoustic allometry. Using post-mortem anatomical examinations, this chapter investigates whether this vocal flexibility arises from morphological adaptations to their vocal tract or from enhanced vocal control, inferred from whether the size of vocal tract structures (vocal folds, upper vocal tract, and trachea) scales proportionally with body size. Findings indicate that body size predicts vocal tract dimensions, though some variation occurs due to age and sex. However, since harbour seals do not exhibit unique morphological adaptations to their vocal tract allowing them to deviate from acoustic allometry, these results suggest that enhanced neural control over vocal structures—a necessary trait for vocal production learning—may be the key mechanism offering them a larger degree of vocal flexibility. Additionally, this chapter provides detailed anatomical illustrations that highlight the major structures of the harbour seal vocal tract and depict key components of the source–filter theory of vocal production.

Chapter 5 examines the effect of the social environment on the acoustic structure of harbour seal pup vocalisations. In human conversation, the Communication Accommodation Theory describes how individuals adjust their speech styles during social interactions, either converging to express social closeness or diverging to signal distance from their conversation partner (Giles & Baker, 2008). Similarly, vocal convergence has been observed in various non-human animal species as a means to strengthen social bonds (Baciadonna et al., 2022; Fischer et al., 2020; Prat et al., 2017). This chapter investigates whether harbour seal pups accommodate their vocalisations by converging towards the calls of individuals they frequently interact with. Specifically, it disentangles the maturational effects related to body size, age, and sex from learned vocal modifications. After recording pup vocalisations daily over several weeks, Bayesian models helped determine how acoustic parameters changed throughout development. Acoustic distances were then computed between pairs of seals, while controlling for differences in body size, age, and sex. The results indicate that the vocalisations of pups housed together are more similar than those of pups housed separately, while still remaining distinct enough for individual recognition. These findings suggest that harbour seals exhibit vocal convergence during social interactions, showing that social influences shape vocal development in this species and highlighting intriguing parallels with humans and other species.

Chapter 6 concludes this thesis with a summary and discussion of the main findings in the context of language evolution. This chapter highlights how work on vocal learning capacities in harbour seals can inform studies on human language evolution. Specifically, it discusses the connection between vocal flexibility and neural control of the vocal apparatus, as well as the role of social interactions in shaping (vocal) communication. Moreover, the chapter explores key methodological contributions and considerations, including our growing understanding of vocal learning, the benefits and challenges of the comparative method, and study limitations when working with captive non-human animals. It also emphasises the power of playback experiments in studying animal behaviour. Finally, this chapter highlights several directions for future research in harbour seals, which may offer additional insights into how language evolved. Notably, future studies could explore the role of sociality in driving communicative complexity, how non-vocal and multimodal signals facilitate social interactions, the extent of vocal flexibility and how it changes during development, and the neural circuits underlying vocal control.

Chapter 2

2 | Rhythm in dyadic interactions

Abstract

This review paper discusses rhythmic interactions and distinguishes them from non-rhythmic interactions. We report on communicative behaviours in social and sexual contexts, as found in dyads of humans, non-human primates, non-primate mammals, birds, anurans and insects. We discuss observed instances of rhythm in dyadic interactions, identify knowledge gaps and propose suggestions for future research. We find that most studies on rhythmicity in interactive signals mainly focus on one modality (acoustic or visual) and we suggest more work should be performed on multimodal signals. Although the social functions of interactive rhythms have been fairly well described, developmental research on rhythms used to regulate social interactions is still lacking. Future work should also focus on identifying the exact timing mechanisms involved. Rhythmic signalling behaviours are widespread and critical in regulating social interactions across taxa, but many questions remain unexplored. A multidisciplinary, comparative cross-species approach may help provide answers.

Adapted from: de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., Bruno, J. H., Trainor, L., & Ravignani, A. (2021). Rhythm in dyadic interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200337. https://doi.org/10.1098/rstb.2020.0337

Introduction

Animals rely on effective signal transfer for communication with conspecifics. The sender must produce a clear signal that can be readily detected by the receiver (Dusenbery, 1992). A variety of sensory systems evolved to accommodate the production and reception of signals, and over time, these systems were fine-tuned by selective evolutionary pressures leading to remarkable species-specific adaptations (e.g., three-ossicle middle ear in mammals adapted for acoustic transmission of high-frequency signals; Tucker, 2017). Signals can be produced in several modalities, including acoustic, visual, chemical, tactile, etc.; each modality has its own (psycho)physical limitations that impose constraints on communication. For example, acoustic signals are effective for long-range information transfer but are affected by signal attenuation and degradation, especially at higher frequencies (Wiley & Richards, 1978). In the visual modality (e.g., signed languages), receivers are required to be in close contact with the signaller to accurately discriminate the signals because obstacles can block them. Regardless of the modality in which signalling occurs, interactive signalling will henceforth refer to a communicative exchange during which signals are transferred, back and forth, over a short time scale.

Timing is key in interactive signalling behaviour (Anichini et al., 2020; Ravignani et al., 2014; Ravignani, Verga, et al., 2019). Individuals take turns in human conversation and entrain to the rhythms of other players while making music (Keller et al., 2014; Nguyen et al., 2021). Individually timed behaviour in groups can lead to impressive collective phenomena such as synchronous movements to musical rhythms at concerts (Trehub, Becker, et al., 2015), cricket choruses (Greenfield, 1994a) and the bioluminescent flashing of fireflies (Buck, 1988). Unfortunately, natural interactions involving many agents are often difficult to study under controlled conditions. Hence, for practical reasons, most studies on interactive rhythms minimise the number of individuals by focusing on dyadic interactions. Some species can achieve a high degree of temporal coordination, and even synchronisation, during conspecific interactions (Backwell et al., 1998; Grafe, 1999; Greenfield, 1994a; Hall, 2009). Therefore, we ask: how does rhythm shape dyadic interactions and how do dyadic interactions shape rhythm?

Here, we review stable interactions in dyads across species, modalities and contexts. This review covers humans and non-human animals—namely insects, anurans, birds, non-human primates and non-primate mammals—interacting mainly in the acoustic and visual modalities, and in social and sexual contexts. We (i) begin by proposing widely applicable definitions for relevant concepts (Table 2.1); (ii) highlight rhythmic dyadic interactions in each group; and (iii) discuss the existing behaviours in the light of Tinbergen's four questions while proposing suggestions for future research.

Table 2.1. Definitions of terms relating to dyadic interactions.

Term	Definition
Signals	"Traits that (1) change another organism's behaviour while benefitting the sender, that (2) are evolved for this function, and that (3) have their effects through the evolved response of the receiver" (Stegmann, 2005, p. 1011)
Rhythm	"Pattern of time intervals between the onset of events" (van den Bosch der Nederlanden et al., 2019, p. 165)
Rhythmic interaction	An interaction where two conspecifics adjust their timing behaviour to each other to create temporal regularities that facilitate the interaction
Entrainment	"Spatiotemporal coordination resulting from rhythmic responsiveness to a perceived rhythmic signal" (Phillips-Silver et al., 2010, p. 5)
Duets	"Joint […] displays where two [partners] coordinate their [signals] with a degree of temporal precision" (Farabaugh, 1982); traditionally, duets occur in the acoustic modality, but we believe the use of the term should be extended to all other communicative modalities
Synchrony	"Precise coincidence of events in time" (Ravignani, 2017, p. 158)

Turn-taking	“Orderly exchange of purely communicative signals or behaviours (e.g. peek-a-boo games in humans) between individuals characterised by principles for the coordination of turn transfer, which result in observable temporal regularities” (Pika et al., 2018, p. 2); we subscribe to the idea that turn-taking is mostly rhythmic in its temporal dimension (Ravignani, Verga, et al., 2019)
Alternation	“Where the regularly repeating signals of two […] individuals are broadcast such that they do not occur at the same time” (Greenfield, 2005, p. 4)
Antiphony	“When […] two animals transmit sounds among themselves in response to preceding signals” (Yoshida & Okanoya, 2005, p. 155)

We first need to distinguish between what constitutes a rhythmic versus a non-rhythmic interaction. A rhythmic interaction is one where two conspecifics adjust their individual timing to each other to create temporal regularities that facilitate the interaction. The rhythmic structure that emerges can be measured and quantified along sequential and temporal dimensions (Ravignani & Norton, 2017). For instance, in piano duets and dyadic finger tapping experiments, a statistical association, either synchronous and/or asynchronous, can be established between the temporal intervals of the two players (Goebl & Palmer, 2009; Loehr & Palmer, 2011). Based on the concepts described in Table 2.1, any interaction that requires timing adjustment between two conspecifics, such as dyadic synchrony and turn-taking, is considered rhythmic in this paper. A non-rhythmic interaction is one where there is no timing adjustment between the members of the dyad. Despite this, individual rhythms can still play an important role in regulating the interaction. For example, male northern elephant seals (*Mirounga angustirostris*) can recognise the rhythmic structure of their rivals’ vocalisations; depending on the individual’s status within the colony, males will either ignore or move away from the vocalizing male (Mathevon et al., 2017). In the absence of a temporal relationship between the signals of two conspecifics, both alternation and antiphony are considered non-rhythmic. In the supplementary materials, readers can find examples of both interactions for all animal clades discussed in this review (Table S2.1) and a discussion on non-rhythmic interactions.

Humans

Humans are a highly social species and perform many types of duets; here, we focus on two: speech and music. In the rhythmic domain, speech generally involves turn-taking, whereas music is typically performed in synchrony, although people can chant speech together and call-and-response is used in music. Such behaviours involve complex social interactions and serve to communicate information, express emotion and socially bond (Feldman, 2007; Juslin & Sloboda, 2013). These acoustic behaviours are performed with others in complex behavioural patterns involving cues including gesture, touch, body sway and dance, but also facial expressions and gaze direction (Chang et al., 2019; Chauvigné et al., 2019; Goldin-Meadow, 1999; Holler et al., 2014).

Infancy and childhood

Human infants cannot survive without carers for a considerable period, and their early development and learning are done in a social context, often primarily with the mother (Feldman, 2007). Within this social context, movement behaviours (Tronick & Cohn, 1989), vocal behaviours (Beebe et al., 1988), gaze (Leong et al., 2017), autonomic functioning (Leclère et al., 2014) and hormone expression (Feldman, 2017) are coordinated in time between infants and their carers. This early temporal coordination may enable infants to (i) regulate physiological and behavioural processes for survival (Feldman, 2017), (ii) develop self-regulation (Feldman et al., 1999), (iii) differentiate the emergence of their self from others (Feldman, 2007), (iv) begin building social and empathetic relations with others (Trainor & Cirelli, 2015), and (v) learn speech and music through continual reciprocal adjustments (Lytle & Kuhl, 2017).

Carers provide rhythmic input to their infants in the forms of singing, patting and rocking, which helps infants to regulate their states (Trehub, Ghazban, et al., 2015). The periodicities found in rhythms of infant-directed singing provide a context for bidirectional entrainment between infants and carers, which can be seen in the coordination of autonomic and brain responses between mothers and infants (Cirelli et al., 2020; Leong et al., 2017; Patel, 2010). Synchronous movement to music during infancy also has important social consequences. Fourteen-month-old infants who are bounced in synchrony to music with an experimenter are much more likely to subsequently help that experimenter compared to infants who are bounced out of sync (Cirelli et al., 2018; Trainor & Cirelli, 2015), suggesting that the origins of empathy and friendship may be found in coordinated rhythmic behaviours.

While the bidirectionality of interactions begins in infancy, carers initially have a broader role in structuring and scaffolding the interaction (Harrist & Waugh, 2002; Maccoby, 1994). As children develop their advancing language, social and cognitive skills support and expand their active participation in complex, rhythmic verbal and non-verbal interactions (Harrist & Waugh, 2002). In the verbal domain, children become increasingly proficient at dialogue and conversation (Casillas & Frank, 2017; Harrist & Waugh, 2002). Active turn-taking during conversation, which reflects partners' attunement to each other, is associated with children's language and social skill development (Stanton-Chapman & Snell, 2011; Tomasello & Farrar, 1986). Children who experience more appropriate turn-taking styles during conversations with their parents are more liked by their peers (Black & Logan, 1995), and also exhibit more appropriate turn-taking with their peers even when the peers are novel acquaintances (Black & Hazen, 1990). In the non-verbal domain, unfamiliar child peers who actively engage in coordinated rhythmic movement activities together show greater cooperation (Kirschner & Tomasello, 2010) and perceive that they are closer and more similar to each other, potentially due to increased intentional communication (Rabinowitch & Knafo-Noam, 2015).

Adulthood

In adults, entrainment during dyadic conversation is observed through temporal adaptations within each individual's speech pattern that occur over multiple timescales (Abney et al., 2014). The precise timing involved in conversational entrainment suggests the presence of underlying rhythmic processes that allow for accurate timing predictions (Wilson & Wilson, 2005). During a conversation, dyads converge their individual speech rhythms as measured via speech rate, prosody and respiratory movement (Buder et al., 2010; McFarland, 2001; Schultz et al., 2016), as well as turn-taking timing (e.g. by minimising silent gaps and avoiding overlap; O'Dell et al., 2012; Stivers et al., 2009; Wilson & Wilson, 2005). The degree of entrainment in conversation is impacted by the interaction context (e.g., friendly versus unfriendly; Abney et al., 2014). Rhythms are crucial in these interactions as their temporal regularity enables prediction and sampling of environmental stimuli (Haegens & Zion Golumbic, 2018).

Coordination in both musical and conversational duets relies on visual and auditory cues (Bishop & Goebl, 2015; Moran et al., 2015) but, in contrast with speech which largely involves turn-taking, musical interactions typically involve synchrony. While the two interacting musicians may sing or play different pitches, their outputs must fit together both rhythmically and harmonically to create a single meaningful joint performance. The high real-time temporal demands of duet music-making require continual adjustments, anticipation and prediction (if an individual reacts to rather than anticipates their partner's output, they will be late and not in sync with them; Keller et al., 2014). The continual adaptation of two people tapping together can be seen in "lag-1 correlations", whereby on a given tap, the individual who is slightly ahead will slow down on their next tap and the individual who is slightly behind will speed up on their next tap (Repp & Su, 2013). Studies of string quartets also show mutual adjustments of timing, with some quartets being more leader-driven and others more egalitarian (Gilboa & Tal-Shmotkin, 2012; Timmers et al., 2014; Wing et al., 2014). However, asynchronies increase when one musician in a duet is replaced with a recording (Demos et al., 2017) and when tempo preferences of two musicians are divergent (Zamm et al., 2016), highlighting the importance of bidirectional coordination. Musicians convey their upcoming intentions through body sway, similar to how people use hand gestures when they speak. Findings from string quartets show that the body sway of one musician predicts the upcoming body sway of another musician (Chang et al., 2017). In short, the better the overall communication is among group members, the higher the quality of their performance is rated (Chang et al., 2019). Music affords an ideal context in which people can synchronise their movements, and when a person experiences even a short period of synchronisation with another person, it has social consequences, leading to increased liking, cooperation and trust (Hove & Risen, 2009; Wiltermuth & Heath, 2009). These social consequences may explain why music-making in groups is universal across human societies and is likely an evolutionary adaptation (Savage et al., 2020).

Non-human primates

Dyadic interactions among primates are essential to determine hierarchies within stable groups (Cheney et al., 2016), and establish alliances and partnerships (Silk et al., 2003). In several monkey and ape species, where vocal communication follows precise temporal and social rules, strong social bonds shape interactive signalling patterns (Lemasson et al., 2011; Lemasson & Hausberger, 2004). For instance, bonobos respond preferentially to conspecifics with whom they have close bonds, and take turns to avoid overlap when vocalizing with those (Levréro et al., 2019). Unfortunately, rhythmic features have generally been understudied in primate communication, with few exceptions.

Primates may form long-lasting, socially monogamous pairs with opposite-sex conspecifics. Tarsiidae, Callicebinae, Hylobatidae and Indriidae include the main species of pair-living primates producing duets (Geissmann, 2000; Torti et al., 2017). These pair-living primates face the need to defend a reduced home range (Bonadonna et al., 2017) and males within these pairs struggle to protect mating exclusivity (Bonadonna et al., 2019). In both gibbons and indris, duets serve to inform neighbouring groups about the occupation of a territory and to defend it during group encounters (Gamba et al., 2016). In indris, duetting mediates group cohesiveness (see Figure 2.1c; Torti et al., 2013), while in gibbons, it can inform about the presence of particular predators (Clarke et al., 2006). Duetting in siamang pairs (*Hylobates syndactylus*) is associated with a rhythmic swinging from branch to branch, which helps maintain entrained vocal displays between males and females (Badraun et al., 1998). Depending on the species, simultaneous singing or avoiding song overlap advertises the strength of a bond and may predict both behavioural coordination and grooming rates (Geissmann & Orgeldinger, 2000).

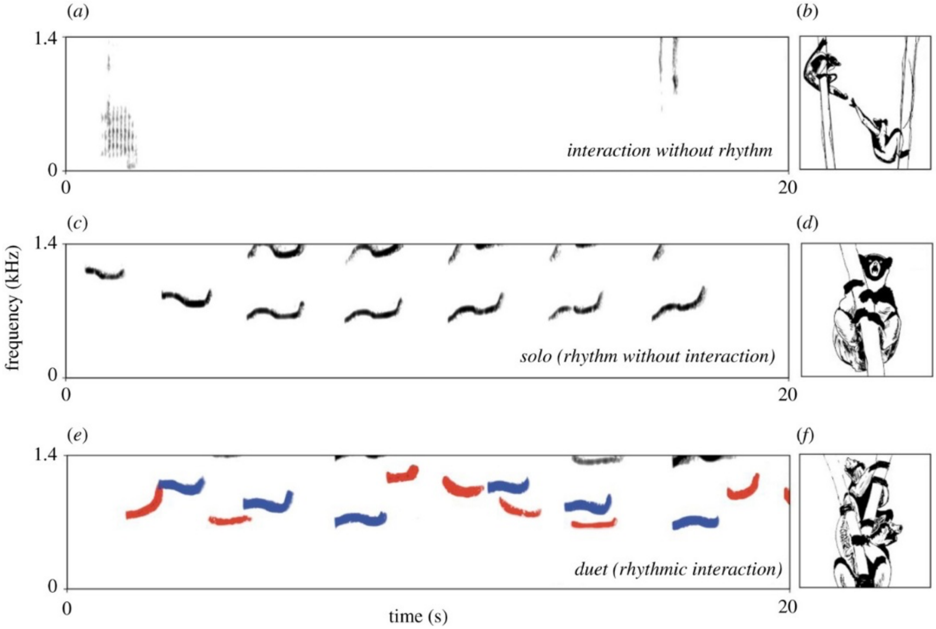


Figure 2.1. Rhythm and social interactions in indris. (a) A spectrogram of a vocal interaction between two males from the same family group that compete for food; the adult male emits a low-pitched grunt (Maretti et al., 2010) followed by a kiss and a wheeze from the younger male (Valente et al., 2019). (c) A spectrogram of part of an unusual solo song by a male indri. (e) A spectrogram of part of a duet by a pair of indris (cohesion song). After dispersing within a territory, they emit a particular song type to regroup in a particular location (Torti et al., 2013). Red shading denotes the fundamental frequency of the female's calls, and blue shading denotes the fundamental frequency of the male's units. (b), (d), and (f) are sketches of the animals represented in the adjacent spectrograms.

Non-primate mammals

Rhythmic interactions in other mammalian species allow individuals to identify conspecifics and maintain affiliations with other group members. We summarise findings of interactive exchanges in animal groups inhabiting different physical and socio-ecological environments including cetaceans, pinnipeds, bats, rodents, elephants, antelopes and meerkats.

Cetacean vocal interactions span alternation and synchrony. Bottlenose dolphins exchange signature whistles for individual recognition and group cohesion (M. C. Caldwell & Caldwell, 1968; Lilly & Miller, 1961). These exchanges are closely coordinated in time such that acoustic overlap is minimal (silent gaps < 1 s; Nakahara & Miyazaki, 2011). Male bottlenose dolphins also engage in coordinated cooperative interactions when coercing females by synchronising their threat vocalisations and matching tempos (Moore et al., 2020).

In pinnipeds, mother–pup recognition is a common dyadic interaction, where mothers and pups use vocal cues to identify and localise each other in the breeding colony, though species differ in whether mother, pup or both individuals vocalize (Charrier, 2020; Insley et al., 2003). Harbour seals may also adjust their call timing and call asynchronously relative to conspecific vocalisations (Ravignani, 2019). Within a breeding colony, this turn-taking strategy may avoid acoustic overlap and make individual pup calls more conspicuous.

Several bat species engage in antiphonal exchanges of stereotyped calls, also for mother–pup recognition (Balcombe & McCracken, 1992; de Fanis & Jones, 1996; Esser & Schmidt, 1989). Moreover, adult white-winged vampire bats (*Diaemus youngi*) respond to contact calls in a duet-like fashion by temporally coordinating their reply (silent gaps < 500 ms), potentially to monitor the spatial positions of conspecifics (Carter et al., 2009).

Few studies reported on rhythmicity in rodent dyadic communication. In Alston's singing mice (*Scotinomys teguina*), depending on the social context, males adjust their signal timing to sing in turns and avoid acoustic overlap (silent gaps around 500 ms; Okobi et al., 2019). Naked mole-rats (*Heterocephalus glaber*) produce soft chirps antiphonally between two or more individuals to identify their social status and maintain affiliations (silent gaps < 400 ms; Yosida & Okanoya, 2009). Middle East blind mole-rats (*Spalax ehrenbergi*) communicate using vibratory signals by engaging in alternating head-drumming duets (silent gaps < 2 s; Rado et al., 1987).

Data on interactive temporal coordination in dyads are limited for the remaining mammal species. Female African elephants (*Loxodonta africana*) use antiphonal rumbling sequences between group members to maintain social distance (Leighty et al., 2008). An African antelope, the klipspringer (*Oreotragus oreotragus*), engages in alarm call duetting and calls are produced in alternation, with female calls closely following those of males (Tilson & Norton, 1981). Finally, meerkats (*Suricata suricatta*) avoid overlapping conspecifics in low-conflict group sunning calls by vocalizing in turns (Demartsev et al., 2018). Here, group turn-taking is an outcome of dyadic interactions between group members and mechanistically relies on two alternating processes: call inhibition and call excitation.

Birds

Duets, which require a certain degree of temporal coordination (Figure 2.1e), are also widespread across birds, especially in mating contexts. Vocal or dance duets occur in 18–20% of all avian species (J. A. Tobias et al., 2016), which accounts for mostly song-duetting species among songbirds and dance-duetting species among non-songbirds (but see Malacarne et al., 1991; Soma & Garamszegi, 2015). Avian duets are generally performed by paired partners or prospective mates, and serve different functions including mate-guarding, joint resource defence and/or mutual courtship (Hall, 2004; Langmore, 1998).

Some barbet species (Capitonidae; Short & Horne, 1983) and the magpielark (*Grallina cyanoleuca*; Ręk & Magrath, 2016) perform multimodal duets, where the two birds simultaneously coordinate vocalisations and body movements. For singing, species-specific duet rules are well documented and show interspecific variation, even among closely related species. For example, in *Thryothorus* wrens, some species produce perfect antiphonal duets without overlap, while others sing in synchrony (Mann et al., 2009). In avian song duetting, each of the sexes often produces a particular phrase at the precise onset or offset of its partner's singing, making it a perfectly timed collective display (Brumm & Slater, 2007). By contrast, rhythmicity in dance duets is still relatively unexplored. For example, in the red-crowned crane (*Grus japonensis*), males and females exhibit sequences of multiple dance elements for their joint display (Takeda et al., 2019), but it is not known whether the paired cranes precisely time and synchronise their dancing with each other.

The degree to which two individuals coordinate their signalling during a duet varies even among dyads of the same species. Although temporal coordination can greatly influence reproductive fitness as better-coordinated signals are more effective for securing mates or reproductive resources (Ręk, 2018), not every pair will be able to achieve the same degree of coordination. In fact, high-level coordination can only be achieved by pairs with longer partnerships (Hall & Magrath, 2007). Indeed, newly established pairs of canebrake wrens (*Cantorchilus zeledoni*) improve the coordination of their duets over time, suggesting that learning enables fine duetting (Rivera-Cáceres et al., 2016).

Multimodal signal coordination in dyadic interactions is not restricted to species known for their duetting. The zebra finch has been intensively studied and disproportionately so relative to other songbirds, but lacks duetting as only males sing (Griffith & Buchanan, 2010). Despite this, mated pairs display tight temporal coordination in both visual and auditory modalities. Zebra finches form strong dyads with a lifelong mate (Hall & Magrath, 2007) where coordinated behaviours serve as an honest signal of the pair's coalition quality. Indeed, the strength of the bond in zebra finch pairs predicts the degree of coordination of movements (Prior et al., 2020) and call exchanges (D'Amelio et al., 2017), which mechanistically rely on predictive timing (Benichov et al., 2016). Pairs use such vocal exchanges to maintain a joint behavioural state (Perez et al., 2015) and coordinate shared parenting duties (Boucaud et al., 2017).

Anurans

Dyadic interactions in anurans occur in mating contexts, where the same-sex duets serve to compete for access to mates and opposite-sex duets serve to attract potential mates (Grafe, 2005). Anurans show a precedence effect, where calls of leading individuals are preferred when two identical calls are presented in close succession (Greenfield, 2005), but exceptions exist (Bosch & Márquez, 2001). Moreover, calls that alternate in precise phase relationships are preferred by females of some species (e.g., the midwife toad *Alytes obstetricans*; Bosch & Márquez, 2001). Precise patterns of signal timing largely influence mate choice and are thus under strong sexual selection in anurans (Grafe, 2005). Males typically form large groups and produce loud alternating advertisement calls to attract females for mating (Wells & Schwartz, 2006). Instances of non-random call timing in males were studied in larger groups, but also in dyads (Wells, 1977). In duetting male pairs of the Neotropical toad (*Rhinella ocellata*), males call in alternation and avoid overlap with each other (J. P. Caldwell & Shepard, 2007). Moreover, the call delay of the responding male varies depending on the call duration of the male that initiated the interaction. This alternation pattern shows a high degree of temporal coordination within bouts. Call timing adjustment between two males of the European tree frog (*Hyla arborea*) can result in near-perfect antiphonal calling with the mean relative phase angles distributed around 180° (Klump & Gerhardt, 1992). The onset of vocalisations can even be adjusted based on the distance separating individuals, as evidenced in bullfrogs (*Rana catesbeiana*; Boatright-Horowitz et al., 2000).

Bullfrog males call more frequently following calls of distant neighbours than of those nearby; nearby neighbours may inhibit the vocal response of the focal individual. In sum, males try to avoid signal overlap in an effort to increase their conspicuousness by paying selective attention to their close neighbours and vary their call timing accordingly (Klump & Gerhardt, 1992). However, precise alternation of vocalisations is not always the norm. A few species signal using overlapped calls. For example, males of the American toad (*Bufo americanus*) signal in synchrony or near-perfect synchrony (Grafe, 1999). This timing strategy may help reduce predation risk (Tuttle & Ryan, 1982) and/or increase the chorus' audibility in order to attract more females to the breeding area (Wells, 1977). Even though they are rarely described, duets between males and females do occur at the beginning of courtship. In the South African clawed frog (*Xenopus laevis*), receptive females produce a vocalisation named "call rapping", which is composed of a rapid series of loud clicks, in response to the male's advertising call (M. L. Tobias et al., 1998). The female call spurs the male to move towards the sound source and produce an intense bout of calling within 1–2 s of the female's signal. A possible explanation for the evolution of female courtship vocalisations in anurans could be found by looking at similarities in terms of breeding biology (M. L. Tobias et al., 1998).

Insects

Observations of non-random timing suggest that insects can adjust the onset of their signals. Temporal relationships of interactive signals are described for the acoustic modality in bush-crickets, grasshoppers, mosquitoes and flies, for the vibratory modality in planthoppers, leafhoppers, stoneflies, stink bugs and lacewings, and for the bioluminescent modality in fireflies (Bailey, 2003; Greenfield, 1994a). Generally, males initiate insect duets and females respond with a fixed latency relative to the male signal (Bailey, 2003). The temporal pattern of the duet and the time window of the female reply are species-specific (Zimmermann et al., 1989), hence allowing species recognition (Bailey, 2003). The duration and complexity of the male signal affects the reply latency of the female (Bailey & Hammond, 2003). The longer and more complex the male signal is, the more time is needed to process the information encoded in the call and assess the male's quality. However, the latency of the reply may also depend on the female's readiness to mate (Bailey & Hammond, 2003).

In the leafhopper *Aphrodes makarovi*, sexual communication and mate recognition are mediated via species-specific and sex-specific vibrational signals. During duetting, the female's response overlaps with the last portion of the male's signal (Kuhelj et al., 2015). Males can only evaluate the non-overlapped part of the female call, and consequently adjust their signal period length to the duration of the female's reply to obtain a longer non-overlapped segment of the female's call (Kuhelj et al., 2016). While competing for access to females in the presence of rivals, male broad-winged bush katydids (*Scudderia pistillata*) produce an acoustic tick that mimics the female's tick signal and response timing (Villarreal & Gilbert, 2014). This indicates to the female when the call has concluded and also confounds eavesdropping males by making it difficult for them to accurately move towards the true female's sound source (Bailey & Hammond, 2003). Not only does this reduce the risk of competing males interrupting the established duet, but it also increases the coordination of the duet itself.

Some insects perform multimodal duets. For temporal coordination to occur in duets of the fruit fly *Drosophila virilis*, females need to detect the male's sound cues and be in close physical proximity (LaRue et al., 2015). During courtship, males tap the female's abdomen and lick the genitalia in a precisely timed manner to coordinate the duet. Females may choose to mate only with males that provide multiple timing cues during the duet. In the katydid *Onomarchus uninotatus*, male and female signals alternate in a clear phase relationship. Males produce a low-pitched call, which receives a female vibrational signal in response (Rajaraman et al., 2015; see Figure S2.1). This unique multimodal duet may have evolved in response to predation; silent flying females were predated at a higher rate than calling males. With females using vibrational signals, the roles are now reversed: males must search for females by moving towards the vibration source and face increased risk of predation by bats.

The same-sex dyadic interactions in insects occur in highly competitive environments, hence the resulting temporal patterns are an outcome of competition between signalling males. In species where females prefer to mate with males leading the call sequence, males alternate calls to avoid losing their leader role (Greenfield, 1994a, 2005). In tarbush grasshoppers (*Ligurotettix planum*), males engage in acoustic "fights" to defend or conquer a mating territory. The male that cannot match the signalling rate and length of the signal of its opponent gives up the fight (Greenfield & Minckley, 1993). The aggressive calls emitted by competing males thus represent an honest signal of male fitness.

Ultimate cause of dyadic interactions: function and phylogeny

Dyads can consist of mates, siblings, parent and offspring, or any two individuals from the same group. Dyadic interactions serve a variety of functions including social bonding, sharing emotion, establishing hierarchies and partnerships, mate-guarding, courtship, joint resource defence, etc. (Cheney et al., 2016; Feldman, 2007; Hall, 2004, 2009; Juslin & Sloboda, 2013; Langmore, 1998; Lytle & Kuhl, 2017; Silk et al., 2003). Studying dyads means that experiments possibly neglect relevant group dynamics, especially in highly social species. However, it also enables researchers to study semi-natural behavioural interactions under more controlled experimental conditions.

In this review, we identified two settings in which dyadic behavioural interactions can occur: social and sexual. Helping behaviour, learning behaviour and parent–offspring recognition were classified as occurring in social settings. Dyadic interactions in sexual settings mostly pertain to opposite-sex conspecifics that are potential mates or already form an established pair, but they can also involve same-sex conspecifics during intrasexual competition occurring during mate search and attraction (Bosch & Márquez, 2001). Rhythmic behaviours have been reported for both contexts in mammals and birds, but in anurans and insects, they have only been reported for sexual settings.

Animals produce many different types of rhythms; hence, they could entrain to conspecific rhythms in ways that can be difficult to observe. For instance, individual rhythms do not need to occur in the same modality to interact rhythmically (i.e., coordination of an infant's brain responses to its mother's singing (Cirelli et al., 2020). Moreover, the dynamics of the interaction can be both unidirectional and bidirectional (Chang et al., 2019; Gilboa & Tal-Shmotkin, 2012; Timmers et al., 2014; Wing et al., 2014). In music, bidirectional interaction improves coordination, while unidirectional interactions create larger asynchronies (Demos et al., 2017). Does this finding in humans carry over to interactions in other species? One could test whether the degree of coordination is affected by unidirectionality in songbird species in which one of the sexes does not sing (e.g. the Java sparrow or zebra finch). It seems that several species overcome such asynchrony by signalling in more than one modality to achieve higher levels of temporal coordination.

Even though multimodal signals are widespread across numerous taxa (Badraun et al., 1998; Bishop & Goebl, 2015; LaRue et al., 2015; Ręk & Magrath, 2016; Short & Horne, 1983), their modalities are often studied separately (Halfwerk et al., 2019). Rhythmic interactive signalling has been reported mainly for the acoustic and visual modalities (Table 2.2). However, birds can communicate using vibratory signals (Ota et al., 2015), but avian rhythmic interactions in the tactile modality have not been described. Similarly, anurans have visual signals (Grafe & Wanger, 2007), but rhythmicity in this modality has not been described during conspecific interactions. Future work on interactive rhythms should investigate the unexplored modalities in these animal groups. In addition, analysing rhythmic patterns of multimodal signals during interactive communication will provide more insight into the functions of signal timing.

Table 2.2. Summary table of rhythmic interactive signalling in dyads across taxa. Although many signalling modalities have been proposed, the distinctions between them are often murky, hence, in the spirit of simplicity and to avoid confusion, we only include the main ones: auditory (A), visual (V) and tactile (T). Bioluminescent signals enhance visual perception (Jeng, 2019) and are grouped under the visual modality. Vibratory signals are substrate-borne signals that are perceived using mechanoreceptors (Tuthill & Wilson, 2016), similar to touch, hence they are considered part of the tactile modality. Cases where no empirical evidence was found are represented by a long dash (—).

Composition of dyad		**Interaction context**	**Taxonomic group**			
			Mammals	*Birds*	*Anurans*	*Insects*
Kin	Parent–offspring	Social	A, V	A, V	—	—
	Social group member	Social	A, V	A, V	—	—
Non-kin	Social group member	Social	A, V	A, V	—	—
	Male-female	Sexual	A, V	A, V	A	A, V, T
	Male-male	Sexual	A, V	A, V	A	A, V, T

Proximate causes of dyadic interactions: ontogeny and mechanisms

A wide repertoire of signals is needed to become an active, grown-up participant in dyadic interactions and precise timing of signals requires extensive practice. The social environment during development plays a crucial role in shaping adult signals. Human children benefit from structured bidirectional interactions scaffolded by their carers (Harrist & Waugh, 2002;

Maccoby, 1994). Similarly, young birds learn to participate in adult interactions. Male zebra finches learn their courtship song from their father, but can also receive feedback from other individuals within their social environment (Carouso-Peck et al., 2020). It has even been argued that learning of species-specific rhythms starts before birth, with possible long-term effects on vocal and social development (Lampen et al., 2019). The social context experienced during development modulates male signal features and female mate preferences in insects, which will constrain the interaction (Desjonquères et al., 2019). Future work on dyadic interactions should accurately describe the social environment in which animals have been raised, as it may influence both rhythm perception and production. Moreover, longitudinal studies should aim to follow developing individuals to investigate the effect of development on rhythmic interactions.

Coordinated signal exchanges in dyadic interactions show reply latencies ranging from the order of milliseconds to seconds (Carter et al., 2009; Nakahara & Miyazaki, 2011; Okobi et al., 2019). The precise temporal coordination of behaviours must be regulated by the presence of underlying rhythmic processes, and many timing mechanisms proposed in the literature are based on the concept of coupled oscillators (Wilson & Wilson, 2005). Human behaviours such as conversational entrainment (Abney et al., 2014) or the synchronous playing of music (Keller et al., 2014) rely on mechanisms that allow us to make accurate timing predictions about the onset of upcoming signals. Pinnipeds and birds also show some affinity with predictive timing. Similar to conversational turn-taking, seal pups may adapt the timing of their calls (Ravignani, 2019), and some songbirds attempt to simultaneously synchronise their vocalisations and body movements (D’Amelio et al., 2017; Prior et al., 2020; Short & Horne, 1983). Insects are also capable of synchronous displays in larger groups, but the neural processes that regulate signal timing are different from those of mammals and birds. Instead of being able to predict the exact onsets of events, signalling in anurans and insects is reactive (Greenfield, 1994b). Similar observable and interactive behaviours can thus be produced using different mechanisms, but we do not know whether these mechanisms are learned and whether they change across modalities.

Conclusion

Rhythmic behaviours are widespread among animal clades and are crucial in regulating social interactions in dyads and larger groups. Unfortunately, in the temporal domain of animal communication, many species remain unexplored and several questions remain unanswered. Within species and across contexts, which features of signal timing are stable, and which are constrained, and to what degree? How are timing mechanisms learned, and do they change from one modality to the next? Lines of investigation that integrate ecological and neuroethological perspectives have begun to resituate rhythmic behaviours within animal communication systems (see Hoffmann et al., 2019), but they have not been described in this paper. Such a multidisciplinary approach would allow researchers to design species-specific experiments to infer how rhythm functions during interactive signalling (Henry et al., 2021). We thus strongly encourage future studies on interactive rhythmic behaviours in the hope of ultimately developing an integrative cross-species framework (Anichini et al., 2020). The comparative method could then provide crucial insights into the evolution and adaptive functions of interactive rhythmic behaviour across taxa (Ravignani et al., 2014; Ravignani, Verga, et al., 2019).

Acknowledgments

We thank the organisers of the Lorentz workshop on Synchrony and Rhythmic Interaction: From Neurons to Ecology, the Lorentz Center, and the NIAS, ILLC and ANU. This work was supported by the Research Foundation Flanders (FWO) [G034720N] to Bart de Boer.; JSPS Grant-in-Aid for Young Scientist [16H06177] and for Scientific Research [20K06809] to M.S.; Junior Fellowship of the Hanse-Wissenschaftskolleg (Institute for Advanced Study) to M.A.; NIMH and NCCIH [R61MH123029] and NIDCD [R21DC016710] to M.L.; the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2019-05416], the Canadian Institutes of Health Research (CIHR) [RTI-2017-00643] and the Canadian Institute for Advanced Research to L.J.T. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions

K.d.R. conceived the manuscript. All authors contributed to writing and revising of the manuscript, and gave final approval for publication.

Supplementary materials

Table S2.1. Examples of rhythmic and non-rhythmic interactions. All non-rhythmic interactions in the table include a rhythmic component. The non-rhythmic interactions described function to maintain affiliations, court prospective mates, and compete with conspecifics. It is possible that if these behaviours are performed in a different context, the receiver will adjust its own rhythm to that of the producer and the interaction then becomes rhythmic. The rhythmic interactions described function for social bonding, increase individual conspicuousness and mate acquisition.

Clade	Non-rhythmic interaction	Rhythmic interaction
Humans	Waving to someone and receiving a smile in return	Neural rhythms of infants entraining to the rhythms of an experimenter bouncing in synchrony with a beat (Trainor & Cirelli, 2015)
Non-human primates	Chest-beating displays in gorillas (Schaller, 1963)	Siamang pairs swinging from branch to branch to maintain entrained vocal displays during duets (Badraun et al., 1998)
Non-primate mammals	Pulsed calls of northern elephant seal for broadcasting social status to nearby males (Mathevon et al., 2017)	Harbour seal pups calling asynchronously relative to calls of conspecific neighbours (Ravignani, 2019)
Birds	Female palm cockatoos observing drumming sequences of males performed during courtship (Heinsohn et al., 2017)	Pairs of the magpie lark simultaneously coordinating vocalisations and body movements (Ręk & Magrath, 2016)
Anurans	Male frog foot-flagging during agonistic interactions with intruders (Grafe & Wanger, 2007)	Males of the American toad calling in synchrony (Grafe, 1999)
Insects	Female katydids, *N. spiza*, orient towards males that produce the leading signal in the calling sequence (Greenfield et al., 1997)	Males of the leafhopper, *A. makarovi*, adjusting the period length of their signals to match the duration of the female's reply during duets (Kuhelj et al., 2016)

Non-rhythmic interactions

A non-rhythmic dyadic interaction is one where there is no coordination between the individual rhythms of the two interacting agents. For instance, male indris competing for food vocalise during agonistic interactions with no rhythmic pattern emerging from the signalling (see Figure 2.1a). However, individual rhythms still play an important role in regulating social interactions. For example, some whale species alternate rhythmically structured calls, matching the type and structure of the call they respond to. Such call sequences can function either as individual signatures, as evidenced in beluga whale (*Delphinapterus leucas*) burst-pulse sounds (Morisaka et al., 2013), or group-identifying calls, as evidenced in killer whales (*Orcinus orca*; Miller et al., 2004) and sperm whale (*Physeter macrocephalus*) click sequences (Schulz et al., 2008). Sperm whale mothers and calves also exchange a different repertoire of click patterns, potentially used for individual identification (Schulz et al., 2011). In the lesser spear-nosed bat (*Phyllostomus discolor*), pups continuously adapt the frequency-time structure of their isolation calls during development to match that of their mothers (Esser & Schmidt, 1989). This acoustic matching may also play a role in mother–pup recognition. Findings in harbour seal pups (*Phoca vitulina*) also show that the rhythmic structure of mother attraction calls resembles the temporal structure of vocalisations in other species known to communicate interactively (Ravignani, Kello, et al., 2019). Although signal timing varies across species, this highlights that individual rhythms are essential in regulating non-rhythmic interactions by helping to identify individuals and groups, inform about the caller’s social status, and probably more.

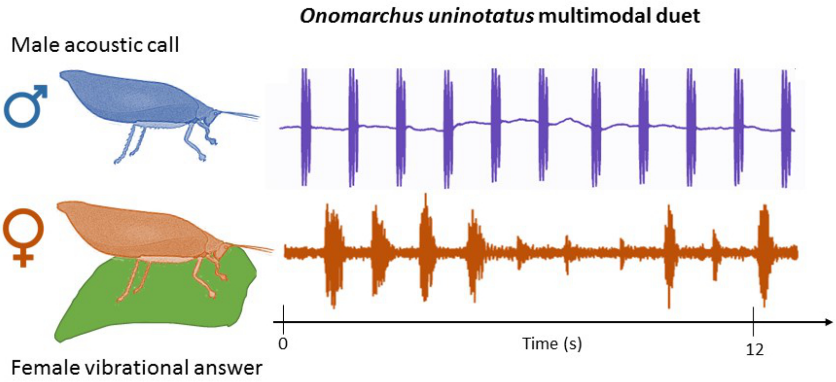


Figure S2.1. Alternating phase relationships found in a multimodal duet of *Onomarchus uninotatus*. Adapted with permission from The Journal of Experimental Biology (Rajaraman et al., 2015).

Chapter 3

3 | Vocal turn-taking in harbour seal pups

Abstract

Turn-taking is a fundamental feature of social communication requiring prediction and adjustment. It enables coordinated social interactions that are characterised by short silent gaps and minimal overlap between turns. This study investigates the presence of turn-taking behaviour in the vocal interactions of harbour seal (*Phoca vitulina*) pups. Consistent with findings from other species, including humans, we found that their vocal interactions are tightly timed, with short response latencies (median = 287.0 ms) and systematic avoidance of overlap, indicating temporal coordination that may facilitate social interactions. Moreover, a negative lag-1 autocorrelation between response latencies, where shorter silences are followed by longer silences and vice versa, suggests that pups actively adjust their call timing based on feedback from the preceding turn. Our playback experiment further revealed that pups adapt their call timing to the tempo of the interaction and duration of conspecific signals, underscoring the role of social dynamics in shaping turn-taking behaviour. Collectively, these findings confirm vocal turn-taking in harbour seal pups and contribute to the growing body of comparative research on this behaviour, highlighting its importance in social interaction and shedding light on the evolutionary connections between human language and animal communication.

Adapted from: de Reus, K., Leonetti, S., Salazar-Casals, A., Keller, P. E., de Boer, B., Holler, J., & Ravignani, A. (2025). *Vocal turn-taking in harbour seal pups.* [Manuscript in preparation].

Introduction

The “interaction engine” hypothesis suggests that interpersonal coordination during human communication relies on a specialised set of social cognitive skills (Levinson, 2006). A key skill is turn-taking: the ability to regulate the temporal structure of social interactions using a variety of cues (e.g., Holler et al., 2018; Keitel & Daum, 2015). Turn-taking enhances the flow of conversation, establishes shared attention, and facilitates coordination. In particular, it allows a rapid exchange of communicative turns, with minimal gaps and overlap (Sacks et al., 1974). Across 10 studied languages, the average silent gap between a question and its response is around 200 ms (Stivers et al., 2009). Remarkably, this rapid exchange is possible despite response latencies for word production averaging around 600 ms (Indefrey & Levelt, 2004), suggesting that turn-taking relies on predictive mechanisms. Such mechanisms enable speakers to plan their next turn while listening to their conversation partner (Corps et al., 2018) and anticipate when to begin speaking based on the integration of signals from several modalities (de Ruiter et al., 2006; Kendrick et al., 2023). Infants begin developing turn-taking skills around 3 months of age, with caregivers synchronising smiles, vocalisations, and facial expressions with infant cues to foster communication and bonding (Bateson, 1975; Hilbrink et al., 2015). Turn-taking thus supports effective communication, strengthens social bonds, and facilitates human development, underscoring its central role in coordinating social interaction.

The study of turn-taking has recently expanded beyond humans to encompass a wide range of non-human animals, including mammals, birds, amphibians, and insects (Abreu & Pika, 2022; de Reus et al., 2021; Demartsev et al., 2018; Pika et al., 2018; Takahashi et al., 2013, 2016). Turn-taking studies in non-human animals enable researchers to adopt a comparative, cross-species approach, providing valuable insights into the communication modalities, social contexts, and cognitive mechanisms that are shared across species. Turn-taking occurs across diverse communication modalities, such as vocalisations, gestures, and bodily signals (Pika et al., 2018). It also plays a role in various behavioural contexts, such as courtship (Hall, 2004), mate attraction (Buck & Buck, 1978), territorial defence (Greenfield & Roizen, 1993), as well as cooperative contexts like social bonding (Demartsev et al., 2018) and group coordination (Leighty et al., 2008), highlighting its significance in social interactions. For example, gestural turn-taking in chimpanzees seems to facilitate cooperation and social bonding (Badihi et al., 2024). Notably, gestural latencies vary across chimpanzee communities and align with the 200 ms gaps observed in humans (see Stivers et al., 2009), reflecting remarkable similarities with language. Adopting a comparative approach to turn-taking can thus potentially shed light on the evolutionary origins of social interaction, language, and communication more broadly (e.g., Levinson, 2016; Levinson & Holler, 2014). However, research on turn-taking in mammals remains predominantly focussed on primates (Abreu & Pika, 2022; Fröhlich et al., 2016; Takahashi et al., 2013). Future studies on non-primate mammals are therefore essential for a more comprehensive understanding of turn-taking behaviour.

Pinnipeds (seals, sea lions, and walruses) are promising candidates for comparative research on turn-taking in social interactions, particularly due to their ability to produce and perceive temporally structured vocalisations and vocal sequences (Ravignani et al., 2016). Their sensitivity to auditory temporal patterns likely evolved to optimise communication in marine environments, where temporal features are transmitted more effectively than spectral ones (Kershenbaum et al., 2016; Mulsow & Reichmuth, 2007). This sensitivity facilitates long-distance communication (Turnbull & Terhune, 1993) and improves signal detection in noisy environments (Moors & Terhune, 2004; Terhune & Ronald, 1986). Pinniped vocalisations, such as barks and growls, are characterised by temporal variation in the millisecond-to-second range (Ravignani et al., 2016). This temporal structuring plays a key role in regulating social interactions across various contexts. For example, northern elephant seal (*Mirounga angustirostris*) males recognise rhythmic features in each other's vocalisations and adjust their behaviour based on the perceived social rank of the caller (Mathevon et al., 2017). Male southern elephant seals (*Mirounga leonina*) display consistent individual variability in their aggressive vocalisations, with stereotyped temporal regularities in syllable composition that likely facilitate individual recognition (Sanvito & Galimberti, 2000). Moreover, Californian sea lions (*Zalophus californianus*) adjust the rhythm and number of barks in a sequence based on the social context, influencing the behaviour of neighbouring individuals (Schusterman, 1977). Similarly, the bark rates of Cape fur seal (*Arctocephalus pusillus*) males during territorial defence and mating signal their arousal state, influence the vigilance of nearby subadult males, and potentially reduce conflict (M. Martin et al., 2021). In mother–offspring interactions of the same species, vocal sequences exhibit asynchronous temporal patterns, possibly enhancing individual recognition and improving the chances of successful reunion in large, noisy colonies (Osiecka et al., 2022). Pinnipeds' rhythmic abilities also extend beyond vocal communication. For instance, a California sea lion famously synchronised its head movements to musical beats (Cook et al., 2013), demonstrating a capacity for predictive timing previously thought to be unique to humans and certain bird species, like sulphur-crested cockatoos (Patel et al., 2009a). Taken together, these studies highlight pinnipeds' sensitivity to rhythmic patterns and their capacity to flexibly alter the temporal structure of their vocalisations, underscoring their potential as promising comparative models for studying the evolution and mechanisms of turn-taking.

The harbour seal (*Phoca vitulina*), a phocid species, provides a valuable model for studying turn-taking and its role in social interactions, particularly during the lactation period when pups are highly vocal. During this time, pups emit mother attraction calls (henceforth, calls) with distinct individual vocal signatures, enabling mothers to recognise their pups shortly after birth (Perry & Renouf, 1988; Renouf, 1984). Pup call sequences exhibit temporal organisation that evolves throughout puppyhood and display features that may indicate interactive communication (Ravignani, Kello, et al., 2019). For example, one study used circular statistics to show that harbour seal pups time their calls asynchronously during vocal interactions, with response phase angles clustering around 90 degrees (Ravignani, 2019). This finding indicates that the pups' responses typically occur about one-quarter of the way through the interval between calls, suggesting a delayed response that may reflect coordinated communication, such as turn-taking. This delayed response was later confirmed in a second study, which demonstrated that response phase angle distributions varied depending on social context (Anichini et al., 2023). Specifically, phase angles clustered around 90 degrees during social interactions, whereas they were more variable when pups vocalized alone. Pups thus seem to strategically adjust the timing of their calls during interactions to avoid overlap. This behaviour would be particularly beneficial in noisy colony environments, as it could increase the likelihood of successful reunions with their mothers. The adaptation and anticipation model of sensorimotor synchronisation (ADAM; van der Steen & Keller, 2013), which was originally developed to test reactive error correction and predictive processes in human interpersonal coordination, was applied to harbour seals to explore the temporal dynamics of their vocal interactions (Anichini et al., 2023). Results indicated that harbour seal pups adjusted their call timing in a complementary manner during real-time interaction in pairs, with one pup adapting its timing to synchronise calling rates, while the other adjusted its timing to emphasise the distinction in calling rates. This complementarity in temporal adaptation leads to gaps in call timing that might reflect turn-taking dynamics. Notably, in a playback experiment, temporal adaptation estimates indicated that these dynamics quickly disappear when a pup interacts with a non-responsive partner (i.e., the recording). Moreover, estimates of temporal anticipation showed that pups initially predict the timing of playback calls, but this dissipates over time, possibly due to habituation, highlighting the role of social contingency in vocal interactions.

Vocal interactions in harbour seal pups do not follow a strict A-B-A-B-A-B calling pattern or perfect alternation. In many calling bouts, a pup could call several times before the other pup responds, suggesting a more flexible pattern of alternation. While previous studies suggest this pattern may still indicate turn-taking behaviour, their statistical methods cannot conclusively confirm this. Specifically, the studies rely on the inter-onset interval (i.e., the time difference between the onsets of successive calls) to calculate phase angles. In the circular statistics approach used, a phase angle of 90 degrees indicates that the timing of the response call is systematically desynchronised with the onset of the previous call. However, a 90 degrees angle does not rule out the possibility of overlap, where the response could begin before the first call ends, which is a key characteristic of turn-taking that this methodology cannot address. A methodology based on the response latency (i.e., the time difference between the offset of one call and the onset of the next) could more reliably determine whether these vocal interactions are characterised by minimal silent gaps between turns and avoidance of overlap (Sacks et al., 1974).

This study uses standard response latency analyses from conventional turn-taking research to determine whether harbour seal pups engage in turn-taking behaviour, focusing on three key characteristics: minimal gaps between turns, avoidance of overlap, and adjustment. Harbour seal pup calls are primarily described in the context of parent–offspring recognition. In this context, pups might adjust their call timing to ensure that the individually identifiable features in their calls are clearly transmitted and recognised by their mothers within the acoustically crowded colony environment, thereby increasing the likelihood of successful parent–offspring reunions (Anichini et al., 2023). However, the function of minimal gaps is less obvious in this context. In human conversation, minimal gaps are thought to enhance communicative efficiency (Levinson, 2016). If similar timing patterns are observed in seal pups, they may indicate a role in facilitating social interaction between neighbouring pups. Moreover, in human communication, rhythm and tempo shape the flow of conversation and affect the conversational style, even with subtle timing differences (Stivers et al., 2009). A playback experiment will thus investigate whether call timing in harbour seal pups is similarly influenced by factors such as rhythm, tempo, and call duration, as well as the type of interaction (e.g., 1- or 2-way communication). The final characteristic, timing adjustment, is reflected in a negative lag-1 autocorrelation of the response latencies, obtained using time series analyses. This type of autocorrelation would indicate mutual coordination of call timing between pups during vocal interactions, similar to what has previously been observed in human telephone conversations (e.g., Pouw & Holler, 2022). Specifically, it would indicate that the timing of the current turn is adjusted based on feedback from the previous turn; a relationship that can be confirmed with linear models. Together, findings of minimal gaps between turns, overlap avoidance, and adjustment could provide evidence for turn-taking in harbour seals and contribute to our evolutionary understanding of the role of turn-taking in regulating social interactions and communication across species.

Materials and methods

Study overview and subjects

This playback experiment took place at Sealcentre Pieterburen, a pinniped rehabilitation facility in the Netherlands, during the summers of 2021 (June 16–August 4) and 2022 (June 27–July 20). Harbour seal pups are born in summer, and are often admitted for rehabilitation because of being orphaned, sick, injured, entangled, or disturbed by humans (Salazar-Casals et al., 2022). This temporary captivity allows researchers to study their behaviours under controlled conditions before the seals are released back into the wild. Over two summers, 37 harbour seal pups (Table S3.1) participated in this experiment, including 5 that were housed alone and 17 in pairs, with some animals tested twice. One pup housed alone was excluded from the statistical analysis due to a lack of responses. All animals in this study were under

1 month old upon arrival, with most aged 3–10 days old. Age was estimated by veterinarians based on the condition of the umbilical cord or umbilicus (Dierauf et al., 1986).

Experimental set-up

Animals were housed individually or with a partner in units equipped with a swimming pool and a resting plateau. Playback stimuli were broadcast via a Yamaha HS7 loudspeaker (flat frequency response: 43 Hz to 30 kHz; Yamaha Corporation, Hamamatsu, Japan), connected to an iPhone 5C (Apple Inc., Cupertino, CA) using an XLR/TRS cable. The sound pressure level (SPL) was set at 87 dB to simulate a harbour seal pup calling from a 2-meter distance (Ravignani, 2019; Sauvé et al., 2015). At the start of each data collection season, SPL was measured using a Voltcraft 322 Datalog Sound Level Meter (automatic range detection: 30–130 dB, decibel weighting filter C, 125 ms evaluation window; Conrad Electronic, Hirschau, Germany). Response vocalisations and playback calls were recorded using a Sennheiser ME-66 unidirectional microphone (frequency response: 40 Hz to 20 kHz; Sennheiser electronic GmbH & Co. KG, Wedemark, Germany), fitted with an MZW-66 foam windshield to protect against water and wind. The microphone was connected to a Zoom H6 digital recorder (48 kHz sampling frequency, 24-bit quantisation) that saved audio as uncompressed WAV files (Zoom Corporation, Tokyo, Japan). Video footage was captured using a Canon LEGRIA HF-G30 camera (1080 p resolution, 50 fps; Canon Inc., Tokyo, Japan), with recordings saved in MP4 format. Since the researcher was not present during the experimental sessions, one seal was marked with a temporary orange dye to ensure accurate attribution of response vocalisations to individuals. Seals were not allowed to swim during the experiment to maximise the number of recorded responses and prevent interference with the experimental set-up. In 2022, we made a minor adjustment to the set-up, replacing the XLR/TRS cable with a much longer one to start the experiment from outside the quarantined area. Our set-up is illustrated in Figure 3.1.

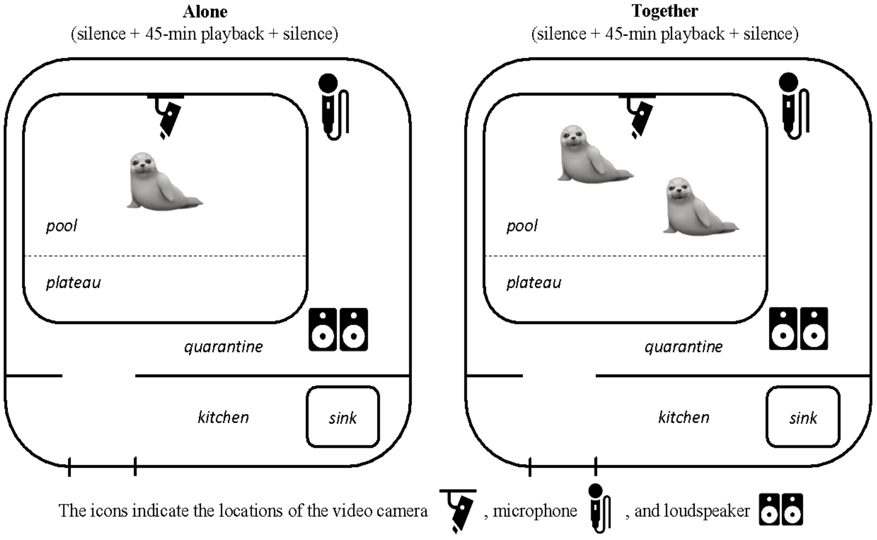


Figure 3.1. Experimental set-up shown for animals housed individually (left) or with a partner (right).

Playback stimuli

Each experimental session began with an initial period of silence lasting 45, 90 or 120 minutes, followed by a 45-minute audio playback and a variable period of silence, depending on when caretakers arrived for the next feeding (Figure 3.1). It quickly became apparent that the pups were more vocal towards the end of each session, likely due to hunger. Consequently, the initial silence duration was gradually increased, and the playback was delayed throughout the experiment to maximise the number of responses recorded during this period. During both silent periods, seals behaved spontaneously, undistracted by experimental stimuli or the presence of caretakers in the housing units. The playback period consisted of 16 randomly ordered experimental trials, with 10 seconds of silence between them. Each trial was one of 16 unique combinations of four experimental variables: (1) tempo (fast = 2 s, slow = 4 s), corresponding to the average inter-onset-interval (IOI) of the trial, (2) rhythmicity (isochronous, random), which defined the regularity of the intervals,

(3) duration[1] (normal ≈ 0.65 s, long ≈ 1.0 s) of the playback call, and (4) sex (female, male) of the pup whose call was played. Isochronous trials had regular IOIs based on tempo, while random trials had variable IOIs sampled from a normal distribution with the same mean IOI and a standard deviation equal to 20% of the IOI (Patel et al., 2009b). Each trial consisted of 51 calls, interspersed with variable silence, totalling 50 IOIs. A 15-second sound fragment was played 10 seconds before and after the playback period to measure room reverberation (see "The reverberation fragment" in the supplementary materials).

Experimental procedure

Each subject underwent one playback session per day for five consecutive days. Sessions were conducted between feeding times, either from 11 am to 3 pm or from 3 pm to 7 pm, allowing two locations to be tested daily. The experiment was set-up immediately after feeding and taken down before the next feeding, with each session lasting about 3–4 hours. Every location was consistently tested during the same time slot, ensuring that sessions were separated by approximately 24h. Seals housed alone interacted only with the playback (1-way interaction), while those housed together interacted both with the playback (1-way interaction) and with each other (2-way interaction).

Data processing and variable extraction

Video recordings were annotated in BORIS v7.10.7 by two independent annotators to ascertain which individual was calling (Friard & Gamba, 2016). Using a custom-written Python script, the video annotations were then synchronised with the audio and exported to Praat TextGrids, prior to annotating the audio. Audio recordings were annotated in the acoustic software Praat v6.1.42 (http://www.praat.org/; Boersma & Weenink, 2021). The audio annotations for every seal were set manually, whereas playback annotations were set automatically using a second custom-written Python script. Calls from seals produced when caretakers or researchers were present inside of the quarantine were not annotated. Although

---

[1] Durations were based on an analysis of over 6000 calls recorded in summer 2017 (de Reus, 2017), with calls selected from one male and one female.

the video annotations were checked for reliability between the two annotators, the audio annotations were conducted only by a single annotator due to time constraints. Data were then imported into Python using the TextGridTools package (Buschmeier & Włodarczak, 2013). The onset and offset of each call were extracted and the data was sorted in chronological order based on the call onsets.

Calling activity in harbour seal pups is bursty and vocal productions tend to occur in bouts (Ravignani, Kello, et al., 2019). The inter-onset-interval (IOI)—the difference between the onsets of two adjacent calls—was computed to help identify calling bouts (Figure 3.2). Bouts are vocal sequences of at least three calls, with each call separated from adjacent calls by an IOI less than or equal to a specified threshold. Individual bout thresholds were obtained for every seal by computing their median IOI, across all recorded dates and interaction contexts. The median IOI was then multiplied by 1,5 to obtain the bout threshold for each seal, as done by Anichini and colleagues (2023), with a maximum threshold of 10 seconds to account for possible memory constraints. If seals were housed together with a partner, the bout threshold for that pair was calculated as the average of the individual thresholds of both seals. This approach ensured that the call timing of both individuals was represented in the bout definition during interactive contexts. Calls were then grouped in bouts based on the obtained thresholds (Table S3.2). Note that 1-way interaction bouts that extended across two playback trials were split, ensuring that each resulting call sequence was correctly assigned to a specific trial. Only bouts with at least 3 turns, or alternations, between callers were selected for the analysis.

For every bout, we computed the response latencies—the time difference between the onset of a call and the offset of the previous call—whenever there was an alternation between callers (Figure 3.2). In conventional turn-taking research, meaning-based analyses would be performed to understand if the response from caller B is contingent on caller A's most recent call or an earlier one. However, since meaning-based analyses have never been done in harbour seals and any analysis would have to be treated with extreme caution to avoid anthropomorphism, response latencies were always derived from alternating calls that were closest to each other in time. Response latencies could be negative, indicating overlap, or positive, indicating no overlap.

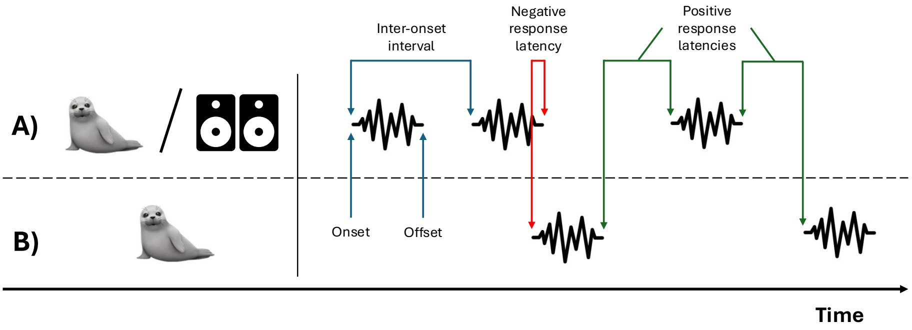


Figure 3.2. Schematic diagram of an interactive bout between callers A and B, illustrating all relevant measures. This calling bout represents a 2-way interaction if caller A is a seal, or a 1-way interaction if caller A is the loudspeaker. Calls are represented by acoustic wave icons. The measures shown include the onset and offset of a call, the inter-onset interval (i.e., the time difference between the onsets of two adjacent calls), and the response latencies between turns. Response latencies can be either negative (i.e., onset of the response occurs before the offset of the previous call) or positive (i.e., onset of the response occurs after the offset of the previous call).

Linear modelling of seal response latencies

Using the *lmer* function (lmerTest R package v3.1.3; Kuznetsova et al., 2017), two different generalised linear-mixed effects models (GLMMs) were fitted to test which variables affected response latencies. The first model tested for the fixed effect of interaction type (1-way vs. 2-way) and included two random intercepts, one for each unique caller and bout. The second model tested for the effect of playback variables from the 1-way interaction, namely the tempo, rhythm, call duration, and caller sex of the playback trial, and also included two random intercepts, one for each unique caller and bout. Note that, for the 1-way interaction in both models, only response latencies from seals were included and the "response latencies" from the playback calls were ignored. Before fitting each model, the most suitable transformation for approximating a normal distribution to the response variable (i.e., response latency) was applied using the *bestNormalize* function (bestNormalize R package v1.9.1; Peterson, 2021). All response latencies were normalised using the *orderNorm* transformation, a rank-based normalisation method that converts data into ranks and then adjusts them to match a normal distribution (Peterson & Cavanaugh, 2020). Once the full models were fitted, backwards model selection was performed based on the Akaike Information Criterion (AIC) using the *step* function (lmerTest R package v3.1.3; Kuznetsova et al., 2017) to find which combination of fixed and random effects offered the best model fit. The best possible model was then fitted using the retained variables and their effects were reported. The normality assumption was visually assessed using QQ plots, and the assumption for equal residual variances was visually assessed using a fitted vs. observed scatterplot.

Autocorrelations of response latencies

For every unique bout, the lag-1, lag-2, and lag-3 autocorrelations were computed from the observed response latencies in both types of interactions. Lag-1 autocorrelation refers to the correlation between consecutive response latencies, lag-2 to the correlation between latencies two turns apart, and lag-3 to the correlation latencies three turns apart. Note that the response latencies of both seal and playback were used for the 1-way interaction. The autocorrelations were performed twice, once on the real, non-altered data, and again on shuffled data, to test the importance of the original data structure, as done in Pouw & Holler (2022). Pairwise comparisons of the lagged autocorrelation distributions were then performed using the Wilcoxon rank sum test. Moreover, a GLMM was built to regress the lag-1 response latencies onto the lag-0 response latencies, with a random intercept for each unique bout, to complement the results from the pairwise comparisons. Similar to the approach used for the previous GLMMs and prior to fitting the model, the response variable was transformed to approximate a normal distribution using the proposed transformation from the *bestNormalize* function (bestNormalize R package v1.9.1; Peterson, 2021). The model assumptions for normality and homogeneity of variances were visually assessed using a QQ plot and fitted vs. observed scatterplot, respectively.

Results

Although 36 pups were originally included in the statistical analysis, only those with sufficient call data were included in the bout analysis. The final dataset comprised 30 pups, which produced a total of 481 bouts, including 294 during 1-way interactions and 187 during 2-way interactions. Most bouts consisted of 3–6 turns, although one bout included as many as 33 turns. The overall median response latency of seals (collapsed across both interaction types) was 87.2 ms (Figure 3.2), with response latencies ranging from -1200 ms to 6000 ms. A GLMM, including random intercepts for each unique caller and bout, was conducted to examine the effect of interaction type on normalised response latencies. Response latencies were significantly longer in 2-way (median = 287.0 ms) compared to 1-way interactions (median = -54.3 ms), $t(194.04) = 6.84$, $p < .001$. Full model results, including parameter estimates, can be found in Table S3.3, and diagnostic plots for model assumptions are shown in Figure S3.1.

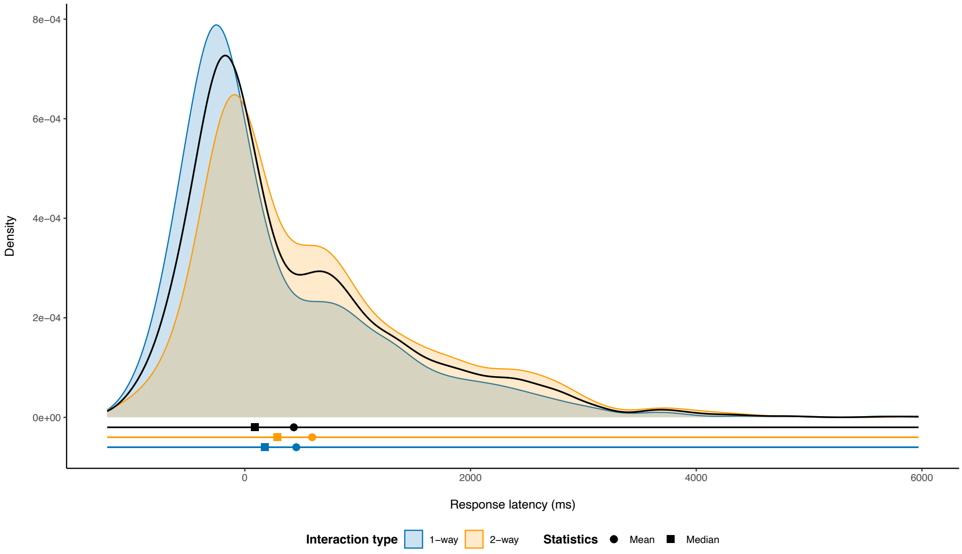


Figure 3.2. Density distribution of response latencies from seals relative to their interaction partner. The data are split by interaction type, showing results for both 1-way interactions (N = 997) and 2-way interactions (N = 991). The black density line represents the distribution of the overall response latencies. The black horizontal line at the bottom of the plot shows the range of the overall response latencies (-1200–6000 ms). The overall median (87.2 ms) is marked by the square and the mean (434.2 ms) is marked by the circle. Negative response latencies indicate turns whose onset began before the offset of the previous turn, thus representing partial overlap.

Effect of playback variables on seal response latencies

Figure 3.3 shows the density plots of the seal response latencies in the 1-way interactions for each playback variable. Wilcoxon rank sum tests revealed that the 1-way seal response latencies were significantly different for tempo ($U = 144645$, $p < .001$), call duration ($U = 154929$, $p < .001$), and sex ($U = 136489$, $p < .05$), but not for rhythm ($U = 115731$, $p = .06$). A second GLMM was conducted to examine the effect of playback variables on normalised response latencies, with random intercepts included for each unique caller and bout. Based on the *step* function using AIC, the best-fitting model included playback tempo and call duration as fixed effects, with a random intercept for caller ID only. Results showed that response latencies were significantly shorter in trials with faster playback tempo ($t(966.14) = -6.34$, $p < .001$) and longer call duration ($t(952.21) = -8.23$, $p < .001$). Reduced model results are presented in Table S3.4, and diagnostic plots (Figure S3.2) confirmed that the assumptions of normality and homogeneity of variance were met.

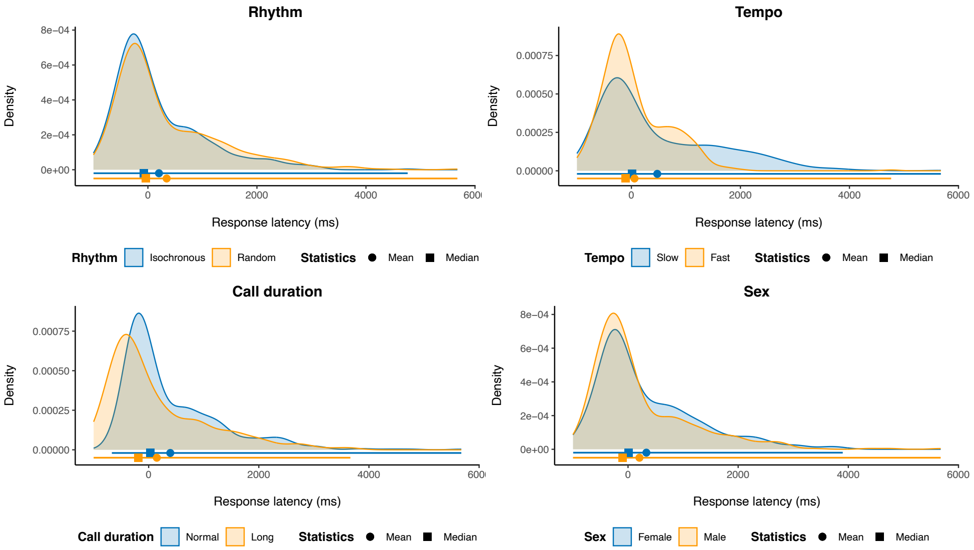


Figure 3.3. Density plots of the seal response latencies in the 1-way interaction, shown by playback variable. The horizontal lines at the bottom of each plot shows the range of latency values for each condition, with the median marked by the square and the mean marked by the circle.

Autocorrelation analysis

Figure 3.4 shows the distributions of lagged autocorrelations computed from each unique bout at lags 1, 2, and 3, for both the original and shuffled series of response latencies. While the average lag-1 autocorrelation was negative for both series, the real series exhibited a much larger negative average compared to the shuffled series, for both 1-way and 2-way interactions. A Wilcoxon rank sum test revealed that the lag-1 autocorrelation was significantly more negative in the real series than in the shuffled series for both 1-way ($U = 29493$, $p < .001$) and 2-way ($U = 13864$, $p < .001$) interactions. Additionally, for both interaction types, the lag-1 autocorrelation of the real series was significantly more negative than the lag-2 (1-way: $U = 14681$, $p < .001$; 2-way: $U = 4872$, $p < .001$) and lag-3 (1-way: $U = 13964$, $p < .001$; 2-way: $U = 3879$, $p < .001$) autocorrelations.

A GLMM was used to examine the relationship between lag-1 response latencies and lag-0 latencies. The model included a random intercept for each bout. The analysis revealed a significant negative relationship between current and previous response latencies, t(2383) = -17.87, $p < .001$). Model results are presented in Table S3.5, and assumptions of normality and homogeneity of variance were confirmed using diagnostic plots (Figure S3.3).

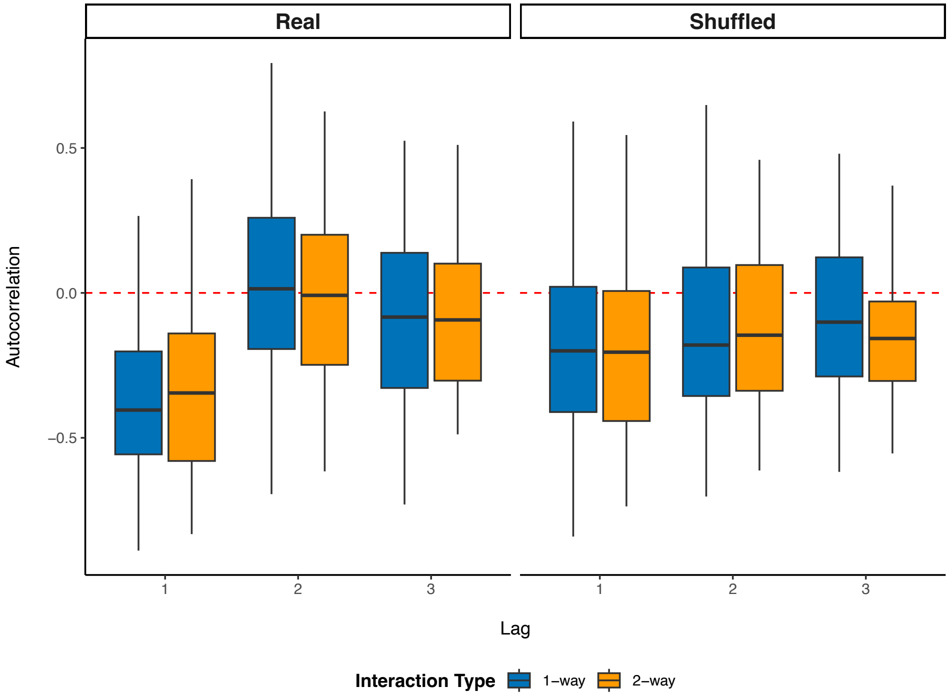


Figure 3.4. Distributions of lagged autocorrelation values for response latencies between turns, shown separately for 1-way (blue) and 2-way (orange) interactions. The left panel displays autocorrelation values based on the original order of response latencies, while the right panel shows values computed from a shuffled order.

Discussion

Our study demonstrates that vocal interactions in harbour seal pups exhibit the key turn-taking characteristics of minimal gaps between turns, avoidance of overlap, and adjustment. Turns during vocal interactions are separated by short, positive silent intervals, with an overall median response latency of 87.2 ms across interaction types. Median response latencies were significantly longer in 2-way interactions (287.0 ms) compared to 1-way interactions (-54.3 ms), which may reflect differences in how latencies were computed between the two interaction types (i.e., ignoring the "response latencies" of the playback in 1-way interactions), as well as differences in call timing predictability or attention. Moreover, the playback experiment indicated that tempo and call duration significantly affect response latencies between turn transitions, suggesting that vocal interaction dynamics influence vocal coordination. Specifically, shorter response latencies were observed at faster tempi, whereby a faster tempo may facilitate quicker vocal exchanges by reducing the time between calls. Similarly, shorter response latencies were observed for longer call durations, suggesting that having more time to prepare a response may contribute to shorter gaps between turns. Finally, in the real, non-shuffled data, response latencies between turns exhibited a significant negative lag-1 autocorrelation, where shorter turn transitions are followed by longer turn transitions and vice versa, suggesting that seal pups use a counter-adjustment mechanism to efficiently time their calls during vocal interactions. Together, these findings provide evidence that harbour seal pups engage in turn-taking behaviour and contribute to a growing body of comparative research on this topic.

The significantly longer median response latencies in 2-way interactions (287.0 ms) compared to 1-way interactions (-54.3 ms) may stem from how the latencies were computed. Specifically, in 1-way interactions, we excluded the latencies of the playback “responding” to the seal. Consistent with the 90-degree anti-synchrony scenario (Anichini et al., 2023; Ravignani, 2019), calls were likely clustered in pairs, with a short call latency within pairs and a long latency between pairs. Therefore, if the playback predominantly initiated a seal response, the response latencies in 1-way interactions would naturally be shorter than in 2-way interactions, where the longer latencies between pairs were not excluded. However, this explanation alone may not be sufficient. Interactions with a live partner are more dynamic and require real-time coordination, which may contribute to longer response latencies due to increased attentional demands and the unpredictability of partner signals. Regardless, longer response latencies are associated with a marked reduction in call overlap between turns. Harbour seal pups may actively avoid call overlap to facilitate maternal recognition of their vocal signatures amidst the noisy colony environment (Anichini et al., 2023). Interestingly, pups show response latencies close to the 200 ms observed in humans (Stivers et al., 2009). While human speech involves pre-planned utterances and predictive cognitive processes (Corps et al., 2018; de Ruiter et al., 2006), it remains unknown if seal vocal responses reflect similar predictive mechanisms. For instance, short response latencies could also arise through simpler, reactive mechanisms (Greenfield, 1994b). Analysing turn-taking using response latencies alone is therefore not sufficient to make inferences about the underlying timing mechanisms. Future studies incorporating phase-based analyses—using an anchor other than the inter-onset-interval, such as the response latency—may offer a more informative approach. Finally, since minimal gaps between turns have been proposed to enhance communicative efficiency in human social interactions (Levinson, 2016), vocal turn-taking in harbour seal pups may support a similar social function. In 1-way interactions, response latencies between turns are shorter and there is regular overlap. While overlap is not uncommon in human face-to-face interactions (Sacks et al., 1974), turn-taking violations (i.e., overlapping turns) in this study may reflect a lack of attention caused by lower engagement with the repetitive nature of playback stimuli, particularly during isochronous trials. Indeed, previous work using ADAM reported lower temporal adaptation during playback experiments compared to conspecific interactions, which was attributed to habituation to the playback stimuli (Anichini et al., 2023). In sum, this study found evidence of minimal gaps and avoidance of overlap between turns in vocal interactions of harbour seal pups, which is suggestive of turn-taking behaviour.

While 1-way interactions may not fully represent naturalistic turn-taking dynamics, playback experiments can help investigate different interaction dynamics, allowing researchers to systematically test how animals adjust call timing under controlled and ecologically valid conditions (Henry et al., 2021). In our playback experiment, response latencies were significantly shorter during trials with faster tempo and longer call durations. Previous work also found that faster playback tempi are associated with shorter response latencies and suggested a form of timing adjustment, which keeps phase constant, ensuring avoidance of call overlap irrespective of tempo (Ravignani, 2019). Notably, tempo plays an important role in human turn-taking, where cultural differences in interactional pace influence perceptions of delayed responses across languages (Stivers et al., 2009). However, it remains unknown whether the same is true across different harbour seal populations. The shorter response latencies observed for longer call durations may indicate that the additional time pups have to anticipate the end of their partner's call contributes to shorter gaps between turn transitions. The increased preparation time may thus allow for more synchronised responses. Interestingly, while seals are sensitive to temporal features such as tempo and call duration, they do not appear to rely on rhythm to adjust their call timing. Moreover, findings show that harbour seal pups display individual-specific variation in their turn-taking behaviour, with caller identity significantly predicting response latencies. Including caller identity in the model rendered sex a non-significant predictor, indicating that individual differences outweigh sex-based effects on response timing. This individual variability in turn-taking has similarly been observed in marmosets, where individual calling rates influenced how quickly they responded to conspecific calls (Grijseels et al., 2024). Harbour seal pups' sensitivity to the interactions' temporal characteristics such as tempo and call duration suggests they favour relative timing adjustments, relying on cues from the immediately preceding turn.

This study revealed a serial dependence (i.e., negative lag-1 autocorrelation) in the real, non-shuffled response latencies of harbour seal pups, providing evidence of mutual timing adjustments during vocal interactions. Importantly, this dependence was specifically reflected in lag-1 autocorrelation and not in higher-order lags, such as lag-2 and lag-3, suggesting that the timing of a turn is primarily adjusted based on feedback from the immediately preceding turn. Additionally, the fact that lag-1 autocorrelations were similar for both 1-way and 2-way interactions indicates that mutual adjustment may not be a necessary condition for turn-taking, with one individual adjusting to the other, but not necessarily vice versa. While the lack of mutual adjustment in 1-way interactions observed here may be explained by the absence of a social partner, timing asymmetries during social interactions could also reflect developmental differences in timing abilities. For instance, in infant-caregiver interactions, mutual adjustment may develop over time, with the caregiver's adjustments being more pronounced at first than the infant's (e.g., Gratier et al., 2015). Moreover, non-mutual coordination may reflect leader-follower dynamics, which have been observed in competitive contexts, such as mate attraction (Grafe, 1996; Greenfield & Roizen, 1993). In contrast, mutual adjustment in humans enables coordinated turn transitions that facilitate cooperative behaviours, like telephone conversations (Pouw & Holler, 2022), where turn-taking can be interpreted using the coupled oscillators framework. For example, the coupled oscillator model suggests that individuals operate as a coupled system, where a speaker's rhythmic cycles, like syllable rate, entrain a listener's endogenous oscillators, allowing the listener to accurately predict when to start its turn (Wilson & Wilson, 2005). Oscillatory dynamics are also evident in other human activities involving temporal synchronisation, such as interpersonal synchronisation in music and dance and finger tapping experiments (Hadley et al., 2015; Repp, 2005; Wing et al., 2014; Zamm et al., 2023). Importantly, oscillator frameworks could also be used to model non-periodic patterns (E. E. Harding et al., 2025), such as the irregular alternation observed in harbour seal vocal interactions. In human conversation, coupled oscillators typically align in an anti-phase relationship to facilitate communication. Similar antisynchronous timing has previously been reported in harbour seal vocal interactions (Anichini et al., 2023; Ravignani, 2019), supporting the idea that oscillatory mechanisms may underlie these interactions. Future studies should therefore test whether turn-taking patterns across species can be interpreted using dynamical systems models, offering a more integrative approach to uncover the potentially shared mechanisms driving this behaviour. While mutual adjustment in humans is typical of cooperative behaviours, cross-species comparisons of turn-taking dynamics could shed light on whether this is indeed more commonly observed in cooperative species.

Taken together, our results suggest that harbour seal pups take turns during vocal interactions and avoid overlapping calls with their immediate neighbours, corroborating previous findings in this species (Anichini et al., 2023; Ravignani, 2019). The remarkably short latency observed in 2-way interactions (287 ms) comes close to the average response duration of around 200 ms observed in humans (Stivers et al., 2009), chimpanzees (Badihi et al., 2024; Fröhlich et al., 2016), bonobos (Fröhlich et al., 2016; Rossano, 2013), and meerkats (Demartsev et al., 2018), suggesting that minimal gaps between turns may support fine-grained vocal coordination across these mammals. Moreover, harbour seal pups display a range of response latencies between -1200 ms and 6000 ms, which falls between the narrower range of human responses (-500 ms to 1500 ms; Stivers et al., 2009) and the broader range seen in chimpanzee gestural communication (-1600 ms to 8640 ms; Badihi et al., 2024). These variations may reflect modality-specific constraints (vocal vs. gestural), as well as differences in behavioural contexts and species-specific biology. While this study focused on turn-taking in the vocal modality, animals, including humans, frequently use multiple modalities such as gesture and facial expressions to communicate (Higham & Hebets, 2013), with the quasi-rhythmic structure of these signals facilitating turn-taking and helping to avoid call overlap (Pouw et al., 2021). Future work integrating both vocal and non-vocal modalities could offer a more comprehensive understanding of how multimodal signals are used to facilitate temporal coordination (e.g., Kendrick et al., 2023; ter Bekke et al., 2024).

In conclusion, this study finds that vocal interactions between harbour seal pups are characterised by minimal gaps between turns, avoidance of overlap, and adjustment. Specifically, vocal interactions exhibit short silent gaps between turns, with pups seemingly using a counter-adjustment mechanism to avoid overlap. Collectively, these findings suggest that harbour seals coordinate vocal interactions by taking turns, possibly to facilitate behaviours such as parent–offspring recognition and social bonding. They further contribute to the growing body of comparative research on turn-taking, offering valuable evolutionary insights into the potentially shared mechanisms underlying this behaviour. Additionally, our playback experiment highlights the role of temporal features, such as tempo and call duration, in shaping social interaction dynamics. Finally, this study advocates for an integrative approach to turn-taking that incorporates vocal and non-vocal signals.

Acknowledgements

We are grateful to Sander van Weerelt for his contribution to the video annotations that helped us attribute vocalisations to each seal. We also thank all the members and volunteers of Sealcentre Pieterburen for their continued support in this research.

Author contributions

Conceptualisation: K.d.R.; Methodology: K.d.R.; Software: K.d.R.; Validation: K.d.R., S.L., A.S.-C., A.R.; Formal analysis: K.d.R.; Investigation: K.d.R, S.L., A.S.-C.; Resources: A.S.-C.; A.R.; Data curation: K.d.R., S.L.; Writing - original draft: K.d.R.; Writing - review & editing: K.d.R., S.L., A.S.-C., P.E.K., B.d.B., J.H., A.R.; Visualisation: K.d.R.; Supervision: B.d.B., J.H., A.R.; Project administration: K.d.R.; Funding acquisition: B.d.B., A.R.

Supplementary materials

Table S3.1. Detailed overview of which animals were tested and when. Adjacent animals whose lines are coloured the same were housed together during the experiment. Note that Oscar and Jenny were housed together with others during the first testing but were housed alone during the second testing.

Name	Sex	Age at arrival	Dates 1st testing	Dates 2nd testing	Number of sessions
Summer of 2021 (from June 16th to August 4th)					
Tiana	F	7	16/06		1
Pongo	M	3	16/06		1
Aladdin	M	7	18–22/06	08–09/07	7
Abu	M	7	18–22/06	08–09/07	7
Jasmine	F	7	18–22/06		5
Kida	F	7	18–22/06		5
Simba	M	7	23–27/06		5
Flynn	M	7	23–27/06		5
Magica	F	7	23–27/06	08–12/07	10
Phocahontas	F	3	23–27/06	08–12/07	10
Jane	F	7	28/06–02/07	13–19/07	10
Tarzan	M	7	28/06–02/07	13–19/07	10
Oscar	F	6	28/06–02/07	13–19/07	10
Naveen	M	7	28/06–02/07		5
Eric	M	7	03–07/07	20–24/07	10
Ariel	F	7	03–07/07	20–24/07	10
Goofy	M	7	03–07/07	21–25/07	10
Dagobert	M	7	03–07/07	21–25/07	10
Gaston	M	7	26–30/07		5
Sjef	M	7	26–30/07		5
Mowgli	M	10	26–30/07		5
Stitch	F	10	31/07–04/08		5
Lilo	F	30	31/07–04/08		5
Summer of 2022 (from June 27th to July 20th)					
Ana	F	7	27/06–01/07		5
Arjen	M	7	27/06–01/07		5
Renato	M	7	27/06–01/07		5
Sander	M	7	27/06–01/07		5

Jenny	F	7	02–03/07	09–13/07	7
Richard	M	7	02–03/07		2
Margriet	F	7	02–06/07		5
Andras	M	7	02–06/07		5
Isa	F	7	04–08/07		5
Vincent	M	7	04–08/07		5
Guus	M	7	07–11/07		5
Hans	M	7	07–11/07		5
Rutger	M	7	12–16/07		5
Brad	M	7	20/07		1

Table S3.2. Overview of all the bout thresholds for each individual and pair. When seals were housed together with a partner, the bout threshold for that pair is equal to the mean value of the pair's respective individual thresholds. Adjacent animals whose lines are coloured the same were housed together during the experiment.

Individual	Partner	Housing	Individual threshold	Final threshold
Pongo	Tiana	Together	3.219	3.219
Aladdin	Abu	Together	4.973	5.275
Abu	Aladdin	Together	5.576	5.275
Jasmine	Kida	Together	5.374	6.013
Kida	Jasmine	Together	6.652	6.013
Flynn	Simba	Together	10.000	10.000
Simba	Flynn	Together	10.000	10.000
Phocahontas	Magica	Together	3.831	3.963
Magica	Phocahontas	Together	4.096	3.963
Oscar	Naveen	Together	8.397	9.199
Naveen	Oscar	Together	10.000	9.199
Tarzan	Jane	Together	3.859	3.690
Jane	Tarzan	Together	3.522	3.690
Goofy	Dagobert	Together	4.392	4.345

Individual	Partner	Housing	Individual threshold	Final threshold
Dagobert	Goofy	Together	4.299	4.345
Ariel	Eric	Together	3.904	3.861
Eric	Ariel	Together	3.817	3.861
Sjef	Gaston	Together	4.961	5.239
Gaston	Sjef	Together	5.518	5.239
Stitch	Lilo	Together	10.000	8.255
Lilo	Stitch	Together	6.510	8.255
Ana	Arjen	Together	6.362	7.741
Arjen	Ana	Together	9.121	7.741
Renato	Sander	Together	4.445	4.627
Sander	Renato	Together	4.809	4.627
Jenny	Richard	Together	4.098	7.049
Richard	Jenny	Together	10.000	7.049
Andras	Margriet	Together	3.094	3.672
Margriet	Andras	Together	4.249	3.672
Vincent	Isa	Together	3.105	5.443
Isa	Vincent	Together	7.782	5.443
Guus	Hans	Together	7.836	8.407
Hans	Guus	Together	8.978	8.407
Oscar		Alone	7.905	7.905
Mowgli		Alone	6.926	6.926
Jenny		Alone	2.457	2.457
Rutger		Alone	10.000	10.000
Brad		Alone	10.000	10.000

Table S3.3. GLMM results for the effect of interaction type on normalised response latencies. Note that this model only included the response latencies of seals relative to their interaction partner (i.e., playback latencies were ignored in the 1-way interaction).

Variable	Estimate	SE	95% CI	P-value
(Intercept)	-0.133	0.055	[-0.241, -0.020]	0.023
Interaction	0.449	0.066	[0.321, 0.578]	< .001

Table S3.4. GLMM results for the effect of playback tempo and call duration on normalised response latencies for the 1-way interactions. Note that this model only included the response latencies of seals relative to the playback (i.e., playback latencies were ignored).

Variable	Estimate	SE	95% CI	P-value
(Intercept)	0.518	0.081	[0.360, 0.685]	< .001
Tempo	-0.390	0.062	[-0.510, -0.269]	< .001
Call duration	-0.514	0.062	[-0.636, -0.390]	< .001

Table S3.5. GLMM results for the effect of lag-1 response latencies on the normalised lag-0 response latencies.

Variable	Estimate	SE	95% CI	P-value
(Intercept)	0.138	0.029	[0.081, 0.197]	< .001
Lag-1	-0.00034	0.00002	[-0.00038, -0.00030]	< .001

The reverberation fragment

This 15-seconds sound fragment was composed of three different elements that together can help measure the reverberation in the room: a sine wave at 440 Hz (similar to the fundamental frequency of harbour seal pup calls; de Reus, 2017), white noise following the inverse repeated sequence technique (Stan et al., 2002) and a logarithmic sine sweep going from 50 Hz up to 10 kHz. It was twice during each session, 10 s before and after the playback period.

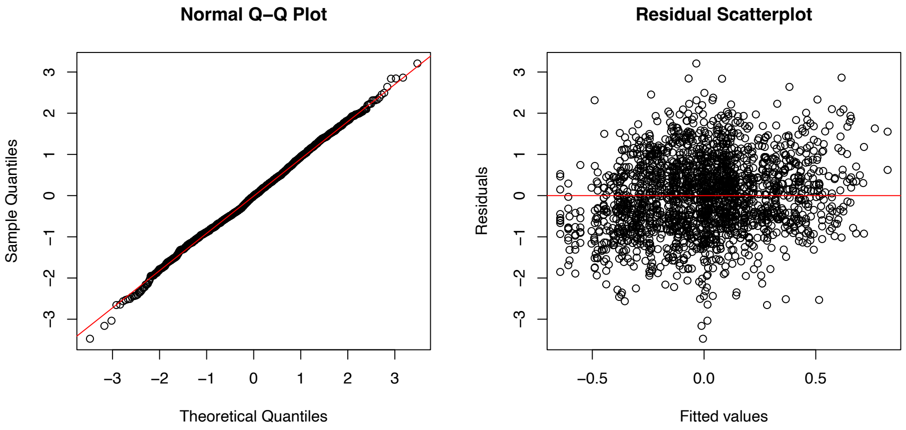


Figure S3.1. Diagnostic plots for assumptions of the interaction model. The left panel shows the QQ plot assessing residual normality, and the right panel shows the residual scatterplot evaluating homogeneity of variance.

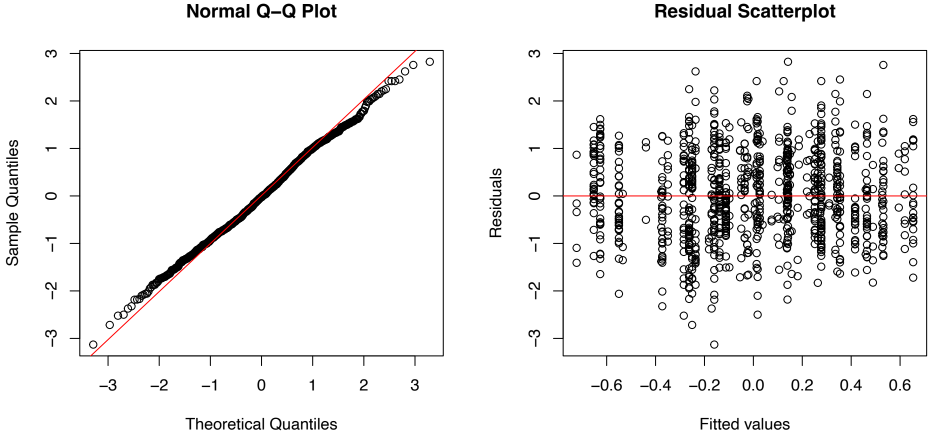


Figure S3.2. Diagnostic plots for assumptions of the playback model. The left panel shows the QQ plot assessing residual normality, and the right panel shows the residual scatterplot evaluating homogeneity of variance.

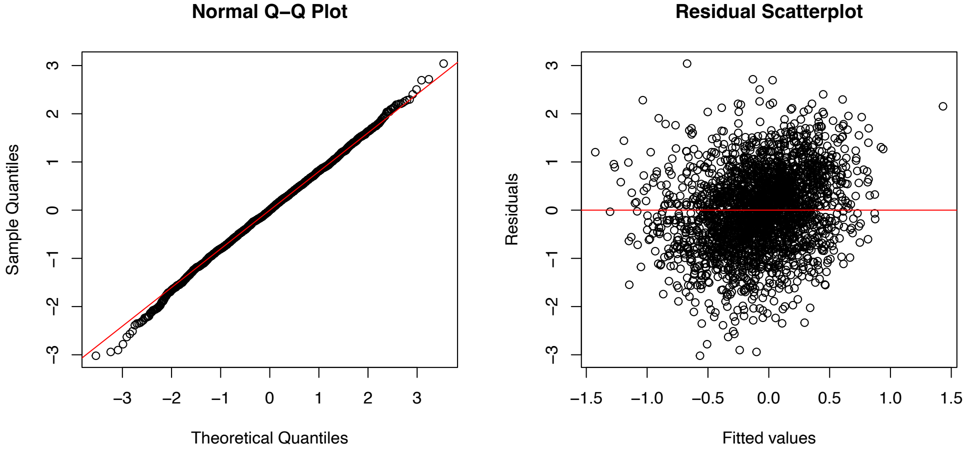


Figure S3.3. Diagnostic plots for assumptions of the autocorrelation model. The left panel shows the QQ plot assessing residual normality, and the right panel shows the residual scatterplot assessing the homogeneity of variance.

Chapter 4

4 | Vocal tract allometry in a mammalian vocal learner

Abstract

Acoustic allometry occurs when features of animal vocalisations can be predicted from body size measurements. Despite this being considered the norm, allometry sometimes breaks, resulting in species sounding smaller or larger than expected for their size. A recent hypothesis suggests that allometry-breaking mammals cluster into two groups: those with anatomical adaptations to their vocal tracts and those capable of learning new sounds (vocal production learners). Here, we tested which mechanism is used to escape from acoustic allometry by probing vocal tract allometry in a proven mammalian vocal learner, the harbour seal (*Phoca vitulina*). We tested whether vocal tract structures and body size scale allometrically in 68 young individuals. We found that both body length and body mass accurately predict vocal tract length and one tracheal dimension. Independently, body length predicts vocal fold length while body mass predicts a second tracheal dimension. All vocal tract measures are larger in weaners than in pups and some structures are sexually dimorphic within age classes. We conclude that harbour seals do comply with anatomical allometric constraints. However, allometry between body size and vocal fold length seems to emerge after puppyhood, suggesting that ontogeny may modulate the anatomy–learning distinction previously hypothesised as clear cut. We suggest that seals, and perhaps other species producing signals that deviate from those expected from their vocal tract dimensions, may break allometry without morphological adaptations. In seals, and potentially other vocal learning mammals, advanced neural control over vocal organs may be the main mechanism for breaking acoustic allometry.

Adapted from: de Reus, K., Carlson, D., Lowry, A., Gross, S., Garcia, M., Rubio-Garcia, A., Salazar-Casals, A., & Ravignani, A. (2022). Vocal tract allometry in a mammalian vocal learner. *Journal of Experimental Biology*, *225*(8), jeb243766. https://doi.org/10.1242/jeb.243766

Introduction

In many species, acoustic signals help mediate social interactions such as competition for mates and territory, and parent–offspring recognition (Bradbury & Vehrencamp, 2011; K. Martin et al., 2017). Signals can encode information about the caller's biology which can be readily deciphered by the receiver, including age (Charlton et al., 2009; Reby & McComb, 2003), sex (Charlton et al., 2009; Vignal & Kelley, 2006), body size (Charlton et al., 2009, 2011; Fitch, 1997; Garcia et al., 2016), hormone levels (Koren & Geffen, 2009) and physical condition (Koren & Geffen, 2009; Wyman et al., 2008).

In particular, body size often shapes mammalian sounds by constraining the geometry of the vocal tract (Fitch, 2000; Reby & McComb, 2003). Acoustic cues relating to the body size of the caller can inform the receiver about the caller's competitive ability and reproductive success (Kuester et al., 1995; Pfefferle & Fischer, 2006; Poole, 1999; Reby & McComb, 2003). For example, in primates and carnivores, there is an inverse relationship between body size and call frequency parameters, where larger animals produce calls with lower frequencies, i.e., have a "deeper" voice (Bowling et al., 2017). This relationship between acoustical call features and body size, where one accurately reflects the other, is known as "acoustic allometry" (Fitch, 1997; Taylor & Reby, 2010). Here, signalling is considered honest when the acoustic parameters of observed vocalisations accurately reflect an individual's body size (Fitch & Hauser, 2003; Zahavi, 1977). Deviations from allometry can generate dishonest signals, with animals sounding unexpectedly small or large for their body size (Garcia & Ravignani, 2020). Dishonest signals may be produced when an animal (1) shows a lack of allometric scaling between its vocal tract and its body size, or (2) shows enhanced control over its vocal organs, which allows it to learn new vocalisations or modify existing vocalisations—an ability known as "vocal learning" (Janik & Slater, 1997; Lattenkamp & Vernes, 2018). Recent work indeed showed that, given a cross-species regression between sounds produced and body size, outlier species seem to cluster either well below the regression line (those with anatomical adaptations) or markedly above it (the vocal learners). This led to a "morphology versus learning hypothesis" (Garcia & Ravignani, 2020; Ravignani & Garcia, 2022): dishonest signals in mammals may arise either from anatomical adaptations or from vocal learning capacities. This prediction has the potential to identify new vocal learners or species with unexpected vocal tract morphology. Vocal learners should therefore be able to violate acoustic allometry while possessing a vocal tract that scales allometrically with the rest of their body. For the first time, we tested this prediction, asking whether vocal tract allometry is present in a vocal learning species that is known to violate acoustic allometry.

Harbour seals (*Phoca vitulina*) are vocal learners that escape acoustic allometry by producing sounds with different frequencies to those expected from their body size, allowing them to transmit dishonest body size information. Indeed, they stand out as outliers in cross-species allometric regressions between body mass and frequency parameters (Ravignani & Garcia, 2022; Figure 4.1). Moreover, previous studies have shown that harbour seals can actively modulate the call frequencies they produce based on auditory experience. In one special case, a human-raised harbour seal, named Hoover, was found to be capable of mimicking human speech sounds (Ralls et al., 1985). In a more recent study on harbour seal pups, young animals were found to be capable of lowering their fundamental frequency (f_0) in the presence of background noise (Torres Borda et al., 2021). Do the environmental noise conditions in which vocalisations are produced have a stronger influence on f_0 values than body size? To address this, we complemented acoustic data from Torres Borda and colleagues (2021) with body mass information and reanalysed it to show that acoustic allometric relationships do indeed break down in this species as a result of the large vocal plasticity observed within individuals (see Figure 4.2 and detailed explanations in the supplementary materials). These re-analyses indicate that, also within species, individual harbour seals may sound bigger or smaller than predicted by body size. Seals can therefore escape the constraints of acoustic allometry, both across and within species.

Harbour seals are particularly vocal during the first few weeks following birth (Perry & Renouf, 1988). Pups produce individually distinctive mother attraction calls (Renouf, 1984), which vary with age, sex and body length (Khan et al., 2006; Sauvé et al., 2015). After weaning, however, these calls disappear entirely from their vocal repertoire, with most vocalisations ceasing aside from occasional clicks and growls (Renouf, 1984). During adulthood, female harbour seals remain almost entirely vocally inactive (van Parijs & Kovacs, 2002), but males start vocalising again, producing underwater calls during the mating season (Hanggi & Schusterman, 1994). The large variation in vocal repertoire observed across individuals, sexes and age classes makes harbour seals ideal candidates to test the morphology versus learning hypothesis, i.e., whether a vocal learning mammal does indeed escape acoustic allometry via learning instead of via anatomical adaptations.

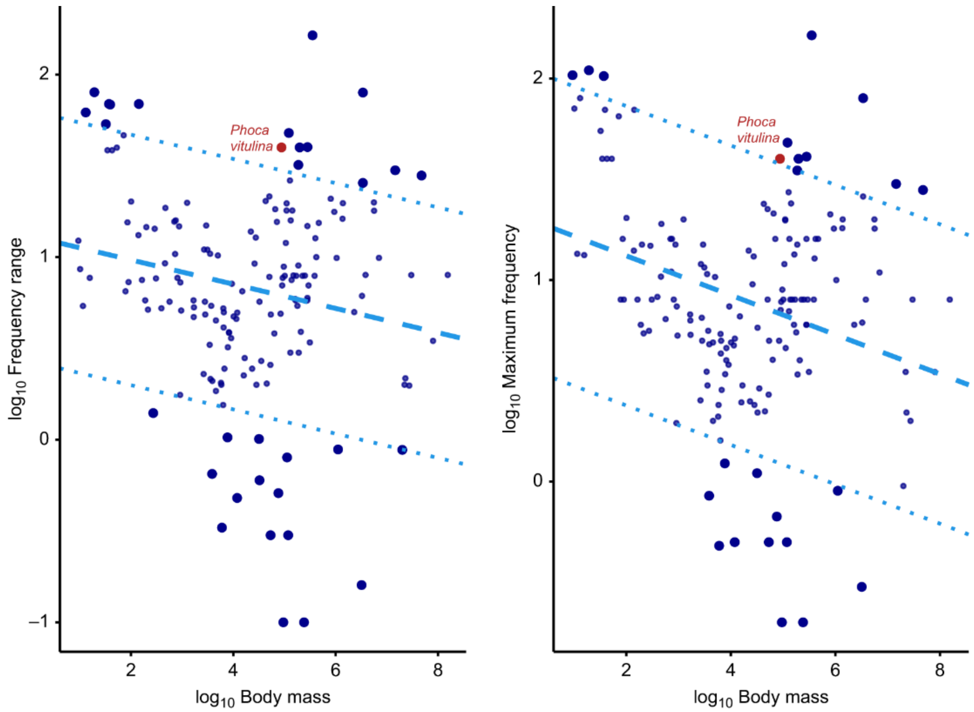


Figure 4.1. Phylogenetic generalised least squares regressions between frequency parameters and body mass across 164 mammalian species. Left, frequency range; right, maximum frequency. All variables are log-transformed and the figure is adapted from Ravignani and Garcia (2022). The dotted lines represent a threshold at 2.5 standard deviations from the main regression lines used to define outliers. Non-outlier species (which show acoustic allometry between frequency parameters and body mass) are represented by small circles, and outlier species (which escape acoustic allometry) are represented by large circles. The two red data points, representing harbour seals, are both outliers.

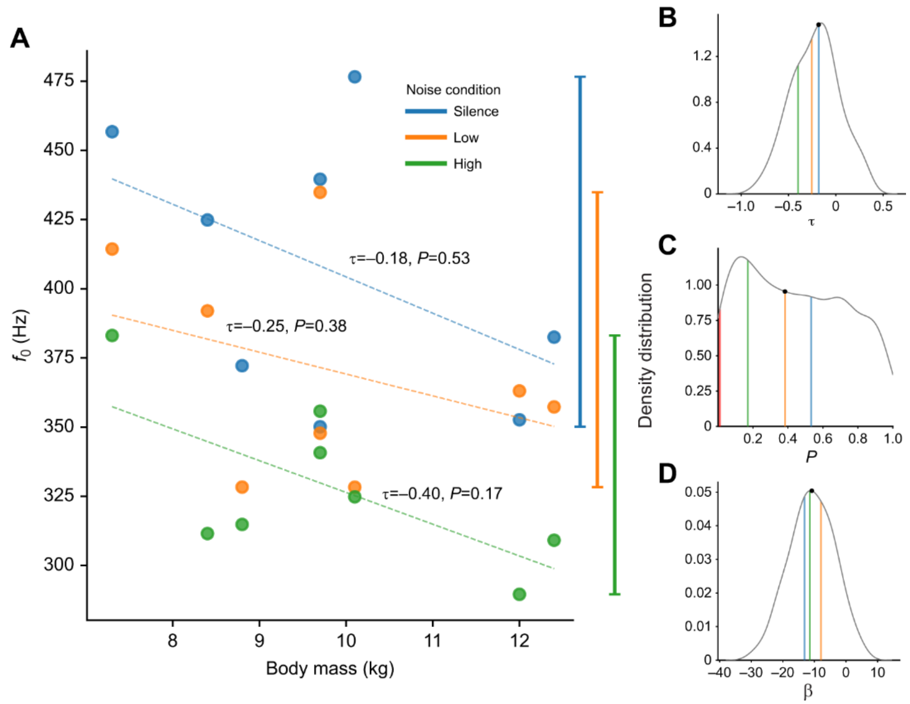


Figure 4.2. Lack of acoustic allometry relationships in harbour seals. (a) Correlations between median fundamental frequency (f_0) for each noise condition (silence, low and high) and body mass. The respective correlation coefficients (τ) and associated P-values for each correlation are reported above the regression line. At first sight, the characteristic inverse relationship between f_0 and body size may seem present, but there is some overlap in the range of f_0 values (whiskers on the right side of the plot) produced by individuals of differing body size between noise conditions. Non-significant *p*-values suggest that, at least in this sample, there is a lack of acoustic allometry. In addition, allometry may break if calls are produced in different noise conditions. In other words, do the environmental conditions in which vocalisations are produced strongly affect the f_0 values, as much as or even more than body mass? (b–d) Density distributions produced by computing 10,000 different combinations of randomly selected median f_0 values (1 of the 3 median frequency values per seal) to assess whether allometric relationships hold across noise conditions. The coloured vertical lines in these plots represent the respective median values for each of the noise conditions. The median value of the distribution is represented by black circle on the density curve. (b) Density distribution of the Kendall rank correlation coefficients. The median value lies around -0.18, pointing to a weak negative correlation. (c) Density distribution of the correlation *p*-values associated with the correlations from B. The median *p*-value is 0.38, which means that in most of the simulated cases we would not reject the null hypothesis (i.e. the correlation is not significantly different from 0). In fact, in only 2.2% of cases (217 out of 10,000) is the correlation significant; this is indicated by the red vertical line. In other words, in 10,000 simulated samples of 8 seals, we generally found no acoustic allometry. (d) Density distribution of the simulated linear regression coefficients (β), where the median value is -10.8 Hz. Given a 5.1 kg difference in body mass between the smallest and the

largest seal, we would expect, on average, a frequency shift of 55.08 Hz. For every individual, we calculated the difference of the median f_0 values between the silent and high noise condition; the median range across all individuals was 73.6 Hz. This suggests that the differences caused by individual variability in f_0 in response to noise conditions are larger than the f_0 differences expected from body mass differences alone. Seals of differing body sizes (e.g., 7 versus 12 kg) could thus potentially produce the same f_0 value. This would mean that, in harbour seal pups, vocal plasticity can outweigh and mask acoustic allometric relationships.

Most mammalian vocalisations are described using the source–filter theory of vocal production. Within this framework, vocal signals are initially produced by a source and are then filtered by the vocal tract before being emitted into the environment (Fant, 1971). In mammals, the source of sound production consists of the vocal folds in the larynx, and the filter is composed of the cavities making up the upper vocal tract (Fant, 1971; Figure 4.3). The vocal folds are shelves of tissue lying across the airway that attach ventrally and laterally to the thyroid cartilage and dorsally to the arytenoid cartilage (Figure 4.4a). When vocalising, the air expelled from the lungs rushes between the vocal folds, causing them to vibrate and produce sound (Elemans et al., 2015). The sound then continues to propagate along the upper vocal tract and is modified by its geometry (i.e., filtered) before being emitted as vocalisation. The source–filter framework highlights which vocal tract structures determine specific features present in acoustic signals. The rate of vibration of the vocal folds determines the f_0 and the cavities of the vocal tract determine formant frequencies (Taylor & Reby, 2010). Measurements of these vocal tract structures can thus be used to estimate specific acoustic features of vocalisations.

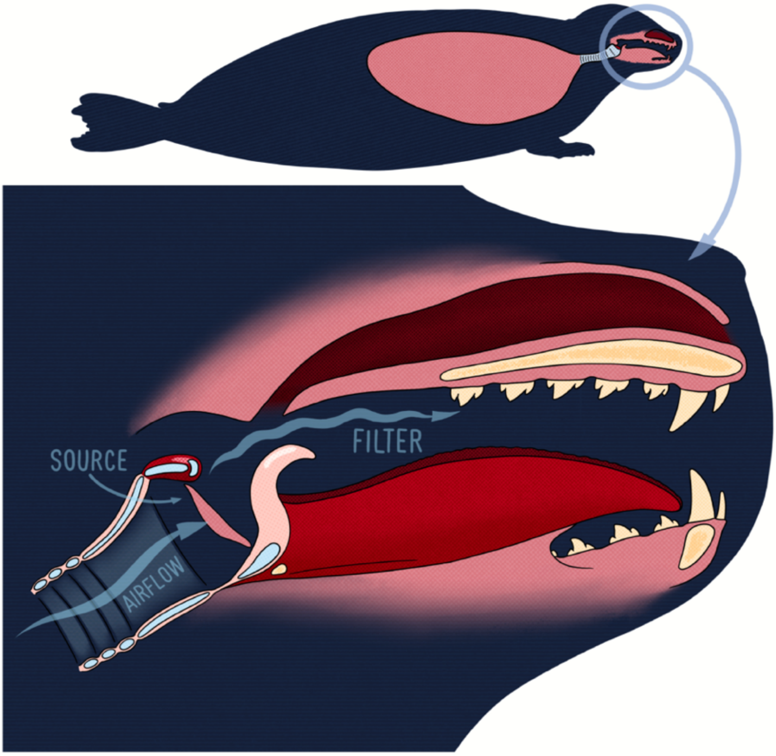


Figure 4.3. Illustration of the source–filter theory of sound production using the vocal anatomy of the harbour seal.

Bioacoustics studies often investigate allometric relationships between acoustic signal features and body size, without consideration of the underlying allometric scaling between body size and vocal anatomy. Most mammals show allometry between body size and upper vocal tract length because the upper vocal tract is constrained by bony structures (Fitch, 1997, 2000; Fitch & Giedd, 1999; Garcia et al., 2016; Plotsky et al., 2013). However, allometry between body size and the size of the vocal folds is less common: the larynx is surrounded by cartilaginous structures and is thus less constrained, suggesting that vocal fold length can be decoupled from overall body size, as found in non-human primates (Fitch, 1997; Fitch & Hauser, 1995; Garcia et al., 2017). In mammals, formants, the acoustic proxy of vocal tract length, are thus often a stronger body size predictor than f_0, the acoustic proxy of vocal fold length (Fitch, 1997; Garcia et al., 2016).

Within the larger framework of the hypothesis above, this study tested for allometric relationships between body size and vocal anatomy measurements in young harbour seals and how these relationships vary with sex and age. Preliminary work found that harbour seal body length correlates with upper vocal tract length and tracheal diameter, but not with vocal fold length (Ravignani et al., 2017). Here, we aimed to expand on these findings by using a larger sample size (353% increase), adding refined anatomical measurements, and comparing different age classes (to test for developmental effects). Based on previous literature, we expected to find allometry between body size and vocal tract structures that are surrounded—and hence constrained—by bony structures, such as vocal tract length. However, based on harbour seal vocal learning abilities (Janik & Slater, 1997; Ralls et al., 1985; Torres Borda et al., 2021), we expected their vocal flexibility to offer favourable grounds to find deviations from body size allometry for vocal tract components surrounded by cartilage, such as the trachea and vocal fold length.

Materials and methods

Sample collection

Larynges were collected during necropsies on 68 young harbour seals, *Phoca vitulina* Linnaeus 1758 (34 males, 34 females). Fifty-two samples came from seals that stranded on the Dutch coastline, the rest from animals found on the German coastline (Schleswig-Holstein). Forty-two animals died in captivity at Sealcentre Pieterburen, Pieterburen, The Netherlands, either naturally during rehabilitation despite intensive care or by means of euthanasia because of the presence of severe clinical signs without any indication of recovery. Euthanasia was performed by trained veterinarians, after sedation, with pentobarbital sodium (100 mg kg^{-1}) using the method described in Greer and colleagues (2001). The other 26 animals died in the wild, either naturally or were mercy killed by trained hunters because they showed severe signs of illness (Table S4.1). No animals were euthanised or mercy killed for the purpose of this study.

At the time of death, the seals studied were aged between 9 days and 12 months (median 6 months). The age of new-born individuals was estimated in number of days by expert seal veterinarians based on the condition of the umbilical cord or the umbilicus. Older individuals with a closed umbilicus were assigned June as their birth month, which is consistent with the majority of harbour seal births in the Wadden Sea (Osinga et al., 2012; Reijnders et al., 2010). Animals aged 1 month or younger were classified as pups, while those between 1 and 12 months of age were classified as weaners, making age a binary variable. Of the 68 individuals included in this study, 14 (8 males, 6 females) were classified as pups and 54 (26 males, 28 females) were classified as weaners. A Fisher's exact test showed no significant association between age and sex ($\chi^2 = 0.765$, $p > .05$), suggesting our sample was balanced between sexes and ages.

Sample treatment and measurements

Post-mortem examinations were performed by veterinarians who all trained at Sealcentre Pieterburen and thereby used the same necropsy protocol (Pugliares et al., 2007). Dutch seals were examined at Sealcentre Pieterburen and German seals were necropsied at the Institute for Terrestrial and Aquatic Wildlife Research (ITAW), Büsum, Germany. Necropsies were performed on either cooled or defrosted carcasses. Body mass, body length and axillary girth were all measured prior to the start of the necropsy. Body length was measured from the tip of the nose to the end of the tail in a non-curvilinear fashion, while the animal was in supine position, and axillary girth was measured as the body circumference directly caudal to the front flippers. The vocal apparatus including the upper vocal tract, the larynx and part of the trachea was then removed and immediately frozen at -20°C. All samples were in a similar condition (i.e., none presented signs of decomposition), comparable to pinniped vocal tracts in Schneider (1962) and Ravignani and colleagues (2017).

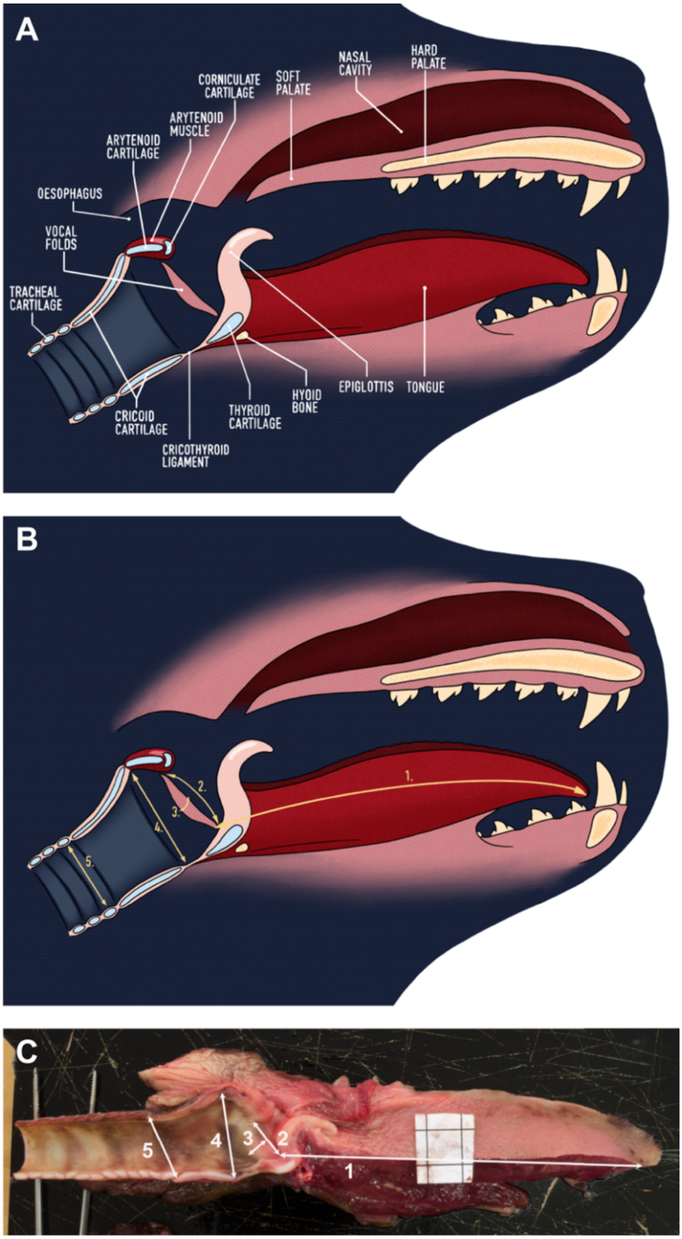


Figure 4.4. Vocal anatomy of the harbour seal. (a) The main anatomical structures composing the vocal tract. (b) The measurements shown on a digital rendering. (c) The measurements shown on a picture of a hemi-larynx from a harbour seal pup. In C, the black square outlined on the piece of white paper serves as a reference and is exactly 1 cm^2. The vocal tract measurements taken include (1) vocal tract length (VTL), (2) vocal fold length (VFL), (3) vocal fold thickness (VFT), (4) subglottic-tracheal dorsoventral distance 1 (STDV1) and (5) subglottic-tracheal dorsoventral distance 2 (STDV2).

Prior to measurement, samples were thawed in a refrigerator at 8°C and each larynx was cut medially to produce two hemi-vocal tracts. The measurements taken on these hemi-vocal tracts (Figures 4.4b and 4.4c) include vocal tract length (VTL), vocal fold length (VFL), vocal fold thickness (VFT), and tracheal measurements in the form of subglottic-tracheal dorsoventral distance (STDV; called subglottic-tracheal anterior–posterior distance, STAP, in Roers et al., 2009) using a calliper to an accuracy of ± 0.01 mm. Although the vocal tract can be divided into lower (below larynx) and upper (above larynx) sections, formants (the resonant frequencies, which often encode information about body size) are only determined by the upper vocal tract (Lester & LaGasse, 2008). VTL will henceforth refer to the length of the upper vocal tract. VTL was measured as the linear distance from the caudal end of the epiglottis to the rostral end of the tongue muscle while the tongue was kept straight. VFL was measured as the distance from the ventral attachment of the vocal fold on the thyroid cartilage to the dorsal attachment of the vocal fold on the arytenoid cartilage. VFT was measured as the distance between the anterior and posterior sides of the vocal folds. The first STDV was measured as the distance between the cricothyroid ligament and the caudal end of the arytenoid. The second STDV was measured as the diameter of the first tracheal ring. All measurements were performed independently by two raters (K.d.R. and A.R.), different from the veterinarians who performed the dissections. For both raters, VTL, VFL and VFT were measured 4 times, twice for each hemi-vocal tract, and STDVs were taken twice, once for each hemi-larynx, because the start and end measuring points were composed of cartilage (as opposed to soft tissue) and, hence, we assumed that the inter-rater reliability for STDVs would be higher than for other measurements.

Statistical analysis

Statistical analyses were performed in RStudio v1.1.463 (R v4.0.4). First, for both raters, the medians for VTL, VFL and VFT were computed from all values reported for every right and left hemi-larynx. Second, using the medians from the first step, the median values for all measurements including STDV1 and STDV2 were computed for each larynx. This provided, for each larynx and rater, five measurements: VTL, VFL, VFT, STDV1 and STDV2. The inter-rater reliability for VTL, VFL, VFT, STDV1 and STDV2 was evaluated using Pearson's correlations. Finally, the overall median values between raters were computed for all measurements. Using these new values, Spearman's correlations between body size and vocal anatomy measurements were then calculated (Table 4.1). For each measurement, normality was assessed using the Shapiro–Wilk test and homogeneity of variance was assessed using an F-test. If both assumptions were met, a two-tailed independent samples *t*-test was computed to check for age and sex differences. When variables were not normally distributed, but samples had equal variance, a Mann–Whitney *U*-test was performed to assess group differences instead.

Predictive modelling was done using generalised linear models (GLMs) with the stats R package (http://www.R-project.org/). A series of models were produced for all anatomical measurements with high inter-rater reliability ($r > 0.70$; Salkind, 2010, p. 627). For every response variable, the full model included the fixed effects body length, body mass, girth, sex, age and the interaction effects of sex with all body size predictors, age with all body size predictors and the interaction of age and sex. The reduced model was then obtained through stepwise regression based on Akaike information criterion (AIC) values. An analysis of variance (ANOVA) test was performed to ensure that the reduced model was not performing significantly worse than the full one. Variance inflation factor (VIF) scores were calculated for all predictors included in the reduced models using the car R package (Fox & Weisberg, 2018). Multicollinearity was considered problematic for subsequent model selection if VIF scores were greater than 5 (Akinwande et al., 2015). For all selected models, deviance explained was calculated from the model output (1 - residual deviance/null deviance) and expressed as a percentage. Plots displaying the predicted effects of every predictor retained in the final models were produced to assess their relationship with the response variable. Diagnostic residual plots were used to verify the model assumptions. Independence of residuals was tested using a Durbin–Watson test (Fox & Weisberg, 2018). Normality of residuals was assessed visually by plotting model fit against the observed data. Homoscedasticity (i.e., constant variance) of residuals was also assessed visually using quantile–quantile plots. Finally, influential data points were assessed by calculating Cook's distance.

Results

Inter-rater reliability for VTL, VFL, VFT and both STDVs was evaluated using Pearson correlations. VTL ($r = 0.94$), VFL ($r = 0.88$), STDV1 ($r = 0.97$) and STDV2 ($r = 0.93$) showed high inter-rater reliability. VFT ($r = 0.59$) showed lower inter-rater reliability and was consequently excluded from further analysis. All correlations were significant at $p < .001$.

All Spearman correlations between body size and vocal anatomy measurements showed positive relationships and significance at the 0.05 level (Table 4.1). There were high correlations between body mass and body length ($r_s = 0.70$), and between body mass and girth ($r_s = 0.86$). Other notable correlations included those between VTL and VFL ($r_s = 0.72$), VTL and STDV1 ($r_s = 0.70$), VFL and STDV1 ($r_s = 0.82$), and VFL and STDV2 ($r_s = 0.76$). Spearman correlations for pups and weaners can be found in Table S4.2.

Table 4.1. Pairwise Spearman correlations of anatomical measurements. VTL, vocal tract length; VFL, vocal fold length; VFT, vocal fold thickness; STDV, subglottic-tracheal dorsoventral distance. Note, all correlations were significant at $p < .05$ after correcting for multiple comparisons using the Holm–Bonferroni method.

Variable	Body length (cm)	Body mass (kg)	Girth (cm)	VTL (mm)	VFL (mm)	VFT (mm)	STDV1 (mm)
Body mass (kg)	0.70						
Girth (cm)	0.53	0.86					
VTL (mm)	0.62	0.69	0.63				
VFL (mm)	0.73	0.79	0.65	0.72			
VFT (mm)	0.44	0.67	0.68	0.52	0.60		
STDV1 (mm)	0.63	0.78	0.65	0.70	0.82	0.69	
STDV2 (mm)	0.58	0.72	0.62	0.67	0.76	0.60	0.81

All anatomical measurements were non-normally distributed but showed equal variances across age and sex groups. A Mann–Whitney *U*-test was used to test for group differences as only the assumption for homogeneity of variance was satisfied. All anatomical measurements were significantly larger in weaners than in pups ($p < .001$; Table 4.2 and Figure 4.5). No significant sex differences were found when considering pups and weaners together ($p > .05$). When considering pups alone, both the normality and homoscedasticity assumptions were met. A two-tailed independent samples *t*-test found significant sex differences for VTL ($t = -3.42$, $p < .05$; Figure 4.6a). Male pups (mean ± s.d. VTL 86.0 ± 2.9 mm) had a larger mean VTL than females (79.8 ± 3.7 mm). When considering weaners alone, variables showed non-normal distribution, but equal variances. A series of Mann–Whitney *U*-tests found that only STDV1 was significantly different across sexes ($U = 218$, $p < .05$; Figure 4.6b). Weaned males (25.1 ± 1.5 mm) had a wider mean STDV1 compared with weaned females (24.2 ± 1.2 mm).

Table 4.2. Mean (± s.d.) anatomical measurements.

Variable	All	Pups	Weaners
Body length (cm)	88.07 ± 8.03	79.86 ± 4.79	90.19 ± 7.32
Body mass (kg)	14.53 ± 3.92	9.81 ± 1.46	15.75 ± 3.39
Girth (cm)	57.55 ± 9.62	46.54 ± 5.66	60.41 ± 8.30
VTL (mm)	91.43 ± 6.77	83.36 ± 4.48	93.53 ± 5.60
VFL (mm)	10.92 ± 1.16	9.26 ± 0.90	11.35 ± 0.77
VFL (mm)	5.15 ± 0.61	4.50 ± 0.43	5.35 ± 0.51
STDV1 (mm)	23.97 ± 1.95	21.35 ± 1.37	24.65 ± 1.43
STDV1 (mm)	17.43 ± 1.82	15.17 ± 1.04	18.02 ± 1.49

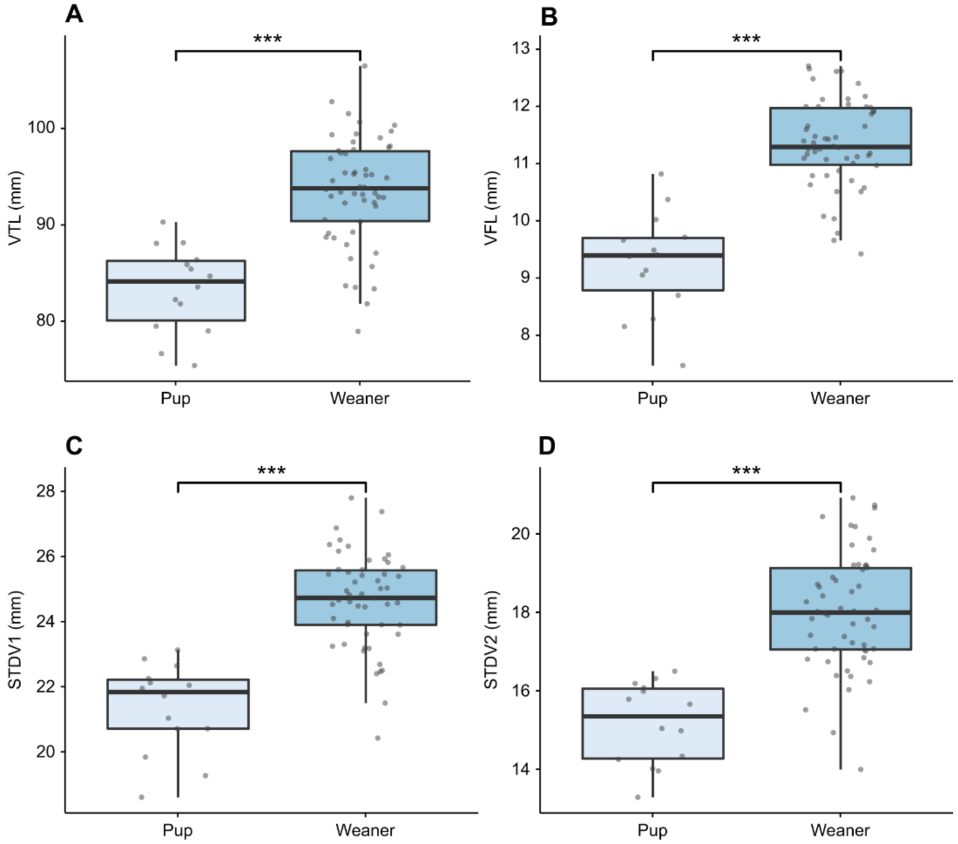


Figure 4.5. Boxplots illustrating the significant differences in anatomical measurements between pups and weaners. (a) VTL, (b) VFL, (c) STDV1 and (d) STDV2. Boxplots show median, upper and lower quartiles and 1.5 × the interquartile range. The level of significance is denoted by asterisks (*** *p* = .001).

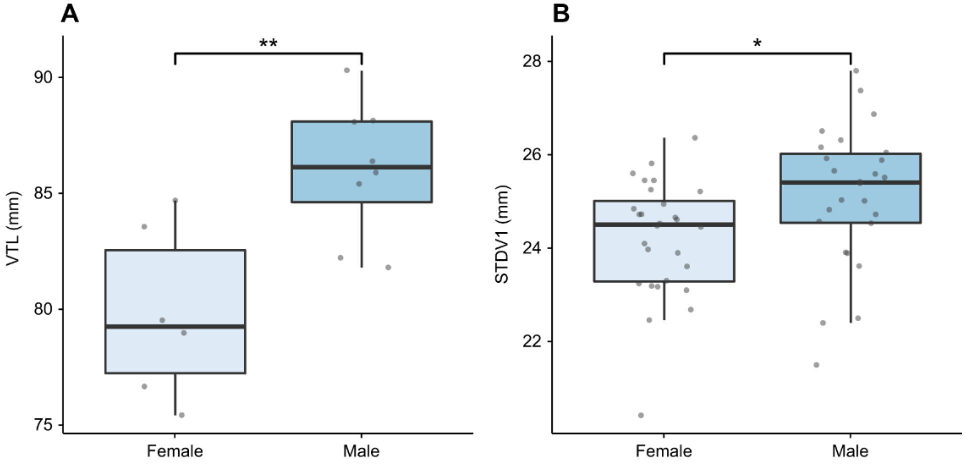


Figure 4.6. Boxplots illustrating the significant differences in anatomical measurements according to sex. (a) VTL in pups. (b) STDV1 in weaners. The level of significance is denoted by asterisks (* *p* = .05 and ** *p* = .01).

A reduced GLM, obtained by stepwise regression based on AIC values, was produced for every vocal tract measurement with high inter-rater reliability, including VTL, VFL, STDV1 and STDV2. All VIF scores were lower than 5 suggesting that multicollinearity was not problematic in the selected models. All model assumptions were satisfied. Moreover, ANOVA testing indicated that the reduced models did not perform significantly worse than the full models ($p > .90$). GLM results showed that most vocal tract dimensions were best explained by body length, body mass, age and sex (Table 4.3). Girth was not retained as a predictor term in any of the selected models. For each model, the predictor estimates with their confidence intervals can be found in Table S4.3 and plots of the predicted effects can be found in Figures S4.1–S4.3. Significant interaction effects are shown in Figure 4.7.

Table 4.3. Selected models for each vocal tract structure. Models included the predictors body length (BL), body mass (BM), age class (A) and sex (S). Predictor terms joined by an asterisk denote an interaction effect. Significant predictor terms are shown as underlined.

Measurement	Selected model	Deviance explained (%)
VTL	BL + BM + A + S + BL*S	59.30
VFL	BL + BM + A + S + A*BM + A*S	74.89
STDV1	BL + BM + A + S	69.99
STDV2	BL + BM + A + S	58.38

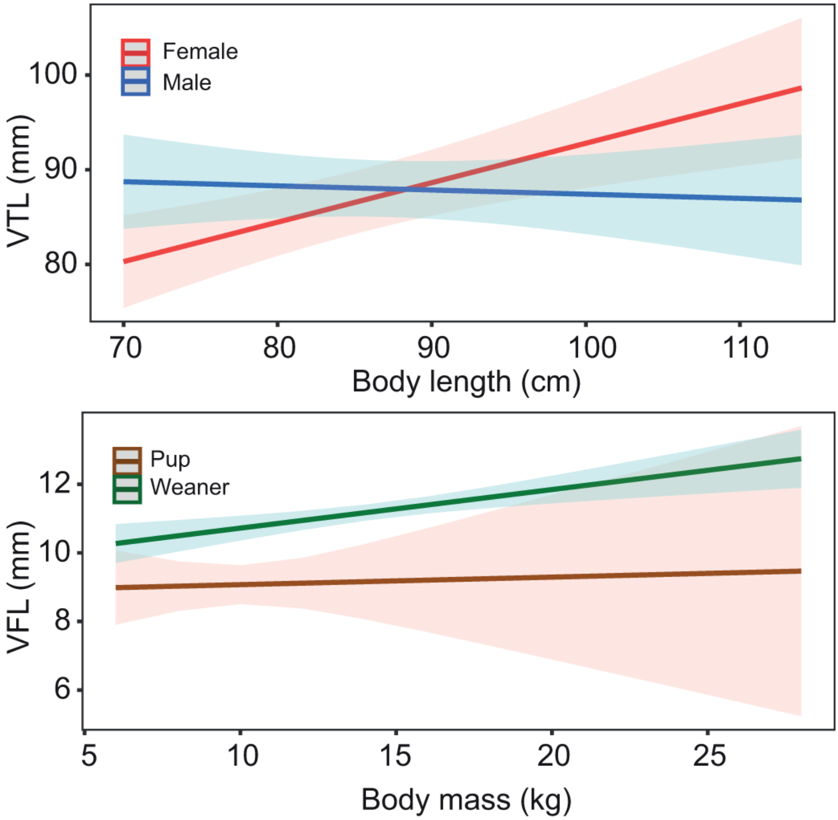


Figure 4.7. Predicted effects of the body length and sex interaction for VTL (top) and the body mass and age interaction for VFL (bottom). The shading around each line of best fit indicates the 95% confidence interval.

Discussion

This study reports on the allometric relationships between body size and vocal tract dimensions in harbour seals. It shows that body length accurately predicts VTL, VFL and STDV1, and body mass predicts VTL and both tracheal measurements (STDV1 and 2). We also find age and sex to be important predictors for the size of vocal tract structures. This is evidenced by significant differences in measurements between age classes and significant sexual differences within age classes.

Previous work showed that upper vocal tract (i.e., filter) dimensions in mammals are predicted by body size measurements (Fitch, 1997, 2000; Fitch & Giedd, 1999; Garcia et al., 2016; Plotsky et al., 2013; Ravignani et al., 2017) and our results provide additional evidence to support such allometry. Although most studies have used body length as a proxy for body size, we found that body mass can also be used to predict VTL in harbour seals. In the first years of life, harbour seals show a linear growth rate for both body length (Hauksson, 2006) and mass (Markussen et al., 1989), suggesting that VTL may develop in a similar fashion during this period. Acoustic proxies for the filter could thus provide a good estimation of a harbour seal's size. In mammals, formant frequencies and formant spacing can be predicted from VTL and vice versa (see Reby & McComb, 2003). Other acoustic proxies include energy quartiles, the frequency of amplitude peaks, and the ratios between these amplitudes (Sauvé et al., 2015). These parameters also encode individual signatures, suggesting that acoustic individuality may partially be an allometric by-product (Ravignani et al., 2017). Harbour seals have the vocal tract predispositions to produce vocalisations that accurately reflect body size whilst also sharing individual-specific information, suggesting that learning does not need to be invoked to explain individuality.

Across mammals, source-related features such as f_0 can sometimes predict body size despite showing weaker allometric scaling than filter-related features (Charlton et al., 2011; Charlton & Reby, 2016; Pfefferle et al., 2007; Reby & McComb, 2003); it was unclear whether this holds for harbour seals (Bowling et al., 2017; Ravignani et al., 2017). Our findings indicate that VFL, which may be used to determine f_0, can be predicted by body size in harbour seals. Moreover, Sauvé and colleagues (2015) reported a decrease in f_0 with an increase in body length of harbour seal pups. Taken together, this suggests that a harbour seal's f_0 can be predicted from vocal anatomy. Previous evidence against allometric scaling for VFL could be explained by low statistical power or lack of testing for age effects on vocal tract measurements (Ravignani et al., 2017). It is indeed notable that age is included in both interactions that were retained in the selected VFL model. Our results, including both pups and weaners, show that allometric scaling between body size and VFL only emerges after weaning, suggesting that VFL may not be constrained in harbour seal pups (Figure 4.7, bottom). This begs the following question: how would escaping acoustic allometry for source-related features be beneficial for pups? Broadcasting honest body size information may be detrimental to harbour seal pups as they are significantly more likely to be displaced by larger conspecifics during agonistic interactions (Neumann, 1999). However, pups may be able to benefit from lowering the f_0 (Torres Borda et al., 2021) of their calls to create an impression of size exaggeration. However, pups may also benefit from increasing the f_0 of their calls to convey distress to the mother (Briefer, 2012). Future playback studies could and should contrast these hypotheses.

Several phocid species use the trachea for sound production (Bryden & Felts, 1974), but this could be a by-product of adaptive modifications to the respiratory tract required for diving (Kooyman & Andersen, 1969; Tyack & Miller, 2002). Our results support the correlation between tracheal diameter and body length found by Ravignani and colleagues (2017), but also provide evidence that tracheal dimensions can be predicted by body mass. Previous literature found that the trachea may potentially convey body size information if its size influences acoustic call features (Ravignani et al., 2017). In humans, a wider tracheal diameter partially predicts turbulence (i.e., unsteady air movements) for large airflows (van den Berg et al., 1957). Applying the same logic to other mammals, larger seals would have wider tracheal dimensions, which, in turn, would make vocalisations noisier. This could explain, for instance, why the harmonics-to-noise ratio decreases as harbour seals get older (de Reus, 2017). Future work on sound production in this species could test this prediction using sound–anatomy correlations and excised larynx set-ups. Moreover, playback experiments could test whether adding noise to vocalisations alters interactive behaviour to determine whether the harmonics-to-noise ratio may encode body size information. Understanding whether and how the trachea is involved in sound production will thus require further research.

As expected, all anatomical measurements were larger for weaners than they were for pups. In the study by Ravignani and colleagues (2017), animals up to 108 days old were classified as pups. However, in the wild, the lactation period for harbour seals ranges from 23 to 42 days, after which the pups are weaned (Renouf, 1991). Hence, for the sake of simplicity, we considered animals up to 1 month old to be pups and animals older than 1 month to be weaners. Through this categorical classification, we were able to identify how allometric trends develop over the harbour seal's early life. At the time of data collection, we had very few larynges from subadults and adults, leading us to not include these data points in our analysis to avoid potential problems caused by small sample size. Future research including larynges from subadults and adults will further extend our knowledge of how vocal allometry develops in harbour seals. There were no sexual differences when considering the sample size as a whole, but significant sexual differences existed within age classes.[2] These differences might be attributed to differing levels of steroid hormones acting on the laryngeal structures in males and females (Aufdemorte et al., 1983; Sauvé et al., 2015). In some mammals, sex hormones affect the structural development of the larynx and the viscoelastic properties of the vocal fold tissue (Beckford et al., 1985; Fitch & Giedd, 1999). At puberty, in these animals, the male larynx descends in the vocal tract causing an elongation of the upper vocal tract, allowing males to convey an exaggerated impression of size (Fitch & Giedd, 1999; Fitch & Reby, 2001). In harbour seal pups of similar body size, males have larger VTLs than females, suggesting that laryngeal descent in males possibly occurs early in life. Once weaned, however, females show a clear increase in VTL, whereas it remains relatively constant in males (Figure 4.7, top), suggesting that VTL differences across sexes may become less pronounced over time. In mammalian males, sex hormone action also causes a rapid increase in cartilage size, leading to an enlarged larynx and an increase in the vibrating portion of the vocal folds (Fitch & Hauser, 2003). This could explain why, in weaners, STDV1 is larger in males than in females. Nevertheless, these findings are somewhat surprising as young harbour seals normally show little sexual dimorphism (Le Boeuf, 1991). In particular, there is a lack of evidence for sexual differences regarding birth mass and growth rates among harbour seal pups (Bowen et al., 1994). In our sample, there were no significant body size differences between sexes ($p > .05$); however, male pups were

---

[2] While no significant sex differences were observed when all individuals were considered together, sex differences emerged when individuals were analysed within their respective age classes. Specifically, male pups had significantly longer vocal tracts than female pups, and weaned males had a wider STDV1 than weaned females.

slightly larger than female pups in body length (males 81.6 ± 4.4 cm, females 77.5 ± 4.5 cm), which could partially explain the VTL differences observed in this age class. Male (9.8 ± 1.58 kg) and female (9.8 ± 1.6 kg) pups did not differ in body mass, but it is important to note that the sampled animals were sick and/or in poor condition; hence, body mass values are not representative of healthy individuals and should be interpreted with caution. In short, based on these observed differences in vocal anatomy across sexes, formants are expected to differ in pups and harmonics-to-noise ratio is expected to differ in weaners. The anatomical structures that determine these acoustic features both show strong allometric scaling; hence, these parameters may provide distinct body size cues across age classes, potentially facilitating the discrimination of male and female conspecific calls. Future research should investigate how sex hormones affect the elastic properties of harbour seal laryngeal tissues. Hormone levels can be measured by taking blood samples from healthy male and female seals at different developmental stages, and results can be combined with magnetic resonance imaging (MRI) mapping of laryngeal tissue elasticity.

The high inter-rater reliability observed for VTL, VFL and both STDVs demonstrates that these quantities can be measured and replicated easily, making them reliable landmarks for vocal tract measurements. However, tissue properties such as the viscoelasticity of certain vocal tract structures, like the vocal folds, are significant obstacles to getting accurate measurements. Indeed, raters struggled to produce precise data for VFT. Future research in the field of pinniped vocal anatomy would benefit from improved measuring techniques using 2D pictures, radiography, MRI and computed tomography scans as this would enable more accurate measurements for structures that are difficult to handle. Finally, future similar studies should include measurements of another vocal tract structure: the corniculate cartilage. Although widely absent in terrestrial carnivores, harbour seals have rather large corniculate cartilages that help close the trachea together with the epiglottis (Adams et al., 2020). These cartilages are located close to the vocal folds and are possibly innervated by the same nerves and controlled by the same muscles. It may be possible that these cartilages play a role in sound production by, for example, lowering the f_0 by adding weight to the vocal folds. Taken together, these suggestions will provide a more precise and detailed picture of the harbour seal's vocal anatomy.

Observed species-specific vocalisations are determined by both a species' vocal anatomy and their capacity for vocal learning (Garcia & Ravignani, 2020; Ravignani & Garcia, 2022). The vocal anatomy generates vocal predispositions by imposing biomechanical constraints, whereas neural processes determine the degree of control species have over their vocal organs (Garcia & Manser, 2020). Particularly, vocal learners, such as the harbour seal, are capable of actively modulating sounds, suggesting that they are less constrained by anatomy and have a refined capacity for vocal motor control. Unfortunately, the relative contribution of the two sound production mechanisms is unclear. In this study, we tested a hypothesis trying to segregate anatomical versus learning mechanisms (Garcia & Ravignani, 2020; Ravignani & Garcia, 2022). As shown here, by testing for allometric relationships between body size and vocal tract structures, one can start to disentangle the respective contributions of vocal anatomy and vocal motor control in shaping acoustic signals. We found that harbour seals are mechanistically constrained by their vocal anatomy, and have large vocal flexibility (Ralls et al., 1985; Torres Borda et al., 2021), which may result in the production of dishonest signals, thus pointing towards extensive volitional control over their vocalisations. In brief, we provide support for the morphology versus learning hypothesis, showing however that this relationship may be mediated by ontogeny.

In summary, we provide evidence of allometry between body size and vocal tract measurements in harbour seals. Body length is a strong predictor for VTL, VFL and STDV1, and body mass is a strong predictor for VTL and both tracheal measurements (STDVs). Age and sex are also important in predicting the dimensions of these anatomical structures. Taken together, the combined findings demonstrate that harbour seal vocal tracts do indeed scale with body size, although allometry between VTL and body size may only emerge after weaning. One could now make inferences about the vocal predispositions of harbour seals (e.g., f_0, formants), based on either their body size or the size of their vocal tract. However, to accurately predict f_0, further studies are needed in harbour seals to determine the range of stress they apply to their vocal folds while vocalising and to infer the tissue density of their vocal folds (Titze et al., 1989). Once such predictions are made, comparing them with data obtained from observed natural vocalisations would shed light on the range of vocal flexibility resulting from their extensive vocal motor control. Although formant spacing could be predicted from vocal tract length (Titze, 1994), it is still difficult to consistently extract formants from harbour seal vocalisations, meaning that predictions cannot currently be compared with observed vocalisations. Finally, a critical next step to directly relate acoustic features to sound production structures is to connect harbour seal vocal anatomy measurements to the vocalisations they produce while alive. Integrating such results with investigations of call function will eventually inform on which vocal structures are responsible for generating the individual- and species-specific information encoded in harbour seal vocalisations.

Acknowledgments

We are grateful to Letty Stupers, Rebecca Andreini and Maria Jose Robles for help in gathering and preparing samples for data collection. We thank Aline Joustra for her illustrations of seal vocal tracts. We also greatly appreciate the extensive comments provided by Taylor Hersh, Laura Verga and Yannick Jadoul. We would also like to thank all members and volunteers of Sealcentre Pieterburen for their continued support in this research. This work was supported by a Max-Planck-Gesellschaft (Max Planck Research Group, MPRG) and a Fonds Wetenschappelijk Onderzoek (FWO) Pegasus Marie Curie fellowship 12N5517N awarded to A.R. K.d.R. was supported by the Fonds Wetenschappelijk Onderzoek project "Interactive vocal rhythms" G034720N awarded to Bart de Boer. The German Ministry of Energy, Agriculture, the Environment, Nature and Digitalization (MELUND) funded the marine mammal stranding scheme and the necropsies in Schleswig-Holstein. Sealcentre Pieterburen funded the seal stranding scheme and the necropsies at Sealcentre Pieterburen. M.G. was supported by a research grant from the Ethologische Gesellschaft e.V., as well as a postdoctoral research grant from the Forschungskredit of the University of Zurich [grant no. FK-21-121].

Author contributions

Conceptualisation: K.d.R., S.G., M.G., A.R.-G., A.S.-C., A.R.; Methodology: K.d.R., S.G., A.R.-G., A.S.-C., A.R.; Software: S.G.; Validation: K.d.R., M.G., A.R.-G., A.S.-C., A.R.; Formal analysis: K.d.R., D.C., A.L.; Investigation: K.d.R., D.C., A.L., S.G.; Resources: S.G., A.R.-G., A.S.-C., A.R.; Data curation: K.d.R.; Writing—original draft: K.d.R., D.C., A.L., A.R.; Writing—review & editing: K.d.R., S.G., M.G., A.R.-G., A.S.-C., A.R.; Visualisation: K.d.R., D.C., A.L.; Supervision: A.R.; Project administration: K.d.R.; Funding acquisition: A.R..

Supplementary materials

Acoustic analyses

We ran several acoustic analyses to show that seals can escape acoustic allometry. The acoustic data used to perform these analyses comes from the study published by Torres Borda and colleagues (2021), during which they observed fundamental frequency changes in harbour seal pups under different noise conditions (silence, low and high noise). The acoustic data was complemented with previously unpublished body mass measurements of the animals in their study. The harbour seal pups were weighed on their day of arrival at Sealcentre Pieterburen, a pinniped rehabilitation centre in the Netherlands, where the animals were also audio recorded. Eight harbour seals participated in the noise playback experiment and the fundamental frequency (f_0) was extracted from the vocalisations they produced during the testing period (all details in Torres Borda et al., 2021).

Grouping the observations by seal ID and noise condition, we computed the median f_0 for each of the 24 groups. We then regressed median f_0 on body size (using body mass as a proxy for body size) for each of the noise conditions (Figure 4.2a). Visually, an inverse relationship between body size and call frequency seems to hold in all three noise conditions, but none of the correlations ($\tau_{silence}$ = -0.18, τ_{low} = 0.25, τ_{high} = -0.40) are significant ($p < .05$). This apparent inconsistency may be explained by a large degree of overlap in the range of f_0 values produced by individuals of differing body size between noise conditions. For instance, an animal of 12.4 kg under silence can produce a similar f_0 value as an animal of 7.3 kg under high noise (see Figure 4.2a), suggesting that acoustic allometry may not hold across noise conditions. Could it be that the environmental noise conditions in which vocalisations are produced more strongly affect the f_0 values than body size? If hypothetically we were to record calls of harbour seal pups on different days and irrespective of environmental noise conditions, the inverse relationship between f_0 and body size may disappear (i.e., acoustic allometry would break) if the individuals can, thanks to their large vocal plasticity, adjust their f_0 depending on the noise conditions.

To assess if allometric relationships do indeed break down across noise conditions, we computed 10,000 different combinations of randomly selected median f_0 values (1 of the 3 median frequency values per seal) and matched each value to the corresponding body mass value. We then performed 10,000 Kendall rank correlations, each among the 8 resulting pairs of f_0 and body mass values. Figure 4.2b shows the kernel density distribution of the resulting correlation coefficients and their associated *p*-values (Figure 4.2c). We find that the median correlation coefficient is -0.18, suggesting a weak negative correlation. The median *p*-value is 0.38, indicating that—in more than half of the cases—we cannot reject the null hypothesis which states that the correlations are generally not significantly different from 0. In only 2.2% of cases (217 out of 10,000) is the correlation significant. We should take care when interpreting the correlation *p*-values as the power of the test statistic is low given the small sample size ($n = 8$), resulting in a higher probability of committing type II errors. Moreover, the body mass values correspond to the measurements taken on the day of the animal's arrival at the Sealcentre; they are not representative of the actual body mass values on the days of testing. Using the same set of random combinations of f_0 values, we also plotted the density distribution for the linear regression coefficients (Figure 4.2d). The median regression coefficient is -10.8 Hz/kg. The difference in initial body mass between the largest and smallest seal is 5.1 kg. This means that across their mass range, we would expect, on average, a 55.08 Hz difference. For every seal, we calculated the range between the median f_0 values of the silent and high noise condition (silence f_0 – high f_0) and find that the median is 73.6 Hz. This suggests that the differences caused by individual variability in f_0 in response to noise conditions are larger than the f_0 differences expected from body mass differences alone. Seals of differing body sizes (e.g., 7 vs. 12 kg) could thus potentially produce the same f_0 value (and they actually do, see Figure 4.2a). Furthermore, we also calculated, for each seal, the f_0 range (maximum – minimum f_0) for all recorded observations from that individual. We find that, across the tested seals, the median f_0 range is 322.6 Hz. Applying the same logic as above, seals with a body mass difference of almost 30 kg (322.6 / 10.8) could all produce similar f_0 values. Finally, we computed and compared two simple generalised linear models, testing if body mass (Model 1: $f_0 \sim$ Body Mass) or noise condition (Model 2: $f_0 \sim$ Noise Condition) was better at predicting f_0. We find that body mass is not a significant predictor of f_0 ($t = -1.78, p = .09$), but noise condition is ($t_{high\ vs.\ low} = 2.10$, $p = .048$; $t_{high\ vs.\ silence} = 3.90, p = .001$). Moreover, Model 1 explained 12.63% of the deviance (calculated as (1 - residual deviance / null deviance) * 100) and Model 2 explained 42.05% of the deviance. An ANOVA test confirmed that Model 2 significantly outperformed Model

1 (F = 10.7, p = .001), showing that environmental noise conditions may have a stronger influence on f_0 than body size.

Table S4.1. List of sampled animals. Seals were from the Netherlands (NL) or Germany (DE). Sex is denoted as F for females and M for males.

ID	Age class	Where from	Sex	Body Length (cm)	Body Mass (kg)	Girth (cm)	Cause of death
1	weaner	NL	F	86	15.6	90	Euthanised
2	weaner	NL	F	99	17.3	83	Died during rehab
3	weaner	NL	M	96	26.8	76	Found dead in the wild
4	weaner	NL	M	96	22.9	71	Euthanised
5	weaner	NL	M	86	19.2	71	Died before rehab
6	weaner	NL	M	84	18.2	69	Euthanised
7	weaner	NL	M	89	19.9	66	Euthanised
8	weaner	NL	F	94	14.2	66	Euthanised
9	weaner	NL	F	92	15.8	65	Euthanised
10	weaner	NL	F	94	14.9	65	Died during rehab
11	weaner	NL	M	104	20.8	64	Died before rehab
12	weaner	NL	F	86	16.5	63	Died during rehab
13	weaner	NL	F	86	15.7	63	Euthanised
14	weaner	NL	M	100	18.37	62	Euthanised
15	weaner	NL	F	93	16.8	62	Euthanised
16	weaner	NL	M	114	18.8	61	Died before rehab
17	weaner	NL	F	87	15.8	61	Euthanised
18	weaner	NL	F	93	17.8	60.5	Euthanised
19	weaner	NL	M	96	16.3	60	Died during rehab
20	weaner	NL	F	82	15.3	60	Euthanised
21	weaner	NL	M	80	14.3	60	Died before rehab
22	weaner	NL	F	88	16.1	59	Died during rehab
23	weaner	NL	F	89	16.9	58	Euthanised
24	weaner	NL	F	71	10	58	Euthanised

ID	Age class	Where from	Sex	Body Length (cm)	Body Mass (kg)	Girth (cm)	Cause of death
25	weaner	NL	M	92	17	57	Euthanised
26	weaner	NL	F	94	14.5	57	Died during rehab
27	weaner	NL	M	86	13.9	57	Died before rehab
28	weaner	NL	F	79	11.9	56.5	Euthanised
29	weaner	NL	M	85	14.6	56	Died during rehab
30	weaner	NL	M	94	14.6	55	Euthanised
31	weaner	NL	F	92	13.7	55	Died during rehab
32	weaner	NL	M	80	13.1	55	Died during rehab
33	pup	NL	F	75	11.9	55	Euthanised
34	pup	NL	M	84	11.79	54	Found dead in the wild
35	weaner	NL	F	93	14	53	Died during rehab
36	pup	NL	M	83	11.47	52	Euthanised
37	weaner	NL	F	93	13.9	51.5	Euthanised
38	weaner	NL	M	86	13	51.5	Died before rehab
39	weaner	NL	F	87	12.4	51	Died before rehab
40	pup	NL	M	86	10.6	51	Euthanised
41	pup	NL	F	81	11.37	49.5	Euthanised
42	pup	NL	F	82	9.3	49	Found dead in the wild
43	pup	NL	M	80	9.46	47	Euthanised
44	pup	NL	M	73	8.6	46	Found dead in the wild
45	pup	NL	F	77	8.5	44.5	Died before rehab
46	pup	NL	F	80	9.63	44	Found dead in the wild
47	weaner	NL	M	87	9.3	44	Died during rehab
48	pup	NL	F	70	8	44	Found dead in the wild
49	pup	NL	M	87	9.43	41	Found dead in the wild
50	weaner	NL	F	77	7.47	40	Euthanised
51	pup	NL	M	80	7.28	38.5	Died during rehab
52	pup	NL	M	80	9.95	36	Died before rehab
53	weaner	DE	M	85.5	19.2	67.5	Mercy killed
54	weaner	DE	F	98.5	17	66	Mercy killed

ID	Age class	Where from	Sex	Body Length (cm)	Body Mass (kg)	Girth (cm)	Cause of death
55	weaner	DE	M	90	14.6	65	Mercy killed
56	weaner	DE	M	101	20.8	64	Mercy killed
57	weaner	DE	M	90	17	63	Found dead in the wild
58	weaner	DE	F	92	20.4	62	Found dead in the wild
59	weaner	DE	F	99	17.8	60.5	Mercy killed
60	weaner	DE	M	86	16	60.5	Found dead in the wild
61	weaner	DE	M	90	16.6	59	Found dead in the wild
62	weaner	DE	F	94	17	58	Found dead in the wild
63	weaner	DE	M	90	14.6	57	Mercy killed
64	weaner	DE	M	96	13.4	56	Found dead in the wild
65	weaner	DE	F	82	11.6	55.5	Found dead in the wild
66	weaner	DE	F	97	14.8	54	Mercy killed
67	weaner	DE	M	81	10.2	51	Mercy killed
68	weaner	DE	F	88.5	12	50	Mercy killed

Table S4.2. Pairwise Spearman correlations for pups and weaners. * indicates *p* < .05 after correcting for multiple comparisons using the Holm-Bonferroni method.

Age class	Variable	Body Length (cm)	Body Mass (kg)	Girth (cm)	VTL (mm)	VFL (mm)	VFT (mm)	STDV1 (mm)
Pups	Body Mass (kg)	0.40						
	Girth (cm)	0.23	0.72					
	VTL (mm)	0.23	-0.03	-0.04				
	VFL (mm)	0.22	0.08	-0.01	0.49			
	VFT (mm)	-0.21	0.11	0.42	0.01	0.16		
	STDV1 (mm)	0.20	0.18	0.08	0.49	0.79*	0.55	
	STDV2 (mm)	0.35	0.36	0.38	0.26	0.71	0.57	0.76*
Weaners	Body Mass (kg)	0.51*						
	Girth (cm)	0.28	0.76*					
	VTL (mm)	0.39*	0.48*	0.39*				
	VFL (mm)	0.58*	0.64*	0.41*	0.54*			
	VFT (mm)	0.16	0.48*	0.50*	0.22	0.34		
	STDV1 (mm)	0.38*	0.61*	0.38*	0.46*	0.67*	0.47*	
	STDV2 (mm)	0.32	0.48*	0.31	0.47*	0.57*	0.30	0.66*

Table S4.3. Generalised linear model (GLM) estimates for all vocal structures. The vocal tract structures tested are vocal tract length (VTL), vocal fold length (VFL), subglottic tracheal dorsoventral distance 1 (STDV1) and subglottic tracheal dorsoventral distance 2 (STDV2). For all models, the reference level for Age Class is "Pup" and the reference level for Sex is "Female".

Vocal structure	Effect	Estimate	St. Err.	2.5%	97.5%	*p*
VTL	Intercept	42.4788	9.2445	23.9898	60.9678	< .001
	Age Class-Weaner	4.6695	1.7801	1.1093	8.2297	< .05
	Body Length	0.4170	0.1184	0.1802	0.6538	< .001
	Body Mass	0.5933	0.2182	0.1569	1.0297	< .01
	Sex-Male	40.7192	12.1685	16.3822	65.0562	< .01
	Body Length*Sex-Male	-0.4610	0.1379	-0.7368	-0.1852	< .01
VFL	Intercept	9.1651	3.0598	3.0455	15.2847	< .01
	Age Class-Weaner	-2.7492	3.2348	-9.2188	3.7204	0.399
	Body Length	-0.0189	0.0396	-0.0981	0.0603	0.635
	Body Mass	0.1050	0.0301	0.0448	0.1652	< .001
	Sex-Male	1.0070	0.3670	0.2730	1.7410	< .01
	Age Class-Weaner*Body Length	0.0562	0.0411	-0.026	0.1384	0.177
	Age Class-Weaner*Sex-Male	-1.1833	0.4071	-1.9975	-0.3691	< .01

STDV1	Intercept	15.389	1.6827	12.0236	18.7544	< .001
	Age Class-Weaner	1.7474	0.4352	0.8770	2.6178	< .001
	Body Length	0.0472	0.0236	0.0000	0.0944	< .05
	Body Mass	0.1887	0.0533	0.0821	0.2953	< .001
	Sex-Male	0.5956	0.2754	0.0448	1.1464	< .05
STDV2	Intercept	9.9726	1.8526	6.2674	13.6778	< .001
	Age Class-Weaner	1.5194	0.4792	0.5610	2.4778	< .01
	Body Length	0.0427	0.0260	-0.0093	0.0947	0.105
	Body Mass	0.1560	0.0587	0.0386	0.2734	< .01
	Sex-Male	0.4523	0.3032	-0.1541	1.0587	0.141

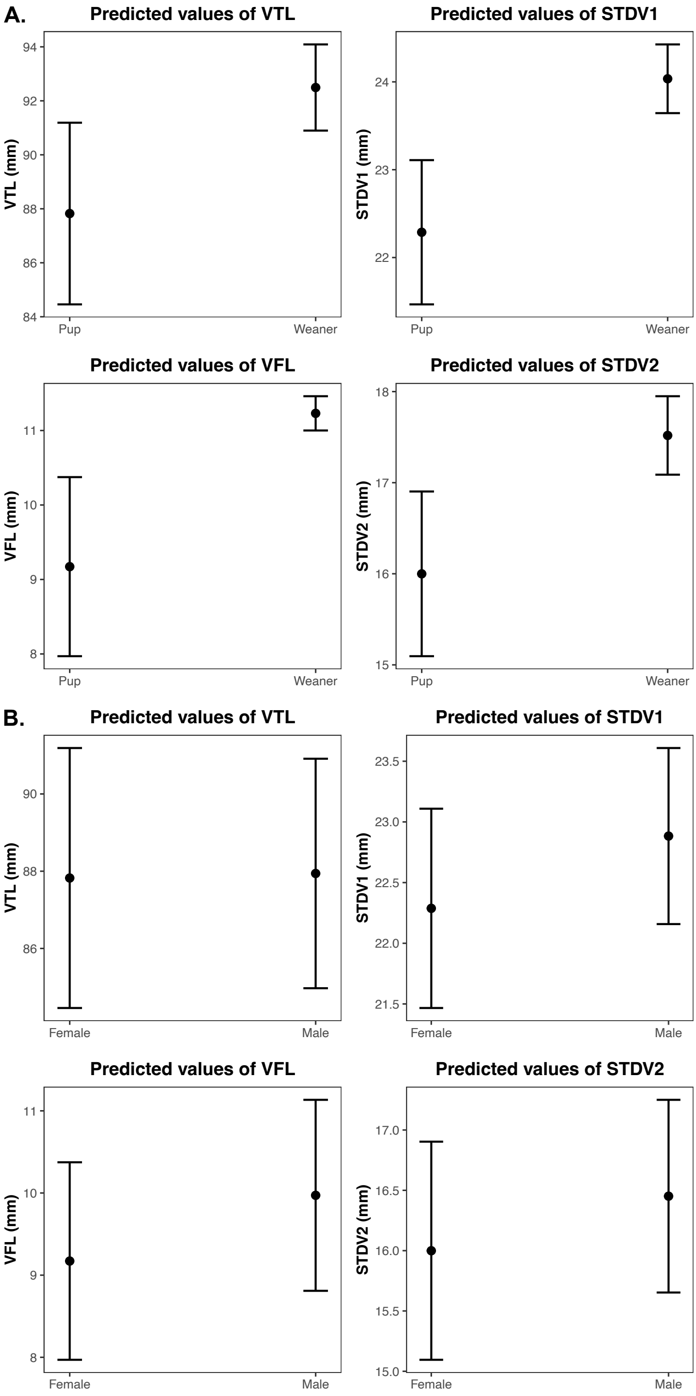


Figure S4.1. Predicted effects of (a) Age class and (b) Sex in each of the GLM models.

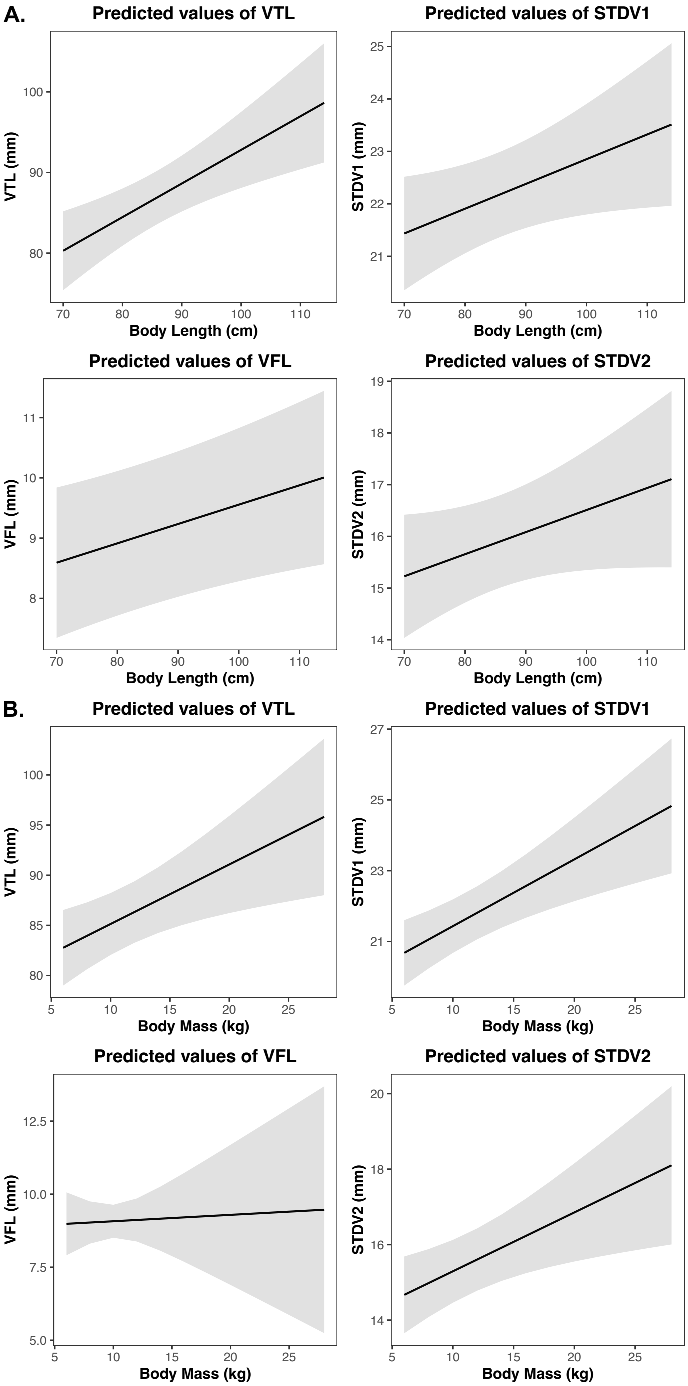


Figure S4.2. Predicted effects of (a) Body Length and (b) Body Mass in each of the GLM models.

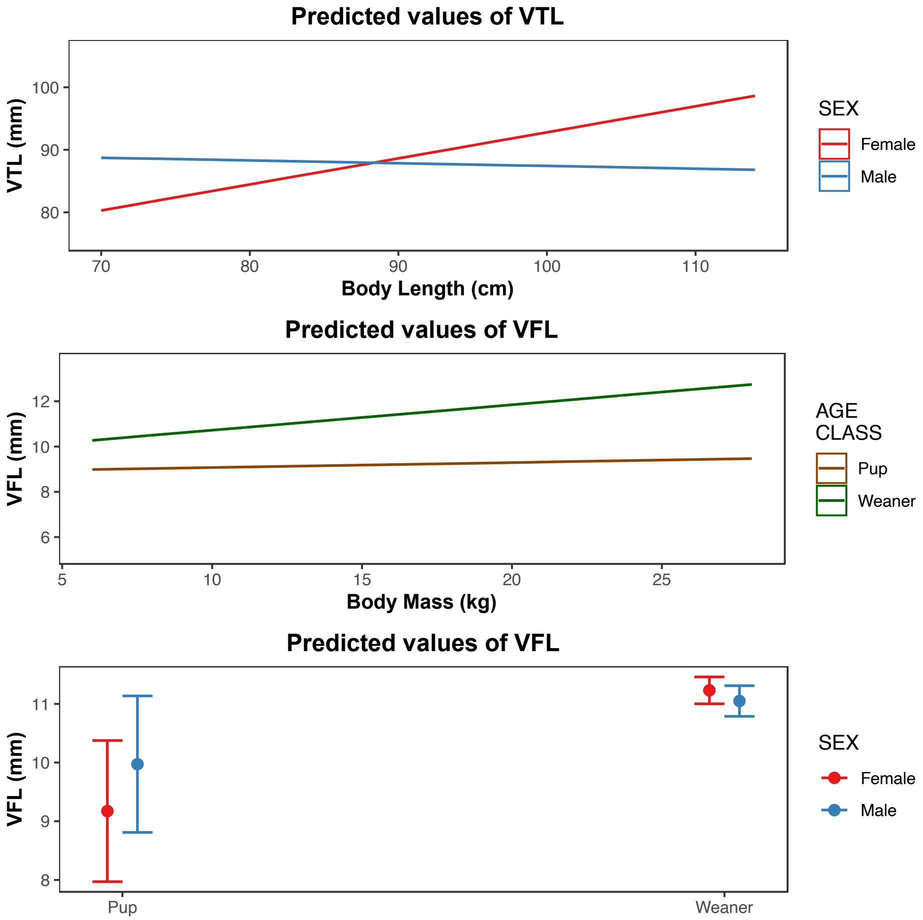


Figure S4.3. Predicted effects of the body length and sex interaction for VTL (top), the body mass and age interaction for VFL (middle) and the age and sex interaction for VFL (bottom).

Chapter 5

5 | Socially driven vocal convergence in harbour seal pups

Abstract

To help answer how language evolved in humans, comparative scientists have looked for animals that share our propensity for social interaction. Vocal accommodation—the ability to modify existing call types—is used to express social closeness or distance during social interactions. Due to its remarkable vocal imitative abilities, the harbour seal was proposed as a promising animal model for comparative work on the evolution of speech, although evidence for socially mediated vocal accommodation during conspecific interactions has been lacking. This study tested harbour seal pups' ability for social vocal accommodation by disentangling the effects of physiological changes due to maturation from the social influences on vocal development. Study results confirmed previous findings that the acoustic structure of harbour seal pup vocalisations was affected by age, sex, body weight, and the interaction between age and sex. Notably, after accounting for maturation effects, harbour seal pups showed vocal convergence towards pups they frequently interacted with, while also remaining sufficiently identifiable to allow for individual recognition. These findings provide key missing evidence of socially driven vocal accommodation in harbour seals, indicating they may share similar predispositions for social interaction with humans.

Introduction

The human "interaction engine" hypothesis suggests that language evolved thanks to the cognitive abilities and social behaviours that predispose us to communicative interaction (Levinson, 2006). To establish if humans share these predispositions with other species, the field of language evolution began studying how interactive communication modulates social behaviour in non-human animals (Bradbury & Vehrencamp, 2011; Fröhlich et al., 2019; Hauser, 1996; Pika et al., 2005; Seyfarth & Cheney, 2003). Since vocal signals can be transmitted over long distances, many animals use them to develop and maintain social relationships, particularly in environments with limited visibility, like oceans and forests (Dudzinski et al., 2009; Marten & Marler, 1977). Studying social behaviours that are maintained through vocal signalling in non-human animals, such as competition for space, access to mates, and parent–offspring recognition can thus provide interesting insights into the evolution of both animal vocal communication and spoken language.

Vocal production learning (henceforth referred to as vocal learning)—the ability to modify or learn vocalisations based on auditory experience—is a fundamental prerequisite for the evolution of speech (Fitch, 2000; Jarvis, 2019). Recent frameworks suggest that vocal learning is best understood as a continuous trait, extending from vocal non-learners to advanced vocal learners (Arriaga & Jarvis, 2013; Petkov & Jarvis, 2012), with evidence for limited vocal learning found in many animals other than humans (Tyack, 2019). While advanced vocal learners can match vocal output to auditory templates, vocal accommodation—the modification of existing call types in response to social and environmental influences—is observed in a wider range of species (e.g., Briefer & McElligott, 2012; Lameira et al., 2010). In light of this, vocal accommodation has become recognised as an integral part of vocal learning frameworks (Ruch et al., 2018; Wirthlin et al., 2019).

Two human-centred theories address the mechanisms and functions of vocal accommodation (Tinbergen, 1963). First, the Interactive Alignment Model (IAM; Pickering & Garrod, 2004) proposes that successful communication involves alignment at multiple levels of linguistic representations (e.g., lexical, syntactic, semantic, phonological). In conversation, such alignment would lead to increased similarity between speakers over time due to an automatic priming mechanism linking production and perception. Second, the Communication Accommodation Theory (CAT; Giles & Baker, 2008) explains how individuals accommodate their communication to that of their conversation partner. According to CAT, people use two types of strategies to express social closeness or distance during interaction, namely convergence (i.e., becoming more similar) and divergence (i.e., becoming less similar). The IAM and CAT both emphasise alignment in communication but from different perspectives. While both theories emphasise alignment in interactive communication, the IAM explains the automatic cognitive processes of alignment, whereas CAT addresses how social motivations influence these processes—either reinforcing or overriding automatic priming based on the desire to signal social closeness or distance. Together, they offer a comprehensive understanding of how cognitive processes and social goals interact to influence real-time vocal modifications during social interactions.

Understanding whether vocal accommodation in non-human animals is also driven by automatic priming mechanisms (as suggested by IAM) and socially mediated motivations (as proposed by CAT) can provide insights into the evolutionary origins of vocal communication systems, such as human speech. However, before we can determine whether the cognitive mechanisms underlying vocal accommodation are shared across species, it is essential to first establish if other species exhibit comparable behaviour with similar functions. Beyond humans, vocal convergence has been reported, among others, in Guinea baboons (Fischer et al., 2020), African penguins (Baciadonna et al., 2022), and Egyptian fruits bats (Prat et al., 2017), and likely serves to form new social bonds and maintain group cohesion. Vocal divergence in non-human animals, like group-specific calls, generally develops over longer timescales and co-occurs with vocal convergence (e.g., diverging from an out-group to increase in-group cohesion; Boughman, 1997; Crockford et al., 2004), but there are reports of animals showing accommodation during vocal exchanges, as found in Diana monkeys (Candiotti et al., 2012) and orange-fronted conures (Balsby & Bradbury, 2009). Since numerous non-human mammals and birds are capable of vocal accommodation (for a review, see Ruch et al., 2018), it may be that the mechanisms underlying vocal accommodation in humans are shared with other species.

Pinnipeds (seals, sea lions, and walruses) are gaining much interest as promising animal models for the comparative study on speech origins (Ravignani et al., 2016). Many pinniped species have an extensive call repertoire and produce vocalisations in different behavioural contexts ranging from competition for mates to parent–offspring recognition (Charrier et al., 2009; Mathevon et al., 2017). Specifically, the harbour seal (*Phoca vitulina*) stands out for its advanced vocal learning abilities. One famous study reported that an adult harbour seal, named Hoover, could imitate human speech by repeating words it heard from its caretaker after it was separated from its mother as a pup (Ralls et al., 1985), suggesting that these animals can match their vocal output to auditory templates. Moreover, harbour seal pups can also show vocal adjustment in response to environmental conditions by lowering their fundamental frequency in the presence of background noise (Torres Borda et al., 2021), potentially to avoid acoustic overlap. However, these two studies represent the only evidence for vocal learning in this species[3], and further evidence of vocal learning—like socially driven vocal accommodation—is still lacking. Testing whether social interactions influence vocal behaviour can help address this knowledge gap and offer deeper insights into their vocal capacities.

Harbour seal vocal communication during early development plays a role in individual recognition and potentially in social bonding as well. In their first weeks of life, harbour seal pups produce vocal signals known as mother attraction calls (MACs) but these disappear entirely from their vocal repertoire once weaned, suggesting these calls play an important role in parent–offspring recognition during lactation (Renouf, 1984). MACs are individually distinctive (Perry & Renouf, 1988; Renouf, 1984) and mothers can discriminate between recordings of different pups shortly after birth (Renouf, 1985). Harbour seal mothers regularly forage at sea during lactation and return to the breeding colony to find their pups (Riedman, 1990). Offspring survival thus depends on the pups' ability to emit individualised signals and the mothers' ability to accurately identify their pups' signals when returning from foraging trips (Insley et al., 2003; Tibbetts & Dale, 2007), thereby avoiding misdirected parental care and possible risks of genetically unrelated adults injuring pups (Harcourt, 1992). However, while the mothers do not vocalise, harbour seal pups find themselves in close proximity to other calling pups during lactation, raising the question of how their social environment impacts their vocal development.

---

[3] This statement refers specifically to vocal production learning, which involves modifying existing or acquiring novel vocalisations. It should not be conflated with forms of contextual vocal learning, such as vocal usage or comprehension learning (see Duengen et al., 2024, for work on these forms in harbour seals).

To fully appreciate the extent to which social interactions influence vocal development, it is important to determine how their vocal signals change as they mature into adults (Schusterman, 2008; Shapiro et al., 2004). Two previous studies characterised how the acoustic structure of harbour seal MACs developed throughout the lactation period (Khan et al., 2006; Sauvé et al., 2015). Study findings showed that acoustic parameters such as the fundamental frequency (f_0), ascending and descending modulation of f_0, the dominant frequency, and call duration were possibly individualised. Both studies also reported that individual differences in the acoustic structure of MACs are, in part, driven by differences in age, sex, and body length. These findings are perhaps unsurprising given that the production of vocal signals depends on vocal anatomy. In harbour seals, age and body length are positively correlated with the length of the vocal tract and of the vocal folds, and with the tracheal diameter (de Reus et al., 2022), indirectly affecting acoustic parameters like the f_0, formant frequencies, and harmonics-to-noise ratio. Moreover, sex hormones lead to differential growth rates of laryngeal structures in males and females (Aufdemorte et al., 1983). Longitudinal studies that disentangle the effects of maturation and learning on the acoustic structure of MACs are thus crucial for understanding the interplay of physiological and social influences on vocal development.

In the present longitudinal study, I focus on how maturation and social interactions affect vocal development in harbour seals, expanding on previous ontogenetic work in this species (Khan et al., 2006; Sauvé et al., 2015) by including daily recordings of a larger number of harbour seal pups and different acoustic parameters. I aim to (i) further characterize how the acoustic structure of MACs develops during their first weeks of life and investigate the effects of age, sex, and body size on acoustic parameters, and (ii) identify evidence for vocal accommodation based on the frequency of social interactions. I hypothesise that several, if not all, acoustic parameters will be influenced by maturational changes related to age, sex, and body size. Moreover, based on CAT, I predict that the MACs of harbour seal pups that frequently interact with each other will become more similar, thereby showing socially driven vocal convergence. Although the present study does not directly test the mechanisms proposed by IAM, such vocal convergence may also be consistent with automatic priming processes linking perception and production. Demonstrating vocal convergence would offer much-needed experimental evidence of socially driven vocal accommodation in harbour seals and provide a more comprehensive understanding of MAC development.

Materials and methods

Experimental design and subjects

Aerial vocalisations were recorded from 64 harbour seal pups of the Eastern Atlantic subspecies (*Phoca vitulina vitulina*) from May 29th to July 28th of 2017. This study was conducted at Sealcentre Pieterburen in the Netherlands, a hospital that admits orphaned, sick, and injured seals for rehabilitation before releasing them back into the wild. Throughout the study, on-site veterinarians assessed the seals' health to determine if they were fit enough to participate. Seals with known respiratory tract issues were not recorded. Each harbour seal pup participating in the study was recorded daily during the first phase of rehabilitation, typically lasting 2–4 weeks depending on individual circumstances.

The subjects, born in the wild, were admitted for rehabilitation between 0–10 days postpartum. Upon arrival at the Sealcentre, veterinarians sexed the pups and estimated their age based on the condition of the umbilical cord and the closing of the umbilicus. Morphological measurements were also taken, including body length and weight, from each individual. The length of the pups was measured twice, first at intake and then at release. Animals were weighed at intake, followed by further weekly measurements during their rehabilitation, and again at release. It is important to note that the growth rate of body weight for harbour seals in rehabilitation is significantly lower than that of their wild-raised counterparts (MacRae et al., 2011; Richmond et al., 2010). Body mass values in this study are therefore not representative of healthy individuals and should thus be interpreted with caution. According to Sealcentre protocols, the pups were either housed in pairs or alone (3 individuals).

Recording material and procedure

Researchers recorded the vocalisations with a Sennheiser ME-66 unidirectional microphone (flat frequency response ranging from 40–20,000 Hz ± 2,5 dB; Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) connected to a Zoom H6 digital recorder (Zoom Corporation, Tokyo, Japan). The recordings had a 48 kHz sampling frequency with 24-bit quantisation and were saved as uncompressed WAV files. To protect the microphone from water splashes, we equipped it with an MZW-66 foam windshield.

All subjects were recorded at intake, when veterinarians performed a physical health check of the animal, prior to being placed in one of three block facilities containing first phase units. Each unit was sampled daily during one of the four feeding sessions (early morning, late morning, afternoon, and evening), prior to handling of the animals, with each session lasting 10 minutes. Researchers made the recordings from a distance of about one meter away from the subjects. Flipper tags and fur patterns were used to accurately distinguish between pups housed in the same unit. During the recording, the researcher called out the pup's name immediately after it produced a vocalisation to ensure that I assigned the vocalisations to the correct individual during acoustic analysis. Recordings stopped once the animals started the next phase of rehabilitation and were moved to different locations.

Data annotation and pre- processing

The dataset consists of roughly 100 hours of harbour seal pup recordings. To my knowledge, the scientific literature on acoustic signalling in harbour seal pups does not describe any call types other than the MAC. Therefore, all calls identified in the recordings were assumed to be MACs. I marked the onset and offset of all calls found in the audio recordings, and labelled MACs with visible formants and no temporarily overlapping sounds (e.g., calls from a partner, banging against a metal gate) as "u", treating them as good-quality calls, while labelling all other calls as "p". Continuous background noise, such as that coming from running water and electrical appliances, was ignored and later addressed during extraction of the acoustic parameters. To minimise potential annotation bias, I annotated the recordings in random order. In total, I annotated 55,840 calls, of which 6,494 were good-quality MACs. Data annotation was done in the software Praat (v6.0.28; Boersma & Weenink, 2017), applying custom settings determined through visual inspection of the spectrogram (i.e., view range = 0–5000 Hz, window length = 0.05 s, dynamic range = 70 dB, time step = automatic). Unfortunately, due to time constraints, these call annotations were not validated by a second annotator.

Calls were removed from the annotated dataset for several reasons. First, out of the 64 harbour seals recorded in this study, six died during rehabilitation. One of these animals was euthanised because of a megaoesophagus, a condition where the oesophagus becomes enlarged and loses its ability to efficiently move food to the stomach (Field et al., 2018; Mace et al., 2012). Animals afflicted by such a condition regurgitate their food and cough regularly to ensure that regurgitated material is not inhaled (Mace et al., 2012). Moreover, they can experience pressure on the airways due to their enlarged oesophagus (Dominguez et al., 1987; Maayan et al., 1990; Panzini & Traube, 1993), potentially leading to the production of unusual sounds. Since its condition may have affected its vocalisations, MACs from this animal were removed from the dataset. Second, calls of harbour seal pups produced during the intake were also removed to ensure that only calls recorded in the same acoustic environment (i.e., their assigned first phase unit) were retained. Finally, one seal was moved from block B to a smaller, separate unit (with different acoustics) during its initial phase of rehabilitation. All calls produced by this animal in the new location were removed to ensure that the acoustic environment remained as consistent as possible, meaning that only calls produced in the main first phase units were retained. Taken together, removing these calls reduced the number of good-quality MACs to 5,674 from 62 animals.

Extraction of acoustic parameters and post-processing

All good-quality MACs were retained instead of including only the 5 highest quality calls, as done in Garcia et al. (2016). By including all observations, I hoped to avoid selection bias and make sure that my results were representative of the natural course of vocal development in harbour seal pups. Prior to extraction of the acoustic parameters, the duration of each good-quality MAC was trimmed by a standardised amount to account for the reverberation in the first phase facilities. A custom Python script using the packages TextGridTools (v1.4.4; Buschmeier & Włodarczak, 2013) and Parselmouth (v0.4.4; Jadoul et al., 2018) was written to extract 11 acoustic parameters (7 spectral and 4 temporal) from the MACs (Table 5.1). Note that, when extracting the fundamental frequency (f_0), the first and last 10% of the call duration were ignored because pitch was not always established at the onset and offset of the call. The reverberation calculations and the advanced Praat settings used to extract each parameter can be found in the supplementary materials.

While recording the animals and annotating the dataset, I noticed that one pup was vocalising at a fundamental frequency (f_0) that was much higher than all of the other individuals. Calls from this seal were later removed from the annotated dataset because its f_0 was estimated at about 1.2 kHz, whereas all other seals had a f_0 below 800 Hz (see Figure S5.1). Removal of this individual further reduced the number of good-quality MACs to 5,492 from 61 animals (30 females).

Table 5.1. Summary of acoustic variables extracted from good-quality MACs.

Acoustic variable	Description
1. f_0	Lowest frequency value (Hz)
2. Minimum f_0	Value at the 5th percentile of f_0 (Hz)
3. Maximum f_0	Value at the 95th percentile of f_0 (Hz)
4. Time of minimum f_0	Location of value at the 5th percentile of f_0 (% of call duration)
5. Time of maximum f_0	Location of value at the 95th percentile of f_0 (% of call duration)
6. Dominant frequency	Frequency of the highest amplitude (Hz)
7. Centre of gravity	Weighted mean of all observed frequencies (Hz)
8. Harmonics-to-noise ratio	Relative proportion of harmonic sound to noise (dB)
9. Maximum intensity	Highest intensity value (dB)
10. Time of maximum intensity	Location of maximum intensity (% of call duration)
11. Duration	Call length (ms)

Interpolation of body size values

Since body weight and length were not measured daily, I interpolated body size values for each seal by fitting a model to the available size data to obtain unique values for each recording day. Although previous work found that young harbour seals grow linearly in body weight (Markussen et al., 1989) and length (Hauksson, 2006), visual assessment of the relationship between weight and length, and age using scatterplots indicated a non-linear pattern, which seems better described by a quadratic function. Using the *lmer* function (lmerTest R package v3.1.3; Kuznetsova et al., 2017), generalised linear mixed effects models (GLMMs) helped predict body size values as a function of age. The weight model included both a random slope and a random intercept. The length model included only a random intercept because each pup was only measured twice, once at intake and again before release. An analysis of variance (ANOVA) was used to compare the fit of linear and quadratic models, with the quadratic model performing significantly better for both weight

($\chi^2 = 485$, $p < .001$) and length ($\chi^2 = 10.7$, $p < .01$). Although more complex models generally provide better fits, the selection of the quadratic model was further supported by visual inspection and confirmed by lower AIC values. The performance of each quadratic model was evaluated by extracting the conditional goodness-of-fit value, which considers the variance explained by all fixed and random effects. The conditional goodness-of-fit was 97.6% for the weight model and 83.3% for the length model. Visual assessment of the model residuals using QQ plots and residual plots showed that the residuals were normally distributed and had equal variances (see Figures S5.2–5.3). Model coefficients were then used to interpolate weight and length values for each seal across the recording period.

Bayesian generalised linear mixed effects models

Bayesian GLMMs were used to estimate how well the combination of the fixed effects age, sex, and their interaction, as well as body weight and length, and a random intercept for each seal could predict each acoustic call variable, as shown in Equation (1). Bayesian models estimate the probability of a hypothesis given the data, directly quantifying uncertainty in parameter values through probability distributions (Migon et al., 2014). This uncertainty is intuitively represented using a credible interval (CI)—a range of values within which the true parameter value is likely to fall. Parameters whose CI do not include zero are considered statistically significant (Hespanhol et al., 2019). The CIs directly reflect the uncertainty in the model's predictions. In contrast, frequentist models estimate the probability of the data given a hypothesis and quantify uncertainty of parameter values using hypothesis tests like p-values (Migon et al., 2014), which are often misinterpreted. In short, Bayesian models can test a hypothesis directly instead of trying to reject the null hypothesis. A second advantage of using Bayesian models is that prior knowledge from previous studies can be incorporated in the models, allowing them to update the probability of a hypothesis (Migon et al., 2014). For example, in this study, I used previous parameter values from Khan et al. (2006) and Sauvé et al. (2015) to inform the prior probabilities for the models pertaining to the fundamental frequency, the dominant frequency, and call duration. Specifically, the informed priors had a normal distribution centred at the mean of the available parameter values and large parameter-specific standard deviations, resulting in weakly informative priors. For all other acoustic variables, I used uniform non-informative priors for each model parameter.

Note that Bayesian GLMMs with uniform, non-informative priors produce parameter values similar to those from frequentist models (Lemoine, 2019).

$$Acoustic\ variable \sim Age + Sex + Weight + Length + Age * Sex + \ (1 \mid ID) \qquad (1)$$

Bayesian modelling was done using the brms R package (v2.21.6; Bürkner, 2017). I specified each model using four Markov Chain Monte Carlo (MCMC) chains, 5000 iterations per chain, and a warm-up period of 50% of the total number of iterations before sampling (van Ravenzwaaij et al., 2018). Using variance inflation factors (VIFs) from the performance R package (v0.10.3; Lüdecke et al., 2021) I checked that model parameters were not multicollinear, with VIFs below 5 indicating low correlation and between 5 and 10 indicating moderate correlation. If VIFs were larger than 10, the parameter with the highest VIF was removed from the model and the model was refitted without that parameter. I performed prior sensitivity analyses for the three models that included informed priors using the priorsense R package (v1.0.2; Kallioinen et al., 2023). If the priors were deemed too strong, their standard deviation was increased, and the model was refitted with the new priors. For all models, I ensured that the MCMC chains converged using trace plots and checked that all R-hat values (i.e., statistical measure of convergence of the MCMC chains to the target distribution) were close to 1.00, but always below 1.1 (Gelman & Rubin, 1992). I also checked for the presence of outliers or influential values using Pareto shape parameters, with values larger than 0.7 indicating observations that greatly affect the posterior. I then extracted parameter estimates for each fixed effect and computed the model’s goodness-of-fit (Lüdecke et al., 2021). Finally, I visually assessed the model assumptions of residual normality and equal variance. No posterior predictive checks were performed for the models as this study aims to explore the size and direction of age, sex, and body size effects on each acoustic feature, as opposed to trying to find the best model fit.

Discriminant function analysis

Harbour seal pups in this study were housed in one of three facilities used for the first phase of rehabilitation. Although seal pups were housed in units either in pairs or alone, they were in acoustic contact with other pups housed in similar units within the same facility. If harbour seal pups are capable of vocal accommodation, I expect pups housed in the same facility to sound more similar to each other compared to pups housed in a different facility. Moreover, despite vocally converging towards seals housed in a shared social environment, I expect seal pups to remain individually identifiable, thereby facilitating reliable mother–offspring recognition. To test both hypotheses, I performed a stepwise discriminant function analysis (DFA) that aimed to (i) extract the acoustic distances between pairs of seals, and (ii) attribute calls to the correct individuals. I first removed seals that produced less than 12 calls, given that 11 acoustic variables were extracted from the MACs and that the number of variables included in a DFA should always be smaller than the lowest number of samples per individual (Mundry & Sommer, 2007). Then, using the function *greedy.wilks* (klaR R package v1.7.3; Weihs et al., 2005), I selected the acoustic variables that best discriminated between individuals.

Pairwise acoustic distances

To compare how the acoustic structure of MACs differed between seals, pairwise acoustic distances were extracted based on the original variables instead of extracting them from the discriminant functions obtained after dimensionality reduction. I first computed a multivariate analysis of variance (MANOVA) to analyse individual differences by regressing seal identity onto the combination of acoustic variables extracted from the MACs. I then included the MANOVA results in the *mvpaircomp* function (biotools R package v4.2; da Silva, 2021) and, using the Wilks method, computed approximated F-values for each possible seal pair. These approximated F-values can serve as a proxy for the acoustic distance between individuals in the original, multivariate feature space, as done in Pfefferle and colleagues (2016). Similarly, for each unique pairwise combination, I determined if the paired seals were housed in the same (within group) or in a different facility (between group) and extracted their respective mean age, weight, and sex differences. Finally, using *lmer* (lmerTest R package v3.1.3; Kuznetsova et al., 2017), I built a GLMM where the pairwise F-values were included as dependent variable, and, as fixed effects, I included the variable group, but also the pairs' respective age, weight, and sex difference to account for their possible effects. I also included two random intercepts for the facilities each of the paired individuals were housed in, thereby accounting for the different acoustics between housing locations. Once the model was fitted, I visually assessed the model assumptions of residual normality and homoscedasticity. If assumptions were not met, the dependent variable was log-transformed and the GLMM fitted anew.

Classification accuracies

I performed a linear discriminant analysis (LDA) using a leave-one-out cross validation method to predict how accurately MACs could be assigned to the correct seals. The LDA was fitted using the *lda* function (MASS R package v7.3.61, Venables & Ripley, 2002) and used all acoustic features to correctly predict which seal produced each MAC. I then computed the classification accuracy score as a percentage. To check how the classification accuracy compared to chance levels, I computed 1000 different LDAs using datasets in which seal identity was randomised. Permuted datasets are used to test the null hypothesis that there is no relationship between the predictors and the response (Mundry & Sommer, 2007). In other words, the cross-validated classification accuracies of the LDAs using the permuted datasets were used to form the distribution of values representing chance levels (i.e., the null distribution). I then compared how often the classification accuracy scores from permuted LDAs were equal or higher to the real score obtained from the original LDA. A probability value below 0.05 would indicate that the original LDA can assign MACs to the correct seals at levels that are significantly above chance.

Results

General call characteristics

The reduced dataset contains 5,492 MACs from 61 animals (30 females). On average, the animals were 16 (13–20) days old, weighed 9.62 (8.21–10.35) kilograms, and measured 80.20 (77.83–81.40) centimetres. These values represent the median and, in parentheses, the range between the first and third quartiles. Table 5.2 reports summary statistics for each acoustic variable, including the median, minimum and maximum, and the first and third quartiles.

Table 5.2. Summary statistics of harbour seal pup mother attraction calls.

Acoustic variable	Minimum	Q1	Median	Q3	Maximum
f_0 (Hz)	272.69	375.08	402.40	432.68	789.68
Minimum f_0 (Hz)	252.93	328.72	359.25	394.76	775.08
Maximum f_0 (Hz)	276.15	409.24	449.67	525.95	797.09
Time of minimum f_0 (%)	10.00	22.33	69.02	81.56	89.99
Time of maximum f_0 (%)	10.05	22.98	38.50	57.03	89.99
Dominant frequency (Hz)	275.39	398.44	451.17	808.59	1710.94
Centre of gravity (Hz)	346.69	554.07	661.41	855.55	1654.22
Harmonics-to-noise ratio (dB)	-2.69	5.85	9.75	14.74	32.56
Maximum intensity (dB)	50.25	75.46	79.74	83.79	93.23
Time of maximum intensity (%)	3.08	35.80	46.23	57.49	97.27
Duration (ms)	128.65	540.13	661.08	802.38	1747.91

Bayesian generalised linear mixed effects models

Bayesian GLMMs were computed for all 11 acoustic variables extracted from the pup MACs. After specifying each model as shown in (1), I noticed that body length had a VIF value above 10 in all of the models. To solve this multicollinearity problem, all models were refitted without the fixed effect of body length. Only low to moderate levels of correlation between model parameters were present in the GLMMs after removal of body length. For the models informed by previous work, prior sensitivity analyses indicated that the original priors were too strong and were consequently readjusted to ensure that their effect on the posterior was not too strong. The adjusted priors and the corresponding prior sensitivity analyses results can be found in Table S5.1 the supplementary materials. The MCMC chains converged for all models and the R-hat values did not exceed 1.1 (Gelman & Rubin, 1992). No outliers were identified in any of the models. The effects of the intercept, age, sex, body weight, and the interaction of age and sex on each acoustic feature can be found in Table 5.3. Model goodness-of-fit values ranged from 0.05-0.49, suggesting that the model parameters did not explain all the variance observed in the data for each acoustic feature. The distribution of the model residuals did not always seem to follow a normal distribution, with the tail ends showing deviations from the QQ-line (Figure S5.4). The residual variance, however, was rather homogenous, given the custom settings used to extract the acoustic features in this study (Figure S5.5).

Age, sex, body weight, and interaction effects on the acoustic structure of MACs

Older individuals produced calls with a lower f_0, but with a higher dominant frequency and centre of gravity (Figure 5.2). The maximum f_0 was observed later in the call as they grew older. Calls of older seals were longer in duration, less harmonic, and had a lower maximum intensity. Male harbour seal pups produced MACs with a lower f_0 (Δ f_0 = 39.09 Hz) and minimum f_0 (Δ min f_0 = 31.99 Hz). Males also produced longer calls (Δ duration = 140.76 ms). Moreover, compared to females, time of maximum f_0 occurred later in male calls, while the time of maximum intensity occurred earlier. Heavier seal pups produced calls with a lower maximum f_0 and centre of gravity (Figure 5.3). They also produced calls of longer duration.

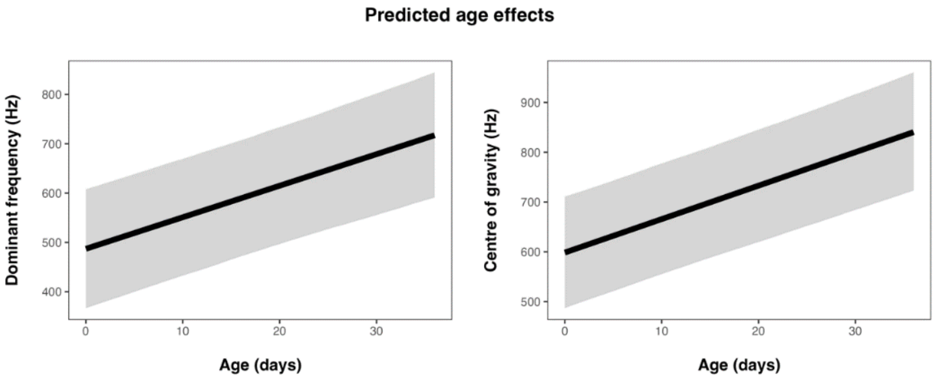


Figure 5.2. Conditional age effects for the dominant frequency and the centre of gravity. The linear effect is represented by the black regression line and the shaded grey area represents the credible interval. Note that the estimates of the other model parameters are kept constant.

Table 5.3. Parameter estimates from the Bayesian GLMMs for each acoustic call feature of harbour seal pup MACs. Values represent the linear regression slopes for age, sex (female referent), body weight, and the interaction between age and sex. The 95% credible interval (95% CI) quantifies uncertainty in the estimates, replacing the error term. Cells shaded in light grey indicate statistically significant effects (Hespanhol et al., 2019), where zero is not included in the 95% CI.

Acoustic variable	**Intercept**	**Age (days)**	**Sex (referent F)**	**Body weight (kg)**	**Age*Sex**	**Goodness of fit (R^2)**
	Est. *[95% CI]*	*Est.* *[95% CI]*	*Est.* *[95% CI]*	*Est.* *[95% CI]*	*Est.* *[95% CI]*	
f_0 (Hz)	432.23 [407.71, 456.20]	-0.89 [-1.53, -0.27]	-39.64 [-58.71, -20.69]	0.40 [-2.58, 3.45]	0.81 [0.37, 1.28]	0.48
Minimum f_0 (Hz)	373.82 [348.31, 398.34]	-1.03 [-1.68, -0.38]	-31.99 [-52.01, -13.50]	2.45 [-0.65, 5.66]	0.52 [0.05, 1.00]	0.43
Maximum f_0 (Hz)	512.71 [474.83, 549.01]	0.47 [-0.59, 1.53]	-19.08 [-44.54, 5.91]	-4.94 [-9.72, -0.02]	-0.52 [-1.40, 0.35]	0.22
Time of minimum f_0 (%)	58.99 [48.90, 69.23]	-0.13 [-0.43, 0.18]	-3.66 [-10.37, 2.95]	0.32 [-0.99, 1.58]	0.13 [-0.18, 0.43]	0.06
Time of maximum f_0 (%)	36.97 [29.41, 44.18]	0.23 [0.00, 0.46]	6.05 [1.09, 10.97]	-0.09 [-0.98, 0.87]	-0.30 [-0.53, -0.06]	0.05
Dominant frequency (Hz)	607.37 [482.01, 732.16]	6.60 [3.25, 9.94]	-50.76 [-121.35, 19.07]	-10.59 [-26.65, 5.93]	-1.68 [-4.20, 0.84]	0.25
Centre of gravity (Hz)	747.23 [640.94, 850.15]	6.75 [3.92, 9.50]	-33.53 [-112.83, 46.13]	-13.99 [-27.08, -0.26]	-1.83 [-3.95, 0.27]	0.32
Harmonics-to-noise ratio (dB)	15.64 [12.77, 18.60]	-0.14 [-0.21, -0.06]	-1.21 [-3.41, 0.99]	-0.20 [-0.55, 0.15]	0.09 [0.05, 0.14]	0.49
Maximum intensity (dB)	80.39 [77.89, 82.83]	-0.08 [-0.16, -0.01]	0.69 [-0.97, 2.33]	0.02 [-0.30, 0.35]	-0.11 [-0.17, -0.04]	0.13
Time of maximum intensity (%)	56.06 [50.04, 62.27]	-0.08 [-0.25, 0.10]	-5.57 [-9.47, -1.75]	-0.67 [-1.46, 0.08]	0.35 [0.18, 0.51]	0.06
Duration (ms)	368.75 [264.56, 464.70]	3.57 [0.86, 6.17]	142.31 [77.67, 208.26]	23.39 [11.00, 36.62]	-8.05 [-10.09, -6.01]	0.27

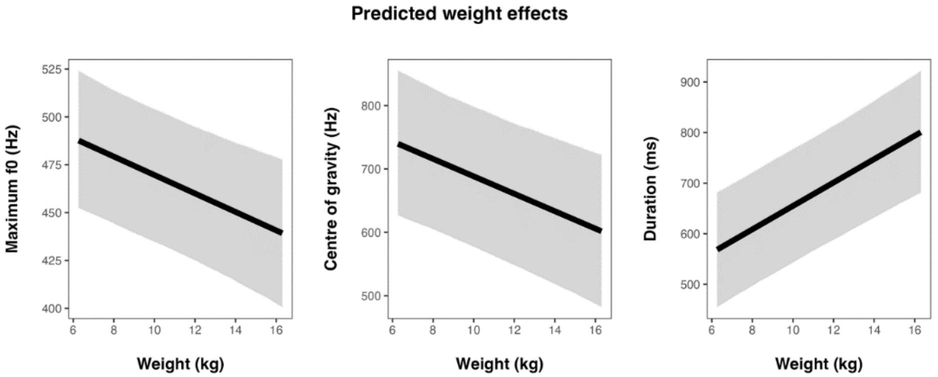


Figure 5.3. Conditional weight effects for the maximum fundamental frequency (maximum f_0), the centre of gravity, and call duration. The linear effect is represented by the black regression line and the shaded grey area represents the credible interval. Note that the estimates of the other model parameters are kept constant.

Several acoustic parameters showed an interaction effect between age and sex (Figure 5.4). As animals aged, the f_0 and the harmonics-to-noise ratio decreased for both sexes, but the decrease was much more pronounced in females. In contrast, for the maximum intensity, the decrease was more pronounced in males. At birth, the time of maximum f_0 occurred earlier in female calls, but the opposite became true as animals grew older. Conversely, the time of maximum intensity at birth occurred earlier in males, while in older individuals, it occurred earlier in females. Finally, the call duration increased over time in females, whereas it decreased in males.

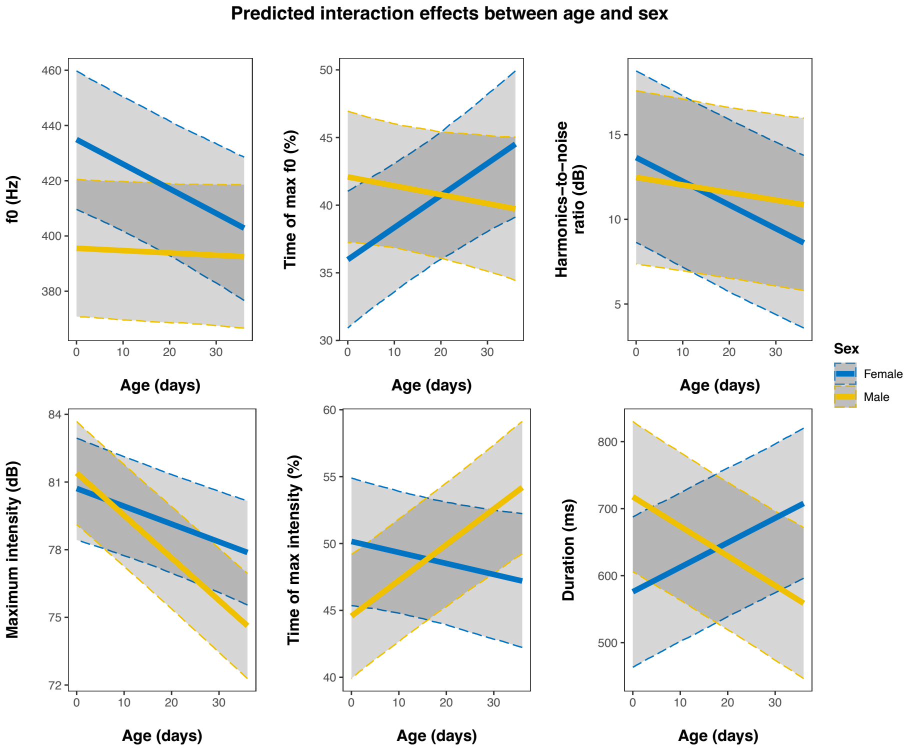


Figure 5.4. Conditional interaction effects between age and sex for f0, the time of maximum f0, harmonics-to-noise ratio, maximum intensity, time of maximum intensity, and call duration. Females are shown in blue and males in yellow. Linear effects are represented by the coloured regression lines, whereas the credible intervals are shaded grey with their limits defined by dashed lines. Note that the estimates of the other model parameters are kept constant.

Discriminant function analysis

A stepwise DFA was performed to compare pairwise acoustic distances within and between groups of seals, and to assess how accurately MACs could be attributed to the correct individuals. After filtering out seals that did not meet the minimum threshold requirement of producing at least 12 calls, the reduced dataset contained 5,441 MACs from 53 harbour seal pups. All of the 11 acoustic parameters were included in the analysis, given that all were useful in discriminating between different individuals, with f0, harmonics-to-noise ratio, centre of gravity, and call duration providing the largest contributions.

Pairwise acoustic distances

After fitting the GLMM using the non-transformed F-values and noticing that the model failed to meet the assumption for homogeneity of variance, the F-values were log-transformed, and the model was refitted. The acoustic distance between seals was significantly different depending on whether they were housed in the same or a different facility ($t = -2.76$, $p < .01$; Figure 5.5). Specifically, the acoustic distance between seals housed in the same facility was smaller than between those housed in different facilities. In addition, sex differences significantly affected pairwise acoustic distances ($t = 2.55$, $p < .05$), with pairs of same-sex individuals exhibiting smaller acoustic distances compared to those between males and females. Assumptions for residual normality and equal variance were visually assessed using QQ and residual plots, respectively (Figure S5.6).

Classification accuracies

The LDA classified 31.32% of all MACs to the correct seals. This cross-validated classification accuracy score was significantly better than the chance levels obtained through permutation ($p = .00$), which ranged from 4.61% to 6.58%.

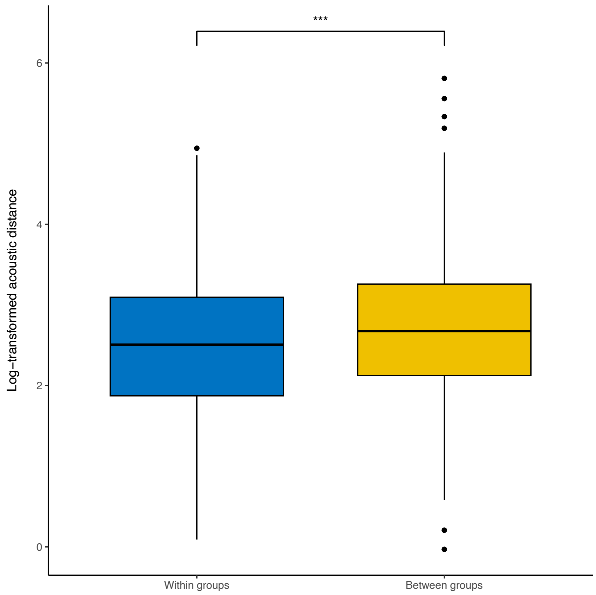


Figure 5.5. Boxplots showing the distributions of the pairwise log-transformed approximated F-values observed between seals housed in the same facility and in a different facility. The distributions are significantly different from each other, as observed from the output of the GLMM. Note that the distributions plotted here do not account for the age, sex, and weight differences between paired seals.

Discussion

The present study showed that the acoustic structure of harbour seal pup MACs is affected by age, sex, body weight, and the interaction between age and sex. Results also showed that the acoustic distance between seals housed in the same facility is shorter than the distance between seals housed in different facilities. This finding indicates that harbour seals are capable of vocal convergence towards familiar conspecifics that they frequently interact with. Finally, results showed that pup MACs can be accurately assigned to the correct individuals, at levels that are significantly above chance.

The first aim of this study was to test if age, sex, body weight, and the interaction between age and sex influence the acoustic parameters that characterise MACs. The results of the Bayesian GLMMs confirmed our hypothesis, as well as some findings from previous vocal developmental work in this species (Khan et al., 2006; Sauvé et al., 2015). Vocal tract structures grow in size as seal pups mature and given that the production of MACs is constrained by their vocal anatomy, maturation will indirectly cause changes to the acoustic structures of MACs (de Reus et al., 2022).

Study findings indicate that older pups and males produced MACs with a lower f_0. The observed decrease in f_0 as pups grow older can be explained by the growth of the vocal folds (Taylor & Reby, 2010). Effects of sex on f_0 could be attributed to differing levels of testosterone across sexes acting on both the size and viscoelastic properties of the vocal folds (Aufdemorte et al., 1983; Beckford et al., 1985). In addition, the action of sex hormones causes a rapid increase in cartilage size, and consequently a larger larynx and longer vocal folds (Fitch & Hauser, 2003), which translates to a lower f_0. These findings partially align with those of Sauvé and colleagues (2015) who also reported that MACs of older pups showed a lower f_0, but instead found that males had a higher f_0 than females. The different sex effect in both studies could be ascribed to differences in experimental design, for example, in terms of the animals recorded (i.e., wild vs. captive), the recording conditions (i.e., handling vs. no handling), and model specifications (with body length vs. with body weight). The current study also reported a decrease in the maximum f_0 with increased body weight, which can likely be explained by the inverse relationship between body size and f_0, observed in both carnivores and primates (Bowling et al., 2017). This explanation is further supported by the presence of body weight and the interaction between age and body weight being retained as model predictors for the size of the vocal folds (de Reus et al., 2022), the anatomical proxy for f_0. A final notable f_0-related effect found in this study is that the location of the maximum f_0 occurs later in the call of older seals. The location of the maximum f_0 can reflect the emotional or motivational state of an animal, with earlier peaks being tied to positive emotions in humans, whereas later frequency peaks indicate increased arousal (Briefer, 2012). Although more psychoacoustic work in non-human animals is needed to understand how acoustic parameters reflect emotions, it could be that an increase in the location of the maximum f_0 signals a more urgent need for food in harbour seal pups. The pups need to grow as much as possible before their mothers leave and they have to learn to feed themselves, hence any additional food they can receive during lactation should increase their chances of survival (K. C. Harding et al., 2005). Future bioacoustics studies should thus aim to describe the contexts in which vocalisations are recorded to better understand the emotional information that acoustic features could potentially convey.

This study also showed that older individuals produce MACs that are less harmonic and have a longer duration. This decrease in harmonicity could be explained by the wider tracheal diameter of older pups (de Reus et al., 2022), potentially causing more unsteady air movements (i.e., turbulence) and noisier vocalisations, as found in humans (van den Berg et al., 1957). The increase in call duration with age could be attributed to their growing lung capacity (Fahlman et al., 2014). Call duration and harmonics-to-noise ratio are also known vocal correlates of arousal (Filippi et al., 2017), hence these findings could also reflect changes in the emotional and motivational state of older pups. Comparing results to previous work, Sauvé and colleagues (2015) found that older individuals produced calls with a longer duration. However, males in this study produced longer MACs than females, whereas in Sauvé and colleagues (2015), the females produced longer calls. Moreover, when looking at the interaction between age and sex in this study, call duration increased in females but decreased in males, whereas Khan and colleagues (2006) found the opposite effect. Given reports of geographical variation in adult call structure across harbour seal populations (Bjørgesæter et al., 2004; van Parijs et al., 2000, 2003), these differential effects for duration could potentially be attributed to the different subspecies studied. Specifically, this study reported on the Eastern Atlantic harbour seal (*P. v. vitulina*), whereas the other studies reported on the Western Atlantic (*P. v. concolor*; Sauvé et al., 2015) and the Eastern Pacific (*P. v. richardii*; Khan et al., 2006) harbour seal. Nonetheless, study findings clearly demonstrate that physiological changes associated with maturation, such as those related to age, sex, and body size, significantly influence the vocal development of MACs across subspecies.

The results of this study provide an extension to previous ontogenetic work done in harbour seals (Khan et al., 2006; Sauvé et al., 2015). The choice for a Bayesian alternative to the usual frequentist GLMMs was motivated by the available prior knowledge from these earlier studies. Unfortunately, due to differences in the acoustic parameters chosen for the analysis, few parameters could be informed by prior knowledge, which constitutes a clear limitation. For example, this study did not look at frequency modulation because visual assessment of the raw data indicated that not all MACs had the "inverted u-shape" described in the previous studies. Khan and colleagues (2006) only extracted parameters from the dominant frequency, and Sauvé and colleagues (2015) also looked at energy quartiles and amplitude ratios. Future work in bioacoustics should aim to define a standardised set of acoustic parameters when characterising non-human animal vocalisations, thereby allowing researchers to better inform their statistical models using prior knowledge. Given the integration of prior knowledge, results are not statistically independent from the studies to which they are compared. Researchers may doubt the validity of comparing findings between studies, especially given their a priori differences in terms of the animals tested (wild vs. captive), recording conditions (handling vs. no handling), and harbour seal subspecies. Prior sensitivity analyses were thus performed to test that the posterior was largely influenced by the data and the likelihood, and not by the prior (Table S5.1), making the priors in this study only weakly informative. Despite their differences, these three studies provide a solid baseline of how harbour seal pup MACs develop during lactation, enabling future comparisons across different populations and species.

The second aim of this study was to test if harbour seal pups showed vocal accommodation. Specifically, it aimed to test if the MACs of pups that repeatedly interact with each other will sound more similar compared to those of pups they do not interact with. As predicted, results showed that the acoustic distances between seals housed in the same facility were significantly smaller than the acoustic distances between seals housed in a different facility. However, to properly determine if harbour seals were capable of vocal accommodation, it was important to account for maturational effects during lactation, such as the changes in acoustic structure related to age, sex, and body size (Khan et al., 2006; Sauvé et al., 2015). The linear mixed-effects model confirmed that, after accounting for maturational effects, the housing conditions significantly predicted pairwise acoustic distances, thereby highlighting the importance of the social environment on vocal development. Specifically, the present study finds that the MACs of harbour seal pups converge towards those of conspecifics as a result of repeated vocal interactions. Such vocal accommodation is consistent with the interaction engine hypothesis, which posits that social interactions play an important role in the emergence of advanced communication systems. Interestingly, the vocal convergence observed between the calls of harbour seal pups suggests that MACs may serve a communicative function that goes beyond parent–offspring recognition. Moreover, these findings suggest that the mechanisms behind interactive vocal communication, which are essential for human speech acquisition, may have deeper evolutionary roots shared across species.

Given that harbour seals have previously demonstrated a remarkable affinity for vocal learning (Ralls et al., 1985), finding evidence of social vocal accommodation in harbour seals is perhaps unsurprising but nonetheless, it constitutes an important contribution to a more comprehensive understanding of their vocal abilities. According to CAT, vocal accommodation is influenced by social motivations and manifests as either convergence or divergence, depending on the desire to express social closeness or distance (Giles & Baker, 2008). While the present study clearly showed that harbour seals are capable of vocal convergence, it remains unclear whether this convergence is driven by positive social interactions with conspecifics. Future work could address this by incorporating a video camera into the experimental set-up and annotating non-vocal behaviours. Paired seals exhibiting more affiliative behaviours (e.g., nuzzling) should also show larger degrees of vocal convergence. If vocal accommodation is found to play a role in the formation and maintenance of social bonds, it could be established that this behaviour is driven by similar social motivations as in humans, allowing us to further explore whether the underlying cognitive mechanisms are shared. It is possible that harbour seals use a similar automatic priming mechanism linking perception and production, as proposed by IAM, particularly considering previous evidence of imitative abilities in this species (Ralls et al., 1985), where seals visibly match vocal output to an auditory template. Neurobiological studies (e.g., Cook et al., 2025; Hoeksema et al., 2021) and vocal learning experiments on trained animals (e.g., Duengen & Ravignani, 2024; Duengen et al., 2024) could provide additional insights into the cognitive mechanisms underlying vocal accommodation in harbour seals.

Finally, the present study also found that MACs could be attributed to the correct seals at levels significantly above chance, indicating that seals diverge from one another and that their MACs are distinct enough to allow for individual recognition. However, it remains unclear which acoustic parameters are used for individual recognition in this species and how these parameters change over time. Future work could test the importance of acoustic parameters for individual recognition through playback experiments with mother–pup pairs. After recording the pup’s MACs, researchers could manipulate recordings one acoustic variable at the time, play the altered sounds back to the mother, and infer each parameter's contribution by assessing the mother's reaction upon hearing the altered MAC. It is interesting to note that pup MACs remain sufficiently individualised to allow for individual recognition, despite vocally converging towards conspecifics their social environment, highlighting the possible competing pressures of individual recognition and social bonding on vocal development.

This longitudinal study on the vocal development of harbour seals expands on earlier ontogenetic work in this species by including more animals and more acoustic parameters, confirming previous findings that MACs are subject to maturational effects and showing acoustic changes with age, sex, body weight, and the interaction between age and sex. Additionally, this study provides evidence of vocal accommodation in response to social influences, showing that the MACs of seal pups can vocally converge towards those of other pups, while remaining individualised enough to allow for individual recognition. These findings offer a more comprehensive understanding of vocal development in harbour seals and further reinforce the harbour seal's potential as a comparative model for studying the evolution of (spoken) language.

Acknowledgments

I am grateful to Letty Stupers and Andrea Ravignani for their help in recording the animals, and to all the members and volunteers of Sealcentre Pieterburen for facilitating these recordings. I also would like to extend a big thank you to Yannick Jadoul for writing the custom Python script used to extract the acoustic parameters from the vocalisations.

Supplementary materials

Reverb calculation

Given that harbour seal pups were housed inside of first phase facilities, the mother attraction calls (MACs) they produced inside these closed units were reflected off the different surfaces, causing reverberation. The reverberation differed depending on the resonant properties of each facility. Unfortunately, during the annotation process, it was difficult to mark exactly where the call ended, and the reverberation began. It is thus not unreasonable to assume that I consistently annotated part of the reverberation during this process. To account for this, I measured the reverberation inside each of the first phase facilities by playing and recording pure tones of 20, 50, and 200 ms, respectively. The recordings were performed twice, using different recorder gains (4 and 5,5) to reflect the gain range used while recording harbour seals MACs in this study. The onset and offset of the recorded tones were then marked in Praat, similarly as when annotating MACs, but this time by two raters. For all possible combinations of pure tones, gains, and raters, I then subtracted the length of the pure tones from the annotation length to obtain a duration estimate for the observed reverberation. The median value was then taken as the estimated reverberation length, which ranged between 108–127 ms depending on the first phase facility.

Praat settings for acoustic parameter extraction

Pitch parameters (f_0, minimum and maximum f_0, jitter)

- *method* = "AC"
- *pitch_floor* = 250
- *pitch_ceiling* = 800*
- *silence_threshold* = 0.1
- *voicing_threshold* = 0.2
- *octave_cost* = 0.1
- *voiced_unvoiced_cost* = 0.3
- *octave_jump_cost* = 0.5

**pitch_ceiling* was set at 1,500 Hz (i.e., 1.5 kHz) for the individual known to produce MACs with a f_0 above 1 kHz

Intensity parameters (maximum intensity)

- *minimum_pitch* = 250
- *time_step* = None
- *subtract_mean* = True

Harmonicity parameters (harmonics-to-noise ratio)

- *method* = "CC"
- *time_step* = 0.01
- *minimum_pitch* = 250
- *silence_threshold* = 0.1
- *periods_per_window* = 5

Spectrogram parameters (dominant frequency, centre of gravity)

- *window_length* = 0.05
- *maximum_frequency* = 10000
- *time_step* = 0.002
- *frequency_step* = 10
- *window_shape* = "GAUSSIAN"

Table S5.1. Prior sensitivity analyses of the Bayesian linear regression models informed by priors, as determined by the *powerscale_sensitivity* function (priorsense R package; Kallioinen et al., 2023). The posterior should be largely influenced by the data and the likelihood, and not by the prior. The model should thus show low prior sensitivity, indicated by values below 0.05. Likelihood sensitivities below 0.05 indicate weak or non-informative likelihood.

Variable	**Prior ($\mu \pm \sigma$)**	**Prior sensitivity**	**Likelihood sensitivity**	**Diagnosis**
Model: f_0				
Age	-0.4 ± 3	0.015	0.178	-
Sex	25.9 ± 50	0.043	0.047	-
Weight	-	0.008	0.161	-
Age*Sex	-1.0 ± 3	0.024	0.172	-
Model: *dominant frequency*				
Age	0.8 ± 6	0.031	0.047	-
Sex	-40.1 ± 60	0.040	0.033	-
Weight	-	0.032	0.061	-
Age*Sex	-4.7 ± 5	0.019	0.069	-
Model: *duration*				
Age	0.4 ± 6	0.034	0.187	-
Sex	-44.1 ± 170	0.044	0.042	-
Weight	-	0.016	0.253	-
Age*Sex	4.9 ± 10	0.044	0.098	-

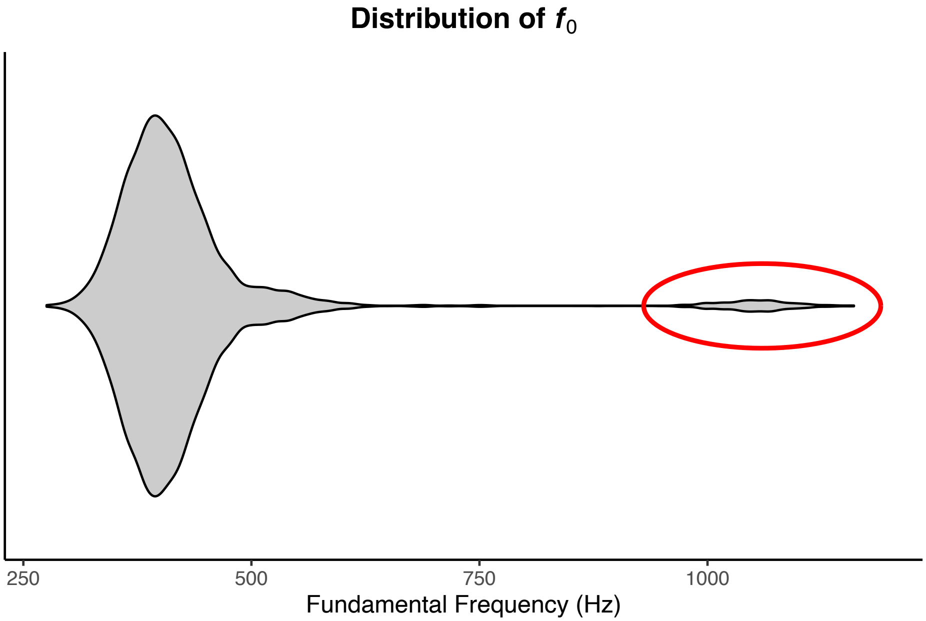


Figure S5.1. Fundamental frequency distribution of all the harbour seal pups in the annotated dataset. The red circle shows all the f_0 estimated from calls from the harbour seal pup named Bigos. Note that calls from this individual were removed from the annotated dataset because its f_0 was consistently 2–3 times higher than all the other seals.

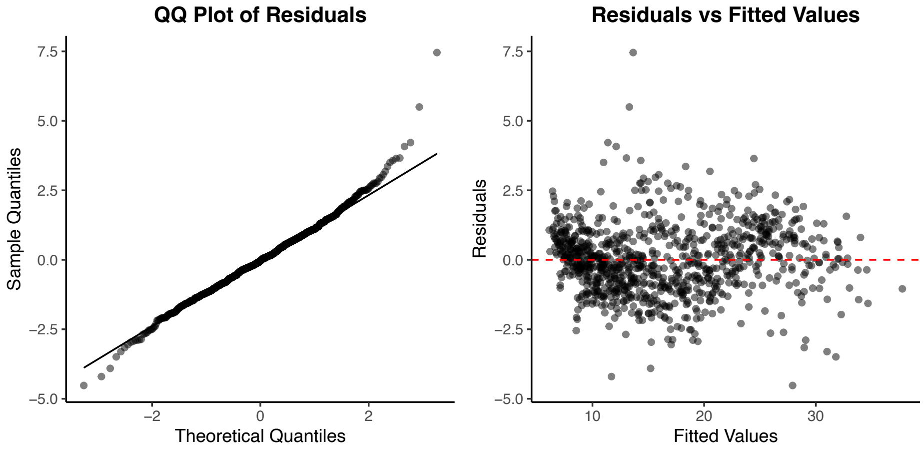


Figure S5.2. Assumptions of the weight model, showcasing the model residuals' relative normality (QQ plot on the left) and equal variance (residual plot on the right).

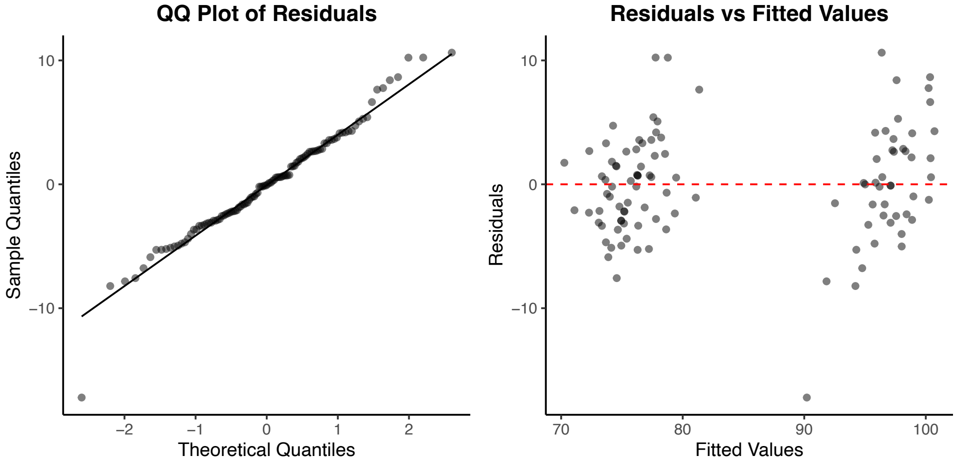


Figure S5.3. Assumptions of the length model, showcasing the model residuals' relative normality (QQ plot on the left) and equal variance (residual plot on the right).

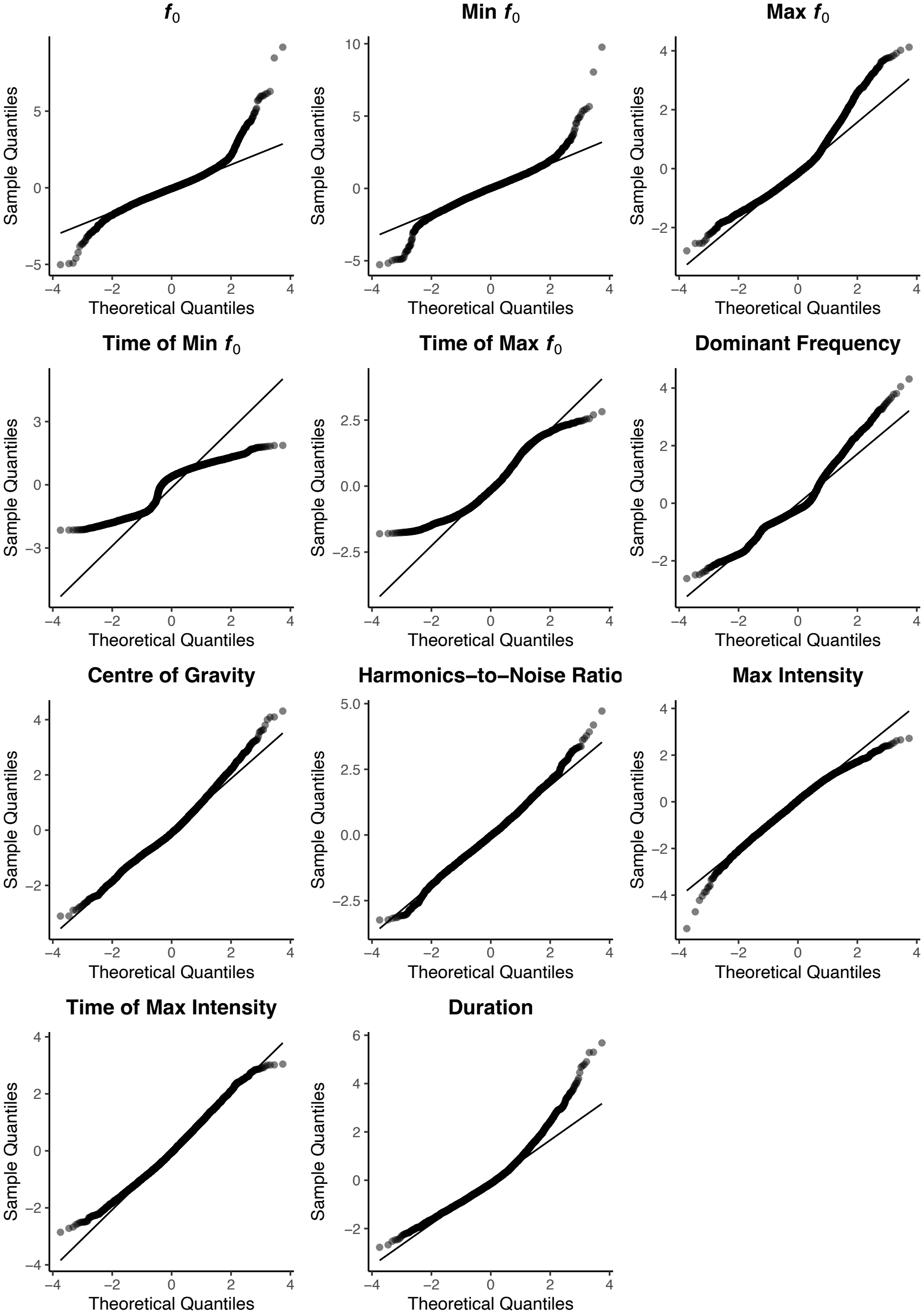


Figure S5.4. QQ plots used to visually assess if the residuals of the Bayesian regression models are normally distributed.

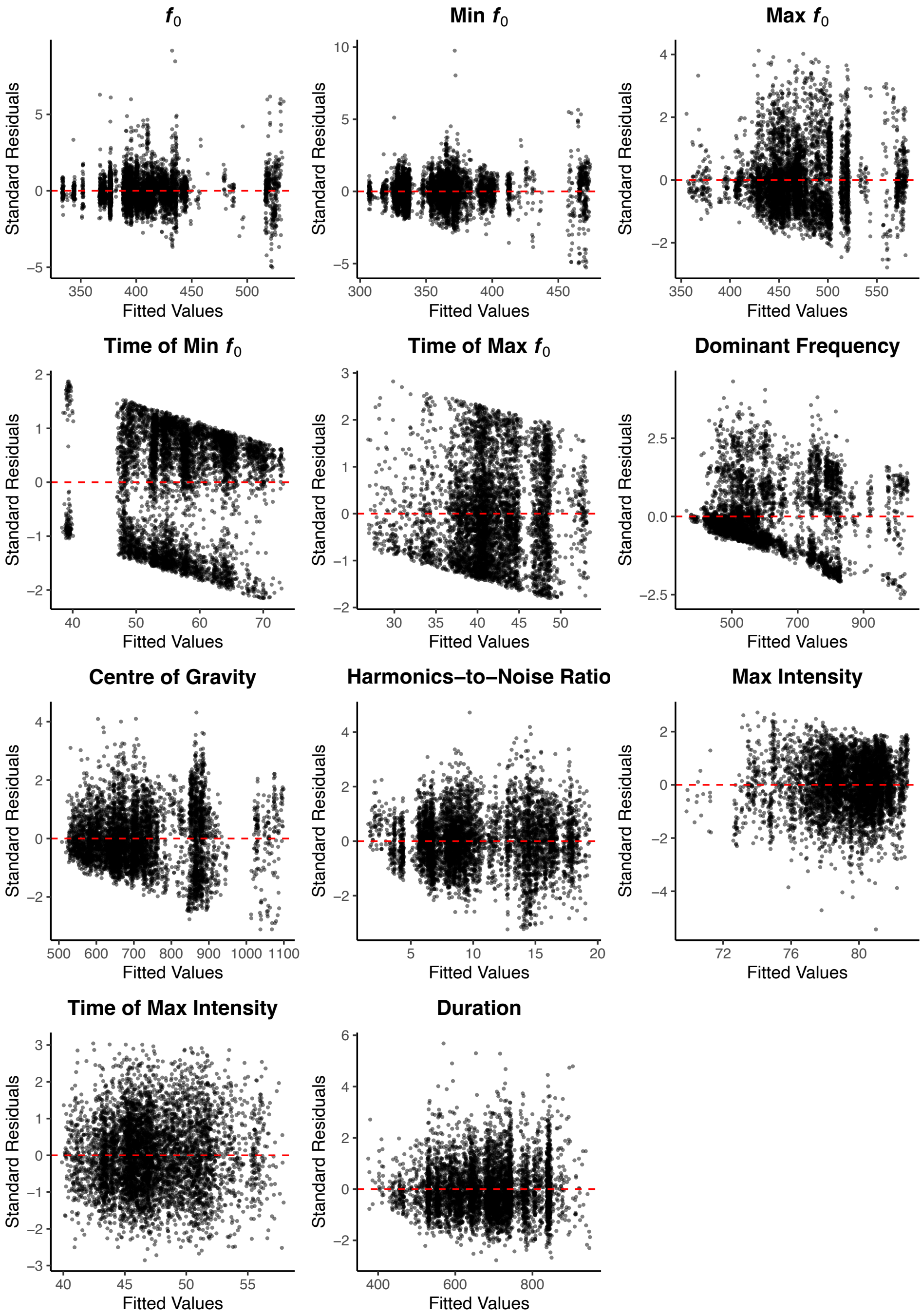


Figure S5.5. Residual plots used to visually assess if the residuals of the Bayesian regression models show equal variance.

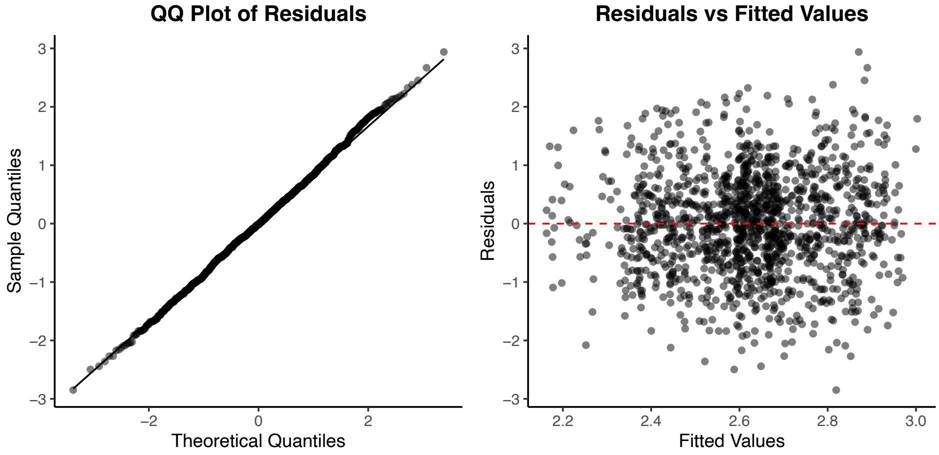
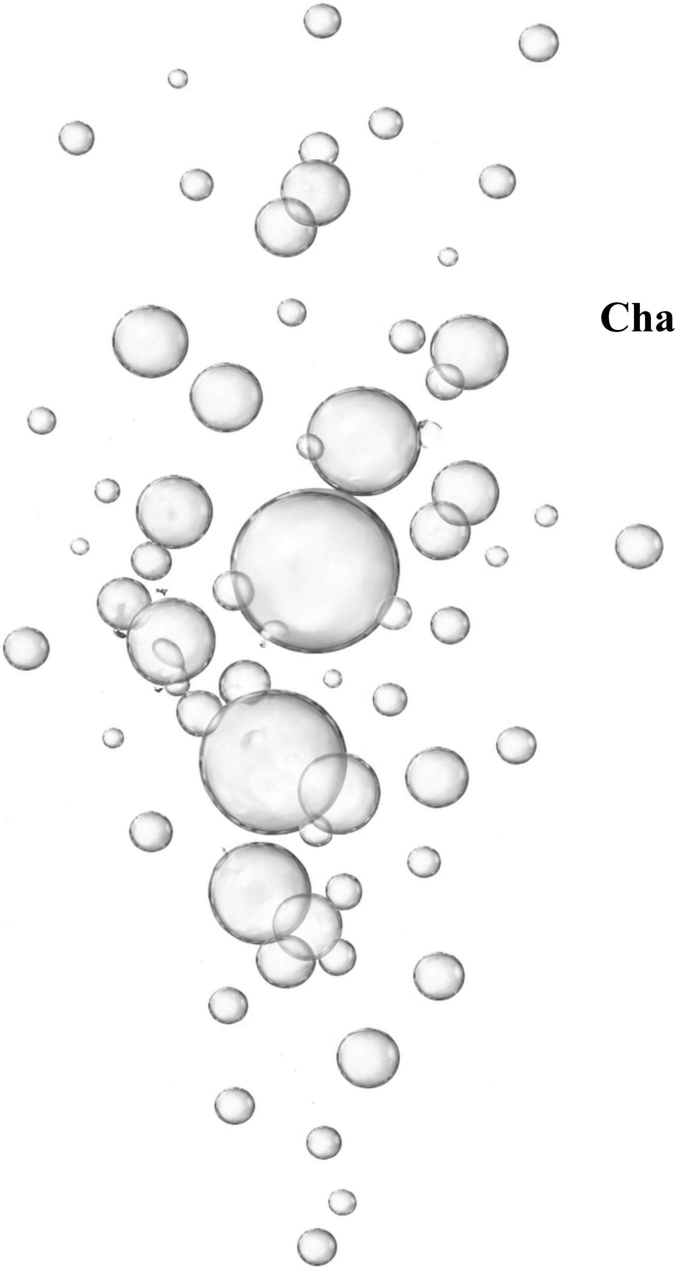


Figure S5.6. Visualisations of the normality (QQ plot on the left) and equal variance (residual plot on the right) assumptions of the log-transformed acoustic distance model.



Chapter 6

6 | General discussion

This thesis set out to address the gap in our understanding of harbour seal vocal communication by applying a recent modular vocal learning framework (Wirthlin et al., 2019) to explore key aspects of vocal variability, flexibility, control, and coordination during social interaction. Notably, it provides a comprehensive reference for key vocal learning components in harbour seals using a combination of non-invasive methodologies, including post-mortem anatomical studies, along with acoustic recordings and behavioural playback experiments. This thesis serves as a foundation for meaningful comparisons between the vocal capabilities of harbour seals and other species, such as humans, supporting future comparative work on the evolution of vocal communication systems, including speech.

This final chapter summarises the main findings and discusses their contributions in light of the previous literature, highlighting their broader implications for language evolution. It also discusses methodological insights and suggests several possible avenues for future work.

Summary of main findings

Chapter 2 demonstrated how the comparative method can be applied to gain insights into communicative behaviours by analysing findings from several animal groups. This chapter focusses on the role of signal timing in structuring social interactions and facilitating effective communication across species. Notably, interactive timing can range from turn-taking in human conversation to impressive collective displays, like cricket choruses (Greenfield, 1994b) and the bioluminescent flashing of fireflies (Buck, 1988). However, timing can be difficult to study in large groups, which is why this study focussed on the smallest possible group size, namely pairs or dyads. This chapter reviewed the scientific literature on signal timing in dyadic interactions across various animal groups, including mammals, birds, anurans, and insects, while considering signals across multiple communicative modalities and behavioural contexts.

This review provides a comprehensive table of definitions of various terms previously used to refer to rhythm or timing in dyadic interactions. Specifically, it distinguished between rhythmic interactions, during which mutual timing adjustments take place between individuals (e.g., Goebl & Palmer, 2009), and non-rhythmic interactions, where there is no timing adjustment between individuals but where the rhythmic patterns of individual signals may still influence social behaviour (e.g., Mathevon et al., 2017). Literature findings were analysed through the lens of Tinbergen's four questions, considering function, phylogeny, ontogeny, and mechanism (Tinbergen, 1963). Dyadic interactions occurred mainly in social contexts, such as cooperation, teaching, and parent–offspring recognition, and in sexual contexts, such as courtship and competition. Findings indicate that for species raised in a social environment, signal timing is a skill learned and fine-tuned over the course of development. Although most timing mechanisms found in the literature are based on the concept of coupled oscillators (Wilson & Wilson, 2005), some species, such as birds and pinnipeds, exhibit a capacity for predictive timing, whereas others, such as anurans and insects, display reactive signalling. This distinction highlights the need for further research on the diverse timing mechanisms used across species. A major contribution of this review constitutes the identification of literature gaps and suggestions for future research, which could help develop an integrative cross-species framework. Notably, it advocates for the inclusion of more species, as well as more studies on tactile and multimodal signals, and on the neural mechanisms underlying signal timing.

Chapter 3 investigated whether harbour seal pups exhibit turn-taking behaviour during vocal interactions with conspecifics. According to the "interaction engine" hypothesis, turn-taking is a fundamental social skill that facilitates coordination between individuals during communication, as seen in human conversation (Levinson, 2006). Similarly, in non-human animals, turn-taking serves to facilitate cooperation, form and maintain social bonds, and strengthen group cohesion (Badihi et al., 2024; Demartsev et al., 2018; Pika et al., 2018). Turn-taking is defined as a rapid exchange of communicative turns, characterised by minimal silent gaps and minimal overlap between turns (Sacks et al., 1974). Notably, turns in human conversations are about 200 ms on average, which is about three times faster than the response latencies observed in word production tasks (Indefrey & Levelt, 2004), suggesting that individuals anticipate when to begin their turn. Moreover, previous work in humans observed that response latencies during telephone conversations showed a negative lag-1 autocorrelation (Pouw & Holler, 2022), highlighting active timing adjustment on a turn-by-turn basis. In harbour seals, previous studies found that pups adopt an antisynchronous timing strategy during vocal interactions to avoid acoustic overlap (Anichini et al., 2023; Ravignani, 2019). However, these studies did not apply the standard methodology based on response latencies often used in human turn-taking research, which limited their ability to infer the extent of overlap and the length of the silent gaps. This chapter thus aimed to address this gap using response latency analyses to test whether harbour seal vocal interactions exhibit the core features of turn-taking behaviour. Specifically, it tested for the presence minimal gaps between turns and avoidance of overlap, reflected by short positive response latencies, and mutual call timing adjustment, reflected by a negative lag-1 autocorrelation of the response latencies. Finally, this study employed a playback experiment to test whether social dynamics, such as changes in tempo, rhythm, call duration, and sex of the caller, affected response timing during communication.

The results of this study showed that the silent gaps between vocalisations of different individuals were on average 287 ms, indicating minimal gaps between turns and no overlap. Results from the playback experiment showed that response latencies were markedly shorter and even overlapped with the playback (-54.3 ms), potentially reflecting differences in call timing predictability or attention depending on whether pups interacted with a real (i.e., another seal) or fake (i.e., loudspeaker) partner. Additionally, the playback experiment showed that harbour seal pups responded more quickly at faster tempi and for longer call durations, suggesting that pups are sensitive to the temporal characteristics of the acoustic sequence and actively adjust their call timing based on the dynamics of the interaction. This study found a significant negative autocorrelation at lag-1, supporting the idea of mutual timing adjustment in turn-taking behaviour, where feedback from the previous turn influences the timing of the next. However, this autocorrelation was observed for both types of interactions, irrespective of whether a pup was interacting with a real or fake partner, indicating that similar turn-taking patterns can be observed with only a single individual adjusting its call timing.

Chapter 4 tested whether the size of structures making up the vocal tract in harbour seals scales proportionally with body size. Body size is known to anatomically constrain the size and shape of the vocal tract, consequently influencing the range of vocalisations an animal can produce (Reby & McComb, 2003). One well-known phenomenon is the inverse relationship between body size and fundamental frequency (f_0), where larger animals (e.g., lions) produce vocalisations with a lower f_0 than smaller animals (e.g., mice; Bowling et al., 2017). When acoustic features of vocalisations accurately reflect body size, we refer to this relationship as “acoustic allometry” (Fitch, 2000; Taylor & Reby, 2010). However, some animals can deviate from this relationship, producing vocalisations that make them sound unexpectedly small or large given their actual body size. Previous studies proposed a “morphology vs. vocal learning hypothesis”, whereby animals deviate from acoustic allometry either through anatomical adaptations to the vocal tract or advanced vocal learning capacities (Garcia & Ravignani, 2020; Ravignani & Garcia, 2022). According to this hypothesis, vocal learners should be able to produce vocalisations that do not reflect their body size while possessing a vocal tract that scales in size with that of their body. This study tested whether harbour seals, which are considered vocal learners (Janik & Slater, 1997; Vernes et al., 2021) due to their remarkable vocal flexibility (Ralls et al., 1985; Torres Borda et al., 2021), exhibit vocal tract allometry.

The study begins by reanalysing data from Torres Borda and colleagues (2021) to demonstrate that harbour seals can deviate from acoustic allometry, showing that harbour seals of differing body sizes can produce similar f_0 values under different environmental conditions. Next, during post-mortem examinations of seals, body size information—such as length, mass, and axillary girth—was recorded and the vocal apparatus was removed to look for specific anatomical adaptations. No specialised vocal tract structures were found that could explain why seals sometimes deviate from acoustic allometry. Several measurements were then taken from the vocal tract, including vocal tract length, vocal fold length, vocal fold thickness, and tracheal dimensions. Using generalised linear models, the study found that body length and body mass strongly predicted vocal tract measurements, suggesting that vocal tract size scales proportionally with body size. Moreover, the study found differences in vocal tract measurements related to age and sex, indicating that their ability to violate acoustic allometry may change during development. For example, acoustic allometry for f_0—the acoustic proxy of vocal tract length—may only emerge after puppyhood, suggesting that pups potentially show a higher degree of vocal flexibility than older animals. Taken together, these findings show that harbour seals can deviate from the well-established acoustic allometry relationship despite the size of their vocal tracts scaling with their body size. Since no anatomical adaptations were found in the vocal tract, the study infers that harbour seals' vocal learning capacities may enable them to escape their anatomical constraints. Specifically, the study proposes that they possess a refined capacity for vocal control, likely due to strong neural connections between the vocal tract and the brain, allowing them to actively modify the shape of their vocal tract and modulate their vocal output.

Chapter 5 investigated whether harbour seal pups adjust the acoustic properties of their vocal output based on social interactions with conspecifics and provided an overview of their vocal development. Humans often accommodate their speech styles during conversation and converge towards each other to express social closeness or diverge from each to signal distance (Giles & Baker, 2008), demonstrating how social interactions influence communication. Similarly, vocal convergence in non-human animals has been reported as a means to strengthen social bonds (Baciadonna et al., 2022; Fischer et al., 2020; Prat et al., 2017). Although previous work showed that harbour seals can accommodate their vocalisations based on environmental conditions (Torres Borda et al., 2021), vocal accommodation as a result of social interaction has not been reported in this species. To date, two studies characterised vocal development in harbour seal pups, showing how individual differences in the acoustic structure of their vocalisations are driven by differences in age, sex, and body size (Khan et al., 2006; Sauvé et al., 2015). Notably, they showed that despite the effects of maturation on vocal anatomy during development, certain acoustic features, such as f_0 and the dominant frequency, remain highly individualised, allowing for individual recognition. The goal of this study was twofold. First, it aimed to provide a clearer overview of vocal development in harbour seal pups, building on the aforementioned studies by including a larger number of harbour seal pups and using daily recordings. Second, it tested whether harbour seal pups converge towards the calls of their conspecifics during social interactions by disentangling the relative contributions of maturation on vocal development from potential learning effects.

This longitudinal study recorded daily vocalisations from 64 harbour seal pups daily over the course of several weeks. Using Bayesian linear models, this study tested how acoustic parameters of vocalisations changed throughout development, with parameter estimates from previous studies used to inform the models (Khan et al., 2006; Sauvé et al., 2015). As expected, the study reported that acoustic parameters changed with age, sex, and body size, emphasising the role of maturation in shaping vocal signals. Next, the study computed the acoustic distances between all possible pairs of seals to assess how similar they sounded to each other. Linear modelling was used to test the hypothesis that pairs of seals housed in the same location sounded more similar to each other than to pairs housed separately, all the while controlling for differences in age, sex, and body size. Additionally, a discriminant function analysis was performed to test if seals were individually recognisable at levels higher than chance. Study findings showed that pups housed together had significantly smaller acoustic distances between vocalisations than pups housed separately, indicating that harbour seal pups show vocal convergence in response to social interactions while remaining distinct enough to allow for individual recognition.

Taken together, this thesis provides several important contributions, namely:

- A detailed analysis of vocal learning capacities in harbour seals, contributing to the growing body of comparative work on vocal learning across species.

- Demonstrates that harbour seals are constrained by their vocal anatomy and suggest their vocal flexibility arises from extensive neural control over their vocal tract articulators.

- Shows that harbour seals coordinate social interactions with conspecifics through turn-taking, exemplified by minimal silent gaps between turns, overlap avoidance, and mutual timing adjustments during vocal exchanges.

- Presents the first evidence that social interactions lead to vocal convergence in harbour seal pups, showing how social influences on vocal development can be disentangled from the effects of maturation.

- Uses a combination of innovative methodologies, including anatomical studies, behavioural experiments, and diverse statistical approaches to study vocal learning in harbour seals.

Theoretical implications

This thesis makes a compelling case for harbour seals as valuable comparative models for understanding the evolution of complex communication systems, like language in humans. The findings presented here provide important insights into speech acquisition and development through the lens of vocal learning (Jarvis, 2019). Results demonstrate that harbour seals exhibit vocal flexibility, vocal control, vocal coordination, and vocal variability in response to social and environmental conditions, all of which are essential components of vocal learning (Wirthlin et al., 2019). These abilities suggest that, like humans, harbour seals—despite being less vocally active—possess the capacity to modify and adapt their vocalisations in socially relevant ways, emphasising how studying vocal communication in harbour seals can shed light on the cognitive and neural mechanisms underlying vocal learning, and could contribute to our understanding of the evolution of speech. Specifically, harbour seals represent a previously underused mammalian model for the comparative study of vocal learning—showing enhanced neural control over vocal production and the ability to learn and use vocalisations during social interactions. Including species like the harbour seal in cross-species comparisons improves our capacity to infer whether speech-related traits arose through homologous or analogous evolution.

The connection between vocal flexibility and neural control

Harbour seals show a remarkable degree of vocal flexibility, as previously exemplified by Hoover, who famously imitated human speech (Ralls et al., 1985) and their ability to lower their fundamental frequency (f_0) in response to environmental noise (Torres Borda et al., 2021). These initial findings implied that harbour seals may not be as constrained by their vocal anatomy as would typically be expected, given acoustic allometry (Bowling et al., 2017). The results presented in Chapter 4 confirm that harbour seal pups can deviate from the well-established inverse relationship between fundamental frequency and body size, showing they can produce vocalisations with f_0 values that do not reflect their actual body size (de Reus et al., 2022). Notably, under varying environmental conditions, pups were found to produce vocalisations that could make them sound 5 kg larger or smaller. One possible function of this vocal flexibility relates to the “size-exaggeration” hypothesis

(Charlton & Reby, 2016). By lowering the f_0 of their vocalisations, pups might create an impression of a larger body size, potentially deterring competition or aggression from larger conspecifics (Neumann, 1999). In contrast, another possible function is that increasing the f_0 may elicit quicker responses from the mother, as higher-pitched vocalisations are often associated with distress (Briefer, 2012). Since deviations from acoustic allometry may arise either through anatomical adaptations to the vocal tract or from vocal learning capacities (Garcia & Ravignani, 2020; Ravignani & Garcia, 2022), it is important to consider if harbour seal vocal tracts possess any specific adaptations that might explain their vocal flexibility. Specifically, this thesis examined whether harbour seals exhibit a descended larynx—an anatomical feature previously considered a potential pre-adaptation for speech because it increases the repertoire of possible speech sounds that may be produced (Lieberman et al., 1969).

The anatomical examinations of the vocal tract performed in Chapter 4 confirmed the absence of a descended larynx or any morphological specialisations that might explain the vocal flexibility of harbour seals (de Reus et al., 2022). Notably, these examinations support previous claims that the vocal anatomy of harbour seals is remarkably similar to that of humans (Ravignani et al., 2016). Like humans, harbour seals possess a simplified larynx, without additional structures such as vocal membranes and laryngeal air sacs (Adams et al., 2020). Interestingly, previous work in humans attributed our increased vocal flexibility to the evolutionary loss of these anatomical structures (Nishimura et al., 2022). Specifically, this simplification of laryngeal anatomy was proposed to minimise acoustic chaos and support precise control of frequencies necessary for phonation. However, this hypothesis alone does not fully explain the vocal flexibility observed in harbour seals. Although non-human primates possess vocal membranes and laryngeal air sacs—and lack a descended larynx—vocal tract modelling in long-tailed macaques (*Macaca fascicularis*) has shown that their vocal tracts can produce human speech (Fitch et al., 2016), suggesting that vocal learning capacities are key in explaining vocal flexibility.

Although several neurobiological adaptations support vocal flexibility in pinnipeds (Cook et al., 2025), fine-tuned neural control over the vocal apparatus is a necessary prerequisite for the production of complex vocalisations, like speech sounds (Jürgens, 2002; Nieder & Mooney, 2019; Simonyan & Horwitz, 2011). Specifically, increased laryngeal control enables flexible vocal production, as it allows individuals to dynamically adjust the shape of the vocal tract. For example, humans can adjust the stiffness and length of their vocal folds, as well as their glottal configuration (Zhang, 2016), thereby modulating the acoustic properties of their vocalisations (e.g., f_0) within anatomical limits. Since harbour seals lack specific anatomical adaptations in their vocal tracts (as shown in Chapter 4), the findings suggest that, like humans, extensive neural control over the vocal tract is likely responsible for the vocal flexibility observed in this species (Cook et al., 2025; de Reus et al., 2022). The "direct connections" hypothesis posits that neural control of the vocal tract emerges from direct connections between the neurons in the brain's motor cortex and the primary motor neurons controlling the laryngeal muscles (Simonyan & Jürgens, 2003). While these direct connections have been observed in vocal learning birds (Striedter, 1994; Wild, 1997), neurobiological studies are still needed to confirm their presence in harbour seals.

Nothing in communication makes sense except in the light of interaction

The "interaction engine" hypothesis suggests that human social interactions are supported by a specialised set of communicative and cognitive skills (Levinson, 2006). Interestingly, the findings of vocal convergence (Chapter 5) and coordination (Chapters 2 and 3) show that harbour seals possess similar predispositions to facilitate social interactions. Along with previous findings in other species, they suggest that these communicative behaviours have deeper evolutionary roots and that the "interaction engine" hypothesis may be applicable to a much broader range of species.

This thesis presents the first evidence that harbour seal pups exhibit vocal convergence during social interactions, indicating that communication is shaped by interaction. The results of Chapter 5 demonstrate that harbour seal pups adjust their vocalisations to sound more similar to those of the conspecifics they frequently interact with. Specifically, the acoustic distance between pairs of seals housed together is significantly shorter than that between those housed separately. Previous studies reported that harbour seal pup vocalisations were highly individualised and their structure was affected by biological factors such as age, sex, and body size (Khan et al., 2006; Sauvé et al., 2015). Hence, the models used to determine the effect of social interaction on their vocalisations controlled for these biological differences. This thesis expands on this earlier developmental work, showing that vocal development is not only subject to biological factors, but also social factors. Notably, it shows that harbour seal pup vocalisations remain uniquely identifiable at levels higher than chance, despite showing vocal convergence towards conspecifics. While these vocalisations have only been described in the context of mother–offspring recognition (Perry & Renouf, 1988; Renouf, 1984), this finding raises the possibility that they may serve a different function beyond individual recognition, such as social bonding. This idea aligns with similar observations in humans and other animals, where communication is not only a means of conveying information, but also a tool for fostering social bonds (Baciadonna et al., 2022; Fischer et al., 2020). For example, during conversation, humans regularly accommodate their speech styles (e.g., accent, intonation, f_0) and converge to that of their partner to signal social closeness (Giles & Baker, 2008). Moreover, despite our capacity for vocal accommodation, every human voice contains individualised markers that are preserved across sound types (Pisanski et al., 2020). Such individual variation may result from subtle anatomical differences between individuals (e.g., Rua Ventura et al., 2013) or from genetically predisposed motor biases influencing how individuals use their voices (Toji et al., 2024), and which persist regardless of environment or experience. In sum, although biology partly dictates vocal output, evidence of vocal convergence in harbour seal pups reinforces the importance of social interaction in shaping communication across species.

This thesis also confirms that harbour seal pups engage in turn-taking behaviour during vocal interactions. Results from Chapter 3 demonstrate that turn transitions during vocal interactions are characterised by short silent gaps and minimal overlap between turns. This finding suggests that pups possess an affinity for interactive timing and aligns with results from the literature review presented in Chapter 2 (de Reus et al., 2022). Moreover, seals actively adjust the timing of their calls during these vocal exchanges on a turn-by-turn basis, as evidenced by the negative lag-1 autocorrelation of Chapter 3. The finding that harbour seal pups avoid acoustic overlap with conspecifics supports previous work, which suggested an antisynchronous timing strategy to facilitate information sharing and aid individual recognition in noisy colony environments (Anichini et al., 2023; Ravignani, 2019). While these earlier studies focused on call timing in the context of mother–offspring recognition, the interactions in Chapter 3 were recorded in the absence of mothers, indicating that turn-taking behaviour in this species might also serve social functions. The results of Chapter 3 resemble human turn-taking, where minimal silent gaps and minimal overlap between turns facilitate orderly communication (Sacks et al., 1974; Stivers et al., 2009). Humans anticipate when to speak and when to remain silent, showing both predictive timing and mutual timing adjustments (Pouw & Holler, 2022). Future studies should help determine if harbour seals similarly use predictive mechanisms when coordinating vocal exchanges. Moreover, in humans and other animals, turn-taking plays an important role in maintaining and developing social bonds (e.g., Badihi et al., 2024; Demartsev et al., 2018; Pika et al., 2018). In harbour seals, the role of social interaction in vocal coordination is supported by results from a previous playback experiment, showing that pups initially predicted the timing of playback calls, but that this anticipation dissipated over time, likely due to habituation to the "non-social" stimuli (Anichini et al., 2023). Additionally, the results of the playback experiment in Chapter 3 showed that harbour seals adjust the timing of their calls in relation to tempo and call duration, highlighting their sensitivity to the interaction dynamics. Finally, it should be noted that turn-by-turn adjustment can also be observed in the absence of social partners, which raises the question of whether mutual timing adjustment is a necessary condition for turn-taking. Regardless, evidence of vocal coordination in harbour seals further supports the idea that communication is contingent on interaction (whether social or non-social).

Methodological contributions and considerations

This interdisciplinary thesis integrates a wide range of concepts, theoretical frameworks, and methodologies to study vocal communication in harbour seals and situate the findings within the context of language evolution. While a comparative approach enriches our understanding by drawing from a variety of scientific disciplines, it is important to recognise the contributions and limitations inherent to this type of work.

On frameworks of vocal learning

The concept of vocal learning has evolved significantly over time, reflecting our growing understanding of this complex trait. The first framework by Janik and Slater (1997) classified vocal learning into two distinct forms: vocal production learning and vocal contextual learning. Over time, the emphasis in the literature shifted almost entirely to vocal production learning because the ability to modify and learn vocalisations based on auditory experience is central to human speech acquisition (Jarvis, 2019). Notably, this focus led to a reconceptualization of vocal production learning as a trait lying on a continuum (Petkov & Jarvis, 2012) and consisting of multiple dimensions (Vernes et al., 2021). These newer frameworks recognise the complexity of vocal production learning and extend its study to a broader range of species. However, despite their strengths, the ongoing reframing of how vocal learning is defined and structured can make it difficult to systematically compare findings across studies and species. For instance, Vernes and colleagues (2021) describe vocal production learning in terms of how vocal output matches an auditory template across three dimensions: model, degree, and timing. In contrast, Wirthlin and colleagues (2019) propose three modules—vocal versatility, vocal production variability, and vocal coordination—which do not map easily onto the former framework. For example, the module *vocal versatility* seems to correspond to the *model* and *degree* dimensions, while *vocal production variability* overlaps more with *timing*, though this mapping is not explicitly defined. These conceptual differences make it challenging to align findings across frameworks and hinder systematic, cross-species comparisons. Given that animal studies often involve long preparation times, newer frameworks may emerge by the time results are published, making it difficult to interpret findings in light of the most current conceptual models. These conceptual adjustments have also been accompanied by changes in how terminology is used in the field. In particular, the term "vocal learning" is now often used to refer exclusively to vocal production learning (e.g., Boughman, 1998; Connor et al., 1998; Jarvis, 2019), leading researchers to overlook how individuals learn to use and comprehend vocalisations within social contexts.

This thesis demonstrates that vocal behaviours in harbour seals are heavily influenced by social interactions (Chapters 3 and 5). Therefore, one of the main contributions of this thesis is addressing the gap in research on vocal contextual learning. Its findings emphasise the need to better acknowledge the contribution of contextual factors (e.g., type of partner, interaction dynamics) in shaping vocal communication. Consequently, to advance our evolutionary understanding of vocal learning and its role in the emergence of advanced vocal communications systems—like speech—future definitions and frameworks must embrace a more inclusive approach that considers both production and context (e.g., Wirthlin et al., 2019).

The comparative conundrum: Advantages and challenges of cross-species comparisons

The comparative approach is a powerful tool for understanding the evolution of complex traits across species, such as communication. By considering the evolutionary relationships between species, we can identify which traits arose through shared ancestry and which evolved independently in different animal groups. Respectively, these relationships allow us to make inferences about the characteristics of extinct common ancestors and the socioecological contexts that drove their evolution. This thesis contributes to our understanding of vocal learning by studying this trait in harbour seals, reinforcing their value as comparative models for language evolution, particularly speech. Its findings demonstrate that social interactions are fundamental in shaping vocal communication (Chapters 3 and 5) and suggest that, like humans, vocal flexibility emerges from extensive neural control (Chapter 4). These results demonstrate how the comparative approach can highlight important aspects of communication and open up exciting future research avenues (Chapter 2).

Despite its strengths, the comparative approach is not without its challenges. Cross-species comparisons are complicated by differences in research designs, environments, and experimental contexts. As noted by Henry and colleagues (2021), to meaningfully compare traits between species, experimental designs must account for species-specific constraints—including physical, perceptual, and motivational factors—in order to be ecologically valid. For example, in this thesis, the playback experiment from Chapter 3 used calls from harbour seal pups that were played at tempi falling within their natural range (Ravignani, 2019). Such careful consideration of the species' natural vocal properties ensures that experiments reflect vocal behaviours that are as naturalistic as possible. To confirm the ecological validity of these findings, they should be corroborated by behavioural observations in the animals' natural environment. A comparative approach thus necessitates a close interplay between field observations and experimental work to understand species-specific constraints, design ecologically valid experiments, and validate experimental findings. Notably, if experiments are not carefully designed and findings are not validated by behavioural observations, conclusions may fall prey to anthropocentric bias—the tendency to interpret non-human animal behaviour through a human-centric lens, distorting our understanding of their actual behaviour. Finally, another challenge in comparative work is to not overlook species that do not share similar life histories with humans, such as those differing significantly in anatomy, behaviour, or cognitive capacity (Kokko, 2017). Such taxonomic chauvinism restricts the scope of research and limits the range of species studied, narrowing our general understanding of how communication systems could have evolved across the animal kingdom. By including a wider variety of species—especially those with different life histories like harbour seals—we move away from the traditional primate-centric approach and gain more insights into our own behaviours and their evolution

In conclusion, while the comparative approach is an invaluable tool for understanding behaviour, such as communication, it is essential to be mindful of its diverse challenges. By refining experimental designs to accommodate species-specific constraints, acknowledging anthropocentric biases, and broadening the scope of research to include less studied species, we maximise the value of the comparative approach and deepen our understanding of how communication evolved across species.

Working with harbour seals undergoing rehabilitation

Some may question why the behavioural work presented in this thesis (Chapters 3 and 5) was conducted on harbour seals undergoing rehabilitation and whether the findings on the role of social interactions in shaping vocal communication can be generalised to natural populations.

The harbour seals part of this study were young pups, admitted into rehabilitation shortly after being born in the wild. Upon arrival, pups were paired and housed in an intensive care unit, with several units constituting one housing location—a controlled environment that provided two unique research advantages. First and foremost, it allowed researchers to track all the social interactions pups experienced during their stay. Specifically, pups housed together could see, hear, and physically interact with each other, and since the units were not acoustically isolated, they could also hear pups calling from other units within the same location, albeit with some signal attenuation. Knowing when and where animals were housed meant researchers knew exactly with whom an individual interacted and for how long, enabling them to control for these interactions in behavioural analyses. In contrast, in the wild, young seals move around the colony during the lactation period, making it impossible for researchers to track every interaction they have with other pups, juveniles, and adults without continuous 24h observations. Second, captive conditions allowed for recording high-quality vocalisations, since animals were housed indoors and could be approached closely. In the wild, the quality of audio recordings is highly influenced by weather conditions like wind and rain, causing greater interference from environmental noise (Wiley & Richards, 1978). Additionally, recordings would need to be made from a much greater distance to avoid disturbing the natural interactions within the colony, meaning vocal signals would be substantially attenuated by the time they are recorded, particularly the higher frequencies (Bradbury & Vehrencamp, 2011). In short, working with harbour seals in a rehabilitation centre enabled highly controlled experiments, but at the cost of (some) ecological validity.

Carrying out behavioural experiments in wild populations also raises several ethical concerns related to animal welfare and working in protected natural areas. These considerations would not only affect the experimental design but also would require proper assessment by an ethical committee before research permits could be issued. Although the rehabilitation centre did not require a research permit thanks to the non-invasive nature of the experiments, animal welfare was ensured through strict care protocols and close coordination with the on-site caretakers and veterinarians. Since animals are admitted into rehabilitation because they are often sick or injured, their health is always the foremost priority when designing experimental protocols and collecting experimental data. For example, researchers were only allowed to enter the units when the animals were being tended to, minimising unnecessary stress from human interaction. Researchers were also required to wear personal protective equipment, such as hospital scrubs, sanitary gloves, masks, and hairnets when present in the facilities to minimise any biosecurity risks. Furthermore, experiments with complicated setups—such as the playback experiment—had to be assembled and disassembled each time to ensure caretakers could continue to easily access the animals. To maintain high biosecurity standards, all equipment was disinfected after every session to prevent any risk of disease transmission between animals or facilities. However, despite these precautions, experimental sessions could be cancelled by the veterinarians due to concerns about the animals' health, leading to the loss of potential data. I feel it is important to clarify that, while this research capitalised on the opportunity to conduct non-invasive behavioural experiments during the seals' brief time in rehabilitation, all healthy animals were released back into the wild.

In summary, although working with in harbour seals in a captive setting presents several limitations, the controlled environment provided valuable behavioural insights that would have been challenging to obtain in the wild. Importantly, this research employed ecologically valid paradigms, and the seals were neither habituated to humans nor trained to perform specific behaviours, as is often the case in long-term captivity. These semi-captive conditions more closely approximated natural contexts, supporting the likelihood that findings would generalise to wild populations. Nonetheless, field observations would admittedly be a critical next step in validating the conclusions of this work.

Leveraging the power of playback experiments

Playback experiments—which involve broadcasting signals and observing responses—have long been an invaluable tool in the study of animal communication (e.g., Narins, 1982). They have been successfully used to investigate a wide range of behaviours, including territoriality, mate choice, predator-prey interactions, and mother–offspring recognition (Deecke, 2006; Fischer et al., 2013). The non-invasive nature of playback experiments makes them particularly suitable for studying behaviour in wild animals without significantly disrupting their natural activities. Additionally, their flexibility allows them to be adapted to various communicative modalities (e.g., acoustic, visual, chemical, electrical), making them applicable across a broad range of species. Finally, recent technological advances enabled the creation of automated playback systems that can respond in real-time, making the playback design interactive and helping to simulate natural communication dynamics (King, 2015).

Playback experiments can be leveraged as a powerful tool for testing a variety of hypotheses in a controlled manner, including individual and kin recognition, the function of signals, and cognitive mechanisms involved in communication. In Chapter 3 of this thesis, an acoustic playback experiment was used to simulate vocal interactions in harbour seal pups. While it cannot fully replicate the complexity and spontaneity of natural social interactions, the experiment used naturalistic stimuli—recorded calls from other pups—and systematically manipulated variables such as tempo, rhythm, call duration, and sex of the caller to test how seals adjusted their vocal responses. The results offered valuable insights into the timing mechanisms they use to coordinate social interactions and their sensitivity to the interaction dynamics. One way to make the current playback experiment more interactive could be to design a “call-triggered playback system” (Narins, 1982), where a seal vocalisation could trigger a random calling bout of 3–6 vocalisations from the loudspeaker. Consistent with previous observations (Anichini et al., 2023; Ravignani, 2019) and the findings of this thesis (Chapter 3), the system could be designed to delay playback vocalisations when the seal responds, thereby avoiding overlap and further increasing the ecological validity of the experiment. Moreover, recorded vocalisations of several different pups could be used for these triggered bouts to mimic interactions with different individuals in the colony. In Chapter 4, results showed that harbour seal pups can flexibly adjust the f_0 of their vocalisations to make themselves sound smaller or larger than they actually are (de Reus et al., 2022). Future playback experiments performed in wild colonies could manipulate the f_0 of calls (see Stansbury & Janik, 2021) to explore the function of such vocal modulation in harbour seals. In short, the flexible design of playback experiments makes them an extremely versatile method that can be tailored to address a wide range of diverse research questions in animal communication.

Future directions

This thesis’ findings of vocal flexibility (Chapter 4), convergence (Chapter 5) and coordination (Chapters 2 and 3) in harbour seals show that vocal learning can emerge in species with more solitary lifestyles, challenging the assumption that communicative complexity is primarily a trait of group-living animals (Freeberg et al., 2012). The results of Chapter 5 show that repeated social interactions lead to vocal convergence in harbour seals. It would be interesting to test whether the degree of vocal convergence is stronger for animals that share stronger social bonds. For instance, animals housed together that spend more time in close proximity and show more affiliative behaviours may exhibit a greater degree of vocal convergence than pairs with fewer affiliative interactions. While these findings would reinforce the idea that vocal learning plays an important role in facilitating social interactions, it is also possible that vocal learning originally evolved for a different purpose (Carouso-Peck et al., 2021; Nowicki & Searcy, 2014). Therefore, one interesting avenue for future research would be to explore why vocal learning evolved in harbour seals and other "less socially complex" species, and to consider alternative explanations for communicative complexity beyond sociality (Ord & Garcia-Porta, 2012).

Admittedly, by focusing exclusively on vocal communication in harbour seals and neglecting other non-vocal signals, this thesis is limited in scope, preventing a full appreciation of the richness of their communication. As discussed in Chapter 2, future research should explore non-vocal signals, like percussive and visual behaviours (e.g., Hocking et al., 2020; Kocsis et al., 2024), and examine how these integrate with vocal signals to facilitate social interactions. For example, the video recordings made in Chapter 3 to attribute vocalisations to the correct seal during the playback experiment could also be annotated to identify non-vocal behaviours (e.g., flipper slapping, sniffing, scratching) that play a role in communication. To isolate potential multimodal signals, one approach could be to examine non-vocal behaviours that co-occur closely in time with vocal productions. While the functional significance of these behaviours during social interactions would need to be tested, this line of inquiry could reveal whether harbour seals possess a well-developed, multimodal communication system and more complex social interactions than previously surmised. Moreover, it would be interesting to explore if these multimodal signals further facilitate response times during turn-taking in harbour seals, as seen in humans (e.g., ter Bekke et al., 2024). Ultimately, this research could provide a more comprehensive understanding of harbour seal communication and facilitate future cross-species comparisons.

Since the findings from Chapter 4 support the idea that the vocal flexibility of harbour seals likely stems from advanced neural control over their vocal tract, future work should confirm whether the “direct connections” hypothesis is also applicable to vocal learning seals (Simonyan & Jürgens, 2003). Specifically, neurobiological studies—involving magnetic resonance imaging and tractography—should aim to test for the presence of direct connections between the brain’s motor cortex and the laryngeal muscles. As previously done in birds (Striedter, 1994; Wild, 1997), similar studies in (presumably) non-vocal learning pinnipeds, such as otariids, should confirm the absence of direct connections to provide solid evidence linking vocal flexibility and neural control. Notably, specific neurobiological adaptations may subserve different aspects of vocal flexibility (Cook et al., 2025; Tyack, 2019), and the extent to which these circuits are integrated may more accurately reflect a species’ vocal learning ability. For instance, neurobiological studies in harbour seals and less vocally flexible pinnipeds could offer valuable insights into how vocal flexibility may arise from neural connectivity between the brain’s auditory and motor systems. In humans and zebra finches, strong auditory-motor coupling allows for real-time error correction of vocal productions following auditory feedback (Liu et al., 2018; Roberts et al., 2017). Based on evidence from the primate lineage, the “gradual audiomotor evolution” hypothesis suggests that the enhanced neural connectivity observed in humans may have evolved gradually (Merchant & Honing, 2014). Interestingly, while this hypothesis focussed on rhythmic entrainment, these same circuits have also been proposed to underpin vocal learning ability. Comparative studies across pinniped species with varying degrees of vocal flexibility could thus test both hypotheses simultaneously, providing critical insights into how neural control of the vocal tract and auditory-motor coupling interact to shape vocal flexibility.

The results from Chapter 4 suggest that pups may be less constrained by their vocal anatomy, indicating that the extent of vocal flexibility in harbour seals may be tied to a developmental window or sensitive period (Duengen & Ravignani, 2023). One promising line of future inquiry would be to integrate anatomical data with audio recordings taken prior to an animal’s death. In the event of death, all previous recordings from that animal would constitute valuable data for studying how its vocal tract anatomy influences the acoustic properties of its vocalisations. Importantly, it could allow researchers to gain insight into its vocal range. Moreover, collecting data from both pups and weaners would help determine if and how vocal flexibility changes with age. Unfortunately, audio recordings were taken from only three of the seals whose vocal tracts were measured in Chapter 4, with too few recorded vocalisations to yield meaningful statistical power. To address this, future work carried out in a captive setting should aim to obtain daily recordings from all animals admitted to the facility. Finally, these studies could be further complemented with vocal tract modelling work simulating the acoustic properties of vocalisations based on the size and shape of the vocal apparatus (e.g., Gamba et al., 2017). If the predictive accuracy of these computational models can be confirmed using real-world acoustic data, they could serve as a valuable tool for assessing changes in vocal flexibility throughout development.

Conclusion

Taken together, this thesis demonstrates that harbour seals can modify and adapt their vocalisations in socially relevant ways, highlighting their comparative value for language evolution research. Specifically, they exhibit several key vocal learning traits essential for human speech acquisition—including vocal flexibility, convergence, and coordination—which offer unique insights into the neural mechanisms of vocal learning and the role of social interactions in shaping communication. Notably, the amphibious lifestyle of harbour seals makes them far more accessible for study than whales and dolphins, and as mammals, they bridge the phylogenetic gap between humans and the well-studied birds. Simply put, by incorporating harbour seals into cross-species comparisons on language evolution, we can gain novel insights into vocal learning and its role in the emergence of complex vocal communication systems, like human speech.

References

Abney, D. H., Paxton, A., Dale, R., & Kello, C. T. (2014). Complexity matching in dyadic conversation. *Journal of Experimental Psychology: General*, *143*(6), 2304–2315. https://doi.org/10.1037/xge0000021

Abreu, F., & Pika, S. (2022). Turn-taking skills in mammals: A systematic review into development and acquisition. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.987253

Adams, A., Vogl, W., Dawson, C., Raverty, S., Haulena, M., & Skoretz, S. A. (2020). Laryngeal and soft palate valving in the harbour seal (*Phoca vitulina*). *Journal of Experimental Biology*, *223*(20), jeb230201. https://doi.org/10.1242/jeb.230201

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, *5*(7), 754. https://doi.org/10.4236/ojs.2015.57075

Anichini, M., de Heer Kloots, M., & Ravignani, A. (2020). Interactive rhythms in the wild, in the brain, and in silico. *Canadian Journal of Experimental Psychology*, *74*(3), 170–175. https://doi.org/10.1037/cep0000224

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

Arnon, I., Carmel, L., Claidière, N., Fitch, W. T., Kirby, S., Okanoya, K., Raviv, L., Wolters, L., & Fisher, S. E. (2025). What enables human language? A bio-cultural framework. *Science*. https://doi.org/10.1126/science.adq8303

Arriaga, G., & Jarvis, E. D. (2013). Mouse vocal communication system: Are ultrasounds learned or innate? *Brain and Language*, *124*(1), 96–116. https://doi.org/10.1016/j.bandl.2012.10.002

Aufdemorte, T. B., Sheridan, P. J., & Holt, G. R. (1983). Autoradiographic evidence of sex steroid receptors in laryngeal tissues of the baboon (*Papio cynocephalus*). *The Laryngoscope*, *93*(12), 1607–1611. https://doi.org/10.1288/00005537-198312000-00013

Baciadonna, L., Solvi, C., del Vecchio, F., Pilenga, C., Baracchi, D., Bandoli, F., Isaja, V., Gamba, M., & Favaro, L. (2022). Vocal accommodation in penguins (*Spheniscus demersus*) as a result of social environment. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1978), 20220626. https://doi.org/10.1098/rspb.2022.0626

Backwell, P., Jennions, M., Passmore, N., & Christy, J. (1998). Synchronized courtship in fiddler crabs. *Nature*, *391*(6662), 31–32. https://doi.org/10.1038/34076

Badihi, G., Graham, K. E., Grund, C., Safryghin, A., Soldati, A., Donnellan, E., Hashimoto, C., Mine, J. G., Piel, A. K., Stewart, F., Slocombe, K. E., Wilke, C., Townsend, S. W., Zuberbühler, K., Zulberti, C., & Hobaiter, C. (2024). Chimpanzee gestural exchanges share temporal structure with human language. *Current Biology*, *34*(14), R673–R674. https://doi.org/10.1016/j.cub.2024.06.009

Badraun, J. C., Mootnick, A. R., Deaner, R. O., Agoramoorthy, G., & McNeese, K. M. (1998). Hand modulation of vocalization in Siamangs *Hylobates syndactylus*. *International Zoo Yearbook*, *36*(1), 84–89. https://doi.org/10.1111/j.1748-1090.1998.tb02888.x

Bailey, W. J. (2003). Insect duets: Underlying mechanisms and their evolution. *Physiological Entomology*, *28*(3), 157–174. https://doi.org/10.1046/j.1365-3032.2003.00337.x

Bailey, W. J., & Hammond, T. J. (2003). Duetting in insects – does call length influence reply latency? *Journal of Zoology*, *260*(3), 267–274. https://doi.org/10.1017/S0952836903003728

Balcombe, J. P., & McCracken, G. F. (1992). Vocal recognition in Mexican free-tailed bats: Do pups recognize mothers? *Animal Behaviour*, *43*(1), 79–87. https://doi.org/10.1016/S0003- 3472(05)80073-9

Balsby, T. J. S., & Bradbury, J. W. (2009). Vocal matching by orange-fronted conures (*Aratinga canicularis*). *Behavioural Processes*, *82*(2), 133–139. https://doi.org/10.1016/j.beproc.2009.05.005

Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101–113. https://doi.org/10.1111/j.1749- 6632.1975.tb41575.x

Beckford, N. S., Schaid, D., Rood, S. R., & Schanbacher, B. (1985). Androgen stimulation and laryngeal development. *Annals of Otology, Rhinology & Laryngology*, *94*(6), 634–640. https://doi.org/10.1177/000348948509400622

Beebe, B., Alson, D., Jaffe, J., Feldstein, S., & Crown, C. (1988). Vocal congruence in mother-infant play. *Journal of Psycholinguistic Research*, *17*(3), 245–259. https://doi.org/10.1007/BF01686358

Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., & Tchernichovski, O. (2016). The forebrain song system mediates predictive call timing in female and male zebra finches. *Current Biology*, *26*(3), 309–318. https://doi.org/10.1016/j.cub.2015.12.037

Berta, A., Churchill, M., & Boessenecker, R. W. (2018). The origin and evolutionary biology of pinnipeds: Seals, sea lions, and walruses. *Annual Review of Earth and Planetary Sciences*, *46*(1), 203–228. https://doi.org/10.1146/annurev-earth-082517-010009

Bishop, L., & Goebl, W. (2015). When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae*, *19*(1), 84–110. https://doi.org/10.1177/1029864915570355

Bjørgesæter, A., Ugland, K. I., & Bjørge, A. (2004). Geographic variation and acoustic structure of the underwater vocalization of harbor seal (*Phoca vitulina*) in Norway, Sweden and Scotland. *The Journal of the Acoustical Society of America*, *116*(4), 2459–2468. https://doi.org/10.1121/1.1782933

Black, B., & Hazen, N. L. (1990). Social status and patterns of communication in acquainted and unacquainted preschool children. *Developmental Psychology*, *26*(3), 379–387. https://doi.org/10.1037/0012-1649.26.3.379

Black, B., & Logan, A. (1995). Links between communication patterns in mother-child, father-child, and child-peer interactions and children's social status. *Child Development*, *66*(1), 255–271. https://doi.org/10.2307/1131204

Boatright-Horowitz, S. L., Horowitz, S. S., & Simmons, A. M. (2000). Patterns of vocal interactions in a bullfrog (*Rana catesbeiana*) chorus: Preferential responding to far neighbors. *Ethology*, *106*(8), 701–712. https://doi.org/10.1046/j.1439-0310.2000.00580.x

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.28) [Computer software]. Phonetic Sciences, University of Amsterdam. http://www.praat.org/

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (Version 6.1.42) [Computer software]. Phonetic Sciences, University of Amsterdam. http://www.praat.org/

Bonadonna, G., Torti, V., De Gregorio, C., Valente, D., Randrianarison, R. M., Pozzi, L., Gamba, M., & Giacoma, C. (2019). Evidence of genetic monogamy in the lemur Indri (*Indri indri*). *American Journal of Primatology*, *81*(6), e22993. https://doi.org/10.1002/ajp.22993

Bonadonna, G., Torti, V., Sorrentino, V., Randrianarison, R. M., Zaccagno, M., Gamba, M., Tan, C. L., & Giacoma, C. (2017). Territory exclusivity and intergroup encounters in the indris (Mammalia: Primates: Indridae: *Indri indri*) upon methodological tuning. *The European Zoological Journal*. https://doi.org/10.1080/24750263.2017.1318184

Bosch, J., & Márquez, R. (2001). Call timing in male-male acoustical interactions and female choice in the midwife toad *Alytes obstetricans*. *Copeia*, *2001*(1), 169–177. https://doi.org/10.1643/0045- 8511(2001)001[0169:CTIMMA]2.0.CO;2

Boucaud, I. C. A., Perez, E. C., Ramos, L. S., Griffith, S. C., & Vignal, C. (2017). Acoustic communication in zebra finches signals when mates will take turns with parental duties. *Behavioral Ecology*, *28*(3), 645–656. https://doi.org/10.1093/beheco/arw189

Boughman, J. W. (1997). Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology*, *40*(1), 61–70. https://doi.org/10.1007/s002650050316

Boughman, J. W. (1998). Vocal learning by greater spear–nosed bats. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. https://doi.org/10.1098/rspb.1998.0286

Bowen, W. D., Oftedal, O. T., Boness, D. J., & Iverson, S. J. (1994). The effect of maternal age and other factors on birth mass in the harbour seal. *Canadian Journal of Zoology*, *72*(1), 8–14. https://doi.org/10.1139/z94-002

Bowling, D. L., Garcia, M., Dunn, J. C., Ruprecht, R., Stewart, A., Frommolt, K.-H., & Fitch, W. T. (2017). Body size and vocalization in primates and carnivores. *Scientific Reports*, *7*(1), 41070. https://doi.org/10.1038/srep41070

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

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x

Briefer, E. F., & McElligott, A. G. (2012). Social effects on vocal ontogeny in an ungulate, the goat, *Capra hircus*. *Animal Behaviour*, *83*(4), 991–1000. https://doi.org/10.1016/j.anbehav.2012.01.020

Brumm, H., & Slater, P. (2007). Animal communication: Timing counts. *Current Biology*, *17*(13), R521–R523. https://doi.org/10.1016/j.cub.2007.04.053

Bryden, M. M., & Felts, W. J. (1974). Quantitative anatomical observations on the skeletal and muscular systems of four species of Antarctic seals. *Journal of Anatomy*, *118*(Pt 3), 589–600.

Buck, J. (1988). Synchronous rhythmic flashing of fireflies II. *Quarterly Review of Biology*, *63*(3), 265–289. https://doi.org/10.1086/415929

Buck, J., & Buck, E. (1978). Toward a functional interpretation of synchronous flashing by fireflies. *The American Naturalist*, *112*(985), 471–492. https://doi.org/10.1086/283291

Buder, E. H., Warlaumont, A. S., Oller, D. K., & Chorna, L. B. (2010). Dynamic indicators of mother- infant prosodic and illocutionary coordination. *Speech Prosody*, 6–9. https://doi.org/10.21437/SpeechProsody.2010-214

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Buschmeier, H., & Włodarczak, M. (2013). TextGridTools: A TextGrid processing and analysis toolkit for Python. *Tagungsband Der 24*, 152–157.

Caldwell, J. P., & Shepard, D. B. (2007). Calling site fidelity and call structure of a neotropical toad, *Rhinella ocellata* (Anura: Bufonidae). *Journal of Herpetology*, *41*(4), 611–621. https://doi.org/10.1670/07-025.1

Caldwell, M. C., & Caldwell, D. K. (1968). Vocalization of naive captive dolphins in small groups. *Science*, *159*(3819), 1121–1123. https://doi.org/10.1126/science.159.3819.1121

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012). Convergence and divergence in Diana monkey vocalizations. *Biology Letters*, *8*(3), 382–385. https://doi.org/10.1098/rsbl.2011.1182

Carouso-Peck, S., Goldstein, M. H., & Fitch, W. T. (2021). The many functions of vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200235. https://doi.org/10.1098/rstb.2020.0235

Carouso-Peck, S., Menyhart, O., DeVoogd, T. J., & Goldstein, M. H. (2020). Contingent parental responses are naturally associated with zebra finch song learning. *Animal Behaviour*, *165*, 123–132. https://doi.org/10.1016/j.anbehav.2020.04.019

Carstairs-McCarthy, A. (1999). *The Origins of Complex Language: An Inquiry into the Evolutionary Beginnings of Sentences, Syllables, and Truth*. Oxford University Press. https://doi.org/10.1093/oso/9780198238225.001.0001

Carter, G. G., Fenton, M. B., & Faure, P. A. (2009). White-winged vampire bats (*Diaemus youngi*) exchange contact calls. *Canadian Journal of Zoology*, *87*(7), 604–608. https://doi.org/10.1139/Z09-051

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language*, *92*, 234–253. https://doi.org/10.1016/j.jml.2016.06.013

Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2019). Body sway reflects joint emotional expression in music ensemble performance. *Scientific Reports*, *9*(1), 205. https://doi.org/10.1038/s41598-018-36358-4

Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences*, *114*(21), E4134– E4141. https://doi.org/10.1073/pnas.1617657114

Charlton, B. D., Ellis, W. A. H., McKinnon, A. J., Cowin, G. J., Brumm, J., Nilsson, K., & Fitch, W. T. (2011). Cues to body size in the formant spacing of male koala (*Phascolarctos cinereus*) bellows: Honesty in an exaggerated trait. *Journal of Experimental Biology*, *214*(20), 3414–3422. https://doi.org/10.1242/jeb.061358

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications*, *7*(1), 12739. https://doi.org/10.1038/ncomms12739

Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2009). The information content of giant panda, *Ailuropoda melanoleuca*, bleats: Acoustic cues to sex, age and size. *Animal Behaviour*, *78*(4), 893–898. https://doi.org/10.1016/j.anbehav.2009.06.029

Charrier, I. (2020). Mother–offspring vocal recognition and social system in pinnipeds. In T. Aubin & N. Mathevon (Eds.), *Coding Strategies in Vertebrate Acoustic Communication* (pp. 231–246). Springer Cham. https://doi.org/10.1007/978-3-030-39200-0_9

Charrier, I. (2021). Vocal communication in otariids and odobenids. In C. Campagna & R. Harcourt (Eds.), *Ethology and Behavioral Ecology of Otariids and the Odobenid* (pp. 265–289). Springer International Publishing. https://doi.org/10.1007/978-3-030-59184-7_14

Charrier, I., Pitcher, B. J., & Harcourt, R. G. (2009). Vocal recognition of mothers by Australian sea lion pups: Individual signature and environmental constraints. *Animal Behaviour*, *78*(5), 1127– 1134. https://doi.org/10.1016/j.anbehav.2009.07.032

Chauvigné, L. A. S., Walton, A., Richardson, M. J., & Brown, S. (2019). Multi-person and multisensory synchronization during group dancing. *Human Movement Science*, *63*, 199–208. https://doi.org/10.1016/j.humov.2018.12.005

Cheney, D. L., Silk, J. B., & Seyfarth, R. M. (2016). Network connections, dyadic bonds and fitness in wild female baboons. *Royal Society Open Science*, *3*(7), 160255. https://doi.org/10.1098/rsos.160255

Chomsky, N. (2014). *The Minimalist Program* (20th ed.). The MIT Press. https://doi.org/10.7551/mitpress/9780262527347.001.0001

Christiansen, M. H., & Kirby, S. (Eds.). (2003). *Language Evolution*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199244843.001.0001

Cirelli, L. K., Jurewicz, Z. B., & Trehub, S. E. (2020). Effects of maternal singing style on mother– infant arousal and behavior. *Journal of Cognitive Neuroscience*, *32*(7), 1213–1220. https://doi.org/10.1162/jocn_a_01402

Cirelli, L. K., Trehub, S. E., & Trainor, L. J. (2018). Rhythm and melody as social signals for infants. *Annals of the New York Academy of Sciences*, *1423*(1), 66–72. https://doi.org/10.1111/nyas.13580

Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLOS One*, *1*(1), e73. https://doi.org/10.1371/journal.pone.0000073

Connor, R. C., Mann, J., Tyack, P. L., & Whitehead, H. (1998). Social evolution in toothed whales. *Trends in Ecology & Evolution*, *13*(6), 228–232. https://doi.org/10.1016/S0169- 5347(98)01326-3

Cook, P., Rouse, A., Sawyer, E., Miller, K., & Berns, G. (2025). *Neurobiological adaptations supporting vocal plasticity have accumulated in the marine carnivores*. Research Square. https://doi.org/10.21203/rs.3.rs-6222519/v1

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*(4), 412–427. https://doi.org/10.1037/a0032345

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine *what* to say but not *when* to say it. *Cognition*, *175*, 77–95. https://doi.org/10.1016/j.cognition.2018.01.015

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group- specific calls: A case for vocal learning? *Ethology*, *110*(3), 221–243. https://doi.org/10.1111/j.1439-0310.2004.00968.x

Crow, T. J. (2000). Schizophrenia as the price that *Homo sapiens* pays for language: A resolution of the central paradox in the origin of the species. *Brain Research Reviews*, *31*(2), 118–129. https://doi.org/10.1016/S0165-0173(99)00029-6

da Silva, A. R. (2021). *biotools: Tools for biometry and applied statistics in agricultural science* (Version 4.2) [Computer software]. https://cran.r-project.org/web/packages/biotools/index.html

D'Amelio, P. B., Trost, L., & ter Maat, A. (2017). Vocal exchanges during pair formation and maintenance in the zebra finch (*Taeniopygia guttata*). *Frontiers in Zoology*, *14*(1), 13. https://doi.org/10.1186/s12983-017-0197-x

de Fanis, E., & Jones, G. (1996). Allomaternal care and recognition between mothers and young in pipistrelle bats (*Pipistrellus pipistrellus*). *Journal of Zoology*, *240*(4), 781–787. https://doi.org/10.1111/j.1469-7998.1996.tb05324.x de Reus, K. (2017). *Talking seals: Vocal development in Eastern Atlantic harbour seal pups Phoca vitulina vitulina*. Royal Veterinary College, University of London.

de Reus, K., Carlson, D., Lowry, A., Gross, S., Garcia, M., Rubio-Garcia, A., Salazar-Casals, A., & Ravignani, A. (2022). Vocal tract allometry in a mammalian vocal learner. *Journal of Experimental Biology*, *225*(8), jeb243766. https://doi.org/10.1242/jeb.243766

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., Bruno, J. H., Trainor, L., & Ravignani, A. (2021). Rhythm in dyadic interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200337. https://doi.org/10.1098/rstb.2020.0337

de Ruiter, J.-P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Deecke, V. B. (2006). Studying marine mammal cognition in the wild: A review of four decades of playback experiments. *Aquatic Mammals*, *32*(4), 461–482. https://doi.org/10.1578/AM.32.4.2006.461

Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., & Manser, M. (2018). Vocal turn-taking in meerkat group calling sessions. *Current Biology*, *28*(22), 3661–3666. https://doi.org/10.1016/j.cub.2018.09.065

Demos, A. P., Carter, D. J., Wanderley, M. M., & Palmer, C. (2017). The unresponsive partner: Roles of social status, auditory feedback, and animacy in coordination of joint music performance. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00149

Desjonquères, C., Maliszewski, J., Lewandowski, E. N., Speck, B., & Rodríguez, R. L. (2019). Social ontogeny in the communication system of an insect. *Animal Behaviour*, *148*, 93–103. https://doi.org/10.1016/j.anbehav.2018.12.002

Dierauf, L. A., Dougherty, S. A., & Lowenstine, L. J. (1986). Survival versus nonsurvival determinants for neonatal harbor seals. *Journal of the American Veterinary Medical Association*, *189*(9), 1024–1028.

Dominguez, F., Hernandez-Ranz, F., Boixeda, D., & Valdazo, P. (1987). Acute upper-airway obstruction in achalasia of the esophagus. *American Journal of Gastroenterology*, *82*(4), 362.

Dudzinski, K. M., Thomas, J. A., & Gregg, J. D. (2009). Communication in marine mammals. In W. F. Perrin, B. Würsig, & J. G. M. Thewissen (Eds.), *Encyclopedia of Marine Mammals* (pp. 260– 269). Academic Press. https://doi.org/10.1016/B978-0-12-373553-9.00064-X

Duengen, D., Fitch, W. T., & Ravignani, A. (2023). Hoover the talking seal. *Current Biology*, *33*(2), R50–R52. https://doi.org/10.1016/j.cub.2022.12.023

Duengen, D., Jadoul, Y., & Ravignani, A. (2024). Vocal usage learning and vocal comprehension learning in harbor seals. *BMC Neuroscience*, *25*(1), 48. https://doi.org/10.1186/s12868-024- 00899-4

Duengen, D., & Ravignani, A. (2023). The paradox of learned song in a semi-solitary mammal. *Ethology*, *129*(9), 445–453. https://doi.org/10.1111/eth.13385

Duengen, D., & Ravignani, A. (2024). *Training experimentally naive seals for vocal learning experiments* (p. 2024.08.27.609954). bioRxiv. https://doi.org/10.1101/2024.08.27.609954

Dusenbery, D. B. (1992). *Sensory Ecology: How Organisms Acquire and Respond to Information*. New York: W.H. Freeman.

Elemans, C. P. H., Rasmussen, J. H., Herbst, C. T., Düring, D. N., Zollinger, S. A., Brumm, H., Srivastava, K., Svane, N., Ding, M., Larsen, O. N., Sober, S. J., & Švec, J. G. (2015). Universal mechanisms of sound production and control in birds and mammals. *Nature Communications*, *6*(1), 8978. https://doi.org/10.1038/ncomms9978

Esser, K.-H., & Schmidt, U. (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae)—Evidence for acoustic learning. *Ethology*, *82*(2), 156–168. https://doi.org/10.1111/j.1439-0310.1989.tb00496.x

Ey, E., & Fischer, J. (2009). The “acoustic adaptation hypothesis”—A review of the evidence from birds, anurans and mammals. *Bioacoustics*. https://doi.org/10.1080/09524622.2009.9753613

Fahlman, A., Loring, S. H., Johnson, S. P., Haulena, M., Trites, A. W., Fravel, V. A., & Van Bonn, W. G. (2014). Inflation and deflation pressure-volume loops in anesthetized pinnipeds confirms compliant chest and lungs. *Frontiers in Physiology*, *5*. https://doi.org/10.3389/fphys.2014.00433

Fant, G. (1971). *Acoustic Theory of Speech Production*. De Gruyter Mouton. https://doi.org/10.1515/9783110873429

Farabaugh, S. M. (1982). The ecological and social significance of duetting. In D. E. Kroodsma & E. H. Miller (Eds.), *Acoustic Communication in Birds* (Vol. 2, pp. 85–124). Academic Press.

Feldman, R. (2007). Parent–infant synchrony: Biological foundations and developmental outcomes. *Current Directions in Psychological Science*, *16*(6), 340–345. https://doi.org/10.1111/j.1467- 8721.2007.00532.x

Feldman, R. (2017). The neurobiology of human attachments. *Trends in Cognitive Sciences*, *21*(2), 80–99. https://doi.org/10.1016/j.tics.2016.11.007

Feldman, R., Greenbaum, C. W., & Yirmiya, N. (1999). Mother–infant affect synchrony as an antecedent of the emergence of self-control. *Developmental Psychology*, *35*(1), 223–231. https://doi.org/10.1037/0012-1649.35.1.223

Field, C. L., Gulland, F. M. D., Johnson, S. P., Simeone, C. A., & Whoriskey, S. T. (2018). Seal and sea lion medicine. In *CRC Handbook of Marine Mammal Medicine* (3rd ed.). CRC Press.

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., Hoeschele, M., Ocklenburg, S., de Boer, B., Sturdy, C. B., Newen, A., & Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *American Journal of Primatology*, *75*(7), 643–663. https://doi.org/10.1002/ajp.22153

Fischer, J., Wegdell, F., Trede, F., Dal Pesco, F., & Hammerschmidt, K. (2020). Vocal convergence in a multi-level primate society: Insights into the evolution of vocal learning. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1941), 20202531. https://doi.org/10.1098/rspb.2020.2531

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, 1213–1222. https://doi.org/10.1121/1.421048

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267. https://doi.org/10.1016/S1364-6613(00)01494-7

Fitch, W. T. (2005). The evolution of language: A comparative review. *Biology and Philosophy*, *20*(2), 193–203. https://doi.org/10.1007/s10539-005-5597-1

Fitch, W. T. (2006). Production of vocalizations in mammals. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 115–121). Elsevier. https://doi.org/10.1016/B0-08- 044854-2/00821-X

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511817779

Fitch, W. T. (2011). The evolution of syntax: An exaptationist perspective. *Frontiers in Evolutionary Neuroscience*, *3*. https://doi.org/10.3389/fnevo.2011.00009

Fitch, W. T. (2018). The biology and evolution of speech: A comparative analysis. *Annual Review of Linguistics*, *4*, 255–279. https://doi.org/10.1146/annurev-linguistics-011817-045748

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech- ready. *Science Advances*, *2*(12), 1–7. https://doi.org/10.1126/sciadv.1600723

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511– 1522. https://doi.org/10.1121/1.427148

Fitch, W. T., & Hauser, M. D. (1995). Vocal production in nonhuman primates: Acoustics, physiology, and functional constraints on "honest" advertisement. *American Journal of Primatology*, *37*(3), 191–219. https://doi.org/10.1002/ajp.1350370303

Fitch, W. T., & Hauser, M. D. (2003). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic Communication* (pp. 65–137). Springer. https://doi.org/10.1007/0-387-22762-8_3

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1477), 1669–1675. https://doi.org/10.1098/rspb.2001.1704

Fox, J., & Weisberg, S. (2018). *An R Companion to Applied Regression* (3rd ed.). SAGE Publications.

Freeberg, T. M., Dunbar, R. I. M., & Ord, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1785–1801. https://doi.org/10.1098/rstb.2011.0213

Frey, R., Reby, D., Fritsch, G., & Charlton, B. D. (2018). The remarkable vocal anatomy of the koala (*Phascolarctos cinereus*): Insights into low-frequency sound production in a marsupial species. *Journal of Anatomy*, *232*(4), 575–595. https://doi.org/10.1111/joa.12770

Frey, R., Volodin, I., Volodina, E., Carranza, J., & Torres-Porras, J. (2012). Vocal anatomy, tongue protrusion behaviour and the acoustics of rutting roars in free-ranging Iberian red deer stags (*Cervus elaphus hispanicus*). *Journal of Anatomy*, *220*(3), 271–292. https://doi.org/10.1111/j.1469-7580.2011.01467.x

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325– 1330. https://doi.org/10.1111/2041-210X.12584

Fröhlich, M., Kuchenbuch, P., Müller, G., Fruth, B., Furuichi, T., Wittig, R. M., & Pika, S. (2016). Unpeeling the layers of language: Bonobos and chimpanzees engage in cooperative turn-taking sequences. *Scientific Reports*, *6*(1), 25887. https://doi.org/10.1038/srep25887

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins: Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809–1829. https://doi.org/10.1111/brv.12535

Gamba, M., Favaro, L., Araldi, A., Matteucci, V., Giacoma, C., & Friard, O. (2017). Modeling individual vocal differences in group-living lemurs using vocal tract morphology. *Current Zoology*, *63*(4), 467–475. https://doi.org/10.1093/cz/zox023

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., Bonadonna, G., Friard, O., & Giacoma, C. (2016). The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00249

Garcia, M., Herbst, C. T., Bowling, D. L., Dunn, J. C., & Fitch, W. T. (2017). Acoustic allometry revisited: Morphological determinants of fundamental frequency in primate vocal production. *Scientific Reports*, *7*(1), 10450. https://doi.org/10.1038/s41598-017-11000-x

Garcia, M., & Manser, M. (2020). Bound for specific sounds: Vocal predisposition in animal communication. *Trends in Cognitive Sciences*, *24*(9), 690–693. https://doi.org/10.1016/j.tics.2020.05.013

Garcia, M., & Ravignani, A. (2020). Acoustic allometry and vocal learning in mammals. *Biology Letters*, *16*(7), 20200081. https://doi.org/10.1098/rsbl.2020.0081

Garcia, M., Wondrak, M., Huber, L., & Fitch, W. T. (2016). Honest signaling in domestic piglets (*Sus scrofa domesticus*): Vocal allometry and the information content of grunt calls. *Journal of Experimental Biology*, *219*(12), 1913–1921. https://doi.org/10.1242/jeb.138255

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In *The Origins of Music* (pp. 103–123). The MIT Press.

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

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gilboa, A., & Tal-Shmotkin, M. (2012). String quartets as self-managed teams: An interdisciplinary perspective. *Psychology of Music*, *40*(1), 19–41. https://doi.org/10.1177/0305735610377593

Giles, H., & Baker, S. C. (2008). Communication accommodation theory. In W. Donsbach (Ed.), *The International Encyclopedia of Communication*. Wiley-Blackwell. https://doi.org/10.1002/9781405186407.wbiecc067

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438. https://doi.org/10.1525/mp.2009.26.5.427

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419–429. https://doi.org/10.1016/S1364-6613(99)01397-2

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, *40*, e46. https://doi.org/10.1017/S0140525X15001247

Gould, S. J., & Vrba, E. S. (1982). Exaptation—A missing term in the science of form. *Paleobiology*, *8*(1), 4–15. https://doi.org/10.1017/S0094837300004310

Grafe, T. U. (1996). The function of call alternation in the African reed frog (Hyperolius marmoratus): Precise call timing prevents auditory masking. *Behavioral Ecology and Sociobiology*, *38*(3), 149–158. https://doi.org/10.1007/s002650050227

Grafe, T. U. (1999). A function of synchronous chorusing and a novel female preference shift in an anuran. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1435), 2331–2336. https://doi.org/10.1098/rspb.1999.0927

Grafe, T. U. (2005). Anuran choruses as communication networks. In P. K. McGregor (Ed.), *Animal Communication Networks* (pp. 277–299). Cambridge University Press. https://doi.org/10.1017/CBO9780511610363.017

Grafe, T. U., & Wanger, T. C. (2007). Multimodal signaling in male and female foot-flagging frogs *Staurois guttatus* (Ranidae): An alerting function of calling. *Ethology*, *113*(8), 772–781. https://doi.org/10.1111/j.1439-0310.2007.01378.x

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01167

Greenfield, M. D. (1994a). Cooperation and conflict in the evolution of signal interactions. *Annual Review of Ecology, Evolution, and Systematics*, *25*, 97–126. https://doi.org/10.1146/annurev.es.25.110194.000525

Greenfield, M. D. (1994b). Synchronous and alternating choruses in insects and anurans: Common mechanisms and diverse functions. *Integrative and Comparative Biology*, *34*(6), 605–615. https://doi.org/10.1093/icb/34.6.605

Greenfield, M. D. (2005). Mechanisms and evolution of communal sexual displays in arthropods and anurans. *Advances in the Study of Behavior*, *35*, 1–62. https://doi.org/10.1016/S0065- 3454(05)35001-7

Greenfield, M. D., & Minckley, R. L. (1993). Acoustic dueling in tarbush grasshoppers: Settlement of territorial contests via alternation of reliable signals. *Ethology*, *95*(4), 309–326. https://doi.org/10.1111/j.1439-0310.1993.tb00480.x

Greenfield, M. D., & Roizen, I. (1993). Katydid synchronous chorusing is an evolutionary stable outcome of female choice. *Nature*, *364*, 618–620.

Greenfield, M. D., Tourtellot, M. K., & Snedden, W. A. (1997). Precedence effects and the evolution of chorusing. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1386), 1355–1361. https://doi.org/10.1098/rspb.1997.0188

Greer, L. L., Whaley, J., & Rowles, T. K. (2001). Euthanasia. In L. Dierauf & F. M. D. Gulland (Eds.), *CRC Handbook of Marine Mammal Medicine* (2nd ed., pp. 729–737). CRC Press.

Griffith, S. C., & Buchanan, K. L. (2010). The zebra finch: The ultimate Australian supermodel. *Emu*. https://doi.org/10.1071/MUv110n3_ED

Grijseels, D. M., Fairbank, D. A., & Miller, C. T. (2024). A model of marmoset monkey vocal turn- taking. *Proceedings of the Royal Society B: Biological Sciences*, *291*(2026), 20240150. https://doi.org/10.1098/rspb.2024.0150

Hadley, L. V., Novembre, G., Keller, P. E., & Pickering, M. J. (2015). Causal role of motor simulation in turn-taking behavior. *Journal of Neuroscience*, *35*(50), 16516–16520. https://doi.org/10.1523/JNEUROSCI.1850-15.2015

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, *86*, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Halfwerk, W., Varkevisser, J., Simon, R., Mendoza, E., Scharff, C., & Riebel, K. (2019). Toward testing for multimodal perception of mating signals. *Frontiers in Ecology and Evolution*, *7*. https://doi.org/10.3389/fevo.2019.00124

Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. *Behavioral Ecology and Sociobiology*, *55*(5), 415–430. https://doi.org/10.1007/s00265-003-0741-x

Hall, M. L. (2009). A review of vocal duetting in birds. In *Advances in the Study of Behavior* (Vol. 40, pp. 67–121). Academic Press. https://doi.org/10.1016/S0065-3454(09)40003-2

Hall, M. L., & Magrath, R. D. (2007). Temporal coordination signals coalition quality. *Current Biology*, *17*(11), R406–R407. https://doi.org/10.1016/j.cub.2007.04.022

Hanggi, E. B., & Schusterman, R. J. (1994). Underwater acoustic displays and individual variation in male harbour seals, *Phoca vitulina*. *Animal Behaviour*, *48*(6), 1275–1283. https://doi.org/10.1006/anbe.1994.1363

Harcourt, R. (1992). Factors affecting early mortality in the South American fur seal (*Arctocephalus australis*) in Peru: Density-related effects and predation. *Journal of Zoology*, *226*(2), 259–270. https://doi.org/10.1111/j.1469-7998.1992.tb03838.x

Harding, E. E., Kim, J. C., Demos, A. P., Roman, I. R., Tichko, P., Palmer, C., & Large, E. W. (2025). Musical neurodynamics. *Nature Reviews Neuroscience*, *26*(5), 293–307. https://doi.org/10.1038/s41583-025-00915-4

Harding, K. C., Fujiwara, M., Axberg, Y., & Härkönen, T. (2005). Mass-dependent energetics and survival in harbour seal pups. *Functional Ecology*, *19*(1), 129–135. https://doi.org/10.1111/j.0269-8463.2005.00945.x

Harrist, A. W., & Waugh, R. M. (2002). Dyadic synchrony: Its structure and function in children's development. *Developmental Review*, *22*(4), 555–592. https://doi.org/10.1016/S0273-2297(02)00500-2

Hauksson, E. (2006). Growth and reproduction in the Icelandic common seal (*Phoca vitulina* L., 1758). *Marine Biology Research*, *2*(1), 59–73. https://doi.org/10.1080/17451000600650038 Hauser, M. D. (1996). *The evolution of communication* (pp. xii, 760). The MIT Press.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., Chomsky, N., & Lewontin, R. C. (2014). The mystery of language evolution. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00401

Hayes, S. A., Costa, D. P., Harvey, J. T., & le Boeuf, B. J. (2004). Aquatic mating strategies of the male Pacific harbor seal (*Phoca vitulina richardii*): Are males defending the hotspot? *Marine Mammal Science*, *20*(3), 639–656. https://doi.org/10.1111/j.1748-7692.2004.tb01184.x

Heinsohn, R., Zdenek, C. N., Cunningham, R. B., Endler, J. A., & Langmore, N. E. (2017). Tool- assisted rhythmic drumming in palm cockatoos shares key elements of human instrumental music. *Science Advances*, *3*(6), e1602399–e1602399. https://doi.org/10.1126/sciadv.1602399

Henry, M. J., Cook, P. F., de Reus, K., Nityananda, V., Rouse, A. A., & Kotz, S. A. (2021). An ecological approach to measuring synchronization abilities across the animal kingdom. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200336. https://doi.org/10.1098/rstb.2020.0336

Hespanhol, L., Vallio, C. S., Costa, L. M., & Saragiotto, B. T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian Journal of Physical Therapy*, *23*(4), 290–301. https://doi.org/10.1016/j.bjpt.2018.12.006

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1381–1388. https://doi.org/10.1007/s00265-013-1590-x

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*, 1492. https://doi.org/10.3389/fpsyg.2015.01492

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Hockett, C. F. (1963). The problem of universals in language. In J. H. Greenberg (Ed.), *Universals of Language* (pp. 1–29). The MIT Press.

Hocking, D. P., Burville, B., Parker, W., Evans, A. R., Park, T., & Marx, F. G. (2020). Percussive underwater signaling in wild gray seals. *Marine Mammal Science*, *36*(2), 728–732. https://doi.org/10.1111/mms.12666

Hoeksema, N., Verga, L., Mengede, J., van Roessel, C., Villanueva, S., Salazar-Casals, A., Rubio- Garcia, A., Ćurčić-Blake, B., Vernes, S. C., & Ravignani, A. (2021). Neuroanatomy of the grey seal brain: Bringing pinnipeds into the neurobiological study of vocal learning. *Philosophical Transactions: Biological Sciences*, *376*(1836), 1–4. https://doi.org/10.1098/rstb.2020.0252

Hoffmann, S., Trost, L., Voigt, C., Leitner, S., Lemazina, A., Sagunsky, H., Abels, M., Kollmansperger, S., Maat, A. T., & Gahr, M. (2019). Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior. *Nature Communications*, *10*(1), 2577. https://doi.org/10.1038/s41467-019-10593-3

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, *25*(5), 1900– 1908. https://doi.org/10.3758/s13423-017-1363-z

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, *133*(3), 692–697. https://doi.org/10.1016/j.cognition.2014.08.008

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*(6), 949–960. https://doi.org/10.1521/soco.2009.27.6.949

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Insley, S. J., Phillips, A. V., & Charrier, I. (2003). A review of social recognition in pinnipeds. *Aquatic Mammals*, *29*(2), 181–201. https://doi.org/10.1578/016754203101024149

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Janik, V., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, *26*, 59–99. https://doi.org/10.1016/S0065-3454(08)60377-0

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. https://doi.org/10.1126/science.aax0287

Jeng, M.-L. (2019). Biofluorescence in terrestrial animals, with emphasis on fireflies: A review and field observation. In H. Suzuki (Ed.), *Bioluminescence—Analytical Applications and Basic Biology*. IntechOpen. https://doi.org/10.5772/intechopen.86029

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews*, *26*(2), 235–258. https://doi.org/10.1016/S0149-7634(01)00068-9

Juslin, P. N., & Sloboda, J. A. (2013). Music and emotion. In *The Psychology of Music* (3rd ed., pp. 583–645). Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00015-8

Kallioinen, N., Paananen, T., Bürkner, P.-C., & Vehtari, A. (2023). Detecting and diagnosing prior and likelihood sensitivity with power-scaling. *Statistics and Computing*, *34*(1), 57. https://doi.org/10.1007/s11222-023-10366-5

Keitel, A., & Daum, M. M. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00108

Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130394. https://doi.org/10.1098/rstb.2013.0394

Kendrick, K. H., Holler, J., & Levinson, S. C. (2023). Turn-taking in human face-to-face interaction is multimodal: Gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1875), 20210473. https://doi.org/10.1098/rstb.2021.0473

Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., … Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, *91*(1), 13–52. https://doi.org/10.1111/brv.12160

Khan, C. B., Markowitz, H., & McCowan, B. (2006). Vocal development in captive harbor seal pups, *Phoca vitulina richardii*: Age, sex, and individual differences. *The Journal of the Acoustical Society of America*, *120*(3), 1684–1694. https://doi.org/10.1121/1.2226530

King, S. L. (2015). You talkin' to me? Interactive playback is a powerful yet underused tool in animal communication research. *Biology Letters*, *11*(7), 20150403. https://doi.org/10.1098/rsbl.2015.0403

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, *31*(5), 354–364. https://doi.org/10.1016/j.evolhumbehav.2010.04.004

Klump, G. M., & Gerhardt, H. C. (1992). Mechanisms and function of call-timing in male-male interactions in frogs. In P. K. McGregor (Ed.), *Playback and Studies of Animal Communication* (pp. 153–174). Springer New York. https://doi.org/10.1007/978-1-4757-6203-7_11

Kocsis, K., Duengen, D., Jadoul, Y., & Ravignani, A. (2024). Harbour seals use rhythmic percussive signalling in interaction and display. *Animal Behaviour*, *207*, 223–234. https://doi.org/10.1016/j.anbehav.2023.09.014

Kokko, H. (2017). Give one species the task to come up with a theory that spans them all: What good can come out of that? *Proceedings of the Royal Society B: Biological Sciences*, *284*(1867), 20171652. https://doi.org/10.1098/rspb.2017.1652

Kooyman, G. L., & Andersen, H. T. (1969). Deep diving. In H. T. Andersen (Ed.), *The Biology of Marine Mammals* (pp. 65–94). Academic Press.

Koren, L., & Geffen, E. (2009). Androgens and social status in female rock hyraxes. *Animal Behaviour*, *77*(1), 233–238. https://doi.org/10.1016/j.anbehav.2008.09.031

Kuester, J., Paul, A., & Arnemann, J. (1995). Age-related and individual differences of reproductive success in male and female barbary macaques (*Macaca sylvanus*). *Primates*, *36*(4), 461–476. https://doi.org/10.1007/BF02382869

Kuhelj, A., de Groot, M., Blejec, A., & Virant-Doberlet, M. (2016). Sender–receiver dynamics in leafhopper vibrational duetting. *Animal Behaviour*, *114*, 139–146. https://doi.org/10.1016/j.anbehav.2016.02.001

Kuhelj, A., Groot, M. de, Blejec, A., & Virant-Doberlet, M. (2015). The effect of timing of female vibrational reply on male signalling and searching behaviour in the leafhopper *Aphrodes makarovi*. *PLOS One*, *10*(10), e0139020. https://doi.org/10.1371/journal.pone.0139020

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lameira, A. R., Delgado, R. A., & Wich, S. A. (2010). Review of geographic variation in terrestrial mammalian acoustic signals: Human speech variation in a comparative perspective. *Journal of Evolutionary Psychology*, *8*(4), 309–332. https://doi.org/10.1556/jep.8.2010.4.2

Lampen, J., McAuley, J. D., Chang, S.-E., & Wade, J. (2019). Neural activity associated with rhythmicity of song in juvenile male and female zebra finches. *Behavioural Processes*, *163*, 45–52. https://doi.org/10.1016/j.beproc.2017.12.003

Langmore, N. E. (1998). Functions of duet and solo songs of female birds. *Trends in Ecology & Evolution*, *13*(4), 136–140. https://doi.org/10.1016/S0169-5347(97)01241-X

LaRue, K. M., Clemens, J., Berman, G. J., & Murthy, M. (2015). Acoustic duetting in *Drosophila virilis* relies on the integration of auditory and tactile signals. *eLife*, *4*, e07277. https://doi.org/10.7554/eLife.07277

Lattenkamp, E. Z., & Vernes, S. C. (2018). Vocal learning: A language-relevant trait in need of a broad cross-species approach. *Current Opinion in Behavioral Sciences*, *21*, 209–215. https://doi.org/10.1016/j.cobeha.2018.04.007

Le Boeuf, B. J. (1991). Pinniped mating systems on land, ice and in the water: Emphasis on the Phocidae. In D. Renouf (Ed.), *The Behaviour of Pinnipeds* (pp. 45–65). Springer Netherlands. https://doi.org/10.1007/978-94-011-3100-1_2

Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., & Cohen, D. (2014). Why synchrony matters during mother-child interactions: A systematic review. *PLOS One*, *9*(12), e113571. https://doi.org/10.1371/journal.pone.0113571

Leighty, K. A., Soltis, J., Wesolek, C. M., & Savage, A. (2008). Rumble vocalizations mediate interpartner distance in African elephants, *Loxodonta africana*. *Animal Behaviour*, *76*(5), 1601–1608. https://doi.org/10.1016/j.anbehav.2008.06.022

Lemasson, A., & Hausberger, M. (2004). Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (*Cercopithecus campbelli campbelli*). *Journal of Comparative Psychology*, *118*(3), 347–359. https://doi.org/10.1037/0735-7036.118.3.347

Lemasson, A., Ouattara, K., Petit, E. J., & Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology*, *11*(1), 362. https://doi.org/10.1186/1471- 2148-11-362

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928. https://doi.org/10.1111/oik.05985

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., & Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences*, *114*(50), 13290–13295. https://doi.org/10.1073/pnas.1702493114

Lester, B. M., & LaGasse, L. L. (2008). Crying. In M. M. Haith & J. B. Benson (Eds.), *Encyclopedia of Infant and Early Childhood Development* (pp. 332–343). Academic Press. https://doi.org/10.1016/B978-012370877-9.00045-1

Levinson, S. C. (2006). On the human 'interaction engine'. In Roots of Human Sociality. Routledge.

Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130302. https://doi.org/10.1098/rstb.2013.0302

Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J.-P., & Lemasson, A. (2019). Social bonding drives vocal exchanges in bonobos. *Scientific Reports*, *9*(1), 711. https://doi.org/10.1038/s41598-018-36024-9

Lieberman, P. H., Klatt, D. H., & Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, *164*(3884), 1185–1187. https://doi.org/10.1126/science.164.3884.1185

Lilly, J. C., & Miller, A. M. (1961). Sounds emitted by the bottlenose dolphin. *Science*, *133*(3465), 1689–1693. https://doi.org/10.1126/science.133.3465.1689

Liu, Y., Fan, H., Li, J., Jones, J. A., Liu, P., Zhang, B., & Liu, H. (2018). Auditory-motor control of vocal production during divided attention: Behavioral and ERP correlates. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00113

Loehr, J. D., & Palmer, C. (2011). Temporal coordination between performing musicians. *Quarterly Journal of Experimental Psychology*, *64*(11), 2153–2167. https://doi.org/10.1080/17470218.2011.603427

Lorenz, K. Z. (1974). Analogy as a source of knowledge. *Science*, *185*(4147), 229–234. https://doi.org/10.1126/science.185.4147.229

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lytle, S. R., & Kuhl, P. K. (2017). Social interaction and language acquisition. In E. M. Fernández & H. S. Cairns (Eds.), *The Handbook of Psycholinguistics* (pp. 615–634). Wiley-Blackwell. https://doi.org/10.1002/9781118829516

Maayan, C., Oren, A., Goldin, E., Dinour, D., Goldberg, M., & Mogle, P. (1990). Megaesophagus and recurrent apnea in an adult patient with familial dysautonomia. *American Journal of Gastroenterology*, *85*(6), 729–732.

Maccoby, E. E. (1994). The role of parents in the socialization of children: An historical overview. In R. D. Parke, P. A. Ornstein, J. J. Rieser, & C. Zahn-Waxler (Eds.), *A Century of Developmental Psychology* (pp. 589–615). American Psychological Association. https://doi.org/10.1037/10155-021

Mace, S., Shelton, G. D., & Eddlestone, S. (2012). Megaesophagus. *Compendium: Continuing Education for Veterinarians*, *34*(2), E1–E8.

MacRae, A. m., Haulena, M., & Fraser, D. (2011). The effect of diet and feeding level on survival and weight gain of hand-raised harbor seal pups (*Phoca vitulina*). *Zoo Biology*, *30*(5), 532–541. https://doi.org/10.1002/zoo.20356

Malacarne, G., Cucco, M., & Camanni, S. (1991). Coordinated visual displays and vocal duetting in different ecological situations among Western Palearctic non-passerine birds. *Ethology Ecology & Evolution*. https://doi.org/10.1080/08927014.1991.9525369

Mann, N., Dingess, K., Barker, K., Graves, J., & Slater, P. (2009). A comparative study of song form and duetting in neotropical *Thryothorus* wrens. *Behaviour*, 1–43. https://doi.org/10.1163/156853908X390913

Maretti, G., Sorrentino, V., Finomana, A., Gamba, M., & Giacoma, C. (2010). Not just a pretty song: An overview of the vocal repertoire of *Indri indri*. *Journal of Anthropological Sciences*, *88*, 151–165.

Markussen, N. H., Bjørge, A., & ØRitsland, N. A. (1989). Growth in harbour seals (*Phoca vitulina*) on the Norwegian coast. *Journal of Zoology*, *219*(3), 433–440. https://doi.org/10.1111/j.1469- 7998.1989.tb02591.x

Marten, K., & Marler, P. (1977). Sound transmission and its significance for animal vocalization. *Behavioral Ecology and Sociobiology*, *2*(3), 271–290. https://doi.org/10.1007/BF00299740

Martin, K., Tucker, M. A., & Rogers, T. L. (2017). Does size matter? Examining the drivers of mammalian vocalizations. *Evolution*, *71*(2), 249–260. https://doi.org/10.1111/evo.13128

Martin, M., Gridley, T., Elwen, S. H., & Charrier, I. (2021). Feel the beat: Cape fur seal males encode their arousal state in their bark rate. *The Science of Nature*, *109*(1), 5. https://doi.org/10.1007/s00114-021-01778-2

Mathevon, N., Casey, C., Reichmuth, C., & Charrier, I. (2017). Northern elephant seals memorize the rhythm and timbre of their rivals' voices. *Current Biology*, *27*(15), 2352–2356. https://doi.org/10.1016/j.cub.2017.06.035

McFarland, D. H. (2001). Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, *44*(1), 128–143. https://doi.org/10.1044/1092- 4388(2001/012)

Merchant, H., & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, *7*, 1–8. https://doi.org/10.3389/fnins.2013.00274

Migon, H. S., Gamerman, D., & Louzada, F. (2014). *Statistical Inference: An Integrated Approach* (2nd ed.). Chapman & Hall.

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., & Solow, A. R. (2004). Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*. *Animal Behaviour*, *67*(6), 1099–1107. https://doi.org/10.1016/j.anbehav.2003.06.017

Moore, B. L., Connor, R. C., Allen, S. J., Krützen, M., & King, S. L. (2020). Acoustic coordination by allied male dolphins in a cooperative context. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1924), 20192944. https://doi.org/10.1098/rspb.2019.2944

Moors, H. B., & Terhune, J. M. (2004). Repetition patterns in Weddell seal (*Leptonychotes weddellii*) underwater multiple element calls. *The Journal of the Acoustical Society of America*, *116*(2), 1261–1270. https://doi.org/10.1121/1.1763956

Moran, N., Hadley, L. V., Bader, M., & Keller, P. E. (2015). Perception of 'back-channeling' nonverbal feedback in musical duo improvisation. *PLOS One*, *10*(6), e0130070. https://doi.org/10.1371/journal.pone.0130070

Morisaka, T., Yoshida, Y., Akune, Y., Mishima, H., & Nishimoto, S. (2013). Exchange of "signature" calls in captive belugas (*Delphinapterus leucas*). *Journal of Ethology*, *31*(2), 141–149. https://doi.org/10.1007/s10164-013-0358-0

Mulsow, J., & Reichmuth, C. (2007). Electrophysiological assessment of temporal resolution in pinnipeds. *Aquatic Mammals*, *33*(1), 122–131. https://doi.org/10.1578/AM.33.1.2007.122

Mundry, R., & Sommer, C. (2007). Discriminant function analysis with nonindependent data: Consequences and an alternative. *Animal Behaviour*, *74*(4), 965–976. https://doi.org/10.1016/j.anbehav.2006.12.028

Nakahara, F., & Miyazaki, N. (2011). Vocal exchanges of signature whistles in bottlenose dolphins (*Tursiops truncatus*). *Journal of Ethology*, *29*(2), 309–320. https://doi.org/10.1007/s10164- 010-0259-4

Narins, P. (1982). Behavioral refractory period in neotropical treefrogs. *Journal of Comparative Physiology*, *148*, 337–344. https://doi.org/10.1007/BF00679018

Neumann, D. R. (1999). Agonistic behavior in harbor seals (*Phoca vitulina*) in relation to the availability of haul-out space. *Marine Mammal Science*, *15*(2), 507–525. https://doi.org/10.1111/j.1748-7692.1999.tb00816.x

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P., & Hoehl, S. (2021). Neural synchrony in mother–child conversation: Exploring the role of conversation patterns. *Social Cognitive and Affective Neuroscience*, *16*(1–2), 93–102. https://doi.org/10.1093/scan/nsaa079

Nieder, A., & Mooney, R. (2019). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190054. https://doi.org/10.1098/rstb.2019.0054

Nishimura, T., Tokuda, I. T., Miyachi, S., Dunn, J. C., Herbst, C. T., Ishimura, K., Kaneko, A., Kinoshita, Y., Koda, H., Saers, J. P. P., Imai, H., Matsuda, T., Larsen, O. N., Jürgens, U., Hirabayashi, H., Kojima, S., & Fitch, W. T. (2022). Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science*, *377*(6607), 760–763. https://doi.org/10.1126/science.abm1574

Nowicki, S., & Searcy, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, *28*, 48–53. https://doi.org/10.1016/j.conb.2014.06.007

O’Dell, M. L., Nieminen, T., & Lennes, M. (2012). Modeling turn-taking rhythms with oscillators. *Linguistica Uralica*, *48*(3), 218–227. https://doi.org/10.3176/lu.2012.3.08

Okobi, D. E., Banerjee, A., Matheson, A. M. M., Phelps, S. M., & Long, M. A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. *Science*, *363*(6430), 983–988. https://doi.org/10.1126/science.aau9480

Ord, T. J., & Garcia-Porta, J. (2012). Is sociality required for the evolution of communicative complexity? Evidence weighed against alternative hypotheses in diverse taxonomic groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1811– 1828. https://doi.org/10.1098/rstb.2011.0215

Osiecka, A. N., Gridley, T., & Fearey, J. (2022). Temporal patterns in Cape fur seal (*Arctocephalus pusillus pusillus*) mother and pup attraction calls. *Belgian Journal of Zoology*, *152*. https://doi.org/10.26496/bjz.2022.103

Osinga, N., Nussbaum, S. B., Brakefield, P. M., & Udo de Haes, H. A. (2012). Response of common seals (*Phoca vitulina*) to human disturbances in the Dollard estuary of the Wadden Sea. *Mammalian Biology*, *77*(4), 281–287. https://doi.org/10.1016/j.mambio.2012.02.005

Ota, N., Gahr, M., & Soma, M. (2015). Tap dancing birds: The multimodal mutual courtship display of males and females in a socially monogamous songbird. *Scientific Reports*, *5*(1), 16614. https://doi.org/10.1038/srep16614

Panzini, L., & Traube, M. (1993). Stridor from tracheal obstruction in a patient with achalasia. *American Journal of Gastroenterology*, *88*(7), 1097–1100.

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195123753.001.0001

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009a). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*(10), 827–830. https://doi.org/10.1016/j.cub.2009.03.038

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009b). Studying synchronization to a musical beat in nonhuman animals. *Annals of the New York Academy of Sciences*, *1169*(1), 459–469. https://doi.org/10.1111/j.1749-6632.2009.04581.x

Perez, E. C., Elie, J. E., Boucaud, I. C. A., Crouchet, T., Soulage, C. O., Soula, H. A., Theunissen, F. E., & Vignal, C. (2015). Physiological resonance between mates through calls as possible evidence of empathic processes in songbirds. *Hormones and Behavior*, *75*, 130–141. https://doi.org/10.1016/j.yhbeh.2015.09.002

Perry, E. A., & Renouf, D. (1988). Further studies of the role of harbour seal (*Phoca vitulina*) pup vocalizations in preventing separation of mother–pup pairs. *Canadian Journal of Zoology*, *66*(4), 934–938. https://doi.org/10.1139/z88-138

Peterson, R. A. (2021). Finding optimal normalizing transformations via bestNormalize. *The R Journal*, *13*.

Peterson, R. A., & Cavanaugh, J. E. (2020). Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, *47*(13–15), 2312–2327. https://doi.org/10.1080/02664763.2019.1630372

Petkov, C. I., & Jarvis, E. (2012). Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, *4*. https://doi.org/10.3389/fnevo.2012.00012

Pfefferle, D., & Fischer, J. (2006). Sounds and size: Identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. *Animal Behaviour*, *72*(1), 43–51. https://doi.org/10.1016/j.anbehav.2005.08.021

Pfefferle, D., Hammerschmidt, K., Mundry, R., Ruiz-Lambides, A. V., Fischer, J., & Widdig, A. (2016). Does the structure of female rhesus macaque coo calls reflect relatedness and/or familiarity? *PLOS One*, *11*(8), e0161133. https://doi.org/10.1371/journal.pone.0161133

Pfefferle, D., West, P. M., Grinnell, J., Packer, C., & Fischer, J. (2007). Do acoustic features of lion, *Panthera leo*, roars reflect sex and male condition? *The Journal of the Acoustical Society of America*, *121*(6), 3947–3953. https://doi.org/10.1121/1.2722507

Phillips-Silver, J., Aktipis, C. A., & A. Bryant, G. (2010). The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music Perception*, *28*(1), 3–14. https://doi.org/10.1525/mp.2010.28.1.3

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). Gestural communication of apes. *Gesture*, *5*(1– 2), 41–56. https://doi.org/10.1075/gest.5.1.05pik

Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1880), 20180598. https://doi.org/10.1098/rspb.2018.0598

Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., & Feinberg, D. R. (2016). Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Scientific Reports*, *6*(1), 34389. https://doi.org/10.1038/srep34389

Pisanski, K., Raine, J., & Reby, D. (2020). Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society Open Science*, *7*(2), 191642. https://doi.org/10.1098/rsos.191642

Pleyer, M., Perlman, M., Lupyan, G., de Reus, K., & Raviv, L. (2025). The 'design features' of language revisited. *Trends in Cognitive Sciences*. *Advance online publication*. https://doi.org/10.1016/j.tics.2025.10.004

Plotsky, K., Rendall, D., Riede, T., & Chase, K. (2013). Radiographic analysis of vocal tract length and its relation to overall body size in two canid species. *Journal of Zoology*, *291*(1), 76–86. https://doi.org/10.1111/jzo.12048

Poole, J. H. (1999). Signals and assessment in African elephants: Evidence from playback experiments. *Animal Behaviour*, *58*(1), 185–193. https://doi.org/10.1006/anbe.1999.1117

Pouw, W., & Holler, J. (2022). Timing in conversation is dynamically adjusted turn by turn in dyadic telephone conversations. *Cognition*, *222*, 105015. https://doi.org/10.1016/j.cognition.2022.105015

Pouw, W., Proksch, S., Drijvers, L., Gamba, M., Holler, J., Kello, C., Schaefer, R. S., & Wiggins, G. A. (2021). Multilevel rhythms in multimodal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200334. https://doi.org/10.1098/rstb.2020.0334

Prat, Y., Azoulay, L., Dor, R., & Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: Playback of conspecifics shapes fundamental frequency usage by pups. *PLOS Biology*, *15*(10), e2002556. https://doi.org/10.1371/journal.pbio.2002556

Prior, N. H., Smith, E., Dooling, R. J., & Ball, G. F. (2020). Familiarity enhances moment-to-moment behavioral coordination in zebra finch (*Taeniopygia guttata*) dyads. *Journal of Comparative Psychology*, *134*(2), 135–148. https://doi.org/10.1037/com0000201

Pugliares, K. R., Bogomolni, A., Touhey, K. M., Herzig, S. M., & Harry, C. T. (2007). *Marine mammal necropsy: An introductory guide for stranding responders and field biologists.* [Report]. Woods Hole Oceanographic Institution. https://doi.org/10.1575/1912/1823

Rabinowitch, T.-C., & Knafo-Noam, A. (2015). Synchronous rhythmic interaction enhances children's perceived similarity and closeness towards each other. *PLOS One*, *10*(4), e0120878. https://doi.org/10.1371/journal.pone.0120878

Rado, R., Levi, N., Hauser, H., Witcher, J., Alder, N., Intrator, N., Wollberg, Z., & Terkel, J. (1987). Seismic signalling as a means of communication in a subterranean mammal. *Animal Behaviour*, *35*(4), 1249–1251.

Rajaraman, K., Godthi, V., Pratap, R., & Balakrishnan, R. (2015). A novel acoustic-vibratory multimodal duet. *Journal of Experimental Biology*, *218*(19), 3042–3050. https://doi.org/10.1242/jeb.122911

Ralls, K., Fiorelli, P., & Gish, S. (1985). Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Canadian Journal of Zoology*, *63*(5), 1050–1056. https://doi.org/10.1139/z85-157

Ravignani, A. (2017). Agree on definitions of synchrony. *Nature*, *545*(7653), 158–158. https://doi.org/10.1038/545158c

Ravignani, A. (2019). Timing of antisynchronous calling: A case study in a harbor seal pup (*Phoca vitulina*). *Journal of Comparative Psychology*, *133*(2), 272–277. https://doi.org/10.1037/com0000160

Ravignani, A., Bowling, D. L., & Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01118

Ravignani, A., Fitch, W. T., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., Scharff, C., Stoeger, A. S., & de Boer, B. (2016). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Frontiers in Neuroscience*, *10*. https://doi.org/10.3389/fnins.2016.00274

Ravignani, A., & Garcia, M. (2022). A cross-species framework to identify vocal learning abilities in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1841), 20200394. https://doi.org/10.1098/rstb.2020.0394

Ravignani, A., Gross, S., Garcia, M., Rubio-Garcia, A., & de Boer, B. (2017). How small could a pup sound? The physical bases of signaling body size in harbor seals. *Current Zoology*, *63*(4), 457–465. https://doi.org/10.1093/cz/zox026

Ravignani, A., Kello, C. T., de Reus, K., Kotz, S. A., Dalla Bella, S., Méndez-Aróstegui, M., Rapado- Tamarit, B., Rubio-Garcia, A., & de Boer, B. (2019). Ontogeny of vocal rhythms in harbor seal pups: An exploratory study. *Current Zoology*, *65*(1), 107–120. https://doi.org/10.1093/cz/zoy055

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

Ravignani, A., Verga, L., & Greenfield, M. D. (2019). Interactive rhythms across species: The evolutionary biology of animal chorusing and turn-taking. *Annals of the New York Academy of Sciences*, *1453*(1), 12–21. https://doi.org/10.1111/nyas.14230

Reby, D., & McComb, K. (2003). Anatomical constraints generate honesty: Acoustic cues to age and weight in the roars of red deer stags. *Animal Behaviour*, *65*(3), 519–530. https://doi.org/10.1006/anbe.2003.2078

Reijnders, P. J. H., Brasseur, S. M. J. M., & Meesters, E. H. W. G. (2010). Earlier pupping in harbour seals, *Phoca vitulina*. *Biology Letters*, *6*(6), 854–857. https://doi.org/10.1098/rsbl.2010.0468

Ręk, P. (2018). Multimodal coordination enhances the responses to an avian duet. *Behavioral Ecology*, *29*(2), 411–417. https://doi.org/10.1093/beheco/arx174

Ręk, P., & Magrath, R. D. (2016). Multimodal duetting in magpie-larks: How do vocal and visual components contribute to a cooperative signal's function? *Animal Behaviour*, *117*, 35–42. https://doi.org/10.1016/j.anbehav.2016.04.024

Renouf, D. (1984). The vocalization of the harbour seal pup (*Phoca vitulina*) and its role in the maintenance of contact with the mother. *Journal of Zoology*, *202*(4), 583–590. https://doi.org/10.1111/j.1469-7998.1984.tb05055.x

Renouf, D. (1985). A demonstration of the ability of the harbour seal *Phoca vitulina* (L.) to discriminate among pup vocalizations. *Journal of Experimental Marine Biology and Ecology*, *87*(1), 41–46. https://doi.org/10.1016/0022-0981(85)90190-X

Renouf, D. (Ed.). (1991). *The Behaviour of Pinnipeds*. Springer Netherlands. https://doi.org/10.1007/978-94-011-3100-1

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, *12*(6), 969–992. https://doi.org/10.3758/BF03206433

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006– 2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452. https://doi.org/10.3758/s13423-012-0371-2

Richmond, J. P., Norris, T., & Zinn, S. A. (2010). Re-alimentation in harbor seal pups: Effects on the somatotropic axis and growth rate. *General and Comparative Endocrinology*, *165*(2), 286–292. https://doi.org/10.1016/j.ygcen.2009.07.007

Riedman, M. (1990). *The Pinnipeds: Seals, Sea Lions, and Walruses* (Vol. 12). University of California Press. https://doi.org/10.2307/jj.8501146

Rivera-Cáceres, K. D., Quirós-Guerrero, E., Araya-Salas, M., & Searcy, W. A. (2016). Neotropical wrens learn new duet rules as adults. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1843), 20161819. https://doi.org/10.1098/rspb.2016.1819

Roberts, T. F., Hisey, E., Tanaka, M., Kearney, M. G., Chattree, G., Yang, C. F., Shah, N. M., & Mooney, R. (2017). Identification of a motor-to-auditory pathway important for vocal learning. *Nature Neuroscience*, *20*(7), 978–986. https://doi.org/10.1038/nn.4563

Roers, F., Mürbe, D., & Sundberg, J. (2009). Predicted singers' vocal fold lengths and voice classification—A study of x-ray morphological measures. *Journal of Voice*, *23*(4), 408–413. https://doi.org/10.1016/j.jvoice.2007.12.003

Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction Studies*, *14*(2), 160–189. https://doi.org/10.1075/is.14.2.02ros

Rua Ventura, S. M., Freitas, D. R. S., Ramos, I. M. A. P., & Tavares, J. M. R. S. (2013). Morphologic differences in the vocal tract resonance cavities of voice professionals: An MRI-based study. *Journal of Voice*, *27*(2), 132–140. https://doi.org/10.1016/j.jvoice.2012.11.010

Ruch, H., Zürcher, Y., & Burkart, J. M. (2018). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*, *93*(2), 996–1013. https://doi.org/10.1111/brv.12382

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn- taking for conversation. *Language*, *50*(4), 696–735.

Salazar-Casals, A., de Reus, K., Greskewitz, N., Havermans, J., Geut, M., Villanueva, S., & Rubio- Garcia, A. (2022). Increased incidence of entanglements and ingested marine debris in Dutch seals from 2010 to 2020. *Oceans*, *3*(3), Article 3. https://doi.org/10.3390/oceans3030026

Salkind, N. J. (2010). *Encyclopedia of Research Design*. SAGE Publications. https://doi.org/10.4135/9781412961288

Sanvito, S., & Galimberti, F. (2000). Bioacoustics of Southern elephant seals. I. Acoustic structure of male aggressive vocalisations. *Bioacoustics*, *10*(4), 259–285. https://doi.org/10.1080/09524622.2000.9753438

Sauvé, C. C., Beauplet, G., Hammill, M. O., & Charrier, I. (2015). Acoustic analysis of airborne, underwater, and amphibious mother attraction calls by wild harbor seal pups (*Phoca vitulina*). *Journal of Mammalogy*, *96*(3), 591–602. https://doi.org/10.1093/jmammal/gyv064

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2020). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59. https://doi.org/10.1017/S0140525X20000333

Schaller, G. E. (1963). *The Mountain Gorilla: Ecology and Behavior*. University of Chicago Press. Schneider, R. (1962). Vergleichende Untersuchungen am Kehlkopf der Robben (mammalia, carnivora, pinnipedia). *Gegenbaurs Morph Jb*, *103*, 177–262.

Schneider, R. (1964). *Der Larynx der Säugetiere*. De Gruyter.

Schultz, B. G., O'brien, I., Phillips, N., McFarland, D. H., Titone, D., & Palmer, C. (2016). Speech rates converge in scripted turn-taking conversations. *Applied Psycholinguistics*, *37*(5), 1201–1220. https://doi.org/10.1017/S0142716415000545

Schulz, T. M., Whitehead, H., Gero, S., & Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: Insights into communication function. *Animal Behaviour*, *76*(6), 1977–1988. https://doi.org/10.1016/j.anbehav.2008.07.032

Schulz, T. M., Whitehead, H., Gero, S., & Rendell, L. (2011). Individual vocal production in a sperm whale (*Physeter macrocephalus*) social unit. *Marine Mammal Science*, *27*(1), 149–166. https://doi.org/10.1111/j.1748-7692.2010.00399.x

Schusterman, R. J. (1977). Temporal patterning in sea lion barking (*Zalophus californianus*). *Behavioral Biology*, *20*(3), 404–408. https://doi.org/10.1016/S0091-6773(77)90964-6

Schusterman, R. J. (2008). Vocal learning in mammals with special emphasis on pinnipeds. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communicative Flexibility: Complexity, Creativity, and Adaptability in Human and Animal Communication* (pp. 41–70). The MIT Press. https://doi.org/10.7551/mitpress/9780262151214.003.0003

Schusterman, R. J., & van Parijs, S. M. (2003). Pinniped vocal communication: An introduction. *Aquatic Mammals*, *29*(2), 177–180. https://doi.org/10.1578/016754203101024103

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, *54*, 145–173. https://doi.org/10.1146/annurev.psych.54.101601.145121

Shapiro, A. D., Slater, P. J. B., & Janik, V. M. (2004). Call usage learning in gray seals (*Halichoerus grypus*). *Journal of Comparative Psychology*, *118*(4), 447–454. https://doi.org/10.1037/0735- 7036.118.4.447

Short, L. L., & Horne, J. F. M. (1983). A review of duetting, sociality and speciation in some African barbets (Capitonidae). *The Condor*, *85*(3), 323–332. https://doi.org/10.2307/1367069

Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social bonds of female baboons enhance infant survival. *Science*, *302*(5648), 1231–1234. https://doi.org/10.1126/science.1088580

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist*, *17*(2), 197–208. https://doi.org/10.1177/1073858410386727

Simonyan, K., & Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research*, *974*(1), 43–59. https://doi.org/10.1016/S0006- 8993(03)02548-4

Soma, M., & Garamszegi, L. Z. (2015). Evolution of courtship display in Estrildid finches: Dance in relation to female song and plumage ornamentation. *Frontiers in Ecology and Evolution*, *3*. https://doi.org/10.3389/fevo.2015.00004

Stan, G.-B., Embrechts, J.-J., & Archambeau, D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio Engineering Society*, *50*(4).

Stansbury, A. L., & Janik, V. M. (2021). The role of vocal learning in call acquisition of wild grey seal pups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200251. https://doi.org/10.1098/rstb.2020.0251

Stanton-Chapman, T. L., & Snell, M. E. (2011). Promoting turn-taking skills in preschool children with disabilities: The effects of a peer-based social communication intervention. *Early Childhood Research Quarterly*, *26*(3), 303–319. https://doi.org/10.1016/j.ecresq.2010.11.002

Stegmann, U. E. (2005). John Maynard Smith's notion of animal signals. *Biology and Philosophy*, *20*(5), 1011–1025. https://doi.org/10.1007/s10539-005-9020-8

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587– 10592. https://doi.org/10.1073/pnas.0903616106

Striedter, G. F. (1994). The vocal control pathways in budgerigars differ from those in songbirds. *Journal of Comparative Neurology*, *343*(1), 35–56. https://doi.org/10.1002/cne.903430104

Takahashi, D. Y., Fenley, A. R., & Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150370. https://doi.org/10.1098/rstb.2015.0370

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology*, *23*(21), 2162–2168. https://doi.org/10.1016/j.cub.2013.09.005

Takeda, K. F., Hiraiwa-Hasegawa, M., & Kutsukake, N. (2019). Uncoordinated dances associated with high reproductive success in a crane. *Behavioral Ecology*, *30*(1), 101–106. https://doi.org/10.1093/beheco/ary159

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology*, *280*(3), 221–236. https://doi.org/10.1111/j.1469- 7998.2009.00661.x

ter Bekke, M., Levinson, S. C., van Otterdijk, L., Kühn, M., & Holler, J. (2024). Visual bodily signals and conversational context benefit the anticipation of turn ends. *Cognition*, *248*, 105806. https://doi.org/10.1016/j.cognition.2024.105806

Terhune, J. M., & Ronald, K. (1986). Distant and near-range functions of harp seal underwater calls. *Canadian Journal of Zoology*, *64*(5), 1065–1070. https://doi.org/10.1139/z86-159

Tibbetts, E. A., & Dale, J. (2007). Individual recognition: It is good to be different. *Trends in Ecology & Evolution*, *22*(10), 529–537. https://doi.org/10.1016/j.tree.2007.09.001

Tilson, R. L., & Norton, P. M. (1981). Alarm duetting and pursuit deterrence in an African antelope. *The American Naturalist*, *118*(3), 455–462. https://doi.org/10.1086/283840

Timmers, R., Endo, S., Bradbury, A., & Wing, A. M. (2014). Synchronization and leadership in string quartet performance: A case study of auditory and visual cues. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00645

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Titze, I. R. (1994). *Principles of Voice Production*. Prentice Hall.

Titze, I. R., Luschei, E. S., & Hirano, M. (1989). Role of the thyroarytenoid muscle in regulation of fundamental frequency. *Journal of Voice*, *3*(3), 213–224. https://doi.org/10.1016/S0892- 1997(89)80003-7

Tobias, J. A., Sheard, C., Seddon, N., Meade, A., Cotton, A. J., & Nakagawa, S. (2016). Territoriality, social bonds, and the evolution of communal signaling in birds. *Frontiers in Ecology and Evolution*, *4*. https://doi.org/10.3389/fevo.2016.00074

Tobias, M. L., Viswanathan, S. S., & Kelley, D. B. (1998). Rapping, a female receptive call, initiates male–female duets in the South African clawed frog. *Proceedings of the National Academy of Sciences*, *95*(4), 1870–1875. https://doi.org/10.1073/pnas.95.4.1870

Toji, N., Sawai, A., Wang, H., Ji, Y., Sugioka, R., Go, Y., & Wada, K. (2024). A predisposed motor bias shapes individuality in vocal learning. *Proceedings of the National Academy of Sciences*, *121*(3), e2308837121. https://doi.org/10.1073/pnas.2308837121

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Torres Borda, L., Jadoul, Y., Rasilo, H., Salazar Casals, A., & Ravignani, A. (2021). Vocal plasticity in harbour seal pups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1840), 20200456. https://doi.org/10.1098/rstb.2020.0456

Torti, V., Bonadonna, G., De Gregorio, C., Valente, D., Randrianarison, R. M., Friard, O., Pozzi, L., Gamba, M., & Giacoma, C. (2017). An intra-population analysis of the indris' song dissimilarity in the light of genetic distance. *Scientific Reports*, *7*(1), 10140. https://doi.org/10.1038/s41598-017-10656-9

Torti, V., Gamba, M., Rabemananjara, Z. H., & Giacoma, C. (2013). The songs of the indris (Mammalia: Primates: Indridae): Contextual variation in the long-distance calls of a lemur. *Italian Journal of Zoology*, *80*(4), 596–607. https://doi.org/10.1080/11250003.2013.845261

Trainor, L. J., & Cirelli, L. (2015). Rhythm and interpersonal synchrony in early social development. *Annals of the New York Academy of Sciences*, *1337*(1), 45–52. https://doi.org/10.1111/nyas.12649

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140096. https://doi.org/10.1098/rstb.2014.0096

Trehub, S. E., Ghazban, N., & Corbeil, M. (2015). Musical affect regulation in infancy. *Annals of the New York Academy of Sciences*, *1337*(1), 186–192. https://doi.org/10.1111/nyas.12622

Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction: Age and gender differences in coordination and the occurrence of miscoordination. *Child Development*, *60*(1), 85–92. https://doi.org/10.2307/1131074

Tucker, A. S. (2017). Major evolutionary transitions and innovations: The tympanic middle ear. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150483. https://doi.org/10.1098/rstb.2015.0483

Turnbull, S. D., & Terhune, J. M. (1993). Repetition enhances hearing detection thresholds in a harbour seal (*Phoca vitulina*). *Canadian Journal of Zoology*, *71*(5), 926–932. https://doi.org/10.1139/z93-120

Tuthill, J. C., & Wilson, R. I. (2016). Mechanosensation and adaptive motor control in insects. *Current Biology*, *26*(20), R1022–R1038. https://doi.org/10.1016/j.cub.2016.06.070

Tuttle, M. D., & Ryan, M. J. (1982). The role of synchronized calling, ambient light, and ambient noise, in anti-bat-predator behavior of a treefrog. *Behavioral Ecology and Sociobiology*, *11*(2), 125–131. https://doi.org/10.1007/BF00300101

Tyack, P. L. (2019). A taxonomy for vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20180406. https://doi.org/10.1098/rstb.2018.0406

Tyack, P. L., & Miller, E. (2002). Vocal anatomy, acoustic communication and echolocation. In A. R. Hoelzel (Ed.), *Marine Mammal Biology: An Evolutionary Approach* (pp. 142–184). Wiley- Blackwell.

Valente, D., De Gregorio, C., Torti, V., Miaretsoa, L., Friard, O., Randrianarison, R. M., Giacoma, C., & Gamba, M. (2019). Finding meanings in low dimensional structures: Stochastic neighbor embedding applied to the analysis of *Indri indri* vocal repertoire. *Animals*, *9*(5), 243. https://doi.org/10.3390/ani9050243

van den Berg, J. W. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, *1*(3), 227–244. https://doi.org/10.1044/jshr.0103.227

van den Berg, J. W., Zantema, J. T., & Doornenbal Jr., P. (1957). On the air resistance and the Bernoulli effect of the human larynx. *The Journal of the Acoustical Society of America*, *29*(5), 626–631. https://doi.org/10.1121/1.1908987

van den Bosch der Nederlanden, C. M., Taylor, J. E. T., & Grahn, J. A. (2019). Neural basis of rhythm perception. In *The Oxford Handbook of Music and the Brain* (pp. 165–186). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198804123.001.0001

van der Steen, M. C., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00253 van Parijs, S. M., Corkeron, P. J., Harvey, J., Hayes, S. A., Mellinger, D. K., Rouget, P. A., Thompson, P. M., Wahlberg, M., & Kovacs, K. M. (2003). Patterns in the vocalizations of male harbor seals. *The Journal of the Acoustical Society of America*, *113*(6), 3403–3410. https://doi.org/10.1121/1.1568943

van Parijs, S. M., Hastie, G. D., & Thompson, P. M. (2000). Individual and geographical variation in display behaviour of male harbour seals in Scotland. *Animal Behaviour*, *59*(3), 559–568. https://doi.org/10.1006/anbe.1999.1307

van Parijs, S. M., & Kovacs, K. M. (2002). In-air and underwater vocalizations of eastern Canadian harbour seals, *Phoca vitulina*. *Canadian Journal of Zoology*, *80*(7), 1173–1179. https://doi.org/10.1139/z02-088

van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychonomic Bulletin & Review*, *25*(1), 143–154. https://doi.org/10.3758/s13423-016-1015-8

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer. https://doi.org/10.1007/978-0-387-21706-2

Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., ten Cate, C., & Janik, V. M. (2021). The multi-dimensional nature of vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200236. https://doi.org/10.1098/rstb.2020.0236

Vignal, C., & Kelley, D. (2006). Significance of temporal and spectral acoustic cues for sexual recognition in *Xenopus laevis*. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1609), 479–488. https://doi.org/10.1098/rspb.2006.3744

Villarreal, S. M., & Gilbert, C. (2014). Male *Scudderia pistillata* katydids defend their acoustic duet against eavesdroppers. *Behavioral Ecology and Sociobiology*, *68*(10), 1669–1675. https://doi.org/10.1007/s00265-014-1775-y

Weihs, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR analyzing german business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme (Eds.), *Data Analysis and Decision Support* (pp. 335– 343). Springer. https://doi.org/10.1007/3-540-28397-8_36

Weissengruber, G. E., Forstenpointner, G., Peters, G., Kübber-Heiss, A., & Fitch, W. T. (2002). Hyoid apparatus and pharynx in the lion (*Panthera leo*), jaguar (*Panthera onca*), tiger (*Panthera tigris*), cheetah (*Acinonyx jubatus*) and domestic cat (*Felis silvestris f. Catus*). *Journal of Anatomy*, *201*(3), 195–209. https://doi.org/10.1046/j.1469-7580.2002.00088.x

Wells, K. D. (1977). The social behaviour of anuran amphibians. *Animal Behaviour*, *25*, 666–693. https://doi.org/10.1016/0003-3472(77)90118-X

Wells, K. D., & Schwartz, J. J. (2006). The behavioral ecology of anuran communication. In P. M. Narins, A. S. Feng, R. R. Fay, & A. N. Popper (Eds.), *Hearing and Sound Communication in Amphibians* (pp. 44–86). Springer. https://doi.org/10.1007/978-0-387-47796-1_3

Wild, J. M. (1997). Neural pathways for the control of birdsong production. *Journal of Neurobiology*, *33*(5), 653–670. https://doi.org/10.1002/(SICI)1097-4695(19971105)33:5<653::AID-NEU11>3.0.CO;2-A

Wiley, R. H., & Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology*, *3*(1), 69–94. https://doi.org/10.1007/BF00300047

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*(6), 957–968. https://doi.org/10.3758/BF03206432

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, *20*(1), 1–5. https://doi.org/10.1111/j.1467-9280.2008.02253.x

Wing, A. M., Endo, S., Bradbury, A., & Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, *11*(93), 20131125. https://doi.org/10.1098/rsif.2013.1125

Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V., Miller, C. T., Pfenning, A. R., Vernes, S. C., Tchernichovski, O., & Yartsev, M. M. (2019). A modular approach to vocal learning: Disentangling the diversity of a complex behavioral trait. *Neuron*, *104*(1), 87–99. https://doi.org/10.1016/j.neuron.2019.09.036

Wyman, M. T., Mooring, M. S., McCowan, B., Penedo, M. C. T., & Hart, L. A. (2008). Amplitude of bison bellows reflects male quality, physical condition and motivation. *Animal Behaviour*, *76*(5), 1625–1639. https://doi.org/10.1016/j.anbehav.2008.05.032

Yoshida, S., & Okanoya, K. (2005). Evolution of turn-taking: A bio-cognitive perspective. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *12*(3), 153–165. https://doi.org/10.11225/jcss.12.153

Yosida, S., & Okanoya, K. (2009). Naked mole-rat is sensitive to social hierarchy encoded in antiphonal vocalization. *Ethology*, *115*(9), 823–831. https://doi.org/10.1111/j.1439-0310.2009.01677.x

Zahavi, A. (1977). The cost of honesty: Further remarks on the handicap principle. *Journal of Theoretical Biology*, *67*(3), 603–605. https://doi.org/10.1016/0022-5193(77)90061-3

Zamm, A., Debener, S., & Sebanz, N. (2023). The spontaneous emergence of rhythmic coordination in turn taking. *Scientific Reports*, *13*(1), 3259. https://doi.org/10.1038/s41598-022-18480-6

Zamm, A., Wellman, C., & Palmer, C. (2016). Endogenous rhythms influence interpersonal synchrony. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 611–616. https://doi.org/10.1037/xhp0000201

Zhang, Z. (2016). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America*, *140*(4), 2614–2635. https://doi.org/10.1121/1.4964509

Zimmermann, U., Rheinlaender, J., & Robinson, D. (1989). Cues for male phonotaxis in the duetting bushcricket *Leptophyes punctatissima*. *Journal of Comparative Physiology A*, *164*(5), 621–628. https://doi.org/10.1007/BF00614504

English summary

How and why did language evolve in humans? Can we learn from studying communication in other animals? Language is a uniquely effective and powerful tool to interact with others and is often considered to be what sets us apart from the rest of the animal kingdom. Despite this, many species have also evolved elaborate communication systems, some of which even show features found in human language. The discovery in the 1980s that harbour seals can imitate human speech promised a new comparative model for language evolution studies. However, we currently miss a comprehensive understanding of the features shared between human language and harbour seal communication to enable effective cross-species comparisons. This thesis provides a long-awaited, interdisciplinary account of some vocal communication features in harbour seals, including vocal flexibility, vocal development, social accommodation, and turn-taking. All the empirical studies in this thesis were performed using non-invasive methods from bioacoustics (audio recordings and a playback experiment) and morphometrics (anatomical measurements). Additionally, this thesis contributes a qualitative cross-species review on the rhythmic patterns observed in social interactions between animal dyads of several mammals, birds, anurans, and insects. Taken together, my findings show that harbour seal communication shares features with that of other species and human language, reinforcing its value as a comparative model for language evolution research. Finally, this work demonstrates how the comparative approach can be applied to communicative behaviours to shed light on the biological and social factors driving their evolution.

Nederlandse samenvatting

Hoe en waarom evolueerde taal in mensen? Kunnen we iets leren van communicatie in andere diersoorten? Taal is een uniek, effectief en krachtig instrument om met elkaar te communiceren en algemeen wordt aangenomen dat het ons onderscheidt van de rest van het dierenrijk. Desondanks hebben andere diersoorten ook uitgebreide communicatiesystemen ontwikkeld, waarbij sommige zelfs kenmerken vertonen die ook in taal voorkomen. De ontdekking in de jaren tachtig dat gewone zeehonden menselijke spraak kunnen imiteren was veelbelovend voor een nieuw vergelijkend model voor taalevolutiestudies. We missen momenteel echter een alomvattend begrip van de kenmerken die menselijke taal en communicatie van gewone zeehonden gemeen hebben om effectieve vergelijkingen tussen de twee mogelijk te maken. Dit proefschrift biedt een langverwacht, interdisciplinair verslag van enkele vocale communicatie kenmerken in gewone zeehonden, inclusief vocale flexibiliteit, vocale ontwikkeling, sociale aanpassing en het wisselen van beurten. Alle empirische onderzoeken in dit proefschrift zijn uitgevoerd met behulp van niet-invasieve methoden uit de bio-akoestiek (audio-opnamen en een afspeelexperiment) en morfometrie (anatomische metingen). Daarnaast levert dit proefschrift ook een kwalitatieve review van de ritmische patronen die worden waargenomen in sociale interacties tussen paren van verschillende zoogdieren, vogels, anuranen en insecten. Mijn bevindingen tonen aan dat de communicatie van gewone zeehonden kenmerken deelt met die van andere soorten en menselijke taal, waardoor de waarde van de gewone zeehond als vergelijkend model voor onderzoek naar taalevolutie wordt versterkt. Ten slotte laat dit werk zien hoe de vergelijkende aanpak kan worden toegepast om licht te werpen op de biologische en sociale factoren die de evolutie van communicatief gedrag aansturen.

Research data management

Ethical note

The experimental chapters in this thesis are based on results of bioacoustics studies in non-human animals. All studies prioritised the health and welfare of the animals and were conducted in accordance with the Dutch and EU regulations on the ethical treatment of animals in behavioural research. All recordings adhered to the guidelines set by the Veterinary Department and were performed in close coordination with the Sealcare Department of Sealcentre Pieterburen.

Data availability

No datasets or code were produced for Chapter 2. The audio and video data from Chapter 3, including detailed documentation of the experimental set-up and procedures, were archived on the institutional servers of the Max Planck Institute for Psycholinguistics and will be preserved for a minimum of 10 years. These data, along with the annotations, preprocessing scripts, datasets, and analysis code are available upon reasonable request. For Chapter 4, a subset of the data, comprising the list of animals with body size information and relevant metadata, is provided in the supplementary materials at DOI: 10.1242/jeb.243766. The full dataset and the analysis code are available upon reasonable request. For Chapter 5, the audio data, annotations, preprocessing scripts, dataset, and analysis code are also available upon reasonable request.

Acknowledgments

It feels unreal to start writing this section of my thesis and realise how far I've come. This project started during the COVID-19 pandemic and took almost five years to complete. It was a true rollercoaster of emotions, with its typical ups and down. The various research projects I developed during this PhD slowly transformed me into an expert on a very niche topic, and this has felt rather lonely at times. But in reality, it has been anything but lonely, because I would not have been able to come this far without the incredible support of the people around me.

I want to begin by thanking **Andrea Ravignani**, my daily supervisor, and the man who made this all possible. I have a vivid memory of the first time we met online to discuss me coming to Sealcentre Pieterburen to carry out a research project on harbour seal vocalisations. After completing that initial project, I kept coming back to be with the seals and continue working with you in Pieterburen, and we even lived in the same house for some time. I then took a teaching job in Rotterdam and regularly drove up and down so that we could continue what we started, be it running around with a microphone, setting up whole playback experiments, performing dissections in the necropsy room, or simply cleaning up seal shit. And suddenly, just before the start of the pandemic, you sent me a grant proposal. At first, I thought you were asking me for feedback on the proposal, but you then surprised me, told me the application was successful, and offered me a PhD position. This all began in 2017, and here I stand almost a decade later defending the research you inspired. Thank you for taking me along on this wild journey, sharing your passion for research, and the opportunity of a lifetime. I also want to thank my two promotors, **Bart de Boer** and **Judith Holler**, for their continued support and guidance in this project. Although neither of you study seals, you both have been essential in helping me navigate the intricacies of doing a double doctorate and framing my research questions to the broader scientific community. I also want to thank all the members of my jury, **Asli Özyürek**, **Wim Pouw**, **Elisa Gonzalez Boix**, **Paul Van Eecke**, **Sylvie Nozaradan**, and **Sonja Vernes**, for their time and consideration. Finally, a very big thank you to all the other scientific colleagues, **Masayo**, **Marianna**, **Marco**, **Marianne**, **Miriam**, **Julia**, **Laurel**, **Anna**, **Silvia**, **Peter**, **Daryll**, **Alice**,

Stephanie, **Maxime**, **Ana**, and **Yannick**, who have collaborated on the projects that are part of this thesis.

Next, I want to thank my direct colleagues from both research groups. To all the amazing members of the **Comparative Bioacoustics group** in Nijmegen, thank you for making me feel part of something bigger and supporting me throughout the years. **Laura**, I am so grateful we became friends. Thank you for all your advice, the amazing dinners, and your incredible passion for brains. I really miss walking past your office every morning and bother you with random discussions. **Yannick**, you were not only my colleague but also my housemate. Thank you for your patience, enjoying all those special beers together, and helping me to further develop my programming skills. **Masha**, thank you for always listening to me and making me laugh with your witty humour and sarcasm. I am grateful for all the emotional support you have given me, especially during the final stretch—you are a true friend. **Diandra**, my vocal learning buddy, it was wonderful debating the ins and outs of this topic together, learning from you how to train seals to participate in behavioural experiments, and afterwards sharing Thai food in Kleve. **Silvia**, I cannot thank you enough for all your work in helping me collect data in Pieterburen and annotating all those recordings. **Sander**, it was a pleasure to supervise you, and I thank for your commitment in finishing up the annotations. **Taylor**, thank you for your wonderful energy and the engaging discussions on rhythm and cetaceans. **Christoforos**, your stay in the group was far too short, but I will never forget all the times we spent galivanting around Nijmegen. To all the other members, **Jelle**, **Lara**, **Anna**, **Marianne**, **Mary**, and **Kinga**, thank you for all the beautiful memories. To all the members of the **Artificial Intelligence Lab** in Brussels, **Yannick**, **Heikki**, **Katie**, **Peter**, **Rodrigo**, **Marnix**, **Mathilde**, and **Limor**, thank you for opening my eyes to world of computational modelling and giving me a very basic understanding of agent-based models and neural networks.

My PhD experience would not have been possible without the support of the **Max Planck Institute for Psycholinguistics**, which kindly hosted me for the duration of this project, covered travel expenses, and provided some of the conference funding. Special thanks to **Kevin** for always offering support and a listening ear. You were my MPI anchor and your insistence on making me follow the IMPRS doctoral training program allowed me to feel like I was a part of the community. Thank you to the people working in operations,

Laura, **Cedric**, **Marie-Luise**, **Lianne**, **Martina**, and **Paulina**, for facilitating all the administrative matters surrounding my stay at your beautiful institute. Finally, this list would not be complete without **Marjolein** and **Julia**, thank you for preparing me to face the world of media. It is thanks to your dedication and hard work that I could present myself well enough to be featured in a CBC documentary and have my work reach millions of people through outlets like Le Monde and National Geographic. Thank you also to the people working the reception for their daily kind welcome, and **Thea** for the numerous salads and sandwiches you made me for lunch.

The MPI is not only a beautiful institute with amazing staff and incredible facilities but constitutes a vibrant community of gifted researchers. Thank you to my fellow co-organizers of the IMPRS Conference 2022, **Laura**, **Sara**, **Cecilia**, **Giulio**, **Gökberk**, **Figen**, **Rong**, and **Kevin**—it was a great learning experience. Besides organising extracurricular activities, working at the MPI is a rewarding experience in itself thanks to the many people that work there. **Caitlin**, I will miss our catch-up coffees lounging on the couch in the institute hallways. **Ronny**, thank you for always letting me pop by your office to share and ask for the latest update. **Orhun**, thank you for indulging all my talk of seals and the interesting discussions about science, philosophy, and politics. **Laura**, thank you for your kindness and support, be it while creating a virtual conference environment in Gather.town or meeting up in the city. **Marlijn**, thank you for all the fun—I hope we continue running into each other during the Vierdaagsefeesten. Thank you to **Ine**, **Candice**, **Sophie**, **Naomi**, **Filiz**, **Alessio**, **Iza**, **Jenni**, and all my other fellow MPI colleagues, for the spontaneous conversations, banter, and afternoon drinks! Finally, thank you to **Limor** and the members of LEADS group, **Oxana**, **Remi**, **Danielle**, **Lukas**, **Loïs**, **Mathilde**, **Lewis**, for having me as an honorary lab member.

I am eternally grateful to the staff and volunteers of **Sealcentre Pieterburen** for supporting this work and creating such a dynamic workplace. People think that working with animals means one gets to cuddle with them all the time, but it means early mornings, late nights, sick animals, intensive care, heartbreak, and lots of cleaning—it is hard, emotional work and far from glamorous. And then these scientists show up wanting to run behavioural bioacoustics experiments… Thank you to the veterinarians, **Ana**, **Anna**, **Stella**, **András**, and all your interns, for helping us adapt our protocols to not interfere with your work and harbour the welfare of the seals in your care. Thank you to the Sealcare nurses, **Liselotte**, **Lisa**, **Marga**, **Bea**, **Vicky**, **Eva**, **Diana**, **Rosanna**, **Emmy**, **Nils**, and all the volunteers for your continued dedication to these animals and the interest you have shown in our work. Thank you to **Sander** for your help in communicating our results to the general public, be it by giving our work a permanent place in your exposition, by helping to organise the Weekend of Science, or by collaborating with the MPI to coordinate press releases. **Aline**, thank you for your beautiful digital illustrations of seal vocal tracts. Over a hundred harbour seal pups participated in my experiments but some names will forever stay with me, **Poffertje**, **Ella**, **Mojito**, **Jan**, **Bigos**, **Phocahontas**, thank you for your contributions.

I also want to thank the people outside of work that helped make my time in Nijmegen a positive experience and (in)directly contributed to this project. I want to begin by thanking **Caitlin** and **Ronny** for introducing me to Nijmegen and without whom the start of the PhD and the COVID pandemic experience would have been drastically different. **Limor**, where to begin. From our first meeting getting fined on a tram in Poland to becoming part of the same lab, working at the same institute, and now publishing papers together, I thank you for all the support, advice, vivid discussions, memories, and lifelong friendship. You are a truly an amazing person, and although we may be scientists, we are also idiots. To **Masha** and your wonderful wife **Dasha**, thank you for your unconditional support and the many dinners. To **Laura**, **Riccardo** and **Carlo**, thank you for all the fun activities, be it extracting seal brains, visiting zoos, or diving. Thank you to **Tzuf**, **Raz**, **Marla**, **Jorn**, **Laurens**, and **Simone** for all the laughs and good times that helped me disconnect from work. Thank you to my tennis friends, **Tim**, **Ana**, **Sigrid**, **Bram**, and **Eline** for allowing me to air my frustrations by hitting tennis balls across the court and forming the best competition team. Stap voor stap en punt voor punt!

I also want to thank my closest friends who unfortunately do not live in Nijmegen but are always there in spirit and whose friendship and support were key to getting this thesis across the finish line. To the Extended Consortio, **Remko**, **Robin**, **Ivi**, **Boy**, and **Anna**, I thank you for always rooting for me. I am grateful for your patience, care, and the occasional reality checks to help set me straight again. I love you and I look forward to our next holiday. **Bob**, thank you for always making light of a situation and the enthusiasm and pride you have shown for this work. To the Yo to the Lo, **Megan**, **Calum**, **Vass**, and **Morten**, I cannot believe that it’s been over 10 years since we met on exchange in Australia and still meet up several times a year. Thank you for your continued presence in my life and for putting up with me.

Anna, I could not have dreamed of a better life partner. We met when I was volunteering in Pieterburen and what a journey it has been since then. From working together with the seals, carrying out experiments together, publishing together, spending lockdown together, and now finally living together, you have shown me the utmost level of love, patience, and selflessness. Your ability to empathise with people, view situations from different perspectives, and capacity to communicate feelings in kind words have challenged and taught me to become a better person, scientist, and writer. From conceptualisation to data collection and analysis to writing, you have provided input at every stage of this work, and this thesis is practically as much yours as it is mine. I cannot express in words how much your unconditional support means to me. I love you.

Last but most definitely not least, I want to thank my family. **Mam** en **Pap**, dit was echt nooit mogelijk geweest zonder jullie. Het is niet in woorden te omschrijven wat jullie allemaal voor mij hebben gedaan in de loop der jaren, van het vieren van successen tot het eindeloos aanhoren van mijn frustraties. Dat betekent heel veel voor mij en ik ben jullie ontzettend dankbaar. **Stijn**, wat kan je lief en vervelend zijn zeg. Ik kan me nog herinneren dat je jezelf zo grappig vond toen je me wakker maakte met je zeehondengeluiden en ik zo boos was toen. En toch ben ik heel blij dat jouw kunst nu onderdeel is van mijn boekje. **Opa**, alhoewel je geen flauw benul hebt van de stof in dit proefschrift, behalve dat het over zeehonden gaat, weet ik dat je altijd heel trots op me bent. **Oma**, je was altijd mijn grootse fan en supporter en ik mis je verschrikkelijk. Dit proefschrift is voor jou.

Curriculum Vitae

Koen de Reus was born in Emmen, the Netherlands, in 1995. In 2016, he obtained a bachelor's degree in Liberal Arts and Sciences from University College Roosevelt in Middelburg, the Netherlands, majoring in Biology, Psychology, and Cognitive Science, with a minor in Statistics. In 2017, he completed a master's degree in Wild Animal Biology at the Royal Veterinary College in London, United Kingdom. His master's thesis was conducted at Sealcentre Pieterburen, where he studied vocal development in Eastern Atlantic harbour seal pups. In the following years, he worked part-time as a tutor in the Life Sciences Department of Erasmus University College in Rotterdam, the Netherlands, while continuing research related to his master's project. In September 2020, he started a joint PhD position with the Artificial Intelligence Lab of the Vrije Universiteit Brussel and the Donders Centre for Cognition. For the duration of this project, he was hosted by the Comparative Bioacoustics group at the Max Planck Institute for Psycholinguistics. This dissertation reports the findings of his PhD research. He is currently working as a lecturer in the Life Sciences Department of Erasmus University College.

Publications

de Reus, K., Benítez-Burraco, A., Hersh, T. A., Groot, N., Lambert, M. L., Slocombe, K. E., Vernes, S., & Raviv, L. (under review). Behavioral markers of domestication in vocal learning mammals. *Scientific Reports*.

Raimondi, T., Haas, C. E., **de Reus, K.**, Mendez-Arostegui, M., Jadoul, Y., & Ravignani, A. (in press). Neonatal auditory input affects vocal development in harbour seals. *Philosophical Transactions of the Royal Society B: Biological Sciences.*

Pleyer, M., Perlman, M., Lupyan, G., **de Reus, K.**, & Raviv, L. (2025). The 'design features' of language revisited. *Trends in Cognitive Sciences*. *Advance online publication*. doi:10.1016/j.tics.2025.10.004.

Rapado-Tamarit, B., Méndez-Aróstegui, M., **de Reus, K.**, Sarraude, T., Pen, I., & Groothuis, T. G. (2024). Age estimation and growth patterns in young harbor seals (*Phoca vitulina vitulina*) during rehabilitation. *Journal of Mammalogy*, gyae128. doi: 10.1093/jmammal/gyae128.

de Reus, K., Benítez-Burraco, A., Hersh, T. A., Groot, N., Lambert, M. L., Slocombe, K. E., Vernes, S. C., & Raviv, L. (2024). Self-domestication traits in vocal learning mammals. In J. Nölle, L. Raviv, K. E. Graham, S. Hartmann, Y. Jadoul, M. Josserand, T. Matzinger, K. Mudd, M. Pleyer, A. Slonimska, & S. Wacewicz (Eds.), *The Evolution of Language: Proceedings of the 15th International Conference (EVOLANG XV)* (pp. 105-108). Nijmegen: The Evolution of Language Conferences. doi:10.17617/2.3587960

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

de Reus, K., Carlson, D., Lowry, A., Gross, S., Garcia, M., Rubio-García, A., Salazar-Casals, A., & Ravignani, A. (2022). Body size predicts vocal tract size in a mammalian vocal learner. In A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, Y. Jadoul, M. Martins, Y. Oseki, E. D. Rodrigues, O. Vasileva, & S. Wacewicz (Eds.), *The evolution of language: Proceedings of the Joint Conference on Language Evolution (JCoLE)* (pp. 154-156). Nijmegen: Joint Conference on Language Evolution (JCoLE). doi:10.17617/2.3398549

Salazar-Casals, A., **de Reus, K.**, Greskewitz, N., Havermans, J., Geut, M., Villanueva, S., & Rubio-Garcia, A. (2022). Increased Incidence of Entanglements and Ingested Marine Debris in Dutch Seals from 2010 to 2020. *Oceans*, 3(3), 389-400. doi:10.3390/oceans3030026.

de Reus, K., Carlson, D., Lowry, A., Gross, S., Garcia, M., Rubio-Garcia, A., Salazar-Casals, A., & Ravignani, A. (2022). Vocal tract allometry in a mammalian vocal

learner. *Journal of Experimental Biology*, 225(8), jeb243766. doi:10.1242/jeb.243766.

Henry, M. J., Cook, P. F., **de Reus, K.**, Nityananda, V., Rouse, A. A., & Kotz, S. A. (2021). An ecological approach to measuring synchronization abilities across the animal kingdom. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 376(1835), 20200336. doi:10.1098/rstb.2020.0336.

de Reus, K., Soma, M., Anichini, M., Gamba, M., de Heer Kloots, M., Lense, M., Bruno, J. H., Trainor, L., & Ravignani, A. (2021). Rhythm in dyadic interactions. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 376(1835), 20200337. doi:10.1098/rstb.2020.0337.

de Heer Kloots, M., Carlson, D., Garcia, M., Kotz, S., Lowry, A., Poli-Nardi, L., **de Reus, K.**, Rubio-García, A., Sroka, M., Varola, M., & Ravignani, A. (2020). Rhythmic perception, production and interactivity in harbour and grey seals. In A. Ravignani, C. Barbieri, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, M. Martins, K. Mudd, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 13th International Conference (Evolang13)* (pp. 59-62). Nijmegen: The Evolution of Language Conferences. doi:10.17617/2.3190925.

de Reus, K., Carlson, D., Jadoul, Y., Lowry, A., Gross, S., Garcia, M., Salazar Casals, A., Rubio-García, A., Haas, C. E., De Boer, B., & Ravignani, A. (2020). Relationships between vocal ontogeny and vocal tract anatomy in harbour seals (*Phoca vitulina*). In A. Ravignani, C. Barbieri, M. Flaherty, Y. Jadoul, E. Lattenkamp, H. Little, M. Martins, K. Mudd, & T. Verhoef (Eds.), *The Evolution of Language: Proceedings of the 13th International Conference (Evolang13)* (pp. 63-66). Nijmegen: The Evolution of Language Conferences. doi:10.17617/2.3190925.

Ravignani, A., & **de Reus, K.** (2019). Modelling Animal Interactive Rhythms in Communication. *Evolutionary Bioinformatics*, 15, 1-14. doi:10.1177/1176934318823558.

Ravignani, A., Kello, C. T., **de Reus, K.**, Kotz, S. A., Dalla Bella, S., Mendez-Arostegui, M., Rapado-Tamarit, B., Rubio-Garcia, A., & de Boer, B. (2019). Ontogeny of vocal rhythms in harbor seal pups: An exploratory study. *Current Zoology*, 65(1), 107-120. doi:10.1093/cz/zoy055.

Ravignani, A., Garcia, M., Gross, S., **de Reus, K.**, Hoeksema, N., Rubio-Garcia, A., & de Boer, B. (2018). Pinnipeds have something to say about speech and rhythm. In C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, & T. Verhoef (Eds.), *Proceedings of the 12th International Conference on the Evolution of Language (EVOLANG XII)* (pp. 399-401). Toruń, Poland: NCU Press. doi:10.12775/3991-1.095.