

Special Issue: Time in the Brain

Review

The Evolution of Rhythm Processing

S.A. Kotz,^{1,5,*} A. Ravignani,^{2,3,5,*,@} and W.T. Fitch^{4,*}

Behavioral and brain rhythms in the millisecond-to-second range are central in human music, speech, and movement. A comparative approach can further our understanding of the evolution of rhythm processing by identifying behavioral and neural similarities and differences across cognitive domains and across animal species. We provide an overview of research into rhythm cognition in music, speech, and animal communication. Rhythm has received considerable attention within each individual field, but to date, little integration. This review article on rhythm processing incorporates and extends existing ideas on temporal processing in speech and music and offers suggestions about the neural, biological, and evolutionary bases of human abilities in these domains.

Commonalities Underlying Rhythm in Music, Speech, and Animal Communication?

Across all cultures in the world, humans synchronize to and move with musical rhythms. Similarly, we seem to neurally synchronize with rhythm in speech, which captures our attention, regularizes speech flow, may emphasize meaning, and facilitates interaction with others [1]. However, compared to music, rhythm in speech is more difficult to define and commonalities across the two domains remain elusive. Discrepancies between these domains begin with the fact that, while rhythmic behavior is often based on quasi-periodic repetition of steady intervals, for musicologists and linguists alike a simple periodicity-based definition is incomplete. One conceptual obstacle arises from conceiving human rhythmic behavior as a single monolithic entity rather than a multi-component phenomenon [2,3]. Clearly defining and empirically differentiating sub-components of rhythmic phenomena across domains (our first comparative task) will allow researchers to specify similarities and differences. We can then attempt to integrate insights from comparative animal work to help resolve the biological and evolutionary foundations of human rhythm cognition (our second comparative task).

This review article aims to compare and integrate conceptions of rhythmic behavior in music, speech, and animal communication. We start by isolating the multiple mechanisms underlying rhythm cognition, identifying four core sub-components. We then discuss how these components are manifested in different domains, identifying commonalities and differences. We next consider how rhythm at the signal level interacts with brain rhythms, and which neural systems support integration of the cognitive mechanisms underlying rhythm processing. Finally, adopting an evolutionary perspective, we reflect upon how comparative data concerning rhythm, although still sparse, can inform our understanding of the biology and evolution of rhythmic sub-components.

Multiple Sub-components Underlying Rhythm Cognition

A crucial first step in analyzing rhythmic behavior involves recognizing that human rhythmic behavior is not monolithic but comprises several different components, each potentially with its own biological basis and evolutionary history [3,4]. At least four separable components underlie human rhythmic behavior. These include periodic motor pattern generation itself, 'beat'

Highlights

Musical rhythm constitutes the sum of multiple constituent behavioral and neural features.

A comparative multi-component view on rhythm in music, speech, and animal communication reveals similarities and differences and may be key to understanding rhythm evolution.

Rhythm production and perception may be anchored in social synchronization across domains and species.

A wider comparative perspective, which incorporates insights from not only primates and birds but also cetaceans, pinnipeds, amphibians, and insects, can inform our understanding of rhythm evolution.

¹Faculty of Psychology and Neuroscience, Department of Neuropsychology and Psychopharmacology, Maastricht University, Universiteitssingel 40, P.O. Box 616, 6200 MD Maastricht, The Netherlands

²Research Department, Sealcentre Pieterburen, Hoofdstraat 94a, 9968 AG Pieterburen, The Netherlands

³Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

⁴Department of Cognitive Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

⁵Equal first authors

@Twitter: @AndreaRavignani

*Correspondence:
sonja.kotz@maastrichtuniversity.nl
(S.A. Kotz),
andrea.ravignani@gmail.com
(A. Ravignani),
tecumseh.fitch@univie.ac.at
(W.T. Fitch).

extraction from complex auditory patterns, and **entrainment** (see [Glossary](#)) of one's own motor output to this inferred beat. Finally, both music and speech typically arrange recurring patterns of events into hierarchical structures; this is termed meter perception.

Before defining these sub-components, we clarify some basic terms. At the heart of rhythmicity is the notion of **periodicity**: a somewhat regular recurrence of events that leads to cycles with a particular repetition rate (frequency) and a specific onset time within a cycle (phase). Periodic processes can be modeled as **oscillators** characterized by frequency and phase. Through entrainment, two or more oscillators can synchronize to match in frequency, phase, or both. **Isochronicity**, where an oscillator is strictly periodic, is often present in music but is not a fundamental requirement of rhythm or for entrainment. Two oscillators may vary in frequency but remain synchronized (just as two joggers running side by side can speed up and slow down in unison). Indeed, a central difference between speech and musical rhythm is that strict isochronicity is much more typical of music (for evolutionary and mathematical perspectives on isochronicity, see [5]). Thus, we should distinguish entrainment of quasi-periodic oscillators from those that are strictly isochronic (such as a metronome).

We now define our proposed sub-components. First, **motor periodicity** is simply the quasi-periodic execution of any repetitive action and is ubiquitous in biology including heartbeat, breathing, running, swimming, chewing, the wake/sleep cycle, and numerous other rhythmic activities. Any of these actions can be characterized by a time-varying frequency and phase, but they need not be synchronized. Second, **beat extraction** involves the perceptual inference of a pulse given a repetitive stimulus, often acoustic. This is a cognitive phenomenon that presumably involves the entrainment of endogenous neural oscillators with some external time-giver (the beat in [Figure 1](#)). Third, we can optionally synchronize motor actions with this inferred beat (as when tapping a finger to a metronome), a phenomenon we term **audiomotor entrainment** (potentially the most biologically unusual feature of human rhythmic behavior). The final element, **meter**, involves higher-level groupings of single events (or beats) into a hierarchical structure (cf. [2,6]) in which some events are stressed ('strong') and others are not ('weak'). Meter is a core element of speech rhythm and **metrical phonology** and also plays an important role in music ([Figure 1](#)). In its simplest case, meter may involve groupings of just two events (syllables or notes), where one event is strong and the other event is weak, but more complex groupings are possible.

We now discuss how these four sub-components manifest themselves in music and speech and explore the degree to which these components represent different manifestations of the same underlying processes, or instead rely upon distinct cognitive/neural mechanisms.

Rhythm in Music

Regular rhythm is a central component of human music. It seems likely that musical rhythm incorporates both primitive biological elements and unusual (apomorphic) elements, probably recently evolved [2,4]. Strictly regular or isochronic entrainment is a widespread peculiarity of musical rhythm, presumably driven by the need for group coherence and predictability in group music making and dancing. However, even in music strict isochronicity may be violated for expressive purposes (termed 'rubato') and may be absent in certain solo styles (e.g., vocal lament).

Three of the four sub-components of rhythm cognition are necessary for ensemble playing or dancing [4]: beat extraction, motor periodicity, and audiomotor entrainment. Beat extraction involves auditory processing only, motor periodicity requires complex motor control, and

Glossary

Allan factor: a hypothesis-free metric that aims to quantify the clustering of temporal events in a signal.

Amplitude envelope: the smoothed time-varying amplitude of a signal; subsumes duration, rhythm, tempo, and stress in the auditory signal.

Audiomotor entrainment: the production of periodic motor actions synchronized to a perceptual beat inferred from a (quasi)-periodic acoustic stimulus (also termed beat perception and synchronization).

Beat extraction: the cognitive process by which the frequency and phase of some external periodic signal are inferred by a listener.

Dynamic attending theory: a theoretical framework describing the presence and interaction of levels of attention to temporal periodicities in auditory signals.

Entrainment: the process by which two or more oscillators can become synchronized in frequency, phase, or both. The 'tick' and 'tock' of a clock have the same frequency but different phases. The hour and minute hand of a clock have different frequencies but coincide at 0° phase every 12 h.

Grouping: a series of events of either identical or non-identical physical properties perceived together, or perceptually clustered, often treated as alternating in specific acoustic features such as stress.

Isochronicity: highly regular periodicity, where the frequency is nearly constant.

Meter: hierarchical structuring of a series of events (which may or may not be strictly isochronous) into higher-order groupings. In music, indicated by the first number of a time signature (e.g., 3/4 indicates grouping of three quarter notes or crotchets, while 4/4 indicates groups of four).

Metrical phonology: the sub-discipline of linguistics that studies how units such as syllables are perceptually grouped into patterns of prominence, where some units are stronger and others weaker (termed stressed and unstressed, respectively). The auditory cues that can underlie such stress groupings are diverse and include loudness, duration, and pitch.

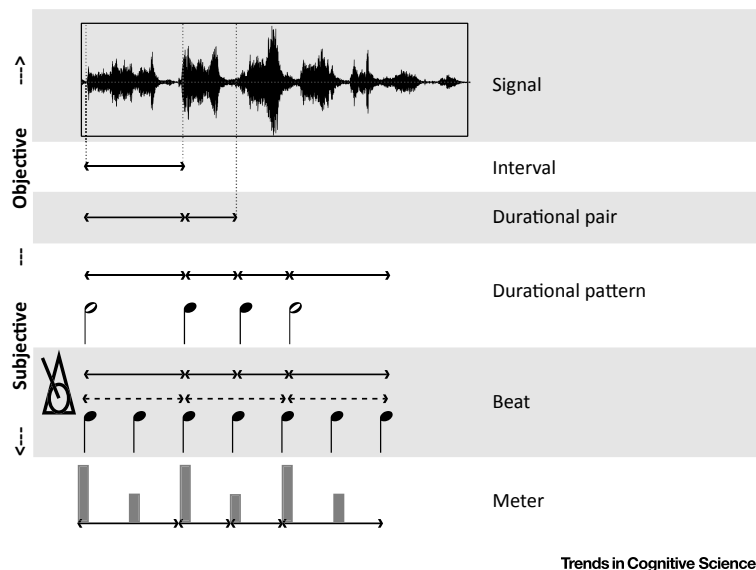


Figure 1. Levels of Temporal and Rhythmic Structuring. Each level shows a graded transition from signal-based, more objective features of rhythm (top) to more subjective, cognitive properties of rhythmic patterns (bottom). Spectral properties of a signal (e.g., intensity peaks) induce acoustic event perception, leading to temporal intervals encompassing two events. The first generalization of an interval is a durational pair and then a durational pattern. This is the transition between timing and rhythm: each individual temporal interval is processed differently depending on its adjacent intervals. A beat can then be superimposed to a durational pattern that can be perceived as having a metrical structure (meter).

Motor periodicity: any recurrent action that has an identifiable frequency and phase that may, however, vary over time; examples include walking, running, chewing, or scratching.

Oscillator: an abstract periodic entity characterized by a rate of repetition (frequency) and a specific moment within a cycle (phase). Oscillators may be stronger or weaker (amplitude), but this is not crucial in the context of the current review.

Periodicity: the quality of having a steady pattern of repetition in time (having a fixed frequency and phase).

Quasi-periodicity: the term applied to oscillators that are not strictly periodic (having a fixed frequency) but are nonetheless stable enough for frequency analysis over short time windows; typical for biological oscillators.

audiomotor entrainment requires the sensory-guided motor synchronization of perceived beat and produced movement.

Even beat (or ‘tactus’) extraction [7] can be challenging: a listener often must infer a beat that is not directly present in the acoustic signal. We typically do not dance to metronomes but rather to complex patterns from which an isochronic beat must be inferred despite both excess events (events not ‘on the beat’) and, in the case of syncopated rhythms, ‘silent events’ (see [8]). This is already non-trivial, although the spectrum of highly regular music (e.g., contemporary dance music) typically has substantial energy at the beat frequency, allowing bottom-up beat extraction by computer algorithms [9].

A substantial literature on the neural basis of synchronization to music [10,11] indicates that rhythm perception typically involves brain regions traditionally associated with motor output and production (see Figure 1 in Box 1). This consistent result strongly suggests important roles for motor regions including both cortical and subcortical areas (e.g., basal ganglia and cerebellum [12]) in beat extraction.

The next step is to entrain one’s own motor output to the extracted beat. If this output is simply tapping one’s finger, this could be as basic as periodically tensing a single muscle to the inferred beat: simple audiomotor entrainment. A rich experimental literature, mostly involving finger tapping, documents robust human abilities to extract a beat from complex auditory stimuli and entrain motion to it [13,14]. This body of research suggests that humans are more inclined to entrain to auditory than visual stimuli [15], while the opposite may be true of non-human primates [16].

Box 1. Cortico-Subcortical Circuitry of Time and Rhythm Processing

The cortical neural circuitry underlying rhythm in music is well described [125] and encompasses areas that mostly overlap with those reported for the production and perception of temporal structure in auditory signals in general [126]. Next to dorsolateral prefrontal, anterior cingulate, insula, supplementary motor (SMA), and temporal cortices, thalamus, basal ganglia, cerebellum, and inferior parietal cortex (IPC) play important roles in timing and rhythm processing (Figure 1). Describing how we produce and perceive a sequence of sounds in time may explain why such a sophisticated cortico-subcortical network subserves both temporal and rhythm processing across domains.

To produce a sequence of events, the prefrontal cortex (PFC) must retrieve memory representations of events (temporal cortex) and activate the respective action plans (SMA proper, motor cortex). It also recruits the slow modulating dynamics inherent to the interplay of the pre-SMA, basal ganglia, and thalamus to initiate action and to establish sequential relations among events (temporal structure). To implement a sequence of events, connections to SMA proper and motor cortices are recruited.

In the perception of an event sequence, an event-based representation of temporal structure is relayed to the thalamus that is then passed on to the pre-SMA and calls for attention (PFC). If attention is oriented towards a sequence of events, the event-based representation can elicit dynamic attending and similar attention-dependent temporal processing oscillations. Attention-dependent temporal processing encodes temporal inter-event relations (intervals) and conveys this information to PFC. Information about the temporal structure of a sequence of events may then be stored, and successive intervals compared, to extract larger patterns. However, regular temporal structure not only allows pattern extraction but also facilitates the stimulus-driven maintenance of attention (IPC), thereby reducing sensitivity to error in the whole system and stabilizing the perceptual process. This may rely on attention-modulating areas such as the inferior parietal lobule and anterior cingulate gyrus.

Consequently, rhythm and temporal processing may interact not only to control the encoding of spatially distributed memory representations into a temporal sequence (production) but also to optimize the processing of dynamic sequential signals during rhythm perception.

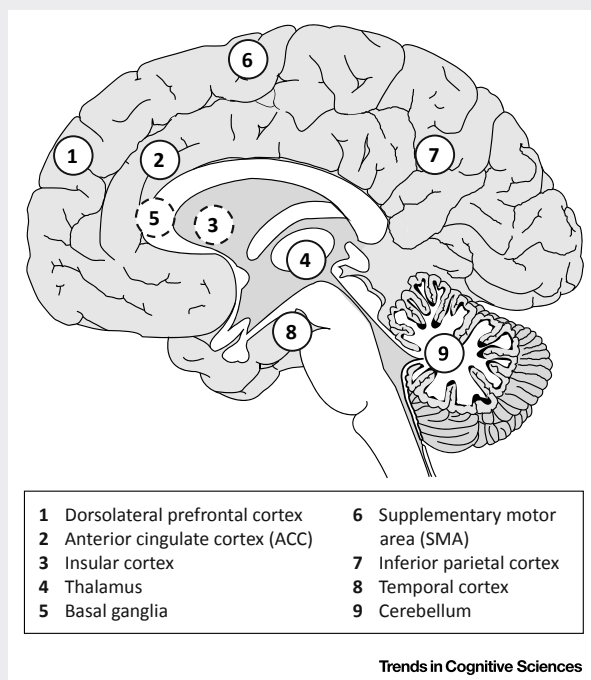


Figure 1. Cortico-Subcortical Brain Network for Time and Rhythm Processing. Medial sagittal view of engaged brain areas.

However, in contrast to finger tapping, for musicians or dancers, entrainment typically involves basing more complex movements on the pulse extracted from a complex auditory stimulus. Typically, musicians performing different parts make different complex movements to a shared inferred pulse, and dancers may well produce movements that have no corresponding event in the acoustic stream. Literature on the cognitive and neural mechanisms underlying these more complex movements in dance is still in its infancy [17–19].

The final element of musical rhythm, **meter**, involves higher-level groupings of single events (or beats) into a hierarchical structure (cf. [2,6]). Musical meter may involve groupings of just two events (e.g., the 2/4 beat typifying Brazilian samba), but groupings of three (waltz time) or four (4/4 or ‘common’ time) are more typical. In music, the first beat of a grouping is always considered strong (although a musical phrase may begin with one or more weak ‘pick up’ notes).

Although metrical aspects of rhythm perception have received less neuroscientific attention than isochronicity, meter perception appears to involve electroencephalographic oscillations at higher metrical levels, even when the stimulus stream lacks differentiation of strong and weak events [20–22]. Furthermore, musicians appear to engage a left hemisphere network in identifying metrical deviations, while non-musicians show right hemisphere biases [23], suggesting that musical training might influence the degree to which musical rhythms are processed and categorized at an abstract level.

Rhythm in Speech

For linguists (specifically metrical phonologists) and poets, ‘speech rhythm’ has connotations that overlap only partly with those of musicologists. In particular, unlike music, speech events are rarely isochronous; however, the hierarchical pattern-processing aspects of speech and musical rhythm overlap considerably.

Multiple properties underpin the temporal structure of speech [24,25], including envelope, periodicity, and acoustic fine structure [26]. The **amplitude envelope** in particular captures information about duration, rhythm, tempo, and stress and, when degraded, renders the speech signal unintelligible [27]. Broadly speaking, speech rhythm involves the systematic patterning of acoustic events encompassing timing, accent, and grouping [28]. Specifically, researchers distinguish between coordinative rhythm, where units are grouped by phonetic cues, and contrastive rhythm specifying the alteration of strong and weak speech events [29]. The latter type of rhythm aligns well with metrical structure in music, where the alteration of strong and weak beats forms the basis of the metrical hierarchy.

A much-debated issue in speech rhythm research concerns its ‘periodicity’, that is, whether or not speech units occur in a temporally periodic manner [25,30]. There is broad agreement that a strict notion of isochronic periodicity seems unsustainable and artificial in speech [31], where frequency is highly variable. Listeners may intuitively perceive isochrony even when the speech signal is at best quasi-periodic [32]. The isochrony typical of music, and its relative absence in speech, represents the most obvious difference between these two domains. However, there is no clear dividing line, since speech rate can vary from very free in conversational speech, to more periodic in poetry, infant-directed speech, and political speeches, to nearly isochronous in rap, chanting, or sung speech [33–35]. Thus, there is a continuum from strict isochrony to very variable **quasi-periodicity**, with different styles occupying different ranges of this gradient.

Turning to meter, there is broad agreement that alterations of stronger and weaker speech events (such as stressed and unstressed syllables) give rise to groupings forming a metrical grid [36]. The segmentation of speech signals benefits from these perceptible stress patterns, both in language acquisition and in word recognition [37]. Furthermore, anticipation of the next stressed syllable in online speech processing appears to direct attention to salient events and facilitate speech comprehension [1,24,37]. Stress patterns can also differentiate lexical items (e.g., the English noun 'REbel' versus the verb 'to reBEL'), playing a role in semantics. Finally, speakers vary their speech rhythm depending on speaking style (e.g., systematically using pauses and accentuations) and speech rate (faster or slower) to influence how they are understood. Thus, in addition to aiding comprehension and reducing processing costs, speech rhythm may influence how we communicate in social interactions (see Figure 1 in Box 2). In sum, speech rhythm forms a core communicative component of spoken language.

What cognitive and neural mechanisms underlie the extraction of rhythmic patterns in speech (Box 1)? The data on metrical perception reviewed above suggest that listeners perceive and entrain to accented events in music and speech alike. When salient events are embedded in

Box 2. Interpersonal Coupled Dynamic Attention

Speech rhythm may both facilitate speech segmentation and comprehension and also influence how we communicate with conspecifics. Does neural entrainment to speech rhythms also apply across the brains of speakers and listeners? In other words, do specific frequencies not only couple with each other (e.g., theta-delta-gamma) in speech perception within one brain but also between a speaker and a listener when they share information through the synchronization of produced and perceived speech frequencies (Figure 1, oscillations 1, 2, and 3)? Do we find evidence in the evolution of rhythm that could help evaluate this suggestion?

Returning to the idea that theta frequency tracks the aperture of articulatory gestures [40,127] and that the vocalic nucleus of the syllable carries the beat in the same frequency range [128], the theta rate of syllables (2–8 Hz) is a core temporal variable in communication. The frequency of the articulatory gesture, a multimodal (visual and auditory) signal produced by the speaker, may couple with the perception of the vocalic nucleus of the syllable by the listener, thereby establishing a shared periodic framework between them. When we speak, orofacial movements carry sound and visual information, with both cycling at a theta rhythm [129]. Seeing a mouth move and perceiving salient events in speech may be central, ancestrally, to using speech rhythm in communication. Comparative work in monkeys [115] has led to the hypothesis that rhythmic facial expressions such as lip-smacks may be the evolutionary precursors of rhythm in speech [130], based on their occurrence at the frequency of syllable production. Furthermore, macaque monkeys can use rhythmic actions (cage rattling) as a communicative signal [131]. These signals also consist of approximately 5 beats per second, reminiscent of natural syllable production in speech. The drumming frequency also overlaps with the frequency of macaque vocalization potentially creating a basic rate for communicative action-perception cycle in primates [131].

In addition to frequency modulations in the theta range (2), speech also carries both faster modulations [>8 Hz; intrinsic sound characteristics (3)] and slower modulations [<2 Hz; (1)], which include the melodic pitch contour of speech and phrasal structure. Future human and comparative work needs to further explore the role of delta and theta coupling, since coupling of slower (delta) and medium (theta) frequencies may underlie how salient events unfold over time, bind our attention, and facilitate not only speech comprehension but also communication [132].

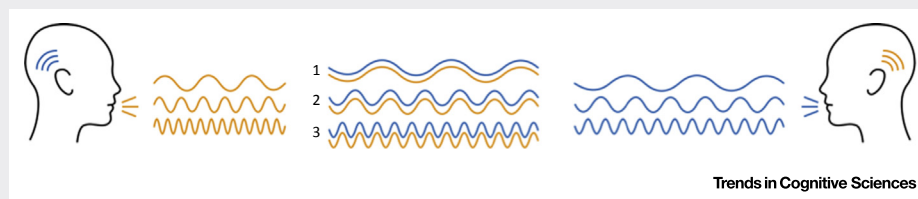


Figure 1. Synchronization of Speech Frequencies between Speaker and Listener. Even if the speaking rate of the respective speaker/listener is faster (left) or slower (right), synchronization of at least three neural frequencies underlying speech rhythm may occur.

metrical grids, this leads to improved neural entrainment where increased regularity may lead to improved content processing in both music and language.

A seminal cognitive theory on rhythmic entrainment is **dynamic attending theory** [38], which posits stimulus-driven periodic allocation of attention in time, arising from the coupling of endogenous (brain-based) oscillators with external rhythmicity in an auditory signal. Although periodicity alone may be adequate for entrainment, the metrical grid referred to above can provide further predictive structure in speech. The syllabic nucleus constitutes a salient feature in speech as it aligns with maximal oral opening during articulation [39,40]. The resulting spectro-temporal increase in signal duration and intensity marks a salient syllable [41]. In turn, hierarchically organized meter-based alternations of stressed and unstressed syllables are thought to entrain an internal oscillator to these properties of the speech signal, leading to improved speech comprehension [24].

At the neural level, entrainment to rhythmic stimuli such as the beat in music or quasi-periodic streams of syllables in speech is hypothesized to involve phase coupling of oscillations in the theta range (3–8 Hz, 150–300-ms period) with gamma-range oscillations around 40 Hz (25–50-ms period), directing attention to the rhythmic properties of a stimulus [42,43]. It has been suggested that such coupling of neural oscillations may provide the underlying neural basis of dynamic attending [44]. While physiological evidence of entrainment in speech has not focused on the metrical grid *per se*, magneto/electroencephalography speech frequency tagging studies suggest that the brain entrains to properties at the rate of phoneme [45] and syllable frequencies [46–49]. Furthermore, a recent magnetoencephalography study showed that larger linguistic units (e.g., phrase boundaries) can also be hierarchically entrained at the slower delta level around 1–2 Hz [50]. Finally, research utilizing event-related potentials (ERPs) reported that regular meter in speech enhances ERP responses, relative to irregular meter. This was shown for phoneme perception, semantic and syntactic comprehension, and again entrainment correlates with improved speech comprehension [51–54].

In sum, there is abundant evidence that when neural rhythms can synchronize to quasi-periodic acoustic input, this leads to increased processing ease and facilitates speech comprehension. The fact that neural coupling to the acoustic stream occurs in both music and speech implies that shared mechanisms might underlie this process.

With this first comparison completed, we now turn to the second comparative question of how these cognitive and neural mechanisms may have arisen in an evolutionary context by investigating rhythm cognition in non-human animals.

Animal Rhythmicity: A Comparative Approach

Lacking a time machine to directly witness early human evolution, a comparative approach is key to understanding how rhythm processing evolved in our species [4,55–59]. Comparisons within humans between cognitive domains, individuals, cultures, and developmental stages are illuminating [60–64], but cross-species comparisons can highlight which behavioral features and neural systems are crucial for a species to develop rhythmic competence and can help reconstruct both when and why these features or systems evolved. The presence or absence of features of rhythmic competence in other animals can be mapped to evolutionary gains (or losses) of the sub-components underlying modern human rhythmicity. Evidence from species closely related to humans (e.g., non-human primates) can elucidate recently emerged traits and common ancestry, while convergent evolution of rhythmic traits in distant species (e.g., birds and crickets) allows analysis of the selective pressures that drive the evolution of rhythmic traits.

For instance, beat extraction and audiomotor entrainment, also termed ‘beat perception and synchronization’ (BPS) [6], to artificial stimuli, such as music, appear to be sparsely distributed in non-human species, being found in a small subset of bird and mammal species. By contrast, the capacity to entrain vocalizations or movements to conspecific displays is more widespread among both vertebrates (including birds and frogs) and invertebrates (e.g., crickets and fireflies). This distribution suggests that basic audiomotor entrainment has evolved multiple times, allowing us to test adaptive hypotheses about the origin of rhythmicity (reviewed in [56]).

Field-based investigations of rhythmic structures in the vocalizations, visual displays, or movements of animals in their natural environment, along with experimental methods used in the laboratory, have been used to study animal capacities to produce, perceive, and entrain to rhythmic patterns [65–70]. Research in field and/or laboratory settings has investigated key sub-components of rhythm such as audiomotor entrainment in primates, other mammals, birds, frogs, and insects [71–80].

Spontaneous rhythmic behaviors are found in humans’ closest living relatives, the African great apes. In particular, chimpanzees perform displays, known as ‘buttress drumming’ [81,82], beating resonant trees in the wild or other resonant objects if housed in captivity [83,84], using their hands and/or feet to produce sequences of percussive sounds. The precise rhythmic structure of these sequences has received little investigation [85]. Equally little is known about the temporal structure of bonobos’ ‘staccato hooting’, an apparently rare periodic vocalization sequence reported in one paper [86].

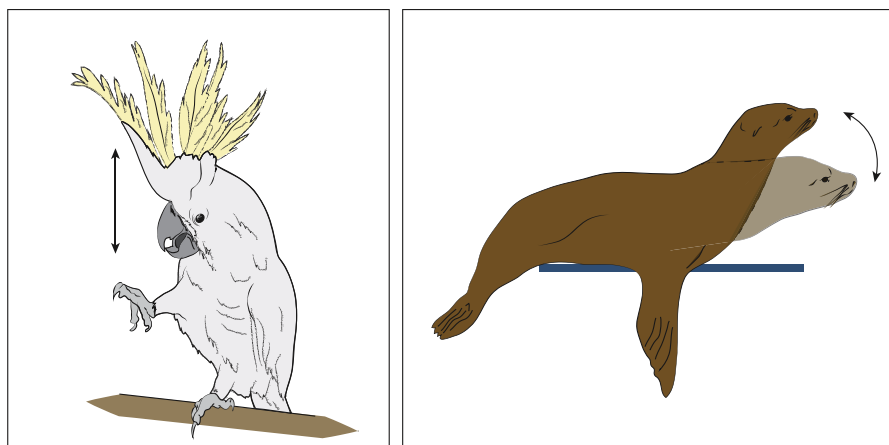
Laboratory work in animals has focused on both motor entrainment to acoustic input (typically music) and perceptual grouping of events in time, relevant to meter. Spontaneous production of roughly isochronous sequences has been reported in zebra finches [87], a popular laboratory species in birdsong research [88–90]. Zebra finches combine song elements (‘syllables’) into complex vocal displays that are temporally structured so that syllables occur at specific time points. Recordings of the onset of notes in bird song can be aligned with an isochronous click sequence [87].

Several parrot species and a sea lion have convincingly been shown capable of audiomotor entrainment experimentally [65–67,91]. A study based on YouTube videos found widespread evidence of audiomotor entrainment to music among multiple parrot species, and in one elephant, but no evidence in dogs, even in dog shows where music played a prominent role [92]. A recent study also indicates no spontaneous entrainment to music by show horses [93]. Although several studies have argued for a modicum of audiomotor entrainment to music by chimpanzees and bonobos [73,77,78], this appears quite limited relative to parrots or sea lions, restricted both to specific individual animals and to preferred motor tempos. Ronan (a California sea lion) provides an excellent example of experimental BPS in a non-human animal [67,94]. Trained by reinforcement to extract a metronomic beat and entrain to it, Ronan then generalized this behavior to novel tempi and to complex musical rhythms. Other than a relatively long training period, Ronan’s fine-grained timed behavior is highly reminiscent of humans’ tapping to music [67,91,94]. In summary, audiomotor entrainment is unusual but nonetheless found in a select group of species, and it is by no means uniquely human (see Figure 1 in Box 3).

Turning to meter, basic perceptual grouping has been observed in multiple animal species, for example, rats, pigeons, zebra finches, and budgerigars. These animals could be trained to rely on alternations of intensity, duration, or pitch to segment an incoming signal into binary sets [95–99]. Crucially, however, no animal has yet been shown to perceive true musical meter, interpreted as hierarchical perceptual grouping of an isochronous sound stream. Limited

Box 3. Audiomotor Entrainment and Vocal Learning

It has been hypothesized that audiomotor entrainment may be mechanistically linked to vocal learning, that is, that a species with a neural circuitry needed for vocal auditory/motor imitation may also be capable auditory/motor entrainment [3]. However, the lack of motor entrainment in multiple vocal learning species such as songbirds suggests that vocal learning is at best necessary, but not sufficient [19]. However, two other taxonomic groups, cetaceans and pinnipeds, may provide further insight [58,133–135], because many aquatic mammals are capable of vocal production learning [3,58,136]. In unusual circumstances, both orcas and harbor seals can learn to imitate human speech sounds [137,138]. Surprisingly little work has been done on pinniped or cetacean rhythmicity to date [139,140]. Some pinniped species show vocal production learning (VPL) [134], and multiple whale and seal species perform underwater displays reminiscent in form and function of songbirds' songs [58,69,111,141,142]. However, and partly contradicting the 'vocal learning' hypothesis, Ronan the sea lion is able to entrain but belongs to a pinniped species with little or no documented capacities for vocal production learning (Figure 1) [19,58,136].



Trends in Cognitive Sciences

Figure 1. Examples of Audiomotor Entrainment in Animals. Snowball the cockatoo (left) and Ronan the sea lion (right) are both capable of entraining to a musical stimulus and varying their head movements as the tempo of music is changed experimentally [6,66,67,91].

neurophysiological evidence seems to support a beat/meter distinction: the mismatch negativity in event-related potential research found evidence for auditory **grouping** in macaques but no evidence for meter or hierarchical beat perception [100–104]. Macaques, and possibly other non-human primates, may thus perceive sequences of isochronous intervals in a rather different way than humans [105–109] (cf. [110]).

The Evolution of Rhythm Cognition

We end by formulating some tentative hypotheses about the evolution of the different sub-components of human rhythmicity. First and most obvious, the capacity to produce periodic motion is nearly universal among animals, indicating that motor periodicity is a basic vertebrate capacity. In contrast, beating periodically on resonant objects to make sounds (drumming) is an unusual trait, shared by a far-flung group of rodents and birds, and a few apes. Among primates it is present in humans and the African great apes (chimpanzees and gorillas) [111], suggesting that this percussive precursor of instrumental music had evolved in our last common ancestor, roughly 10 million years ago.

The capacity for perceptual beat extraction also seems to be common, at least among birds and mammals, and where it has been studied in nonhuman primates, seems to rely on similar neural mechanisms [11,102,112]. Crucially, macaques also show neural oscillations in the theta

Box 4. Rhythm in Interaction: Chorusing and Turn-Taking

Rhythmic behavior does not exist in a vacuum; it is typically embedded in a social context [143,144]. Even the rhythmic structure of ‘animal monologue’, such as those performed by songbirds, is shaped by their functional constraints and necessities to convey information (Figure 1A); the partially isochronous structure in zebra finches’ songs is a good example [87]. Another component of rhythmic behavior in social contexts emerges at shorter timescales and concerns interactive rhythms (Figure 1B): how does the signal timing of one individual’s output influence the signal timing of a conspecific?

Two strands of research have investigated temporal properties of animal interactivity with comparative focus: chorusing and turn-taking. Interestingly, chorusing work has often attracted comparisons to music [56], with animal turn-taking being related to human speech and language [145] (see Box 2). There is a strong overlap between what researchers in animal chorusing (Figure 1B) and turn-taking (Figure 1C) investigate, less so between the theoretical framing they provide. The comparative study of turn-taking focuses on four elements of human conversation: (TT1) cognitive flexibility, (TT2) individual roles, (TT3) absolute time delay between offset of a call and onset of the next, and (TT4) relative function of two adjacent turns [145]. The comparative study of animal chorusing is instead concerned with: (CO1) temporal flexibility, (CO2) emergent coordination patterns across multiple individuals, (CO3) relative phase of individual calls with respect to the calls of other individuals, and (CO4) evolutionary function(s) of interactive timing [56]. Although these foci are different in detail, they are clearly conceptually related.

One possible way to enrich the dialogue between disciplines studying animal interactive timing would involve bottom-up, hypotheses-free methods [68,70,146]. We suggest that the structure of animal interactive timing can first be quantified, and then similarities be drawn to speech or music rather than seeking specific music-related or speech-related features in animal timing. For instance, the temporal structure of vocalizations can be compared across species and domains; crucially, these rhythmic similarities seem to derive from signals’ (i) degree of interactivity (e.g., a solo versus a duet) and (ii) functional role (e.g., displaying learning plasticity versus enhancing pair bonding), more than phylogenetic relatedness [68,70]. If quantified via the Allan factor, the hierarchical temporal structure of animal duets is closer to interactive jazz improvisation, and a songbird ‘solo’ is closer to a human spoken monolog, than these are to other vocalizations in the same species [68].

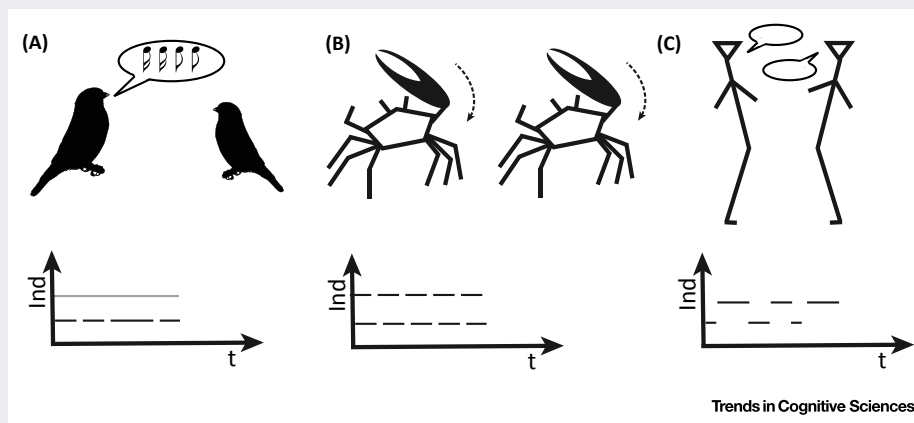


Figure 1. Different Types of Social Timing. (A) Temporal structure of a songbird ‘solo’ (dashed black line in the graph below the birds), while another bird is listening (gray line in the graph) and potentially evaluating it as a suitable mate. (B) Synchronous waving of fiddler crabs (animal choruses are not limited to acoustics!). (C) Timing of human conversational turn-taking. Individual activity (ind) on the vertical axes, and time (t) on the horizontal axes. Figure partly adapted from [59].

and gamma range that seem to play a similar role in dynamic attention in macaques and humans (e.g., [113]), suggesting that the basic machinery for neural entrainment was present in our common ancestor (~25 million years ago), if not before. However, current data suggest that this neural entrainment encompasses meter in human infants, but not non-human primates, suggesting that metrical hierarchical grouping of events in time may be a derived feature specific to human rhythmicity [101,114].

A fascinating hypothesis concerning the evolution of rhythmicity in speech starts with the observation that certain primate facial displays, such as lip-smacking, occur in the same theta frequency range as human speech [115,116]. These displays are typically, but not always, nearly silent [117] and consist of complex, synchronized movements of the lips, jaw, and tongue that are highly similar to speech movements [118]. Human data show that neural entrainment to visual components of speech enhances auditory perception, but only in the speech-typical theta frequency range, suggesting that the origin of speech rhythmicity may lie in pre-existing perceptual neural oscillations, to which lip-smacking, and later speech movements, became 'tuned' during evolution [119] (Box 2).

Motor entrainment to an auditory beat is a pervasive feature of human musical rhythmicity, shared based on current experimental data with parrots and at least one pinniped species. It is apparently lacking in several mammal species (dogs and horses), despite intensive training and exposure to music. There is currently no solid evidence for audiomotor entrainment to music in many other bird and mammal species, and evidence for such entrainment in non-human primates remains tenuous. This suggests that the capacity for audiomotor entrainment requires more than just motor periodicity and perceptual beat extraction. Consistent with this, human newborns perceive a beat, and produce periodic movements, but do not synchronize the extracted beat with motor output until after age 2 years [120]. Nonetheless, audiomotor entrainment in the form of conspecific display synchronization occurs in multiple species of insects and frogs, suggesting both that this has evolved convergently several times and that the neural circuitry allowing entrainment of auditory oscillators with motor planning oscillators is neither particularly complex nor difficult to evolve.

Finally, although the affective and functional role of audiomotor entrainment in other species remains unstudied, it is clear in humans that engagement in rhythmic group activity is a human universal [60] and can lead to enhanced social bonding [121–123]. So, both the mechanisms allowing human audiomotor entrainment, and its affective, interactive, and adaptive basis, remain fascinating open questions (see Figure 1 in Box 4).

Concluding Remarks and Future Directions

Recognizing that rhythm is not a monolithic concept and that sub-components of rhythm may involve different cognitive and neural mechanisms with different evolutionary histories helps unify divergent conceptions of rhythm in music and speech. Comparing facets of rhythm across cognitive domains, it is clear that isochronicity plays a stronger role in music than in speech. However, the degree of isochronicity represents a continuum [5], with multiple styles of speech (e.g., poetry, chant, or rap) showing strong periodicity and a few musical styles little [124]. Turning to meter, music and speech share clear common ground at the metrical level, possibly due to shared neural mechanisms for building up hierarchical structure from auditory sequences (Box 1).

Interestingly, at both neural and cognitive levels, current data suggest more overlap than separation for facets of rhythm across cognitive domains, suggesting that shared resources may be deployed in the perception of any signal with some regular temporal structure. Furthermore, the existence of non-human species sharing key components of human rhythm cognition offers the potential for important insights into the neural and evolutionary basis of rhythm perception (see Outstanding Questions). More comparative work is needed to elucidate which rhythmic components engage which neural systems and to clarify the neurobiological prerequisites for different facets of rhythmic behavior. Lastly, rhythm in humans has a strong affective component: sharing rhythmic behaviors such as singing, dancing, chanting, or talking

Outstanding Questions

Which genetic networks underlie the neural mechanisms for musical and speech rhythm perception and production in humans?

Are features of musical rhythm expectations of speech rhythm (or vice versa)?

Is the study of drummed and whistled speech a way of getting to the core of rhythm across domains and species?

Which socio-ecological factors put selective pressures on the evolution of specific rhythmic sub-components?

Does the development of beat perception and synchronization piggyback on purely physical isochronicity?

What are the cross-species differences in beat perception and synchronization abilities when comparing vocal and non-vocal learners?

What is the structure of spontaneous rhythms in animals (especially mammals), which brain circuits does it rely upon, and how is it affected by social interaction?

Why do some animal species spontaneously synchronize with conspecifics while other species entrain to artificial periodic stimuli (e.g., Ronan the sea lion)?

Do non-human animals spontaneously group identical isochronous stimuli into binary strong-weak pairs (the 'tick-tock' effect in humans) in perception?

Are there differences between trained and spontaneous rhythmic behaviors in animals?

How do rhythmic skills develop over an animal's lifespan?

Which techniques from time series and music information retrieval are best suited to unveil rhythmic patterns in animal vocalizations?

together can increase social bonding (Box 2). This aspect of rhythm remains unexplored in non-human species at present. Nonetheless, human rhythm production and perception appear to build upon more basic timing and sequencing capacities shared with many non-human animals. Understanding these mechanistically will be crucial to further progress in understanding how rhythm cognition evolved in our own, highly rhythmic, species.

Acknowledgments

The authors thank Michael Schwartz, two anonymous reviewers, and the editor for helpful comments and suggestions on the manuscript and Nadja Kavcic-Graumann and Andrea Gast-Sandmann for graphical support. S.A.K. was supported by the Bial Foundation and the Biotechnology and Biological Sciences Research Council (BB/M009742/1). A.R. was supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 665501 with the research Foundation Flanders (FWO) (Pegasus² Marie Curie fellowship 12N5517N awarded to A.R.). W.T.F. was supported by Austrian Science Fund (FWF) DK Grant 'Cognition & Communication' (W1262-B29).

References

- Kotz, S.A. and Schwartz, M. (2016) Motor-timing and sequencing in speech production: a general-purpose framework. In *Neurobiology of Language* (Hickok, G. and Small, S. L., eds), pp. 717–724, Academic Press
- Fitch, W.T. (2013) Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7, 68
- Patel, A.D. (2006) Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.* 24, 99–104
- Fitch, W.T. (2012) The biology and evolution of rhythm: unravelling a paradox. In *Language and Music as Cognitive Systems* (Rebuschat, P. et al., eds), pp. 73–95, Oxford Scholarship Online
- Ravignani, A. and Madison, G. (2017) The paradox of isochrony in the evolution of human rhythm. *Front. Psychol.* 8, 1820
- Fitch, W.T. (2015) The biology and evolution of musical rhythm: an update. In *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff* (Toivonen, D.I. et al., eds), pp. 293–324, MIT Press
- Lerdahl, F. and Jackendoff, R. (1983) An overview of hierarchical structure in music. *Music Percept.* 1, 229–252
- Fitch, W.T. and Rosenfeld, A.J. (2007) Perception and production of syncopated rhythms. *Music Percept.* 25, 43–58
- Scheirer, E.D. (1998) Tempo and beat analysis of acoustic musical signals. *J. Acoust. Soc. Am.* 103, 588–601
- Grahn, J.A. and Brett, M. (2007) Rhythm and beat perception in motor areas of the brain. *J. Cogn. Neurosci.* 19, 893–906
- Merchant, H. et al. (2015) Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. B* 370, 20140093
- Nozaradan, S. et al. (2017) Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex* 95, 156–168
- Repp, B.H. (2005) Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bull. Rev.* 12, 969–992
- Repp, B.H. and Su, Y.-H. (2013) Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. Bull. Rev.* 20, 403–452
- Repp, B.H. and Penel, A. (2004) Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol. Res.* 68, 252–270
- Gómez, J. et al. (2018) Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Ann. N. Y. Acad. Sci.* Published online April 30, 2018. <http://dx.doi.org/10.1111/nyas>
- Brown, S. et al. (2006) The neural basis of human dance. *Cereb. Cortex* 16, 1157–1167
- Fitch, W.T. (2016) Music, dance, meter and groove: a forgotten partnership. *Front. Hum. Neurosci.* 10, 64
- Ravignani, A. and Cook, P. (2016) The evolutionary biology of dance without frills. *Curr. Biol.* 26, R878–R879
- Abecasis, D. et al. (2005) Differential brain response to metrical accents in isochronous auditory sequences. *Music Percept.* 22, 549–562
- Nozaradan, S. et al. (2011) Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240
- Nozaradan, S. et al. (2012) Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32, 17572–17581
- Vuust, P. et al. (2005) To musicians, the message is in the meter: pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* 24, 560–564
- Kotz, S.A. and Schwartz, M. (2010) Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399
- Jadoul, Y. et al. (2016) Seeking temporal predictability in speech: comparing statistical approaches on 18 world languages. *Front. Hum. Neurosci.* 10, 586
- Rosen, S. (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. B* 336, 367–373
- Shannon, R.V. et al. (1995) Speech recognition with primarily temporal cues. *Science* 270, 303–304
- Patel, A.D. (2010) *Music, Language, and the Brain*, Oxford University Press
- Nolan, F. and Jeon, H.-S. (2014) Speech rhythm: a metaphor? *Philos. Trans. R. Soc. B* 369, 20130396
- Abercrombie, D. (1967) *Elements of General Phonetics*, Edinburgh University Press
- Bowling, D.L. et al. (2013) Social origins of rhythm? Synchrony and temporal regularity in human vocalization. *PLoS One* 8, e80402
- Turk, A. and Shattuck-Hufnagel, S. (2013) What is speech rhythm? A commentary on Arvaniti and Rodriguez, Krivokapić, and Goswami and Leong. *Laborat. Phonol.* 4, 93–118
- Hawkins, S. (2014) Situational influences on rhythmicity in speech, music, and their interaction. *Philos. Trans. R. Soc. B* 369, 20130398
- Obermeier, C. et al. (2016) Aesthetic appreciation of poetry correlates with ease of processing in event-related potentials. *Cogn. Affect. Behav. Neurosci.* 16, 362–373
- Menninghaus, W. et al. (2015) Rhetorical features facilitate prosodic processing while handicapping ease of semantic comprehension. *Cognition* 143, 48–60

36. Liberman, M. and Prince, A. (1977) On stress and linguistic rhythm. *Linguist. Inq.* 8, 249–336
37. Beier, E.J. and Ferreira, F. (2018) The temporal prediction of stress in speech and its relation to musical beat perception. *Front. Psychol.* 9, 431
38. Large, E.W. and Jones, M.R. (1999) The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119
39. Greenberg, S. et al. (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. *J. Phon.* 31, 465–485
40. Ghitza, O. (2013) The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4, 138
41. Kochanski, G. and Orphanidou, C. (2008) What marks the beat of speech? *J. Acoust. Soc. Am.* 123, 2780–2791
42. Schroeder, C.E. and Lakatos, P. (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18
43. Giraud, A.-L. and Poeppel, D. (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511
44. Henry, M.J. and Hermann, B. (2014) Low-frequency neural oscillations support dynamic attending in temporal context. *Timing Time Percept.* 2, 62–86
45. Di Liberto, G.M. et al. (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25, 2457–2465
46. Luo, H. and Poeppel, D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010
47. Astheimer, L.B. and Sanders, L.D. (2009) Listeners modulate temporally selective attention during natural speech processing. *Biol. Psychol.* 80, 23–34
48. Buiatti, M. et al. (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuro-electric responses. *Neuroimage* 44, 509–519
49. Doelling, K.B. et al. (2014) Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768
50. Ding, N. et al. (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158
51. Cason, N. and Schön, D. (2012) Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* 50, 2652–2658
52. Kotz, S.A. and Schmidt-Kassow, M. (2015) Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex* 68, 48–60
53. Roncaglia-Denissen, M.P. et al. (2013) Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One* 8, e56000
54. Rothermich, K. et al. (2012) Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia* 50, 232–244
55. Honing, H. et al. (2015) Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140088
56. Ravignani, A. et al. (2014) Chorusing, synchrony and the evolutionary functions of rhythm. *Front. Psychol.* 5, 1118
57. Ravignani, A. et al. (2017) The evolution of rhythm cognition: timing in music and speech. *Front. Hum. Neurosci.* 11, 303
58. Ravignani, A. et al. (2016) What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* 10, 274
59. Ravignani, A. (2015) Evolving perceptual biases for antisynchrony: a form of temporal coordination beyond synchrony. *Front. Neurosci.* 9, 339
60. Savage, P.E. et al. (2015) Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8987–8992
61. Cirelli, L.K. et al. (2018) Rhythm and melody as social signals for infants. *Ann. N. Y. Acad. Sci.* Published online March 7, 2018. <http://dx.doi.org/10.1111/nyas.13580>
62. Hannon, E.E. et al. (2018) The developmental origins of the perception and production of musical rhythm. *Child Dev. Perspect.* Published online February 26, 2018. <http://dx.doi.org/10.1111/cdep.12285>
63. Rohrmeier, M. et al. (2015) Principles of structure building in music, language and animal song. *Philos. Trans. R. Soc. B* 370, 20140097
64. Ravignani, A. et al. (2018) Evolving building blocks of rhythm: how human cognition creates music via cultural transmission. *Ann. N. Y. Acad. Sci.* Published online March 6, 2018. <http://dx.doi.org/10.1111/nyas.13610>
65. Hasegawa, A. et al. (2011) Rhythmic synchronization tapping to an audio–visual metronome in budgerigars. *Sci. Rep.* 1, 120
66. Patel, A.D. et al. (2009) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830
67. Cook, P. et al. (2013) A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* 127, 1–16
68. Kello, C.T. et al. (2017) Hierarchical temporal structure in music, speech and animal vocalizations: jazz is like a conversation, humpbacks sing like hermit thrushes. *J. R. Soc. Interface* 14, 20170231
69. Ravignani, A. (2018) Comment on “Temporal and spatial variation in harbor seal (*Phoca vitulina* L.) roar calls from southern Scandinavia” [J. Acoust. Soc. Am. 141, 1824–1834 (2017)]. *J. Acoust. Soc. Am.* 143, 1–5
70. Ravignani, A. et al. (2018) Ontogeny of vocal rhythms in harbour seal pups: an exploratory study. *Curr. Zool.* Published online July 7, 2018. <http://dx.doi.org/10.1093/cz/zoy055>
71. Benichov, J.I. et al. (2016) Finding the beat: from socially coordinated vocalizations in songbirds to rhythmic entrainment in humans. *Front. Hum. Neurosci.* 10, 255
72. Gamba, M. et al. (2016) The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Front. Neurosci.* 10, 249
73. Large, E.W. and Gray, P.M. (2015) Spontaneous tempo and rhythmic entrainment in a bonobo (*Pan Paniscus*). *J. Comp. Psychol.* 129, 317
74. Hartbauer, M. and Römer, H. (2016) Rhythm generation and rhythm perception in insects: the evolution of synchronous choruses. *Front. Neurosci.* 10, 223
75. Sasahara, K. et al. (2015) A rhythm landscape approach to the developmental dynamics of birdsong. *J. R. Soc. Interface* 12, 20150802
76. ten Cate, C. et al. (2016) Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front. Psychol.* 7, 730
77. Hattori, Y. et al. (2013) Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Sci. Rep.* 3, 1566
78. Hattori, Y. et al. (2015) Distractor effect of auditory rhythms on self-paced tapping in chimpanzees and humans. *PLoS One* 10, e0130682
79. Hoeschele, M. and Bowling, D.L. (2016) Sex differences in rhythmic preferences in the budgerigar (*Melopsittacus undulatus*): a comparative study with humans. *Front. Psychol.* 7, 1543
80. Yu, L. and Tomonaga, M. (2015) Interactional synchrony in chimpanzees: examination through a finger-tapping experiment. *Sci. Rep.* 5, 10218
81. Arcadi, A.C. et al. (1998) Buttress drumming by wild chimpanzees: temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness. *Primates* 39, 505–518
82. Arcadi, A.C. et al. (2004) A comparison of buttress drumming by male chimpanzees from two populations. *Primates* 45, 135–139

83. Ravignani, A. *et al.* (2013) Primate drum kit: a system for studying acoustic pattern production by non-human primates using acceleration and strain sensors. *Sensors* 13, 9790–9820
84. Dufour, V. *et al.* (2017) The extraordinary nature of Barney's drumming: a complementary study of ordinary noise making in chimpanzees. *Front. Neurosci.* 11, 2
85. Babiszewska, M. *et al.* (2015) Social, contextual, and individual factors affecting the occurrence and acoustic structure of drumming bouts in wild chimpanzees (*Pan troglodytes*). *Am. J. Phys. Anthropol.* 156, 125–134
86. Bernejo, M. and Omedes, A. (2000) Preliminary vocal repertoire and vocal communication of wild bonobos (*Pan paniscus*) at Lilungu (Democratic Republic of Congo). *Folia Primatol.* 70, 328–357
87. Norton, P. and Scharff, C. (2016) "Bird song metronomics": isochronous organization of zebra finch song rhythm. *Front. Neurosci.* 10, 309
88. James, L.S. and Sakata, J.T. (2017) Learning biases underlie "universals" in avian vocal sequencing. *Curr. Biol.* 27, 3676–3682.e4
89. Lampen, J. *et al.* (2017) Neural activity associated with rhythmicity of song in juvenile male and female zebra finches. *Behav. Processes* Published online December 13, 2017. <http://dx.doi.org/10.1016/j.beproc.2017.12.003>
90. Tchernichovski, O. and Lipkind, D. (2017) Animal communication: origins of sequential structure in birdsong. *Curr. Biol.* 27, R1268–R1269
91. Rouse, A.A. *et al.* (2016) Beat keeping in a sea lion as coupled oscillation: implications for comparative understanding of human rhythm. *Front. Neurosci.* 10, 257
92. Schachner, A. *et al.* (2009) Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836
93. Fitzroy, A.B. *et al.* (2018) Horses do not spontaneously engage in tempo-flexible synchronization to a musical beat. In *ICMPC15/ESCOM10: Abstract book (electronic)* (Parncutt, R. and Sattmann, S., eds), pp. 423–424, Centre for Systematic Musicology, University of Graz
94. Wilson, M. and Cook, P.F. (2016) Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon. Bull. Rev.* 23, 1647–1659
95. de la Mora, D.M. *et al.* (2013) Do humans and nonhuman animals share the grouping principles of the iambic-trochaic law? *Atten. Percept. Psychophys.* 75, 92–100
96. Hagmann, C.E. and Cook, R.G. (2010) Testing meter, rhythm, and tempo discriminations in pigeons. *Behav. Processes* 85, 99–110
97. Spierings, M. *et al.* (2017) Selective auditory grouping by zebra finches: testing the iambic-trochaic law. *Anim. Cogn.* 20, 665–675
98. Toro, J.M. and Hoeschele, M. (2016) Generalizing prosodic patterns by a non-vocal learning mammal. *Anim. Cogn.* 20, 179–185
99. Toro, J.M. and Nespore, M. (2015) Experience-dependent emergence of a grouping bias. *Biol. Lett.* 11, 20150374
100. Crowe, D.A. *et al.* (2014) Dynamic representation of the temporal and sequential structure of rhythmic movements in the primate medial premotor cortex. *J. Neurosci.* 34, 11972–11983
101. Honing, H. *et al.* (2012) Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS One* 7, e51369
102. Merchant, H. and Honing, H. (2014) Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* 7, 274
103. Merchant, H. *et al.* (2015) Sensorimotor neural dynamics during isochronous tapping in the medial premotor cortex of the macaque. *Eur. J. Neurosci.* 41, 586–602
104. Selezneva, E. *et al.* (2013) Rhythm sensitivity in macaque monkeys. *Front. Syst. Neurosci.* 7, 49
105. Bower, F.L. *et al.* (2016) Disentangling beat perception from sequential learning and examining the influence of attention and musical abilities on ERP responses to rhythm. *Neuropsychologia* 85, 80–90
106. Donnet, S. *et al.* (2014) Monkeys time their pauses of movement and not their movement-kinematics during a synchronization-continuation rhythmic task. *J. Neurophysiol.* 111, 2138–2149
107. Honing, H. and Merchant, H. (2014) Differences in auditory timing between human and nonhuman primates. *Behav. Brain Sci.* 37, 557–558
108. Merchant, H. and Bartolo, R. (2018) Primate beta oscillations and rhythmic behaviors. *J. Neural Transm.* 125, 461–470
109. Takaura, K. and Fujii, N. (2016) Facilitative effect of repetitive presentation of one stimulus on cortical responses to other stimuli in macaque monkeys – a possible neural mechanism for mismatch negativity. *Eur. J. Neurosci.* 43, 516–528
110. Ayala, Y.A. *et al.* (2017) Monkeys share the neurophysiological basis for encoding sound periodicities captured by the frequency-following response with humans. *Sci. Rep.* 7, 16687
111. Fitch, W.T. (2015) Four principles of bio-musicology. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140091
112. Zarco, W. *et al.* (2009) Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J. Neurophysiol.* 102, 3191–3202
113. Leopold, D.A. and Logothetis, N.K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature* 379, 549
114. Winkler, I. *et al.* (2009) Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci.* 106, 2468–2471
115. Ghazanfar, A.A. *et al.* (2013) Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proc. Natl. Acad. Sci.* 110, 1959–1963
116. MacNeilage, P.F. (1998) The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–511
117. Bergman, T.J. (2013) Speech-like vocalized lip-smacking in geladas. *Curr. Biol.* 23, R268–R269
118. Ghazanfar, A.A. *et al.* (2012) Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr. Biol.* 22, 1176–1182
119. Ghazanfar, A. and Poeppel, D. (2014) The neurophysiology and evolution of the speech rhythm. In *The Cognitive Neurosciences* (5th edn) (Gazzaniga, M.S. and Mangun, G.R., eds), pp. 629–638, MIT Press
120. Zentner, M. and Eerola, T. (2010) Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci.* 107, 5768–5773
121. Lang, M. *et al.* (2017) Sync to link: endorphin-mediated synchrony effects on cooperation. *Biol. Psychol.* 127, 191–197
122. Pearce, E. *et al.* (2015) The ice-breaker effect: singing mediates fast social bonding. *R. Soc. Open Sci.* 2, 150221
123. Tarr, B. *et al.* (2016) Silent disco: dancing in synchrony leads to elevated pain thresholds and social closeness. *Evol. Hum. Behav.* 37, 343–349
124. Frigyesi, J. (1993) Preliminary thoughts toward the study of music without clear beat: the example of 'flowing rhythm' in Jewish 'nusah'. *Asian Music* 24, 59–88
125. Grahn, J.A. (2012) Neural mechanisms of rhythm perception: current findings and future perspectives. *Topics Cogn. Sci.* 4, 585–606
126. Schwartz, M. and Kotz, S.A. (2013) A dual-pathway neural architecture for specific temporal prediction. *Neurosci. Biobehav. Rev.* 37, 2587–2596
127. MacNeilage, P.F. and Davis, B.L. (2001) Motor mechanisms in speech ontogeny: phylogenetic, neurobiological and linguistic implications. *Curr. Opin. Neurobiol.* 11, 696–700
128. Port, R.F. (2003) Meter and speech. *J. Phon.* 31, 599–611
129. Chandrasekaran, C. *et al.* (2009) The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436

130. Ghazanfar, A.A. and Takahashi, D.Y. (2014) Facial expressions and the evolution of the speech rhythm. *J. Cogn. Neurosci.* 26, 1196–1207
131. Remedios, R. *et al.* (2009) Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. *Proc. Natl. Acad. Sci.* 106, 18010–18015
132. Hasson, U. and Frith, C.D. (2016) Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philos. Trans. R. Soc. B* 371, 20150366
133. Janik, V.M. (2014) Cetacean vocal learning and communication. *Curr. Opin. Neurobiol.* 28, 60–65
134. Reichmuth, C. and Casey, C. (2014) Vocal learning in seals, sea lions, and walruses. *Curr. Opin. Neurobiol.* 28, 66–71
135. Lattenkamp, E.Z. and Vernes, S.C. (2018) Vocal learning: a language-relevant trait in need of a broad cross-species approach. *Curr. Opin. Behav. Sci.* 21, 209–215
136. Patel, A.D. (2014) The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol.* 12, e1001821
137. Abramson, J.Z. *et al.* (2018) Imitation of novel conspecific and human speech sounds in the killer whale (*Orcinus orca*). *Proc. R. Soc. B* 285, 20172171
138. Ralls, K. *et al.* (1985) Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. *Can. J. Zool.* 63, 1050–1056
139. Handel, S. *et al.* (2012) Hierarchical and rhythmic organization in the songs of humpback whales (*Megaptera novaeangliae*). *Bioacoustics* 21, 141–156
140. Schneider, J.N. and Mercado, E., III (2018) Characterizing the rhythm and tempo of sound production by singing whales. *Bioacoustics* 1–18 <http://dx.doi.org/10.1080/09524622.2018.1428827>
141. Sabinisky, P.F. *et al.* (2017) Temporal and spatial variation in harbor seal (*Phoca vitulina* L.) roar calls from southern Scandinavia. *J. Acoustic. Soc. Am.* 141, 1824–1834
142. Nikolich, K. *et al.* (2016) Quantitative classification of harbor seal breeding calls in Georgia Strait, Canada. *J. Acoustic. Soc. Am.* 140, 1300–1308
143. Ravignani, A. and Kotz, S. (2017) Evolution of temporal interaction: a comparative approach to social timing. *PeerJ Preprints* 5, e3275v1
144. Schirmer, A. *et al.* (2016) The socio-temporal brain: connecting people in time. *Trends Cogn. Sci.* 20, 760–772
145. Pika, S. *et al.* (2018) Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B* 285, 20180598
146. Roeske, T.C. *et al.* (2018) Multifractal analysis reveals music-like dynamic structure in songbird rhythms. *Sci. Rep.* 8, 4570