



UvA-DARE (Digital Academic Repository)

Precursors of Music and Language in Animals

ten Cate, C.; Honing, H.

DOI

[10.31234/osf.io/4zxtr](https://doi.org/10.31234/osf.io/4zxtr)

Publication date

2022

Document Version

Final published version

License

CC BY

[Link to publication](#)

Citation for published version (APA):

ten Cate, C., & Honing, H. (2022). *Precursors of Music and Language in Animals*. PsyArXiv. <https://doi.org/10.31234/osf.io/4zxtr>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Precursors of Music and Language in Animals

Carel ten Cate¹ and Henkjan Honing²

¹Institute of Biology Leiden (IBL), Leiden Institute for Brain and Cognition (LIBC), Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands

²Music Cognition Group (MCG), Institute for Logic, Language and Computation (ILLC), University of Amsterdam, PO Box 94242, 1090 CE Amsterdam, The Netherlands

Abstract

Language and music are universal human traits, raising the question for their evolutionary origin. This chapter takes a comparative perspective to address that question. It examines similarities and differences between humans and non-human animals (mammals and birds) by addressing whether and which constituent cognitive components that underlie the human ability for language and music can be found in non-human animals. It first provides an introduction to the nature and meaning of vocalizations and non-vocal communicative sounds in non-human animals. Next it reviews experimental and observational evidence of animal perception of various frequency and temporal dimensions of sounds. Many animal species show perceptual and cognitive abilities to distinguish between or to generalize auditory stimuli. This includes evidence of the presence of one or more of the constituent cognitive components on which the human abilities for language and music are based, or that may have served as precursors for these components. At the same time, there are also important differences among animal species in their abilities. Hence contrasts are not limited to those between humans and other animal species. The differences between humans and other species, as well as those among non-human species, might result from specific biases and the weight or priority certain species give to attending to certain features of an acoustic signal, or because different species use particular mechanisms to different degrees.

Keywords

Evolution; capacity for music; capacity for language; comparative studies; birds; mammals; primates; sound perception; animal cognition

29.1 Introduction

Both language and music are universal human traits (Arbib, 2013; Hagoort, 2019), suggesting they were already present in the ancestors of modern humans. This raises the question for the evolutionary origins of these traits. Comparative research, examining whether and which other species show similar capabilities, or behaviours that may be considered evolutionary precursors of language and music, addresses this question. At first sight, the answer may seem simple. After all, many other species produce sounds. These vary from simple or more complex vocalizations to producing sounds by other means, such as drumming on a substrate. There is abundant evidence that animals use such sounds to communicate and exchange information (Bradbury & Vehrencamp, 2011; Fishbein et al., 2020). But to what extent is such behaviour really comparable to human language and music? To answer this question, we need to probe deeper. Human language and music are both complex behaviours arising from constituent components that together make that we can use spoken language to communicate infinite types of messages by finite means, or produce or enjoy a bewildering variety of vocal or instrumental sounds full of musical sensory and emotional information that can be appreciated as music.

Our starting point is that the search for the precursors of human language and music is, in fact, the search for the presence of the constituent cognitive components of language and music in other species. The cognitive skills that might allow for language to arise, consist of components such as auditory-guided vocal learning, sophisticated vocal production and advanced perceptual abilities, the presence of a semantic repertoire, a system of syntactical rules, etc. We will refer to this set of linguistic and cognitive skills as the capacity for language (or *linguisticity*; Haspelmath, 2020), to distinguish between these perceptual and cognitive skills and the acoustical object of language itself. Similarly, we will refer to the capacity for music as *musicality* (Honing et al., 2015), i.e. the constituent components that enable music perception and production, for instance, cognitive skills like relative pitch perception, beat perception, tonal encoding of pitch, and metrical encoding of rhythm. So, rather than asking whether other animal species 'have' language or music, comparative research asks whether and which of the components of the capacity for language or music are present in humans and non-human species (Fitch, 2005, 2006). Such studies may concentrate on species closely related to humans, such as great apes or monkeys. The presence of one or more of the above mentioned components may indicate that they were already present in the shared ancestor of humans and apes or monkeys. Comparative studies may also indicate traits that may have served as precursors for more advanced or complex human abilities and allow experiments to examine the neurobiological or genetic basis of such traits. Apart from tracing such so-called homologies in related species, more distantly related species can be examined for the presence of similar, analogous traits, arisen by independent convergent evolution (ten Cate & Healy, 2017). Comparing features of birdsong with language or music is an example of this approach (Honing, 2018). Songbird songs are often melodiously varied, seemingly complexly structured and acquired by vocal learning (e.g., Searcy et al., 2021). As such, they share more similarities with language and music than the less melodic, more simply structured and non-learned vocalizations of most monkeys and apes (Hage et al., 2016; Watson et al., 2015). Studies of songs may thus provide insight in the mechanisms required for such complex traits and provide hypotheses for the selection pressures giving rise to them.

29.1.1 Comparative research

Comparative studies can be done in several ways. One approach is to examine the sounds made by animals and look for shared features or parallels with language or music. To study these, one can, for example, examine whether different vocalisations differ in meaning, comparable to words or sentences, ask how the structure of a sequence of sounds compares to syntactic structures in language or rhythmic structures in music, or whether harmonic sounds are recognized by their pitch (like in music) or by their spectral structure (like in speech). The presence of such features can indicate that similar sensory or cognitive mechanisms may underlie their perception and production and those needed for language and music in humans. However, one needs to be cautious with drawing such conclusions. That a sound produced by an animal has certain features in common with language or music may be incidental and a result of human interpretation, rather than indicating shared mechanisms per se. Animal sounds showing, for example, a specific rhythmic pattern (e.g., in the call of the indri, a lemur species; De Gregorio et al., 2021) or that contain tones based on a harmonic series (e.g., in the hermit thrush; Doolittle et al., 2014), need not indicate an ability of the animal to perceive or produce rhythms or harmonic sounds in general, as is common in humans. To show this, it is necessary to demonstrate the perception or production of such patterns outside and beyond what is realized in the species-specific sound patterns (ten Cate et al., 2016). This requires a second approach: using controlled experiments to address whether animals can (learn to) distinguish and generalize artificially constructed sounds that differ in specific linguistic or musical features. The two approaches, observational-analytical and experimental, are complementary: the first one may hint at presence of a certain ability, while the second one can test its existence and the limits of the capacity (see **Figure 29.1**).

Below, we first provide some general background on animal sounds, in particular their function, with a focus on birds and mammals. Next follows a number of sections in which we discuss to what extent a specific sound feature shown by one or more animal species shows a similarity with a linguistic or musical feature, suggestive of the presence of a similar underlying cognitive ability. We also discuss experimental approaches focussing on testing the presence of a certain ability, and conclude what can be said (or not) about presence of that ability in animals. We end this chapter with some thoughts on the extent to which various abilities can be seen as precursors for language and/or music, and whether they may or may not indicate the presence of common roots for both cognitive capacities.

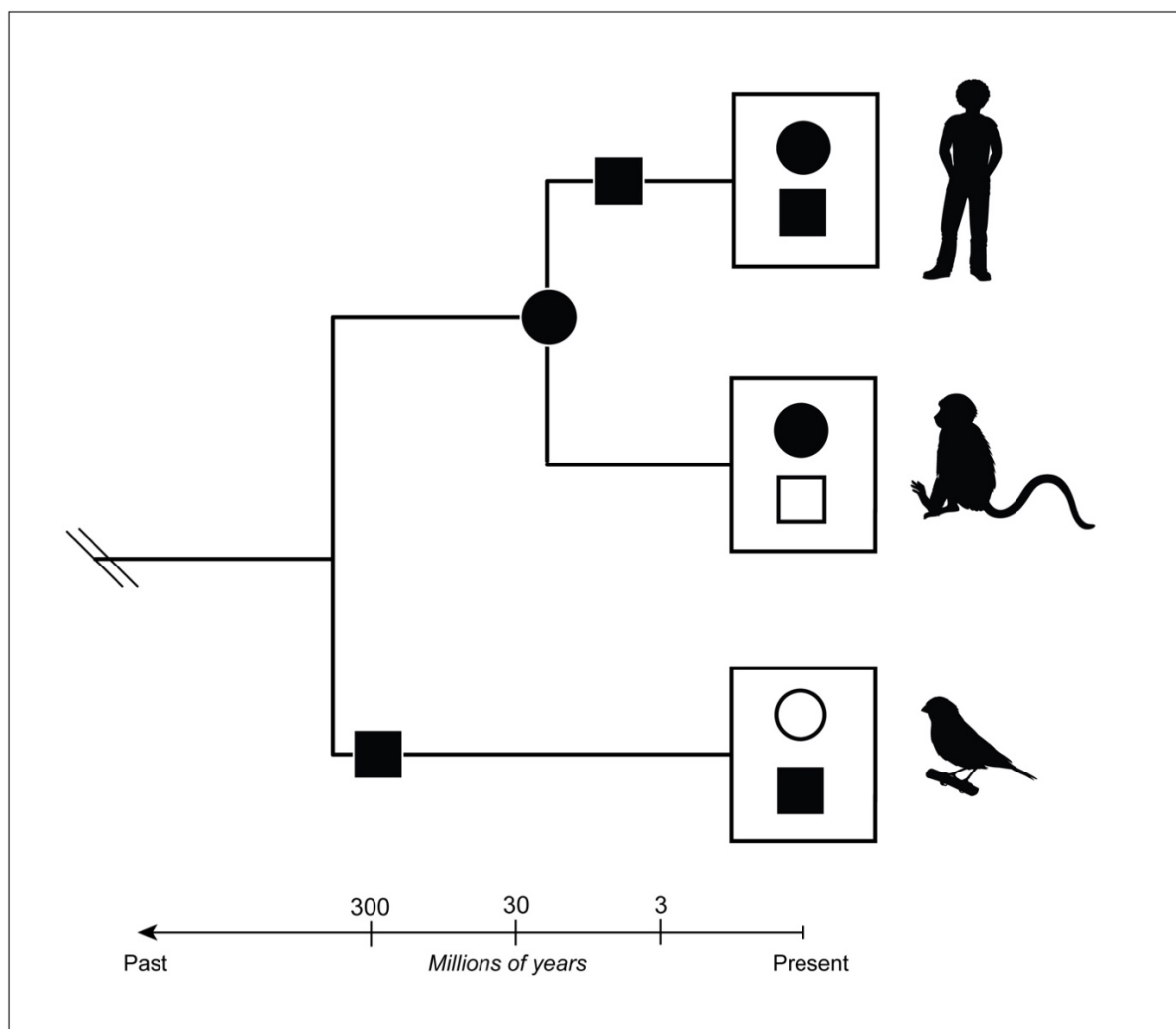


Figure 29.1 Diagrammatic representation of the comparative approach (as discussed in this chapter). It shows a hypothetical phylogenetic tree that illustrates the evolution of several traits that humans may share with monkeys and birds. Filled shapes represent a hypothetical trait (such as vocal learning or beat perception); open shapes indicate the absence of that trait. The position on the phylogenetic tree dates the possible evolutionary origin of such a trait. N.B. Circle: homologous trait, present in human and monkeys, originating from a shared ancestor; Square: an independently evolved trait, similar in humans and birds by convergence.

29.2 Natural vocalizations and non-vocal sounds

Most mammal and bird species communicate by using a range of vocalizations or sometimes non-vocal sounds, mostly as signals linked to specific contexts. Whatever the nature of the sound or the context, it is important to realize that these sounds could only have evolved because they serve a biological function: in the end they contribute to the survival and reproduction of the individuals. This is because sound production always comes with costs: an animal needs to invest resources into developing a sound producing organ, the neurological system to control it, as well as perceptual abilities to analyse sounds. Producing sounds also takes energy and making sounds has the negative side effect to indicate the location of the producer to a potential predator. Therefore, sound production and perception can only evolve and remain present in a species if the benefits in terms of increased reproductive success outweigh the costs. Selective pressures will thus act upon sound signals such that their characteristics are adapted to a specific biological function. This also results in

different sound characteristics and repertoires among different species, although some basic principles are shared between species.

Vocalizations are often classified as either 'calls' or 'songs'. The term 'calls' is mostly used for vocalisations that are generally short, consisting of one or a few units (syllables or elements). They often have a species-specific structure, showing limited inter- and intra-individual variability. Many, if not all, mammal and bird species produce calls. In addition, some species may also produce vocalizations often classified as 'song'. This term is often used for longer and more variable vocalisations. This includes the conspicuous songs of songbirds (Catchpole & Slater, 2008) but song-like vocalisations are also known from whales, such as the humpback whale (Payne & McVay, 1971). Calls often, though not always, develop without learning. In contrast, the songs of several animal groups, most notably songbirds (Catchpole & Slater, 2008), hummingbirds (e.g., Johnson & Clark, 2020) and whales (e.g., Tyack, 2020) are acquired by learning (imitation) from a parent or other individual. Such so-called vocal production learning, characteristic of human language development, is rare among non-human animal species (Vernes et al., 2021; See also chapter X, this volume). Among birds, it is also present in parrots and one duck species (ten Cate, 2021), while among mammals it is present in dolphins and whales, some pinnipeds, elephants and some bat species (Janik et al., 2021), but not in non-human primates. Vocal production learning may also give rise to individual differences as well as cultural changes and local dialects in vocalizations (Searcy et al., 2021; Zandberg et al., 2021).

29.2.1 Calls

Some calls may already be produced by new born individuals, such as begging calls. The intensity of such calls may vary depending on the hunger state of the individual and they affect the parental feeding rate. When adult, animals may produce a range of calls in various contexts. A well-known category are the alarm calls given by many birds and mammals when a predator is around. In some species, such as the vervet monkey, different types of calls are used depending on the predator present – a leopard, eagle or snake (Seyfarth et al., 1980). Playback experiments with these calls give rise to different types of responses in recipients: a leopard call makes them rush into the trees, an eagle call induces them to hide in thick bushes and a snake call results in approaching and mobbing (Seyfarth et al., 1980). This has often been interpreted as a form of referential signalling, comparable to humans shouting the words 'leopard', 'eagle' or 'snake' to each other. However, later studies indicate that the same call types are also used in other contexts, for example, eagle calls are also given during within-group aggression (Price et al., 2015). Hence these calls are not directly or uniquely linked to a specific referent or message. One interpretation of the findings is that different contexts may produce a similar internal state in the monkeys and that this internal state gives rise to the production of a specific call (Fischer & Price, 2017). The sound may get its meaning through being produced in a specific context. Another difference between the alarm calls and human words is that in humans the labels to refer to different predators are arbitrary. Each language may use a different word for 'eagle', but such arbitrariness is absent in the monkeys in which the call type-context relationship is present from birth and the call is produced without learning being involved. A second example of call variation and use are the calls made by chimpanzees in a feeding context. In Edinburgh Zoo chimpanzees produce somewhat different 'grunts' when they are given apples and when given bread (Slocombe & Zuberbühler, 2005). Playback of these sounds demonstrated that the chimpanzees have formed an association between the call variants and the type of food (Slocombe &

Zuberbühler, 2005). These calls thus serve to indicate to group members the presence of a particular type of food at the location of the caller. Structurally the calls are very similar and are on different ends of a graded continuum – one might say that they are different intonations of the same call, rather than different calls. So, while humans use a repertoire of different (learned) words to relate to different objects, primates often have a limited repertoire of (non-learned) calls within which there may be meaningful variation. However, that a call shows graded variation needs not imply that the message or interpretation is also graded, as recipients may have *categorical perception*, in which the perception of a continuously varying signal shows distinct transitions within the continuum. This is similar to humans who, confronted with an artificially varying ba/da continuum show a specific threshold at which the perception of the sound flips from ‘ba’ to ‘da’. A similar phenomenon can be observed for music in the temporal domain, where the continuum of physical time seems to be clustered around integer-ratio rhythmic categories (Desain & Honing, 2003; Jacoby & McDermott, 2017). Such categorical perception is one of the cornerstones in human speech perception and its presence in other species, ranging from chinchillas (e.g., Burdick & Miller J D, 1975) to Japanese quail (e.g., Kluender et al., 2016) and budgerigars (e.g., Dooling & Brown, 1990) indicates that it most likely preceded and enabled the evolution of speech sound repertoires, rather than evolved in consort with speech – an example of the insights that can be gained from comparative research. Interestingly however, presence of a limited ‘vocabulary’ of calls does not mean that animals are unable to differentiate perceptually among a much wider range of signals. Various apes, for instance Kanzi, a bonobo (Savage-Rumbaugh et al., 1986) have demonstrated that they can make links between arbitrary, in this case visual, symbols and objects or actions. Animals can also be trained to link spoken words to many objects. Some dogs, such as the border collie Rico (Kaminski et al., 2004) have demonstrated a remarkable ability to link up to hundreds of arbitrary human spoken labels to particular objects, while dogs and other species also can link commands or other sounds to performing specific actions. So, in contrast to the lack of flexible vocal production enabling an extensive range of arbitrary vocalizations, the ability to perceptually distinguish many, arbitrary, sounds and making connections to objects and actions is present in many if not all species. A prime example of this is provided by famous grey parrot Alex (Pepperberg, 1999), which also possessed a rich repertoire of word imitations and showed the cognitive ability to use these appropriately.

29.2.2 Songs

Because of being learned and their complexity and melodious nature, the songs of songbirds are suggestive of structural and functional similarities with both language and music. In later sections we discuss in more detail whether some features of songs indicate the presence of similar underlying cognitive capabilities in humans and non-human animals. Here we focus on the communicative significance of songs. They are mostly produced in relation to reproduction, and many studies demonstrate songs can function to attract mates or to keep competitors out of a territory (Catchpole & Slater, 2008). They have this impact because there is often a relation between some song characteristics or singing styles and relevant physical qualities of the singer, which is thus being ‘advertised’. In canaries, for example, females are attracted to males singing a particular type of syllable. This so-called ‘sexy syllable’ (Vallet & Kreutzer, 1995) is difficult to produce by males, which means that its presence in songs is a signal of male quality. Another example is provided by male water pipits, which have songs that may contain a ‘snarr’ element (Rehsteiner et al., 1998). This loud harsh broadband element is energetically costly to produce and hence is shown more by heavier and most likely also stronger males and thus signals the competitive qualities to other males. As a result,

territories of such males show less overlap with those of other males. For many bird species, also for non-songbirds, there is evidence of such types of relations between certain sound characteristics or singing styles and the competitive qualities of the singer. The idea is that the importance of demonstrating the singer's competitive qualities is thus one prominent factor driving the complexity of songs. Apart from this, song features are also under selection to avoid confusion with other species and to transfer the message in a specific environment. Intraspecific variation in songs may also assist the identification of important conspecifics, such as partners, kin or territorial neighbours, just like human voice characteristics can provide information of several aspects of a speaker. So, a song thus may provide a receiver with a lot of information, but not because it is providing a sentence-like message, but because acoustic features are linked to species- and individual-specific biologically meaningful information. Selection on vocal characteristics may also result in songs that sound music-like to humans. The songs of the thrush nightingale, for instance, contain fragments, 'phrases', of repeated notes that over the phrase increase in amplitude or in note intervals, just like a *crescendo* or an *accelerando* that is common in many styles of human-produced music (Rothenberg et al., 2014). As nightingale songs are produced to attract females, it has been suggested that such features are not only features that are attractive to human listeners, but also to female nightingales. That may well be so, but we have to keep in mind that females are primarily likely to respond because there might be information about physical or other qualities of a male encoded in such features. Perhaps an acceleration is correlating with physical qualities, which may have driven the evolution of both its acoustical character as well as the bird's preference for it. So, in order to consider such a trait as indicating a general sensitivity to particular acoustical patterns, rather than it being a preference strictly linked to judging song quality, it needs to be demonstrated that nightingales can also recognize a *crescendo* and an *accelerando* in arbitrary acoustic stimuli.

29.2.3 Non-vocal sounds

Non-vocal sounds, using some substrate for sound production, are also present in several animal species. Woodpecker species show species-specific drumming patterns in their reproductive season (Dodenhoff et al., 2001), often produced on structures that amplify the sound. These sounds may serve a similar role as the songs of other species – attracting mates and advertising territories. Great apes, such as gorillas and chimpanzees (e.g., Babiszewska et al., 2015), may drum on their chest, on tree buttresses, or other structures to impress potential rivals. Several mammal species, including rodent, deer, shrews, marsupials and elephants, communicate by foot drumming on the ground, thus transmitting similar messages as other species do by vocalizing (Randall, 2001). The palm cockatoo even uses a stick or other object to hit a branch, producing a more or less regular isochronous series of taps (Heinsohn et al., 2017). This is also a display produced in a reproductive context. Hence, these non-vocal instrumental sounds serve similar functions to the vocal sounds produced by animals.

29.3 Components of language and music

We now turn to some specific components of language and music to see whether and what similarities can be found in various animal species.

One way of categorizing the sensitivities of animals to the building blocks of language and music is to group these sensitivities along the frequency/spectral and temporal dimensions of sound. Although speech and music share many acoustic features, music appears to take advantage of a different set of acoustic features than speech. In humans the frequency dimension is central to music/melody

perception, while for understanding speech the temporal dimension appears to be most fundamental (Albouy et al., 2020; Shannon et al., 1995). With respect to the frequency dimension of speech, humans attend primarily to the spectral structure (which enables the distinction between the different vowels and consonants), while for music the attention appears to be less on a spectral quality (e.g., the sound of a guitar versus that of a flute), but instead on the melodic and rhythmic patterns. As such, it might well be that humans are an exception in that they can interpret the same sound signal in (at least) two distinct ways: as speech or as music (*cf.* speech-to-song illusion; See chapter X, this volume). In other animals such distinction is not observed (as yet). In humans, melody and speech are processed along specific and distinct neural pathways (Albouy et al., 2020; Norman-Haignere et al., 2015, 2022) and it could be that brain networks that support musicality are partly recycled for language (Peretz et al., 2018). This could imply that both language and music share one precursor. In fact, it is one possible route to test the Darwin-inspired conjecture that musicality precedes music and language (Fitch, 2013a; Hoeschele et al., 2018; Honing, 2021). Below we will discuss the potential components of such a precursor.

29.3.1 Animal perception of frequency dimensions

29.3.1.1 Frequency and pitch

Pitch perception is critical for the perception of speech and music, or, more generally, for auditory scene analysis: how to extract relevant information from a complex sound signal (A. S. Bregman, 1990). The perceived pitch of a complex sound corresponds often to its fundamental frequency. However, it is good to realize that pitch is a *percept*. Humans perceive pitch in harmonic tones, even when the fundamental is missing. Pitch perception develops in the first months of human life (He & Trainor, 2009), hence its ontogeny might be influenced by both biological and environmental factors.

Perceiving the pitch of the missing fundamental (of harmonically structured tones) has been shown in a few non-human animals, including starlings (Cynx & Shapiro, 1986), cats (Heffner & Whitfield, 1976), and monkeys (Tomlinson & Schwarz, 1988), as such suggesting a biological basis for inferring pitch from a harmonic spectrum. Most mammals and avian species perceive, memorize and recognize the frequency of vocalizations of their conspecifics. Frequency often contains biologically relevant information (e.g., larger individuals producing lower frequency calls), so distinguishing between frequencies and memorizing them can be important to an animal. Frequency perception is often done in an absolute way, that is, mammals and birds can identify the frequency of a pure tone. To recognize a sound by its pitch (a percept derived from fundamental frequency) is often referred to as *absolute pitch* (AP).¹ A range of studies have demonstrated that birds are particularly good at distinguishing different tones by their absolute frequency (Hoeschele, 2017). (N.B. By using pure tones in these experiments, pitch and frequency become the same thing.) For instance, when 40 different pure sine tones between 980 and 5660 Hz were divided in 8 frequency ranges, several songbirds, budgerigars and pigeons, could all be trained to allocate each single tone to its correct frequency range. This is in contrast to humans and rats that were trained on the same tones. Both species showed difficulties in allocating pure tones to the correct frequency ranges. They only

¹ It has to be noted that AP generally refers not only to being able to remember a sound by its pitch or F_0 , but also being able to label it (is it a C or a D?) without a reference, as effortlessly as most of us (humans) label colors (Levitin & Rogers, 2005). The first part of the skill (i.e. memory for pitch) is widespread in humans (Schellenberg & Trehub, 2003) and non-human animals (Hoeschele, 2017), the second part is rare (estimated as 1 in 10000 people; Levitin & Rogers, 2005).

managed to do so when the number of ranges was reduced from eight to three. Interestingly, humans with full AP¹ were able at distinguishing all eight ranges (Weisman et al., 2012). So, the main difference in pitch perception seems not to be between humans and other non-human animals, but between mammals (including humans) and birds.

Another aspect of the spectral structure of sound is the variation it may show in the energy distribution of its overtones or partials. For vocalizations, this energy distribution, or *spectral envelope* (Patel, 2017), is determined by the resonance properties of the vocal tract, and is particularly important for human speech production and perception. For instance, different vowels are produced by specific energy distributions, referred to as *formants*: a formant being a local maximum in the spectrum. The ratio between the frequency ranges of these formants characterizes the vowel that is being perceived (like the [i] or [u] in **Figure 29.2**). The spectral structure in the calls of many species may also show variation in the spectral envelope of the different overtones, and this variation has shown to be of communicative significance in many mammal and bird species. It is thus no surprise that various species, such as ferrets (e.g., Bizley et al., 2013), dogs (e.g., Root-Gutteridge et al., 2019), budgerigars (e.g., Henry et al., 2017), zebra finches (Ohms et al., 2010) and other species, also are capable of distinguishing natural human vowels or artificially constructed ones.

In humans, the voice characteristics differ substantially between infants and adults and between male and female voices, due to the maturation and size of the vocal tract. Hence the fundamental frequency of different vowels, as well as the formant peak frequencies differ – both being lower in male than in female and infant voices (Huber et al., 1999). Nevertheless, humans can distinguish between spoken words, such as *wet* and *wit*, irrespective of the speaker because we use the frequency ratio of the formants that characterize the vowel sound. This may seem a specialized speech-related ability, but a study on zebra finches (Ohms et al., 2010) showed that, when trained on either male voices or female voices to discriminate the human words *wit* and *wet*, they could generalize this discrimination to voices of the other sex. Zebra finches also have the capacity to distinguish harmonically structured sounds by either attending to the frequency of the fundamental or to that of the formant by attending to the spectral envelope (Burgering et al., 2018). Furthermore, zebra finches (Burgering et al., 2019), but also dogs (Root-Gutteridge et al., 2019), can learn to categorize different words spoken by different speakers by either word or speaker. These experiments show that several species perceive both pitch as well as spectral envelope (Heimbauer et al., 2011; Salmi et al., 2021), and have the flexibility to attend to one or the other depending on the context.

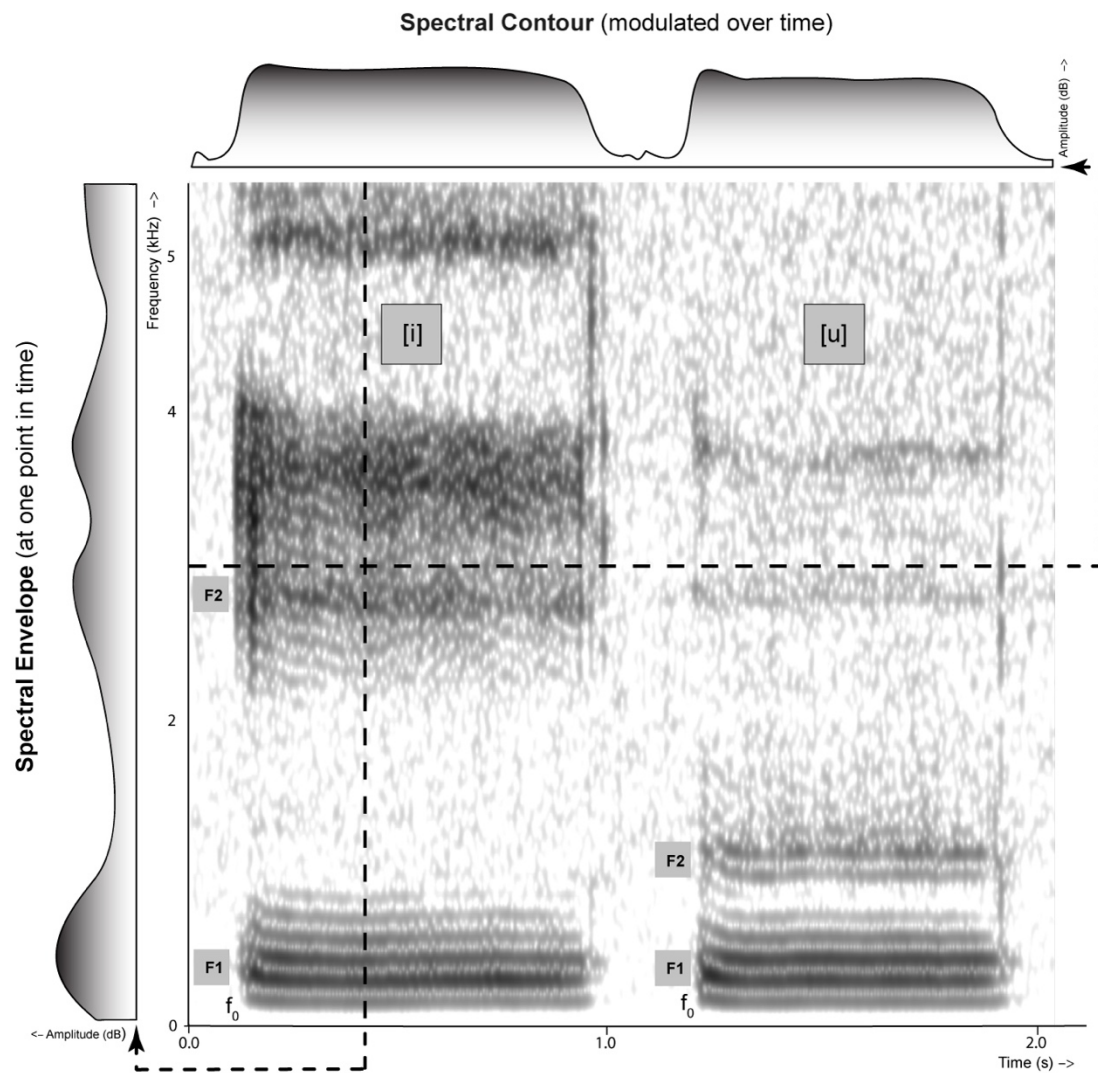


Figure 29.2 Spectrogram showing two speech vowels: [i] and [u]. The fundamental frequency (f_0) is indicated, while labels F1 and F2 mark the formants (local maxima at lower frequency bands), with their peak frequency ratio characterizing the nature of the vowel. The energy distribution across the Y-axis is referred to as *spectral envelope* (or *spectral shape*), depicted as an envelope curve at one point in time. The energy distribution across the x-axis is referred to as *spectral contour* (i.e. *spectral change over time*), here depicted for one frequency-band.

In music spectral quality or *timbre* (sound colour) often plays a secondary role (McAdams & Siedenburg, 2019). Humans remember and reproduce a melody primarily based on the intervallic structure or pitch contour (McDermott et al., 2008); a melody played on a piano or a flute (i.e. two different timbres) is generally considered the same melody. In fact, humans can easily distinguish between triadic chords (major, minor, diminished, etc.) even when the timbre of these chords changes. In contrast, black-capped chickadees, while able to discriminate between triadic chords, fail to generalize to novel timbres (Hoeschele et al., 2012, 2014). As such, they are not attending to the pitch-relationships of the chords, as humans tend to do.

Returning to pitch perception, it can be questioned, though, whether AP is a constituent component of music and/or language (Gingras et al., 2015). More fundamental appears to be *relative pitch* (RP): recognizing a transposed melody or intonation pattern as the same. In humans *transposability* is an important aspect of musicality. We can easily recognize a melody that is transposed up a few semi tones as the same melody, similarly, tonal language speakers recognize the tones irrespective of speakers' fundamental pitch. Birds have great difficulties in classifying pitch-shifted sounds. Zebra finches, that are trained to discriminate two different songs, fail to recognize these songs when the frequencies are shifted up or down by 8% or more (Nagel et al., 2010). Starlings trained to discriminate a series of subsequent pure tones ascending in frequency from a descending series, fail to do this when the frequencies are shifted outside the training range (MacDougall-Shackleton & Hulse, 1996). However, they can learn to do this after more extensive training. In addition, and in contrast to zebra finches, starlings are capable of recognizing frequency transposed versions of starling songs (M. R. Bregman et al., 2012). Next to starlings (MacDougall-Shackleton & Hulse, 1996) and ferrets (Yin et al., 2010), RP perception is also shown in black-capped chickadees. They sing a two-syllable 'fee-bee' song in which the two elements have a fixed relative pitch interval, but may vary in absolute frequency or pitch height (Weisman & Ratcliffe, 1989). The chickadees are sensitive to the frequency ratio of the elements, but clearly more so in modified versions of their natural song than with synthetic tones. Again, this suggests that birds attend to more than pitch alone in an acoustic signal. While there is thus some evidence for RP (Weisman & Ratcliffe, 2004), most birds seem to rely on AP. But starlings, when given the opportunity to use other information than frequency alone (i.e., by using rich spectral sounds instead of a single sine tone – as is used in virtually all pitch perception experiments cited above), tend to use the spectral information in their generalizations. Hence, a starling might not focus on the absolute frequencies, but more on the *spectral envelope* –the distribution of energy in the spectrum– and the change thereof from moment to moment, the *spectral contour* (M. R. Bregman et al., 2016) (*cf.* Figure 1). This is comparable to the way humans listen to speech (Shannon, 2016). However, it remains unclear how wide spread this listening strategy is in birds and other animals.

29.3.1.2 Octave equivalence

Octave equivalence (or octave generalization) refers to the perceived similarity of tones that are separate by an octave or doubling in frequency, also referred to as *pitch chroma* (Shepard, 1964). This can be seen as a weak form of RP. Pitch chroma has been shown in humans to be processed differently from *pitch height*, suggesting them to be different 'percepts' that are processed in different locations in the auditory cortex (Warren et al., 2003). In animals pitch height perception is wide spread (see under Frequency above), but pitch chroma perception has only been shown sporadically, for instance in monkeys (Wright et al., 2000) and in a captive bottlenose dolphin, which, when mimicking sounds, transposed them to a lower octave when the original was in a frequency range outside its vocal range (Richards et al., 1984). In contrast, octave generalization was shown to be absent in budgerigars (Wagner et al., 2019), black-capped chickadees (Hoeschele et al., 2013) and starlings (Cynx, 1993), suggesting it not to be crucial for these avian species and unlikely to be linked to vocal mimicry. Hence, it could well be that octave generalization is restricted to mammals.

So, with respect to RP, the results from animal observations and experiments seem somewhat mixed, but various mammal and bird species seem to perform poorer at this ability than humans.

29.3.1.3 Consonance

In music, consonance and dissonance are categorizations of simultaneous or successive sounds, that refer to the perceived quality of two tones with a certain pitch, as well as that between two spectra, i.e. two simultaneously sounding sounds. There is a vast literature theorizing about the mathematical and/or physical basis of this perceived quality. Some consider consonance a natural phenomenon that is inherent in those acoustical phenomena that have a harmonic structure (such as vibrating strings), with low-integer relationships between the tones making them sound consonant (Bowling et al., 2017). Hence, it is reasonable to think that species that are largely dependent on acoustic signals are sensitive to consonance. On the other hand one can focus on consonance as a perceptual or cognitive construct that might be (partly) learned (McDermott et al., 2016). Several studies have identified consonance perception in animals (Chiandetti & Vallortigara, 2011; Hulse et al., 1995; Izumi, 2000; Sugimoto et al., 2010), but also the absence thereof (McDermott & Hauser, 2004; Wagner et al., 2020), or have more complex outcomes. The vocalizations of some species suggest they are sensitive to successive consonance. For example, great tits have stereotyped songs consisting of a small number (usually 2-3) of subsequent tones. Attractive males are more likely to produce small-integer ratios between these tones, suggesting this ability to act as a sign of quality, reflecting, e.g., social status, dominance and reproductive potential (Richner, 2016). Male hermit thrush songs also consist of tones with harmonically, integer related fundamental frequencies (Doolittle et al., 2014). However, it is unknown whether these bird species can actually *perceive* these integer relationships. Overall, the available, but still limited, literature does not reveal a pattern that can unambiguously underpin the biological basis for consonance perception in simultaneous or successive sounds, despite the prominence of this idea in many theories (Harrison, 2021; Helmholtz, 1954; Purves, 2017).

In summary, the available literature suggests many animal species have a wide spread perceptual sensitivity to at least some aspects related to frequency and pitch perception that are important to human speech and music perception. This is not to say that non-human animals can perceive the same detailed sound distinctions that humans can, but it indicates the presence of taxonomically wide spread and sometimes quite advanced precursors of human sensitivities. It is also clear that the nature of the stimuli, the experimental paradigms, and the animal species concerned, can strongly affect the experimental outcomes (Hoeschele, 2017). Clearly, the topic needs more species to be studied before general conclusions can be made.

29.3.2 Animal perception of temporal dimensions

29.3.2.1 Rhythm and tempo generalization

The rhythmic behaviours exhibited by different species vary wildly, from humans dancing to the beat of the music, the rhythmic swaying of chimpanzees to bird vocalizations or cricket chirps containing precisely timed rhythmic patterns (Bouwer et al., 2021). Many birds (e.g., collared doves; Slabbekoorn & ten Cate, 1999), crickets and frogs (Gerhardt & Huber, 2002) are also sensitive to disruptions of species-specific rhythmic patterns. However, a sensitivity to modifications of the rhythmic pattern of species-specific vocalizations need not indicate a sensitivity to rhythms in general, constructed from artificial sounds. To study rhythmic perception, it might be helpful to decompose a rhythmic signal in its constituent components, such as rhythmic pattern, metrical structure, tempo and timing (Honing & Bouwer, 2019). With regard to *rhythmic patterns* most

animals can discriminate between regular (isochronous) and irregular (non-isochronous) sequences constructed from identical sound pulses. Rats (Celma-Miralles & Toro, 2020) and starlings (Hulse et al., 1984; Humpal & Cynx, 1984) show tempo generalizations when having to differentiate between regular and irregular test sequences at rates that were not trained. This suggests that they base their discrimination on the whole sequence, and are sensitive to the stimulus being isochronous (and not just memorize a single time interval). However, in contrast, bird species like zebra finches (van der Aa et al., 2015, but see Rouse et al., 2021) and pigeons (Hagmann & Cook, 2010) seem to be less capable of tempo generalizations. Zebra finches can discriminate between isochronous and non-isochronous sequences, but generalize less to sequences that are presented at another tempo (or rate), while pigeons appear not to be able to discriminate at all between isochronous and non-isochronous sequences. They might respond to the absolute duration of specific intervals, hence to local cues rather than to the global pattern regularity, to discriminate a regular from an irregular series of pulses (ten Cate et al., 2016).

Interestingly, jackdaws are not only able to distinguish between two rhythmic patterns, but are also able to generalize to other tempi, and are capable of maintaining the discrimination when the sounds making up the patterns were changed (varying timbre or pitch), suggesting that the jackdaws recognize relative and global rhythmic patterns (Reinert, 1965). This is in contrast to zebra finches and budgerigars (ten Cate et al., 2016), who seem to have a limited ability to use the rhythmic structure to distinguish between stimuli.

29.3.2.2 Metrical structure and beat perception

Metrical structure is present in both language and music. Surprisingly, the perception of metrical regularity in a varying rhythm is only shown in a few animals (Fitch, 2013b). While beat and meter perception are considered fundamental to musicality (Honing, 2012), these are difficult to trace back in the animal world. In the few species that are studied, it appears to be mostly vocal learners that are sensitive to the *beat* of the music: a regular pulse to which one can dance and/or synchronize. Note that a beat can be induced in a listener by a time-varying, non-isochronous rhythm, suggesting that beat processing is a cognitive rather than an acoustic or sensory phenomenon (Honing, 2012). Seminal examples are a sulphur-crested cockatoo (Patel et al., 2009) and a grey parrot (Schachner et al., 2009) that are capable of synchronizing to the beat of human music and, importantly, maintaining synchrony (e.g., by head-bobbing in sync with the beat) when the same music is played at a different tempo. The observation that this behaviour was initially only shown in vocal learning species (Schachner et al., 2009) gave rise to the ‘vocal learning and rhythmic synchronization’ hypothesis (Patel, 2006, 2021) suggesting that a tight neural connection between auditory and motor areas, as is the case in vocal learners, is a prerequisite for beat perception and synchronization. Interestingly, however, a California sea lion (not considered a vocal learner) was also able to synchronize head movements to a variety of musical fragments, as well as showing generalization over different tempi (Cook et al., 2013; Rouse et al., 2016). Hence, it seems that perceiving a beat in a complex stimulus (i.e. music) and being able to synchronize to it, is not restricted to humans and parrots, and might well be more wide-spread than previously thought (Wilson & Cook, 2016). While only recently experimental research is finding some support for this claim, within the primates, only humans show clear evidence for beat perception (Winkler et al., 2009). For example, while monkeys are sensitive to the isochrony of a rhythmic stimulus, they lack beat perception (Honing et al., 2012, 2018). One possible interpretation is that beat perception and synchronization evolved gradually

within the primates, peaking in humans and present only with limitations in chimpanzees (Hattori & Tomonaga, 2019), bonobos (Large & Gray, 2015), and other non-human primates (Honing & Merchant, 2014; Merchant & Honing, 2014).

29.3.3 Animal perception of higher order acoustic patterns

To be able to detect arbitrary rhythmic patterns, sense isochrony, perceive a beat in a varying rhythm or recognize a melody, can be seen as higher order cognitive abilities, which require the detection of patterns over a series of sounds, to some extent irrespective of the precise nature of the sounds themselves. These abilities may also include phenomena like perceptual grouping in which subsequent sound units are together identified as a higher order sound unit, the detection of a melody irrespective of the instruments playing it (i.e., spectral quality), the detection of intonation patterns (prosody) in language irrespective of what is said, up to advanced abilities like the use of syntax rules as present in languages. For all of these phenomena, demonstrated to be present in humans and providing basal building blocks for the development and use of language and the production and appreciation of music, there is evidence of at least some precursors being present in non-human species.

29.3.3.1 Perceptual grouping

Gestalt psychologists noted already long ago that a continuously repeated series of two tones which differ in frequency, amplitude or duration are systematically perceived (grouped) as a series of duplets. When subsequent tones differ only in duration they are perceived as a duplet which ends with the longer tone: an iambic pattern. When the tones differ in frequency or amplitude, the higher pitched or loudest one is perceived as the initial one of a duplet: a trochaic pattern. This type of perceptual grouping is common to music as well as speech perception (Iversen et al., 2008). Young infants spontaneously recognize trochaic patterns and attention to such patterning may act as a cue to learning linguistic structures (Mueller et al., 2020). Both rats (de la Mora et al., 2013) and zebra finches (Spierings et al., 2017) also show such a grouping bias. Although this indicates that perceptual grouping is not limited to humans, to establish whether this bias is wide spread awaits testing in other species, as does the question whether other grouping biases identified for humans are present in non-human animals.

29.3.3.2 Prosodic patterns

Acoustic cues can provide key information for interpreting linguistic utterances. Intonation patterns can indicate whether a sentence is perceived as a statement or a question, while the lexical stress pattern may differentiate between word categories, for example, between the noun 'PREsent' and the verb 'preSENT' in English. Also different languages differ in intonation or stress patterns. A variety of experiments demonstrated the ability of various animals to distinguishing different human speech utterances by their prosodic patterns. Cotton-top tamarins (Ramus et al., 2000), rats (Toro et al., 2003), Java sparrows (Watanabe et al., 2006), zebra finches (Spierings & ten Cate, 2014) and budgerigars (Hoeschele & Fitch, 2016) can use prosodic patterns, for instance to distinguish between a question or a statement in the same language, or the distinction between sentences in one language from that in another one with a different prosodic pattern (see Mol et al., 2017 and Mueller et al., 2020, for reviews). More generally, a sensitivity to prosodic variation is likely to be wide spread among animals. This is based on a homology between humans and many animal species as their vocalizations can be modulated in similar ways by their arousal or emotional state (Briefer, 2018;

Filippi, 2016). This produces similar vocal prosodic patterns across species and is linked to being sensitive to such variations. Such sensitivity to emotional prosody may have been at the basis for the evolution of linguistic prosodic patterns. As noted by Huron (2015), the emotional responses that humans have to some types of musical patterns may also find its basis in this sensitivity (see also Juslin & Laukka, 2003). Identifying such patterns provides a potentially useful strategy for discerning the musical aspects of vocalizations that are likely shared across species.

29.3.3.3 Pattern perception and rule learning

An important feature of language is the presence of syntax, a series of grammar rules which determine the structuring, and thereby the meaning, of sentences. A young infant will implicitly learn the grammatical rules of the language spoken around it. The rules present in any specific language may differ in their level of complexity. These levels have been described formally as a hierarchical system; the Chomsky hierarchy. While some animal vocalizations, in particular songs, may have a seemingly complex structure in which different element types are produced in a long and varied sequence, there is consensus that the 'rules' that characterize the element sequencing in various species are limited in complexity compared to those of language as judged by the Chomsky hierarchy (Berwick et al., 2011). However, the finding that the structure of the vocalizations of a species is constrained in complexity need not indicate that the perception of regularities or sequencing rules in strings of acoustic elements is also constrained (ten Cate, 2017). These perceptual abilities have been tested in a range of species, using a paradigm called 'artificial grammar learning' (AGL), in which arbitrary sound items are organized according to particular patterns, 'grammars'. After exposing an animal to example strings of a particular type, it is next tested with novel items or sequence violations to assess whether the animal has learned the underlying pattern. This has shown that various songbird species (e.g., starlings, zebra finches), primates (marmosets, macaques) and rats can readily learn to detect adjacent dependencies (i.e. whether particular sound items are systematically preceded or followed by another particular type). Some species can also detect dependencies with varying transitional probabilities among items or even non-adjacent dependencies between different items, in which two linked items are separated by one or more arbitrary intervening items (reviewed by Petkov & ten Cate, 2020). More complex rules, in particular those requiring abstraction of a pattern, such as identifying structures as XXYY, XYXY, XYX, etc., irrespective of the nature of the X and Y type of sounds, are clearly a more difficult problem to many species. For instance, zebra finches trained to distinguish XYX from XXY strings using one set of sound items and next tested with novel sounds items arranged in the same structure failed to classify these novel strings according to their underlying structure. They had learned to distinguish the training sounds by memorizing them all separately. However, using the same stimuli and procedure, budgerigars did classify the novel strings by whether they had an XYX or XXY structure (Spierings & ten Cate, 2016), thus revealing a clear species difference for this ability. Another ability sometimes claimed to be present in animal species is that of detecting recursive patterns (e.g., Abe & Watanabe, 2011; Gentner et al., 2006), in which one structure is embedded within a similar one, such as in the human sentence 'the starling, the cat observed, flew away'. However, subsequent studies have demonstrated that these claims were premature and that the results can be explained without referring to recursive processing (Beckers et al., 2012; Van Heijningen et al., 2009). Thus, while a range of studies have tested for this and similar abilities in various species and several claimed the existence of learning syntactic rules, the tests done often leave open whether the results show unambiguous evidence of higher order processing abilities (ten Cate & Petkov, 2019, but see Ferrigno et al., 2020). Therefore, while several species

show a basal ability to process strings based on various transitional abilities, evidence for detecting more abstract sequencing rules from exposure to example strings is still limited and if present still far from the syntactic abilities of humans.

29.4 Conclusion

What our overview shows is that although human language and music have no direct parallel among non-human animals, many animal species show evidence of the presence of one or more of the constituent cognitive components on which these human abilities are based, or presence of abilities that may have served as precursors for these components. This includes a sensitivity to various spectral dimensions, the ability to detect specific rhythmic patterns, sense isochrony or perceive a beat in a varying rhythm, the perception of relative pitch, the ability for perceptual grouping, the detection of intonation patterns irrespective of its actual sound elements, up to advanced abilities like the detection of (a rudimentary) linguistic syntax. For all of these phenomena there is evidence for at least some precursor being present in some non-human species. There are thus some clear similarities between humans and non-human animals with respect to features that can be considered central in the capacity for language and music. At the same time, there are also important differences among animal species, for instance in the ability to detect rhythms or to learn a syntactic rule. Hence contrasts are not limited to those between humans and other animal species. This is in itself not surprising, as species differ in many ways. Therefore, and in contrast to some of the older literature on the capacity for language and music, which stresses human uniqueness, we argue that it could well be that many animal species (including humans) in fact share a similar basis for many of the perceptual and cognitive abilities to distinguish between or to generalize auditory stimuli. The differences that are observed between humans and other species, as well as those among non-human species, might be a result of specific biases and the weight or priority certain species give to attending to certain features of an acoustic signal, or because different species use particular mechanisms to different degrees. Also, the similarities between humans and non-human animals may have been obscured because the specific experimental procedures and stimuli used usually focus on one specific parameter, ignoring that other parameters may also be perceived and affect what individuals will attend to (ten Cate et al., 2016). Another key problem in the search for the precursors of language and music in animals is, simply, that too few species have been investigated in a too diverse range of paradigms to arrive at a solid interpretation of, and generalization across, the existing findings. All of this also makes some of our own conclusions preliminary or even premature. Hence, the challenge for future research is to overcome this problem and we expect that, a) if the stimuli used are more diverse and richer in their sound quality, b) if procedures allow for testing the presence of less preferred but nevertheless available perceptual mechanisms in listeners (humans and animals alike), and c) if a wider array of species is tested using more similar procedures, the differences between human, mammalian and avian cognition might well turn out to be smaller, and more one of degree rather than of kind, than is currently suggested in the literature.

References

- Abe, K. & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, *14*(8), 1067–1074. <https://doi.org/10.1038/nn.2869>
- Albouy, P., Benjamin, L., Morillon, B. & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody (Supplementary Materials). *Science*, *367*(6481), 1043–1047. <https://doi.org/10.1126/science.aaz3468>
- Arbib, M. A. (Ed.). (2013). *Language, Music, and The Brain: A Mysterious Relationship*. MIT Press.
- Babiszewska, M., Schel, A. M., Wilke, C. & Slocombe, K. E. (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. <https://doi.org/10.1002/ajpa.22634>
- Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. (2012). Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *NeuroReport*, *23*(3), 139–145. <https://doi.org/10.1097/WNR.0b013e32834f1765>
- Berwick, R. C., Okanoya, K., Beckers, G. J. L. & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. <http://www.sciencedirect.com/science/article/pii/S1364661311000039>
- Bizley, J. K., Walker, K. M. M., King, A. J. & Schnupp, J. W. H. (2013). Spectral timbre perception in ferrets: Discrimination of artificial vowels under different listening conditions. *The Journal of the Acoustical Society of America*, *133*(1), 365–376. <https://doi.org/10.1121/1.4768798>
- Bouwer, F. L., Nityananda, V., Rouse, A. A. & ten Cate, C. (2021). Rhythmic abilities in humans and non-human animals: a review and recommendations from a methodological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1835), 20200335. <https://doi.org/10.1098/rstb.2020.0335>
- Bowling, D. L., Hoeschele, M., Gill, K. Z. & Fitch, W. T. (2017). The nature and nurture of musical consonance. *Music Perception*, *335*(1), 118–121.
- Bradbury, J. W. & Vehrencamp, S. L. (2011). *Principles of Animal Communication, 2nd Edition*. Sinauer Associates.
- Bregman, A. S. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound*. MIT Press.
- Bregman, M. R., Patel, A. D. & Gentner, T. Q. (2012). Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition*, *122*(1), 51–60. <https://doi.org/10.1016/j.cognition.2011.08.008>
- Bregman, M. R., Patel, A. D. & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences*, *113*(6), 1666–1671. <https://doi.org/10.1073/pnas.1515380113>
- Briefer, E. F. (2018). Vocal contagion of emotions in non-human animals. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873). <https://doi.org/10.1098/rspb.2017.2783>
- Burdick, C. K. & Miller J D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*, *58*, 415–427.
- Burgering, M. A., ten Cate, C. & Vroomen, J. (2018). Mechanisms underlying speech sound discrimination and categorization in humans and zebra finches. *Animal Cognition*, *21*(2), 285–299. <https://doi.org/10.1007/s10071-018-1165-3>

- Burgering, M. A., Vroomen, J. & ten Cate, C. (2019). Zebra finches (*Taeniopygia guttata*) can categorize vowel-like sounds on both the fundamental frequency (“pitch”) and spectral envelope. *Journal of Comparative Psychology*, 133(1), 106–117. <https://doi.org/10.1037/com0000143>
- Catchpole, C. K. & Slater, P. J. B. (2008). *Bird song; Biological themes and variations* (2nd Editio). Cambridge University Press.
- Celma-Miralles, A. & Toro, J. M. (2020). Discrimination of Temporal Regularity in Rats (*Rattus norvegicus*) and Humans (*Homo sapiens*). *Journal of Comparative Psychology*, 134(1), 3–10. <https://doi.org/10.1037/com0000202>
- Chiandetti, C. & Vallortigara, G. (2011). Chicks like consonant music. *Psychological Science*, 22(10), 1270–1273. <https://doi.org/10.1177/0956797611418244>
- Cook, P. F., 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>
- Cynx, J. (1993). Auditory frequency generalization and a failure to find octave generalization in a songbird, the European starling (*Sturnus vulgaris*). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, 107(2), 140–146. <https://doi.org/10.1037/0735-7036.107.2.140>
- Cynx, J. & Shapiro, M. (1986). Perception of missing fundamental by a species of songbird (*Sturnus vulgaris*). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, 100(4), 356–360. <http://www.ncbi.nlm.nih.gov/pubmed/3802780>
- De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A. & Gamba, M. (2021). Categorical rhythms in a singing primate. *Current Biology*, 31(20), R1379–R1380. <https://doi.org/10.1016/j.cub.2021.09.032>
- de la Mora, D. M., Nespore, M. & Toro, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic-trochaic law? *Attention, Perception & Psychophysics*, 75(1), 92–100. <https://doi.org/10.3758/s13414-012-0371-3>
- Desain, P. W. M. & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3), 341–365. <https://doi.org/10.1068/p3370>
- Dodenhoff, D. J., Stark, R. D. & Johnson, E. V. (2001). Do woodpecker drums encode information for species recognition? *Condor*, 103(1), 143–150. [https://doi.org/10.1650/0010-5422\(2001\)103\[0143:DWDEIF\]2.0.CO;2](https://doi.org/10.1650/0010-5422(2001)103[0143:DWDEIF]2.0.CO;2)
- Dooling, R. J. & Brown, S. D. (1990). Speech perception by budgerigars (*Melopsittacus undulatus*): Spoken vowels. *Perception & Psychophysics*, 47(6), 568–574. <https://doi.org/10.3758/BF03203109>
- Doolittle, E. L., Gingras, B., Endres, D. M. & Fitch, W. T. (2014). Overtone-based pitch selection in hermit thrush song: Unexpected convergence with scale construction in human music. *Proceedings of the National Academy of Sciences*, 11(46), 1–6. <https://doi.org/10.1073/pnas.1406023111>
- Ferrigno, S., Cheyette, S. J., Piantadosi, S. T. & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians. *Science Advances*, 6(26), 1–11. <https://doi.org/10.1126/sciadv.aaz1002>
- Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: A

- comparative approach to the emergence of language. In *Frontiers in Psychology* (Vol. 7, pp. 1–19). <https://doi.org/10.3389/fpsyg.2016.01393>
- Fischer, J. & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience and Biobehavioral Reviews*, *82*, 22–31. <https://doi.org/10.1016/j.neubiorev.2016.10.014>
- Fishbein, A. R., Fritz, J. B., Idsardi, W. J. & Wilkinson, G. S. (2020). What can animal communication teach us about human language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789). <https://doi.org/10.1098/rstb.2019.0042>
- Fitch, W. T. (2005). The Evolution of Language: A Comparative Review. In *Biology & Philosophy* (Vol. 20, Issues 2–3). <https://doi.org/10.1007/s10539-005-5597-1>
- Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition*, *100*(1), 173–215. <https://doi.org/10.1016/j.cognition.2005.11.009>
- Fitch, W. T. (2013a). Musical protolanguage: Darwin’s theory of language evolution revisited. In J. J. Bolhuis & M. Everaert (Eds.), *Birdsong, speech, and language: Exploring the evolution of mind and brain* (pp. 489–503). MIT Press.
- Fitch, W. T. (2013b). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, *7*, 1–16. <https://doi.org/10.3389/fnsys.2013.00068>
- Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*(7088), 1204–1207. <https://doi.org/10.1038/nature04675>
- Gerhardt, H. C. & Huber, F. (2002). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. University of Chicago Press.
- Gingras, B., Honing, H., Peretz, I., Trainor, L. J. & Fisher, S. E. (2015). Defining the Biological Bases of Individual Differences in Musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 221–250. <https://doi.org/10.1098/rstb.2014.0092>
- Hage, S. R., Gavrilov, N. & Nieder, A. (2016). Developmental changes of cognitive vocal control in monkeys. *Journal of Experimental Biology*, *219*(11), 1744–1749. <https://doi.org/10.1242/jeb.137653>
- Hagmann, C. & Cook, R. G. (2010). Testing meter, rhythm, and tempo discriminations in pigeons. *Behavioural Processes*. <https://doi.org/10.1016/j.beproc.2010.06.015>
- Hagoort, P. (Ed.). (2019). *Human Language: From Genes and Brains to Behavior*. The MIT Press.
- Harrison, P. M. C. (2021). Three Questions concerning Consonance Perception. In *Music Perception* (Vol. 38, Issue 3, pp. 337–339). University of California Press. <https://doi.org/10.1525/MP.2021.38.3.337>
- Haspelmath, M. (2020). Human Linguisticity and the Building Blocks of Languages. *Frontiers in Psychology*, *10*(January), 1–10. <https://doi.org/10.3389/fpsyg.2019.03056>
- Hattori, Y. & Tomonaga, M. (2019). Rhythmic swaying induced by sound in chimpanzees (Pan troglodytes). *Proc Natl Acad Sci U S A*, *117*(2), 936–942. <https://doi.org/10.1073/pnas.1910318116>
- He, C. & Trainor, L. J. (2009). Finding the pitch of the missing fundamental in infants. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(24), 7718–8822. <https://doi.org/10.1523/JNEUROSCI.0157-09.2009>

- Heffner, H. & Whitfield, I. C. (1976). Perception of the missing fundamental by cats. *Journal of the Acoustical Society of America*, 59(4).
- Heimbauer, L. A., Beran, M. J. & Owren, M. J. (2011). A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content. *Current Biology*, 21(14), 1210–1214. <https://doi.org/10.1016/j.cub.2011.06.007>
- 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. *Sci. Adv.*, 3. <https://doi.org/10.1126/sciadv.1602399>
- Helmholtz, H. (1954). *On the sensations of tone*. Dover Publications.
- Henry, K. S., Amburgey, K. N., Abrams, K. S., Idrobo, F. & Carney, L. H. (2017). Formant-frequency discrimination of synthesized vowels in budgerigars (*Melopsittacus undulatus*) and humans . *The Journal of the Acoustical Society of America*, 142(4), 2073–2083. <https://doi.org/10.1121/1.5006912>
- Hoeschele, M. (2017). Animal Pitch Perception: Melodies and Harmonies. *Comparative Cognition and Behavior Reviews*, 12(1), 5–18. <https://doi.org/10.3819/CCBR.2017.120002>
- Hoeschele, M., Cook, R. G., Guillette, L. M., Brooks, D. I. & Sturdy, C. B. (2012). Black-capped chickadee (*Poecile atricapillus*) and human (*Homo sapiens*) chord discrimination. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, 126(1), 57–67. <https://doi.org/10.1037/a0024627>
- Hoeschele, M., Cook, R. G., Guillette, L. M., Hahn, A. H. & Sturdy, C. B. (2014). Timbre influences chord discrimination in black-capped chickadees (*Poecile atricapillus*) but not humans (*Homo sapiens*). *Journal of Comparative Psychology*, 128(4), 387–401. <https://doi.org/10.1037/a0037159>
- Hoeschele, M. & Fitch, W. T. (2016). Phonological perception by birds: budgerigars can perceive lexical stress. *Animal Cognition*, 19(3), 643–654. <https://doi.org/10.1007/s10071-016-0968-3>
- Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y. & ten Cate, C. (2018). Searching for the origins of musicality across species. In H. Honing (Ed.), *The Origins of Musicality* (pp. 149–170). MIT Press. <https://doi.org/10.1098/rstb.2014.0094>
- Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H. & Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Animal Cognition*, 16(4), 599–609. <https://doi.org/10.1007/s10071-013-0597-z>
- Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, 1252(1), 85–91. <https://doi.org/10.1111/j.1749-6632.2011.06402.x>
- Honing, H. (Ed.). (2018). *The Origins of Musicality*. The MIT Press. <https://mitpress.mit.edu/books/origins-musicality>
- Honing, H. (2021). Unravelling the origins of musicality: Beyond music as an epiphenomenon of language. *Behavioral and Brain Sciences*, 44(E78), 66–69. <https://doi.org/10.1017/S0140525X20001211>
- Honing, H. & Bouwer, F. L. (2019). Rhythm. In P. J. Rentfrow & D. Levitin (Eds.), *Foundations in Music Psychology: Theory and Research* (pp. 33–70). The MIT Press.
- Honing, H., Bouwer, F. L., Prado, L. & Merchant, H. (2018). Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat. *Frontiers in Neuroscience*, 12(475).

<https://doi.org/10.3389/fnins.2018.00475>

- Honing, H. & Merchant, H. (2014). Differences in auditory timing between human and non-human primates. *Behavioral and Brain Sciences*, 27(6), 557–558.
<https://doi.org/10.1017/S0140525X13004056>
- Honing, H., Merchant, H., Háden, G. P., Prado, L. & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE*, 7(12), 1–10.
<https://doi.org/10.1371/journal.pone.0051369>
- Honing, H., ten Cate, C., Peretz, I. & Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1664), 20140088. <https://doi.org/10.1098/rstb.2014.0088>
- Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A. & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America*, 106(3), 1532–1542. <https://doi.org/10.1121/1.427150>
- Hulse, S. H., Bernard, D. J. & Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology - General*, 124(4), 409–423. <https://doi.org/10.1037/0096-3445.124.4.409>
- Hulse, S. H., Humpal, J. & Cynx, J. (1984). Processing of rhythmic sound structures by birds. *Annals of the New York Academy of Sciences*, 423, 407–419.
<http://www.ncbi.nlm.nih.gov/pubmed/6588804>
- Humpal, J. & Cynx, J. A. (1984). Discrimination of Temporal Components of Acoustic Patterns by Birds. *Annals of the New York Academy of Sciences*, 423(1), 600–602.
<https://doi.org/10.1111/j.1749-6632.1984.tb23466.x>
- Huron, D. (2015). Affect induction through musical sounds : an ethological perspective. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370(1664).
- Iversen, J. R., Patel, A. D. & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America*, 124(4), 2263–2271.
<https://doi.org/10.1121/1.2973189>
- Izumi, A. (2000). Japanese monkeys perceive sensory consonance of chords. *Journal of the Acoustical Society of America*, 108(6), 3073–3078.
- Jacoby, N. & McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, 27, 359–370.
<https://doi.org/10.1016/j.cub.2016.12.031>
- Janik, V. M., Knörnschild, M. & Janik, V. M. (2021). *Vocal production learning in mammals revisited*.
- Johnson, K. E. & Clark, C. J. (2020). Ontogeny of vocal learning in a hummingbird. *Animal Behaviour*, 167, 139–150. <https://doi.org/10.1016/j.anbehav.2020.07.010>
- Juslin, P. & Laukka, P. (2003). Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code? *Psychological Bulletin*, 129(5), 770–814.
- Kaminski, J., Call, J. & Fischer, J. (2004). Word learning in a domestic dog: Evidence for “fast mapping.” *Science*, 304(5677), 1682–1683. <https://doi.org/10.1126/science.1097859>
- Kluender, K. R., Diehl, R. L., Killeen, P. R., Fitzgerald, P. G., Sandiford, M. & Barrett, P. J. (2016). Japanese quail can learn phonetic categories. *Science*, 237(4819), 1195–1197.
- Large, E. W. & Gray, P. M. (2015). Spontaneous Tempo and Rhythmic Entrainment in a Bonobo.

- Journal of Comparative Psychology*. <https://doi.org/10.1037/com0000011>
- Levitin, D. J. & Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, 9(1), 26–33. <https://doi.org/10.1016/j.tics.2004.11.007>
- MacDougall-Shackleton, S. A. & Hulse, S. H. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, 110(2), 139–146. <https://doi.org/10.1037//0735-7036.110.2.139>
- McAdams, S. & Siedenburg, K. (2019). Perception and cognition of musical timbre. In P. Rentfrow & D. J. Levitin (Eds.), *Foundations of Music Psychology: Theory and Research* (pp. 71–120). MIT Press. https://www.mcgill.ca/mpcl/files/mpcl/mcadams_2019_foundmuspsychol.pdf
- McDermott, J. H. & Hauser, M. D. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition*, 94(2), B11-21. <https://doi.org/10.1016/j.cognition.2004.04.004>
- McDermott, J. H., Lehr, A. J. & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, 19(12), 1263–1271. <http://pss.sagepub.com/content/19/12/1263.short>
- McDermott, J. H., Schultz, A. F., Undurraga, E. A. & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, 535, 21–25. <https://doi.org/10.1038/nature18635>
- Merchant, H. & Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7(8 JAN), 1–8. <https://doi.org/10.3389/fnins.2013.00274>
- Mol, C., Chen, A., Kager, R. W. J. & ter Haar, S. M. (2017). Prosody in birdsong: A review and perspective. *Neuroscience and Biobehavioral Reviews*, 81, 167–180. <https://doi.org/10.1016/j.neubiorev.2017.02.016>
- Mueller, J. L., ten Cate, C. & Toro, J. M. (2020). A Comparative Perspective on the Role of Acoustic Cues in Detecting Language Structure. *Topics in Cognitive Science*, 12(3), 859–874. <https://doi.org/10.1111/tops.12373>
- Nagel, K. I. K., McLendon, H. M. & Doupe, A. J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of Neurophysiology*, 104(3), 1426–1437. <https://doi.org/10.1152/jn.00028.2010>
- Norman-Haignere, S., Kanwisher, N. G. & McDermott, J. H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, 88(6), 1281–1296. <https://doi.org/10.1016/j.neuron.2015.11.035>
- Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., Schalk, G. & Kanwisher, N. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, 1–15. <https://doi.org/10.1016/j.cub.2022.01.069>
- Ohms, V. R., Gill, A., Heijningen, C. A. A. Van, Beckers, G. J. L. & Ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings. Biological Sciences / The Royal Society*, 277(1684), 1003–1009. <https://doi.org/10.1098/rspb.2009.1788>
- Patel, A. D. (2006). Musical Rhythm, Linguistic Rhythm, and Human Evolution. *Music Perception*, 24(1), 99–104. <https://doi.org/10.1525/mp.2006.24.1.99>
- Patel, A. D. (2017). Why Doesn't a Songbird (the European Starling) Use Pitch to Recognize Tone Sequences? The Informational Independence Hypothesis. *Comparative Cognition and Behavior Reviews*, 12(1), 19–32. <https://doi.org/10.3819/CCBR.2017.120003>

- Patel, A. D. (2021). Vocal learning as a preadaptation for the evolution of human beat perception and synchronization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1835). <https://doi.org/10.1098/rstb.2020.0326>
- Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, 19(10), 827–830. <https://doi.org/10.1016/j.cub.2009.03.038>
- Payne, R. S. & McVay, S. (1971). Songs of humpback whales. *Science*, 173(3997), 585. <http://www.sciencemag.org/content/173/3997/585.short>
- Pepperberg, I. M. (1999). *The Alex studies. Cognitive and communicative abilities of grey parrots*. Harvard University Press.
- Peretz, I., Vuvar, D. T., Armony, J. L., Lagrois, M.-É. & Armony, J. L. (2018). Neural overlap in processing music and speech. In H. Honing (Ed.), *The Origins of Musicality* (Vol. 370, Issue 1664, pp. 205–220). The MIT Press. <https://doi.org/http://dx.doi.org/10.1098/rstb.2014.0090>
- Petkov, C. I. & ten Cate, C. (2020). Structured Sequence Learning: Animal Abilities, Cognitive Operations, and Language Evolution. *Topics in Cognitive Science*, 12(3), 828–842. <https://doi.org/10.1111/tops.12444>
- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K. & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, 5, 1–11. <https://doi.org/10.1038/srep13220>
- Purves, D. (2017). *Music as biology. The tones we like and why*. Harvard University Press.
- Ramus, F., Hauser, M. D., Miller, C., Morris, D. & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464), 349–351. <https://doi.org/10.1126/science.288.5464.349>
- Randall, J. A. (2001). Evolution and function of drumming as communication in mammals'. *American Zoologist*, 41(5), 1143–1156. <https://doi.org/10.1093/icb/41.5.1143>
- Rehsteiner, U., Geisser, H. & Reyer, H. U. (1998). Singing and mating success in water pipits: One specific song element makes all the difference. *Animal Behaviour*, 55(6), 1471–1481. <https://doi.org/10.1006/anbe.1998.0733>
- Reinert, J. (1965). Takt- und Rhythmusunterscheidung bei Dohlen. *Zeitschrift Für Tierpsychologie*, 22(6), 623–671. <https://doi.org/10.1111/j.1439-0310.1965.tb01683.x>
- Richards, D. G., Wolz, J. P. & Herman, L. M. (1984). Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, 98(1), 10–28. <https://doi.org/10.1037/0735-7036.98.1.10>
- Richner, H. (2016). Interval singing links to phenotypic quality in a songbird. *Proceedings of the National Academy of Sciences of the United States of America*, 113(45), 201610062. <https://doi.org/10.1073/pnas.1610062113>
- Root-Gutteridge, H., Ratcliffe, V. F., Korzeniowska, A. T. & Reby, D. (2019). Dogs perceive and spontaneously normalize formant-related speaker and vowel differences in human speech sounds. *Biology Letters*, 15(12). <https://doi.org/10.1098/rsbl.2019.0555>
- Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M. & Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hearing Research*, 308, 71–83. <https://doi.org/10.1016/j.heares.2013.08.016>

- Rouse, A. A., Cook, P. F., Large, E. W. & Reichmuth, C. (2016). Beat Keeping in a Sea Lion As Coupled Oscillation: Implications for Comparative Understanding of Human Rhythm. *Frontiers in Neuroscience*, 10(September). <https://doi.org/10.3389/fnins.2016.00257>
- Rouse, A. A., Patel, A. D. & Kao, M. H. (2021). Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, 118(29). <https://doi.org/10.1073/pnas.2026130118>
- Salmi, R., Jones, C. E. & Carrigan, J. (2021). Who is there? Captive western gorillas distinguish human voices based on familiarity and nature of previous interactions. *Animal Cognition*, 0123456789. <https://doi.org/10.1007/s10071-021-01543-y>
- Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D. & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General*, 115(3), 211–235. <https://doi.org/10.1037/0096-3445.115.3.211>
- Schachner, A., Brady, T. F., Pepperberg, I. M. & Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Current Biology : CB*, 19(10), 831–836. <https://doi.org/10.1016/j.cub.2009.03.061>
- Schellenberg, E. G. & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, 14(3), 262–266. <http://pss.sagepub.com/content/14/3/262.short>
- Searcy, W. A., Soha, J., Peters, S. & Nowicki, S. (2021). Variation in vocal production learning across songbirds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1836). <https://doi.org/10.1098/rstb.2020.0257>
- Seyfarth, R. M., Cheney, D. L. & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210(4471), 801–803. <https://doi.org/10.1126/science.7433999>
- Shannon, R. V. (2016). Is birdsong more like speech or music? *Trends in Cognitive Sciences*, 20(4), 245–247. <https://doi.org/10.1016/j.tics.2016.02.004>
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J. & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304. <https://doi.org/10.1126/science.270.5234.303>
- Shepard, R. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, 36, 2346–2353. <https://doi.org/10.1121/1.1919362>
- Slabbekoorn, H. & Ten Cate, C. (1999). Collared Dove Responses to Playback: Slaves to the Rhythm. *Ethology*, 105(5), 377–391. <https://doi.org/10.1046/j.1439-0310.1999.00420.x>
- Slocombe, K. E. & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology*, 15(19), 1779–1784. <https://doi.org/10.1016/j.cub.2005.08.068>
- Spierings, M. J., Hubert, J. & ten Cate, C. (2017). Selective auditory grouping by zebra finches: testing the iambic–trochaic law. *Animal Cognition*, 20(4), 665–675. <https://doi.org/10.1007/s10071-017-1089-3>
- Spierings, M. J. & ten Cate, C. (2014). Zebra finches are sensitive to prosodic features of human speech. *Proceedings of the Royal Society B: Biological Sciences*, 281(1787), 1–7. <https://doi.org/10.1098/rspb.2014.0480>
- Spierings, M. J. & Ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *PNAS*. <https://doi.org/10.1073/pnas.1600483113>

- Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriya, Y., Takeshita, H., Nakamura, T. & Hashiya, K. (2010). Preference for consonant music over dissonant music by an infant chimpanzee. *Primates*, 51(1), 7–12. <https://doi.org/10.1007/s10329-009-0160-3>
- ten Cate, C. (2017). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic Bulletin & Review*, 24, 91–96. <https://doi.org/10.3758/s13423-016-1091-9>
- ten Cate, C. (2021). Re-evaluating vocal production learning in non-oscine birds. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 376(20200249).
- ten Cate, C. & Healy, S. (Eds.). (2017). *Avian Cognition*. Cambridge University Press.
- ten Cate, C. & Petkov, C. I. (2019). The grammatical abilities of animals: a comparative overview. In P. Haagoort (Ed.), *Human Language* (pp. 687–700). MIT Press.
- ten Cate, C., Spierings, M. J., Hubert, J. & Honing, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Frontiers in Psychology*, 7(May), 1–14. <https://doi.org/10.3389/fpsyg.2016.00730>
- Tomlinson, R. W. W. & Schwarz, D. W. F. (1988). Perception of the missing fundamental in nonhuman primates. *Journal of the Acoustical Society of America*, 84(2).
- Toro, J. M., Trobalon, J. B. & Sebastián-Gallés, N. (2003). The use of prosodic cues in language discrimination tasks by rats. *Animal Cognition*, 6(2), 131–136. <https://doi.org/10.1007/s10071-003-0172-0>
- Tyack, P. L. (2020). A taxonomy for vocal learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1789). <https://doi.org/10.1098/rstb.2018.0406>
- Vallet, E. & Kreutzer, M. (1995). Female canaries are sexually responsive to special song phrases. *Animal Behaviour*, 49(6), 1603–1610. [https://doi.org/10.1016/0003-3472\(95\)90082-9](https://doi.org/10.1016/0003-3472(95)90082-9)
- van der Aa, J., Honing, H. & ten Cate, C. (2015). The perception of regularity in an isochronous stimulus in zebra finches (*Taeniopygia guttata*) and humans. *Behavioural Processes*, 115, 37–45. <https://doi.org/10.1016/j.beproc.2015.02.018>
- Van Heijningen, C. A. A., De Visser, J., Zuidema, W. & Ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), 20538–20543. <https://doi.org/10.1073/pnas.0908113106>
- 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. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 376, pp. 1–15). <https://doi.org/10.1098/rstb.2020.0236>
- Wagner, B., Bowling, D. L. & Hoeschele, M. (2020). Is consonance attractive to budgerigars? No evidence from a place preference study. *Animal Cognition*, 23(5), 973–987. <https://doi.org/10.1007/s10071-020-01404-0>
- Wagner, B., Mann, D. C., Afroozeh, S., Staubmann, G. & Hoeschele, M. (2019). Octave equivalence perception is not linked to vocal mimicry: Budgerigars fail standardized operant tests for octave equivalence. *Behaviour*, 156(5–8), 479–504. <https://doi.org/10.1163/1568539X-00003538>
- Warren, J. D., Uppenkamp, S., Patterson, R. D. & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 100(17), 10038–10042. <https://doi.org/10.1073/pnas.1730682100>

- Watanabe, S., Yamamoto, E. & Uozumi, M. (2006). Language discrimination by Java sparrows. *Behavioural Processes*, 73(1), 114–116. <https://doi.org/10.1016/j.beproc.2006.01.013>
- Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V. & Slocombe, K. E. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, 25(4), 495–499. <https://doi.org/10.1016/j.cub.2014.12.032>
- Weisman, R. G., Mewhort, D., Hoeschele, M. & Sturdy, C. B. (2012). New perspectives on absolute pitch in birds and mammals. In E. A. Wasserman & T. R. Zentall (Eds.), *The Oxford Handbook of Comparative Cognition (2 ed.)* (Oxford Han, pp. 67–79). Oxford University Press.
- Weisman, R. G. & Ratcliffe, L. (1989). Absolute and relative pitch processing in black-capped chickadees, *Parus atricapillus*. *Animal Behaviour*, 38(4), 685–692. [https://doi.org/10.1016/S0003-3472\(89\)80014-4](https://doi.org/10.1016/S0003-3472(89)80014-4)
- Weisman, R. G. & Ratcliffe, L. (2004). Relative Pitch and the Song of Black-Capped Chickadees
Chickadees, like people, have a strong sense of relative pitch. These birds use skillful, precise pitch changes to advertise their quality and attract mates. *American Scientist*, 92(6), 532–539.
- Wilson, M. & 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. *Psychonomic Bulletin & Review*, 23(6), 1647–1659. <https://doi.org/10.3758/s13423-016-1013-x>
- Winkler, I., Háden, G. P., Ladinig, O., Sziller, I. & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America*, 106(7), 2468–2471. <https://doi.org/10.1073/pnas.0809035106>
- Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M. & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology. General*, 129(3), 291–307. <http://www.ncbi.nlm.nih.gov/pubmed/11006902>
- Yin, P., Fritz, J. B. & Shamma, S. a. (2010). Do ferrets perceive relative pitch? *The Journal of the Acoustical Society of America*, 127(3), 1673–1680. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2856516&tool=pmcentrez&render type=abstract>
- Zandberg, L., Lachlan, R. F., Lamoni, L. & Garland, E. C. (2021). Global cultural evolutionary model of humpback whale song. *Philosophical Transactions of the Royal Society B*, 376(1836). <https://doi.org/10.1098/rstb.2020.0242>

Biographical paragraph

Carel ten Cate is Professor of Animal Behaviour at Leiden University (the Netherlands). His research focuses on behavioural development and communication by vocal and visual signals in species ranging from birds and fish to humans. He uses birds for comparative studies on various cognitive abilities. He has published extensively on whether and to what extent birds possess mechanisms similar to those relevant for speech, language and music processing in humans. He is editor of the book 'Avian Cognition' (Cambridge University Press, 2017).

Henkjan Honing is Professor of Music Cognition at both the faculty of Humanities and the faculty of Science of the University of Amsterdam. His research focusses on the topic of *musicality*: the cognitive, biological, and environmental factors that contribute to our capacity for music. He is editor of the book 'The Origins of Musicality' (The MIT Press, 2018) and published several books for the general public, including 'The Evolving Animal Orchestra' (The MIT Press, 2019) and 'Music Cognition: The Basics' (Routledge, 2022).

Glossary

Absolute pitch (AP) refers to being able to recognize a sound by its pitch and label it (is it a C or a D#?) without a reference, as effortlessly as most humans can label colours. The first part of the skill, memory for pitch, is widespread in humans and non-human animals, the second part, categorizing and labelling a pitch, is rare (estimated in humans as 1 in 10000).

Beat perception (and synchronization) refers to the sensation of a regular pulse, to which one can dance and/or synchronize, induced by a rhythmically varying acoustic signal. Beat perception and synchronization (BPS) can be seen as a form of entrainment, whereby periodic motor and auditory rhythms become temporally coordinated, with temporal alignment of movements to auditory beats.

Calls refer to animal vocalizations that are usually rather short, often monosyllabic and generally, but not always develop without learning. Many animal and bird species have a repertoire of different calls, functionally linked to specific contexts such as begging calls or alarm calls.

Convergent evolution occurs when traits with a different evolutionary origin become similar during evolution, such as the wings of butterflies and birds. Such traits are indicated as being **analogous**.

Homologous traits are traits that have the same evolutionary origin, due to being derived from a common ancestor.

Pitch perception can be separated in two components, which are referred to as *pitch height* and *pitch chroma*. The first aspect refers to the perception of frequency on a log linear scale. The latter aspect refers to a circular way of perceiving pitch, where pitch repeats each time frequency doubles. It is also referred to as **octave equivalence** or octave generalization, a phenomenon that can be observed in humans and just a few nonhuman animals.

Prosody of speech refers to how tone, stress and intonation vary over a speech utterance.

Relative pitch (RP) refers to being able to recognize a transposed melody or intonation pattern as the same. Humans can easily recognize a melody that is transposed up a few semi tones as the same melody. Birds have great difficulties in classifying pitch-shifted sounds.

Songs are animal vocalizations that are usually longer and acoustically more complex than calls, consisting of a variety of notes or syllables ordered in particular species-specific sequences. In several animal groups (e.g., songbirds, some whales) song are acquired by vocal learning (imitation) from other individuals. Functionally songs are mostly linked to reproduction (mate attraction, advertising territoriality).

Spectral envelope and **spectral contour** refer to the amplitude of the different frequencies present in a sound signal. Spectral envelope (or spectral shape) refers to the energy distribution of one point in time, spectral contour refers to the energy distribution over time (see Fig. 29.2).