ESIC2024 Posted: 15/07/2024

Acoustic Processing and the Origin of Human Vocal Communication

Nicholas Bannan[1], Robin I. M. Dunbar[2], Alan R. Harvey[3,4], Piotr Podlipniak[5]

[1]Honorary Research Fellow, The Conservatorium of Music, University of Western Australia, Perth, Australia
[2]Department of Experimental Psychology, University of Oxford, Radcliffe Observatory Quarter, Oxford, UK
[3]School of Human Sciences and Conservatorium of Music, The University of Western Australia, Crawley, Australia
[4]Perron Institute for Neurological and Translational Science, Nedlands, Australia
[5]Department of Musicology, Adam Mickiewicz University in Poznań, Poland
Email: n.bannan@bigpond.com

Abstract

Humans have inherited from their remotest mammalian ancestors an integration of the sensory and motor systems that permits the exchange of signals and information, led by an instinctive response to harmonicity. The transition, from capacity for animal communication involving calls, facial expression and gestures, to modern human culture that embraces language, music, and dance, has resulted from anatomical adaptations such as upright posture, a distinct oro-facial and respiratory tract arrangement, and important changes in neural architecture, connectivity and plasticity. A key development on which modern human vocal abilities depend was the achievement of a feedback loop between the perception and production of voiced sound that allows precise matching, imitation, and meaningful variation, of the acoustic properties: fundamental frequency, duration, amplitude, and timbre; all contributing to the capacity to coordinate simultaneous participation. These are the four specific properties that influence our sensations of sound. They play parallel and complementary roles in both language and vocal music, represented in blended synchrony such as unison song and collective speech as well as temporally organised and coordinated sequences such as call-and-response and polyphony. In turn, responding especially to the properties of the Harmonic Series, such features inform the musical employment of tools – musical instruments – that make use of materials available in the local environment (skin, bone, stone, bamboo, wood, etc.) to provide increased range and loudness as well as alternative timbres to those produced by the voice. We propose a framework by which the adaptations essential to such universal behaviours may have arisen, drawing on an extensive and varied literature to explore the potential of this position for illuminating the nature of human musicality and its relationship to language development.

Keywords: Evolution, hearing, musicality, singing, social bonding, dimorphism.

This article sets out to examine the widespread literature which addresses the phenomenon of human auditory processing and its evolved role in perception, representation, vocalisation, and communication. This suite of abilities manifests in language, song, as well as in other, non-linguistic behaviours, both vocal (e. g. laughter, crying, humming: Provine, 2017; Provine et al., 2009; Gracanin, 2018; Dunbar, 2022) and in the employment of extrasomatic tools such as musical instruments, and response to the acoustic properties of the environment. There is an enormous and varied literature reaching back to the Ancient Greeks (Franklin, 2002) and Chinese (von Falkenhausen, 1992) that has sought to describe and account for human musicality and its potential (see also Wood, 2022). Since Darwin (1871), and especially during the last thirty years or so, a multi-disciplinary explosion of research interest in this phenomenon (Wallin 1992; Wallin et al., 2000; Mithen, 2005; Bannan, 2012a; Hoeschele et al., 2015; Honing et al., 2015; Harvey, 2017), has emerged in theorisation and experimental study that have developed on evolutionary lines.

Research in this field is proceeding in a range of disciplines in which historical misunderstandings arising from differences of interpretation can prove a bar to progress. This paper therefore sets out to provide an overarching framework within which to set a range of more specialised contributions. Central to the synthesis we attempt is an examination of the role of harmonicity in the evolution of mammalian aural perception and communication, and its surviving role in both music and language.

Figure 1: Acoustic processing and the origin of human vocal communication

A web of connections

Theoretical models

Perception **Production**

Range **Timbre**

Frequency **Formants**

Hearing **Vocalizing**

Multi-sensory integration with vision and touch **Dance and instrumental performance**

Distance, direction and threat **Fight or flight, emotional response**

***

Pythagorean theory: ratios **Aristoxenus: empirical experience**

Instinct **Vocal Learning** **Cultural transmission**

Darwinian sexual selection

Harmonicity **Tonotopic neural organisation** **Scale systems**

Octave equivalence

Vocal dimorphism

Adult female and child vocal range **Adult male vocal range**

Wordless singing **Song** **Normal speech**

Social bonding **Credibility of signalling**

Vocalization in childbirth **Mother-infant dyad**

Commencing with a brief historical overview (1), we explore the similarities and differences between animal vocalisations and human song and speech (2, 3). We then discuss the phylogeny of human hearing and sound production (4, 5, 6), and the ontogeny of human vocal communication (7, 8, 9), leading to consideration of issues of universality (10) and the distinction between song and language (11). Finally, we address in the light of the preceding the issue of harmonicity (12), and responses to the properties of the Harmonic Series (HS) evident in human music-making that also underpin vocal learning in the acquisition of language. A diagrammatic representation of the principal components of the field is given in Figure 1, above.

A historical framework: Darwin, Helmholtz, Lamarck, Baldwin and acoustic communication

The 18th and 19th Centuries saw developments in a variety of disciplines that for the first time addressed the evolution of human communicative ability from a scientific rather than a mythic or religious perspective. Rousseau thought early communication must initially have had both musical and linguistic content (see Scott, 1998) while von Humboldt (1836), anticipating Darwin’s synthesis, conceived of mankind as ‘a singing species’. Helmholtz (1885), on acoustic grounds, illustrated the auditory and vocal overlap between speech and song. Darwin, aware through personal communication with them of the positions of both Humboldt and Helmholtz (Bannan, 2017), proposed, in exemplifying the theory of Sexual Selection (Darwin, 1871), that human vocality had its origins in song, through which males and females attracted and comforted one another, and from which language later descended. His reluctance to explore further the implications of Helmholtz's acoustical theories for the relationship between harmonic organisation and vowel sounds, stating that he did not understand them (Bannan, 2017), left unexplored a key component of the relationship between auditory perception and vocal production.

While Helmholtz focused on the proximal explanation of musical perception (Helmholtz, 1836/1895), Darwin pointed to the ultimate causes of the origin of music (Darwin, 1871). They both thought of music as an acoustic and behavioural phenomenon. However, the picture they painted at that time lacked a role for cultural inheritance and invention as important factors influencing the evolution of the musical mind. For Helmholtz, music, including its structure, was only an acoustic entity external to the listener. Darwin, in turn, saw the origins of music as the result of sexual selection, in which the factors determining human musical abilities were the choices of females. However, the communicative nature of music requires all participants within the group to access the same communication system, which is difficult to explain by the appearance of a random mutation in one individual. An explanation for the emergence of musical abilities that would be in accordance with Darwin's laws of selection is the model proposed by Baldwin (1896a). In the late nineteenth century, several researchers independently (Baldwin himself; Morgan, 1896; Osborn, 1896) proposed a mechanism by which a culturally invented trait comes under genetic control under certain conditions. According to these authors, in order for this to occur, a culturally invented adaptive trait – a phenotypic adaptation – must be biologically costly (i.e., must be achieved by strenuous learning which is energy- and time-consuming) and must last many generations. If these conditions are fulfilled, sooner or later a random mutation arises which predisposes an individual to learn this trait faster and less strenuously, and which is then preferred by natural selection. Therefore, although Baldwinian evolution resembles Lamarckian "inheritance of acquired traits' (Lamarck, 1815/2011) it is in fact a perfect example of Darwinian natural selection (Jablonka and Lamb, 2005). However, since this process necessitates interactions between cultural and genetic evolution it is in fact a specific type of gene-culture coevolution (Dawkins, 1976: Lumsden and Wilson, 1980). Since both music (Brown and Jordania, 2013; Mehr et al., 2019; Savage et al., 2015; Yurdum et al. 2023) and natural speech (Clark and Misyak, 2009; Greenberg, 1966) are composed both of universal features observable in all musical cultures and conventional traits that are culturally variable, their origins would be consistent with Baldwinian evolution (Podlipniak, 2017). The same seems to be true for language (Deacon, 2003, 2010; Dor and Jablonka, 2010). The cultural differences between the structures of the range of world music and national languages suggest additionally that neither speech nor music are entirely reducible to acoustics (Shannon 2005): this was certainly the case prior to the development of recording technology. In other words, music and speech are to an important extent cognitive phenomena. Therefore, in order to explain the origins of both music and speech, one must examine the relations between all of the many ontological levels they involve such as acoustics, the physiology of the vocal apparatus, including its neurophysiology, and outcomes such as emotion, cognition, and social interaction, including aesthetic assessment, as well as pre-human foundations in instinctive responses such as fight/flight behaviors. Such an approach characterizes proposals that accentuate the role of gene-culture coevolution in the origins of music (Benítez-Burraco and Nikolsky, 2023; Harvey, 2017; Killin, 2016, 2017; Patel, 2018, 2023; Podlipniak, 2016, 2017, 2021; Shilton, 2022; Tomlinson, 2015).

The challenge falling to the generation immediately following Darwin was, then, to reconcile the evolutionary position set out in The Descent of Man (1871) with the wider implications of Darwin's foundational theory of natural selection (1859), without which the anatomical and perceptual developments on which human behaviour depends could not be explained. This led, for instance, to a perspective on the primacy of song and speech that Spencer (1890), perhaps consistent with his coinage 'survival of the fittest', saw as the reverse of Darwin's own: that music descended from an existing capacity for language through its communicative association with different emotions. Speculation on this issue needed to avoid a Lamarckian interpretation (Lamarck, 1815) treating as heritable capacities acquired by individuals within their own lifespan. Baldwin's solution (1896) conceived of evolution operating at both the genetic and cultural levels: the 'Baldwin effect' proposed that a culturally invented trait is transformed into an instinctive trait by the means of natural selection (Podlipniak, 2017).

But subsequent developments have not built systematically on these achievements. The Spencerian preference for the precedence of language extended into the 20th Century resonances from the traditions of Classical philosophy and Christian theology: 'in the beginning was The Word'. Evolutionary theories of the acquisition of speech (Pinker, 1994, 1997; Aitchison, 1996) presented for a century or so an exclusive inter-relationship between language use and the development of general intelligence (Burkart et al., 2017). Gardner's (1983) model of mutually independent intelligences (visual-spatial, linguistic-verbal, logical-mathematical, bodily-kinesthetic, musical, interpersonal, intrapersonal, naturalistic) illustrated an alternative framework whereby specific cognitive, emotional, and social abilities emerged in response to contrasting environmental pressures and stimuli. The implications of multiple intelligence theory, while influential in education research and practice, were slower to underpin theorisation regarding the relationship between human cognitive and emotional faculties. For instance, Aitchison (1996, p. 149), in reference to the acquisition of syntactic structures that capture cognitive processes, refers to mathematics as revealing an independent capacity for human generativity in addition to language. But she makes no admission that musical behaviour also presents generative capacities, as has been recognised by commentary on language origin since von Humboldt (1836) and Darwin (1871); nor that it independently employs the vocal-auditory system to do so.

Complex animal vocalisation and possible links to human song and speech

An avenue for research in human communicative behaviour that built on Darwin's selection theories and has developed as a research focus within several disciplines is the comparative study of human and animal behaviour. The human multi-sensory feedback mechanism permits a unique capacity for synchronizing signals that is equivalent in the medium of sound to the flocking of birds in movement through the medium of air, and the shoaling of fish through coordinated movement in the medium of water (Bannan, 2000). Stewart (1998) reduced to four mathematical 'rules' the principles on which such behaviours depend: efficiency (the rules are simpler than the behaviours they generate); consistency (protecting against the consequences of 'rogue' mutation); adaptability (small changes in the rule cause big changes in behaviour); dependence (sensitivity to the group protects the individual) (Stewart, 1998, p.196-7). While human physical movement in dance, marching and other forms of bodily coordination parallels in many ways the flocking and shoaling to which Stewart refers, Bannan (2000) applied Stewart's rules to the parameters of human vocalisation. The four elements that we control as features of vocal learning (pitch, timbre, duration, and amplitude (Bannan, 2008)) are each capable of conveying meaning where the remaining components are unchanged, whether this represents perceived effort, salience, or categorical alternative (for instance, where the alteration of the formant frequencies of the vowel changes one word into another, such as 'boot' to 'bot'; in the durational stresses that determine meaning in certain languages; or in the emotional incorporation of a musical pause, or changing tempo that involves the shortening or lengthening of durations). Sensitivity in the perception of such variation governs both the capacity for simultaneous coordination, as in choral blend (Ternström and Sundberg, 1989), and for imitation available to subsequent recall (vocal learning, Merker, 2012), as well as the perception of both the carrier signal (the word or note) and the additional information it can convey (emotion, effort, credibility). Cummins (2002) has revealed how such variations can be detected in the analysis of synchronised speech (for instance, oaths and prayers), and proposed that the capacity for such synchrony had a developmental role within the continuum that embraces the properties of both language and music (Cummins, 2013).

The anatomic and neurological evidence for tonotopic pitch processing in mammals

Crucial stages in the development of animal communication can be traced in the anatomical evolution of auditory perception, and especially the relationship between the peripheral system, the processing of signals in the auditory cortex, and the specialization of neural and anatomical apparatus necessary for increasingly sophisticated and malleable vocalization. Given such factors as body size, environment, threats from predators, and intra-species organisation, what do mammals listen for, and how do their perceptions benefit their survival? Evolved characteristics compared across species reveal variation in the capacity for auditory scene analysis (Bregman 1990), the relationship between natural stimuli and neural events cleverly termed the ‘acoustic biotope’ (Smolders et al., 1979; Johannesma et al., 1982). Evolutionary sequences are only deducible for these features as evident in comparisons between living examples, since fossil evidence provides insufficient and degraded data. Hence the significance of inter-species comparisons of perceptual responses to acoustic phenomena for helping to adduce the bases of human ability.

Essential components of the physical system include the tympanic membrane (ear drum), middle ear ossicles and the cochlea containing the basilar membrane and the organ of Corti, which together have developed so as to transduce soundwaves into electrical signals that are then sent to specialized auditory processing regions in the brain (Bendor and Wang, 2005; Harvey, 2017, p. 18-19; Matsui, 2023). Distributed areas of the auditory system analyse pitch/frequency relationships (both fundamental and formant), rhythm/beat, timbre and sound intensities. In animals in which responses have been measured, a relationship has been detected whereby the peripheral receptors distributed along the length of the basilar membrane are ‘tuned’ to different frequencies. This map of specific frequencies is transmitted onto brain areas that process them in a manner that is responsive to the proportional vibratory rates of the Harmonic Series (henceforth HS): a relationship known as tonotopy. While speculation regarding the organisation of the aural system on tonotopic lines stretches back to Classical philosophy and Enlightenment speculation (Ruben, 2020), tonotopy was defined and explored anatomically by Békésy (1928), and evidence located of the perception of aural harmonics along the basilar membrane of the guinea pig by Tonndorf (1958). A correlation of the increased sensitivity of tuning and tonotopic relationships in humans (Sumner et al., 2018) may be considered a precondition of the capacity of our species for vocal learning from infancy, as well as for unison participation and the production of a range of vowel sounds (see also Moerel et al., 2013, 2015).

Tonotopy and time (or temporal) coding are the two ways of coding information about sound frequency in the human auditory system (Oxenham, 2013). While time coding in mammals applies to low frequencies (e.g., not higher than 3.5 kHz in guinea-pigs (Palmer and Russell, 1986)), tonotopic coding allows the discrimination of high frequency sounds up to about 20 kHz in humans which provide important cues for speech recognition (Trine and Monson, 2020). Since the representations of pitch and timbre in the human auditory cortex partially overlap with tonotopic representations of sounds frequencies (Allen et al., 2022), tonotopy seems to be an important feature influencing our experience of both pitch and timbre.

A significant finding in the comparison across species of the relationship between the peripheral and neural systems involved in aural perception has revealed differing patterns of response to contrasting acoustic stimuli. Wang (2013) investigated evolutionary differences between processing of sounds according to harmonicity (spectral features) as opposed to periodicity (temporal components), suggesting a ‘mirror relationship’ he identified for these phenomena. He proposed that during the early developmental period, the auditory cortex is flooded with both exogenous harmonics from the acoustic environment and intrinsic harmonics generated within the auditory system itself (Wang 2013, p. 1). Different species balance this ‘mirror relationship’ between harmonicity and the iterative nature of temporal cues in ways that suggest contrasting responses to sound stimuli, correlating to such factors as body size and sensitivity to specific ranges.

Comparative research across a range of species including humans provides the basis for deducing the precise perceptual factors that distinguish human vocal ability and account for its evolution. Shofner (2023) compared the spectral and temporal representations of stochastic, complex sounds which underlie the perception of pitch strength in humans and chinchillas. He concluded that the enhanced cochlear tuning in humans predicts differences in how the spectral and temporal cues of complex sounds are encoded and perceived by humans compared to other mammals (Shofner 2023, p. 154). The findings of Shera and Charaziak (2019) argue that sharper tuning reflects a longer cochlea, and that although mammals with longer basilar membranes may possess sharper tuning, this is determined by the spatial spread of excitation along the basilar membrane, i.e., the space-constant, not simply the basilar membrane length itself (Shera and Charaziak, 2019). Shofner (2023, p. 156) concludes that the representation of harmonic structure in the excitation pattern is more strongly evident in humans than in chinchillas, ferrets (Walker et al., 2019), or rabbits (Wagner et al., 2022). Note here, and this may be important for much of the discussion that follows, that there are also evolutionary conserved elements at the periphery, particularly the presence of nonlinear processing and harmonic distortions in the basilar membrane and organ of Corti within the cochlea itself (Cooper, 1998; He and Ren, 2021). The significance of these harmonic distortions has rarely been recognized, yet in the context of harmonic sounds and chroma they may be relevant because an important product of these distortions is the generation in the receptor apparatus itself of defined harmonic frequencies above the fundamental (F0), the largest and most prominent of these being 2F0, the octave directly above F0 (He and Ren, 2021). The human auditory system evolved to promote vocal learning sensitive to and capable of exploiting all four parameters of sound.

Other factors that contribute to an understanding of the adaptive nature of human aural perception include: the development from early infancy of psychophysiological responses to musical frequencies, as distinct pitches, contour, and intervallic differences (Trehub 2003; Zeloni and Pavani, 2022); the role of harmonicity in separating significant signals from background noise (Moerel et al., 2013; McPherson, Grace and McDermott, 2022); and the differences and similarities in human and macaque responses to aural stimuli that seem tuned to endow preference for conspecific vocalizations (Bodin et al., 2012).

In summary: the feedback loop of the human auditory-vocal system teaches through imitation, both permitting and responding to vocal learning. Key here are links between the auditory perceptual process and relevant parts of the motor system. While similar capacities have been found in the acoustic communication of birds (Elie et al., 2020), they do not play a part in the vocal development of our nearest genetic relatives, or in mammals generally; and birds have not been found to exhibit octave equivalence (Hoeschele et al., 2013; Wagner et al., 2015). That tonotopy is so common in the mammalian auditory systems (Mann and Kelley, 2011) indicates that the recognition of HS properties is crucial not only for all humans but also for all mammals (if not for the majority of animals). This is, of course, not surprising as harmonic sounds are mainly produced by living creatures. Additionally, the appearance of singing (and speech) in our evolutionary lineage was related to the species-specific use of harmonic sounds in vocalizations (volitionally controlled and more elaborate than in the case of affective prosody) (Ackermann et al., 2014).

A picture of what is distinctive about the anatomical evolution of human vocalization and its reliance on sophisticated hearing and motor coordination embraces the following:

1. Our sensations of sound differ from our sensations of visual stimuli in terms of the instinctive search for communicative content (Kant wrote (1798) that "Blindness cuts us off from things, but deafness cuts us off from people"). In other words, the processing of sound adapted to decode information intentionally coded by means of acoustic parameters. Although visual communication has also been observed in primates, the use of vocalizations necessitates the recognition of temporal patterns (the changes of acoustic parameters in time that are crucial for sound communication).

2. The recognition of harmonic sounds, including vowel categories (Nazzi and Cutler,

2019), should generate similar sensations in all humans (unfortunately, sensations are very difficult to compare, being parts of our conscious experience). It is important to note here that there is evidence of octave equivalence and the processing of harmonics in human auditory cortex (Moerel et al., 2013, 2015), and similar observations have been made in at least some primates (Wright et al., 2000).

3. As music is, like speech, a culturally flexible tool of communication, the same sensations can be used differently leading to different aesthetic responses (e.g., the interval of a minor second can be experienced as unstable being a part of an unresolved chord, or as stable, when experienced in relation to the tonic in jazz music). The lack of distinction between sensations and aesthetic assessments has proved a potential source of confusion, for instance regarding whether 'octave equivalence' is universal (Jacoby et al., 2019; see also discussion of this feature below).

4. Alterations in the processing of sensorimotor coordination of the oral cavity, pharynx and larynx were crucial for the evolution of our musicality. However, these changes in connectivity had also to influence the evolution of human auditory processing (Norman-Haignere et al., 2019). The control of vocal tracts, in the case of vocal learning, demands a dedicated kind of auditory feedback and sophisticated motor control. Compared to monkeys, there is in humans greater connectivity between the laryngeal motor cortex and parietal sensory cortices, allowing far greater sensorimotor modulation and integration (Kumar et al., 2016), and there are unique projections from the 'phonatory region' of motor cortex to the brainstem nuclei controlling vocalisation (Zarate, 2013).

Central processing and the neurobiology of human musicality

In addition to increasing sophistication in the control of the peripheral apparatus involved in vocalisation, musical interaction has co-evolved with a neural architecture that highlights its fundamental significance to human culture and well-being. Sensory and affective aspects of music are processed in a complex distributed network within the brain, often with a bias to the right hemisphere, involving multiple areas within the cerebral cortex and regions within the limbic system – hippocampus, amygdala and cingulate cortex (e.g., Peretz and Zatorre, 2005; Levitin, 2007; Patel, 2008; Harvey, 2017, 2018; Koelsch, 2018). There are morphological differences between the left and right human auditory cortex, and functionally the right side differs in frequency and spectral resolution (Harvey, 2017). Even more regions (e.g. motor and premotor cortex, basal ganglia, cerebellum, and cranial nerve nuclei in the brainstem) are then involved in vocalisation or other motor outputs including synchronised movement and/or playing some form of musical instrument (della Bella et al., 2015; Karpati et al., 2016; McPherson et al., 2018). Musical training alters neural connectivity, enhancing interhemispheric transfer of information and increased left hemisphere activity (Manzano and Ullén, 2018; Leipold et al., 2021), plasticity in auditory and motor systems that is even evident in mature adults (Worschech et al., 2023). Important for synchronisation, arousing and rewarding music activates anticipatory and reward centres in the ventral striatum (e.g., Blood and Zatorre, 2001; Menon and Levitin, 2005; Salimpoor et al., 2011; Ferreri et al., 2019) and high emotional valence is associated with increased activity in many regions in frontal and prefrontal cortex as well as in the insula (e.g. Pereira et al., 2011; Altenmüller et al., 2014; Koelsch, 2018; Harvey, 2020).

Given the remarkable array of circuits and interconnections that mediate the sensorimotor processing of music, classic excitatory and inhibitory neurotransmitter systems such as glutamate and gamma-aminobutyric acid (GABA) are of course involved. However, the rewarding, arousing, mood enhancing and socializing aspects of human musicality, impacting both brain and body, involve a number of other neuromodulators (Chanda and Levitin, 2013). The major systems known to be affected by the communal aspects of human musicality are the dopaminergic (Zatorri and Salimpoor, 2013; Ferreri et al., 2019), opioid (Dunbar et al., 2012; Tarr et al., 2015, 2017; Weinstein et al 2016; Mallik et al., 2017; Dunbar, 2023) and oxytocinergic (Kreutz et al., 2014; Keeler et al., 2015; Yuhi et al., 2017; Harvey, 2020) systems.

While each system has tended to be studied separately from the others, recent preclinical and clinical research is beginning to point to numerous neuromodulator interactions, at multiple levels (receptor, synapse and circuit), that potentially affect functionality (Paletta et al., 2022; see also Arnold et al., 2024). For example, many of these neuromodulators act via receptors that are well-known to interact with each other at the cellular level, affecting downstream signalling and network function (Borroto-Escuela et al. 2017). Such potential interactions relevant to music processing can be between some dopamine, oxytocin and serotonin receptors. It is also now known that certain ligands can bind to a secondary site on another receptor to modify its function (Foster and Conn, 2017). This may be important from the perspective of understanding how music drives social cooperativity and sensorimotor coordination because perhaps the two most important and relevant peptides, β-endorphin and oxytocin, have been reported to interact in this way with possible consequences for many aspects of social memory and synergistic behaviour (e.g. Dal Monte et al., 2017; Pearce et al., 2017; Meguro et al., 2018; Salighedar, et al., 2019; Putnam and Chang, 2022).

Where such neural responses prove universal and adaptive, the brain architecture, neurochemistry and physical integration they depend upon must have a long history that defined the emergence of human musical behaviour and has left traces of the sequence by which they occurred. Comparisons of these features in humans to their equivalent in other species reveals the specific cognitive, social and emotional correlates of cooperative musical engagement that are unique to our species.

Human acoustic perception: range, sensitivity, categorical discrimination, and subjectivity

While investigation into the phenomenon of human auditory perception continues to employ comparisons with other species, we need to consider the specific elements that characterise human acoustic response and its consequences for vocalisation, as well as for the additional universal behaviours associated with extrasomatic control in which breath is employed non-vocally, or the hands and feet control in musical performance (Espi-Sanchis and Bannan, 2012). Our abilities descend from those that were determined by the auditory scene analysis (Bregman 1990) which permitted our ancestors to survive and reproduce. What may these have comprised? This section will address the factors and their significance: range; pitch; intensity; timbre; the Doppler effect; developmental sexual dimorphism; the role of culture; and problems and controversies thrown up in the attempt to explain these categories within an evolutionary framework.

Human hearing has developed in correlation with body size and the range of acoustic phenomena in the environment. Range (fundamental frequency, measured in Hertz) determines onto what total capacity for aural perception our more limited vocal repertoire maps: some four-five octaves, stretching from around 50 Hz for the deepest male bass to about 1600 Hz for the most piercing scream of the child or woman. Young humans can perceive sound signals within a frequency range from ca. 20 Hz up to ca. 20 kHz, equivalent to ten octaves: from just below the lowest pitch on the piano to just over two octaves above the highest (and discriminate with precision up to c. 4kHz). Musical training clearly influences clarity of pitch perception (Micheyl et al. 2006).

Within such a range, we possess considerable sensitivity to change of pitch. In many languages, pitch is conceptualised as analogous to height, and changes in pitch are ‘up’, or ‘down’. Pre-verbal infants have been found cross-culturally to associate changes in pitch height with visual experience of upward and downward motion in space (Dolscheid et al., 2014). But tones an octave apart may be labelled as having the same pitch, for which the term chroma is employed. In laboratory tests, however, human subjects do not always detect pitch changes with precision. Neuhoff et al. (2002) suggested that incorrect judgments of direction may be partly to a failure in detection, but also due to “conceptual errors”: “lack of knowledge of the appropriate labels for rising and falling pitch.” In trials of listener discrimination of pitch difference (higher, lower, the same), Semal and Demany (2006) revealed inconsistencies in normal listeners in their detection and identification of the direction of pitch changes that implied differences in individuals’ capacity to determine this. Rather than this representing a deficiency, they proposed that the detection ability exists in the brain of every normal listener but that, for some listeners, they do not respond to very small shifts (2006, p. 3913). By comparison, trained musicians can distinguish two tones that differ by only 0.1%, i.e., a 1000 Hz tone from a 1001 Hz tone (Friauf, 2014).

Aural response to pitch reveals itself early in the human life-cycle. A capacity for pitch discrimination has been recorded in foetuses in the final trimester (Lecanuet et al 2000) and newborn infants process intervals in pitch as well as detecting the beat in music (Stefanics et al., 2009; Winkler et al., 2009). Trainor and Desjardins (2002) illustrated the role of raised vocal pitch in infants' acquisition of vowel categories. The feedback loop that permits babies to respond vocally develops early, evident in pitch-matching between mothers and infants between 3 and 6 months of age (Kessen et al., 1979), providing foundations for the capacity for complex vocal learning. In the context of sensorimotor feedback, deficits in connectivity between an apparently normally functioning adult human auditory cortex and frontal cortex contribute to what some have termed 'congenital amusia' (Hyde et al., 2006; Moreau et al., 2013).

In the case of sensitivity to change of sound intensity (measured in decibels), the perception of change in the amplitude of sounds has clear survival advantages (Ghazanfar et al., 2002). Tajadura-Jiménez et al. (2002) elicited emotional discrimination in listening subjects that captured arousal responses where rising intensity levels conveyed an approaching sound source. Approaching sounds were categorised as being more salient: threatening sounds as eliciting increased danger more intensely than pleasant sounds increased expectation: "Sound sources that are rapidly approaching one, invading the margin of safety surrounding the body, need to be identified quickly in order to determine their threat value" (Tajadura-Jiménez et al., 2002, p. 39).

A significant aspect of response to approaching and receding sounds is caused by the Doppler effect, the apparent change in frequency of a wave in relation to an observer moving relative to the wave source. Doppler, the Austrian physicist who first studied the phenomenon systematically, initially published on it in relation to the frequency of light travelling from stars (Doppler, 1842), and its application to sound was subsequently illustrated in a public experiment in which two brass bands performed, one on a station platform, the other in a moving railway train (Buys Ballot, 1845). Listeners verified the alterations in pitch related to the train approaching and receding. It required such an experiment to demonstrate scientific understanding of a phenomenon naturally influential on the auditory system of animals such as bats (Yin and Müller, 2019) and dolphins (Ryabov, 2019). One might speculate that Doppler shift would also have been experienced by our early ancestors, alongside the binaural capacity to detect the direction of a potential approaching threat: our ears telling us where to locate what to look at, and to estimate distance (Heffner and Heffner 2018).

The discrimination of timbre plays a part both in auditory scene analysis and the modelling of distinct target sounds that permit the learning of language. Timbre conveys aspects of sound sources such as emotion and intention in animal vocalisation, and harmonicity. Since, both naturally in the process of aural perception and through a range of techniques in the laboratory, timbre is analysed in relation to the factors of pitch and intensity that combine in complex ways to yield myriad acoustic outcomes, its study has been the source of debate and controversy (McAdams, 2019). The Acoustical Society of America defined timbre in the 1960s as follows: "the attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar" (Wei et al., 2022). Arguably to a greater extent than other acoustic parameters, the perception of timbre is a subjective experience influenced by the prior experience of the listener: through learned familiarity with language, music and the sonic environment.

Timbre plays a significant role in unvoiced sounds, such as whispering, in which fundamental frequency is absent and formants modify white noise (hhh) to present an acoustically compromised simulacrum of speech (Tartter, 1989). Song cannot be whispered. Developmentally, the ability to whisper, which infants find harder to achieve than voiced speech, emerges once voiced speech has been acquired (Traunmüller and Eriksson 2000). Whispering and silent mouthing therefore present themselves as variants of language appropriate to certain situations in which they may communicate appropriately (Cirillo 2004). Babies prefer the calming effect of voiced speech and song to their mothers' whispering (Spence and Freeman, 1996). It would seem unlikely that whispering represents a stage in the evolution of language, though it is clearly an associated outcome.

Vocal dimorphism in humans and its evolutionary significance

Both in terms of timbre and, most significantly, range, human voices present a distinctive characteristic: developmental sexual dimorphism. While children of both sexes have voices of similar range from birth to the onset of adolescence, those of adult men drop by around an octave on the attainment of maturity. The significance and adaptive rationale for this has been investigated by Bannan et al. (2024). The phenomenon lends to human culture a capacity for enriched social participation, as well as presenting a lifelong dimorphic distinction in human adults.

Turning to the consequences of these features of human auditory perception for vocal production, we encounter inconsistent interpretations of the relationship between hearing and communication that have resulted from the complex evolutionary relationships involved. Fant (1960), drawing on his predecessors from Helmholtz onwards, codified source-filter theory as the model by which we understand the production of meaningful vocal signals. Sounds initiated predominantly on the out-breath with a fundamental frequency produced by the larynx are open to variation through filtering achieved by the muscles which control the posture of the larynx itself (Unteregger et al., 2020), and the shape of the upper respiratory tract through which vibrations pass via the nose and mouth over a mobile tongue and through flexible lips. This procedure makes possible the spectrum of vowels employed in speech as well as voiced consonants such as <l>, <m>, <n>, <v> and <z>. Its role in music, especially where voices are combined, gives rise to some of the ambiguities involved in how it determines what we hear, and that are revealed in different interpretations of abilities regarding whether they are genetically inherited instincts or culturally derived (Jacoby et al., 2019; Ellis, 1885).

An example of such differences of interpretation is that of missing fundamental perception, the apparent detection by the listener of a low-pitched sound evoked by higher pitches taken to be its harmonics, but in reality acoustically absent. It was described as "an illusory percept resulting from the brain's attempt to reconstruct a coherent harmonic spectrum" (Justus and Bharucha, 2002, p454), which may suggest that it is primed by previous experience of musical style, and is therefore dependent on cultural transmission. However, ten Cate and Honing (2022, p. 8) found that a 'biological basis for inferring pitch from the harmonic spectrum' is universal in both humans (including young infants) and is also present in rabbits (Wagner et al., 2022). Note here that humans apparently prefer octave intervals slightly above the 2:1 frequency, this enlargement subjectively resolved by late-stage auditory processing (Jaatinen et al., 2023).

Human vocal response to auditory perception endows further rich and complex experiences that may be open to ambiguous categorisation, subjective preference, and contextual illusion. The impressions 'high' and 'low' are essential to the interpretation of prosody (Podlipniak, 2022a), and what it conveys, but musical testing reveals that the detection of differences is not always accurate (Semal and Demany, 2006; Demany et al., 2021). The overlapping acoustic properties of timbre and vowel category can prove aurally confusing, especially given different spoken accents and dialects (Van Heuven, 1985). The interaction of timbral and pitch characteristics of sound were cleverly revealed in the synthesized 'Shepard tones' (Shepard, 1964) which produce the aural illusion of a continually rising pitch that in fact goes nowhere.

Octave equivalence permits children to exchange and participate in vocalisation with both of their parents: but in what ways are the chroma of pitches an octave apart 'the same'? It is not only octaves that can be fused into musical presentation perceived as singular: intervals such as the perfect fourth and fifth lend themselves to parallel performance, as in the organum of medieval Catholic chant, and folk styles around the world; while singing in parallel thirds signals agreement and affection in Classical opera and popular music alike. If discrimination as a survival instinct evolved to detect 'what' and 'how many' potential threats were responsible for the production of a sound, what explains the capacity to disguise this factor, and to gain intense satisfaction from it? Other acoustic phenomena that may elicit awe and pleasure include response to echo, reverberation, and the performance of antiphony that replicates these. How, again, do we explain these in terms of attribution of origin? The multi-sensory factors involved in the perception of voice illustrate the potential alteration in aural response according to the timing of visual signals. This 'McGurk effect' (McGurk and McDonald, 1976), named after the researcher who first analysed it systematically in spoken language, has also been revealed in sung performance (Laeng et al., 2021).

Research into octave equivalence characterises the nature/nurture dichotomy. Sergeant (1983), influenced by a pedagogical motivation rather than by evolutionary theory, devised experiments to ascertain whether the octave is 'percept or concept'. The experiments involved were carried out with percussion instruments (pitched bars) that children explored in order to match which presented the same chroma. Training prior to the experiment was carried out in the visual medium, in which subjects compared pictures intended to discriminate between the perception of shape and subject matter rather than size and colour. These extrasomatic sound sources contrast markedly with the instinctive and adaptively 'original' medium of the voice, whose workings are invisible to the user. Sergeant's conclusion, that OE is culturally learned, would therefore appear largely an outcome of the research method employed. Similarly, the research of Jacoby et al. (2019) that tested for octave equivalence through measuring the responses of Tsinamé subjects to high-pitched, artificial sounds may have stretched too far the features of range and timbre by which OE operates effectively. Walker et al. (2019), in their cross-species investigation into pitch perception, found that the effects of octave equivalence are strongest between neighbouring octaves. Perhaps relevant again here is the previously mentioned phenomenon of cochlear distortion, resulting in the generation of harmonic frequencies above F0, the most prominent at 2F0 (the octave above) and with a less noticeable harmonic at 3F0 (Cooper, 1998; He and Ren, 2021).

In experimental studies that have explored the relationship between human music and its specialised adaptation to fit the preferences of animals such as cats (Snowdon et al 2010) and tamarins (Snowdon et al., 2015), it proved possible to illustrate the relationship of musical parameters such as range and tempo to the evolved perceptual capacities that relate to body-size and resting heart rate, as well as the measurable range of hearing of non-human species. Snowdon's (2021) overview of animal response to human music reveals the extent to which humans interact with species that share the acoustic range to which they are sensitive, implying that this factor may in a variety of ways have also contributed to shaping human responses. Might a 'war of the frequency spectrum' have determined the sensitivity and categorical analytical capacity of human hearing? Extracting signals from background noise, such as in the 'cocktail-party effect', may well have originated as a response to intra-species competition (Popham et al., 2018) and, by extension, separation of human vocalisation from the sounds made by other species. This is a topic to which we will return in examining the nature of harmonicity, below.

An influential feature of human acoustic processing is response to the physical environment. The effect on collective singing, for instance, of changes to the perceived room acoustic (Fischinger et al., 2015) illustrates the immediate affordances that singers make in relation to their individual perception of their contribution to the overall sound. Such abilities are presumably part of the toolkit for vocalisation that accompanied the emergence of human capacity for vocal learning. Speculation on response to the resonance of natural spaces, including amplification and echo, has illuminated the contribution to the study of music evolution in the field of archaeomusicology. Examples of this include Watson and Keating's (1999) investigation of the acoustic properties of Stonehenge as potentially significant to the site's design for ritual responses to sound; Till's (2014) analysis of the potential link between acoustic response and the placement of examples of cave art; and Ouzman's (2001) examination of the resonances associated with the positioning of rock art in Southern Africa.

Inherited and acquired responses to pitch in music and language

As we have seen, unison singing, both at equal F0 and at octave unison, can be claimed as a universal, signature behaviour unique to our species, analogous in its collective synchrony to the flocking of birds and the shoaling of fish. Although unison singing has been partially elicited in cockatiels (Seki, 2021), this kind of vocalization has not been observed in natural conditions. Recognising that the function of parrots' vocal production learning can vary (Benedict et al., 2022), and different species interpret their vocalization according to different distinctive features (Bregman et al., 2016), further research is needed to answer the question of whether mimicking F0 by parrots is based on human-like mechanisms focusing on HS or simply on faithful imitation of sounds. This human ability can be referred to as "spectral synchronization" and underpins both the capacity for unison performance and more complex exploitation of the HS in homophonic and polyphonic interaction that depends both on singing together and on adjusting to other voices through the negotiation of appropriate intervals that coalesce to produce choral textures. For this to be possible, participants need to be able to predict with accuracy and confidence not only "when" but also "what" to sing, in terms of the spectral characteristics of their contribution, adjusting both to an agreed F0 (keynote) that determines the coherence of combining with the complementary F0 of other, related harmonics, and also to the timbral characteristics of the vocal material being sung. Significantly, the capacity for fine-tuned combinations between individuals is crucial for music, but not for language. Except in the case of collective performance of poetry and prayers (though in such cases "musical" features of speech become crucial for synchrony), this culturally widespread tendency to express vocally together by means of music but not language strongly supports a social origin for music, a group cultural behaviour with evidence of an underlying genetic basis (Pamjev et al., 2012; Le Bomin et al., 2016).

The evolution of this ability must have been achieved by gradual changes in the evolution of vocal learning (based at least partly on a Baldwin-type mechanism). Starting from rough imitation of only temporal patterns, hominins slowly acquired more complex control of their vocalizations before they achieved the level of learning that allowed them to predict, imitate and synchronize with the vocalizations of others. There is disagreement as to the extent to which these earlier vocalizations were music-like, in lacking referentiality and grammatical morphology, or a rudimentary mixture of music and language ('musilanguage', Brown, 2000; Harvey, 2017), but in modern humans speech depends less on the exploitation of the capacity for harmonicity than music, since musical syntax is partly based on pitch (Podlipniak, 2022b), while language syntax is not.

Infant display of response to tonal significance suggests a genetic factor already present that is open to the shaping of culturally influenced learning. But differences between cultures, and dependent assumptions regarding what may be salient in musical stimuli, require careful evaluation. In their study of infant responses to pitch relationships, Zeloni and Pavani (2022) considered the effect of dissonance on infant subjects. There are factors involved that require disambiguation: for instance, the difference between perceiving a change in pitch in terms of (intervallic) pitch class, assuming a resultant (musical) meaning, and detecting a heightening or lessening of effort or intensity. Is a minor second best characterised as "dissonant departure from the maintained unison", or "dissonant alternative in intervallic difference to a major second"? Such a distinction may help us ascertain whether a minor second is heard as the smallest musical interval to intentionally convey 'motion', in comparison with both stasis and with the more definite motion covered by a major second. One would like to have an insight into whether the emotional perception associated with response to the two intervals is more a feature of the contrast between 'maintaining pitch centricity/stasis' and 'disrupting stasis' than it is a feature of the size of the intervals. Which explanation best accords to the instinctive and therefore genetically determined (culturally unlearned) nature of the phenomenon? In a comparison of whining in dogs, this tends normally to be associated with a falling pitch. Systematic discrimination between rising and falling semitones in terms of emotional associations might build on Zeloni and Pavani's (2022) study in a way that could be compared to the pitch contour both of human vocalisation and animal calls presented in

Scherer (1992). In turn, such considerations would illuminate the nature of human response to harmonicity.

Building on such early responses to pitch perception, children acquire language equally well through the imitation of both male and female parents, despite the fact that the acoustic models are pitched around an octave apart (Bannan et al., 2024). This illustrates the instinctive capacity to map pitches of different F0 range as if they are the same. Children do not in their voice production tend to differentiate between words that they have learned by copying their father compared to their mother (Siegel et al., 1990, McRoberts et al., 1997): all language is presented in the distinct, personal voice of the child. This suggest that some complex affordances are made in order for this process to occur, based on linking the similarity of formant frequencies resonating over the different F0 of adult male and female voices that would be consistent with the universality and instinctiveness of the perceptual response to octave equivalence and other HS properties that govern the categorical discrimination of vowels. The presence of cochlear distortion and 2F0 harmonics may also be important here (Cooper, 1998; He and Ren, 2021) as is the presence of groups of neurons in human auditory cortex tuned to multiple frequencies that are one octave apart (Moerel et al., 2013; 2015). The presence of such spectrally tuned responses may represent a “biological substrate for perception of pitch chroma” (Moerel et al., 2015, see also Demany and Armond, 1984, Demany et al., 2021).

Gender, female adaptive characteristics, and human ability

A feature of the development of human vocalisation that has been relatively overlooked is physical dimorphism in our species. The relationship between the development of dimorphism and the perception and production of vocal sound through response to octave equivalence, and the dependent mapping of resonant frequencies of the HS, has been examined by Bannan et al. (2024). Two other features of human development related to vocal output have also influenced social and cultural behaviours: the phenomenon, identified as unique in humans, of lachrymose crying; and maternal response to the pain of childbirth.

Human infants are born with voices of very similar range, and retain these until the onset of adolescence (Sergeant et al., 2005; Sergeant and Welch, 2009). During puberty, small changes affect the emergent adult female vocal range and quality (Gackle, 1991). Far greater changes in both F0 and formant frequencies take place, sometimes over an extended period, in the male voice, resulting in an average drop in range of around an octave (Cooksey, 1997). This separation gives to males a more dominant voice entirely different to that of the child. By contrast, the female voice usually remains capable of unison combination with children in singing, and occupies the same range in imitative exchanges with infants (Kessen et al., 1979). Dimorphism with these consequences is consistent with Fitch’s (2002) examination of the descent of the larynx in humans, and with the role of dominance in adult male vocalisation (Puts et al., 2007). These factors strongly support explanation consistent with sexual selection (Puts et al., 2016; Valentova et al 2019; Aung et al., 2023), whereby the voices of men and women present agencies of both sexual attraction and mate retention.

Human crying is a multisensory affect that presents both audible and visual signals, and can compromise normal vocalisation. It would appear to have developed subsequent to the last common ancestors with the chimpanzee (Gracanin et al., 2018). It can be interpreted as a credible signal that recruits empathy and support. Significantly, it is normally associated with the behaviour of children and adult women to a greater extent than men, and different responses may accompany the crying of men compared to women (Warner and Shields, 2007). Sadness is, though, often communicated in the medium of music, in which its experience can be positively enjoyed by listeners of both genders (Schubert, 2016).

A further dimorphic trait comprises hormonal effects on the female voice associated with menarche, menstruation, pregnancy, and menopause. The larynx is highly sensitive to hormonal change, which affects range, timbre, and efficiency. The condition of pregnancy introduces a sequence of changes that may affect the voice, of which the most dramatic can be involuntary vocalisation occurring during the delivery phase. The vocalisations women produce when giving birth may be quite different to any sounds they otherwise make. This instinctive response can be surprising. As Aitchison reminds us, "… mostly, sound emissions are under our conscious control" (Aitchison, 1996, p. 88). While this is normally the case as humans develop from infancy towards adulthood, it also suggests that a distinctive aspect of human vocal learning is the capacity to remain silent, whether awaiting our turn, or suppressing the sound-making reflex. From that perspective, a significant and gender-specific feature of human vocalisation relates to its role in childbirth.

For health professionals, 'voice is a sensitive signal reflecting functions of the respiratory, endocrine, gastrointestinal, and nervous systems' (Koçak and Atan, 2022, p. 485). The pain of childbirth was, in the study of Weisenberg and Caspi (1989, p. 17), rated as high by most women (see also Melzack, 1984). In the analysis of Koçak and Atan (2022, p. 485), which adopted caesarean delivery as a control against which to compare vaginal delivery, the latter is seen as a process that forces pregnant women to make intensive use of their voice. Danford et al. (2022) observed that the manner in which vocalisation can increase pain tolerance (Swee and Schirmer, 2015) is especially evident in women's coping with childbirth. Maternal expectations of vocalising as well as the policies of health providers vary according to such factors as culture and education (McKay and Roberts, 1990; Weisenberg and Caspi, 1989). Where mothers are uninhibited, they report feeling that the noises they made were 'animal instincts' (McKay and Roberts, 1990, p. 269).

Lā and Sundberg (2010) studied the influence in women of sex steroid hormones on the larynx with a specific focus on concentrations of estrogen, progesterone and testosterone during the last trimester of pregnancy (Lā and Sundberg, 2010, p. 1). Changes to the range and quality of the singing and speaking voice are noticed more by some women than others, and can be the temporary result of dryness and increased viscosity in the mucosal tissue. Similar changes were reported by Hamdan et al., (2009), which also focused on the relationship between F0 and change to acoustic features such as shimmer, relative average perturbation and noise-to-harmony ratio.

In summary, vocalisation during childbirth presents an insight into an evolved characteristic resulting from adaptations essential to the acquisition of human behaviour, including language and culture. Few species draw attention to parturition through behaviour likely to advertise its occurrence. Thus, the endocrine and neurobiological response associated with loud vocalisation in the delivery phase, where women are traditionally isolated, is quite different to that associated with men's collective singing of sea-shanties and battle cries to lessen the potential pain of the hunt or warfare. The risks of human childbirth relate to the trade-off between infant head size at birth and the development of the greater brain-to-body ratio essential to such uniquely human features as speech and self-consciousness. Neoteny demands an unusual level of commitment on the part of human mothers and their familial support to raise infants that remain helpless for such a long period while possessing such enormous cognitive potential (Gould, 1977). While gaze (e.g., Kim et al., 2014) and touch are also important in maternal-infant attachment and preverbal instruction, the role of female vocalisation in the key stages of birth and nurturing, and its consequences for the cultural evolution of singing, merits further investigation (Woodward, 2019; Parncutt, 2019; Dissanayake, 2008), as does increasing evidence for the power of self-selected music to decrease the discomforts of childbirth and post-partum pain (Simalvi et al., 2014; Maleki and Youseflu, 2023; Arnold et al., 2024). In this regard, the peptide oxytocin has analgesic properties (Yang et al., 2022), and given its importance in parturition and lactation its potential impact on the pain of childbirth, and the subsequent memory of that event (Hu et al., 2019; Harvey, 2020), deserves further investigation.

Musicality as an adaptive correlate of social bonding

Given the key adaptive development of upright posture (Niemitz, 2010) and its consequences for bipedal locomotion (Vaughan, 2003), a candidate physical foundation on which vocal exchange and simultaneous participation would depend is indeed the capacity for rhythmic synchrony (Merker, 1999). This has been studied as a component of the vocal communication of a species with which humans share a common ancestor some 18 million years ago (Dawkins, 2005, p.126), the lar gibbon.

In an experimental study of lar gibbon communication, Raimondi et al. (2023) analysed the link between male and female contributions to call duets, and the relationship between isochrony and synchrony whereby an individual's vocal rhythm affects its partner's (2023, p. 6). They provide evidence that the two sexes influence one another's onsets during duets, which brings about changes of tempo. Such duet coordination potentially requires mutual learning and fine-tuned adaptation, evident in the higher organization and coordination of established pairs compared to new ones (Raimondi, 2023, p. 8). This could represent a precursor stage to full vocal learning (Merker, 2012): an acquired performative response sensitive to the partner's role within a duet bout. As such, it may be seen as a shared ancestral trait (Raimondi 2023, p. 9) necessary to the kinds of coordinated vocal displays that Merker (1999) proposed in his theory of synchronous chorusing as a component of vocal communication prior to language.

Synchronized chorusing, in addition of course to harmonization, is an essential sensorimotor component of the unique human capacity for communal music-making. Furthermore, as reviewed recently, "humans tend to temporally coordinate when engaging with each other", and mirror not just vocalizations but facial expressions, gesture, gait, etc. (Hoehl et al., 2021). There is also synchronisation of brain wave activity (Schirmer et al., 2016; Müller et al., 2021), and evidence of complex network synchronisation in groups playing musical instruments (Shahal et al., 2020) or singing in choirs (Delius and Müller, 2023). These psychophysiological phenomena likely underpin the adaptive function of music as a driver of social bonding and within group cooperation, as discussed by a great many contributors to the field.

Aspects of this significant contribution have been embraced in the work of, amongst others: Durkheim (1912); Roderer (1984); Storr (1992); McNeill (1995); Ridley (1996); Merker (1999); Freeman (2000); Mithen (2005); Peretz (2006); Dissanayake (2008); Levitin (2007); Bannan (2012b); Dunbar (2012, 2023); Morley (2013); Tarr et al. (2015, 2017); Harvey (2017); Podlipniak (2017); Patel (2018); Harvey (2020); Cross (2021); Savage et al. (2021); and Jordania (2022). These studies cover a number of different views of bonding from the dyadic level, including mother-infant and partner-bonding, at the neurological up to the community level, and illustrating patterns of involvement that include call-and-response, group chorusing, dance and ritual.

Debate regarding the universality of music in human cultures

Since Darwin's initial 1871 proposal that song both preceded and acted as the foundation for speech, research has sought to align the anatomical and psychological evidence for the adaptive role of music in cultural evolution with the evidence for its universality in contemporary cultures. This section reviews the issues associated with the definitions of the difference between speech and song, and the methods that have been applied to discerning whether specific aspects of musicality are universal.

Four psychoacoustic parameters (Bannan, 2008) vary vocally to expressive affect both instinctively and under voluntary control by widespread species as aspects of communication: pitch, duration, loudness, timbre. Measurement of two of these, pitch and duration, is conventional both in theory and in the variety of methods that have developed for the notation of music, as well as the new technologies available for research analysis. But, while loudness and timbre may in some circumstances be equally important to the production and perception of music, their capture in research and in musical notation has been less easily codified, often reliant on traditional Italian terms and their abbreviations, despite their variation being intentionally and meaningfully expressive.

Investigation into human vocalisation has tended to be influenced by assumptions of precedence: was there a twin divergence from a single, unitary adaptation; or does the evidence point to 'music first'? A simple thought experiment conveys the potential for doubt arising from the terminology involved: is the talking drum coded for the listener as closest to music or speech? And, by comparison: in the case of didjeridu performance, does the manner in which a single F0 generally determines all the pitch content, so that the conventional description of melody cannot apply, render it closer to speech than to music?

Cultural differences indeed present challenges to research. As a result of the observed cultural diversity of music, ethnomusicologists have often postulated that there are no universal aspects of music (e.g. List, 1984). Psychological and social research has for financial and geographical reasons tended to focus on subjects to which the acronym WEIRD (Western, Educated, Industrialized, Rich and Democratic) has been applied (Schulz et al., 2018). This represents a potential inhibition to the convincing development of models of universality, and research has been carried out in non-Western cultures such as Amazonia (Jacoby et al., 2019, McPherson et al., 2020), Africa (Fritz et al., 2009; Cameron et al., 2015; Le Bomin et al., 2016) and New Guinea (Milne et al., 2022) intended to test what aspects of musical response may be universal, and the extent to which cultural differences transmit contrasting experience. Indeed, sensitivity to theoretical assumptions in aspects of musicological investigation has become associated with the decolonisation agenda (Ewell, 2021).

A vital contribution to the philosophy and convincing presentation of data relevant to this field has been the relatively recent commitment to testing and analysing vocal repertoires and responses across populations other than the WEIRD subjects that have dominated investigation in the cognitive and social sciences. WEIRD subjects are on a global scale statistically unrepresentative of the alternative populations for which they stand in, and themselves represent products of a specialised cultural and social pattern (Schulz et al., 2018). Assembly and analysis of extensive cross-cultural corpuses promises to provide both a rich representation of the varied musical achievements of the human species, and a clearer picture of what these represent.

The speech-music distinction: evolutionary aspects and universality

Bickerton (1990) posed the ‘continuity paradox’: ‘language must have evolved out of some prior system, and yet there does not seem to be any such system out of which it could have evolved’ (Bickerton, 1990, p8). A candidate for such a missing link is a form of musical communication that emerged out of animal calls (ten Cate and Honing, 2022): a musilanguage (Brown, 2001), Hmmmm (Mithen, 2005), ‘the notion of a pre-linguistic, sophisticated, multimodal communication prototype’ (Harvey, 2017, p. 112). Language may also have evolved by co-opting certain mechanisms specific to different and previously distinct forms of communication (Podlipniak, 2022b), such as affective prosody, affective gestures, synchronous chorusing, pantomime, and sound symbolism, which may together have composed the hominin primordial communicative niche. Harvey (2017) proposes that ‘to get to the heart of the question as to why music is a human universal, we must confront the tricky, oft-debated issue of the evolutionary relationship between music and language’ (Harvey 2017, p. 97).

Recent studies have focused cross-culturally on precisely this issue: how do people distinguish aurally between musical and linguistic vocalisations, and how reliable are their perceptions? Theorization about the relationship between speech and musicality in human vocal evolution can be represented as a continuum between information exchange that is non-vocal (gesture, expression) and vocalization that has no semantic meaning (purely playful or instinctive sounds). However, some attempts at convincing accounts of the brain evolution involved in the acquisition of language largely eschew any mention at all of musicality. For instance, Aboitiz (2018) describes the development of vocal plasticity and the elaboration of proto-speech as founded on sounds that may have replaced gestures, an information-dependent model that contrasts with the social synchrony apparent in the models of Dunbar (2012, 2023), Savage et al. (2021) and others. Nevertheless, Aboitiz, interpreting the proposal of Baddeley (2007) for a phonological loop linking perception to production in working memory and thus enabling imitation and vocal learning, presents a sequence of five overlapping stages through which this may have occurred, and which led to abilities unique among humans. The sequence can be summarised (Aboitiz, 2018, p. 9-10):

1 A ventral auditory pathway that processes vocalizations and associates them with visual stimuli;

2 Increased neural control of vocalizations and orofacial movements via the laryngeal and orofacial motor cortex;

3 Activation of an auditory-vocal reciprocal loop that enables the learning of complex vocal utterances by imitation;

4 Strengthening of the associations between the ventral auditory pathway, processing the sound characteristics of vocalizations, and the ventral visual pathway processing information about objects, events, and actions;

5 The development of a dorsal pathway unique to humans that may have contributed to the transmission of lexical-semantic information and some elements of syntax.

Aboitiz (2018, p. 10) acknowledges the possibility that social bonding, group coordination and prosody may also have played a role in the development of human vocal learning, including the gradual consolidation of the phonological loop (Aboitz, 2018, p. 11) that permitted the communication of increasingly complex meanings. His focus is on how the mechanics on which vocalisation depends evolved, rather than on what the selective pressures may have been that influenced this. The two perspectives are combined within the Baldwinian model.

The multi-modal, multi-sensory processing of acoustic information associated with the acquisition of vocal learning is revealed in the manner in which ambiguities can be resolved or confused. The McGurk effect was initially explored in relation to consonant perception in speech, whereby the perception of heard syllables ‘ga’, ‘ba’ and ‘pa’ could be confused or resolved according to the precise timing of lip-movements shown on film in relation to their sound (McGurk and McDonald, 1976). In experiments derived from the implications of the McGurk effect but adapted to sung examples, Laeng et al. (2021) illustrated that the perceived pitch of singers’ voices can also vary according to their match to filmed jaw and lip movements and eye-dilation in the performers. The variability, perhaps the subjectivity, of acoustic perception revealed by these studies poses questions regarding the reliability and modal independence of human pitch perception. Perhaps therefore unsurprisingly, experimental methods illustrate that seeing a singer aids the comprehension of song lyrics (Jesse and Massaro, 2010). These studies illustrate both the complexity and the unpredictability associated with human engagement in the rich and subjective perception of vocal communication.

Two recent publications have commenced the analysis of large, varied international corpuses that they have assembled with the intention of illuminating the differences between speech and song. Albouy et al. (2023) analysed vocal recordings in both song and speech gathered from 369 subjects representing 18 different languages in 21 societies covering 12 language families. Each subject was recorded in normal, casual speech; and singing a song of their choice representing their language/culture. Recordings were subjected to digital analysis employing a machine-learning program to determine similarities and differences in spectro-temporal modulation. This revealed contrasting ‘hotspots’ in song compared to speech that identified cross-cultural regularities across countries, language families and world regions.

To ascertain whether these features were similarly discernible by listeners, 80 raters were asked to distinguish on a 5-point scale their evaluations of whether they identified each example as speech or song. Comparison between the two data sets suggest that the classification algorithm and human raters relied on the same acoustic cues to achieve categorisation (Albouy et al., 2023, p. 19), and that these result from the perception of differences in how human vocal musculature is used for speaking or singing (Albouy et al., 2023, p. 20).

Ozaki et al. (2022) comprises a team of 81 international co-authors, each of whom provided material recorded following a detailed protocol to ensure quality and consistency across examples. The recordings comprised four elements: (1) singing of a song representative of the contributor’s language/culture; (2) recitation of the song’s lyrics; (3) normal speech represented in a verbal description of the song; (4) instrumental performance of a transcription of the song. Contributors manually segmented these examples to preserve the inter-onset intervals appropriate to their language. This permitted analysis to be carried out digitally to ascertain similarities and differences arising in the following features: pitch height; temporal rate; pitch stability; timbral brightness; pitch interval size; pitch declination. This represents Stage One of their report. Full analysis of the data is in progress.

Reviewing existing evidence for the hemispherical distribution in the brain of the processing of speech and melody, Albouy et al. (2023) propose that 'speech and song are represented in distinct neural substrates not because of domain-specific aspects, but rather because of their tendency to utilize opposite ends of the spectro-temporal continuum' (Albouy et al., 2023, p. 23). This observation supports that of Kolinsky et al. (2009) regarding the differential neural processing of vowels and consonants, and would support the conclusion that music and speech have divergent functional roles (Albouy et al., 2023, p. 24). Building on a distinction explored by Caramazza et al. (2000), Kolinsky et al. (2009) suggest the adaptations necessary for such functions: 'only humans have elaborated the supralaryngeal articulations that, by inserting consonants into the vocalic carrier, allow the emergence of a rich set of meaningful contrasts' (Kolinsky et al., 2009, p. 2). A parallel neural laterality identified emotional speech prosody with processing in the right hemisphere (Pell, 2005). Recent imaging studies reveal separate processing regions for speech and music in human non-primary auditory cortex (Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015), yet even normal speech reveals patterns of musical modulation (Robledo et al. 2016). In this regard, when imaging activity in the cerebral cortex as a whole it is interesting to note considerable areas of overlap when generating sentences or melodies (Brown et al., 2006; see also Perrachione et al., 2013).

The studies of Ozaki et al., (2022) and Albouy et al., (2023) address perceptual issues arising from the acoustic presentation of song and speech in recordings of individuals. The influence on the two media of differences of function, whereby song may be more strongly associated with collective participation rather than individual utterance, was not a focus of the research. This theme remains significant in accounts of the role of music in group behaviour (Dunbar, 2012; Harvey, 2017; Haiduk and Fitch, 2022; Savage et al., 2021).

The role of the Harmonic Series (HS) in music and language

Western musical theory has been divided for more than two millennia between advocates of the Pythagorean model of pitch relationships derived from the proportions of the HS (Di Stefano et al. 2022), and the view of Aristoxenus that practice and preference in the employment of scales and intervals arises from empirical experience (Cazden, 1958). Divergent interpretations of the role of the HS in vocal perception can subsequently be traced to the influence of three close contemporaries: Charles Darwin, Hermann Helmholtz and Alexander Ellis. Both Helmholtz and Ellis worked on the categorical perception of phonetics in language and its relationship to harmonic frequencies. For Helmholtz, harmonics determined musical perceptions in speech. Darwin's musically literate sons, George and Francis, were familiar with Helmholtz's work and its implications for clarifying their father's 1871 speculation on the precedence, on the lines of sexual selection, of the precedence of musical vocalization. But Darwin senior was hesitant to develop his proposal on acoustic lines that he did not understand (Bannan, 2017). Ellis, himself the translator into English of Helmholtz's magnum opus (Helmholtz, 1885) and inventor of the cents system for expressing precise tuning in relation to the division of the 12 semitones of the equally-tempered octave, was also the designer of a method for notating the phonetics of speech (Stock, 2007). But his view of the tunings employed in music was that they did not arise from responses to HS perception, but rather to human preferences and cultural transmission. These contrasting positions remain divergent to this day, and have been taken to represent more than merely scientific interpretation.

The properties of the HS are clearly not just a 'Western' discovery or cultural construct. They were known to the ancient Greeks (Franklin, 2002; Wood, 2022), associated especially with the figure of Pythagoras, and widely throughout Asian science, both Indian (Krishnaswamy, 2003) and Chinese (von Falkenhausen, 1992), and the cultures to which this spread. They formed the theoretical basis for the construction and tuning of instruments, whether plucked, bowed, struck, or dependent on reed- or lip-vibration (Lawson, 2024). Nevertheless, the appreciation of non-Western musics within music histories and theoretical writings had previously tended to imply a distinction between accepted practice and the unfamiliar:

> The musical scale is not one, not "natural," not even founded necessarily on the laws of the constitution of musical sound so beautifully worked out by Helmholtz, but very diverse, very artificial, and very capricious (Ellis, 1885, p526).

Ellis's (1885) presentation of the natural/unnatural was in many ways provided as a corrective to the cultural prejudices of his time, which accepted Western music as the pinnacle of the art and tended to hear the music of other cultures as disorganised or ugly. For instance, Budasz (2004) records the dismissive response ('blasts') of listeners to a 1730 performance by Brazilian musicians in Portugal; and Pigott (1891) sums up the resistance of European audiences to oriental music:

> The decorative arts of the East–of India, China, Japan–have rarely failed to charm and captivate the Western mind … With this appreciation compare our attitude towards Eastern music–what a contrast! … It is clear that whilst our eyes welcome the Entirely Strange, our ears do not (Piggott, 1893, p. 319).

Ellis was supported in his research by the exceptional pitch discrimination of Alfred Hipkins (1895), employing the laboratory resources of the Broadwood piano manufacturers for whom he worked. Ellis would have been fully aware of the laws of physics represented in the HS, but his view of non-Western tunings was that they were all equally valid (see also Parncutt and Hair, 2018; Phillips and Brown, 2022); and could not therefore be assumed to derive from the HS, either formally (as a consequence of measurement and theory) or instinctively (as a function of human hearing). Ellis's argument for the freedom with which tuning systems can be constructed is paralleled by Garcia's (1894) proposal that there is an infinite number of vowel sounds. Garcia, inventor of the laryngoscope and the most successful singing teacher of his era, was able to observe the complex muscular interactions within the vocal tract, the tongue and the lips that were capable of generating infinite timbral variants. What we are left with, therefore, is the conundrum: if both tunings and vowel sounds can take an infinite number of forms, what governs the categorical perception that tends to limit these so that they can be employed meaningfully through sharing, imitation, and recall? In terms of the evolutionary origin of human pitch perception, decoupling the arrangement of musical scales from any relationship with HS responses merely invites an infinite regression: if not 'then', 'when'? The same applies to the categorical perception of vowels. The influence of Ellis's well-informed choice to express all scales as being equal without underlying cause has been employed to stifle debate on how human preference for organised sound could have arisen at all, and this has tended to fuel objections to universalism. But how might we deduce an ultimate underlying cause?

Parncutt and Hair (2018) present a closely argued dismissal of the traditional Western understanding of intervals and concords as deriving from frequency ratios on lines associated with the formulation of Pythagoras. Their principal argument is that the tuning of instrumental and vocal lines analysed digitally illustrates that performances deemed acceptable by listeners vary by considerable amounts from the assumed target pitches; that such imprecise tunings are often preferred to those predicted by precise frequency ratios; and that listeners cannot discern whether intervals are tuned as they assume them to be. Their conclusion is 'Bye Bye Pythagoras'. They argue (2018, p. 475) that 'no known brain mechanism is sensitive to ratios in musical contexts' (but see Foss et al., 2007), and that 'Musical scales and intervals, although they depend in part on acoustic factors, are primarily psychocultural entities–not mathematical or physical. Intervals are historically and culturally variable distances that are learned from aural traditions'.

A re-examination of the kinds of acoustic experience to which their dismissal of frequency ratios applies reveals gaps in their account that suggest the baby may be in danger of going out with the bathwater. The rigor of their argument - essentially, that modern science undermines confidence in the oversimplified logarithmic ratios represented in the Pythagorean tradition - is highly convincing, at least in application to the realities of human culture, the subjectivity of aural perception (Michael and Gilman, 2021), and cultural preferences. But phenomena in the natural world as well as the practices of musicians in widespread cultures present a case for the ongoing role of frequency ratios that cannot be ignored.

In a parallel study which sets out to propose an alternative to the Pythagorean system that purports to derive scale tunings from harmonicity, Phillips and Brown (2022) explore the idea that tuning systems more naturally adopt properties that derive from a form of intracultural trial-and-error. Indeed, they focus explicitly on the consequences of ‘inaccurate’ and ‘imprecise’ pitching as the active procedures that may account for scale intervals in a given culture. There is a circularity to the presentation of this argument against harmonicity associated with Pythagorean ratios: in order to express the tuning of real-life musical examples, they are labelled as failing to conform to the accepted yardstick of 12-tone equal temperament that is itself historically derived from a mathematical, linear variant of HS relationships. Hello, Pythagoras, nice to have you back …

The models of Phillips and Brown (2022) and Parncutt and Hair (2018) fail, in their exclusive focus on tuning, to capture the complex interactive features of timbre, spectral features and prosodic expressivity that meaningfully play on perceptual affordances (Michael and Gilman, 2021; Marjieh et al., 2024) and illustrate the subjective, even emotional, response to all four parameters of vocal sound rather than only the F0. In relation to the cultural evolution model of Mehr et al. (2021), the characteristics of human vocal signalling may be taken to represent two different kinds of ‘credibility’: those associated with the perceived vocal norm of the signaller (range, clarity, power); and those that convey state of mind or its deceptive representation (timbre, tempo, prosodic features of contour).

The language employed in the approach of Phillips and Brown (2018), ‘imprecise’ and ‘inaccurate’ pitching relative to an assumed 12-note equal temperament model that they nevertheless reject as an invalid and recent invention, represents a circularity of argument in evolutionary terms. Progress needs to be made in resisting the Western norm as the measure for different musical styles that developed entirely independently (and even chronologically earlier), but an agreed language for doing so has yet to be devised; though the locally-informed methods of Arom (2022), in which representatives of a culture are able to manipulate digital analyses in search of satisfactory outcomes that meet with their culturally-derived aesthetic judgements, perhaps offer a way forward. Recent research has found that these cultural preferences are influenced by acoustics. For example, the sensation of consonance can be affected by both HS and spectral features, the interplay of which can lead in turn to the appearance of different musical scales (Marjieh et al., 2024). Nonetheless, extensive evidence that the HS is foundational to the core elements of human music and language transmission, both through explicit knowledge and application of its properties and through instinctive or implicit exploitation of them, remains persuasive.

12a. Comparative responses to the Harmonic Series across species

If we are to claim that HS response played a part in perceptual and productive adaptations that shaped human capacity for vocalisation, there should be evidence of parallels or pre-adaptations in other species. While there are close similarities between human examples and the characteristics of pitch and duration contour of vocal communication in a wide range of species (Scherer, 1992), HS responses need to provide evidence of categorical perception that determines sensitivity to the critical aspects of pitch and formant.

Passive responses, where perception is not matched by production, are common in domestic animals that are trained to act on human commands. These tend to be in the medium of speech rather than music: ten Cate and Honing (2022) distinguish between the cognitive and auditory skills involved in the two media of human vocal communication through adopting the terms musicality (Honing et al., 2015) and linguisticality (Haspelmath, 2020). One aspect of their proposal for comparison between human abilities and those of animals is the extent to which harmonic sounds are recognised by their pitch (as in music) or by their spectral structure (as in speech) (ten Cate and Honing, 2022, p. 3). There is a depth and complexity of passive response to human language of which some species, especially dogs, are capable despite their anatomical inability to vocally reproduce the sounds that they can clearly perceive (ten Cate and Honing, 2022, p. 6).

Feedback loops develop where there is an iterative balance between acoustic perception and production (Baddeley, 2007). Some animals such as dogs (Cuaya et al., 2022) possess sophisticated aural perception allowing categorical discrimination between elements of human vocal signals (pitch, vowels, consonants) that permit them to make appropriate behavioural responses. But they are entirely incapable of imitating what they hear. By comparison, human capacity for vocal learning represents a unique integration of auditory feedback with vocal motor control (González-Garcia and Rendón, 2017) in which response to harmonicity plays a part in vocalisation.

While the role of Pythagorean ratios in underpinning the nature and analysis of human pitch organisation has been questioned by Parncutt and Hair (2018), Cardoso (2013), accounting for pitched signals in animal communication, argued that frequency ratios rather than scalic relationships represent a truer impression of their acoustic production. '… the assumption of linearity … is artificial from the perspectives of sound perception and of certain aspects of sound production' he argued; 'logarithmic perception of frequency should be related to physical properties of sound, whereby certain frequency ratios (harmonics) have a higher proportion of pressure peaks in phase, and many sounds in nature, notably animal vocalisations, have harmonic overtones' (Cardoso, 2013, p. 1529). Cardoso considered the extent to which signals represent the body size of their sources: 'since logarithmic scales reflect the exponential relation between anatomy and resonating frequency' … 'applying identical tension to vibrating structures (e.g., vocal cords or syrinx labia) of different sizes should result in changing their vibrating frequency by identical ratios rather than identical differences' (Cardoso, 2013, p. 1530). Cardoso re-analysed a corpus of animal signal studies in order to ascertain whether his model of logarithmic rather than linear relationships conveys a more accurate account of frequency changes in relation to the mechanical adjustment of vocal tract resonance. This in turn provided a clearer understanding of rate of iteration of syllables in relation to frequency differences in terms of the effort involved in attaining these (Podos, 1997). 'Since adjusting the vocal tract moves larger masses (oropharyngeal cavity, beak) than changing fundamental frequency … adjusting the resonance of the vocal tract is the limiting factor for fast frequency modulations' (Cardoso, 2013, p. 1530).

A fruitful avenue for analysing acoustic change in non-human species arises where they have encountered competition in their environment from humanly produced noise (Bateman et al., 2003). Dolphins (Fouda, 2018) and birds (Hu and Cardoso, 2010) modify their calls in perceiving acoustic competition from the environment, including humanly produced sounds, that requires them to change their pitch or duration: 'The average minimum frequency of urban birds was higher than that of nonurban conspecifics in all species studied' (Hu and Cardoso, 2012, p. 865). Anthropogenic noise can have negative effects on normal social behaviour in bottle-nosed dolphins (Sørensen et al., 2023).

If in some environments, both natural and anthropogenic (e.g., urban), there is a 'war of the frequency spectrum', what evidence is there that this may previously have represented an adaptive pressure that shaped signal quality? Where research is revealing that anthropogenic noise has proved disruptive of the lives and communication of non-human species (Bateman et al., 2023), how might proto-human vocalisation in deep history have responded to acoustic competition in the rainforest and savannah environments of our distant ancestors, both by day and at night? Balancing such a capacity to outperform by means of changing the frequency of the signal and increasing amplitude, early humans seem also to have acquired the parallel ability to blend in a controlled collective presentation in which no individual stands out (Merker, 1999).

12b. Ethnomusicological perspectives on responses to the Harmonic Series

While HS responses are evident in the perception and vocalisations of a range of animal species, this does not imply that they are clearly and exclusively determinant of human vocal production and generativity. Discerning their influence below the surface, however, may prove a fruitful approach to the extent to which they play a role. Human vocalists do not automatically produce patterns of harmonically related pitches: it requires effort and practice. Western-music opera singers rehearse the accurate and evenly connected arpeggios that are key components of the repertoire they need to master (McHenry et al., 2016), and the equivalent of such careful practice has been documented in the learning of Inuit singing-games (Nattiez, 1999), Tuvan overtone singing (Zemp, 1989) and other repertoires in which the vocal techniques employed may vary considerably (Trân Quang and Bannan, 2012). Singing in-tune is a practised art. The role of the HS in human vocal production is perhaps more naturally evident in the differentiation of timbral variants related to facial expression and emotional state that become harnessed as distinct vowels (Manén, 1974). The use of these is culturally acquired by the speakers of any particular language, and the tendency to produce the voice in a characteristic manner dependent on vocal learning presents itself as the accent which permits listeners to detect where speakers may come from that can correlate with responses in the laboratory to ambiguous musical intervals (Deutsch, 1999).

12c. Harmonicity, the Pythagorean tradition, and human musicality

Research both in animal species and in a variety of human societies illustrates that perception of harmonicity represents an instinctive response that can be harnessed to underpin expectation, behaviour, and, in humans, cultural assumptions. In historical terms, the influence this has lent to speculation on the natural world, and within varied fields of human experience, represents a long and widespread tradition. We can envisage both weak and strong effects of harmonicity: as a powerful metaphor that can be employed to conceptualize relationships; and as an actual causal factor in human development. The interactions between these two layers of reality explain the hold that harmonicity theory has upon our understanding of acoustic response.

Lavenda (1985) illustrated how, in the nineteenth century development of atomic theory, hypotheses depended on the working acceptance of ‘convenient fictions’, a term derived from an observation of Mach regarding ‘an ideal mental-economical unity’ of abstraction inferred from and then imposed upon sensations (1886/1914, p. 24). While Pythagorean ratio theory may have been treated as a musical reality prior to the invention of digital instrumentation able to discern with accuracy the precise pitch relationships it was traditionally assumed to capture, it may perhaps continue to be viewed as a ‘convenient fiction’ on similar lines. Certainly, its influence on musical thinking should not be overlooked.

In ancient China, where a tradition of music theory developed independently of that of the Pythagoreans in Greece (von Falkenhausen, 1992), frequency ratios were employed at the heart of a system for maintaining fairly the weights and measures on which trade depended (Vogel, 1994). The system of Hanshu integrated the concept of lü (harmonic ratios) into the properties of number, pitch, length, capacity, and weight. The pitch of a musical pipe presented the standard against which measurements in all other domains could be derived. For instance, the frequency could be transferred to bronze bells for measuring grain, the capacity of which could be checked by comparing their pitch (von Falkenhausen, 1992).

The conceptual power of Pythagorean ratios extended its influence to cosmology and the scientific examination of matter. When Kepler devised the presentation of a heliocentric solar system, the assumption that the relative positions of the sun and planets may relate to scale tunings (Gingerich, 1992) contributed to estimations of distance. Both Galileo and Copernicus followed this example. When Mendeleev (1869) commenced the laying-out of atomic relationship in the Periodic Table, he assumed a kind of scalic system in which the octave played a part in representing iterative functions represented on the vertical axis. Scaled relationships of this kind are far from rare in nature. Both the size distribution and the internal structure of primate (including human) social groups, for example, have a distinctive scaled structure with a scaling ratio of 1:3 created by the efficiency of information flow around networks (Dunbar, 2020; West et al., 2023). These ‘convenient fictions’ are not without explanatory power, not least because they have often leant towards spiritual associations as well as factual interpretations.

Such historical and cultural examples contextualise the thorny issues relating to the nature of harmonicity raised by Parncutt and Hair (2018) as well as Phillips and Brown (2022). However, the ultimate intention of examining these arguments is to weigh up the evolutionary potential of the human abilities in question: could sensitivity to harmonic series properties have played an adaptive role in human development? Bannan et al. (2024) set out to illustrate the kind of sequence that may account for the relationship between the perception and production of F0 and formant resonances in the long evolutionary journey traceable back to common ancestors with the gibbons and chimpanzees (see also Raimondi et al., 2023). It would be consistent with viewing octave equivalence and the HS response as adaptive, and foundational to human musicality, were exploitation of their properties to be revealed as both implicit and explicit. Wood's (2022) account of tuning systems illustrates how ancient theoretical understanding based on mathematical experiment developed in widely separated cultures. Implicit understanding evident in musical styles and techniques is no less impressive (Lawson 2024). Examples would include: the repertoire of timbral gestures and variants that represent Australian native mastery of the didjeridu, and also the signals played on lip-blown instruments from the conch shell to the shofar to the military bugle (Briney and Montgomery, 2015). These two strands – explicit application of HS theorisation in instrument construction and performance techniques; and implicit exploitation of HS properties in human vocalisation – will be explored in this section.

Explicit examples would include those in which either the principles of construction (relative size; positioning of finger-holes; the length of organ pipes) represented conscious application of HS ratios; or those in which playing techniques have developed to exploit HS properties. Musical instruments that comprise activation of columns of air present the potential for melodic variation through adjustment of air pressure and direction that produces frequencies related to the F0 which may be, to a greater or lesser extent, well-tuned harmonics. Espi-Sanchis has shown how found objects can produce a range of pitches in response to such blowing techniques (Espi-Sanchis and Bannan, 2012), for instance in cases such as the Lekolilo flute and its cousins in southern Africa that have equivalents in many cultures around the globe. Flutes activate air columns through blowing across a hole in a manner that initiates oscillation: the lips need not touch the instrument. Variation in air pressure and direction can produce overblown pitches that conform to the HS. For ease of playing, the lips can embrace a mouthpiece containing a fipple that achieves oscillation in a similar fashion. Simple instruments with such design are found on every continent and examples have been traced back to about 40,000 BCE (Conard et al., 2009).

Oscillation can also be achieved by inserting a single reed instead of a fipple into the mouthpiece (bülban, chalumeau, clarinet, saxophone); or by placing two reeds in opposition either in contact with the lips (oboe, duduk, piffero, zurna); or freely vibrating within a mouthpiece or the body of the instrument (crumhorn, sheng, mouth organ). The capacity for overblowing, and the pitches of the HS available, vary according to the technique for sound production and the shape (conical, cylindrical) of the instrument.

An instrumental family that almost certainly derived originally from performance on 'found objects' (Espi-Sanchis and Bannan, 2012) requiring minimum modification (such as shells, animal horns and tubes of wood or bamboo) comprises those instruments that respond to and amplify lip-vibration ('blowing a raspberry'). These can be extremely loud, and are employed for ritual functions (the Tibetan dungchen, the Hebrew shofar) and military signalling (bugles, posthorns). Lip-vibrated instruments generally give rise to clear and distinct overtones derived from the HS, which lend themselves to musical melody, polyphonic harmony (e.g., alphorn ensembles) and the precise clarity of contour essential to unambiguous military signals.

An interesting facet of the acoustic potential of wind-blown instruments is the capacity to adapt airflow or combine it with vocalisation to access complex pitched effects known as multiphonics. These have become part of the technique of jazz musicians (Riera, 2014) and wind techniques of modern Western art music (Bartolozzi, 1967), though evidence of the technique can also be heard in didjeridu performance (Fletcher, 1996). Multiphonics would appear both to cross the implicit/explicit divide, and to represent a perceptual puzzle: are they heard as variants of pitch, or of timbre? We will consider this issue in dealing with vocal subharmonics and 'fry', below.

Explicit examples of the presentation of harmonics include the alternate performance technique on stringed instruments such as the guitar (Absil, 2018), harp (Waltham, 2010) and the Chinese guzheng (Gaywood, 1996) in which harmonics are produced by plucking in a manner that activates vibration at specific nodal ratios of the string length that simultaneously dampens linearly proximate positions so as to result in a pure, bell-like timbre.

A further illustration of HS response is the issue of sympathetic vibration: a sought-after characteristic in instrumental design, especially in synchronised multiple-participant performance, in which it confers acoustic advantages (resonance, amplification, sustaining). Where instrument-making has been systematically studied, the application of HS properties both to the achievement and avoidance of harmonicity is evident. Examples of the former can be found in different traditions of Indian music (Weisser and Demoucron, 2012; Demoucron and Weisser, 2012). The latter comprises instruments in which sympathetic vibration is an unwelcome feature to be avoided, such as in the tuning of pianos (Giordano, 2015).

We now turn to examples of the implicit incidence of harmonics in human vocalisation. Acoustic research in both sung and spoken corpuses has revealed that low pitches (designated subharmonics (Lindestad et al., 2001) or undertones (Kob et al., 2011) in vocal music and vocal fry (Childres and Lee, 1991) in speech) can be generated by logarithmic rather than linear functions that can be learned as varied postures of the laryngeal musculature. The subharmonic series can be accessed through an achievement of 1:2 and even 1:3 ratios of frequency variation in relation to normal modal voice production, achieving the capacity to deepen the voice considerably. In language, pitch range is associated with dominance (Puts et al 2006), and vocal fry has become a normalised element of speech in both men (Abdelli-Beruh et al., 2014) and women (Chao and Bursten, 2021). In music, the more carefully and precisely phonated equivalent gives rise to specialised solo and ensemble repertoires that have been studied in women singers in Southern Africa (Dargie 2018) as well as expert male performers (together with some females) in North-East Asia (Mongolia and Tuva) (Trân Quang, 2002). Through x-ray videography and computer voice analysis (Trân Quang, 2002; Zemp, 1989), the scientific basis for achieving these kinds of performance, with their distinct timbres and harmonic characteristics, has become fully understood in both anatomical and acoustic terms. We now have an explicit capacity to share, learn and understand these techniques. But their implicit transmission to new generations via vocal learning may well be thousands of years old, associated as they are with the achievement of exalted states that serve magical and religious functions. So, the distinction between linear scale systems and the logarithmic pitch relationships that underpin such vocal styles may be both ancient and instinctive (see also Nikolsky et al., 2020).

Acoustic analysis of the signals of widespread species illustrates commonalities within which human behaviour both relates to models shared across other species, and possesses a distinction of pattern-making representative of the properties of music (Kershenbaum et al., 2020; Jermyn et al. (2023; Elhilali 2019; Scherer 1992). Inter-species comparisons on these lines may help to reveal the deep-time evolutionary processes that bequeathed the suite of abilities on which human vocalisation depends. What these have led to is fine-tuned discrimination that assumes great significance in the ways in which we listen and respond. For McDermott et al. (2010), human response to relative pitch – the capacity to assign meaning to, learn and recall the precise relative intervals of musical material even when it is transposed to a different range – has the characteristics of an adaptation: ‘it is a defining property of music perception, it is effortlessly heard by humans from birth, suggesting an innate basis, and there are indications that it might be unique to humans’ (McDermott et al., 2010, p. 1949).

12d. A spectrum embracing both harmonicity and inharmonicity

The evidence points towards an adaptive role for the detection of acoustic modulation that responds to degrees of harmonicity representing a continuum onto which sounds can be mapped, from those that most closely reveal HS relationships to those that most completely remove or confuse them. Adaptive responses that commence instinctively can be refined via vocal learning and enculturation, as well as in their application to the construction of musical instruments.

Milne et al. (2022), compared responses to musical stimuli of participants in urban Australia and an isolated region of Papua New Guinea. They elicited responses to the feature of stability in melodies and chord progressions, recording a universal effect of psychoacoustical roughness on perceived musical stability (Milne et al., 2022, p. 11; see also Yurdum et al. 2023). Popham et al. (2018) propose that sorting harmonicity of a signal against background noise accounts for the 'cocktail party effect' whereby we hear salient speech despite unpromising acoustic conditions.

Response to harmonicity is evident in the relationship between vowels, voiced consonants and unvoiced consonants in speech, and between sung vocal techniques and language (Ikhtisamov, 1988; Dargie, 2018; Nikolsky 2015, 2020; Trân Quang, 2002; Trân Quang and Bannan, 2012). A fascinating example occurs in the whistled speech adopted where speakers are too widely separated to hear one another, and employ whistled pitch contours that map onto the vowel content of phonetic sequences in language and are capable of being heard over distances (Classe, 1956; Meyer, 2021). Similarly, abrupt extensions of over an octave between distinct vocal ranges can be found in mountainous areas in which such sounds can be employed to bridge separations and steep differences in height. These may involve moving from modal register to falsetto, which occurs also in yodelling and its equivalent around the world (Trân Quang and Bannan, 2012; Nikolsky, 2020). There would seem to be a discernible role for harmonicity in the 'beau geste effect' (Jordania, 2022, after Wren 1924) whereby the acoustic signal of simultaneously produced vocalisation both appears amplified and, in consequence, provides the impression of a greater number of performers. Depending on the intention of the signal (warning, greeting, dominance, social cooperation), might the achievement of collective amplification through both dissonant and consonant pitch combinations provide an insight into the role of harmonicity in establishing human capacity for combining signals?

Even where 'choices' that define a tuning system are revealed to be inharmonic, the perception of pitch may nevertheless relate to harmonic consequences (Marjieh et al. 2024). While certain vocal repertoires (e.g., Barbershop (Averill 1999); ensemble styles in Sardinia (Castellengo et al., 2001)) achieve inter-performer fusion through conscious and deliberate HS reinforcement, instrumental distribution such as employed by the resonant metallophones of gamelan may have been shaped by precisely avoiding HS properties that would have the effect of unwanted sympathetic interference between constituent sound sources. For instance, the 'stretched octaves' for which Demany et al. (2021) found preferences may explain how, in widely separated cultures, the tuning of instruments results from the preservation of timbral distinctiveness through the avoidance of harmonic fusion. In such forms of polyphonic music, HS ratio avoidance preserves the independence of individual contributions. Wood's (2022) systematic survey of the earliest records of tuning systems in Chinese, Ancient Greek and Indian music theory illustrates well the nature of the spectrum linking harmonicity to the inharmonic, the control of which in instrumental design and practice depends entirely on understanding of the properties of the HS while also leading to informed aesthetic preferences that can represent departure from HS norms. An example of this is the provision for gamelan orchestras in Indonesia of highly distinct tuning differences with the intention of giving the music-making of each ensemble in its location its own unique sound through which it can be recognised (Vetter, 1981).

Li (2006) illustrated the crucial role of inharmonicity in the tuning of Javanese gamelan instruments in order to achieve the appropriate blend of instruments within the ensemble that does not reinforce sympathetic or mutually resonant properties. Strumolo (2007) investigated similar properties in the tuning of the Classical Thai Renat Ek. Schneider and Leman (2002) considered the correct balance between harmonic and inharmonic responses that allows the optimum tuning of carillon bells.

The horn parts of Classical symphonies depend largely on the availability of pitches generated within the HS of the (predominantly) tonic key. Where composers were impelled to transcend this limitation, one encounters the splendid inharmonicity of those pitches outside the series that one hears in performances on period instruments of, for instance, the virtuoso solo horn writing in the ‘Quoniam tu solus’ movement of Bach’s B minor Mass, or the shaded horn harmonies encountered in Beethoven’s Eroica Symphony. Whether or not performers and listeners were schooled in the neo-Pythagorian understanding of intervals dismissed as illusory by Parncutt and Hair (2018), it is unquestionable that these musical experiences are perceived due to familiarity with HS relationships.

Conclusions: An integrated model of human responses to the Harmonic Series

A convincing model for the gradual adaptive processes that endowed the varied abilities on which human musicality depends needs to account for the complex interactions between the physical and mental operations which control meaningfully the parameters of the fundamental frequency of harmonic sound (pitch), amplitude (loudness), duration (rhythm) and timbre that combine in vocal communication. Rhythm would appear to derive from human bipedal locomotion, and its consequent freeing of the hands to permit gestural movement and percussive effect. Mother-infant interaction, exemplified by unison imitation, and its legacy of intimacy re-experienced in aspects of mate attraction and relationship (Dissanayake, 2008), plays an early part in socialisation that gives way to group play and collective vocalisation which sustains acquired repertoire and endows membership of the social unit. From this emerges creativity, individuality, and the potential for leadership roles, including novel vocal character in males as their voices mature and deepen through adolescence, when imitation and participation in complementary ranges responds to the natural harmonicity of octave equivalence. The roles and capacities of female and male voices can, according to cultural norms in operation, both diverge in function and remain acoustically harnessed within the overall multi-generational group.

In terms of the impact most immediately made on the listener, the timbre of a signal is of at least equal importance to that of its pitch and rhythmic characteristics, if not demonstrably greater. Switch on the radio to a piece of music in full flow and it is the timbre that defines what is being presented: whether it is a brass band, a gamelan ensemble, a church organ, or an a cappella folksong. Pitch and duration operate within this prevailing timbral character to generate the stylistic features of musical structure. Timbre, through the distinct categories of phonemes (alphabets), both vowels that preserve the harmonic features of song and consonants which initiate, divide, conclude and articulate linguistic utterance, plays a different role in the medium of speech to that it lends to music.

Human acoustic communication emerged in the full complexity of music and language as a consequence of its vital significance: initially, to survival; subsequently, to cultural enrichment. In the search for clarity of meaning and fitter tools for survival, vocal communication that already permitted learning, imitation, recall, and reference (Podlipniak, 2022a) combined with categorical functions that laid the foundations for spoken language. Once articulacy became yoked to grammar and syntax, a virtual Babel of disparate, geographically-separated proto-languages would have provided a powerful Baldwinian incubator for accelerated vocal learning. In turn, procedures emerging from the repertoire of sung and spoken generation were utilised to inform the musical employment of tools – musical instruments – that make use of materials available in the environment, or skilled manufacture of its products (skin, bone, stone, bamboo, wood, metal, glass), to provide increased range and loudness as well as alternative timbres to those produced by the voice (Lawson 2024).

Acknowledgments

The authors collaborated on the writing of this article in response to their involvement at the invitation of Frontiers in Psychology as co-editors of the Special Issue on “The Adaptive Role of Musicality in Human Evolution” (2023-4). We would like to express our gratitude to authors whose work was accepted in the Special Issue and which contributed as a consequence to the perspectives covered in our article: Chirag Rajendra Chittar et al.; Steven Brown and Elizabeth Phillips; Steven Jan; Yoshimasa Seki; Joseph Jordania; Joshua Bamford et al.; Eirimas Velička.

WORKS CITED

Abdelli-Beruh, N. B., Wolk, L., and Slavin, D. (2014). Prevalence of vocal fry in young adult male American English speakers. J. Voice 28, 185-190.

Aboitiz, F. (2018). A brain for speech. Evolutionary continuity in primate and human auditory-vocal processing. Front. Neurosci. 12, 174.

Absil, F. (2018). Guitar harmonics; an overview in diagram. https://fransabsil.nl/archpdf/guitar_harmonics.pdf

Ackermann, H., Hage, S. R., and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. Behav. Brain Sci. 37, 529-546.

Albouy, P., Mehr, S.A., Hoyer, R.S., Ginzburg, J., and Zatorre, R.J. (2023). Spectro-temporal acoustical markers differentiate speech from song across cultures. BioRxiv. https://doi.org/10.1101/2023.01.29.526133

Altenmüller, E., Siggel, S., Mohammadi, B., Samii, A., and Münte, T.F. (2014). Play it again, Sam: brain correlates of emotional music recognition. Front. Psychol. 5, 114.

Angulo-Perkins, A., Aubé, W., Peretz, I., Barios, F.A., Armony, J.L., and Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. Cortex 59, 126-137.

Arnold, C.A. Bagg, M.K., and Harvey, A.R. (2024). The psychophysiology of music-based interventions and the experience of pain. Front. Psychol. 15, 1361857.

Arom, S. (2022). Observation and experimentation in ethnomusicological research. In J. L. Friedman (ed), Music in human experience: Perspectives on a musical species, 91-104. Cambridge, UK: Cambridge Scholars Press.

Aung, T., Hill, A.K., Pfefferle, D., Mclester, E., Fuller, J. et al. (2023). Group size and mating system predict sex differences in vocal fundamental frequency in anthropoid primates. Nature Commun. 14, 4069.

Averill, G. (1999). Bell tones and ringing chords: Sense and sensation in barbershop harmony. The World of Music 41, 37-51.

Baddeley, A. (2007). Working memory, thought, and action. Oxford: Oxford University Press.

Baldwin, J. M. (1896). A New Factor in Evolution. The American Naturalist, 30, 441-451.

Ballot, B. (1845). Akustische versuche auf der niederländischen eisenbahn, nebst gelegentlichen bemerkungen zur theorie des hrn. prof. doppler. Annalen der Physik 142, 321-351.

Bannan, N. (2017). Darwin, music and evolution: new insights from family correspondence on The Descent of Man. Musicae Scientiae 21, 3-25.

Bannan, N. (Ed.). (2012a). Music, language, and human evolution. Oxford: Oxford University Press.

Bannan, N. (2012b). Harmony and its role in human evolution. In N. Bannan (ed.), Music, language, and human evolution, 288-339. Oxford: Oxford University Press.

Bannan, N. (2008). Language out of music: The four dimensions of vocal learning. Aust. J. Anthropol. 19, 272-293.

Bartolozzi, B. (1967). New sounds for woodwinds. Oxford: Oxford University Press.

Bateman, P. W., Gilson, L. N., and Bradshaw, P. (2023). Not just a flash in the pan: short and long term impacts of fireworks on the environment. Pacific Conservation Biology. Clayton, Australia: CSIRO Publishing.

Békésy G. von. (1928). Zur Theorie des Horens; Die Schwingungsform der Basilarmembran. Phys Z. 29, 793–810.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. Nature, 436, 1161-1165.

Benedict, L., Charles, A., Brockington, A., & Dahlin, C. R. (2022). A survey of vocal mimicry in companion parrots. Sci. Rep. 12, 20271.

Benítez-Burraco, A., and Nikolsky, A. (2023). The (Co)Evolution of Language and Music Under Human Self-Domestication. Human Nature, 34, 229–275.

Bickerton, D. 1990. Language and species. Chicago: University of Chicago Press.

Blood, A.J. and Zatorre, R.L. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc. Nat. Acad. Sci. (USA) 98, 11818-11823.

Bodin, C., Trapeau, R., Nazarian, B., Sein, J., Degiovanni, X., Baurberg, J., et al. (2021). Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques. Curr. Biol. 31, 4839-4844.

Borroto-Escuela, D.O., Carlsson, J., Ambrogini, P., Narváez, M., Wydra, K., Tarakanov, A.O., et al. (2017). Understanding the role of GPCR heteroreceptor complexes in modulating the brain networks in health and disease. Front. Cell Neurosci. 11, 37

Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. Proc. Natl. Acad. Sci. (USA), 113, 1666-1671.

Briney, B., and Montgomery, K. (2015). Intersecting history and pedagogy: using the natural trumpet to teach beginning brass students. International Trumpet Guild Journal, 34-7. http://www.bfbresources.com/wp-content/uploads/2017/12/201503Natural.pdf

Brown, S. (2000). The "musilanguage" model of human evolution. In: Origins of Music. MIT Press, Cambridge MA, pp 271-300.

Brown, S. (2001). Are music and language homologues? Ann. NY Acad. Sci. 30, 372-374.

Brown, S., Martinez, M.J., and Parsons, LM. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. Eur. J. Neurosci. 23, 2791-2803.

Budasz, R. (2004). A Música no Tempo de Gregório de Mattos / Music in the time of Gregório de Mattos. Brazil: Universidade Federal do Paraná.

Burkart, J. M., Schubiger, M. N., and van Schaik, C. P. (2017). The evolution of general intelligence. Behav. Brain Sci. 40, e195.

Cameron D. J., Bentley J., and Grahn J. A. (2015). Cross-cultural influences on rhythm processing: reproduction, discrimination, and beat tapping. Front Psychol 6, 366.

Caramazza, A., Chialant, D., Capasso, R., and Miceli, G. (2000). Separable processing of consonants and vowels. Nature 403, 428-430.

Cardoso, G. C. (2013). Using frequency ratios to study vocal communication. Animal Behav. 85, 1529-1532.

Castellengo, Michèle, Bernard Lortat-Jacob, and Gilles Léothaud. (2001). Pitch perception: five voices with four Sardinian singers. Int. Symp. Musical Acoustics, pp. 351-355.

Chanda, M.L. and Levitin, D.J. (2013). The neurochemistry of music. Trends Cog. Sci. 17, 179–193.

Chao, M., and Bursten, J. R. (2021). Girl talk: Understanding negative reactions to female vocal fry. Hypatia 36, 42-59.

Childers, D. G., and Lee, C. K. (1991). Vocal quality factors: Analysis, synthesis, and perception. J. Acoust. Soc. Am. 90, 2394-2410.

Cirillo, J. (2004). Communication by unvoiced speech: the role of whispering. Anais da Academia Brasileira de Ciências, 76, 413-423.

Classe, A. (1956). Phonetics of the silbo Gomero. Arch. Linguistic. 9, 44–61.

Conard, N. J., Malina, M., and Münzel, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany. Nature, 460, 737-740.

Cooksey, John M. (1997). Voice transformation in male adolescents. Bodymind and Voice: Foundations of Voice Education. John Cooksey, Leon Thurman and Graham Welch, eds, 718–738. Collegeville, MN: VoiceCare Network.

Cooper, N.P. (1998). Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea. J. Physiol. 509, 277-288.

Costall, A. (1986). Evolutionary gradualism and the study of development. Hum. Devel. 29, 4-11.

Cross, I. (2021). Music, attachment, and uncertainty: Music as communicative interaction. Behav. Brain Sci. 44, e66.

Cuaya, L. V., Hernández-Pérez, R., Boros, M., Deme, A., and Andics, A. (2022). Speech naturalness detection and language representation in the dog brain. NeuroImage 248, 118811.

Cummins, F. (2003). Practice and performance in speech produced synchronously. J. Phonetics, 31, 139-148.

Cummins, F. (2013). Joint speech: The missing link between speech and music? PERCEPTA-Revista de Cognição Musical 1, 17.

Dal Monte, O., Piva, M., Anderson, K.M., Tringides, M., Holmes, A.J. and Chang, S,W,C. (2017). Oxytocin under opioid antagonism leads to supralinear enhancement of social attention. Proc. Natl. Acad. Sci. (USA) 114, 5247-5252.

Danford, K., Roosevelt, L., Vroom, A., Harris, L., and Zielinski, R. (2022). Impolite Birth: Theatre Voice Training and the Experience of Childbirth. Voice and Speech Review, 1-15.

Dargie, D. (2018). A Report on the Dave Dargie Collection in the International Library of African Music. Accessed at https://tranquanghaisworldthroatsinging.com/category/dave-dargie/.

Darwin, C. 1871. The descent of man, and selection in relation to sex. London: John Murray.

Darwin, C. 1859. The origin of species. London: John Murray.

Dawkins, R. 2005. The ancestor's tale: a pilgrimage to the dawn of life. London: Phoenix.

Dawkins, R. (1976). The selfish gene. Oxford: Oxford University Press.

De Manzano, O., and Ullén, F. (2018). Same genes, different brains: neuroanatomical differences between monozygotic twins discordant or musical training. Cereb. Cortex 28, 387-394.

Deacon, T. W. (2003). Multilevel selection in a complex adaptive system: The problem of language origins. In B. H. Weber and D. J. Depew (Eds.), Evolution and Learning: The Baldwin Effect Reconsidered. Life and mind, pp. 81–106. Bradford Book; MIT Press.

Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. Proc. Natl. Acad. Sci. (USA) 107 (Supplement 2), 9000–9006.

Delius, J. A., & Müller, V. (2023). Interpersonal synchrony when singing in a choir. Front. Psychol. 13, 1087517.

Della Bella, S., Berkowska, M., and Sowiński, J. (2015). Moving to the beat and singing are linked in humans. Front. Hum. Neurosci. 9, 663.

Demany, L., and Armond, F. (1984). The perceptual reality of tone chroma in early infancy. J. Acoust. Soc. Am. 76, 57-66.

Demany, L., Monteiro, G., Semal, C., Shamma, S., and Carlyon, R. P. (2021). The perception of octave pitch affinity and harmonic fusion have a common origin. Hearing Res. 404, 108213.

Demoucron, M., and Weisser, S. (2012). Bowed strings and sympathy, from violins to indian sarangis. Acoustics. Gent, Belgium: Musée des Instruments de Musique.

Deutsch, D. (1999). The processing of pitch combinations. In The Psychology of Music, pp349-411. New York: Academic Press.

Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? Musicae scientiae 12 169-195.

Di Stefano, N., Vuust, P., and Brattico, E. (2022). Consonance and dissonance perception. A critical review of the historical sources, multidisciplinary findings, and main hypotheses. Phys. Life Rev. 43, 273-304.

Dolscheid, S., Hunnius, S., Casasanto, D., and Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. Psychol. Sci. 25, 1256–1261.

Doppler, C. (1842/1907). Über das farbige Licht der Doppelsterne. Leipzig: Ostwalds Klassiker.

Dor, D., and Jablonka, E. (2010). Plasticity and canalization in the evolution of linguistic communication: an evolutionary developmental approach. In R. K. Larson, V. Deprez, and H. Yamakido (Eds.), The Evolution of Human Language, pp. 135–147. Cambridge University Press.

Dunbar, R. I. M. (2012). On the evolutionary function of song and dance. In N. Bannan (Ed.), Music, Language, and Human Evolution, 201–214. Oxford: Oxford University Press.

Dunbar, R. I. M. (2020). Structure and function in human and primate social networks: Implications for diffusion, network stability and health. Proc. Roy. Soc. London 476A, 20200446.

Dunbar, R. I. M. (2022). Laughter and its role in the evolution of human social bonding. Phil. Trans. Roy. Soc. Series B 377, 20210176.

Dunbar, R. I. M. (2023). The origins and function of musical performance. Front. Psych. – Perf. Sci. 14, 1257390.

Dunbar, R.I.M., Kaskatis, K., MacDonald, I. and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. Evol. Psychol. 10, 688-702

Durkheim, E. (1912/1995). The Elementary Forms of Religious Life. Edited and translated by Karen E. Fields. New York: Free Press.

Elhilali, M. (2019). Modulation representations for speech and music. In Timbre: Acoustics, perception, and cognition, Edited by Kai Siedenburg, Charalampos Saitis, Stephen McAdams, Arthur N. Popper, Richard R. Fay. Cham, Switzerland: Springer, pp 335-359.

Elie, J. E., Hoffmann, S., Dunning, J. L., Coleman, M. J., Fortune, E. S., and Prather, J. F. (2020). From perception to action: the role of auditory input in shaping vocal communication and social behaviors in birds. Brain Behav. Evol. 94, 51-60.

Ellis, Alexander. 1885. On the Musical Scales of Various Nations. London: Journal of the Society of Arts. https://soundandscience.de/text/musical-scales-various-nations

Espi-Sanchis, P., and Bannan, N. (2012). Found objects in the musical practices of hunter-gatherers: implications for the evolution of instrumental music. In N. Bannan (ed), Music, Language, and Human Evolution, Pp.173-198. Oxford: Oxford University Press.

Ewell, P. (2021). Music theory's white racial frame. Music Theory Spectrum 43, 324-329.

Fant, G. (1960). Acoustic theory of speech production. The Hague: Mouton.

Ferreri, L., Mas-Herrero, E., Zatorre, R.J., Ripollés, P., Gomez-Andres, A., Alicart, H., Olivé, G. et al. (2019). Dopamine modulates the reward experiences elicited by music. Proc. Natl. Acad. Sci. (USA) 116, 3793-3798.

Fitch, W.T. (2002). Comparative vocal production and the evolution of speech: Reinterpreting the descent of the larynx. The Transition to Language. Alison Wray, ed. Pp. 21–45. Oxford: Oxford University Press.

Fletcher, N. (1996). The didjeridu (didgeridoo). Acoustics Australia 24, 11-16.

Foss, A. H., Altschuler, E. L., and James, K. H. (2007). Neural correlates of the Pythagorean ratio rules. Neuroreport 18, 1521-1525.

Foster, D.J., and Conn, P.J. (2017). Allosteric modulation of GPCRs: insights and potential utility for treatment of schizophrenia and other CNS disorders. Neuron 94, 431-446

Fouda, L., Wingfield, J. E., Fandel, A. D., Garrod, A., Hodge, K. B., Rice, A. N., and Bailey, H. (2018). Dolphins simplify their vocal calls in response to increased ambient noise. Biol. Lett. 14, 20180484.

Franklin, J. C. (2002). Musical Syncretism and the Greek Orientalizing Period. Archäologie früher Klangerzeugung und Tonordnungen, 441-451.

Freeman, W. (2000). A neurobiological role of music in social bonding. N Wallin, B. Merker, and S. Brown (ed.s). The Origins of Music, 411-424. Cambridge MA: MIT Press.

Friauf, E. (2014). Hearing. e-Neuroforum 5, 51-52.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., and Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. Curr. Biol. 19, 573–576.

Gackle, L. (1991). The adolescent female voice characteristics of change and stages of development. Choral Journal 31, 17–25.

Garcia, M. (1894). Hints on singing. London: Ascherberg, Hopwood and Crew.

Gardner, H. (1983). Frames of Mind: A Theory of Multiple Intelligences. New York: Basic Books.

Gaywood, H.R.A. (1996) Guqin and Guzheng: the historical and contemporary development of two Chinese musical instruments, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/4894/

Ghazanfar, A. A., Neuhoff, J. G., and Logothetis, N. K. (2002). Auditory looming perception in rhesus monkeys. Proc. Natl. Acad. Sci. (USA) 99, 15755-15757.

Giordano, N. (2015). Explaining the Railsback stretch in terms of the inharmonicity of piano tones and sensory dissonance. J. Acoust. Soc. Am., 138, 2359-2366.

Gingerich, O. (1992). Kepler, Galilei, and the Harmony of the World. Music and Science in the Age of Galileo, 45-63.

González-García, N., and Rendón, P.L. (2017). fMRI mapping of brain activity associated with the vocal production of consonant and dissonant intervals. JoVE 123, e55419.

Gould, S.J. (1977). Ontogeny and Phylogeny. Cambridge, MA: The Belknap Press.

Gould, S.J., and Eldredge, N. (1993). Punctuated equilibrium comes of age. Nature 366, 223-227.

Gracanin, A., Bylsma, L.M., and Vingerhoets, A.J.J.M. (2018). Why only humans shed emotional tears: Evolutionary and cultural perspectives. Human Nature, 29, 104–133.

Greenberg, J. H. (1966). Language Universals. De Gruyter Mouton.

Haiduk, F., and Fitch, W. T. (2022). Understanding design features of music and language: The choric/dialogic distinction. Front. Psychol. 13, 786899.

Hamdan, A. L., Mahfoud, L., Sibai, A., and Seoud, M. (2009). Effect of pregnancy on the speaking voice. J. Voice 23, 490-493.

Harvey, A.R. (2017). Music, Evolution, and the Harmony of Souls. Oxford University Press, Oxford.

Harvey, A.R. (2018). Music and the meeting of human minds. Front. Psychol. 9, 762

Harvey. A.R. (2020). Links between the neurobiology of oxytocin and human musicality. Front. Hum. Neurosci. 14, 350.

Haspelmath, M. (2020). Human linguisticality and the building blocks of languages. Front. Psychol. 10, 3056.

He, W., and Ren, T. (2021). The origin of mechanical harmonic distortion within the organ of Corti in living gerbil cochleae. Commun. Biol. 4, 1008.

Heffner, H. E., & Heffner, R. S. (2018, May). The evolution of mammalian hearing. AIP Conference Proceedings (Vol. 1965, No. 1). AIP Publishing.

Helmholtz, H. F. (1885). On the Sensations of Tone (Translated by A. J. Ellis. Reprinted in 1954). New York: Dover.

Hipkins, A. J. (1895). 'The standard for musical pitch'. Royal Soc. Arts J. 44, 335.

Hoehl, S., Fairhurst, M., and Schirmer, A. (2021). Interactional synchrony: signals, mechanisms and benefits. Soc. Cog. Affect. Neurosci. 16, 5-18.

Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., and ten Cate, C. (2015). Searching for the origins of musicality across species. Phil. Trans. Roy. Soc. B: Biol. Sci. 370, 20140094.

Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H., and Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. Animal Cognition 16, 599-609.

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. Phil. Trans. Roy. Soc. B: Biol. Sci. 370, 20140088.

Hu, Y., and Cardoso, G. C. (2010). Which birds adjust the frequency of vocalizations in urban noise? Animal Behaviour 79, 863-867.

Hu, J., Wang, Z., Feng, X., Long, C., and Schiller, D. (2019). Post-retrieval oxytocin facilitates next day extinction of threat memory in humans. Psychopharmacol. 236, 293–301.

Humboldt, W. von (1836/1999). Humboldt: 'On language': On the diversity of human language construction and its influence on the mental development of the human species. Cambridge: Cambridge University Press.

Hyde, K.L., Zatorre, R.J., Griffiths, T.D., Lerch, J.P., and Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. Brain 129, 2562-2570.

Ikhtisamov, K. S. (1988). On the problem of comparative investigation of two-part throat singing and instrumental music of Turkic and Mongol peoples. In Folk musical instruments and instrumental music (2). Y. Gippius, ed. Pp 197–216. Moscow: Sovetskii Kompozitor.

Jaatinen, J., Väntänen, V., Salmela, V., and Alho, K. (2023). Subjectively preferred octave size is resolved at the late stages of cerebral auditory processing. Eur. J. Neurosci. 58, 3686-3704

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., and McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. Curr. Biol. 29, 3229-3243.

Jablonka, E., and Lamb, M. J. (2005). Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life. MIT Press.

Jermyn, A. S., Stevenson, D. J., & Levitin, D. J. (2023). 1/f laws found in non-human music. Sci. Rep. 13, 1324.

Jesse, A., and Massaro, D. W. (2010). Seeing a singer helps comprehension of the song's lyrics. Psychon. Bull. Rev. 17, 323-328.

Johannesma, P., and Aertsen, A. (1982). Statistical and dimensional analysis of the neural representation of the acoustic biotope in frog. J. Mes. Syst 6, 399-421.

Jordania, J. (2022). Music, homosexuality, and cannibalism as predator avoidance evolutionary strategies in early humans. In J. Friedman (ed.), Music in Human Experience: Perspectives on a Musical Species, 139-59. Cambridge, UK: Cambridge Scholars Publishing.

Justus, T. C., and Bharucha, J. J. (2002). Music perception and cognition. Stevens' Handbook Exp. Psychol. 1, 453-492.

Kant, I. (1798/1996). Anthropology from a Pragmatic Point of View. Translated by Victor Lyle Dowdell. Southern Illinois University Press.

Karpati, F.J., Giacosa, C., Foster, N.E., Penhune, V.B. and Hyde, K.L. (2016). Sensorimotor integration is enhanced in dancers and musicians. Exp. Brain Res. 234, 893-903.

Keeler, J.R., Roth, E.A., Neuser, B.L., Spitsbergen, J.M., Waters, D.J., and Vianney, J.M. (2015). The neurochemistry and social flow of singing: bonding and oxytocin. Front. Hum. Neurosci. 9, 518.

Kershenbaum, A., Demartsev, V., Gammon, D. E., Geffen, E., Gustison, M. L., Ilany, A., and Lameira, A. R. (2021). Shannon entropy as a robust estimator of Zipf's Law in animal vocal communication repertoires. Meth. Ecol. Evol. 12, 553-564.

Kessen, W., Levine, J., and Wendrich, K. A. (1979). The imitation of pitch in infants. Infant Behav. Dev. 2, 93-99.

Killin, A. (2016). Rethinking music's status as adaptation versus technology: a niche construction perspective. Ethnomusicology Forum, 25, 1–24.

Killin, A. (2017). Plio-Pleistocene Foundations of Hominin Musicality: Coevolution of Cognition, Sociality, and Music. Biological Theory, 12, 222–235.

Kim, S., Fonagy, P., Koos, O, Dorsett, K., and Strathearn, L. (2014). Maternal oxytocin response predicts mother-to-infant gaze. Brain Res. 1580, 133-142.

Kob, M., Henrich, N., Herzel, H., Howard, D., Tokuda, I., and Wolfe, J. (2011). Analysing and understanding the singing voice: recent progress and open questions. Curr. Bioinform. 6, 362-374.

Koçak, Ö., and Atan, D.O.Ğ.A.N. (2022). Effect of delivery mode on objective and subjective aspects of voice. Eur. Rev. Med. Pharmacolog. Sci. 26, 485-490.

Koelsch, S. (2018). Investigating the neural encoding of emotion with music. Neuron, 98, 1075-1079.

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., and Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. Cognition 112, 1-20.

Kreutz, G. (2014). Does singing facilitate social bonding? Music and Med. 6, 51-60.

Krishnaswamy, A. (2003). On the twelve basic intervals in South Indian classical music. In Audio Engineering Society Convention 115. Audio Engineering Society.

Kumar, V., Croxson, P.L., and Simonyan, K. (2016). Structural organization of the laryngeal motor cortical network and its implication for evolution of speech production. J. Neurosci. 36, 4170-4418.

Lã, F. M. B., and Sundberg, J. (2012). Pregnancy and the singing voice: reports from a case study. J. Voice 26, 431-439.

Laeng, B., Kuyateh, S., and Kelkar, T. (2021). Substituting facial movements in singers changes the sounds of musical intervals. Sci. Rep. 11, 22442.

Lamarck, J.-B. (1815). Histoire naturelle des animaux sans vertèbres, présentant les caractères généraux et particuliers de ces animaux, leur distribution, leurs classes, leurs familles, leurs genres, et la citation des principales espèces qui s'y rapportent; précédée d'une introduction offrant la détermination des caractères essentiels de l'animal, sa distinction du végétal et des autres corps naturels; enfin, l'exposition des principes fondamentaux de la zoologie, vol. 1. Paris: Déterville.

Lanyon, S. J. (2006). A saltationist approach for the evolution of human cognition and language. In Cangelosi, A., Smith, A. D., and Smith, K. (Eds.). The evolution of language: Proceedings of the 6th International Conference (Evolang6), 176-183. Singapore: World Scientific.

Lavenda, B. H. (1985). Brownian motion. Scientific American 252, 70-85.

Lawson, G. (2024). Sound Tracks: Uncovering Our Musical Past. London: Penguin.

Le Bomin, S., Lecointre, G., and Heyer, E. (2016). The evolution of musical diversity: the key role of vertical transmission. PLoS One 11, e0151570.

Lecanuet, J. P., Graniere-Deferre, C., Jacquet, A. Y., and DeCasper, A. J. (2000). Fetal discrimination of low-pitched musical notes. Devel. Psychobiol. 36, 29-39.

Levitin, D. (2007). This Is Your Brain on Music: The Science of a Human Obsession. New York: Plume (Penguin).

Li, G. (2006). The effect of inharmonic and harmonic spectra in Javanese Gamelan tuning (1): a theory of the Sléndro. In: Proceeding AMTA'06 proceedings of the 7th WSEAS international conference on acoustics and music: theory and applications. World Scientific and Engineering Academy and Society, Stevens Point (pp. 65-71).

Leipold, S., Klein, C., and Jänke, L. (2021). Musical expertise shapes functional and structural brain networks independent of absolute pitch ability. J. Neurosci. 41, 2496-2511.

Lindestad, P. Å., Södersten, M., Merker, B., and Granqvist, S. (2001). Voice source characteristics in Mongolian "throat singing" studied with high-speed imaging technique, acoustic spectra, and inverse filtering. J. Voice 15, 78-85.

List, G. (1984). Concerning the Concept of the Universal and Music. The World of Music, 26(2), 40–49.

Lumsden, C. J., and Wilson, E. O. (1980). Translation of epigenetic rules of individual behavior into ethnographic patterns. Proc. Natl. Acad. Sci. (USA) 77, 4382–4386.

Mann, Z. F., and Kelley, M. W. (2011). Development of tonotopy in the auditory periphery. Hearing Res. 276, 2-15.

McAdams, S. (2019). The perceptual representation of timbre. In K. Siedenburg et al. (eds.), Timbre: Acoustics, perception, and cognition, 23-57. Switzerland: Springer.

McDermott, J. H., Keebler, M. V., Micheyl, C., and Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. J. Acoust. Soc. Am. 128, 1943-1951.

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. Nature 264, 746-748.

McHenry, M. A., Evans, J., and Powitzky, E. (2016). Effects of bel canto training on acoustic and aerodynamic characteristics of the singing voice. J. Voice 30, 198-204.

McKay, S., and Roberts, J. (1990). Obstetrics by ear: Maternal and caregiver perceptions of the meaning of maternal sounds during second stage labor. J. Nurse-Midwif. 35, 266-273.

McNeill, W. H. (1995). Keeping together in time. Dance and drill in human history. Cambridge, Mass.: Harvard University Press.

McPherson, T., Berger, D., Alagapan, S., and Frölich, F. (2018). Intrinsic rhythmicity predicts synchronization-continuation entrainment performance. Sci. Rep. 8, 11782.

McPherson, M.J., Dolan, S.E., Durango, A., Ossandon, T., Valdés, J., Undurraga, E.A., et al. (2020). Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. Nature Comm. 11: 2786.

McRoberts, G. W., and Best, C. T. (1997). Accommodation in mean f0 during mother–infant and father–infant vocal interactions: A longitudinal case study. J. Child Lang. 24, 719-736.

Mach, Ernst. 1886/1914. The Analysis of Sensations and the Relation of the Physical and the Psychical, trans. C. M. Williams, rev. and suppl. Sydney Waterlow. La Salle, IL: Open Court.

Maleki, A., and Youseflu, S. (2023). The effect of music-based interventions on short-term postpartum episiotomy pain: A systematic review and meta-analysis. Heliyon 9, e14785

McPherson, M.J., Grace, R.C., and McDermott, J.H. (2022). Harmonicity aids hearing in noise. Atten. Percept. Psychophys. 84, 1016–1042.

Mallik, A., Chanda, M.L., and Levitin, D.J. (2017). Anhedonia to music and mu-opioids: evidence from the administration of naltrexone. Sci. Rep. 7, 41952.

Manén, L. (1974). The Art of Singing: a manual of bel canto. London: Faber Music.

Marjieh, R., Harrison, P. M., Lee, H., Deligiannaki, F., & Jacoby, N. (2024). Timbral effects on consonance disentangle psychoacoustic mechanisms and suggest perceptual origins for musical scales. Nat. Comm. 15, 1482.

Matsui, T. (2023). Sound Processing in the Auditory Periphery: Toward Speech Communication and Music Comprehension. In Acoustic Communication in Animals: From Insect Wingbeats to Human Music (Bioacoustics Series Vol. 1), pp. 219-231. Singapore: Springer Nature Singapore.

Melzack, R. (1984). The myth of painless childbirth (the John J. Bonica lecture). Pain, 19, 321-337.

Meguro, Y., Miyano, K., Hirayama, S., Yoshida, Y., Ishibashi, N., Ogino, T., Fujii, Y, et al. (2018). Neuropeptide oxytocin enhances μ opioid receptor signaling as a positive allosteric modulator. J. Pharmacol. Sci. 137, 67-75.

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. Behav. Brain Sci. 44, e60.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., et al. (2019). Universality and diversity in human song. Science, 366, eaax0868.

Mendeleev, D. (1869). Osnovy khimii [Principles of Chemistry]. St. Petersburg.

Menon, V., and Levitin, D.J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. Neuroimage 28, 175–184.

Merker, B. (1999). Synchronous chorusing and the origins of music. Musicae Scientiae, 3(1_suppl), 59-73.

Merker, B. (2012). The vocal learning constellation: Imitation, ritual culture, encephalization. In Music, Language and Human Evolution, Nicholas Bannan, ed. Oxford: Oxford university Press, Pp. 215-60.

Meyer, J. (2021). Environmental and linguistic typology of whistled languages. Ann. Rev. Linguistics 7, 493-510.

Michael, D. D., and Gilman, M. (2021). The myth of intonation as an objective measure of singing quality. Australian Voice 22, 49-62.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. Hearing Res. 219, 36–47.

Milne, A., Smit, E., Sarvasy, H., and Dean, R. (2022). Evidence for a universal mechanism associating auditory roughness and musical dissonance. Sydney: MARCS Institute for Brain, Behaviour and Development, Western Sydney University. psyarxiv.com.

Mithen, Steven. (2005). The Singing Neanderthals: The Origins of Music, Language, Mind and Body London: Weidenfeld and Nicolson.

Moerel, M., De Martino, F., Santoro, R., Uqurbil, K., Goebel, R., et al. (2013). Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. J. Neurosci. 33, 11888-11898.

Moerel, M., De Martino, F., Santoro, R., Yacoub, E., and Formisano, E. (2015). Representation of pitch chroma by multi-peak spectral tuning in human auditory cortex. Neuroimage 106, 161-169.

Moreau, P., Jolicœur, P., and Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. Brain Cog. 81, 337-344.

Morgan, C. L. (1896). Habitat and Instinct. Edward Arnold Press.

Morley, I. (2013). The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality. Oxford: Oxford University Press.

Müller, V., Ohström, K. R. P., and Lindenberger, U. (2021). Interactive brains, social minds: neural and physiological mechanisms of interpersonal action coordination. Neurosci. Biobehav. Rev. 128, 661-677.

Nattiez, J.J. (1999). Inuit throat-games and Siberian throat singing: A comparative, historical, and semiological approach. Ethnomusicology 43, 399-418.

Nazzi, T., and Cutler, A. (2019). How Consonants and Vowels Shape Spoken-Language Recognition. Annual Review of Linguistics, 5(1), 25–47. https://doi.org/10.1146/annurev-linguistics-011718-011919

Neuhoff, J. G., Knight, R., and Wayand, J. (2002). Pitch change, sonification, and musical expertise: Which way is up? Proceedings of the 2002 International Conference on Auditory Display. Kyoto, Japan.

Niemitz, C. (2010). The evolution of the upright posture and gait—a review and a new synthesis. Naturwissenschaften 97, 241-263.

Nikolsky, A. (2020). The pastoral origin of semiotically functional tonal organization of music. Front. Psychol. 11, 1358.

Nikolsky, A., Alekseyev, E., Alekseev, I., and Dyakonova, V. (2020). The overlooked tradition of “personal music” and its place in the evolution of music. Front. Psychol. 10, 3051.

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: Prehistoric. Front. Psychol. 6, 1405.

Norman-Haignere, S., Kanwisher, N.G., and McDermott, J.H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. Neuron 88: 1281-1296.

Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., et al. (2019). Intracranial recordings from human auditory cortex reveal a neural population selective for song. bioRxiv, 696161.

Osborn, H. F. (1896). A mode of evolution requiring neither natural selection nor the inheritance of acquired characters. Trans. NY Acad. Sci. 15, 141–142.

Ouzman, S. (2001). Seeing is deceiving: rock art and the non-visual. World Archaeology 33, 237-256.

Oxenham, A. J. (2013). Revisiting place and temporal theories of pitch. Acoust. Sci. Technol. 34, 388–396.

Ozaki, Y., Tierney, A., Pfordresher, P., Mcbride, J., Benetos, E., Proutskova, P., et al. (2022). Similarities and differences in a global sample of song and speech recordings [Stage 1 Registered Report]. https://doi.org/10.31234/osf.io/jr9x7

Paletta, P., Bass, N., Kavaliers, M., and Choleris, E. (2022). The role of oxytocin in shaping complex social behaviours: possible interactions with other neuromodulators. Phil. Trans. Roy. Soc. Series B 377, 20210058

Palmer, A. R., and Russell, I. J. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. Hearing Res. 24, 1-15.

Pamjav, H., Juhász, Z., Zalán, A., Németh, E., and Damdin, B. (2012). A comparative phylogenetic study of genetics and folk music. Mol. Genet. Genomics 287, 337-349.

Patel, A. D. (2023). Human Musicality and Gene-Culture Coevolution: Ten Concepts to Guide Productive Exploration. In E. H. Margulis, P. Loui, & D. Loughridge (Eds.), The Science-Music Borderlands: Reckoning with the Past and Imagining the Future, pp. 15–37. Boston. MA: MIT Press.

Patel, A. D. (2018). Music as a transformative technology of the mind: An update. In H. Honing (ed.), The origins of musicality, 113-126. Cambridge, Mass.: MIT Press.

Patel, A.D. (2008). Music, Language, and the Brain. Oxford: Oxford University Press.

Parncutt, R. (2019). Mother schema, obstetric dilemma, and the origin of behavioral modernity. Behav. Sci. 9, 142.

Parncutt, R., & Hair, G. (2018). A psychocultural theory of musical interval: Bye bye Pythagoras. Music Perception, 35, 475-50.

Pearce, E., Wlodarski, R., Machin, A., and Dunbar, R.I.M. (2017). Variation in the β-endorphin, oxytocin, and dopamine receptor genes is associated with different dimensions of human sociality. Proc. Natl. Acad. Sci. (USA) 114, 5300-5305.

Pell, M. D. (2006). Cerebral mechanisms for understanding emotional prosody in speech. Brain and Language 96, 221-234.

Pereira, C.S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S.L., and Brattico, E. (2011). Music and emotions in the brain: familiarity matters. PLoS One, 6, e27241

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

Peretz, I. and Zatorre, R.J. (2005). Brain organization for music processing. Ann. Rev. Psychol. 6, 89-114.

Perrachione, T.K., Fedorenko, E.G., Vinke, L., Gibson, E., and Dilley, L.C. (2013). Evidence for shared cognitive processing of pitch in music and language. PLoS One 8, e73372.

Piggott, F. T. (1891). The music and musical instruments of Japan. London: Batsford.
Pinker, S. (1997). How the mind works. New York: Norton.
Pinker, S. (1994). The language instinct: How the mind creates language. New York: Harper Collins.
Podlipniak, P. (2022a). Pitch syntax as an evolutionary prelingual innovation. Musicae Scientiae, 26, 280-302.
Podlipniak, P. (2022b). Pitch syntax as part of an ancient protolanguage. Lingua, 271, 103238a.
Podlipniak, P. (2021). The role of canalization and plasticity in the evolution of musical creativity. Front. Neurosci, 15, 267.
Podlipniak, P. (2017). The role of the Baldwin effect in the evolution of human musicality. Front. Neurosci. 11, 542.
Podlipniak, P. (2016). The evolutionary origin of pitch centre recognition. Psychology of Music, 44, 527–543.
Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). Evolution, 51, 537-551.
Popham, S., Boebinger, D., Ellis, D. P., Kawahara, H., and McDermott, J. H. (2018). Inharmonic speech reveals the role of harmonicity in the cocktail party problem. Nature Comm. 9, 2122.
Provine, R. R. (2017). Laughter as an approach to vocal evolution: The bipedal theory. Psychon. Bull. Rev. 24, 238-244.
Provine, R. R., Krosnowski, K. A., and Brocato, N. W. (2009). Tearing: Breakthrough in human emotional signaling. Evol. Psychol.. 7, 52–56.
Punam, P.T., and Chang, S.W.C. (2022) Interplay between the oxytocin and opioid systems in regulating social behaviour. Phil. Trans. Roy. Soc. Series B 377, 20210050.
Puts, D. A., Gaulin, S. J., and Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. Evol. Human Behav. 27, 283-296.
Puts, D. A., Hodges, C. R., Cárdenas, R. A., and Gaulin, S. J. (2007). Men's voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. Evol. Human Behav. 28, 340-344.
Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., et al. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. Proc. Roy. Soc. B: Biol. Sci. 283, 20152830.
Raimondi, T., Di Panfilo, G., Pasquali, M., Zarantonello, M., Favaro, L., Savini, T., et al. (2023). Isochrony and rhythmic interaction in ape duetting. Proc. Roy. Soc. Series B 290, 20222244.
Ravignani, A., and De Boer, B. (2021). Joint origins of speech and music: testing evolutionary hypotheses on modern humans. Semiotica 2021, 169-176.
Ridley, M. (1996). The Origins of Virtue: Human Instincts and the Evolution of Cooperation. New York: Viking.
Riera, P. E., Proscia, M., and Eguia, M. C. (2014). A comparative study of saxophone multiphonics: musical, psychophysical and spectral analysis. J. New Music Res. 43, 202-213.
Robledo, J.P., Hurtado, E., Prado, F., Román, D., and Cornejo, C. (2016). Music intervals in speech: Psychological disposition modulates ratio precision among interlocutors' nonlocal f0 production in real-time dyadic conversation. Psych. Mus. 44, 1404–1418.
Roederer, J.G. (1984). The search for a survival value of music. Music Perception 1, 350-356.
Ruben, R.J. (2020). The developing concept of tonotopic organization of the inner ear. J. Assoc. Res. Otolaryngol. 21, 1-20.
Ryabov, V.A. (2019). Some aspects of reflection of dolphin FM signals (whistles) in an experimental tank. Acoust. Phys. 65, 771-777.
Shahal, S., Wurzberg, A., Sibony, I., Duadi, H., Shniderman, E., et al. (2020). Synchronization of complex human networks. Nature Comm. 11, 3854.
Salighedar, R., Erfanparast, A., Tamaddonfard, E.and Soltanalinejad, F. (2019). Medial prefrontal cortex oxytocin-opioid receptors interaction in spatial memory processing in rats. Physiol. Behav. 209, 112599.
Salimpoor, V.N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R.J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion in music. Nat. Neurosci. 14, 257–264.
Savage, P.E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., and Fitch, W.T. (2021). Music as a coevolved system for social bonding. Behav. Brain Sci. 44, e59.
Scherer, K.R. (1992). Vocal affect expression as symptom, symbol, and appeal. In Papousek, H., Jürgens, U., and Papoušek, M. (Eds.). (1992). Nonverbal vocal communication: Comparative and developmental approaches, 43-60. Cambridge University Press, Cambridge.
Schirmer, A., Meck, W. H., & Penney, T. B. (2016). The socio-temporal brain: Connecting people in time. Trends Cog. Sci. 20, 760-772.
Schneider, A., and Leman, M. (2002). Sonological and psychoacoustic characteristics of carillon bells. In Proceedings of the 16th Meeting of the FWO Research Society on Foundations of Music Research, Ghent University: IPEM, 2002.
Schubert, E. (2016). Enjoying sad music: Paradox or parallel processes? Front. Hum. Neuroscie. 10, 312.
Schulz, J., Bahrami-Rad, D., Beauchamp, J., and Henrich, J. (2018). The origins of WEIRD psychology. Available at SSRN 3201031.
Scott, J.T. (1998) The harmony between Rousseau's musical theory and his philosophy. J. Hist. Ideas 59, 287-308.
Seki, Y. (2021). Cockatiels sing human music in synchrony with a playback of the melody. PLoS One 16, e0256613.
Semal, C., and Demany, L. (2006). Individual differences in the sensitivity to pitch direction. J. Acoust. Soc. Am. 120, 3907-3915.
Sergeant, D. (1983). The octave-percept or concept. Psychology of Music, 11(1), 3-18.
Sergeant, D.C., Sjölander, P.J., and Welch, G.F. (2005). Listeners' identification of gender differences in children's singing. Res. Stud. Mus. Educ. 24, 28–39.
Sergeant, D.C., and Welch, G.F. (2009). Gender differences in long-term average spectra of children's singing voices. J. Voice 23, 319–336.
Seyfarth, R. M., and Cheney, D. L. (1980). The ontogeny of vervet monkey alarm calling behavior: a preliminary report. Zeitschrift für Tierpsychologie 54, 37-56.
Shannon R. V. (2005). Speech and music have different requirements for spectral resolution. Int. Rev. Neurobiol. 70, 121-134.
Shepard, R. N. (1964). Circularity in judgments of relative pitch. J. Acoust. Soc. Amer. 36, 2346–2353
Shera, C.A., and Charaziak, K.K. (2019) Cochlear frequency tuning and otoacoustic emissions. Cold Spring Harb Perspect Med 9, a033498.
Shilton, D. (2022). Sweet participation: The evolution of music as an interactive technology. Music & Science, 5, 20592043221084710.
Shofner, W.P. (2023). Cochlear tuning and the peripheral representation of harmonic sounds in mammals. J. Comp. Physiol. A 209, 145-161.
Siegel, G.M., Cooper, M., Morgan, J.L., and Brenneise-Sarshad, R. (1990). Imitation of intonation by infants. J. Speech Lang. Hear. Res. 33, 9-15.
Simalvi, S., Kaygusuz, I., Gumus, I., Usluogullari, B., Yilirim, M., and Kafali, H. (2014). Effect of music therapy during vaginal delivery on postpartum pain relief and mental health. J. Affect. Disord. 156, 194-199.
Smolders, J.W., Aertsen, A.M., and Johannesma, P.I. (1979) Neural representation of the acoustic biotope. A comparison of the response of auditory neurons to tonal and natural stimuli in the cat. Biol. Cybern 35, 11-20.
Spence, M. J., and Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. Infant Behav. Dev. 19, 199-212.
Stefanics, G., Háden, G.P., Sziller, I., Balázs, L., Beke, A., and Winkler, I. (2009). Newborn infants process pitch intervals. Clin. Neurophys. 120, 304-308.

Swee, G., and Schirmer, A. (2015). On the importance of being vocal: Saying “ow” improves pain tolerance. J. Pain 16, 326-334.
Snowdon, C.T. (2021). Animal signals, music and emotional well-being. Animals 11, 2670.
Snowdon, C.T., Teie, D., and Savage, M. (2015). Cats prefer species-appropriate music. Appl. Anim. Behav. Sci. 166, 106-111.
Snowdon, C.T., and Teie, D. (2010). Affective responses in tamarins elicited by species-specific music. Biol. Lett. 6, 30-32.
Sørensen, P.M., Haddock, A., Guarino, E., Jaakkola, K., McMullen, C., et al. (2023). Anthropogenic noise impairs cooperation in bottlenose dolphins. Curr. Biol. 33, 749-754.
Spencer, H. (1890). The origin of music. Mind 15, 449-468.
Steels, L. (2011). Modeling the cultural evolution of language. Physi. Life Rev. 8, 339-356.
Stewart, I. (1998). Life's other secret: The new mathematics of the living world. New York, NY, USA: John Wiley.
Stock, J P. (2007). Alexander J. Ellis and his place in the history of ethnomusicology. Ethnomusicology 51, 306-325.
Storr, A. (1992). Music and the mind. London: HarperCollins.
Strumolo, W. (2007). Relating Spectrum and Tuning of the Classical Thai Renat Ek. New York University. http://www. lifesci. sussex. ac.
Sumner, C.J., Wells, T.T., Bergevin, C., and Shera, C.A. (2018) Mammalian behavior and physiology converge to confirm shaper cochlear tuning in humans. Proc. Natl. Acad. Sci. (USA) 115, 11322-11326.
Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., and Västfjäll, D. (2010). Embodied auditory perception: the emotional impact of approaching and receding sound sources. Emotion 10, 216.
Tarr, B., Launay, J., Cohen, E., and Dunbar, R.I. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. Biol. Lett. 11, 20150767.
Tarr, B., Launay, J., Benson, C., and Dunbar, R.I.M. (2017). Naltrexone blocks endorphins released when dancing in synchrony. Adapt. Hum. Behav. Physiol. 3, 241-254.
Tartter, V. C. (1989). What’s in a whisper? J. Acoust. Soc. Am. 86, 1678-1683.
ten Cate, C., and Honing, H. (2022). Precursors of music and language in animals. PsyArXiv. doi:10.31234/osf.io/4zxtr.
Ternström, S., and Sundberg, J. (1989). Formant frequencies of choir singers. J. Acoust. Soc. Am. 86, 517-522.
Till, R. (2014). Sound archaeology: terminology, Palaeolithic cave art and the soundscape. World Archaeol. 46, 292-304.
Tomlinson, G. (2015). A million years of music: the emergence of human modernity. Boston, MA: MIT Press.
Tonndorf, J. (1958). Localization of aural harmonics along the basilar membrane of guinea pigs. J. Acoust. Soc. Am. 30, 938-943.
Trainor, L. J., and Desjardins, R.N. (2002). Pitch characteristics of infant-directed speech affect infants’ ability to discriminate vowels. Psychonom. Bull. Rev. 9, 335-340.
Trân Quang, H. (2002). Polyphony in one Throat. In Proceedings of the First International Symposium on Traditional Polyphony. Rusudan Tsurtsumia and Joseph Jordania, eds. Pp. 274-287. Tbilisi: International Research Center for Traditional Polyphony of Tbilisi State Conservatoire.
Trân Quang, H., & Bannan, N. (2012). Vocal traditions of the world: towards an evolutionary account of voice production in music. In Music, Language and Human Evolution (pp. 142-172). Oxford University Press.
Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. J. Acoust. Soc. Am. 107, 3438-3451.
Trehub, S. E. (2003). The developmental origins of musicality. Nat. Neurosci. 6, 669-673.
Trine, A., and Monson, B. B. (2020). Extended High Frequencies Provide Both Spectral and Temporal Information to Improve Speech-in-Speech Recognition. Trends in Hearing, 24, 2331216520980299.
Unteregger, F., Wagner, P., Honeggar, F., Potthast, S., Zwicky, S., and Storck, C. (2020). Changes in vocal fold morphology during singing over two octaves. J. Voice 34, 165-169.
Valentova, J.V., Tureček, P., Varella, M.A.C., Šebesta, P., Mendes, F.D.C., Pereira, K.J., et al. (2019). Vocal parameters of speech and singing covary and are related to vocal attractiveness, body measures, and sociosexuality: a cross-cultural study. Front. Psych. 10, 2029.
Van Heuven, V. J. (1985). Some acoustic characteristics and perceptual consequences of foreign accent in Dutch spoken by Turkish immigrant workers. In: Dutch Linguistics at Berkeley, papers presented at the Dutch Linguistics Colloquium held at the University of California, Berkeley (pp. 67-84).
Vaughan, C.L. (2003). Theories of bipedal walking: an odyssey. J. Biomech. 36, 513-523.
Vetter, R. (1981). Flexibility in the Performance Practice of Central Javanese Music. Ethnomusicology 25, 199-214.
Vogel, H.U. (1994). Aspects of metrosophy and metrology during the Han period. Extrême-Orient Extrême-Occident, 135-152.
von Falkenhausen, L. (1992). On the early development of Chinese musical theory: the rise of pitch-standards. J. Am. Orient. Soc.112, 433-439.
Wagner, B., Mann, D.C., Afroozeh, S., Staubmann, G., and Hoeschele, M. (2019). Octave equivalence perception is not linked to vocal mimicry: budgerigars fail standardized operant tests for octave equivalence. Behaviour 156, 479-504.
Walker, K.M., Gonzalez, R., Kang, J.Z., McDermott, J.H., and King, A. J. (2019). Across-species differences in pitch perception are consistent with differences in cochlear filtering. Elife 8, e41626.
Wallin, N.L. (1991). Biomusicology: Neurophysiological, neuropsychological, and evolutionary perspectives on the origins and purposes of music. New York: Pendragon Press.
Waltham, C. (2010). Harp. In Thomas D. Rosser (ed), The science of string instruments, 145-166. New York: Springer.
Wang, X. (2013). The harmonic organization of auditory cortex. Front. Syst. Neurosci. 7, 114.
Warner, L. R., & Shields, S. A. (2007). The Perception of Crying in Women and Men: Angry Tears, Sad Tears, and the "Right Way" to Cry. In U. Hess and P. Philippot (Eds.), Group dynamics and emotional expression, pp. 92–117. Cambridge, UK: Cambridge University Press.
Watson, A., and Keating, D. (1999). Architecture and sound: an acoustic analysis of megalithic monuments in prehistoric Britain. Antiquity 73, 325-336.
Weinstein, D., Launay, J., Pearce, E., Dunbar, R.I., and Stewart L. (2016). Group music performance causes elevated pain thresholds and social bonding in small and large groups of singers. Evol. Hum. Behav. 37, 152-158.
Weisenberg, M., and Caspi, Z. (1989). Cultural and educational influences on pain of childbirth. J. Pain Sympt. Manag. 4, 13-19.
Wei, Y., Gan, L., and Huang, X. (2022). A Review of Research on the Neurocognition for Timbre Perception. Front. Psychol. 13, 869475.
Weisser, S., and Demoucron, M. (2012). Shaping the resonance: sympathetic strings in hindustani classical instruments. Proc. Meet. Acoust. 163ASA, 15, p. 035006. Acoustical Society of America.
West, B., Dunbar, R.I.M., Culbreth, G., and Grigolini, P. (2023). Fractal structure of human and primate social networks optimizes information flow. Proc. Roy. Soc. Lond. 479A, 20230028.
Winkler, I., Háden, G.P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat of music. Proc. Natl. Acad. Sci. (USA) 106, 2468-2471.
Wood, J. (2022). Tapping the Source. London: Composers Edition.

Woodward, S.C. (2019). Fetal, neonatal, and early infant experiences of maternal singing. In Welch, G. F., Howard, D. M., and Nix, J. (Eds.), The Oxford handbook of singing, 431-54. Oxford, UK: Oxford University Press.
Worschech, F., James, C.E., Jünemann, K., Sinke, C., Krüger, T.H.C., et al. (2023). Fine motor control improves in older adults after 1 year of piano lessons: Analysis of individual development and its coupling with cognition and brain structure. Eur. J. Neurosci. 57, 2040-2061.
Wren, P.C. (1924). Beau Geste. London: John Murray.
Wright, A.A., Rivera, J.J., Hulse, S.H., Shyan, M., and Neiworth, J.J. (2000). Music perception and octave generalization in rhesus monkeys. J. Exp. Psychol. Gen. 129, 291-307.
Yang, L-N., Chen, K., Yin, X-P., Liu, D. and Zhu, L-Q. (2022). The comprehensive neural mechanisms of oxytocin and analgesia. Curr. Neuropharmacol. 20, 147-157. Doi: 10.2174/1570159X19666210826142107.
Yin, X., and Müller, R. (2019). Fast-moving bat ears create informative Doppler shifts. Proc. Natl. Acad. Sci. (USA) 116, 12270-12274.
Yuhi, T., Kyuta, H., Mori, H.A., Murakami, C., Furuhara, K., Okuno, M., Takahashi, M., Fuji, D. and Higashida, H. (2017). Salivary oxytocin concentration changes during a group drumming intervention for maltreated school children. Brain Sci. 7, pii: E152
Yurdum, L., Singh, M., Glowacki, L., Vardy, T., Atkinson, Q. D., Hilton, C. B., Sauter, D., Krasnow, M. M. and Mehr, S. A. (2023). Universal interpretations of vocal music. Proc. Natl. Acad. Sci. 120, e2218593120.
Zarate, J.M. (2013). The neural control of singing. Front. Hum. Neurosci. 7, 237.
Zeloni, G., and Pavani, F. (2022). Minor second intervals: A shared signature for infant cries and sadness in music. i-Perception 13, 20416695221092471.
Zemp, H. (1989). Filming Voice Technique: The Making of "The Song of Harmonics". The World of Music 31, 56–85.