The Biology and Evolution of Rhythm:
Unravelling a Paradox

W Tecumseh Fitch
University of St Andrews

To appear in Rebuschat, et al (Eds): "Language and Music as Cognitive Systems" Oxford University Press.
version 1.3: Final
8858 words

Introduction: Why can't dogs dance?

Periodicity is a ubiquitous feature of all living things, and coupled biological oscillators entrain to each other readily (Glass and Mackey, 1988). Despite this, humans are rare if not unique in our ability to entrain our musical motor output to that of others during singing, dancing and playing in ensembles. This presents something of a paradox concerning human rhythmic entrainment and all that goes with it: why should a phenomenon seemingly so basic be (apparently) so rare in nature? The paradox, put simply, is this: if periodicity and entrainment are ubiquitous features of all living organisms, why can't dogs dance? In this paper I will explore this paradox from multiple comparative viewpoints, exploring similarities and differences between humans and other animals, between different aspects of music (harmony and rhythm), and between music and spoken language. Although my approach will entail asking many more questions than we can, at present, answer, my overall goal is to help focus interest on a number of phenomena that can, at least in principle, yield to empirical investigation of several sorts. Because of the wide range of disciplines involved, my style will be synoptic rather than prolix, and the references are intended as entry points into the literature rather than being exhaustive.

I will suggest that the "paradox of rhythm" can be resolved by recognizing that human rhythmic behaviour comprises several different components, each with their own biological basis and evolutionary history. Thinking of rhythm as a unitary, monolithic cognitive entity would be misleading and unproductive, while a "divide-and-conquer" strategy provides a fruitful way forward (Fitch, 2006b, Patel, 2006). I suggest at least three separable components underlying the human capacity for rhythmic behaviour. These include periodic motor pattern generation itself (an ancient and ubiquitous phenomenon), pulse (or "beat") extraction from complex patterns (a form of perceptual cognition that is shared with speech, at least), and entrainment of one's own motor output to this inferred "beat" (which may be the most biologically unusual feature of human rhythmic behaviour). After reviewing the comparative evidence suggesting that these features are separable, and that the pattern perception component is shared between music and speech, I suggest that "rhythm" incorporates both basic, primitive biological elements and highly unusual (apomorphic) elements, probably recently evolved. Isochronicity, I suggest, is not a fundamental component of musical rhythm. Rather, it is a probable (but not inevitable) interaction of the pattern extraction and entrainment components of the "rhythm faculty", driven by a desire for group coherence and predictability. From the viewpoint advocated here, the mosaic pattern of shared and unique features characterizing rhythm, like that of speech, music and language more generally, is consistent with Darwin's model of language evolution via a "musical protolanguage": an intermediate communication system that was more like music than language.

Defining "Rhythm"

The Greek word "rhythmos" derives from *rhein* "to flow", and the English word "rhythm" is used so frequently, and in so many ways, as to sometimes actually impede communication by causing misunderstanding (Sachs, 1952). In lay terms, we talk about the rhythm of the seasons or the rhythm of our heartbeat, which are simply examples of periodic behaviour. Thus, repetition at some relatively steady interval - quasi-periodicity - is one of the most basic connotations of the word "rhythm" as normally used. However, for musicologists, rhythm means more than mere periodicity, and is intertwined with notions of a "beat" (quasiperiodic tactus), metre and stress, when a pattern of recurring events is structured into hierarchical chunks (phrases, or measures) that follow certain rules of number, subdivision and accent. Not only are such structured patterns produced intentionally by a performer, but the extraction of such patterns from a "raw" stimulus by the listener is a crucial part of musical rhythm perception, which allows a listener to tap their foot, dance to the beat, or play along with the pattern. Thus, the connotations of "rhythm" in a musical context are far more complex and specific than simple periodicity. Finally, for linguists (specifically metrical phonologists) and poets "speech rhythm" has connotations that only partially overlap with those of musicologists. In particular, the periodicity of speech is limited (one does not "play along with," or dance to, speech rhythms), but the hierarchical pattern-processing aspects of speech and musical rhythm appear to be largely shared,.

Based on these and other considerations I will not try to define rhythm *per se* but rather to specify three separable components typical of musical rhythmic behaviour, and then explore the degree to which these components are just different sides of the same thing (as suggested by Sachs, 1952), or whether they are better thought of as separate components that rely on distinct cognitive/neural mechanisms. The musical context I have in mind is a ubiquitous one in human behaviour, typifying virtually any musical happening. Such situations, in the Western concert hall, include tapping one's foot or swaying back and forth to the beat. In Indian classical music venues, concertgoers often "dance" along with the performers using fluid hand motions. Far more typical of the world's music is dancing, which involves extracting the beat from a patterned auditory stimulus, and then moving one's entire body to this inferred beat (alone, or in concert with one or more others). Finally, any sort of ensemble playing involves the processes engaged in such examples, but with the additional feature that the motor output produced is itself musically structured and produces audible output, allowing *mutual* entrainment between the members of the ensemble. I will use this last example - ensemble playing - as my example context. Furthermore, to abstract away from many complex questions of pitch extraction, playing in tune, and generating harmony, the ensemble I have in mind is purely percussive: all the instruments (bells, shakers, drums, sticks, hand claps) either lack defined pitch, or produce a fixed, unchangeable pitch. Such an ensemble can produce extremely complex musical patterns, presenting considerable cognitive challenges both for listeners and performers, but "playing out of tune" or "out of key" is impossible.

To be concrete, let us use the example of a West African drumming ensemble playing a complex polyrhythmic pattern, such as *agbekor* (a popular traditional rhythm in Ghana and neighbouring lands (Agawu, 1987, Zabana, 1997)). Similar social and cognitive principles typify a huge range of musical performance styles throughout Africa and the New World (von Hornbostel, 1928, Jones, 1954, Merriam, 1959, Locke, 1987, Pressing, 2002).

Three components of rhythmic behaviour

I hypothesize three components as necessary for ensemble playing and dancing: pulse extraction, beat entrainment, and motor pattern generation. The first involves auditory processing only, the second cross-modal (auditory to motor) integration, and the last sensorily-guided motor behaviour. First, I illustrate these for clarity, in the concrete context of an agbekor drumming ensemble. Then I will turn to an exploration of the biological and evolutionary roots of these hypothesized mechanisms.

In agbekor, a central role is played by a bell pattern, played on a "gakogui" double bell, shown in Fig 1a (Zabana, 1997, Toussaint, 2003). Partly because bells are loud, and also because the patterns they play tend to be highly repetitive, with limited scope for improvisation, they provide the basic rhythmic pattern to which the other instruments and players respond and entrain. The bell player(s) can also, of course, entrain to other instruments, and in many cases a leader (e.g. the master drummer, or in other contexts the lead singers or master dancer) might provide the pulse. Nonetheless, we can simplify the situation in a natural way by assuming that the bell player "leads" and the other instruments follow. Thus, for example, the master drummer may set the tempo and start the ensemble off by drumming the bell pattern.

Crucially, the bell pattern does not directly provide a "beat" (a colloquial term I will use here to evoke both the Western musicological term "tactus," and the "pulse" of ethnomusicologists (Arom, 1991)). This bell pattern is syncopated, and related in a quite complex fashion to the underlying pulse(s) inferred by other participants. The beat can be seen most directly by the footfalls of dancers responding to the music (and often of the players as well), but *none* of the instruments in the ensemble play a simple pulse on their instrument. Thus, there is nothing like the steady "four on the floor" drumbeat pattern so typical of much Western dance music. Instead, every participant in such a rhythmic ensemble must *infer* a beat that *is not actually directly present* in the acoustic signal. Thus, the first and most basic cognitive activity is one of **pulse extraction**: the inference of a pulse or "beat", given a patterned and repetitive acoustic stimulus (in European music, following (Lerdahl and Jackendoff, 1983), this process might be termed *tactus extraction*). This first cognitive requirement is as crucial for dancing as for playing along, and indeed even a passive listener cannot be said to understand the music if they cannot accomplish this first non-trivial step.

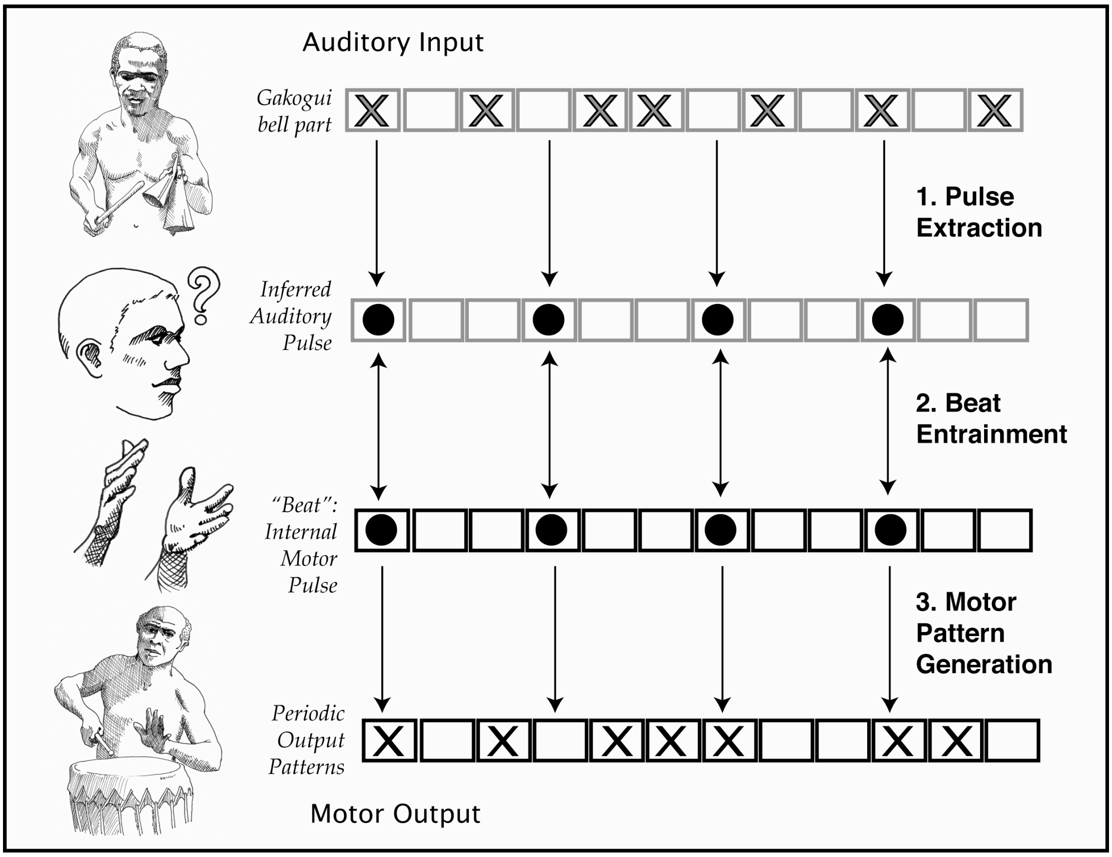

Figure 1: A schematic view of the three cognitive processes hypothesized to underlie ensemble playing and/or dancing, using a West African "agbekor" bell pattern and drumming ensemble as an example. Gray boxes represent auditory components, black boxes motor components. Xs indicate the onsets of externally observable events, while circles indicate the timing of mental events. Although the diagram represents a simplified, one-way movement of information from the bell pattern to a periodic motor output, in reality information can flow in both directions. In particular, the bell player may respond to the playing of the master drummer, constituting "mutual entrainment".

Pulse extraction is made more complex in much African drumming music by the fact that it is polyrhythmic: there are multiple ways to parse the musical surface into an underlying metre and pulse. This rhythmic ambiguity allows such music to be extremely cognitively complex, despite a strictly-kept time, the relative simplicity of the patterns played, and specific rules regarding variation. The formal structure of the agbekor bell pattern mirrors that of the diatonic scale (and indeed provides a rich range of rhythmic possibilities, comparable to that exploited in diatonic harmony in Western classical music, see below). Thus, there is a considerable range of complexity covered by the notion of "pulse extraction", and the simple notion of tapping your foot to an insistent bass drum pulse does not even begin to do it justice (Arom, 1991). We will return to these issues below, but for now let us assume that the participant has inferred at least one possible pulse from the acoustic pattern, enough to potentially dance along with.

The next step is to entrain one's own motor output to the extracted pulse. If one's motor output is simply tapping one's foot, this could be as basic as periodically tensing a single leg muscle "to the beat": this would represent what I will call **beat entrainment** and illustrates entrainment in its simplest form. More typically, for the musicians or dancers, entrainment involves using the pulse extracted from a complex auditory stimulus to drive one's own *complex* motor output. Such entrained **motor pattern generation** is a perfect example of complex cross-modal integration. Information from one cortical domain (auditory cortex, centred in the temporal lobe) must be coherently propagated to another (motor cortex, in the frontal cortex just anterior to the central sulcus). It involves both an entrainment of one's own internally-generated pulse to that extracted from the acoustic stimulus, and the use of this to generate a (potentially highly complex) motor output. This output is not simply a slavish reaction to, or acoustic mirror of, that extracted beat. Thus, the phenomenon of entrainment can be cloven into a synchronization component (where the listener's motor pulse is brought into alignment with that inferred from an acoustic signal) and a pattern output component, where the internal pulse is used to drive motor output. Note that the motor output can be simple or complex, and that in the case of complex output there are typically many different possibilities that are all, equally correctly, spoken of as "entrained". One can be able to play one drum part and not others (or a dancer may be unable to produce any of the instrumental parts, but produce various complex dances), but all of these performances may be equally "entrained". Thus, while the output of the pulse extraction and beat entrainment stages may converge on the same output for all participants (*the* pulse), the pattern output stage may be wonderfully divergent.

Summarizing, an ensemble player (or dancer) in an agbekor session must listen to the bell pattern (shown in Figure 1a), extract the underlying pulse from this pattern (Fig. 1b), and then synchronize their own internal cognitive pulse to this (Fig 1c). While this alone would be enough to constitute "entrainment", most musical behaviour involves a final step, where this internal pulse is used to drive a complex motor output, a "motor score", that could be implemented either via dancing, clapping or singing to this internalized beat, or by producing a different musical output (as symbolized in Fig 1d, with a notation of a master drum part).

As already noted, each of these steps might be, in some sense, just a different manifestation of a single "rhythmic ability", which you either have or you don't (Sachs, 1952). Indeed to accomplished aficionados of a given musical style, the process just described certainly feels unitary. One does not, consciously, go through separate steps of pulse extraction, beat synchronization, and pattern generation. However, there is good reason to suspect that this subjective unity is misleading, and that neurally and biologically, rhythm is a composite system. Certainly, the perceptual and motor aspects can be separated, in the sense that one can hear the beat, internally, without manifesting it via any body movements of one's own. More crucially, one can infer a beat, and attempt to produce motor output entrained with it, but fail (e.g. because the ensemble is playing too fast, or because the part is too complex and has not yet been fully automatized). However, this last distinction I have made, between simple beat entrainment and complex pattern entrainment, may seem less motivated until we delve into the question from a biological viewpoint.

Coupled oscillators and entrainment in nature: The "Paradox of Rhythm"

Oscillators are everywhere in nature, both in the physical world and the biological. At the physical level planetary motion gives us an important set of periodic phenomena (years, days, tides, seasons,...), and biology is rife with periodic functions including the familiar heartbeat and breathing cycles, but also including a set of biological clocks, tracking days and seasons, that are entrained to the physical timekeepers just mentioned. Such "biological clocks" are present in the simplest forms of life, such as bacteria, algae, yeast and other single-celled organisms. Thus both periodicity, and entrainment of biological clocks to external, physical timekeepers, are ubiquitous in biological systems.

It has also become clear with advances in our understanding of nonlinear dynamics that any form of coupling between nonlinear oscillators can easily lead to phase locking and mutual entrainment (e.g., between heartbeat & respiration, or two vocal folds: Glass and Mackey, 1988, Fitch et al., 2002). This leads to an apparent paradox: one that we might term the "**paradox of rhythm**". Periodicity and entrainment seem to be among the most basic features of living things, yet the human ability (and proclivity) to entrain our motor output to auditory stimuli appears to be very rare: there is no credible evidence of even simple beat entrainment among other primates or mammals (Merker, 1999, 2000). Among *all* living organisms, mutual entrainment remains rare in its simplest forms (Greenfield, 1994), and it is perhaps unique to our species in even the mildly complex forms observed in a musical ensemble playing a simple piece. The paradox, put simply, is this: if periodicity and entrainment are ubiquitous features of all living organisms, why can't dogs or other animals keep a beat, or dance? Of course, one answer to this question is to note that absence of evidence does not provide strong evidence of absence. Perhaps dogs *can* dance, and we just can't *tell* that they're dancing. More pertinently, there remains much to be learned about rhythmic behaviour in great apes, and it is certainly too early to state firmly that they are incapable of entrainment (c.f., Fitch, 2006b). A wide variety of bird species show duetting and various bodily movements suggestive of entrainment, but have not been thoroughly investigated from a musicological viewpoint. Nonetheless, it is safe to say that among primates, particularly the well-studied chimpanzees, evidence of entrainment is very sparse, and no published experimental data suggests entrainment. Thus I will cautiously proceed from the assumption that humans are unusual in our proclivity for mutual entrainment of complex patterns, and that this represents a quite marked difference from our nearest living relatives, even if the difference is not absolute, but rather a (large) quantitative difference.

It is interesting to note that cross-modal integration, in any form, was believed to be unique to humans, until the late 1960's, and this was seen as a crucial advance in cognitive organization between chimpanzees and humans (Cole et al., 1961, Geschwind, 1970). However, numerous studies have now shown cross-modal transfer between sensory modalities in apes and monkeys (Ettlinger and Blakemore, 1969, Davenport and Rogers, 1970, Savage-Rumbaugh et al., 1988, Ghazanfar and Logothetis, 2003). The degree to which this is flexible and general, or tied to specific modalities (e.g. face/voice perception) remains unclear. Note that there is good evidence in both humans and other mammals for cross-sensory flexibility in vision, touch and audition: humans born blind use their occipital cortex for Braille reading (Cohen et al., 1997), and ferrets whose brains are "rewired" in utero, so that visual information is directed to temporal cortex (which contains primary auditory cortex in normal mammals) develop a working visual cortex in the temporal lobe (Sur et al., 1990, von Melchner et al., 2000). These data suggest that sensory cortices are in some sense interchangeable during ontogeny. This may be because the structure of sensory cortices is relatively consistent across modalities, which all have a well-developed granular input layer (in layer IV of neocortex). This similarity may ease the computational burden of cross-modal transfer between sensory domains.

In contrast, cross-modal sensory/motor integration may provide a greater neuro-computational challenge. The cytoarchitectonic layout of primary sensory areas is very different from that of the motor cortex, which is agranular and includes a special "output" cell type, the giant Betz cells in layer V that are a cytoarchitectonic indicator of mammalian motor cortex (for discussion see Shipp, 2005). In addition to this neural evidence, research in robotics has demonstrated sensory-guided motor behaviour to be one of the most challenging aspects of developing machines that behave with a modicum of intelligence. Although any living vertebrate out-performs current robots, this is almost certainly due to a half-billion years of intense selection, which has built up a large collection of neural tricks and shortcuts (hard-coded "looming detectors" and the like), rather than a general solution to the sensory-motor integration problem (Brooks, 1989). Because the entire purpose of sensory systems is to guide adaptive behaviour, it is not surprising that brains are good at using this information. But since natural selection has no foresight, we expect (and find) partial, domain-specific "Good Tricks" (Dennett, 1995) to the problem of sensory-guided action, and a palimpsestic inheritance and re-use of such tricks during vertebrate evolution. What would be surprising is the discovery of a general-purpose transform capable of transforming *any* complex sensory input into a complex motor output. But this is precisely what humans appear to have evolved (though see Patel et al., 2005).

Summarizing, the generation of a complex patterned motor output entrained to an *acoustic* pattern is not in any sense assured by the ubiquity of periodicity, or a capacity for cross-modal sensory transfer. Furthermore, even extraction of a beat, and entrainment to that beat in a pure motor fashion (for example in dancing) does not assure the ability to *musically* entrain, for one can be a dancer (motor -> motor) but not an instrumentalist (motor -> acoustic). This is due to the fact that playing an instrument is a motor act, but a crucial aspect of this act is a careful attention to the instrument's acoustic output. Part of learning to play an instrument is the gradual (and difficult) process of closing the motor/auditory loop by playing, directly analogous to babbling in the vocal domain. Playing involves an additional level of complexity over the (already challenging) dancing problem.

Thus, the "paradox of rhythm" already seems less paradoxical if we subdivide the cognitive capacities underlying rhythmic performance into different subcomponents. I will now turn to a consideration of the animal comparative data, in an effort to ascertain which (if any) of these sub-components was present before the emergence of humans.

Rhythm in our nearest cousins: The African great apes

Questions about the evolutionary timing of mechanisms underlying rhythm can be recast in phylogenetic terms as follows. Humans and chimpanzees (our nearest living cousins) shared a most-recent common ancestor ("last common ancestor" or LCA) between 5- 7 million years ago (Chen and Li, 2001). This species was neither a chimpanzee nor a human, but shared many characteristics with both (Carroll, 2003). We can use the comparative method, examining the behaviour of living humans, chimpanzees, and other apes, to reconstruct this LCA. From this reconstruction, we can determine which (if any) rhythmic capacities it possessed.

A striking fact, as soon as we search for rhythmic acoustic generation in other animals, is that both chimps (an informal term I will use to denote both chimpanzee *Pan troglodytes* and bonobos *Pan paniscus*) and gorillas (*Gorilla gorilla*) show periodic acoustic behaviour. These African great apes frequently create sounds by pounding with their limbs (most typically hands, but also often with their feet) on resonant objects. This behaviour is termed "drumming" by ape researchers (e.g., Schaller, 1963, de Waal, 1988, Arcadi et al., 2004). This is strikingly similar, in many ways, to human hand drumming (found in virtually all human cultures). Ape drumming thus seems a plausible homologue of human drumming (Fitch, 2006b). Ape drumming appears throughout wild and captive chimpanzee populations (Arcadi et al., 1998, Arcadi et al., 2004), and appears in gorillas who have never seen others perform the behaviour (Schaller, 1963), and thus appears to represent a genetically-based species-typical characteristic of African apes. Drumming functions both in playful behaviour among young animals and in display behaviour by adult males. While gorillas typically drum on their own (or sometimes others') bodies, chimpanzees tend to drum on external objects. Chimpanzee individuals seek out hollow or resonant objects with particular acoustic properties, and return to those objects repeatedly to drum, a behaviour clearly indicating a linkage between the auditory and motor systems (Kirschner et al., in prep). That drumming behaviour is found in these species, and in humans, but not in orangutans, suggests that the underlying mechanisms represent an African great ape homology, present in the common ancestor of humans, chimpanzees and gorillas, but not before (thus around 10 million years ago). Although detailed studies of the similarities and differences in drumming behaviour between humans, chimpanzees and gorillas are underway (Kirschner et al., in prep), it is safe to say that this fascinating example of a potentially homology has gone little appreciated and remains poorly studied and understood; this is made more surprising by the fact that examples of such drumming are very rare outside of our particular primate lineage (c.f., Fitch, 2006a). In the three-way distinction I have sketched above, I see ape drumming behaviour as providing clear evidence that the last performance stage - mapping a quasiperiodic motor pulse into an auditorily-monitored acoustic motor behaviour - was already present long before the emergence of humans as a separate lineage.

In contrast, there is currently little evidence of entrainment in ape drumming: "Free rhythm ... is doubtless the earlier quality. Strictness comes with man" (p. 20-21 Sachs, 1953). Despite such statements, this area of animal cognition remains poorly explored. In general, chimpanzees perform their drumming displays alone (it is often the dominant male who produces the drumming at the climax of a display including a loud, stereotyped vocal component as well as vigorous motor activity), so there would be little opportunity for, and thus evidence of, social entrainment. While young gorillas sometimes exchange drumming bouts during play, often as a prelude to play chasing or other playful interaction, we have no observations of synchronized drumming and none are reported in the literature (e.g., Schaller, 1963). There is better evidence for mutual entrainment in vocalizations, for example the "staccato hooting" of bonobos (de Waal, 1988), or the elaborate interlocking duets of male and female gibbons and siamangs (Geissmann, 2000), suggesting that at least some degree of vocal entrainment may be possible in these species, but at present there is no definitive evidence of clear, periodic entrainment in ape drumming. Thus the safest assumption at present appears to be that beat entrainment either emerged, or underwent strong positive selection, sometime after the split between chimps and humans around 6 million years ago. However, experimental data on chimp entrainment is sorely needed to test this supposition.

Finally, what of pulse extraction (inference of a "beat" from a complex acoustic pattern)? Here we know even less about animals. Part of the problem is methodological - the clearest way to ascertain if pulse extraction has occurred is to have the subject tap along to a complex rhythmic pattern and examine their pattern of abilities and errors (e.g., Fitch and Rosenfeld, 2007). But this obviously requires subjects who both have beat entrainment and can produce at least a simple motor output, in addition to the auditory perceptual abilities underlying pulse extraction. Although there is a long history of experimentation with animal timekeeping, in the sense of rats learning to press a bar with a certain periodicity (e.g., Collyer and Church, 1998), this ability concerns maintenance of a motor pulse, not entrainment of that pulse to an auditory stimulus. The clearest experiment, for an animal, would be to do tempo matching, where complex patterns at the same or different tempos are presented, and the animal must respond appropriately whether the tempos are the same or different. Given the difficulties training macaques on a far simpler auditory "same/different" task (see Brosch et al., 2004) such an experiment would be very challenging to perform, and it is perhaps not surprising that results of such experiments have not, to my knowledge, been published. More promising species for such an operant task would be birds. Songbirds can be readily trained in an auditory operant tasks (e.g., Gentner et al., 2006) (though note that pigeons apparently fail to master even visual same/different tasks (e.g., Huber, 2001)). Thus, no clear comparative pronouncements can be made about pulse extraction at present. My guess is that animals such as songbirds can extract a pulse from a complex acoustic pattern, but this is speculative, and the situation is far less obvious for nonhuman mammals, such as rats or chimpanzees. Thus our main source of "comparative" data for pulse extraction from acoustic patterns must come from a comparison, within humans, of language and music, to which I will now turn.

In summary, there is good preliminary evidence for rhythmic motor output in great apes, and it supports the distinction I have made between beat entrainment and pattern generation. The comparative data suggest that what is missing in chimpanzees is the beat entrainment component, allowing a tentative hypothesis that this is the last of the three components to appear in human evolution.

Within-Human Comparisons in Music and Language

A long-running debate in human psychology concerns the "modularity", or lack thereof, in the cognitive mechanisms underlying language (Fodor, 1983, Bates, 1994). Do language and music rely on domain-specific, or general purpose, cognitive and computational mechanisms (Peretz and Morais, 1989, Patel, 1998, Peretz, 2006, Patel, 2008)? Again we must recognize that "language" has many components, and that the answer to this question may vary from one to the other. While long term memory and fast learning mechanisms (e.g. for word meanings) may rely on precisely the same cognitive mechanisms as memorizing other facts about the world (e.g., Markson and Bloom, 1997), syntax might rely on dedicated mechanisms unique to the syntactic domain (Chomsky, 1986, Crain, 1991). There is no contradiction here, and the starting point for any reasonable discussion of modularity in language and music is thus a clear delineation of the specific cognitive abilities under discussion (as advocated in Hauser et al., 2002, Fitch et al., 2005, Patel, 2006). Here I compare the cognitive mechanisms underlying rhythmic musical performance discussed above with those involved in harmony and in speech, particularly in the domain of metrical phonology (c.f. Goldsmith, 1990). I wish to draw particular attention to the abilities involved in acoustic pattern perception, which I will argue have an important shared, or general, component.

A major challenge confronting comparisons between language and other cognitive domains is that we know so much about language, at a detailed formal level, compared to other domains of human cognition. In the case of musical rhythm, however, our understanding seems to me adequate to attempt such a comparison in a relatively balanced fashion (e.g., Longuet-Higgins and Lee, 1984). But before this it is worth asking whether there may be even more general pattern-perception capabilities, typical of all perceptual domains. This, of course, was the basic claim of the Gestalt psychologists, starting with Wilhelm Wundt (Wundt, 1908), and continuing to the present day (e.g., in Lerdahl and Jackendoff, 1983, Bregman, 1990). While I think it is reasonable to assume, along with Gestalt theorists, that certain aspects of perception are universal, this does not exclude the possibility of quite specific perceptual processes that apply to audition and not to, say, vision, touch or smell. The auditory system has its own rather specific problems to deal with, and hard-wired solutions to low-level problems like source-localization have no equivalent in the visual or tactile domains (Masterton and Imig, 1984, McAlpine and Grothe, 2003). Thus it seems reasonable to proceed on the assumption that pattern-extraction capabilities specific to the auditory domain *could* exist (e.g., Patel et al., 2005), and let the data tell us if we're incorrect.

Comparisons Within Music: Pitch and Rhythm

A nice example of a domain in which Gestalt principles, and more general mathematical (group-theoretic) constraints might govern auditory perception is provided by a fascinating similarity across the rhythmic and tonal domains (Pressing, 1983, Schloss, 1985). This potential example of a pan-musical perceptual principle is nicely illustrated by the isomorphism between the agbekor bell pattern already discussed and the ubiquitous diatonic scale of Western music. As Fig 2 shows, the pattern of alternating large and small pitch intervals in the diatonic scale is identical to the pattern of durational intervals in the agbekor pattern. This isomorphism seems very unlikely to result from chance (c.f. Toussaint, 2003). As Toussaint points out, the agbekor bell pattern (which he refers to by the name of the Cuban rhythm, "bembé") is one of the most popular and widespread of all such patterns, found widely throughout sub-Saharan Africa and diverse musical cultures in the New World. We thus have an extremely popular and successful rhythmic pattern with the same structure to the diatonic system which is probably the most widespread scale on Earth today. Both rely upon an underlying division of a cycle (either the rhythmic period, or the octave) into twelve equal divisions; both use either 1 or 2 of these intervals as the spacing between events/notes; and both distribute these intervals in a directly isomorphic fashion. This suggests, perhaps surprisingly, that similar pattern-extraction mechanisms may apply in the domains of rhythm and tonality, despite gross differences in the musical surface analyzed, providing support for the idea of broad cognitive/aesthetic principles applicable across musical domains.

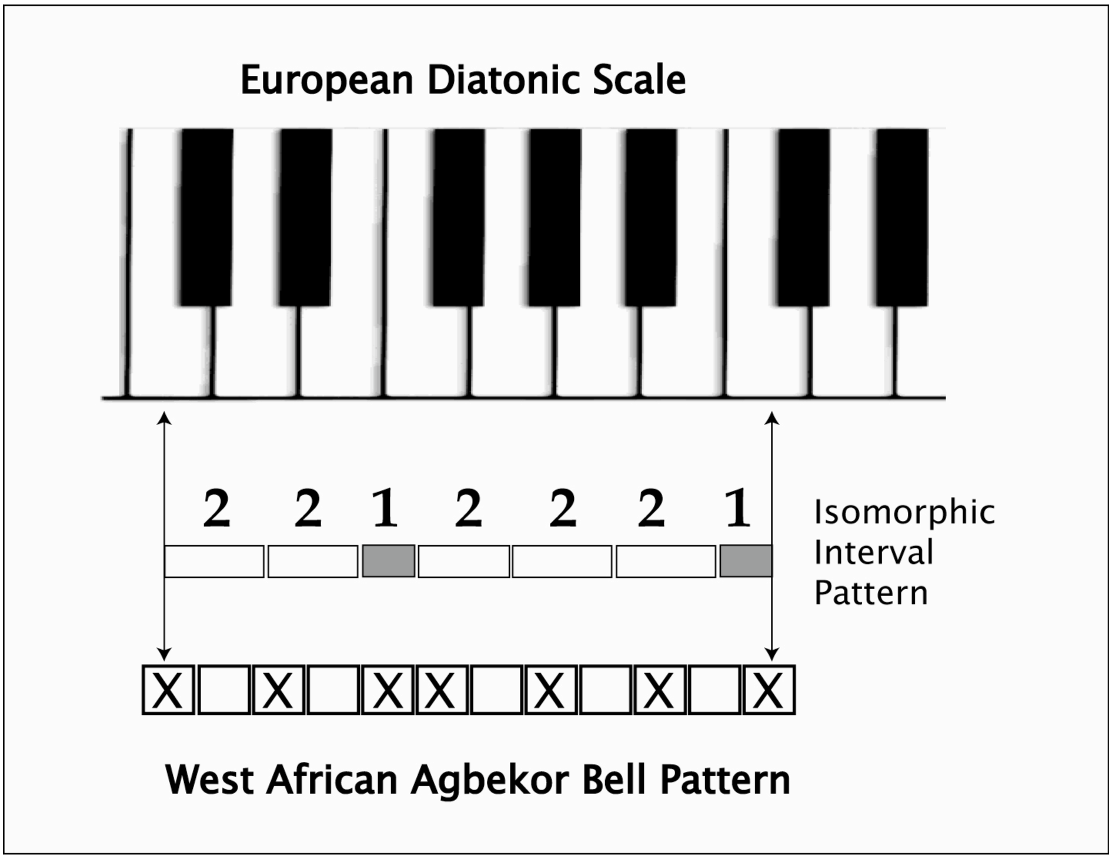

Figure 2: Isomorphism between the (logarithmically-spaced) pitch intervals of the diatonic scale with the temporal intervals of the widespread "agbekor" bell pattern.

What is the cause of this similarity, an isomorphism linking two very different musical domains (tonality and rhythm) and musical cultures (European and West African)? Both the rhythmic and tonal examples have received quite extensive mathematical treatment, and there are multiple ways of characterizing them mathematically, probably all valid to some degree (for an introduction to this literature see Cassirer, 1944, Pressing, 1983, Zweifel, 1996, Toussaint, 2003)). Of course, mathematical similarity does not necessarily indicate any fundamental causal similarity: there are a huge variety of natural phenomena well-modelled by normal curves or power laws (distribution of star sizes, human heights, word frequency, personal incomes or neuron interconnectivity) bearing no fundamental relationship at the level of the phenomena that generate these different phenomena (e.g., Keller, 2005). While these isomorphisms therefore render plausible suggestions that fundamental neural, cognitive or perceptual constraints generate both the diatonic scale and the bell pattern, they by no means guarantee this conclusion (c.f. London, 2002). It is quite possible that there is some fundamental constraint at the level of neural circuitry (perhaps a developmental constraint on how nerve networks are epigenetically wired together), but that the actual networks involved employ quite different neurons in different areas of the brain. Thus, we cannot expect brain imaging to necessarily resolve this question. Note further that if these similarities result from truly *perceptual* constraints, as opposed to musical or auditory constraints, we should expect similar isomorphisms in the domain of visual patterns in art (e.g. quilts, textiles, mosaics, pottery patterns, and similar plane patterns). But, despite some detailed mathematical analyses of such patterns (e.g., Washburn and Crowe, 1998, 2004) and a long history of suggestive allusions (e.g., Gombrich, 1984) I know of no evidence that this is the case.

Thus, the discovery of isomorphisms in the rhythmic and tonal domains raises fascinating questions about the nature of human perception and aesthetic cognition that cannot be answered at present. Although this discovery certainly suggests that there are deeper commonalities in musical cognition than a modularist account might suggest (e.g., Peretz and Morais, 1989, Peretz and Zatorre, 2005) this has not been demonstrated. Nor can we conclude from such findings that there are "general Gestalt principles" that govern such phenomena, given the lack of evidence for similar phenomena in non-auditory perceptual domains. Although the isomorphism is a fact, its causal basis remains a fascinating mystery, a mystery whose resolution may yield deep insights into (at least) musical cognition if explored experimentally at the perceptual, cognitive, and neural levels.

Comparative Data 3: Hierarchy in Speech and Rhythm

I now turn to a comparison between "rhythm" in spoken language and music, a cross-domain similarity that appears to me to represent another fascinating isomorphism between two superficially very different domains (c.f. Patel, 2008). Linguists use the term "rhythm" in a rather different fashion from musicologists. For them, the term connotes an alternating system of accents combined with patterns of grouping, of the sort typical in poetry. Thus, for example, we describe the prosodic rhythm of Shakespeare's sonnets as "iambic pentameter" because the syllables are arranged in pairs ("feet") with accents distributed in a weak-STRONG fashion (termed "iambs"), and organized into lines of five iambs (hence "penta-"). For example:

When I / do COUNT / the CLOCK / that TELLS / the TIME, ... (Sonnet 12)
or
Shall I/ com PARE / thee TO / a SUM/mer's DAY? ... (Sonnet 18)

Here, although the feet (separated above by "/") and the accented syllables will occur at roughly equal intervals, there is no suggestion that they should fall precisely on some isochronic beat. Indeed, if read in strict musical time, the output would seem somewhat forced and mechanical. Although human perceivers often *think* that syllables or stresses are evenly spaced or isochronous, which has led to the distinction between "stress-timed" and "syllable-timed" languages, repeated acoustic analyses have shown that neither type of speech in fact exhibits isochrony (Roach, 1982, Dauer, 1983, Ramus et al., 1999).

Thus, a clear distinction between musical rhythm and poetic rhythm is the degree to which they connote (or require, in many forms of music) a strict, metronomic "beat" governing the occurrence of events. Nonetheless, there are other fundamental similarities to consider, particularly if we compare spoken poetry with musical rhythm. Both have a metre (a rule or rules determining the size of the sets into which events are perceptually grouped). Both have a notion of "accent" (also called "stress" or "prominence"), and rules about how such accents should be distributed. Furthermore, both have larger units into which these groups are arranged (e.g. phrases in music, lines or sentences in speech), and further rules governing relations among these phrases (e.g. phrases can be marked by phrase-final lengthening; non-adjacent phrases may be required to rhyme, etc). Finally, all of these "rules" are actually defeasible; they can be overridden for aesthetic or semantic purposes. Thus they are better thought of as constraints that provide some underlying conceptual structure and predictability, but not as rules strictly determining the structures actually produced. Together, these formal regularities comprise the **hierarchical structure** of speech prosody and music, and seem to represent one of the clearest examples of overlap between these domains. Very recent neuroimaging data support the notion of shared processing of this aspect of speech and music rhythm (Geiser et al., 2008)

Social Origins of Isochrony?

If the rhythm of speech, as studied by metrical phonologists, shows that isochronicity is dispensable, why then is it so ubiquitous in the music of the world? I think the *exceptions* to regularity in music help to understand the rule, for there are actually numerous forms of music in the world that lack isochrony. Most of these are *solo* forms. After the most comprehensive review of "free rhythm" (musical styles lacking a pulse or metre) I know of, Clayton concludes that such music "is unlikely to be performed by an ensemble (at least not in a manner demanding strict synchronization), or to be danced to" (p 330, Clayton, 1996). This suggests that isochronicity is actually a by-product of a more fundamental need for synchronized interactions *between different individuals*. When an individual is speaking, producing music or dancing alone, there is no need for such synchronization: "As long as singers stand alone ... the urge for strictness in rhythm and tempo is very weak" (p. 35 Sachs, 1953). Indeed, in a solo situation, a far freer approach to tempo and metre is possible. This broadens expressive potential, allowing text meaning to dominate (as in many ritualistic/religious "songs" in Jewish, Islamic and Persian traditions (Frigyesi, 1993)), or allowing a freer approach to melodic improvisation, as in the unmetrical *alap* introduction typical of North Indian classical performance (Clayton, 1996).

But as soon as multiple individuals attempt to blend together, either producing separate events synchronously, with some syncopation, or even anti-synchronizing in call-and-response fashion, some predictable tempo becomes invaluable (c.f. McNeill, 1995, Merker, 1999). Anyone who has witnessed the cacophony of a room full of young children given drums for the first time, or attempted to dance to a beginners' drum circle keeping irregular time, will understand intuitively what I mean. More formally stated, if the goal of an ensemble is to cooperatively generate phase coherent patterns between multiple periodic events, produced by multiple individuals, the simplest way to achieve this is to chose a single base frequency (a "tempo") for all phases. By either following some timekeeper (a leader, or a metronome) or mutually entraining to one another, ensembles sharing such an underlying frequency can easily achieve phase locking even for quite complex, interlocking parts. Although it is formally possible to achieve phase coherence among different phases at different frequencies (e.g. in "cross time" such as 2 against 3, or more complex polyrhythms), even in this case the frequencies should be small integer multiples of one another, and thus they connote a shared base frequency. This provides another reason to separate beat entrainment from complex, entrained pattern generation.

Furthermore, the fundamentally social nature of the drive I hypothesize underlying such (in principle rather simple) entrainment returns us to questions about the biology and evolution of such a capacity. Chimpanzees are highly-skilled at many complex motor actions. They exhibit most of the nonhuman tool use known in nature, and their own drumming behaviour reveals their ability to generate such patterns based on a quasiperiodic pulse. However, chimpanzees show little evidence, relative to humans, of a strong cooperative urge underlying complex group activities (e.g., Hare and Tomasello, 2004, Melis et al., 2006). Even "cooperative hunting" in chimpanzees (Boesch and Boesch, 1989, Boesch and Boesch-Achermann, 2000), which provides the best example of ape group action, can be seen as the self-organizing activity of multiple independent individuals, who share a common goal ("catch the monkey") but no shared plan for achieving this goal (Busse, 1978). In contrast, human rhythmic behaviour may represent just one expression of a more general "cooperative urge" - a motivation to share experience, activities, and emotions with others - that is so typical of our species, and so unusual in the natural world (Dissanayake, 2000, Cross, 2003, Tomasello et al., 2005). This, I suggest, provides a ready explanation for our ability, and proclivity, to entrain to an isochronic beat. Recent data (Kirschner and Tomasello, submitted) suggest that a capacity to entrain body movements to an isochronic beat is already present in 2.5 year-old children, though in young children it depends on an adequate social context (a human drummer) to be reliable. These data suggest, again, that it is the capacity for entrainment to a *strict periodic pulse* that came latest in human evolution, but that this capacity might reflect some more general social, cooperative urge, rather than any specifically perceptual or musical mechanism.

This hypothesis has implications for discussions of the adaptive value of music. While many authors have noted the value of a componential approach to the *mechanisms* underlying music (Justus and Hutsler, 2005, McDermott and Hauser, 2005, Fitch, 2006b, Patel, 2006), it is not as frequently recognized that this applies just as well to the *functions* that they serve(d) (whether full-fledged adaptations or not). We cannot assume that behaviours of some extinct hominid which served as precursors to modern human music were themselves "music". Thus, when we (e.g., McDermott and Hauser, 2005, Patel, 2006) pose questions about adaptive functions of specific mechanisms like pitch or rhythm perception in terms of "natural selection for music", (or alternatively "for language"), we implicitly downplay the possibility that the selective value of such mechanisms may have been unrelated to either. In this vein, the "social convergence" hypothesis proposed above is not wedded in any specific way to music: musical synchronization might be just one of a host of different types of behaviours that are enabled by a general-purpose ability to predict and synchronize with the movements of others. Isochrony, rather then being an adaptation for music, may simply be a culturally discovered aid to synchronization.

While this hypothesis remains speculative given the dearth of data on mutual entrainment in other animals (c.f., Merker, 1999), it is consistent with the available evidence and (more importantly) testable with new data that could be readily gathered. Two contrasting sets of predictions can be generated. First, noting the crucial role of cross-modal, audiomotor, integration in synchronization, (Patel, 2006) suggested that synchronization may be an offshoot of vocal learning - what he termed the "vocal learning and rhythmic synchronization" hypothesis. This is a plausible hypothesis, and makes the strong prediction that synchronization will *not* be observed in chimpanzees, gorillas or other "drumming" mammals (e.g. kangaroo rats), where vocal learning is lacking. It also suggests that synchronization *will* be found in birds, seals or cetaceans who possess well-developed vocal learning capabilities. In contrast, the "social convergence" hypothesis predicts synchronization abilities should be best in group living species, particularly those who engage in cooperative group behaviour (e.g. social canids, elephants, white pelicans), and perhaps in chimpanzees as well (accepting at face value the characterization of their hunting as "cooperative" Boesch and Boesch, 1989, Boesch and Boesch-Achermann, 2000).

The biological basis of music/language similarities: Darwin's theory of language evolution

I end by discussing the phylogenetic trajectory that might have led to the mosaic of shared and dissimilar rhythmic components discussed in this paper. I will focus on the comparison between music and language, but many similar questions could be raised for the relationship between rhythm and tonality in human music, or indeed between human music and the complex, learned vocalizations called "song" in other species (c.f., Fitch, 2006b).

One place to start is with a consideration of the multiple "design features" of language and music (Hockett, 1960, Fitch, 2006b). Spoken language and song share all the design features intrinsic to the audio/vocal modality, along with the features of being complex and being learned. As we have seen above, speech and musical rhythm also share a quite specific form of prosodic hierarchical structure. More generally, although both music and language primarily rely on the audiovocal channel, each can potentially be expressed via other channels (e.g. dance for music, and gesture or signed language for linguistic communication). Probably as result, both music and language can be written as well: both are "supramodal" systems. Finally, both systems are culturally transmitted and thus subject to local diversification into dialects or styles, and these styles can themselves become "badges" of one's provenance, and thus potential indicators of an individual's kinship or group affiliations. All of these factors, combined with the existence of complex, learned songs in songbirds, led Darwin to suggest that the first stage of the evolution of language involved something closer to music: a "musical protolanguage" (Darwin, 1871). A strange quirk of recent scholarship in the evolution of language and music is that Darwin's model has been almost universally overlooked (and indeed repeatedly reinvented). This oversight is all the more surprising given that "The Descent of Man" is hardly an obscure work, and more importantly by the fact that (at least in my view), it provides the strongest available model for language evolution.

Darwin's model, slightly updated in light of contemporary research, posits that, during an initial stage of human evolution since our deviation from chimps, proto-humans developed the capacity for complex vocal learning which underlies both speech and song. This system, akin to bird- or whale-song, was not meaningful in any propositional or denotational sense - and thus more like song than spoken language. However, neither was it identical to modern music: there is no reason to suppose that it used discrete pitches or isochrony. The uses of this vocal communication system may have been various (c.f., Roederer, 1984, Morley, 2002, Mithen, 2005), including sexual (courtship and territorial advertisement), as proposed by Darwin, parent-offspring communication (Trainor, 1996, Trehub and Trainor, 1998, Dissanayake, 2000, Falk, 2004), or perhaps group bonding (Brown, 2000, Hagen and Bryant, 2003) or advertisement (Merker, 1999, 2000). Although none of these functions need be exclusive, the best data currently available support the parent-offspring, and specifically mother-infant, function at present (for a review see Fitch, 2006b). Only group advertisement hypotheses (Merker, 1999) entail isochrony at this first stage. Thus, in essence, Darwin's hypothesis posits that the first stage of language evolution was "bare" phonology: phonology without meaning. The crucial mechanisms here are vocal learning (shared by speech and song, and one of the crucial differences between humans and chimpanzees), and hierarchical parsing (as shared by spoken language and musical rhythm). The properties shared by phonology and syntax (e.g. generativity, infinite use of finite means) would already be present at this stage, but complex syntax would not be.

A non-intuitive aspect of Darwin's model is that it suggests a two-stage evolution of complex syntax. In the first "musical" stage, we would have only hierarchical structure, in the specific sense discussed above including notions of grouping, metre, stress/accent, and phrasal prosody. An additional non-intuitive assumption of this model would be that meaning in protolanguage, to the extent that it existed, would have been quite general and probably context-dependent: although there might have been love songs, hunting songs and lullabies, there would have been no word for "love", "hunt" or "sleep". This "holistic" aspect of meaning in the hypothesized protolanguage, as specified in (Jespersen, 1922, Wray, 1998, Kirby, 2000, Arbib, 2005), is the one that has attracted the most criticism (see Tallerman, 2005), and it certainly contrasts with the intuitive notion of cavemen grunting out monosyllables signifying meat, fire, spear, and the like (as posited in the more popular "lexical protolanguage" model of Bickerton, 1990). By the Jespersen/Wray hypothesis, true words would appear only in the last stage of language evolution, by a slow cultural process of "fractionation" of complex holistic phrases into more compositional subunits. By hypothesis, this would be driven by the underlying, pre-existing compositionality of thought, via an accumulation of "mistakes" where children perceive a compositional structure in adult speech that is not (from the speaker's viewpoint) really there. As non-intuitive as this proposition may seem at first blush, the existence of such fractionation processes can be empirically observed both in child language acquisition and historical linguistics (Jespersen, 1922, Wray, 1998), and a simple and elegant computer simulation by Kirby demonstrates that a combination of phonological complexity (ala Darwin's musical protolanguage) and semantic/conceptual compositionality is all that is required for such a process to proceed (Kirby, 2000). The end product yields most of the complex features that differentiate modern syntax from phonology - long-distance dependencies, case-marking, "movement" phenomena, etc - with most of the complexities of such phenomena "inherited" from the interface to the conceptual system.

Most of the data reviewed in this paper concerning rhythm are consistent with Darwin's hypothesis. Far more important, most of the open questions touched upon can serve as tests of the "musical protolanguage" model of language evolution: as we learn more about entrainment, hierarchical pattern parsing and so on, and begin to understand the neural and ultimately genetic bases of such capacities, Darwin's hypothesis makes clear predictions about what we should find. In particular, our ability to crudely date the "selective sweep" that drove particular alleles involved in particular cognitive capacities (as in Enard et al., 2002) will allow us to order their times of occurrence during hominid evolution. If Darwin is correct, we should find that phonological and musical skills tend to co-occur in modern humans, share neural and genetic mechanisms, and that these mechanisms (both vocal learning and hierarchical parsing) were selected earlier than those involved in word learning, pragmatics, or skill at propositional expression. The lexical protolanguage hypothesis predicts, in contrast, a very early fixation of word learning capacities, with hierarchical processing selected last. Although the day when such data are available may seem far off, the extraordinary progress in molecular biology and comparative genomics in the past decade suggests that it should be within most readers' lifetimes. Thus Darwin's hypothesis does what any good evolutionary scenario should do: makes testable predictions, and points to sources of empirical data that would otherwise remain unexplored.

Conclusion

I have suggested in this essay that by analyzing human "rhythm" into component mechanisms (pulse extraction, beat entrainment, and motor pattern generation), we can resolve an apparent paradox about its limited distribution among animals. Darwin's model resolves a similar, but broader, puzzle about the relation between music and language, because it is simultaneously a model of language evolution and of the evolution of music (and specifically song). By this model, music and song represent, in a sense, "living fossils" of an earlier communicative stage of humanity. This supposition neatly solves a core puzzle about music: why should something so apparently useless as music be so ubiquitous in human cultures, attract so much of our time and interest, and seem to have such deep and powerful effects on our emotions? Why should these "powers" of music seem to transcend cultural boundaries so much more readily that spoken language (which serves, quite effectively, to isolate human cultures from one another). While I do not condone the naïve view that music is a "universal language of emotion", both the emotional power and relative transcendency of music are among its most obvious features. By positing music as the vestige of a once-central communication system, ousted when language became our "main" system, Darwin provides a coherent resolution to this puzzle.

Although one may dislike the connotations of the idea that music is a "living fossil", as a musician I find the notion of musical protolanguage a profoundly attractive one. If this model is correct, the "sidelining" of music by language has given our species a rich, unfettered playground for creative expression. Freed from any necessary adaptive, communicative functionality, the emancipated musical faculty, in myriad strands of human musical culture, has been allowed to explore vast realms of aesthetic pleasure of profound intellectual interest and emotional depth. The diversity and greatness of the cultural achievements thus produced requires no utilitarian or adaptive function to justify the respect and admiration, as well as the devotion and passion, that music so widely evokes in humans around the globe.

Acknowledgements

I gratefully acknowledge the insightful comments of Sebastian Kirschner, Bjorn Merker, Daniel Mietchen, Ani Patel, Patrick Rebuschat and W. Andrew Schloss on earlier versions of this manuscript.

References

Agawu, V. K. (**1987**). "The Rhythmic Structure of West African Music," Journal of Musicology 5, 400-418.

Arbib, M. A. (**2005**). "From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics," Behavioral and Brain Sciences 28, 105–167.

Arcadi, A. C., Robert, D., and Mugurusi, F. (**2004**). "A comparison of buttress drumming by male chimpanzees from two populations," Primates 45, 135 - 139.

Arcadi, C., Robert, D., and Boesch, C. (**1998**). "Buttress drumming by wild chimpanzees:

temporal patterning, phrase integration into loud calls, and preliminary evidence for individual distinctiveness.," Primates 39, 505-518.

Arom, S. (**1991**). *African polyphony and polyrhythm* (Cambridge University Press, Cambridge).

Bates, E. (**1994**). "Modularity, domain specificity and the development of language," Discussions in Neuroscience 10, 136-149.

Bickerton, D. (**1990**). *Language and Species* (Chicago University Press, Chicago, IL).

Boesch, C., and Boesch, H. (**1989**). "Hunting behavior of wild chimpanzees in the Taï National Park," American Journal of Physical Anthropology 78, 547-573.

Boesch, C., and Boesch-Achermann, H. (**2000**). *The chimpanzees of the Taï forest* (Oxford University Press, Oxford).

Bregman, A. S. (**1990**). *Auditory Scene Analysis. The Perceptual Organization of Sound.* (MIT Press/Bradford Books, Cambridge, MA).

Brooks, R. A. (**1989**). "A robot that walks: Emergent behaviors from a carefully evolved network," Neural Computation 1, 253-262.

Brosch, M., Selezneva, E., Bucks, C., and Scheich, H. (**2004**). "Macaque monkeys discriminate pitch relationships," Cognition 91, 259-272.

Brown, S. (**2000**). "The "Musilanguage" model of music evolution," in *The Origins of Music*, edited by N. L. Wallin, B. Merker, and S. Brown (The MIT Press, Cambridge, Mass.), pp. 271-300.

Busse, C. D. (**1978**). "Do Chimpanzees Hunt Cooperatively?," American Naturalist 112, 767-770.

Carroll, S. B. (**2003**). "Genetics and the making of *Homo sapiens*," Nature 422, 849-857.

Cassirer, E. (**1944**). "The Concept of Group and the Theory of Perception," Philosophy and Phenomenological Research 5, 1-36.

Chen, F.-C., and Li, W.-H. (**2001**). "Genomic Divergences between Humans and Other Hominoids and the Effective Population Size of the Common Ancestor of Humans and Chimpanzees," American Journal of Human Genetics 68, 444-456.

Chomsky, N. (**1986**). *Knowledge of language: Its nature, origin, and use* (Praeger, Westport, CT).

Clayton, M. R. L. (**1996**). "Free Rhythm: Ethnomusicology and the Study of Music without Metre," Bulletin of the School of Oriental and African Studies, University of London 59, 323-332.

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Falz, L., Dambrosia, J., Honda, M., Sadato, N., Gerloff, C., Catala, M. D., and Hallett, M. (**1997**). "Functional relevance of cross-modal plasticity in blind humans," Nature 389, 180-183.

Cole, M., Chorover, S. L., and Ettlinger, G. (**1961**). "Cross-modal Transfer in Man," Nature 191, 1225-1226.

Collyer, C. E., and Church, R. M. (**1998**). "Interresponse intervals in continuation tapping," in *Timing of Behaviour: Neural, Psychological and Computational Perspectives*, edited by D. A. Rosenbaum, and C. E. Collyer (MIT Press, Cambridge Massachusetts), pp. 63-87.

Crain, S. (**1991**). "Language acquisition in the absence of experience," Behavioral and Brain Sciences 14, 597-650.

Cross, I. (**2003**). "Music, cognition, culture and evolution," in *The Cognitive Neuroscience of Music*, edited by I. Peretz, and R. J. Zatorre (Oxford University Press, Oxford), pp. 42-56.

Darwin, C. (**1871**). *The Descent of Man and Selection in Relation to Sex* (John Murray, London).

Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," Journal of Phonetics 11, 51-62.

Davenport, R. K., and Rogers, C. M. (**1970**). "Intermodal equivalence of stimuli in apes," Science 168, 279-280.

de Waal, F. B. M. (**1988**). "The communicative repertoire of captive bonobos (Pan paniscus), compared to that of chimpanzees.," Behaviour 106, 183-251.

Dennett, D. C. (**1995**). *Darwin's Dangerous Idea* (Simon & Schuster, New York).

Dissanayake, E. (**2000**). "Antecedents of the temporal arts in early mother-infant interaction," in *The Origins of Music*, edited by N. L. Wallin, B. Merker, and S. Brown (The MIT Press, Cambridge, Mass.), pp. 389-410.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., and Paäbo, S. (**2002**). "Molecular evolution of FOXP2, a gene involved in speech and language," Nature 418, 869-872.

Ettlinger, G., and Blakemore, C. B. (**1969**). "Cross-modal transfer set in the monkey," Neuropsychologia 7, 41-47.

Falk, D. (**2004**). "Prelinguistic evolution in early hominins: Whence motherese?," Behavioral and Brain Sciences 27, 491-450.

Fitch, W. T. (**2006a**). "On the biology and evolution of music," Music Perception 24, 85-88.

Fitch, W. T. (**2006b**). "The biology and evolution of music: A comparative perspective," Cognition 100, 173-215.

Fitch, W. T., Hauser, M. D., and Chomsky, N. (**2005**). "The evolution of the Language Faculty: Clarifications and implications," Cognition 97, 179-210.

Fitch, W. T., Neubauer, J., and Herzel, H. (**2002**). "Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production," Anim. Behav. 63, 407-418.

Fitch, W. T., and Rosenfeld, A. J. (**2007**). "Perception and production of syncopated rhythms," Music Perception 25, 43-58.

Fodor, J. A. (**1983**). *The Modularity of Mind.* (MIT Press, Cambridge, MA).

Frigyesi, J. (**1993**). "Preliminary Thoughts toward the Study of Music without Clear Beat: The Example of "Flowing Rhythm" in Jewish "Nusah"," Asian Music 24, 59-88.

Geiser, E., Zaehle, T., Jancke, L., and Meyer, M. (**2008**). "The Neural Correlate of Speech Rhythm as Evidenced by Metrical Speech Processing," Journal of Cognitive Neuroscience 20, 541-552.

Geissmann, T. (**2000**). "Gibbon song and human music from an evolutionary perspective," in *The Origins of Music*, edited by N. L. Wallin, B. Merker, and S. Brown (The MIT Press, Cambridge, Mass.), pp. 103-123.

Gentner, T. Q., Fenn, K. M., Margoliash, D., and Nusbaum, H. C. (**2006**). "Recursive syntactic pattern learning by songbirds," Nature 440, 1204-1207.

Geschwind, N. (**1970**). "The organization of language and the brain," Science 170, 940-944.

Ghazanfar, A. A., and Logothetis, N. K. (**2003**). "Facial expressions linked to monkey calls," Nature 423, 937-938.

Glass, L., and Mackey, M. C. (**1988**). *From clocks to chaos: the rhythms of life* (Princeton University Press, Princeton).

Goldsmith, J. A. (**1990**). *Autosegmental and metrical phonology* (Blackwell, Oxford).

Gombrich, E. H. (**1984**). *The Sense of Order: a Study in the Psychology of Decorative Art* (Phaidon, London).

Greenfield, M. D. (**1994**). "Cooperation and conflict in the evolution of signal interactions," Annual Review of Ecology and Systematics 25, 97-126.

Hagen, E. H., and Bryant, G. A. (**2003**). "Music and dance as a coalition signaling system," Human Nature 14, 21-51.

Hare, B., and Tomasello, M. (**2004**). "Chimpanzees are more skillful in competitive than cooperative cognitive tasks," Anim. Behav. 68, 571-581.

Hauser, M., Chomsky, N., and Fitch, W. T. (**2002**). "The Language Faculty: What is it, who has it, and how did it evolve?," Science 298, 1569-1579.

Hockett, C. F. (**1960**). "Logical considerations in the study of animal communication," in *Animal Sounds and Communication.*, edited by W. E. Lanyon, and W. N. Tavolga (American Institute of Biological Sciences., Washington, D.C.).

Huber, L. (**2001**). "Visual categorization in pigeons," in *Avian Visual Cognition*, edited by R. Cook (Comparative Cognition Press, Medford, MA).

Jespersen, O. (**1922**). *Language: Its Nature, Development and Origin* (W. W. Norton & Co., New York).

Jones, A. M. (**1954**). "African Rhythm," Africa: Journal of the International African Institute 24, 26-47.

Justus, T. C., and Hutsler, J. J. (**2005**). "Fundamental issues in the evolutionary psychology of music: Assessing innateness and Domain specificity," Music Perception 23, 1-27.

Keller, E. F. (**2005**). "Revisiting "scale-free" networks," BioEssays 27, 1060-1068.

Kirby, S. (**2000**). "Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners," in *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, edited by C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Cambridge University Press, Cambridge), pp. 303-323.

Kirschner, S., Call, J., and Fitch, W. T. (**in prep**). "Drumming Behaviour in African Apes."

Kirschner, S., and Tomasello, M. (**submitted**). "Joint Drumming: Social Context Facilitates Synchronization in Preschool Children."

Lerdahl, F., and Jackendoff, R. (**1983**). *A generative theory of tonal music* (MIT Press, Cambridge, Massachusetts).

Locke, D. (**1987**). *Drum Gahu: The rhythms of West African drumming* (Talman, New York).

London, J. (**2002**). "Some non-isopmorphisms between pitch and time," Journal of Music Theory 46, 127-151.

Longuet-Higgins, H. C., and Lee, C. S. (**1984**). "The rhythmic interpretation of monophonic music," Music Perception 1, 424-441.

Markson, L., and Bloom, P. (**1997**). "Evidence against a dedicated system for word learning in children," Nature 385, 813-815.

Masterton, R. B., and Imig, T. J. (**1984**). "Neural mechanisms for sound localization," Annu Rev Physiol 46, 275-287.

McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines--do mammals fit the model?," Trends in Neuroscience 26, 347-350.

McDermott, J., and Hauser, M. D. (**2005**). "The origins of music: Innateness, uniqueness, and evolution," Music Perception 23, 29-59.

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

Melis, A. P., Hare, B., and Tomasello, M. (**2006**). "Engineering cooperation in chimpanzees: tolerance constraints on cooperation," Anim. Behav. 72, 275-286.

Merker, B. (**1999**). "Synchronous chorusing and the origins of music," Musicae Scientiae Special Issue 1999-2000, 59-74.

Merker, B. (**2000**). "Synchronous chorusing and human origins," in *The Origins of Music*, edited by N. L. Wallin, B. Merker, and S. Brown (The MIT Press, Cambridge, Mass.), pp. 315-327.

Merriam, A. P. (**1959**). "Characteristics of African Music," Journal of the International Folk Music Council 11, 13-19.

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

Morley, I. (**2002**). "Evolution of the physiological and neurological capacities for music," Cambridge Archaeological Journal 12, 195-216.

Patel, A. (**1998**). "Syntactic processing in language and music: Different cognitive operations, similar neural resources?," Musical Perception 16, 27-42.

Patel, A. D. (**2006**). "Musical rhythm, linguistic rhythm, and human evolution," Music Perception 24, 99-104.

Patel, A. D. (**2008**). *Music, Language, and the Brain* (Oxford, New York).

Patel, A. D., Iversen, J. R., Chen, Y., and Repp, B. H. (**2005**). "The influence of metricality and modality on synchronization with a beat," Experimental Brain Research 163, 226–238.

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

Peretz, I., and Morais, J. (**1989**). "Music and modularity," Contemporary Music Review 4, 279-293.

Peretz, I., and Zatorre, R. J. (**2005**). "Brain organization for music processing," Annual Review of Psychology.

Pressing, J. (**1983**). "Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and Western tonality," Studies in Music 17, 38-61.

Pressing, J. (**2002**). "Black Atlantic Rhythm: Its Computational and Transcultural Foundations," Music Perception 19, 285-310.

Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition 73, 265-292.

Roach, P. (**1982**). "On the distinction between 'stress-timed' and 'syllable-timed' languages," in *Linguistic Controversies,* edited by D. Crystal, pp. 73-79.

Roederer, J. G. (**1984**). "The search for a survival value for music," Music Perception 1, 350-356.

Sachs, C. (**1952**). "Rhythm and Tempo: An Introduction," Musical Quarterly 38, 384-398.

Sachs, C. (**1953**). *Rhythm and Tempo: A Study in Music History* (Norton, New York).

Savage-Rumbaugh, E. S., Sevcik, R. A., and Hopkins, W. D. (**1988**). "Symbolic Cross-Modal Transfer in Two Species of Chimpanzees," Child Development 59, 617-625.

Schaller, G. B. (**1963**). *The Mountain Gorilla* (University of Chicago Press, Chicago).

Schloss, W. A. (**1985**). "On the Automatic Transcription of Percussive Music: From Acoustic Signal to High-level Analysis," in *CCRMA* (Stanford University, Stanford, CA), p. 126.

Shipp, S. (**2005**). "The importance of being agranular: a comparative account of visual and motor cortex," Philosophical Transactions of the Royal Society of London 360, 797-814.

Sur, M., Pallas, S. L., and Roe, A. W. (**1990**). "Cross-modal plasticity in cortical development: differentiation and specification of sensory neocortex," Trends Neurosci 13, 227-233.

Tallerman, M. (**2005**). "Did our ancestors speak a holistic protolanguage?," Lingua??, ??

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (**2005**). "Understanding and sharing intentions: The origins of cultural cognition," Behavioral & Brain Sciences 28, 675-735.

Toussaint, G. (**2003**). "Classification and phyogenetic analysis of African ternary rhythm timelines," in *BRIDGES: Mathematical Connections in Art, Music and Science* (University of Granada, Granada, Spain), pp. 25-36.

Trainor, L. J. (**1996**). "Infant Preferences for Infant-Directed Versus Noninfant-Directed Playsongs and Lullabies," Infant Behaviour and Development 19, 83-92.

Trehub, S. E., and Trainor, L. J. (**1998**). "Singing to infants: Lullabies and play songs," Advances in Infant Research 12, 43-77.

von Hornbostel, E. M. (**1928**). "African Negro Music," Africa: Journal of the International African Institute 1, 30-62.

von Melchner, L., Pallas, S. L., and Sur, M. (**2000**). "Visual behaviour mediated by retinal projections directed to the auditory pathway," Nature 404, 871-876.

Washburn, D. K., and Crowe, D. W. (**1998**). *Symmetries of Culture: Theory and Practice of Plane Pattern Analysis* (University of Washington Press, Seattle).

Washburn, D. K., and Crowe, D. W. (**2004**). *Symmetry Comes of Age: The Role of Pattern in Culture* (University of Washington Press, Seattle).

Wray, A. (**1998**). "Protolanguage as a holistic system for social interaction," Language & Communication 18, 47-67.

Wundt, W. M. (**1908**). *Grundzuge der Physiologischen Psychologie*.

Zabana, K. (**1997**). *African Drum Music: Slow Agbekor* (Afram Publications, Accra, Ghana).

Zweifel, P. F. (**1996**). "Generalized Diatonic and Pentatonic Scales: A Group-Theoretic Approach," Perspectives of New Music 34, 140-161.