Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 3

Musicae Scientiae
Spring 2007, Vol XI, n° 1, 3-26

© 2007 by ESCOM European Society for the Cognitive Sciences of Music

Contagious heterophony : A new theory about the origins of music

STEVEN BROWN
Department of Psychology, Simon Fraser University

• **ABSTRACT**

Two of the most salient features of music are the blending of pitch and the matching of time. I propose here a possible evolutionary precursor of human music based on a process I call "contagious heterophony". Heterophony is a form of pitch blending in which individuals generate similar musical lines but in which these lines are poorly synchronized. A wonderful example can be found in the howling of wolves. Each wolf makes a similar call but the resultant chorus is poorly blended in time. The other major feature of the current hypothesis is contagion. Once one animal starts calling, other members of the group join in through a spreading process. While this type of heterophonic calling is well-represented in nature, synchronized polyphony is not. In this article, I discuss evolutionary scenarios by which the human capacity to integrate musical parts in pitch-space and in time may have emerged in music. In doing so, I make mention of neuroimaging findings that shed light on the neural mechanisms of vocal imitation and metric entrainment in humans, two key processes underlying musical integration.

While discussions of the origins of music were commonplace during the 18th and 19th centuries (Condillac, 1746 ; Rousseau, 1781 ; Spencer, 1857, 1890 ; Darwin, 1871, 1872), the topic seemed to fall out of favour during the 20th century. However, an interest in the evolution of music underwent a resurgence at the turn of the 21st century (Wallin, Merker & Brown, 2000 ; Brown, 2000 ; Miller, 2000 ; Huron, 2001 ; Cross, 2001 ; Morley, 2002 ; Hauser & McDermott, 2003 ; Hagen & Bryant, 2003). This article presents a new hypothesis about how music evolved. The first parts discuss a proposal for an evolutionary precursor of both song and speech based on a vocal process that I call "contagious heterophony". The last part describes how this precursor may have given rise to the unique discourse arrangements of human music and speech. Along these lines, I describe the distinction between "integration" and "alternation" in the evolution of musical and linguistic discourse, respectively. Before presenting this hypothesis in detail, however, I will make a brief foray into historical theories of music evolution in order to provide an historical context for the ideas to be elaborated.

"Sound" vs. "Sounding" Theories of Harmony

In thinking about acoustic theories of the origins of music in general — and of harmony in particular — we can distinguish two general classes of models. I will refer to these as the "sound" and "sounding" theories of harmony. The "sound" theory can be traced back to Jean-Philippe Rameau (*e.g.*, 1722) and his classic theory of the *corps sonore*, or resonating body. Rameau put forward a theory that was considered in its day to be no less of a revolution for music theory than were Newton's laws for physics. Rameau's idea was that when objects are set into vibration, such as through the plucking a violin string, not only is there a principal vibration of the object at the frequency that we perceive phenomenally as the pitch of the sound, but there is also a series of vibrations, called harmonics or overtones, that occur simultaneous with the principal vibration. These harmonics occur at frequencies that are integer multiples of the fundamental frequency. According to Rameau's view, then, *harmonics* are the source of harmony. Rameau saw harmony as being an intrinsic feature of sound itself, and therefore the primary feature of music, from which all the other musical features were derived. Psychoacoustic and physiological theories from Helmholtz onward have looked to the harmonic series as a physical foundation for the origins of the musical intervals and the perception of consonance in music (*e.g.*, Tramo, Cariani, Delgutte & Braida, 2001).

There are problems with this view despite its reductionistic appeal. The theory of the *corps sonore* is an acoustic theory that applies to all physical objects with minimal specificity for music. Every object, for example a wine glass, when set into vibration will generate a series of resonant frequencies not terribly different from those of a conventional musical instrument (and in fact the "glass harmonica" is a musical instrument based on such principles). But at a more fundamental level, I suggest that the origins of music are not to be found in sound-*perception* mechanisms but rather in sound-*generation* mechanisms, and it is here where "sounding" theories come into play. The "sounding" theory of harmony can be seen as having its roots in ideas put forth by Jean-Jacques Rousseau (*e.g.*, 1781) on the importance of melody to the evolution of music. For Rousseau, not only is melody the most essential feature of music but so is *the voice* (Thomas, 1995). For Rousseau, there can be no stronger link than that between music and voice, and this makes music into a quintessentially *communicative* activity, unlike the resonantly-beautiful clang of the wine glass with all its accompanying harmonics. For Rousseau, melody in music has its origins not just in sound but in communication, and more specifically in the vocal expression of emotion : "Melody, by imitating the inflections of the voice, expresses complaints, cries of sadness or of joy, threats, and moans ; all the vocal signs of the passions are within its scope" (Rousseau 1781/1998 :322). That said, Rousseau did not address the very important question of why human music is in fact so full of acoustic blending, or harmony, as he was so intent on refuting Rameau's claim that harmony was the fundamental element of music (see Thomas [1995] for a discussion of the vociferous rivalry between Rameau and Rousseau).

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 5

I would like to propose a "sounding" theory of harmony that is true to Rousseau's commitment to vocal expression but which goes beyond his exclusive emphasis on solo vocalization. According to this viewpoint, harmony is derived not from the harmonics of vibrating bodies per se but from *the blending of voices*, in other words choral blend. Another way of saying this is that *harmonization* is the source of harmony. True to Rousseau's original ideas about melody, the "sounding" theory of harmony is a vocal theory that applies exclusively to communication sounds produced by animals. It places an emphasis on production processes rather than perceptual processes alone. Throughout this article, I will use the terms "blend" and "chorus" to refer to the collective vocalizations of group-living or pair-bonded species for cooperative purposes. It is this capacity for intentional, cooperative blending that theories of music evolution have to explain. Hence, choral blend in my sense does not refer to a situation of mere overlap or simultaneity but to one in which animals in group-living or pair-bonded species vocalize together with common communicative intent, generally for territorial purposes. By this definition, the "dawn chorus" of birds — in which birds from multiple species sing simultaneously and in which multiple males from a single species compete with one another to be heard by potential mates — does not qualify as an example of choral blend but merely one of simultaneity. Likewise, a group of several hundred people on the floor of the New York Stock Exchange screaming out stock bids at the same time does not qualify as a reasonable example of a chorus in humans. Indiscriminate use of the "chorus" concept in ethology has created many problems for those interested in music origins. In this article, I reserve use of the words "chorus" and "blend" to refer to the collective, cooperative, group-level blending processes that occur in human choruses. Simultaneity, and even synchrony, is not sufficient.

One of the strongest arguments for giving priority to production over perception in the evolution of human musical communication is the observation that the auditory organs/pathways of humans and chimpanzees show strong similarity but that the vocal pathways show profound differences, permitting, on the one hand, a species capable of song and articulate speech, and on the other, a species incapable of either one. While peripheral changes such as descent of the larynx have been frequently cited as critical factors in this evolution (Lieberman, Klatt & Wilson, 1969), nervous-system changes are now recognized as being equally significant (Fitch, 2000). For example, the primary motor cortex is a critical mediator of voluntary vocalization in humans. Bilateral lesions of the primary motor cortex or the associated corticobulbar pathway lead to debilitating motor speech disorders, known collectively as pseudobulbar palsy or spastic dysarthria (Murphy, 2005). In addition, transient lesions of the motor cortex by means of transcranial magnetic stimulation can lead to severe speech disruption and even total speech arrest (Epstein, Meador, Loring, Wright, Weissmann, Sheppard, Lah, Puhalovich, Gaitan & Levy, 1999; Stewart, Walsh, Frith & Rothwell, 2001). For non-human primates, there is no evidence that the primary motor cortex is involved in vocalization at all, as lesions here

do not affect the production of calls (Kirzinger & Jürgens, 1982). Moreover, while humans are thought to have a direct neural projection from the primary motor cortex to the nucleus ambiguus — the major brainstem centre for the control of phonation — nonhuman primates lack a direct connection (Jürgens, 1992; Simonyan & Jürgens, 2003). Perhaps most importantly, a major fiber pathway in humans connects the auditory association cortex of the superior temporal lobe with the vocal-planning centres of the inferior frontal lobe (Catani, Howard, Pajevic & Jones, 2002), thereby creating the neural underpinnings for vocal imitation. The details of this connectivity in the monkey are still being debated, as it is unclear if the ventral prefrontal areas receiving auditory input are at all vocal areas (but see Petrides and Pandya, 2001). The main behavioural outcome of these species differences is that no primate except the human is capable of phonatory vocal learning (Egnor & Hauser, 2004) such as that underlying the acquisition of song and speech through vocal imitation. Interestingly, the observation of "call convergence" in male chimpanzees (Mitani & Gros-Louis, 1998) — in which the chorus calls of animals having the strongest social bonds converge in acoustic structure — reveals a kind of adult-stage vocal plasticity that might be a phylogenetic precursor to vocal imitation in humans.

One offshoot of the sounding theory of harmony is the idea that the capacity to perceive and produce harmony is essentially contained within a melodic system rather than within a specific and dissociable harmony module. I have proposed that the human capacities for melody and harmony co-evolved as a single neural system (see Brown, Parsons, Martinez, Hodges & Fox, 2004). This view tends to lean on the side of Rousseau in saying that harmony is derived from (or contained within) a basic melodic system, and a vocal one at that. This is attested to by the fact that the principal form of pitch blending in world musics is not homophony (*i.e.*, the musical texture in which all voices move more or less together, as in Western chordal music), but instead multi-layered polyphony, in which short melodic motifs are woven together in an interlocking fashion. Perhaps the best-described examples of this are the complex polyphonic choruses and instrumental ensembles of the Pygmies of central Africa (Arom, 1991). Thus, it is not unreasonable to speculate that the human harmony system emerged from a basic melodic system in which individual parts were temporally blended with one another following developments in temporal processing (see the discussion of "vertical integration" below). If this were indeed the case, it would reverse the traditional evolutionary progression proposed by Curt Sachs (1943), and argue that polyphony (or at least heterophony) preceded monophony (*i.e.*, the musical texture in which all voices sing the same part) in human history.

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 7

CONTAGIOUS HETEROPHONY

With this commitment to a "sounding" theory of music evolution in mind, I would like to propose a possible evolutionary precursor of human music and speech, one bearing significant similarities to the calling behaviour of a number of contemporary animal species. This precursor is a vocal-blending system based on what I call "contagious heterophony". I define this term as follows : a group vocalization in which each individual produces a variation on a similar kind of call but in which the members of the group call asynchronously ; group-wide vocalizing emerges through a sequential process of spreading or contagion. An example of this is found in the howling of wolves. Each wolf makes a similar call but the resultant chorus is poorly blended in time, thereby creating a type of heterophony. We can identify many other examples of such chorusing in nature, including the group roar of howler monkeys, the pant hoot chorus of chimpanzees, the roar of lions, or the heterophonic singing style of humpback whales. The other major feature of the current hypothesis is contagion. Interestingly, contagion is a prominent feature of all of the examples of heterophonic vocalizing just mentioned. Once one animal starts calling, other members of the group join in through a spreading process. Harrington (1989) has analyzed the process of chorus howling by packs of wolves in the wild. That wolf howling is a contagious process is suggested by the finding that, while chorusing often begins with a solo howl, entry of other wolves into the chorus "is accelerated rather than paced at a relatively constant rate" (Harrington, 1989 :124). This is demonstrated in Figure 1a, which shows that the time interval of entry of wolves into the chorus gets successively shorter. The end result of this successive addition of parts to the chorus is an increase of texture and a filling of acoustic space, as shown clearly in the continuous sonogram in Figure 1b. As Harrington (1989) has shown, howls become shorter, higher, and more frequency-modulated as the chorus progresses. There is a great increase in spectral variation as chorus proceeds, making the wolf chorus into a kind of jam session.

The concept of contagious heterophony can be seen as providing important insights into the nature of music. First, a system based on contagious heterophony places music firmly within the domain of group vocalization from its origin. This jibes well with more than a century's worth of ethnomusicological research showing that music is, in its most basic sense, a means of group communication (Merriam, 1966, Dissanayake, 2000), one that is useful in solidifying social groups (Brown, 2000 ; Hagen & Bryant, 2003). Likewise, most of the animal forms of contagious heterophony mentioned above are thought to play important roles in group function, including territory maintenance, group identity, and social cohesion. This is especially well-established for duetting species, where chorusing provides an important means of marking territories acoustically and maintaining pair bonds (*e.g.*, gibbons [Geissmann, 2000], indris [Pollock, 1986], Australian magpies [Brown & Farabaugh, 1991]). Second, we can see the origins of basic tonal systems in these

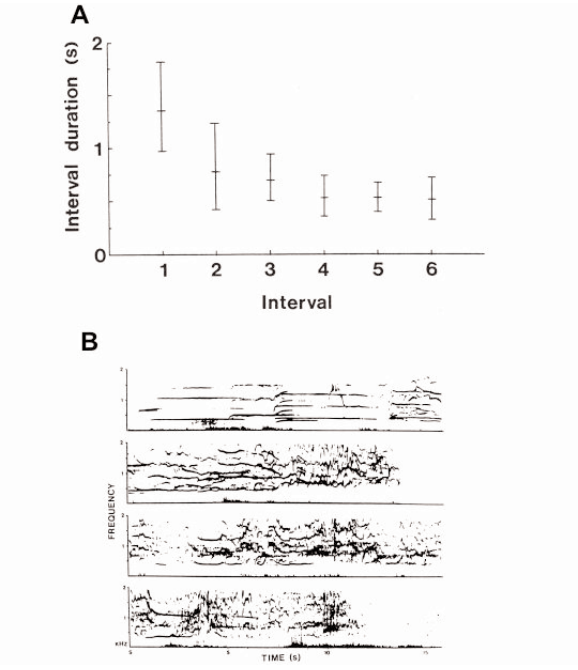

Figure 1.

Wolf chorusing. ***A.*** *The time of entry of successive individual wolves at the beginning of each chorus is shown. Interval duration is the time in seconds between the beginning of successive howls in a chorus. Interval 1 is the period between the first and second howls in a chorus, interval 2 is the period between the second and third howls, and so on.* ***B.*** *Continuous sonogram of a wolf chorus, which begins in the upper frame and runs down, left to right, through each successive frame. Notice the increase in frequency modulation as the chorus progresses. Reprinted from Harrington (1989) with permission of the author.*

calls. Wolf howls generally begin with a simple ascending interval towards a sustained note — often a fifth or sixth — followed by several chromatically-descending portamentos. While lacking the melodic complexity of whale song or some elaborate forms of birdsong, such calls do provide elements of simple tonality. Third, we can see the rudiments of true, collective blending in such a system, leading ultimately to what might become the capacity for homophony in modern humans. At the times of accelerated entry of multiple wolves into a group chorus and at the point where multiple animals are sustaining howl-notes simultaneously, this overlapping of parts provides what is perhaps the best analogue of a musical chord in animals. It is interesting to consider the territorial consequences of such cooperative and complex blending. Harrington (1989) has pointed to a kind of "Beau Geste effect" of wolf chorusing in which complex blending can lead to a strong overestimation of group size (at least by human listeners), thereby enhancing the territorial function of howl-

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 9

ing. Fourth, the fact that all members of a group performing these contagiously-heterophonic vocalizations create quite similar calls means that acoustic complexity can evolve out of variations in blending rather than variations in individual calling alone. In other words, small individual repertoires need not impose a constraint on creating group-level complexity. My suspicion is that this point applies equally well to many forms of human music-making. Finally, the role of contagion in such a system provides not only a mechanism by which group participation may come about but also an important clue as to how this kind of system might operate both psychologically and neurally. Contagion is a kind of resonance behaviour in which a common state of group activity is achieved through matching or imitation. Think about when people stand up during the ovation at a concert. A standing ovation is often initiated with just one person getting up. Soon thereafter, everyone in the house is on their feet through a kind of contagious process of imitation. The end result is that all the people in the room have made a change in their state (from sitting to standing), and have achieved a new state of resonance. More to the point of vocalization, the declamatory recitation of slogans at political rallies or the singing of songs at sporting events often occurs in a contagious fashion, with a few initiators spurring on a large number of followers. Along similar lines, the singing of "Happy Birthday" at a party is often initiated by a single leader and key-setter, in response to whom people spontaneously converge on a given set of pitches. Those who cannot match these pitches with their voice are said to be tone deaf.

To sum up, I argue that a vocal system based on contagious heterophony provides a reasonable precursor for the systems that will become both song and speech in modern humans. Interestingly, Geissmann (2002) argued that duetting is an ancestral trait for gibbons and that solo singing evolved secondarily through an abandonment of duetting in those species that have exclusively solo calling. In like form, I suggest that group calling may have been an ancestral trait of humans. But how did music evolve from the kind of heterophonic precursor that I have proposed ? In other words, how do we go from a system of asynchronous blending (such as wolf howling) to a system of tightly-coordinated blending such as that which occurs, for example, in musical polyphony ? What I would like to propose is that the route from contagious heterophony to human communication followed at least two complementary paths, as shown in Figure 2 : 1) vertical *integration*, in which well-blended, temporally-aligned performance styles arise (*i.e.*, monophony, homophony, polyphony), and 2) horizontal *alternation*, in which well-timed but alternating vocalization processes arise (typical of human speech). A useful example to consider is monophonic singing, such as the singing of "Happy Birthday" mentioned in the last paragraph. At least two skills are necessary for monophonic singing to occur in a group of individuals : first, an ability to match pitch ; second, an ability to match event-onsets in time. This highlights the related concepts of *vocal imitation* and *metric entrainment*, which will be addressed in sequence in the following two sections.

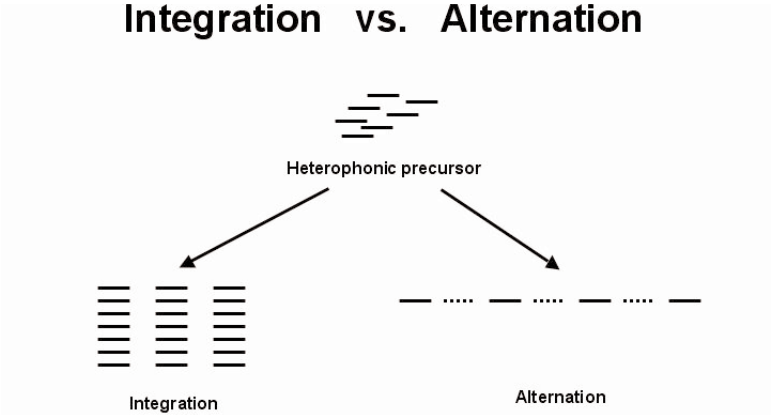

Figure 2.

The distinction between integration and alternation. An evolutionary progression is shown in which a vocal precursor based contagious heterophony emerges into an integrated form of vocalization (such as is found in much music) or an alternating form of vocalization (as is found in most forms of speech). For integration, three sets of integrated parts are shown in sequence in order to demonstrate rhythmic organization in time.

Vocal Imitation : Pitch Matching

Although most vertebrates have the capacity to vocalize, very few species have the ability to imitate sounds vocally. Among the principal exceptions are songbirds, humpback whales, and human beings (Janik & Slater, 1997, 2000). Vocal imitation in humans is important not only for the establishment of large and flexible acoustic repertoires in speech (*e.g.*, phonemes, syllables) and music (*e.g.*, pitch sets) but also for the ability to precisely mimic linguistic, musical or environmental sounds in real time, as is important in the learning of novel melodies. It is now well-established that vocal imitation during human development provides a foundation for the acquisition of communication sounds, such as speech phonemes, intonational melodies, and musical pitches (Poulson, Kymissis, Reeve, Andreators & Reeve, 1991 ; Kuhl & Meltzoff, 1996 ; Papousek, 1996 ; Studdert-Kennedy, 2000 ; Trehub, 2001). In other words, vocal learning in humans is imitative, dependent upon interactions with adult role models during critical periods in neurocognitive development. This process involves the ontogenesis of neural pathways that can effectively translate heard communication sounds into motor plans for vocalization.

While the development of these neural pathways in humans is poorly understood, a large database has amassed in the understanding of analogous processes in the songbird brain. This is not the appropriate place to review the avian neurobiology literature, however the major point to get across is the distinction between, on the one hand, a neuromotor system that is necessary for vocalization itself and, on

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 11

the other, a separate system that does not directly control vocalization but which is necessary for vocal imitation. These two systems — which in the songbird are referred to as the posterior and anterior forebrain pathways, respectively — are interconnected with one another but can be functionally dissociated by means of lesions. While lesions to the posterior forebrain pathway subsequent to song development can cause generalized muteness (Simpson & Vicario, 1990), lesions to the anterior forebrain pathway during critical periods in development can effectively knock out the ability of a songbird to acquire its species-specific song patterns without affecting its general ability to vocalize (or hear, for that matter) (Scharff & Nottebohm, 1991 ; Brainard, 2004). Jarvis (2004) pointed out that this two-pathway system in the bird has striking analogies to a similar arrangement in the human brain, where the primary motor cortex is akin to the posterior forebrain pathway involved in direct vocal control, and the premotor cortex (such as Broca's area) is akin to the anterior forebrain pathway involved in vocal imitation. Hence, in both humans and songbirds, vocal imitation may be mediated by a part of the brain that is distinct from that involved in direct vocal control. One prediction is that such an area, unlike the direct motor pathway, should receive inputs from auditory areas, as seems to be well-established for Broca's area in humans.

Beyond this requirement for vocal imitation during song and speech development is the retention of this capacity throughout the life span. Humans are consummate vocal imitators (Skoyles, 1998). Some people even make a living by vocally shadowing people's speech as a form of entertainment. In experimental settings, people can shadow continuous prose with a mere 70 millisecond (msec) delay (Bailly, 2002 ; see also Marslen-Wilson, 1973 and Porter & Lubker, 1980 for reports of longer latencies). Such a latency is significantly faster than any kind of reaction time seen with manual responses, thereby suggesting a highly direct connection between audition and vocal production. The human ability to vocally imitate can sometimes become overactive in certain clinical conditions. One such condition is echolalia, which can be defined as the "unsolicited repetition of another's utterances" (Murphy, 2005 :359). It is a prominent feature of autism and Tourette's syndrome, as well as several neurodegenerative conditions (Schuler, 1979). Spontaneous echolalic imitation can occur with a delay of only 270 msec from the offset of a stimulus (Fay and Coleman, 1977).

Imitation also plays an important role in traditional forms of musical performance throughout the world. Two well-described examples of this are the "lift-up-over sounding" style of out-of-phase singing (called *dulugu ganalan*) of the Kaluli people of the southern highlands of Papua New Guinea (Feld, 1994), and the vocal "throat games" (called *katajjaq*) of the Inuit people of Canada wherein one person imitates the sounds of another while standing nearly mouth to mouth with that person (Nattiez, 1999). In jazz as well as pop music, there are traditions of on-stage "imitation rituals" in which musicians do alternating improvisations *à deux* (on different instruments with different timbres), with one player improvising a line and the other

repeating it (identically, or with somewhat different phrasing, such as with grace notes or with a coda attached). I would also mention anecdotally that vocal shadowing plays a large role in choral singing, as there is a strong tendency to listen to the people of your voice (baritones in my case) singing next to you. This can be beneficial if you are uncertain about the notes, but it can also be detrimental if someone of your voice makes errors, as there is a tendency to fall prey to these errors. Finally, tone deafness represents an impaired ability to match pitch with the voice. Like many other developmental and acquired disorders of the voice, tone deafness is much more prevalent in males than in females (Howard and Angus, 1998). There are suggestions that tone deafness may be congenital (Ayotte, Peretz & Hyde, 2002; Peretz, Ayotte, Zatorre, Mehler, Ahad, Penhune & Jutras, 2002) and may thus have a genetic basis (Kalmus & Fry, 1980; Drayna et al., 2001). However, alternate interpretations are equally plausible, including the necessity for musical exposure and vocal practice during a critical period in childhood brain development. Vocal pitch matching, like absolute pitch perception, might be a use-it-or-lose-it skill. While I know of no published ethnographic analyses of tone deafness, my suspicion is that this trait will be found most prominently in people who were not sung to and who were not encouraged to sing as children.

A system of "mirror neurons" in the brain is thought to provide one explanation for the existence of imitative behaviours such as pitch matching. Mirror neurons are a class of cells, first discovered in the hand-controlling region of the premotor cortex of the Rhesus monkey, that fire not only during object-directed hand movements per se but during *observation* of the same type of hand movements by another organism. A mirror system is essentially an observation/execution matching system (Rizzolatti, Fadiga, Gallese & Fogassi, 1996; Rizzolatti, Fadiga, Fogassi & Gallese, 1999). In the writings of Rizzolatti and his colleagues, the focus of such a system has generally been on visuomanual matching. However, I believe that a strong case can be made for the idea that such a system would be the ideal foundation for *audiovocal* matching functions such as song and speech. An interesting feature of the mirror-neuron theory of resonance behaviours is the role assigned to Broca's area in the resonance system. Broca's area is a part of the inferior frontal lobe of the brain that for more than a century has been thought to play a critical role in speech production. Rizzolatti & Arbib (1998) pointed out that the region of monkey brain first shown to contain mirror neurons is the homologue of Broca's area in the human. They thus argued that Broca's area in the human evolved from a premotor system involved in gestural communication. In support of this, Iacoboni, Woods, Brass, Bekkering, Mazziotta & Rizzolatti (1999), in a brain imaging study, showed that the frontal operculum (*i.e.*, the lower part of Broca's area) was active when subjects had to observe and then imitate finger movements.

My colleagues and I have carried out brain imaging experiments that provide insight into the role of this brain area for imitative song production (Brown et al., 2004). Among the vocal tasks that subjects performed in this study was a melody

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 13

repetition task in which they listened to a series of unfamiliar, tonal, monophonic melodies, and had to sing back each one after it was played. Each melody was six seconds in duration, followed by a six-second period for response generation. The results of this study clearly showed that the frontal operculum was activated bilaterally for this task. (Many other brain areas were activated as well, but they are more related to fundamental auditory and vocalization processes). It is tempting to speculate that this part of Broca's area is specialized for *template-matching* processes that underlie pitch imitation. And in fact, another feature that makes the notion of template matching important from a neurobiological standpoint is the parallel activation of this vocal-planning area during perceptual tasks in which there is no vocalization. Template matching is clearly an essential process in performing discrimination tasks, and a large number of studies have demonstrated that the frontal operculum is active during perceptual discrimination tasks for music, including pitch discrimination (Zatorre, Evans & Meyer, 1994; Griffiths, Johnsrude, Dean & Green, 1999; Zatorre & Binder, 2000), chord discrimination (Maess, Koelsch, Gunter, & Friederici, 2001), duration discrimination (Griffiths et al., 1999), time-interval discrimination (Rao, Mayer & Harrington, 2001), discrimination for chords, keys and timbres (Koelsch, Gunter, Cramon, Zysset, Lohmann & Friederici, 2002), and discrimination of errors in melody and harmony during score reading (Parsons, 2001). Overall, I propose that a mirror function for Broca's area can provide as much enlightenment for imitative audiovocal processes underlying music and speech as it does for visuomanual matching processes underlying a proposed gestural origin for language (Rizzolatti & Arbib, 1998). This might suggest that the song system of the human brain evolved from a vocalization system based on imitation, perhaps one derived from a heterophonic precursor.

One caveat is in order. Mirror neurons were first invoked to explain "action understanding" (rather than direct imitation) by arguing that an understanding of the behaviour of others involves a kind of motor simulation of the same action performed by the self. Behavioural imitation, then, would flow out of such a system for action understanding. Whether this theory turns out to be the correct explanation for imitation or not (Jacob & Jeannerod, 2005), and whether vocal imitation in humans involves mirror neurons or not, we do have to recognize that 1) there must exist some system in the human brain that functions to translate heard sounds into motor targets for phonation and articulation, 2) it operates with incredible speed, and 3) this system is a species-specific trait of humans among primates. Hence, the proposed precursor of music and speech based on contagious heterophony should be thought of as a stereotyped system with a genetically-restricted repertoire of sounds and minimal capacity for direct pitch matching. Humans, then, evolved a capacity for phonatory and articulatory imitation that not only permitted the generation of a rich repertoire of sounds but a tremendous capacity for vocal mimicking throughout the life span. I would argue that the inferior part of Broca's area (the frontal operculum) is one critical component of the system involved in pitch matching during

vocal imitation, in other words a system that receives information about perceived pitches and generates phonatory targets based on them.

Finally, it is useful to keep in mind a distinction between sequentiality and simultaneity when talking about imitation. In a conventional sense, imitation is about mimicking the action of someone else *after* they have performed an activity, in other words about sequentiality. However, for activities whose motor patterns are established in long-term memory for all participants, such as the singing of "Happy Birthday", imitation may simply involve the convergence of all participants towards the particular pattern initiated by one individual. So while this may be a situation of simultaneity, imitation may be important during the early stage when a pattern, such as a musical key, is being established. Likewise for the shouting of slogans or the singing of songs at a sporting event. Imitation might apply mainly to the triggering and convergence processes.

Temporal Coordination : Integration and Alternation

While vocal imitation during childhood development is clearly important for acoustic imprinting in both music and speech, vocal imitation during adulthood seems to be more relevant to music than to speech. That is because music often involves an *integration* of parts (such as during monophonic singing) whereas speech generally involves an *alternation* of parts. One of the main proposals of this article is that both types of communication arrangement evolved from a common heterophonic vocalization precursor, one having an imprecise arrangement of parts (Figure 2). The distinctive discourse arrangements of music and speech differ principally in their mechanism of temporal coordination. Simply put, a system of integration requires timing of event *onsets*, where a system of alternation requires timing of event *offsets*. These will be discussed in sequence.

• **Vertical Integration.** The basic problem we encounter in moving from a stage of contagious heterophony to a stage of highly-synchronized monophony, homophony or polyphony is to align beats temporally to create a synchronized output. As mentioned before, the overlap of voices found in heterophonic vocalizations like wolf howling provides features of a blending system but lacks a synchronization of parts, the kind that is found in human chorusing. But, in the same way that monophonic singing of "Happy Birthday" requires a key-setter for pitch matching, it also requires a timekeeper to generate both a rhythm and a tempo. As in the discussion of pitch imitation above, we can think about temporal matching as occurring by means of simultaneity or sequentiality. For example, the learning of novel drum patterns or dance rhythms would be unthinkable without a general capacity for motor imitation. This is the domain of sequentiality. However, entrainment with an existing timekeeper — such as synchronizing one's movements to a drum beat — is the

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 15

domain of simultaneity. Whether entrainment to a beat requires a capacity for imitation in humans is an open question. The one thing that is certain is that this capacity is rare in nature. Among mammals, it appears to be unique to humans, a capacity that no doubt evolved jointly for dance and music.

My colleagues and I have recently investigated the brain areas involved in metric entrainment in the context of dance movement (Brown, Martinez & Parsons, 2006). In this study, subjects moved their feet in a patterned manner along a flat, inclined surface while their brains were being scanned using positron emission tomography. Two of the experimental conditions will be mentioned here, one in which subjects entrained their foot movements to the beat of music, and one in which they moved their feet in a self-paced manner (with no music) at essentially the same tempo. While both tasks engaged nearly identical parts of the motor system for the control of the lower extremity, the entrainment task led to a large increase in activity in the midline of the cerebellum (vermis III), an area that was activated to a lesser extent in the self-paced condition. Importantly, this large increase in cerebellar activity was not due to the presence of music in the entrainment condition, as subtraction of all the brain activity due to music listening alone (as measured in a separate brain scan) did not reduce the intensity of the activation. Hence, the difference was a relatively pure reflection of audiomotor entrainment itself. Whether the vermis of the cerebellum is involved in vocal entrainment as well as body entrainment is not known. In our aforementioned study of vocalization (Brown et al., 2004), there was low-level activation in this region when subjects had to harmonize in synchrony with another musical part (a form of vocal entrainment) but not when subjects had to repeat back melodies (a form of vocal alternation and imitation). At a conceptual level, we would expect a neural timekeeper to be a superordinate entity that works upon multiple types of sensory signals and controls multiple motor effectors, as humans can synchronize their movements to sounds, flashes of light, taps on the skin, etc., and can entrain movement of almost any body part to these signals. The cerebellar vermis is at least one candidate for a brain area mediating metric entrainment in humans.

It is worth mentioning that synchronous-like calling occurs in certain insect and frog species (Greenfield, 1994 ; Pollack, 2000), and hence that the neural requirements for integration might be quite minimal. Given this, it seems surprising that integrated calling has not evolved more frequently in the animal world, and that it appears absent even in highly complex vocalisers like songbirds and humpback whales. What this suggests is a dissociation between phonatory imitation and entrainment : 1) phonatory imitation occurs in songbirds and humpback whales in the absence of entrainment ; 2) entrainment occurs in insects and frogs in the absence of phonatory imitation ; and 3) humans appear to be unique in that they combine phonatory imitation with metric entrainment in their vocalizations. While pitch imitation and entrainment — processes that involve sequentiality and simultaneity, respectively — may require different cognitive processes, it is also possible that they rely jointly on human imitative capacity. Again, it seems quite clear that

mimicking temporal patterns in sequence requires a capacity for imitation. Whether synchronization to a beat in real time does as well is something in need of further investigation.

Summing up the ideas about music from the last two sections, we can see that a shift from a precursor system based on heterophonic vocalizing to the modern system of musical integration requires the evolution of at least two key capacities : 1) pitch matching, as mediated in part by the frontal operculum ; and 2) entrainment, as mediated in part by the midline of the cerebellum. The frontal operculum may play a role in rhythm processing in addition to pitch imitation, as demonstrated by the previously-mentioned studies showing activation here during discrimination tasks for rhythm and duration. In addition, the cerebellum may play a role in imitation in addition to synchronization, as Penhune, Zatorre & Evans (1997) showed activation in the cerebellum, including the vermis, when subjects had to mimic acoustic rhythmic patterns by means of finger tapping.

Beyond this discussion of fundamental cognitive capacities, I think that music has taken a highly conservative approach to the process of phrase generation. Compared to speech, music has tended to become a simple system based on extensive repetition. Lomax (1968) noted long ago that song was far and away the most redundant of human vocalization systems (see also Richman, 2000). Repetition is certainly one of the most salient features of musical form cross-culturally. Ostinato-based systems such as African polyphony (Arom, 1991) exemplify this process quite nicely. The combination of metric entrainment and phrasal redundancy makes music into a *cyclic* system, in which repetition is found not only in the use of isochronous time intervals but in the redundancy of musical phrases as well. These two processes are critical in establishing integration for music. At the same time, they probably limit the overall information content of music, as the system is oriented toward repetition rather than novelty. From a social standpoint, vertical integration of parts is an ideal device for promoting group-level communication, as the goal of such a process is the creation of a common and coordinated acoustic output. This relatively egalitarian means of musical production permits all participants to have equivalent roles, thereby fostering role symmetry. This is optimal for mediating group cohesion, cooperation and catharsis (Brown, 2000).

• **Horizontal Alternation.** While vertical integration is the quintessential group-musical device, alternation is the dominant mode of speech discourse, one that is found in many forms of choral music as well, mainly antiphonal and responsorial forms. The principal form of vertically-integrated speech that I know of consists of the chanting of short slogans at political rallies. How do we move from a system of strong simultaneity in the heterophonic precursor to one of true turn-taking, such as that which underlies human conversation ? The basic problem that the heterophonic precursor presents is a sense of semantic confusion due to an overlapping of parts, such as when many people speak at the same time. The basic goal of an alternation

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 17

system, therefore, is to effect a *segregation of parts* so that overlap is prohibited and alternation becomes the rule for communication. While alternation can occur in many different manners, the optimal means is to have the individual parts abut one another through a process that could be called "stringing". In stringing, not only are the parts sequential and segregated but one part starts just as another one ends so that there is a continuous flow of interaction with minimal gaps or overlap ; in other words, coordination is with regard to event offsets. Note that, compared to the cyclic and redundant nature of the integration system, an alternation system is highly linear and generative. Such a system is ideally suited for exchanging information in a way that is inconceivable for a cyclic system based on repetition. As such, it relies strongly on role *asymmetry*, and hence a differentiation of roles, such as that found in the leader/chorus distinction in responsorial musical forms. Again, this diversity of parts is optimal for the exchange of information. Alternation tends to be favoured by dyads, which is the typical social arrangement of human speech. Unfortunately, I do not know of any human neuroimaging studies that have explicitly examined the timing arrangement underlying alternation in either speech or music. Most people would argue that alternation is regulated more by *semantic* considerations than acoustic cues alone. Researchers who perform discourse analysis tend to focus on semantic cues that lead to a trade-off in speaking roles rather than to acoustic signalling cues alone (Yule, 1996). This is unquestionably an area in need of further research. It is interesting to point out that many forms of duetting in animals involve basic processes of alternation (*e.g.*, Australian magpies, Brown & Farabaugh, 1991). Geissman's (2002) proposal — following Wickler and Seibt (1982) — that alternation in gibbon duetting may have arisen through a process called "song splitting", in which a song common to both sexes became divided up between two mated partners, is one reasonable mechanism by which alternation might have evolved in humans, a mechanism completely compatible with my heterophonic precursor model.

Figure 3 presents a proposal of the three major routes that were taken from the precursor system based on contagious heterophony to the contemporary forms of human vocal communication, pointing out that the integration system (labelled 1) is one of pitch blending, metric entrainment, redundancy, groupishness, social cohesion and role symmetry, while the alternation system (labelled 3) is one of segregation, stringing, generativity, dyads, information exchange, and role asymmetry. By positing that musical discourse and linguistic discourse evolved from a joint heterophonic system, I am arguing that integration and alternation are variations on a common theme. And in fact, these two forms of discourse co-exist in many musicolinguistic contexts throughout the world, as shown by an important third route sitting in between the two extremes of integration and alteration : namely, *alternation with integration* (labelled 2 in the figure). Alternation with integration is essentially the leader/chorus arrangement found in various forms of choral music throughout the world. The figure points out that in such a discourse arrangement, the chorus

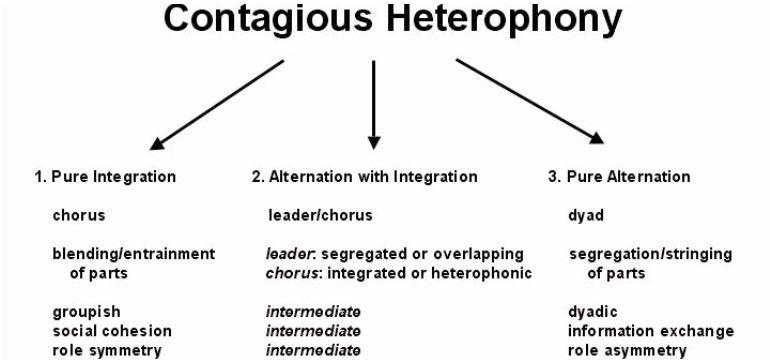

Figure 3.

A comparison between three forms of discourse arrangement in human communication: integration, alternation, and alternation with integration.

can be either highly integrated (as in pure integration) or highly heterophonic (as in the heterophonic precursor). In addition, the leader's interaction with the chorus can either be highly segregated (as in pure alternation) or highly overlapping (as in the heterophonic precursor). It is important to note that when the chorus is highly integrated and the leader is highly segregated from the chorus, then the leader/chorus arrangement becomes that of a "pseudo-dyad", in other words, a group approximation of two individuals having a conversation. From a social standpoint, alternation with integration is intermediate between pure integration and pure alternation in terms of the groupish/dyadic distinction and role symmetry/asymmetry spectrum (see Figure 3). For example, at one extreme, we have choruses that are highly integrated and which recite the exact same refrain in response to each of the leader's calls, which themselves are variable for each verse. So while the leader is providing new information with each verse, the (integrated) chorus is not.

Conclusions

I have presented ideas in this essay based on the importance of "sounding" to the origins of music. To talk about harmony is to talk about the interpersonal processes of musical performance through which individuals volitionally and cooperatively blend musical lines in pitch-space and time. I have presented a proposal for a vocal precursor of human song and speech based on contagious heterophony, and suggested that an audiovocal matching system located in Broca's area may provide the neural underpinnings for pitch-imitation processes and that the cerebellar vermis may play a role in metric entrainment for the integration of parts. While the rudiments of tonal systems may be found in contagious-heterophonic vocalizations such as wolf

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 19

howling, what is more important to me is the group arrangement of such vocalizations rather than their musical properties. I believe that even the most unpitched, broadband roar of a howler monkey chorus provides more insight into the origins of music than does the most melodious birdsong solo. Finally, I discussed the possible emergence of discourse in music and speech through different types of processes, leading to a focus on either integration (mainly music), alternation (mainly speech) or alternation with integration (mainly speech-based music). Overall, I see harmony as being a direct acoustic offshoot of the fact that we are a cooperative, group-living species. What kind of communication process could possibly be more central to the collectivity of human culture than the coordinated blending of voices that underlies human chorusing?

ACKNOWLEDGEMENT

I am grateful to Dr. Peter Pfordresher (University at Buffalo), three anonymous reviewers, and one non-anonymous reviewer for their very helpful suggestions for revising the manuscript.

Address for correspondence:
Steven Brown, Ph.D.
Department of Psychology
Simon Fraser University
Robert C. Brown Hall
8888 University Drive
Burnaby, BC
Canada V5A 1S6
Phone: 1 (604) 291 3013
Fax: 1 (604) 291 3427
e-mail: stebro@sfu.ca

• REFERENCES

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

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia : A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238-51.

Bailly, G. (2002). Close shadowing natural versus synthetic speech. *International Journal of Speech Technology*, *6*, 11-9.

Brainard, M. S. (2004). Contributions of the anterior forebrain pathway to vocal plasticity. *Annals of the New York Academy of Sciences*, *1016*, 377-94.

Brown, E. D., & Farabaugh, S. M. (1991). Song sharing in a group-living songbird, the Australian magpie, *Gymnorhina tibicen*. 3. Sex specificity and individual specificity of vocal parts in communal chorus and duet songs. *Behaviour*, *118*, 244-74.

Brown, S. (2000). Evolutionary models of music : From sexual selection to group selection. In F. Tonneau & N. S. Thompson (Eds.), *Perspectives in ethology. 13 : Behavior, evolution and culture* (pp. 231-81). New York : Plenum Publishers.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). The neural basis of human dance. *Cerebral Cortex*, *16*, 1157-67.

Brown, S., Parsons, L. M., Martinez, M. J., Hodges, D. A., & Fox, P. T. (2004). The song system of the human brain. *Cognitive Brain Research*, *20*, 363-75.

Catani, M., Howard, R. J., Pajevic, S., & Jones D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage, 17*, 77-94.

Condillac, E. (1746/1971). *An essay on the origin of human knowledge*. English translation by R. G. Weyant (1756), reprinted in facsimile form (1971). Gainesville : Scholars' Facsimiles & Reprints.

Cross, I. (2001). Music, cognition, culture, and evolution. In R. J. Zatorre ; & I. Peretz (Eds.), *The biological foundations of music* (pp. 28-42). New York : New York Academy of Sciences.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London : J. Murray.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London : J. Murray.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In. N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389-410). Cambridge, MA : MIT Press.

Drayna, D., Manichaikul, A., de Lange, M. Snieder, H.,& Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, *291*, 1969-72.

Egnor, S. E., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, *27*, 649-54.

Epstein, C. M., Meador, K. J., Loring, D. W., Wright, R. J., Weissmann, J. D., Sheppard, S., Lah, J. J., Puhalovich, F., Gaitan, L., & Davey, K. R. (1999). Localization and characterization of speech arrest during transcranial magnetic stimulation. *Clinical Neurophysiology*, *110*, 1073-79.

Fay, W. H., & Coleman, R. O. (1977). A human sound transducer/reproducer : Temporal capabilities of a profoundly echolalic child. *Brain and Language*, *4*, 396-402.

Feld, S. (1994). Aesthetics as iconicity of style (uptown title) ; or, (downtown title) "Lift-up-over Sounding" : Getting into the Kaluli groove. In C. Keil, & S. Feld, *Music Grooves : Essays and Dialogues* (pp. 109-50). Chicago : University of Chicago Press.

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 21

Fitch, W. T. (2000). The evolution of speech : A comparative review. *Trends in Cognitive Sciences, 4*, 258-67.

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In. N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 103-23). Cambridge, MA : MIT Press.

Geissmann, T. (2002). Duet-splitting and the evolution of gibbon songs. *Biological Reviews*, *77*, 57-76.

Greenfield, M. D. (1994). Synchronous and alternating choruses in insects and anurans : Common mechanisms and diverse functions. *American Zoologist*, *34*, 605-15.

Griffiths, T., Johnsrude, I., Dean, J. L., & Green, G. G. R. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound ? *NeuroReport*, *10*, 3825-30.

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

Harrington, F. H. (1989). Chorus howling by wolves : Acoustic structure, pack size and the Beau Geste effect. *Bioacoustics*, *2*, 117-36.

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty : A comparative perspective. *Nature Neuroscience*, *6*, 663-68.

Howard, D. M., & Angus, J. A. S. (1998). A comparison between singing pitching strategies of 8 to 11 year olds and trained adult singers. *Logopedics Phoniatrics Vocology*, *22*, 169-76.

Huron, D. (2001). Is music an evolutionary adaptation ? In R. J. Zatorre, & I. Peretz (Eds.), *The biological foundations of music* (pp. 43-61). New York : New York Academy of Sciences.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526-28.

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition : A critique. *Trends in Cognitive Sciences*, *9*, 21-5.

Janik, V. M., & Slater, P. J. B. (1997). Vocal learning in mammals. *Advances in the Study of Behavior*, *26*, 59-99.

Janik, V. M., & Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, *60*, 1-11.

Jarvis, E. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, *1016*, 749-77.

Jürgens, U. (1992). On the neurobiology of vocal communication. In H. Papousek, U. Jürgens & M. Papousek (Eds.), *Nonverbal vocal communication : Comparative and developmental approaches* (pp. 31-42). Cambridge, UK : Cambridge University Press.

Kalmus, H., & Fry, D. B. (1980). On tune deafness (dysmelodia) : Frequency, development, genetics and musical background. *Annals of Human Genetics*, *43*, 369-82.

Kirzinger, A., & Jürgens, U. (1982). Cortical lesion effects and vocalization in the squirrel monkey. *Brain Research*, *233*, 299-315.

Koelsch, S., Gunter, T. C., v. Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks : A cortical "language-network" serves the processing of music. *NeuroImage*, *17*, 956-66.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech : Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, *100*, 2425-38.

Lieberman, P. H., Klatt, D. H., & Wilson, W. H. (1969). Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science*, *164*, 1185-87.

Lomax, A. (1968). *Folk song style and culture*. New Brunswick : Transaction Books.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area : An MEG study. *Nature Neuroscience*, *4*, 540-45.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*, 522-23.

Merriam, A. P. (1964). *The anthropology of music*. Evanston, Ill. : Northwestern University Press.

Miller, G. F. (2000). *The mating mind : How sexual selection shaped the evolution of human nature*. New York : Doubleday.

Mitani, J. C., & Gros-Louis, J. (1998). Chorusing and call convergence in chimpanzees : Tests of three hypotheses. *Behaviour*, *135*, 1041-64.

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

Murphy, J. R. (2005). *Motor speech disorders : Substrates, differential diagnosis, and management*. 2nd Ed. St. Louis : Elsevier Mosby.

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

Papousek, M. (1996). Intuitive parenting : A hidden source of musical stimulation in infancy. In I. Deliège, & J. Sloboda (Eds.), *Musical beginnings : Origins and development of musical competence* (pp. 88-112). Oxford : Oxford University Press.

Parsons, L. M. (2001). Exploring the functional neuroanatomy of music performance, perception and comprehension. In R. J. Zatorre, & I. Peretz (Eds.), *The biological foundations of music* (pp. 211-31). New York : New York Academy of Sciences.

Penhune, V. B., Zatorre, R. J., & Evans, A. C. (1997). Cerebellar contributions to motor timimg : A PET study of auditory and visual rhythm production. *Journal of Cognitive Neuroscience*, *10*, 752-765.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia : A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 185-91.

Petrides, M., & Pandya, D. N. (2001). Comparative cytoarchitectonic analysis of the human and macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience*, *16*, 291-310.

Pollack, G. (2000). Who, what, where ? Recognition and localization of acoustic signals by insects. *Current Opinion in Neurobiology*, *10*, 763-67.

Pollock, J. I. (1986). The song of the Indris (*Indri Indri*; Primates : Lemuroidea) : Natural history, form, and function. *International Journal of Primatology*, *7*, 225-67.

Porter, R. J., & Lubker, J. F. (1980). Rapid reproduction of vowel-vowel sequences : Evidence for a fast and direct acoustic-motor linkage in speech. *Journal of Speech and Hearing Research*, *23*, 593-602.

Poulson, C. L., Kymissis, E. Reeve, K. F., Andreators, M., & Reeve, L. (1991). Generalized vocal imitation in infants. *Journal of Experimental Child Psychology*, *51*, 267-79.

Rameau, J.-P. (1722/1971). *Treatise on harmony*. Philip Gossett, translator. New York : Dover Publications.

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 23

Rao, S. M., Mayer, A. R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, *4*, 317-23.

Richman, B. (2000). How music fixed "nonsense" into significant formulas : On rhythm, repetition, and meaning. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 301-14). Cambridge, MA : MIT Press.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*, 188-94.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1999). Resonance behaviors and mirror neurons. *Archives Italiennes de Biologie*, *137*, 85-100.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-41.

Rousseau, J.-J. (1781/1998). Essay on the origin of languages. In *Essay on the origin of languages and writings related to music* (pp. 289-332). Translated and edited by John T. Scott. Hanover : University Press of New England.

Sachs, Curt (1943). *The rise of music in the ancient world, east and west*. New York : W. W. Norton and Company.

Scharff, C., & Nottebohm, F. (1991). A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system : Implications for vocal learning. *Journal of Neuroscience*, *11*, 2896-913.

Schuler, A. L. (1979). Echolalia : Issues and clinical applications. *Journal of Speech and Hearing Disorders*, *44*, 411-34.

Simonyan, K., & Jürgens, U. (2003). Efferent subcortical projections of the laryngeal motorcortex in the rhesus monkey. *Brain Research*, *974*, 43-59.

Simpson, H.B., & Vicario, D. S. (1990). Brain pathways for learned and unlearned vocalizations differ in zebra finches. *Journal of Neuroscience*, *10*, 1541-56.

Skoyles, J. R. (1998). Speech phones as a replication code. *Medical Hypotheses*, *50*, 167-73.

Spencer, H. (1857). The origin and function of music. *Fraser's Magazine*, *56*, 396-408.

Spencer, H. (1890). The origin of music. *Mind*, *15*, 449-68.

Stewart, L., Walsh, W., Frith, U., & Rothwell, J. C. (2001). TMS produces two dissociable types of speech disruption. *Neuroimage*, *13*, 472-78.

Studdert-Kennedy, M. (2000). Imitation and the emergence of segments. *Phonetica*, *57*, 275-83.

Thomas, D. A. (1995). *Music and the origins of language : Theories from the French enlightenment*. Cambridge, UK : Cambridge University Press.

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in western tonal music. In R. J. Zatorre and I. Peretz (Eds.), *The biological foundations of music* (pp. 92-116). New York : New York Academy of Sciences.

Trehub, S. E. (2001). Musical predispositions in infancy. In R. J. Zatorre and I. Peretz (Eds.), *The biological foundations of music* (pp. 1-16). New York : New York Academy of Sciences.

Wallin, N. L., Merker, B., & Brown, S. (Eds.) (2000). *The origins of music*. Cambridge, MA : MIT Press.

Wickler, W., & Siebt, U. (1982). Song splitting in the evolution of duetting. *Zeitschrift für Tierpsychologie*, *59*, 127-40.

Yule, G. (1996). *Pragmatics*. Oxford : Oxford University Press.

Zatorre, R. J., & Binder, J. R. (2000). Functional and structural imaging of the human auditory cortex. In A. W. Toga, & J. C. Mazziotta (Eds.), *Brain mapping : The systems* (pp. 365-402). San Diego : Academic Press.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, *14*, 1908-19.

Musicæ Scientiæ/XI-1/RR 16/03/07 13:56 Page 25

• Heterofonía contagiosa: una nueva teoría sobre los orígenes de la música

Dos características destacadas de la música son la combinación de alturas y los ajustes de tiempo. Proponemos aquí una posible teoría precursora de la evolución de la música humana, basada en un proceso que denominamos "heterofonía contagiosa". La heterofonía es una forma de combinación de alturas en la cual los individuos generan líneas musicales similares pero en las que estas líneas están rudamente sincronizadas. Un excelente ejemplo se puede encontrar en el aullido de los lobos. Cada lobo emite una llamada similar, pero el coro resultante está pobremente combinado desde el punto de vista temporal.

La otra característica básica de la presente hipótesis es el contagio. Cada vez que un animal emite una llamada, otros miembros del grupo se unen a él, en un proceso expansivo. Mientras este tipo de llamada heterofónica está claramente representada en la naturaleza, la polifonía sincronizada no lo está. En este artículo discutimos los escenarios evolutivos en los cuales la capacidad humana de integrar los elementos musicales altura-espacio y tiempo pueden emerger en la música. Con este fin, hacemos mención de hallazgos de formación de imágenes neurales que aportan luz sobre los mecanismos neurales de imitación vocal y entrenamientto métrico en los humanos, dos procesos clave que subyacen en la integración musical.

• Eterofonia contagiosa: una nuova teoria sulle origini della musica

Due tratti salienti della musica sono la combinazione delle altezze e l'uniformazione del tempo. Propongo qui un possibile precursore evoluzionistico della musica umana basato su un processo che ho chiamato "eterofonia contagiosa". L'eterofonia è una forma di fusione delle altezze nella quale gli individui generano linee simili, ma scarsamente sincronizzate. Se ne può ritrovare un sorprendente esempio nell'ululato dei lupi. Ogni lupo emette un richiamo simile, ma il coro che ne risulta è ben poco armonizzato nel tempo. L'altro aspetto principale della presente ipotesi è il contagio. Una volta che un animale inizia ad emettere il proprio richiamo, altri membri del suo gruppo vi si aggiungono, in un processo di propagazione. Mentre questo tipo di richiamo eterofonico è ben rappresentato in natura, la polifonia sincronizzata non lo è. Nel presente articolo discuto scenari evoluzionistici secondo i quali la capacità umana di integrare le parti musicali nello spazio tonale e nel tempo potrebbe essersi tradotta nella nascita della musica. A tale scopo, accenno alle scoperte della visualizzazione cerebrale (*neuroimaging*) che gettano nuova luce sui meccanismi neuronali di imitazione vocale e di sincronizzazione metrica negli esseri umani, due processi-chiave alla base dell'integrazione musicale.

• Hétérophonie contagieuse : Une nouvelle théorie sur les origines de la musique

Marier les hauteurs et harmoniser les temps sont deux des composantes principales de la musique. Je propose ici un précurseur possible de l'évolution de la musique humaine, fondé sur le processus que j'appellerai « l'hétérophonie contagieuse ». L'hétérophonie est une manière de marier les hauteurs : les individus créent des lignes de musique semblables mais sans les synchroniser complètement. On trouve un magnifique exemple de ceci dans le hurlement des loups. Chaque loup lance un appel semblable aux autres, mais le chœur qui en résulte n'est pas bien ajusté dans le temps. L'autre versant de mon hypothèse est la contagion. Lorsqu'un animal lance son appel, d'autres membres du groupe le rejoignent peu à peu ; il y a une sorte de processus de propagation. Ce type d'appel hétérophonique est assez courant dans la nature alors que la polyphonie synchronisée ne l'est pas. Dans cet article, je traiterai de scénarios possibles de l'évolution qui pourraient expliquer comment la capacité humaine à intégrer des parties musicales dans l'espace des hauteurs et dans le temps a pu donner lieu à la musique. Ce faisant, je parlerai des découvertes de l'imagerie neurologique qui éclairent les mécanismes neuraux de l'imitation vocale et de l'organisation métrique chez les êtres humains ; ce sont là deux processus qui sont à la base de l'intégration musicale.

• Ansteckende Heterophonie : eine neue Theorie zum Ursprung der Musik

Zwei der auffälligsten Merkmale von Musik sind die Kombination von Tonhöhen und ihre zeitliche Anpassung. Ich schlage hier einen möglichen evolutionären Vorläufer für menschliche Musik vor, der auf einem Prozess basiert, den ich „ansteckende Heterophonie" nenne. Heterophonie ist eine Form der Tonhöhenkombination, bei der Individuen ähnliche musikalische Linien erzeugen, die dabei jedoch kaum synchronisiert sind. Ein großartiges Beispiel dafür ist das Heulen von Wölfen. Jeder Wolf ruft auf gleiche Weise, der daraus resultierende Chorus ist jedoch kaum zeitlich aufeinander abgestimmt. Das andere Hauptmerkmal der hier vertretenen Hypothese liegt in der Ansteckung. Sobald ein Tier zu rufen beginnt, fallen die anderen Mitglieder des Rudels durch einen Prozess der Ausbreitung mit ein. Während diese Art des heterophonen Rufens in der Natur häufig vorkommt, trifft dies nicht für synchronisierte Polyphonie zu. In diesem Aufsatz diskutiere ich evolutionäre Szenarien, wie sich die menschliche Fertigkeit, musikalische Parameter im Hinblick auf Tonhöhe und Zeit zu integrieren, in der Musik entwickelt haben könnten. Dabei erwähne ich Ergebnisse der bildgebenden Neurowissenschaft zu zwei der wichtigsten Prozesse für die musikalische Integration, nämlich die neuronalen Mechanismen der vokalen Imitation und des metrischen Entrainments bei Menschen.