Chapter 9

Emotional communication in monkeys: Music to their ears?

Charles T. Snowdon and David Teie

Abstract
The question of whether music is uniquely human is controversial with equivocal results on whether nonhuman animals can appreciate music. We argue that, although human music is unique to our species, the emotional aspects of music have a long evolutionary history. We first present a theory of the emotional origins of music and then show that music/emotional features can be seen in the vocalizations of cotton-top tamarins, a small monkey from the New World. When we composed music comprising our hypothesized emotional features in the frequency range and tempos of tamarin vocalizations, and played these compositions to the monkeys, we could induce calming and arousing emotional states in the monkeys, although human-based music with similar features had little or no effect on the monkeys. Our results suggest that other species incorporate similar emotional features into their calls as are used in music to influence emotions. The results also suggest that calls are not merely communicating the caller’s emotional state but may be used to induce emotional responses in listeners. Furthermore, attempts to play human-based music to other species as enrichment are likely to be misguided, with human-based music either being irritating or merely ineffective.

Introduction
There has been much recent interest in the evolution of music (see reviews by Fitch 2006; McDermott and Hauser 2005; Patel 2008) with potential parallels to music found in bird song (Baptista and Keister 2005; Marler and Slabbekorn 2004). “Song” is also described in whales and the duets of gibbons, but whether these other species have anything that can be defined as music in their own behavioral repertoires or respond to music (of any type) has been little studied. Several studies on nonhuman animals suggest that other species rely solely on absolute pitch with no little or no ability to transfer a melody learned in one key or octave to another key or octave (review in McDermott and Hauser 2005). The one exception is from Wright et al. (2000) who found that rhesus macaques could generalize from one octave to another. Interestingly, this was the only animal study to use a diatonic scale. Pigeons trained to discriminate between Bach’s Toccata and Fugue in D Minor for organ versus Stravinsky’s Rite of Spring for orchestra could eventually discriminate between the pieces, but learned slowly and did not reach high levels of performance (Porter and Neuringer 1984). Studies of responses of cotton-top tamarins (Saguinus...
and common marmosets (*Callithrix jacchus*) to human-based music found that both species preferred slow tempos to fast ones, and preferred Mozart to rock music. However, when either type of human-based music was tested against silence, both species of monkeys preferred silence (McDermott and Hauser 2007). Several authors (Fitch 2006; McDermott and Hauser 2005; Patel 2006) have suggested that if anything resembling music is to be seen in other species it will only appear in species capable of vocal learning. From this perspective it is not surprising that pigeons and nonhuman primates are unable to respond to music or can learn to discriminate between musical pieces only with great difficulty.

Rhythmic behavior is not commonly observed in nonhuman species although neonatal humans detect musical beats and can anticipate the downbeat even when it is not marked by stress or other spectral features (Winkler et al. 2009). Two recent studies have suggested rhythmic behavior in nonhuman species. A sulphur-crested cockatoo (*Cacatua galerita eleonora*) bobbed its head and tapped its foot in synchrony to tempi ranging from 81–130 beats per second (Patel et al. 2009). Remedios et al. (2009) reported rhythmic drumming by rhesus macaques (*Macaca mulatta*) using cage doors and other objects, but did not show synchronization to an external beat as in the cockatoo. Monkeys reacted as much to drumming as to natural vocalizations and brain areas activated by conspecific sounds and drumming showed overlapping activation in the amygdala and caudal auditory cortex. This latter finding is particularly striking since rhesus macaques are not thought to be vocal learners, a precondition hypothesized to underlie any form of music parallels in other species.

Some authors (e.g., Levitin 2009; Juslin and Västfjäll 2008) argue that music is one of the best forms of emotional communication known and that musical components of speech (prosody) provide honest emotional signals. Despite the limited data on musical abilities in nonhuman species there has been great interest in the structure of acoustic signals that communicate emotional state in nonhuman species. Morton (Morton 1977; Owings and Morton 1998) has argued that high-pitched, pure-tone sounds are common to friendly or appeasing contexts whereas low, loud, noisy (broad band) sounds are common to expressions of threats and aggression.

Studies of communication between human parents and infants have identified prosodic (musical) features in speech that can influence the behavioral state of the infant (Fernald 1992). Several short, upwardly rising staccato calls lead to increased arousal. Long descending intonation contours have a calming effect and behavior can be stopped with a single short plosive note. Interestingly, the same features appear when the calls or whistles used by humans to control behavior of working animals (dogs and horses) are analyzed (McConnell 1991; McConnell and Baylis 1985). The convergence of signal structure that humans use to communicate with both infants and nonhuman animals suggests that these signals can be used to induce behavioral change in others. In the case of humans who are attempting to manage the behavior of infants and animals, they need not be experiencing the emotion they are trying to induce. We know very little about the effects of natural animal signals on inducing emotions in other animals, a point we will address later in this chapter.

Musical structure is said to affect the behavior and physiology of humans. Infants as young as 2 months old spend more time looking at a speaker that provided consonant compared with dissonant music (Trainor et al. 2002; Zentner and Kagan 1996) and after hearing dissonant music, it was difficult for infants to attend subsequently to consonant music. Mothers across several cultures could identify lullabies from other cultures and, when asked to sing a nonlullaby in the presence or absence of an infant, mothers in various cultures sang in a higher key and with slower notes to their infants than when singing the same song in the absence of infants (Trehub
et al. 1993). Studies on the effects of music on emotions showed that for adults certain types of classical music so called “high-uplifting” music such as Kreisler’s *Liebesfreud* or Satie’s *Picadilly* led to increased activity, reduced depression and increased norepinephrine levels whereas “low-uplifting music” such as Albinoni’s *Adagio for Strings and Organ* or Satie’s *Gnossiennes No 4* led to an increased sense of well-being (Hirokawa and Ohira 2003). The specific components of these pieces that led to different states (e.g., tempo, dissonance, timbre) were not identified. In parallel to these results, Panksepp and Bernatzky (2002) reported that chickens exposed to amplitude modulated wide band instrumental music showed increased growth and reduced separation induced vocalizations as well as displayed increased levels of the neurotransmitters norepinephrine and dopamine and their metabolites. All these results taken together suggest that, although many specific components are yet to be identified, musical components of pitch and tempo can be used specifically to alter emotional, behavioral, and physiological states in infant and adult humans as well as in companion animals and even in chickens.

If this is the case, why then are monkeys responsive to tempo but indifferent to human music (McDermott and Hauser 2007)? A major issue of testing animals with music composed for humans is that tempos and pitch ranges of human music may not be relevant for another species. There may also be species typical ways of expressing different emotional states (e.g., increasing versus decreasing pitch contours).

We have been interested in developing a comprehensive theoretical approach to the emotional effects of music. In order to test the theory we had to find a population that has not been influenced by prior experience with human music and had no cultural expectations for what music should be like. We used a nonhuman primate species, the cotton-top tamarin, previously studied by McDermott and Hauser (2007). The monkeys in our colony had never been exposed to music and thus had no prior experience with music (We were not required to use music as an enrichment method in our colony and in fact our results described in this chapter, coupled with those of McDermott and Hauser, suggest that playing human-based music to primates may have a negative effect and should not be used as an enrichment tool). However, knowing that this species preferred silence to either form of human-based music tested by McDermott and Hauser (2007), we analyzed the structure of tamarin vocalizations and constructed musical pieces that we predicted would be appropriate for tamarins. In the next section we present our general theory and in subsequent sections we evaluate the theory with respect to natural vocalizations of tamarins and then follow with an experiment that tested monkeys with both human and tamarin appropriate music.

**A theory of the emotional origins of music**

We designed species-specific music to test the validity of the theory of the emotional origins of music. This theoretical model recognizes that music, as we know it, is a human construct made by and for humans based on our species’ developmental experience, vocalizations, perceptions, and cognition, but that music might have had its evolutionary origins in emotional communication in other species. We predicted that by modifying musical characteristics to conform to the development, vocalizations, and perceptions of another species, we could create music that would elicit appropriate affective responses from members of the other species.

There is no generally accepted alternative theory of the origin and affective processes of music. Although the neural pathways and emotional responses of many musical elements have been identified, such as the fear response to dissonance (Blood et al. 1999) the reasons for those responses have been elusive.
Unpleasant (permanently dissonant) music contrasted with pleasant (consonant) music showed activations of amygdala, hippocampus, parahippocampal gyrus, and temporal poles. These structures have previously been implicated in the emotional processing of stimuli with (negative) emotional valence; the present data show that a cerebral network comprising these structures can be activated during the perception of auditory (musical) information. Pleasant (contrasted to unpleasant) music showed activations of the inferior frontal gyrus, the anterior superior insula, the ventral striatum, Heschl’s gyrus, and the Rolandic operculum. (Koelsch et al. 2006.)

Why should we have an innate fear-related response to dissonance? We have attempted to establish a logical connection between each of the many separately identifiable elements of music and the emotional response that it triggers.

Outline of theory

The human auditory system provides particularly clear processing of certain types of sounds, such as the sound of the human voice (see “Resonator-enhanced periodic sounds” section). Stable auditory images are compared to templates of recognition and, when matches are identified, appropriate emotional responses are triggered (Morris et al. 1999). We propose that music presents patterns of sounds that are given priority status in primary auditory processing and are close enough to matching some of the acoustic templates of recognition to trigger emotional responses. Of the templates that musical elements resemble, we hypothesize that some are innate, such as the sound of a human scream, and others are formed by the sounds heard by the fetus in the womb when the fetal brain structures are plastic and being organized. We formulated the species-specific music for the cotton-top tamarins based on the theory that human music presents, concurrently and consecutively, a quickly changing array of acoustic triggers that are specific to *Homo sapiens*. Even a simple example of a single instrument playing a simple 14-s melody may present eight different acoustic stimuli that are each capable of inducing an emotional response (see “Appendix”). Listening to music may be compared to viewing a series of rapidly alternating pictures that induce fear, affection, excitement, and tranquility. The accumulation of neurochemical responses to auditory stimuli may contribute to the enhanced emotional state induced by music.

A variety of musical elements including compositional techniques, instrument modifications, and performance practices were incrementally introduced over the centuries of the development of music. We submit that only those elements and modifications that induced emotional responses were adopted into general usage and these are examined next.

The origins of these elements can be classified into four categories. The first two categories involve responses triggered by auditory processing and recognition: (1) limbic system development and (2) emotionally generated vocalizations. The other two categories trigger responses unrelated to sound that have been exapted by music: (3) visuospatial processing and (4) linguistic processing. There is some redundancy in both the limbic formation and emotional vocalization categories due to the inclusion of prosodic features in spoken language of characteristics that are found in emotionally generated vocalizations such as the loud, high-pitched sound of a scream used to verbalize an alarm call.

The following is an outline of the four categories and their respective musical elements:

1. **Limbic system development memory:** tempo of pulse (pulse is understood as the regular, underlying beat that defines the meter), amplitude contour of pulse, meter, notes, syllabic contour, melodic rhythm, melodic accents, phrase length, phrase contour, and melodic frequency range.
2. *Emotionally generated vocalizations*: resonance-enhanced periodic sound, timbre, tonality, frequency range, melodic contour, melodic rhythm, accents of melody, loudness, rate of syllabic repetition, vocal tract variables, and tempo range.

3. *Visuospatial*: pattern, chaotic movement, directed movement over time (melody), perfect/celestial movement, harmony, counterpoint, and structure (see Patel 2008, chapter 4 for further discussion).


The following are descriptions of those aspects of our theory that were involved in the compositions for tamarins.

**Limbic system development memory**

This category includes those elements of music that are most universal. It includes musical representations of sounds that are heard during the time when the limbic structures are formed in the developing brain. The following conditions allow for the formation of lasting fetal acoustic memories. The human fetus is able to hear at 24 weeks, providing 4 months of constant sound exposure (Birnholz and Benacerraf 1983) prior to birth. The sound of the maternal heartbeat is 25 db above basal noise, dominating the fetal environment (Querleu et al. 1988). The maternal voice is heard in the uterus nearly four times more strongly than it is heard externally (Richards et al. 1992).

In *utero* research and analysis has shown consistent evidence that the fetus responds to the sound of the mother's heartbeat (Porcaro et al. 2006). And infants also respond to the prosodic features of speech (see later).

The combination of three features of human fetal development make it possible for the sounds of the womb to provide a lasting template of recognition: (1) The dearth of competing sensory information in the fetal environment allows sound to be a primary source of varied and ever-present information entering the developing brain; (2) information that is well-organized when incoming while a brain structure is plastic will tend to remain organized in the brain; and (3) the limbic structures are well-developed at birth. The limbic fibers, the cingulum and the fornix, are two of the most dominant tracts in the fetal brain and their entire trajectories are already developed at 19 gestational weeks. “Early formation of the limbic system is well known, and it is expected that limbic fibers are well formed at 19–20 gestational weeks” (Huang et al. 2006). A logical conclusion to the summed effects is that brain structures responsible for emotions and that are well developed at birth may remember and later respond to sounds that resemble those of the fetal environment.

In light of anatomical studies that have emphasized the interconnections between ventral limbic circuits and the motor control connections between striatum and motor cortex (Gunnar and Nelson 1992), we propose that the acoustic information that pervaded the development of structures responsible for our emotions as well as structures near the brainstem responsible for repetitive movement is the source and origin of pulse, meter, and rhythm in music. McDermott (2008) identified several universal properties of music: pulse, hierarchal organization of scales (tonality), infant-directed song, dance, and meter. To McDermott’s list we would add: amplitude contour of pulse instrument, use of resonator-enhanced periodic sounds, prevalence of discrete single-frequency units (musical notes), varied pitches and rhythms in the melodies (prosody), and the 200–900-Hz frequency range of melodic instruments. All of these universal features of music can be traced to the fetal acoustic environment that informed the developing structures of the limbic system.
A regular and repeated pulse is one of the universal traits of music even though it is not found in human vocalizations. The repetition rates of musical pulses (40–240 beats per minute) coincide with the slowest (respiration) and fastest (footfalls of running) pulses that can be heard in the womb. “Intrauterine recordings in humans and animals have confirmed that prenatally audible sounds include the mother’s vocalizations and breathing, heartbeat, body movements, footfalls and digestion” (Parncutt 2006). Instruments that create amplitude contours that resemble that of a heartbeat commonly keep the musical pulse. The regular and constantly repeated sounds of maternal respiration create a repetitive acoustic framework that is broader and less specific than the heartbeat and possibly provides a basis for the recognition of more generalized pulse in music.

The construction of a drum enables it to create a heartbeat-like amplitude contour of the pulse instrument. Drums have been similarly constructed in many different cultures. (1) The onset of sound is graduated by a cushioned beater, a stretched animal skin, or both. (2) The decay of sound is elongated with a resonating chamber. (3) The graduated onset and elongated decay of a strike make the amplitude contour resemble that of the heartbeat as heard in the womb (onset 0.02s, decay 0.06s).

Meter is a repeated pattern of strong and weak beats. The combinations of strong and weak pulses found in the primary meters are derived from the sounds of respiration combined with the sound of the heartbeat. Strong-weak is duple meter. 1, 2, 3, 4 (1 strongest–2 weak–3 strong–4 weak) is known as “common time” in Western music. When respiration and heartbeat are combined: 1 inhalation + heartbeat–2 heartbeat alone–3 exhalation + heartbeat–4 heartbeat alone, the result is common time and this is consistent with normal human heart and respiratory rates (four heartbeats/respiratory cycle). We suggest that the prevailing duality of pulse in Western music is the same duality found in the human rhythms of heartbeats, breathing, and walking. A weak beat placed in the silence between the duple pulses of the heart (LUBB, dub, silence) creates a triple meter: ONE, two, (three).

The mother’s speech that is heard in the womb consists primarily of single-frequency segments created by the vowels between the consonants (Querleu et al. 1988). These units provide the singular basis for notes in music. Mammalian vocalizations generally consist of syllables that have contoured frequencies (sliding pitches) such as a cat’s meow or a dog’s submissive whimper as well as the human vocalizations such as moaning and weeping that originate in structures such as the ventromedial prefrontal cortex, the anterior cingulate cortex, the extended amygdala, and the ventral striatum (Parvizi et al. 2001). Despite this preference for contoured frequencies in emotional vocalizations, human music contains a preponderance of discrete single-frequency units. The acoustical properties of the womb attenuate frequencies unevenly. Due to the absorption of sound by the surrounding tissues in the womb, higher frequencies from external sources are subject to more attenuation than lower frequencies. Consequently the consonants of speech are nearly inaudible in the womb but the “melody” of the pitches created by the vowels between the consonants is quite audible. Speech is produced in predominantly consonant intervals and contains implied tonalities (Bowling et al. 2009; Schwartz and Purves 2004). As a consequence, the melodies heard in the womb are primarily harmonically consonant.

A spoken sentence is heard in the womb as a pattern of these discrete pitches in a variety of melodic contours and rhythms. The prosody of languages forms the bases for melodic treatment in music. Newborns of French mothers prefer the sound of the French language to Russian (Mehler et al. 1988). The newborns still prefer the French language when the speech is filtered to remove the consonant and vowel sounds, retaining only the melody, but they do not show a preference for the melody of the French language when played backwards, implying that a fetus
is able to recognize intervallic relationships and melodic contours. Evidence for fetal learning of the melodic contours of maternal speech is also found in the cries of newborns that emulate the melodic contours of the mother’s language (Mampe et al. 2009). Words and combinations of words create recognizable rhythms that are found in the melodic rhythms of musical motives. Cultures whose languages have accented syllables also have corollary accents in their melodies. For example, the definite articles in the Germanic and Romance languages (the sea, die See, la mer) are heard in the musical upbeats at the beginning of many melodies. The music of cultures whose languages do not contain definite articles, rarely have musical upbeats to their melodies. Note the preference for beginning melodies on the beat in the music of Mussorgsky (Russian) and Dvorak (Czech). A number of other commonalities have been found between the melodic rhythms of a culture and the speech rhythms in its language (Huron and Ollen 2003; Patel and Daniele 2003).

The frequency range of melodic instruments in a wide variety of cultures is roughly 200–900 Hz, the same as the frequency range of an adult human female voice.

**Emotional vocalizations**

All of the acoustic characteristics of emotionally generated vocalizations have been incorporated into emotionally charged speech and music. Knowledge of the acoustic characteristics of a species’ emotional vocalizations tends to be innate and universally shared among members of that species (Herzog and Hopf 1984). A primary role of the first phase of auditory processing is to compare incoming sounds to templates of recognition and signal an appropriate response when an acoustic “match” is identified (Griffiths and Warren 2002). One of the keys to music’s ability to induce affective states in humans is by creating sounds that are approximate matches to commonly shared templates of recognition. The emotional vocalizations of affection and submission are quiet and consist of purer waveforms whereas the emotional vocalizations of threats and alarms are loud and have complex waveforms (Morton 1977). Broadly speaking, the variations of timbre, frequency range, and amplitude found in music are based on these parameters of emotional vocalizations.

Most sounds in the natural world are either nonperiodic (broadband sound of wind or a waterfall) or transient (clicks and pops). Animal vocalizations produce a characteristic resonator-enhanced periodic sound created by the vibration of the vocal folds of the larynx with overtones added by the vocal tract (Fitch and Reby 2001). Mammals have a broad range of resonance enhancements and many mammals, including primates, have additional overtone modifications created by changes in the shape of the mouth (vowels) (Fitch and Hauser 1995). The human auditory system is predisposed to filter out nonperiodic and transient noise. When the input to the cochlea is a periodic sound the neural activity pattern of the sound oscillates. In contrast, the sensation produced by such a sound does not flutter or flicker; indeed, “periodic sounds produce the most stable auditory images” (Patterson et al. 1992, p. 4).

This predisposition allows music to given priority status in auditory processing since nearly every pitched musical instrument produces a resonator-enhanced periodic sound. The importance of adding modifying resonance to the periodic sound in musical instrument manufacturing is indicated by the modifications made to the electric guitar. The amplification that is provided by the resonating body of an acoustic instrument is not necessary on an electric guitar, however the sound from the magnetic pickup was judged to be too pure by the early inventors who then developed and installed modifiers that electronically added overtones to the fundamental periodic sound (Poss 1998).
The dissonant–consonant intervals used in speech and further adapted by music are derived from and directly related to the complex/threat and pure/affection polarity of primitive emotional vocalizations. This acoustic dichotomy is one of the axes of music. The amygdala responds to emotional vocalizations (Fecteau et al. 2007) and also generates a fear response to dissonance (Ball et al. 2007).

Activity in right parahippocampal gyrus and precuneus regions correlated with increasing dissonance, whereas activity in orbitofrontal, subcallosal cingulate and frontal polar cortex correlated with decreasing dissonance (equivalent to increasing consonance). The unpleasantness of increasing dissonance was confirmed by subjects’ analog ratings; these ratings also covaried with rCBF changes in similar paralimbic regions. (Blood et al. 1999.)

This dissonant/consonant emotional alignment is consistent with the nonhuman acoustic alignment of affective calls identified by Morton because simultaneous tones spaced at dissonant intervals create an out-of-phase auditory competition between the overtones that result in complex waveforms. The aligned overtones of consonance create waveforms that are relatively pure. The complex overtones of a chord containing a dissonance may trigger a fear response because it resembles the complex voiceprint of a threat containing a periodic sound with a broad array of overtones. Conversely, consonant harmonies that have purer waveforms may be interpreted as sound that resembles an affectionate vocalization. In another expression of this pure/complex waveform polarity, the dissonant intervals found in warning cries and threats have been translated into their dissonant musical-melodic counterparts and the consonant intervals of affective communication are the foundation of harmonious melodies. These opposing responses to dissonance and consonance have been incorporated into music in a manner that provides the aesthetic perceptions of conflict/resolution and movement when a dissonant appoggiatura is followed by a consonant resolution.

It should be noted that complex waveforms induce fear responses only when associated with the periodic sounds of musical notes or vocalizations; complex waveforms “white noise” that are not associated with periodic sounds do not induce fear responses.

**Sympathetic arousal of affective states through vocalizations**

Variations in the resonating chambers of the vocal tract transmit information relevant to the affective state of the vocalizing individual. During in-group communication these variations will tend to elicit a sympathetic emotional response in the listener. The three variables of vocalizations are: (1) vocal folds of the larynx tighten and loosen to raise and lower the pitch, providing the source of the periodic sound; (2) the higher or lower placement of the larynx in the throat shortens or lengthens the resonating cavity providing distinguishable formant patterns; and (3) the shape of the mouth creates patterns of overtones that are recognizable as vowels.

These variations may sympathetically induce emotions in the listener. The added resonance of a lowered larynx is capable of inducing a lowered larynx in an in-group listener (feels a “lump” in the throat). Similarly, the adrenaline-heightened state of someone screaming is capable of inducing an adrenaline-heightened state in the listeners. These responses have even been demonstrated in human responses to affective calls of other species (Belin et al. 2008). This phenomenon of sympathetic emotional response to resonance-enhanced periodic sound may account for some of the emotional connectivity of music.

Table 9.1 provides a list of mood-related features of musical vocal production that are based on human emotional vocalizations. Musical instruments have been developed that present timbres resembling those of affective vocalizations. Players of instruments that have highly variable
timbres, such as the violoncello and saxophone, are well schooled in the varying techniques and consistently strive to produce timbres that are in keeping with the perceived emotional intent of the music.

The two remaining categories of responses to music are the exaptations of visuospatial and linguistic processing. Because these responses are uniquely human, they will not be described in this chapter (see Patel 2008 for an extended discussion of both of these). Appreciation of these aspects of music is enabled by the allocation of neural resources through exposure and attention. Consequently, the musical elements in these categories are not found in the music of all cultures.

In summary, this theory proposes that: (1) the human auditory system is predisposed to efficiently and clearly process certain types of sounds such as those resembling the human voice, and most musical instruments create sounds that benefit from this clear processing; (2) that each element in music induces an independent emotional response by presenting an acoustic stimulus that is a match for a preexisting template of recognition; (3) as these elements are recognized by a listener in music, concurrently and consecutively, they trigger appropriate emotional, neurochemical responses; and (4) this accumulation of recognition responses may be one of the cornerstones of emotional response to music. This theory is not meant to account for all aspects of music but rather to address how music might have evolved from emotional signals. Clearly music has undergone a rapid expansion and development so that some music may even be deliberately composed to avoid inducing any affective state (e.g., Cage)

**Habituation**

Habituation occurs when the reticular activating system allows conscious identification of the source of a nonthreatening sound to effectively disable an attentive response to that sound. To the extent that music presents acoustic stimuli that resemble, but do not perfectly imitate, emotional vocalizations; this may allow music to escape conscious identification and, therefore, habituation. Music that incorporates constantly changing instrumentation, patterns, and keys also tends to avoid the repeated recognition required for habituation.

**Table 9.1** Mood-related features of vocal production

<table>
<thead>
<tr>
<th>Soothing/affective:</th>
<th>Lively/affective:</th>
<th>Sympathetic/expressive:</th>
<th>Threat</th>
<th>Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raised larynx</td>
<td>Raised larynx</td>
<td>Lowered larynx</td>
<td>Frontal vowels</td>
<td>Open vowels</td>
</tr>
<tr>
<td>Pure waveform</td>
<td>Vocal waveform</td>
<td>Vocal waveform</td>
<td>Complex waveform</td>
<td>Penetrating waveform</td>
</tr>
<tr>
<td>Closed vowel (oo)</td>
<td>Open vowel (ah)</td>
<td>Mid-closed vowel (oh)</td>
<td>Open vowels</td>
<td>Open vowels</td>
</tr>
<tr>
<td>Moderate tempo</td>
<td>Quick, short notes</td>
<td>Slow descending phrases</td>
<td>Accented notes</td>
<td>Sustained notes</td>
</tr>
<tr>
<td>High vocal range</td>
<td>Mid-high vocal range</td>
<td>Mid-low vocal range</td>
<td>Low vocal range</td>
<td>High vocal range</td>
</tr>
<tr>
<td>Open consonant intervals</td>
<td>Major diatonic intervals</td>
<td>Minor diatonic intervals</td>
<td>Dissonant intervals</td>
<td>Dissonant intervals</td>
</tr>
<tr>
<td>Quiet</td>
<td>Moderate</td>
<td>Quiet</td>
<td>Moderately loud</td>
<td>Loud</td>
</tr>
</tbody>
</table>
Combined stimuli

The elements of music must be presented in combination in order for them to induce sufficient emotional responses to bring about observable behavioral changes. In a complex and interdependent system it is often impossible to test the effectiveness of each element separately. For example, if a human scream consisting of a loud, broadband waveform with an open vowel at 200–300 Hz triggers an attentive response, this does not imply that a presentation of a 200–300 Hz signal alone would induce ¼ of that attentive response. A property of auditory recognition at the neurological level known as combination sensitivity provides that the response to a given set of acoustic characteristics is greater than the sum of the responses to each individual characteristic presented independently, akin to arguments made by Gestalt psychologists (Kanwal et al. 2004).

Combined stimuli induce an enhanced emotional state in the listener by creating sequences of neural activations. One of the ways that music induces emotional responses may be through a constant interplay between the dissonance of threats and the consonance of affection and between the familiar and the novel. Primates are particularly sensitive to novelty as well as to familiarity (Wilson and Rolls 1993) and both novelty and familiarity are capable of triggering an attentive response. If new affective neural responses are introduced before previously introduced responses have dissipated the overlap will result in an emotionally “charged” state. The thrill of a musical climax may be partly dependent upon the neural activation that precedes it. This would help to explain why the tamarins in our tests did not respond during the playback of the 30-s selections (see later), but did respond during the 5 min after the music was presented. (This could also be a function of the greater frequency of behaviors observable during a longer duration.)

Chaos/order polarity

An evaluation of the intervallic relationships and rhythmic variations found in human emotional vocalizations and tamarin calls in a variety of affective states indicates the possibility for an expansion of Morton’s rules. The chaos/order polarity found in vocalization waveforms also governs intervals and rhythms. Chaos would be associated with threat and be expressed not only by complex overtones, but also by dissonant intervals and irregular rhythms, whereas order would be associated with affiliation and be expressed by simple wave forms, consonant intervals and regular rhythms.

Musical analysis of tamarin calls

Before we could create music to test the theory outlined previously, we needed to have an understanding of the natural repertoire of tamarins. We used archival and contemporary recordings of cotton-top tamarins in a variety of social contexts. The recorded calls were then presented to a panel of five professional musicians with no previous experience with cotton-top tamarins and no knowledge of the social contexts in which the calls were given. Each musician played a different instrument and was unaware of the premises of this study. The calls were played back at a slower speed to put the pitches and tempos within the range of human music. The panel of musicians was asked to place the calls into five categories of timbres, from pure to harsh. We included these categories in our musical evaluations of the tempo, note duration, direction of pitch change, intervals between pitches, and harmonic versus dissonant components of each of the calls. We then compared the call types with previous work that related specific call structures to social contexts (Campbell and Snowdon 2007; Cleveland and Snowdon 1982). We only presented musicians with calls with multiple syllables or units.
We identified five clusters of call types: Cluster 1 included Small Initially Modulated Whistle, Terminally Modulated Whistle, Ascending Multi-Whistle and the Partial Quiet Long Call. (Call names from Campbell and Snowdon 2007; Cleveland and Snowdon 1982.) These calls had a narrow frequency range (1.9–3.2 kHz related to “head voice” or falsetto as with human infant directed calls) with relatively long notes ranging from 150–500 ms and ascending notes at intervals of 2nds, 3rds, and 4ths of the first note. The structures were simple with two to four notes per call. Each of these calls is used in contexts of affiliation and positive social interaction.

Cluster 2 consisted of Type A and Type H Chirps and Hooked Chatter. All contained short notes of descending intervals of 3rds, 4ths, 5ths, and 6ths. Frequencies ranged from 3.2–9.4 kHz and were also in a “head voice.” The repeated notes were given at six to eight notes per second. All of these calls are given in contexts of high arousal or in response to threats.

Cluster 3 consisted of Chevron Chatter and Type E Chirp Chatter. These calls contain motives that begin and end on the same pitch with an interval of a major second in between. The notes are produced in staccato rhythm at 12–14 notes per second. Both calls are used in contexts of high fear and high arousal.

Cluster 4 consisted of Type F Chirp + Whistles, Normal Long Calls, and Large Modulated Multi-Whistles. These are full-voiced (normal-voiced) calls ranging from 1.1–2.35 kHz with note durations of 200–1100 ms with two to four notes per call. All have rising pitches with no slowly descending slides. The first two calls are vocal responses to hearing the calls of an unfamiliar group and the last is given in aggressive play with the broad context of all of the calls being confident threat.

Cluster 5 calls were the only ones that contained triple meters (note, note, silence). The Rapid Whistle contained ascending intervals in 3rds and 4ths. Notes were approximately 100 ms duration with three to four notes per call with a frequency range of 0.4–1.4 kHz. This call is used by young animals in play mounting and by adults before sex. The Type B Trill has repeated descending intervals of thirds with a frequency range from 1.6–3.2 kHz and is used in retrieving infants, so both calls are used in contexts of approach.

These analyses confirmed many of our theoretical principles but also identified some important differences between human emotional vocalizations and those of tamarins. Natural human tempos range from slow breathing (18/min) to the footfalls of running (340/min) and most of the tempos of human-based music fall within this range (40–208 beats per minute), the intermediate tempos of heartbeats and walking being the most common. A normal human’s vocal range spans approximately two octaves, ranging from 65–262 Hz for a low-voiced man, to 294–1175 for a high-voiced woman. In contrast, the monomorphic tamarin displays a vocal range spanning nearly five octaves from 262–7040 Hz. Most tamarin vocalizations are approximately two times faster than and three octaves above the range of comparable female human affiliation vocalizations. Many tamarin calls displayed clear chordal structures with key centers on a diatonic scale. Tritones, minor seconds, noise, and short staccato notes were more common in calls relating to fear and threat. The pitch variations and aspirant onsets of tamarin vocalizations are from two to three times faster than human vocalizations. In addition, tamarin affiliative vocalizations are characterized by a series of ascending notes (see Cluster 1) whereas human affiliation vocalizations are characterized by descending intervals or glides.

**Composing musical stimuli**

The differences between human and tamarin structures led us to hypothesize that tamarins might be unlikely to respond to human-based music but might instead be likely to respond
to music created to match the frequency range, tempos and pitch changes used by tamarins in their communication. This led us to an experimental design that would present tamarins with samples of human-based music with features modeling human affiliation vocalizations and human threat vocalizations, but also to create a set of tamarin-based musical stimuli that would include the hypothesized universal features of affiliative and threat-based emotional vocalizations embedded in a context of pitches, tempos and pitch changes relevant to tamarins and a pulse based on the resting heart rate of an adult tamarin (200–220 beats per minute).

Before describing the compositions it is important to note that we did not mimic the vocalizations of the tamarins. The musical compositions that we used are not imitations since it would be of little value to simply create a playback study of tamarin vocalizations which could be done much more efficiently and would be much less interesting. Note that since tamarin vocalizations are three times higher in pitch but only two times faster in tempo and since affiliative calls have ascending rather than descending notes, we cannot simply transpose human music to the tamarin range.

Just as linguists have specific definitions of language that exclude nonhuman species, musicologists may question whether any music composed to affect a nonhuman species is in fact music. Nonetheless, in our compositions we have used standard compositional techniques such as repeating phrases augmented with a second parallel line using standard counterpoint with resolution to a chord at the end of each of the phrases in the key used in composition. The threat vocalizations of the tamarins use harsh timbre, irregular rhythms, and dissonant intervals in the frequency range of 1.6–2 kHz. We created music using harsh timbres, irregular rhythms, and dissonant melodic intervals in the same frequency range, but avoided imitating the tamarin calls.

We selected two examples of human-based affiliation music and two examples of human-based threat music and created two samples of tamarin-based affiliation music and two examples of tamarin-based threat music. A summary of the characteristics of each piece is presented in Table 9.2.

**Affiliation vocalization-based music**

A selection from Barber’s “Adagio for Strings” was used as one human-based control. It contained 40 notes/min, and frequencies ranged from 82–622 Hz which correspond to the tempo

| Table 9.2 Frequency, tempo, and timbre of stimuli presented |
|---------------------------------|-----------------|----------------|
| **Affiliative-based music**     | **Frequency range (Hz)** | **Notes per minute** | **Timbre** |
| Human 1                         | 82–622           | 40              | Pure tone  |
| Human 2                         | 196–587          | 35              | Pure tone  |
| Tamarin 1                       | 1900–3700        | 52              | Pure whistle |
| Tamarin 2                       | 1800–4100        | 82              | Pure whistle |
| **Threat-based music**          |                  |                 |            |
| Human 1                         | 82–124           | 206             | Harsh/noisy |
| Human 2                         | 73–262           | 307             | Harsh/noisy |
| Tamarin 1                       | 1400–7900        | 554             | Harsh/noisy |
| Tamarin 2                       | 1300–5400        | 496             | Harsh/noisy |
and low-frequency range of human moans. Of the 26 pitch movements in the melody, 21 moved in stepwise motion and the other five were descending minor thirds, intervals that mimic the melodic contours of moaning. The timbre of the Adagio was relatively pure (string instruments have a wide range of timbres available, and those used in this recording are fairly pure wave-form sounds). The second piece was an excerpt from “The Fragile” by Nine Inch Nails. This excerpt contained the relatively pure waveform sounds of a softly played solo piano. The pitches changed at a rate of 35 notes per minute in regular rhythm. The music consisted of two simultaneous lines. The frequencies of the melodic line ranged from 196–587 Hz while the bass line frequencies range from 73–98 Hz. The two lines moved primarily in contrary motion with a descending melodic line (in the range of an adult female voice) accompanied by an ascending bass line (in the frequency range of an adult male voice). The four-note ascending scale in the bass was repeated four times, providing the accompaniment for four sentence-length melodic phrases.

The first tamarin affiliative vocalization-based music sample contained 52 notes per minute, approximating the rate of frequency variation in the normal long call and affiliation calls with frequencies ranging from 1.9–3.7 kHz. This piece was designed to capture the quality of the pure, high sounds common to the range and timbre of tamarin affiliation vocalizations. It was written in ABA form and played on the cello using the technique of artificial harmonics which produced a pure flute-like quality. It was constructed entirely of motivic pairs, since the most easily perceived pattern in music is the immediately repeated pattern and most mammalian vocalizations consist of repeated calls. All motives were either two or four note motives in regular rhythms in keeping with observed intragroup tamarin vocalizations that tended to consist of two, four, and eight note groups in regular rhythms. Most motives were descending to mimic the descending frequencies of calming vocalizations used by humans to control behavior in working animals.

The A section of the second tamarin affiliation piece in B flat consisted of two simultaneously sounding lines of artificial harmonics on the cello. One line was similar to the melody from the first piece and the second line included a musical representation of the rising frequency with diminishing amplitude motive common to affiliative tamarin vocalizations. The B section introduced a lowered sixth into the tonality and a drum-like pulse was created by thumping the cello body with the thumb. The tamarin heart rate was used as a basis for the tempo of this pulse.

**Fear/threat vocalization-based music**

We used an instrumental excerpt from “Of Wolf and Man” by Metallica as one human-based example of arousing music. The acoustic characteristics of this music include a strong bass and an enhanced broadband timbre similar to that found in human fear/threat vocalizations in a fast march tempo (120 beats per minute) and a minor modality, frequently using dissonant intervals of minor seconds and the tritone. The second piece was an excerpt from “The Grudge” by Tool. This excerpt contained pitch changes at the rate of 81 per minute and a pedal drum pulse of 220 beats per minute. The frequencies of the bass line range from 73–147 Hz and the upper line ranges from 220–262 Hz. Two guitars play repeated motives consisting of minor seconds and thirds using quick-onset picked notes in moderately complex timbre accompanied by simultaneously rhythmic pedal bass drum and tom-toms.

The first tamarin threat vocalization-based music was constructed in ABA form. The A section evokes the complex, broad band timbre of threats played on the cello *sul ponticello* (bowing near the bridge to add harshness to the sound) in a fast tempo (554 beats per minute) with ascending pitch crescendo motives that bear a musical similarity to the ascending pitch crescendo motives
found in fear/threat calls. These were accompanied by an underlying irregular rhythmic unvo-
calized aspirant “ch,” also a broad band timbre. The B section had two lines of music, each
consisting only of the notes B flat and B natural (7.5 and 7.9 kHz), creating quickly alternating
unison–minor second intervals. The second piece was based on three elements of mobbing
vocalizations: quick, descending intervals, long, broadband sounds, and antiphonal, chorused
calls that are often used when perceived threat is present. The form was AbbABCABB where
the second, third, and fourth statements of the B section were augmented with a countermelody
that, combined with the original melody, created dissonant harmonies. The broad band timbre
of the theme area in the B section was created with a rapid repetition of bow strokes (tremolo)
which is commonly found in “agitato” classical music. The modality of A minor was introduced
in the opening figure and in the first two notes in the B theme, and then the dissonant D sharp
dominated the melodic contour of the section. The C section was based on quickly repeated
notes and was comprised of four different lines, two of them with added tritone double stops.
The ascending crescendo motive became gradually higher and louder and the intervals between
the repetitions of the motive were shortened as seen in spontaneous fear/threat calls of many
animals (e.g., Schehka et al. 2007).

Playback tests to tamarins

We have described the methods and results previously (Snowdon and Teie 2010) and will pro-
vide a summary here. We tested seven pairs of tamarins with each of the eight stimuli (two each
of human affiliation-based, human threat-based, tamarin affiliation-based and tamarin threat-
based). All stimuli were edited to approximately 30 s duration, with variation allowing for chords
to be resolved at the end of each sample. Tamarins heard each stimulus only once. We gathered
behavioral data during a 5-min baseline period when the animals were calm and then presented
the 30-s stimulus and gathered data for an additional 5 min following the stimulus. We aver-
aged the data for the two examples of each of the four types of music. We found no differences
among the four types of music during the baseline or during the stimulus presentation. However,
in the 5 min after presentation tamarins showed significant increases in movement, anxious
behavior and social behavior to the tamarin threat-based music compared with the tamarin
affiliation-based music. When we compared responses in the 5 min following the stimulus with
baseline behavior, the affiliative-based tamarin music produced significant decreases in move-
ment, orientation and social behavior and a significant increase in foraging behavior. In contrast
the threat-based tamarin music led to significant increases in orientation relative to baseline.
There were no significant differences in behavioral responses between human affiliation-based
music and human threat-based music in post-tests, but there were two significant effects of
human music compared with baseline responses. Anxious behavior decreased following playing
of human affiliation-based music and movement decreased following playback of the human
threat-based music. Note that this latter result may be explained by the tamarin’s response to
pulse. The fast pulse of human threat-based music was the same as the resting heart rate of an
adult tamarin (note the 220 beats per minute pulse in “The Grudge” by Tool.)

Summary and implications

We have developed a theory for how aspects of music can influence the affective or emotional
state of a listener and we have shown that the musical features hypothesized to influence emo-
tions in humans can also be effective in a nonhuman primate. However, we had to create music
that was in the frequency range and tempo that is appropriate to the cotton-top tamarin.
Species-appropriate music was effective in eliciting the predicted behavioral responses from the tamarins, whereas human music was generally ineffective or influenced responding in a direction opposite to human responses. We did not explicitly test preferences in our study so we cannot directly compare our results to those of McDermott and Hauser (2007). However, we and others who have listened to the tamarin music do not find it particularly pleasant and one might suspect that tamarins might react to human music in a similar way. This has important implications for those who use music as an enrichment tool for nonhuman primates in laboratories or for pets during the workday of their owners. (We are currently testing domestic cats with human and cat-based music.) We cannot simply assume that our musical tastes will be appreciated by other species.

There are several additional implications resulting from our research. First, the results add to those that have shown that prosodic features of human speech can directly affect the behavior of nonverbal infants as well as companion animals (Fernald 1992; McConnell 1991). Second, we have found distinctly musical elements in the vocalizations of tamarins that parallel the features we had hypothesized to convey and influence emotional states in humans. Chaotic music is characteristic of fear and threat contexts and has features of complex overtones, dissonant intervals and irregular rhythms. Music expressing order is characteristic of affiliation and has features of simple wave forms, consonant intervals, and regular rhythms. Third, these same features, when incorporated into music at the pitch range and tempo appropriate for tamarins, has effects on tamarin behavior similar to what we would expect in humans, providing validation of our theory with another species. Because humans have a long individual history of exposure to music and most likely have developed strong preferences (tastes), it would be difficult to find a musically-naive human population on which to evaluate the theory, but testing another species with no prior exposure to music allows the test of theoretical predictions in a music-naïve population.

The similarities between human and tamarin music and the finding of musical structures in tamarin vocalizations suggests the emotional components of music may have an evolutionary history that predates humans and confirms the views of those authors who have hypothesized emotional communication as an early precursor to human music (Fitch 2006; Juslin and Västfjäll 2008; McDermott and Hauser 2005).

Finally, music can induce emotional contagion in listeners (Juslin and Västfjäll 2008) and the findings that music composed for monkeys can also induce emotional responses in tamarins, suggests that the vocal signals of animals may not merely have evolved to communicate the emotional state of the caller, but may also serve to induce an emotional contagion in listeners which may effectively manipulate or manage the behavior of listeners (Owings and Morton 1998; Owren and Rendall 1997). The emotional contagion of music can be made available to another species when the compositions reflect an understanding of the sonic templates of recognition of that species.

Acknowledgments

This chapter is an equal collaboration between both authors. In general, the aspects of this chapter that are related to human psychology and primate behavior and research are attributable to Snowdon and the theory is attributable to Teie. We are grateful for the critical comments of Eckart Altenmüller and an anonymous reviewer that have strengthened this chapter. We thank Elizabeth Abbs for testing the animals. Research supported by funds from the University of Wisconsin Graduate School Research Committee and a Hilldale Professorship to CTS.
Appendix

Our theory proposes that emotions stimulated by music may involve many subtle and subcon- 
scious recognition responses that occur concurrently and consecutively. Each individual emotional 
response involves its own neurochemical signal that may result in a summary hybrid emotion. In 
addition to these there may be a number of additional emotional responses related to learning and 
familiarity; the following list is meant to augment, not supplant other reasons that music elicits 
emotions. The principle of accumulated response is demonstrated in this example: there are nine 
independently recognizable acoustic features that may be capable of inducing attentive responses 
in these first four measures of the prelude in G for unaccompanied cello by J.S. Bach (Fig. 9.1).

1. The cello creates a resonance-enhanced periodic sound that is given priority treatment in 
auditory processing.

2. The repeated G resembles the pulse.

3. The notes are discrete and spaced in time intervals that are consistent with human speech as 
heard and learned in the womb.

4. The first three notes present consonant intervals that have been shown to induce positive 
emotional responses.

5. The F# in the third measure presents a dissonant interval that has been shown to induce 
negative emotional responses. The G is played on an open string and is still vibrating when 
the F# is played, resulting in both consecutive and concurrent dissonance.

6. The pattern of the melodic outline that is established in the first measure is repeated in the 
subsequent measures. Pattern recognition has been shown to trigger positive "aha" emotional 
responses.

Each of the listed features (1)–(6) heard in this music is an example of acoustic information 
that is capable of eliciting an emotional response that has been experimentally verified. The last 
three are untested, hypothetical emotional responses to perceived movement in music:

7. The movements of the interior and superior lines may be heard and followed as contrapun-
tal movements. Following concurrent movements has been shown to increase thalamic 
activity and attentiveness (O’Connor et al. 2002).

8. The change in direction in the upper line in measure (4) may be recognized visuospatially, 
triggering an attentive response. Reaction to change in direction has been shown to elicit an 
attentive emotional response (Saenz et al. 2002).

9. The symmetrical arch of each pattern and the first four measures may be perceived as per-
fected, celestial movement.

Figure 9.1 Suite for Unaccompanied Violoncello #1 in G, BWV 1007.
References


EMOTIONAL COMMUNICATION IN MONKEYS: MUSIC TO THEIR EARS?


REFERENCES
