Multimodal Affective Interaction: A Comment on Musical Origins

Steven Robert Livingstone
Information Technology and Electrical Engineering, University of Queensland

William Forde Thompson
Communication, Culture & Information Technology, University of Toronto

The rigors of establishing innateness and domain specificity pose challenges to adaptationist models of music evolution. In articulating a series of constraints, the authors of the target articles provide strategies for investigating the potential origins of music. We propose additional approaches for exploring theories based on exaptation. We discuss a view of music as a multimodal system of engaging with affect, enabled by capacities of symbolism and a theory of mind.

Key words: multimodal, music origins, music cognition, affect, theory of mind

Justus and Hutsler (2005) and McDermott and Hauser (2005) provide very sensible advice for investigating the bioevolutionary origins of music. Both articles highlight the evidentiary burden that should be placed on evolutionary theories of music, particularly those implicating processes of adaptation. The authors outline strategies of identifying musical properties that are innately determined, domain-specific, and uniquely human and suggest that evolutionary theories should be restricted to a consideration of such properties.

According to Justus and Hutsler, the most compelling evidence of adaptation for a cognitive domain is the demonstration of domain-specific innate constraints. If aspects of music have parallels in domains such as speech, then "there will always be the possibility that the mechanism in question originated exclusively through selection pressures in the other [domain]" (p. 16). This reasoning explains why there are unresolved debates about whether music originated from language, language originated from music, or music and language originated from a common ancestor that shared properties of both systems (Brown, 2000). McDermott and Hauser argue that theories of musical origins should focus on characteristics of music that are uniquely human. If such properties are observed in nonhuman animals, then they probably evolved as general-purpose mechanisms that were co-opted for use in music. The authors also assume that phenotypic similarity among species often reflects a common genotype.

Such strategies allow researchers to constrain theories of music as an adaptation to environmental pressures; they provide less guidance for theories that implicate exaptation or secondary adaptations. Early evidence for music is accompanied by evidence for other forms of human expression that may have been relevant to music in ways that are not obvious today (Cross 2001; Henshilwood et al. 2002; Mithen 1996; Morley, 2002). Moreover, the many uses of music across cultures and historic times suggest that music is not a purely auditory phenomenon. Even modern-day Western performers and listeners experience music in an integrated, multimodal way (Thompson, Graham, & Russo, 2005). Affective cues in music are integrated with those in accompanying gestures and facial expressions, and parallels exist in the nature of such cues in different domains such as melody and prosody. This overlap of cues in music and prosody suggest that these domains have a common ancestry (Brown, 2000).

More generally, parallels and convergences among domains raise the possibility that some properties of music are exaptations of an earlier system of affective communication that integrated processes from several domains within and beyond the auditory system. Although exaptations are phenotypic changes with no corresponding change in genotype (Justus & Hutsler, p. 4), strategies exist for identifying and exploring underlying behaviors and functions involved in exaptation. Most essential is the identification of musical properties that are observed across cultures and historic periods. Such analyses allow us to delineate the scope

Music Perception volume 24, issue 1, pp. 89–94, issn 0730-7829, electronic issn 1533-8312

of activities under discussion and prevent us from inadvertently defining music from the perspective of contemporary Western forms (Cross, 2003).

Looking beyond the surface features of Western music toward abstract processes and properties that connect the diversity of musical forms and behaviors, it is difficult to identify structural features or functions common to all music. Once we reach a point of abstraction that connects all musical phenomena, the resultant constructs describe behaviors and functions that extend beyond music to other domains such as speech, and to other modalities. As we grasp at greater levels of abstraction, there appears to be no clear point where all music is captured, but nothing else.

One of the most salient properties of music is its association with affect (Juslin & Sloboda, 2001), although not everyone agrees on the significance of the association. Cross (2003) has questioned conceptions of music that connect it necessarily with affect, arguing that its uses are not restricted to the communication or induction of emotion. However, while music is not always ostensibly about emotion, we believe it is inherently connected with affect and it is this connection that allows music to function flexibly over a wide range of contexts. Without affect as a mediating factor, the variegated uses of music would have to be explained by semantic links as in language, or iconic links as in pictures. Both explanations seem doubtful. Cross (2005) has reasoned that music has semanticity that is merely "non-consensual," but such a conception understates the stark difference between music and language in their capacity to represent ideas. Even music that performs a generic function such as *The Wedding March* from Wagner's opera *Lohengrin* is experienced in fundamentally different ways than verbal messages about marriage. Music is used in a variety of contexts not because it represents ideas directly, but because its capacity to express or induce nuanced affective states allows it to function flexibly across diverse contexts (e.g., funerals and religious ceremonies, sporting events, courting rituals).

If affect is pivotal to the various uses of music, then one might expect to observe a correlation between musicality and emotional intelligence. Emotional intelligence has been described as a set of abilities for perceiving, using, understanding, and managing emotions (Mayer & Salovey, 1997; Salovey & Grewal, 2005). *Perceiving* emotions, the most basic component of emotional intelligence, relates to the detection and decoding of emotion in faces, voices, and cultural artifacts. *Using* emotions involves the ability to utilize emotions to assist in general cognitive tasks and problem solving. *Understanding* emotions is the ability to conceptualize emotions and understand differences or connections between them (e.g., between sadness and grief). *Managing* emotions refers to the regulation of emotions in self and others.

Resnicow, Salovey, and Repp (2004) administered an emotional intelligence test to twenty-four listeners, who also identified the intended emotions of piano performances. The researchers observed a significant correlation between emotional intelligence and identification of intended emotions ($r = 0.54$). In another study, Thompson, Schellenberg and Husain (2004) examined the effects of music lessons on the ability to decode affective connotations of speech prosody. Six-year-old children were assigned to music lessons, drama lessons, or no lessons for one year, and assessed for their ability to decode emotional connotations of vocal utterances. Children assigned to the keyboard condition outperformed other children, suggesting that some forms of music lessons lead to enhanced emotional sensitivity. In the same study, musically trained and untrained adults identified emotions conveyed in tone sequences that mimicked utterance prosody. Trained adults ($M = 45\%$) were better than untrained adults ($M = 29\%$) at discerning the emotional connotations of the prosodic patterns.

An important step in examining the role of affect in music is to identify acoustic and structural aspects of music that are reliably associated with affective connotations. Figure 1 illustrates a cumulative analysis of structural (score-based) cues for musical emotion based on studies reviewed by Schubert, 1999, Gabrielsson and Lindstrom, 2001, and Gabrielsson and Juslin, 2003

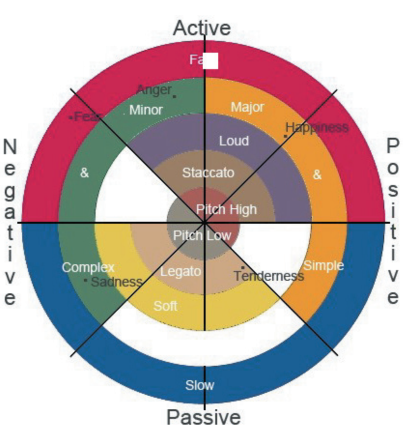

FIG. 1. Meta-analysis of affective cues in music. Cues near the periphery of the circle are based on the results of the greatest number of studies; cues near the center are based on the results of the fewest number of studies. As examples, "energetic" music tends to have a faster tempo (red, 20 studies) and "happy" music tends to be composed in a major mode (orange, 19 studies).

(adapted from Livingstone & Brown, 2005). Affect is represented in two dimensions: activity (arousal) and valence. The position of cues on each axis reflects the number of studies showing the association, and cues are represented in quadrants if the corresponding association was observed in multiple independent studies. Notable features of the meta-analysis include the observation that fast tempi are strongly related to activity but are associated with both positive and negative valence (Husain, Thompson, & Schellenberg, 2002; Thompson, Schellenberg, & Husain, 2001). The immediacy of cues as a means of affective communication is evident in their reflective symmetry. For example, a positively valanced work has a major mode and simple harmonies, while a negatively valanced work has a minor mode and complex harmonies.

Many affective cues have parallels in nonmusical and nonauditory domains, suggesting a broad scope of inquiry (Ilie & Thompson, 2006; Thompson & Balkwill, 2006). Such parallels raise the possibility that music is a carrier of abstract affective cues that are also instantiated in other domains and behaviors, from tone of voice to a slow-paced gait to a sequence of visual impressions. Affective interaction is not specific to music, but is relevant to many expressive forms, including visual arts, dance, and prosody. Identifying and defining the scope of affective cues remains a fundamental challenge that requires perceptual, cross-cultural, and cross-modal studies.

Several studies have examined cross-modal commonalities in the communication of affect. Krumhansl and Schenck (1997) exposed groups of participants to different dimensions of a choreographed work: dance only, music only, dance and music. Judgments of emotion were highly similar in the three conditions. In another study, Thompson, Graham, and Russo (2005) observed that judgments of the affective quality of music were influenced by the facial expressions of performers, and recent evidence suggests that auditory and visual cues to affect are integrated preattentively and unconsciously (Thompson & Russo, 2006).

Dissanayake (2000) and Cross (2003) have emphasized the multimodal nature of music. Cross stressed a form of "floating intentionality" of musics, in which "their 'aboutness' . . . might be 'about' different things at different times" (Cross, 2003, p. 81). We concur with the model proposed by Cross, but posit a central role of affect in musical origins. We also agree with Dissanayake's model of "rhythmic-modality" and its underlying affective motivation, but feel it is premature to make the strong claim that musical origins are tied specifically to genes selected for caregiver-infant interactions. Not only is evidence unsettled for musicality in mammals that demonstrate strong mothering instincts, but many complex mothering behaviors may be driven by experientially based mechanisms (Fleming, O'Day, & Kraemer, 1999). We propose that a general capacity for affective interaction, coupled with the emergence of a theory of mind, enabled a host of symbolic systems, including music. Musical aspects of caregiver-infant interaction are one outcome of these functions, and provide infants with early exposure to music-like structures.

The appearance of Homo sapiens 150,000 to 100,000 years ago marked a significant shift in human evolution, heralding an explosion in cultural phenomena. These developments coincided with the maturation of a complete "Theory of Mind" (ToM) (Baron-Cohen, 1999; Burns, 2004). A ToM refers to the cognitive capacity to "[understand] people as mental beings who have beliefs, desires, emotions, and intentions and whose actions and interactions can be interpreted and explained by taking account of these mental states" (Astington & Baird 2005, p. 3). This capacity provides a means of hypothesis testing for the behaviors of others. Its development is closely interrelated with linguistic capacity (Astington & Jenkins, 1999) and parent-child social interaction in joint-attention activities (Tomasello, 2003).

One of the significant consequences of a ToM is the ability to construct and employ symbols: abstractions or meta-representations (Suddendorf, 1999) of objects, people, social constructs, and their interrelations. This use of symbolism allowed for the development of symbolic language whereby representational elements could be grouped into larger, hierarchical structures. The capacity for symbolic thought led to the diverse range of systems that covers all cultural modalities: visual art, language, dance, religion, and music. While a ToM is often discussed as a model of complex social exchange (Mithen, 1996) and shared intentionality (Tomasello et al., 2005), it also provides a mechanism for reflective discourse. Through this capacity an individual can construct a statement or thought, and then make an evaluative statement regarding it (Lohmann et al., 2005). Mithen (1996) refers to this capacity as "reflexive consciousness." Great apes exhibit a precursory form of a ToM but lack the cognitive skill required for the exchange of psychological states, including emotions, experiences, and activities.

Music and related arts may be construed as systems of affective interaction and reflexive thinking. In this context, a ToM is a derivative of the Aristotelian position in which our experience with artworks is a form of affective sandboxing; a means of pursuing affective exploration / hypothesis testing in a safe environment (Davies, 2001). This contemplative process enables the refinement and progression of individual works and compositional styles, allowing the creator to assess and modify the affective quality of their creation. These processes affirm who and what we are (e.g., a heavy-metal enthusiast, a political activist, a Hindu) through affiliate communication and expressions of individuality. In infancy we lay the foundations of music with the learning of affective cues; but our ability to create music emerges from our capacity to engage with symbolic hierarchical systems, our desire for affective social communication, and a reflexive motivation to explore our identity.

Author Note

Preparation of this paper was supported by a scholarship awarded to the first author from the Australasian CRC for Interaction Design, which is established and supported under the Australian Government's Cooperative Research Centres Programme, and by a discovery grant awarded to the second author from the Natural Sciences and Engineering Research Council of Canada.

Address correspondence to: Bill Thompson, Room 3030, CCT Building, University of Toronto at Mississauga, 3359 Mississauga Road, Mississauga. E-MAIL *b.thompson@utoronto.ca*

References

ASTINGTON, J. W., & JENKINS, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35,* 1311–1320.

ASTINGTON, J. W., & BAIRD, J. A. (2005). Introduction: Why language matters. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 3–25). Oxford: Oxford University Press.

BARON-COHEN, S. (1999). The evolution of a theory of mind. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 261–277). New York: Oxford University Press.

BROWN, S. (2000). The "musilanguage" model of music. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

BURNS, J. K. (2004). An evolutionary theory of schizophrenia: Cortical connectivity, metarepresentation and the social brain. *Behavioral and Brain Sciences, 27*, 831–855.

CROSS, I. (2001). Music, mind and evolution. *Psychology of Music, 29*, 95–102.

CROSS, I. (2003). Music and evolution: Consequences and causes. *Contemporary Music Review, 22*(3), 79–89.

CROSS, I. (2005). Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical communication* (pp. 27–43). Oxford: Oxford University Press.

DAVIES, S. (2001). Philosophical perspectives on music's expressiveness. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 23–44). Oxford: Oxford University Press.

DISSANAYAKE, E. (2000). *Art and intimacy: How the arts began.* Seattle: University of Washington Press.

FLEMING A. S., O'DAY D. H., & KRAEMER G. W. (1999). Neurobiology of mother-infant interactions: Experience and central nervous system plasticity across development and generations. *Neuroscience and Biobehavioral Reviews, 23,* 673–685.

GABRIELSSON, A., & JUSLIN, P. N. (2003). Emotional expression in music. In R.J. Davidson, K.R. Scherer, & H.H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 503–534). Oxford: Oxford University Press.

GABRIELSSON, A., & LINDSTRÖM, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). Oxford: Oxford University Press.

HENSHILWOOD, C. S., D'ERRICO, F., YATES, R., JACOBS, Z., TRIBOLO, C., DULLER, G. A. T., et al. (2002). Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science, 295,* 1278–1280.

HUSAIN, G., THOMPSON, W. F., & SCHELLENBERG, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities: Re-examination of the "Mozart effect." *Music Perception*, *20*, 151–171.

ILIE, G., & THOMPSON, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23,* 319–329.

JUSLIN, P., & SLOBODA J. A. (2001). *Music and emotion: Theory and research.* Oxford: Oxford University Press.

JUSTUS, T., & HUTSLER, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception, 23*, 1–27.

KRUMHANSL, C. L., & SCHENCK, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae, 1,* 63–85.

LIVINGSTONE, S. R., & BROWN, A. R. (2005). Dynamic response: Real-time adaptation for music emotion. *ACM International Conference Proceeding Series, 123*, 105–111.

Lohmann, H., Tomasello, M., & Meyer, S. (2005). Linguistic communication and social understanding. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 245–265). Oxford: Oxford University Press.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–31). New York: Basic Books.

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

Mithen, S. (1996). *The prehistory of the mind.* London: Thames and Hudson.

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

Resnicow, J. E., Salovey, P., & Repp, B. (2004). Is recognition of emotion in music performance an aspect of emotional intelligence? *Music Perception, 22,* 145–158.

Salovey, P., & Grewal, D. (2005). The science of emotional intelligence. *Current Directions in Psychological Science, 14*, 281–285.

Schubert, E. (1999). *Measurement and time series analysis of emotion in music* (Doctoral dissertation, University of New South Wales). *Dissertation Abstracts International, 67*, 7.

Suddendorf, T. (1999). The rise of the metamind. In M. C. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 219–260). Oxford: Oxford University Press.

Thompson, W. F., & Balkwill, L.L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*(1/4), 407–424.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 156*(1/4), 177–201.

Thompson, W. F., & Russo, F. A. (2006, July). *Preattentive integration of visual and auditory dimensions of music.* Paper presented at the Second International Conference on Music and Gesture, Royal Northern College of Music Manchester, UK.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Mood, arousal, and the Mozart effect. *Psychological Science, 12*, 248–251.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46–64.

Tomasello, M. (2003). The key is social cognition. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 47–58). Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.