Embodied Music Cognition and Mediation Technology
By Marc Leman
Cambridge, MA: MIT Press, 2007

Reviewed by Peter E. Keller
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Petr Janata
University of California, Davis

Over the past 25 years our understanding of how music interacts with the human mind and brain has grown rapidly, and multimedia technologies have augmented the ways in which we engage with music. In *Embodied Music Cognition and Mediation Technology*, Marc Leman examines how these developments might be unified into something that is simultaneously a theory of music cognition and a blueprint for the music mediation technology of the future. Mediation refers to the mappings between the intentions and desires on the part of active musical participants and the technology that renders the music. The main mediating principle elaborated on in the monograph, which is more intellectual discourse than textbook, is rooted in the belief that musical interactions are socially charged, embodied affairs. Thus, individuals understand music in the same way that they understand others' intentions during social interaction, and expressive intentions are attributed to music because patterns of sonic energy evoke bodily gestures that are meaningful to an individual due to his or her personal history as an active participant within a cultural environment.

The first three chapters contextualize the embodied music cognition approach. In Chapter 1, Leman sets the scene by making clear the challenges that face those who are concerned with how subjective musical experiences are linked to physical sound patterns. Chapter 2 then deals with the diversity of paradigms that are relevant to the business of interdisciplinary music research. Here Leman adeptly identifies relationships between trends in music research, such as the emergence of systematic musicology, and landmark developments in the discipline of psychology, such as the advent of the Gestalt school and cognitivism. He also charts the progress made in the fields of technology, information theory, and computational modeling, with well selected philosophical matter visited along the way. This serves as a historical prelude to the birth of the modern embodied cognition paradigm, which asserts that, "knowledge does not emerge from passive perception, but from the need to act in an environment" (p. 43). In Chapter 3, Leman expands upon this ecological theme with a view toward music mediation technology, developing the premise that mediating technology should exploit the way in which individuals naturally engage themselves with music.

The ensuing chapters delve into the details of what embodied music cognition means. Chapters 4 and 5 build a case for why it makes sense to think about engagement with music in terms of corporeal articulations and action-based ontologies, and how these lead to pleasurable experiences with music, whereas Chapters 6 and 7 describe how this type of framework might play itself out in musical instrument and music retrieval technologies. Throughout these chapters, Leman articulates a framework in which performer/music/listener interactions can be structured/mediated. The framework contends with the formidable challenges inherent in mapping between the intentions, actions, and percepts of individuals and very specific musical signals. Ultimately, the problem is one of identifying relationships between semantics and musical structure, and then specifying the technological requirements for accomplishing the translation from one to the other.

Leman breaks the problem down into three interacting conceptual levels, which he talks about as first-person, second-person, and third-person descriptions. Third-person descriptions are objective representations of the structural features of the music, whereas first-person descriptions are subjectively assigned semantic labels that refer to expressive intentions. According to Leman, previous approaches to understanding music (e.g., traditional musicology) have fixated upon these two levels of description without giving adequate treatment to the "rules" that govern the mapping between objective representations and subjective interpretations. Such rules are needed to achieve his scientific goal of developing a complete theory of music, as well as his practical goal of developing a successful mediation technology. The key to Leman's solution is the proposal that an understanding of musical intentions requires third-person and first-person descriptions to be linked via second-person descriptions, which are corporeal in nature. At this intermediate level, expressive bodily gestures from an individual's repertoire of actions are used to describe moving sonic forms in a manner that the individual can interpret based on his or

Music Perception volume 26, issue 3, pp. 289–292, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2009.26.3.289

her personal experience of interacting with others in the world. On the one hand, this level relates dynamic musical features such as melodic trajectories, rhythm, and sound intensity to gestures, while on the other it links kinesthetic/synesthetic spaces to affective/emotive spaces via semantic descriptors. One should note that semantic descriptors actually imply relationships across the different levels. For example, describing a piece of music as "driving" connotes structural properties at the sensory level (fast tempo, repetitiveness of pitch, strong accentuation), at the gestural level (e.g., vigorous pounding of the air with a fist), and at the semantic/affective level by implying a high level of arousal.

Leman assumes that corporeal articulations arise in response to music (though actual movement can be inhibited) because the human motor system resonates with patterns of physical sonic energy. Consistent with the view that perception involves covert action simulation, Leman claims that this behavioral resonance occurs due to the tight coupling of action and perception. As he points out, the most compelling evidence for action/perception coupling comes from behavioral studies of imitation and neuroscientific studies of the so-called "mirror system" in the brain. The former have revealed that people are predisposed to produce actions that they observe (e.g., newborn infants imitate facial expressions) and the latter have identified a frontal-parietal cortical network that may drive this tendency. Interestingly, neuroimaging research published since Leman's book went to press has shown that the human mirror system is activated more strongly by complementary actions than by imitative actions (Newman-Norlund, van Schie, van Zuijlen, & Bekkering, 2007). This finding should be welcomed by Leman, as it can be taken to imply that the mirror system is able to support the mapping of different gestures to the same piece of music.

How and why does an understanding of musical intentions emerge from such a sensorimotor resonance system? With regard to why music is perceived to be meaningful, Leman simply (and reasonably) appeals to the human bias to attribute intentionality to things that move in biologically plausible ways. Musical sounds produced, or potentially producible, through human movement are clearly such things. The question concerning how specific semantic meaning is attached to particular patterns of sound is the most forbidding hurdle for any theory of music. Leman attempts to clear it by proposing that corporeal articulations evoked by musical sound patterns are subject to a process of "cognitive filtering" that yields linguistic descriptions based on the individual's action-oriented ontology. Such filtering involves the retrieval of semantic information about beliefs, values, and goals that are associated with similar gestures in the individual's repository of intentional actions. Through this process, musical sounds may "acquire the character of goals with valence (positive or negative character) and activity (high or low energy)" (p. 97). Thus, following Broeckx (as cited in Leman, 2007), Leman claims that music becomes a virtual social agent with whom an individual can interact in a meaningful and affective manner. Furthermore, the individual may come to feel that he or she is generating the music to the extent that inferred corporeal articulations appeal strongly to one's own action-oriented ontology. Deeply immersive, pleasurable musical experiences may stem from the feeling that one is interacting with a familiar social partner. This process might be considered a close cousin of processes underlying a perception-action model of empathy (Preston & de Waal, 2003), with the main difference being that music serves as an intermediary between two individuals.

The notion that empathic involvement with music has social overtones provides the vehicle with which Leman escorts his readers into the realm of pleasure. He claims that interactions with music engender feelings of social connectedness, and provide a training ground for the process of forming social bonds. These bonds are cemented because the mirroring that underscores social interaction—whether it is real or virtual—fosters liking between interaction partners (as much work on nonconscious mimicry has demonstrated). Thus, Leman transports Aristotle's well-known observation that imitation is pleasurable into the music domain. Placing Aristotle on stage reveals nicely that aspects of Leman's theory that may, at first glance, appear to consist of "New Age" concepts are in fact rooted firmly in antiquity. Leman does, however, entertain some options regarding pleasure that apparently were not recorded by the ancients. One such idea is that the link between imitation and pleasure may be related to the role of the former in seduction. In short, music may provide grounds for practicing in isolation the moves usually made to "evoke interest" in social contexts. Although Leman's discourse on music and pleasure concentrates upon such fairly sophisticated social factors, some more solitary and primitive alternatives could play a role. For example, it has been claimed that the low, loud sounds encountered in techno dance music evoke vestibular responses that are intensely pleasurable (Todd & Cody, 2000).

For Leman, the promise of new technologies lies in the ability to mediate interactions between the participant and the music across the first-, second-, and third-person levels of description. Chapters 6 and 7 describe how this interaction might be viewed as a multimodal experience by engaging all of the senses through the use of multiple

media that support an immersive musical experience. In Chapter 6, Leman offers a detailed example of a mapping between a performer's movements and a listener's corporeal interpretation of the music. The performance of a brief, ~30 s, piece of Chinese guqin music is analyzed with respect to the performer's movements and the resulting acoustic cues. Two participants' manipulations of a sophisticated joystick device as they listened to the excerpt are also analyzed and related to the sonogram of the performance. These detailed analyses highlight two of the formidable and somewhat problematic issues that are largely side stepped in this book. One of these is the tremendous effort required to annotate, in this case manually, the performance details associated with every note. If a claim is to be made, quite reasonably, that communicative aspects of musical experiences need to be represented at this level of detail and understood as combinations and sequences of gestures that carry the intentions, the task of describing every snippet of the millions of pieces of music that are available for search and retrieval is prohibitively immense, though principles of large-scale human computation could be applied to this problem as demonstrated by the Tagatune game in which pairs of players enter semantic descriptors of a piece of music they are listening to with the objective of determining whether they are listening to the same piece (Law, von Ahn, Dannenberg, & Crawford, 2007).

The second significant issue is highlighted by the analysis of the gestures made by different listeners as they listened to the same piece. Leman acknowledges that music interaction systems of the future need to take into account the individual, but the significance and importance of this challenge are underestimated. In the example provided in Chapter 6, the velocity profiles of each individual's gestures were relatively consistent across two repetitions of the melody, but the differences between the two listeners were considerable. If corporeal articulations are to be a major component of technology for searching through musical databases, how is one to contend with the variability in how individuals move with music? One need only imagine dancers on a dance floor. In spite of the fact that every individual is hearing the same music, the range of evoked movements is considerable. In other words, a many-to-one mapping exists between the corporeal expressions of individuals and the single piece of music. Considering the multitude of many-to-one mappings that will exist across a musical repertoire, as well as the likely limitations in movement repertoires of individuals, it is difficult to envision how a requisite degree of mapping specificity can be achieved for the purpose of music retrieval applications.

Chapter 7 describes how the themes of the earlier chapters might play themselves out in music retrieval applications, examining how the options that users have for structuring their music search queries might be broadened. A scheme is envisioned in which the metadata that traditionally accompanies music in databases, such as composer, artist, song title, or genre information, can be augmented with tags that describe dynamic properties of the music, where these dynamic properties reflect both an action-based ontology and affective descriptors. Leman draws on Laban's theory of effort and descriptions of forms of movement (Laban & Lawrence, 1947), and Broeckx's idea of relating structural descriptors of the environment to musical cues, as examples of how a listener who is unable to use the technical jargon of music might use semantic descriptors that are more accessible and intuitive. Here, as in earlier sections of the book, Clynes' ideas (Clynes, 1977, 1992) regarding the use of patterns of expressive touch as a window into emotion and music seem underrepresented, though they are strongly coupled to major themes in this book.

Leman describes his approach variously as a "theory," a "powerful model," and—in his conclusion—a "working hypothesis," acknowledging that the proposed basis of musical communication in intended actions is not yet an established fact. Indeed, depending on the desires and intentions of the reader, the book can be seen as mediating the transfer of knowledge about music, cognition, and technology with varying degrees of depth and precision. Nevertheless, the coherence of the book's central message is likely to prevail regardless of the reader's background and predilections, thanks, in large part, to the dovetailing of ideas in helpful preambles and summaries that flank major sections. A practical advantage of reiterating the overarching framework throughout the book is that it is possible to read chapters in isolation. Overall, Leman's embodied approach to music cognition is timely, given that likeminded approaches have recently been applied successfully in domains such as motor control (Jeannerod, 2006), language comprehension (Glenberg, 2007), and social interaction (Knoblich & Sebanz, 2008), and it heralds an exciting new era for music cognition and technology.

Correspondence concerning this article should be addressed to Peter E. Keller, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, Leipzig 04103, Germany. E-mail: keller@cbs.mpg.de or Petr Janata, Center for Mind and Brain, University of California, Davis, 267 Cousteau Place, Davis, CA 95616. E-MAIL: pjanata@ucdavis.edu

References

Clynes, M. (1977). *Sentics: The touch of emotions*. New York: Doubleday/Anchor.

Clynes, M. (1992, September). *Time-forms, nature's generators and communicators of emotion.* Paper presented at the Institute of Electrical and Electronics Engineers International Workshop on Robot and Human Communication, Tokyo, Japan.

Glenberg, A. M. (2007). Language and action: Creating sensible combinations of ideas. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 261-370). Oxford: Oxford University Press.

Jeannerod, M. (2006). *Motor cognition: What actions tell the self*. New York: Oxford University Press.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*, 2021-2031.

Laban, R., & Lawrence, F. C. (1947). *Effort*. London: Macdonald & Evans.

Law, E., von Ahn, L., Dannenberg, R., & Crawford, M. (2007, September). *Tagatune: A game for music and sound annotation.* Paper presented at the International Conference on Music Information Retrieval, Vienna, Austria.

Leman, M. (2007). *Embodied music cognition and mediation technology.* Cambridge, MA: MIT Press.

Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M., & Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience, 10*, 817-818.

Preston, S. D., & de Waal, F. B. M. (2003). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1-20.

Todd, N. P., & Cody, F. W. (2000). Vestibular responses to loud dance music: A physiological basis of the "rock and roll threshold"? *Journal of the Acoustical Society of America, 107*, 496-500.