BOOK REVIEW

***The Psychology of Music in Multimedia*,** edited by **Siu-Lan Tan, Annabel J. Cohen, Scott D. Lipscomb, and Roger A. Kendall. Oxford, New York: Oxford University Press, 2013, 432 pp., ISBN: 978-0-19-960815-7. $60.00**

Reviewed by Pantelis N. Vassilakis, Ph.D., Columbia College Chicago, Chairman, Audio Arts & Acoustics Department

Numerous colleagues and I have been eagerly anticipating the publication of this volume since we first became aware of its preparation at relevant academic conferences. The result is impressive. The contributors represent a wide variety of disciplinary and conceptual approaches to the topic and the work provides rigorous substance to the ever-growing realization that the presence of an image changes what we 'hear' and the presence of a sound changes what we 'see.' A total of 20 authors contribute 17 chapters that bring together research findings relevant to audio-visual perception in a cohesive, intelligent, and critical manner, drawing meaningful inferences, and outlining important follow-up research questions.

The Psychology of Music in Multimedia is well-edited, laid out, and contextualized, filling several gaps in the related literature. It represents a first systematic attempt to bring together the vast majority of significant research (primarily empirical) addressing cross-modal perception of audio-visual composites, in a volume that is quite ambitious in scope. The four editors have gone to great lengths in the book's opening and closing chapters to situate each contribution in the context of the others, tease out common threads or potential conflicts, and make a convincing and critically alert argument for the conceptually common objectives of a rather diverse field of inquiry. Beyond successfully articulating the endeavor's accomplishments and coherent identity, the editorial chapters also present a model for future editorial work willing to self-reflectively and critically expand from providing intelligent summaries into distilling potential new knowledge, not explicitly springing from any of the individual contributions alone. The contributing authors, who include all editors, have gone to analogous lengths to provide focused, detailed presentations and analyses of their specific topics, whether on modeling, exploring, or analyzing audio-visual composites' creation, presentation, and perception/cognition. Furthermore, each chapter engages in a review of a substantial segment of the related literature, with the combined "References" sections constituting a valuable resource in itself.

As a whole, the work makes an earnest effort to engage both the empirical and the reflective/speculative research branches of the overarching topic in a productive 'conversation.' This is largely explicit in the Foreword, provided by Nicholas Cook, and in the opening and closing editorial chapters, but less so in the remainder of the text. In what the editors assess as a reflection of the "scope and infancy of the entire endeavor" (p. 392), contributed chapters illustrate an overall subtle awareness of the rest of the contributors' work and a still non-uniform awareness/appreciation of directly or indirectly relevant research. The occasional 'local' absence from specific chapters of direct reference to key relevant works is one manifestation of this observation (*e.g.* absence of D. Huron's work from Chapter 4, A. D. Patel's work from Chapters 5 & 8, C. L. Krumhansl's work from Chapter 11, or S. M. Eisenstein's work from Chapters 10 & 11). The benefit of a bird's-eye view enables the editors to set up a highly coherent stage in their introductory chapter, presenting significant, interrelated directions and questions. Many return in the closing editorial chapter to remind us that, even though all questions

posed in the introduction may have, at some level, been explored, many are still open and still important.

The table of contents clearly reflects the organizational principles that bring this volume together, the opening chapter further elucidates and justifies these principles, and both are freely available on the publisher's website, along with an extensive *Companion Website* to several of the chapters. The latter [http://www.oup.co.uk/companion/tan] has grown from approximately 60 files, at the time the book went to press, to over 130 files, at the time of the review. It includes audio-visual materials, sound files, and animations, primarily drawn from the laboratory stimuli used in some of the empirical studies described in the volume. In addition to enhancing the effectiveness of the ideas presented in the book, this online companion serves as a valuable teaching and learning tool for anyone seeking a deeper understanding of empirical investigations of music in multimedia.
The editorial opening and closing chapters, Chapter 1, '*Introduction,*' and Chapter 17, '*Future research directions for music and sound in multimedia,*' frame rather independent contributions organized into four parts (the editorial closing chapter constitutes *Part V* "*Future Research Directions*"). Whether from the behavioral, cognitive, neurophysiological, musicological, social, or other perspective represented in this volume, all relevant research efforts attempt to ultimately contribute to the tackling of a single larger question: "*what are the bases of our intellectual and emotional responses to audio, visual, and audio-visual experiences?*" Every chapter in the volume offers important insights into this question, representing a rich variety of perspectives.

Part I "Models and Multidisciplinary Perspectives" *(Chapters 2-6)*

Three of the four editors contribute to *Part I*, a reflection of the theoretical considerations and empirical standpoint that motivated the creation of this volume. Annabel J. Cohen's Chapter 2, '*Congruence-Association Model of music and multimedia: Origin and evolution*' contrasts associationist (largely top-down) and Gestalt (largely bottom-up) theoretical perspectives and incorporates them in the Congruence-Association Model (CAM) to address the meaning and structural dimensions of media cognition. In walking the readers through the model's >20-year-long continuing development and application in a large variety of experimental contexts, the author also reveals a more general path towards addressing the increasingly complex theoretical considerations confronting empirical research on the topic. The evolution of CAM, currently in its fourth incarnation, reflects a progressive concern with and ability to move from additive/synchronic/passive to relational/diachronic/active contexts. To the reviewer, two of the notable contributions of this development relate to a) addressing the synchronic-diachronic tension that confronts our intellectual and affective responses to any experience and b) explorations of music's contributions in diegetic versus non diegetic narrative contexts.

In Chapter 3, '*Experimental semiotics applied to visual, sound, and musical structures*' Roger A. Kendall and Scott D. Lipscomb introduce and empirically explore a model that addresses the associationist and Gestalt perspectives not in the context of a dichotomy but as points on a referentiality continuum, outlined by what they refer to as "index," "icon," and "syntax." Influenced by semiotics, the key advantage of this approach is in its increased ability to explore the experienced close relationship between meaning and structure, where interpretation of a film's meaning relies not only on pre-existing or story-defined associations but also on cross-modal contour resemblances/analogies and on syntactical features of temporally organized filmic elements.

Mark Shevy's Chapter 4, '*Integrating media effects research and music psychology,*' shifts the focus on processes underlying audience demand and preference for certain kinds of media, approached

from the perspectives of communication theory, psychology of social behavior, and media studies. As the author notes, investigations on the psychology of music in multimedia can only benefit from more extensive incorporation of insights gained within media studies research and from the richness of everyday life situations and concerns such research offers as context.

In Chapter 5, '*Musical analysis for multimedia: A perspective from music theory,*' David Bashwiner presents a rather comprehensive and convincing account of the numerous insights gained through music theory's score-based approach to describing, analyzing, and understanding music's contributions to films. Implying that a musical score (*i.e.* a symbolic abstraction) is "music itself," may be contestable. However, as this chapter demonstrates, the value of exploring the question(s) of interest also from within the notational frame of reference is uncontestable. Differences observed in the methodologies and raw materials employed by music psychology and music theory investigations support the author's call for deeper interdisciplinarity, a call that is an underlying secondary theme of the entire volume.

Chapter 6, '*Emotion and music in narrative films: A neuroscientific perspective,*' is co-authored by Lars Kuchinke, Hermann Kappelhoff, and Stefan Koelsch, contributing expertise in the study of emotion, film theory, and brain imaging respectively. An excellent editorial choice on placement, this chapter offers an interdisciplinary "response" to the explicit calls for interdisciplinarity in the two preceding chapters. It provides a fascinating account of the neurophysiological mechanisms associated with visual, auditory/musical, and multimodal processing, and an inspiring attempt to link our knowledge about these mechanisms to issues of emotion, within the context of filmic experiences. In their closing theoretical notes, the authors well argue for the need to expand explorations of engagement and absorption in films/multimedia to include examination of the psychological concepts of empathy and embodiment. An exciting conclusion to the first part of an exciting book!

Part II "Cross-Modal Relations in Multimedia" *(Chapters 7-9)*

In Chapter 7, '*Perceived congruence between auditory and visual elements in multimedia,*' Shin-ichiro Iwamiya examines the relationship between judged meaning and perceived congruence (*i.e.* degree of formal and semantic matching appropriateness) of audio-visual composites, both generally and with respect to audiences' cultural backgrounds. He proceeds by considering the dynamic relationship between aural and visual pattern time-variance and the relationship between degree of semantic congruence and perceptual tolerance to formal incongruence in crossmodal contexts. The latter explorations are rather introductory, functioning more as teasers to important questions on the communication potential of formal incongruence, whether introduced as significant onset delays or through the pairing of opposing aural and visual contours. The author also reviews a sizable portion of relevant research, previously available only in Japanese.

In Chapter 8, '*How pitch and loudness shape musical space and motion,*' Zohar Eitan explores correspondences of sonic and spatial features, both in static (the term "steady" may have been more appropriate, for its temporal implications) and dynamic (*i.e.* time-varying) contexts. Rather than 'in' multimedia, sound and music are approached 'as' multimedia, with the author revealing the clearly perceived and consistently communicatable spatial and kinetic attributes of both. Throughout the chapter, kinetic attributes of sound are discussed in terms of pitch and loudness contours (*i.e.* temporal patterns of pitch/loudness changes), of which steady pitch/loudness is a special case. In this context, the perception of pitch and loudness, whether steady or time-variant, is essentially temporal and closely related to the perception of rhythm, understood by the reviewer not as a distinct entity but as a quality emerging through the layering of, for example, pitch, loudness, timbre, and/or duration contours. The author's closing comments on the temporal nature of pitch, loudness, and rhythm appear

to originate from a different theoretical standpoint on rhythm and, for the reviewer, cloud rather than clarify the chapter's many contributions.

Chapter 9, '*Cross-modal alignment of accent structures in multimedia,*' by Scott D. Lipscomb, reviews research on the determinants of salient moments in the aural and visual modalities, including a body of empirical research that has ventured into experimental contexts complex enough to validly represent real-world multimedia experience. The primary objective is to systematically explore the perception of cross-modal accent structure alignment through time. Issues, therefore, of the relationship between formal and semantic/associative audio-visual congruence (also addressed in Chapters 2, 3, & 7) return. They are accompanied by insightful commentary on ways in which complex and extended experimental contexts necessitate modifications to our understanding and discussion of accent structure alignment in multimedia. The latter opens up possibilities for shifting the focus from accent structure alignment and fitness, to accent structure relationships and range of possible effects/interpretations.

Part III "Interpretation and Meaning" *(Chapters 10-11)*

Chapter 10, '*Music videos and visual influences on music perception and appreciation: Should you want your MTV?,*' by Marilyn G. Boltz, explores ways in which visual information influences our perception of sonic/musical stimuli, in one of the rare contexts (music videos) where an audio-visual composite's sonic elements may assume a primary role. Visual features in the performers' physical appearance, gestures, and degree of ensemble coordination are examined for their impact on audience perception of musical performance and of the performers' musical ability. The author also reviews the theoretical foundations of audio-visual interactions, at once echoing, complementing, and proceeding rather independently from analogous reviews in the rest of the book. She then focuses on how information from both modalities may combine in a single interpretive judgment and affective response, a subject returning in the volume's penultimate chapter. Of particular interest are the author's closing comments on a) "ironic contrasts," a semantically/affectively productive outcome of incongruence that is more likely to arise in diegetic rather than non-diegetic sonic contexts, and b) the perception/interpretation of partial incongruence in complex audio-visual contexts.

In Chapter 11, '*Music and memory in film and other multimedia: The Casablanca effect,*' Berthold Hoeckner and Howard C. Nusbaum discuss the observed higher potential of music to cue visual recall rather than the other way around, in the context of a broader exploration on the role of memory during a film's narrative construction and experience. The authors' investigation of the various types of memory involved in audio-visual information processing is supported by discussions of the analogously nonlinear and fragmented representation of time in films and in all human experience. The chapter closes with stimulating reflections on the integration of an experience's sensory elements into a single memory and on the power of schemata (*i.e.* knowledge structures based on previous experience) to influence both the experience and the memory of an event. Potential future directions suggested by the authors can, in parallel, also acknowledge the converse: the continuous power of every new experience and of event recall to influence schemata.

Part IV "Applications: Music and Sound in Multimedia" *(Chapters 12-16)*

Chapter 12, '*Children's media: the role of music and audio features,*' by Sandra L. Calvert, explores the impact of sonic/musical features on attention, comprehension, and memory of information presented in multimedia, given young persons' routinely fragmented focus on prolonged tasks/experiences. She discusses the creative opportunities presented by the availability of user-friendly multimedia authoring tools, which can support developmental and cross-cultural research on

multimedia perception, and closes by reminding readers of the important role of silence, as a temporal space for reflection and imaginative activity, throughout youth's developmental stages.

In Chapter 13, '*Playing with sound: The role of music and sound effects in gaming*,' Mark Grimshaw, Siu-Lan Tan, and Scott D. Lipscomb begin with a description of the still evolving technological developments on computer and video-game audio. This is followed by a thorough review of relevant research that addresses the effects of sound on player performance, physiological responses, and overall game experience, revealing the variety of roles fulfilled by music and sound effects in interactive games. The increasing sonic complexity of games appears to earn and demand increasingly more of the player's attention. Game designers rely more and more on purely sonic cues to interact with players, whose performance needs to, in turn, increasingly rely on sound and music perception acuity. The authors devote appreciable time exploring ways to improve the quality and reliability of data from research on players' gaming experiences, highlighting the potential impact of such improvements on the interactivity effectiveness of future games.

Mark Shevy and Kineta Hung focus Chapter 14, '*Music in television advertising and other persuasive media*,' on multimedia contexts specifically designed to influence behavior. Although music's importance is not contested, the authors acknowledge that understanding and systematically predicting its specific impact remains a complex question. They explore the *Elaboration Likelihood Model* as an overarching framework that can support addressing this question in a much more nuanced and sophisticated manner than analogous efforts from around the first half of the last century. In addition to identifying the problems involved in seeing music's role as one of simply confirming or hindering persuasive messages already set through textual or visual information, the authors call for future research that addresses persuasion in the increasingly prevalent interactive context of social media.

Chapter 15, '*Auditory icons, earcons, and displays: Information and expression through sound,*' by Agnieszka Roginska, addresses auditory displays as means of conveying visual and affective information through sound. She reviews work that takes advantage of auditory processing mechanisms to discover patterns in and map complex data sets that have been converted into sound signals via sonification algorithms. Such algorithms provide for environments that support interaction, feedback, alerting, monitoring, navigation, and reality augmentation, which may alternatively be accomplished through *earcons* (*i.e.* rather arbitrarily selected sonic substitutes to non-sonic information) or *auditory icons* (*i.e.* sonic analogies to non-sonic information, with some natural, whether formal or associational, resemblance to that information). The author concludes by highlighting the potential of such research to maximize communication of information and interaction possibilities with devices that involve continuously shrinking visual displays.

With motion pictures providing the primary context, Mark Kerins begins Chapter 16, '*Understanding the impact of surround sound in multimedia*,' with a brief history of multichannel sound. This is followed by an extensive literature review of relevant empirical studies, exploring the impact on our experience of multimedia of auditory presentation mode in general and of multichannel sound presentation manner in particular. The author carefully distinguishes between "being able to tell the difference" and "having a distinctly different experience," when perceptually comparing film scene instances presented with single- versus various types of multi-channel sound. He also proposes an understanding of the audio-visual experience that, in the reviewer's interpretation, is ontologically different from the experience of its aural and visual components in isolation, and perceptually emerges during their temporal intersection. This is an intellectually stimulating and, from an analytical perspective, controversial position that provides *Part IV* with a closing as exciting as that of all preceding parts.

The themes and data presented in *Part I*, return throughout the book and pave the way for eventual explorations that can further address issues such at the following, distilled by the reviewer:

a. music's contributions to the productive tension inherent in the synchronic-diachronic dialectic confronting the experience of all time-based media or, for that matter, all human experience;
b. the meaning generation potential of blurring the viewers' perception of the boundaries between diegetic and non-diegetic sonic contexts;
c. "iconic" or "syntactical" relationships in the case of scenes with minimal visual movement/changes but rich "internal" or implied action;
d. the manner and contextual framework(s) within which an audio, visual, or audio-visual component's temporal structure and degree of complexity may support eliciting, versus suggesting, versus pointing to an emotional response;
e. the communication and affective potential of audio-visual temporal misalignments, beyond questions of fitness within isolated audio-visual composites;
f. the relationship of the music's and the story's perceived temporal unfolding to film music's apparent ability to support a perceptual back-and-forth among different levels of narrative and experiential time.

To effectively tackle such questions, empirical investigations may have to tip the "reliability versus validity" scale, which all empirical research routinely struggles to balance, towards the "validity" side. Such a move will involve reduced possibilities for unilateral interpretations of findings and for (statistical) inference. In the reviewer's opinion, this is not necessarily a negative, considering that the power of the experiences we are trying to elucidate does itself rest on their potential fluidity and tendency to resist consistent readings. This observation is far from the customary "laments" of non-empiricists on the limits of analytical and empirical investigations of systems. As the reviewed book illustrates, such investigations can be enriched and empowered by embracing multiple frames of reference, incorporating converging methodologies, and exploring dynamic definitions of what constitutes "a part" and, consequently, relationships of parts, within a system. Rather, this observation reflects the reviewer's suspicion that we may still be exploring just a subset (*i.e.* film as an '*audio-visual composite*') of the system responsible for the intellectual and emotional responses we are trying to elucidate (*i.e.* film as '*story*'). In their closing paragraphs, the editors highlight the need for highly competent interdisciplinarity in all relevant future endeavors, a position echoed and exemplified throughout the book. If the system of interest is indeed films/multimedia as *'stories,'* such interdisciplinarity will also have to involve full (re)examination of relevant speculative literature (*e.g.* Eisler, 1947; Gorbman, 1987; Chion, 1994; Brown, 1994) but also further, first-hand engagement of broader literature on story-telling, interpretation, and the human experience of time (*e.g.* Gadamer, 1989; Grondin, 2008; Ricoeur, 1988). This may be particularly necessary in the case of current film-music theoretical postulates on a) narrative construction, as it relates to the synchronic-diachronic dialectic, or b) the nature and productive potential of "*absorption*" in a story, understood not in terms "lulling" (*e.g.* Gorbman, 1987) but in terms of engagement, empathy, and embodiment.

The reviewed book's empirical focus on the topic is unique, its engagement of both speculative and empirical research, within a variety of disciplines, is commendable, and its publication could not have been timelier. It captures and brings to the forefront an academic and professional momentum of interest in cross modal perception of audio-visual composites that permeates research and application endeavors in a large variety of contexts, including immersive environments, simulations, film, gaming,

product design, and marketing. The result is a resource I consider a must-have for experts, students, and practitioners of the topic alike. The same would be the case for an eventual follow-up publication that would further take up the challenges explicitly or implicitly identified throughout the current volume.

References

Brown, R. S. (1994). *Overtones and undertones*. Los Angeles: University of California Press.

Chion, P. (1994). *Audio-Vision: Sound on screen*. C. Gorbman Ed. and Trans. New York, NY: Columbia University Press (originally published in 1990).

Eisler, H. (1947). *Composing for the Films*. New York, NY: Oxford University Press.

Gadamer, H. G. (1989). *Truth and Method* (2nd ed.). J. Weinshamer and D. Marshall Trans. New York, NY: Continuum (originally published in 1960).

Gorbman C. (1987). *Unheard melodies: Narrative film music*. London: BFI Publishing.

Grondin, S. (Ed.). (2008). *Psychology of Time*. Bingley, UK: Emerald Group Publishing Ltd.

Ricoeur, P. (1988). *Time and Narrative, Vol. #3*. K. Blamey and D. Pellauer Trans. Chicago, IL: The University of Chicago Press.