Modeling Multimedia Cognition: A Review of Nicholas Cook's *Analysing Musical Multimedia*

Scott D. Lipscomb

It is perhaps surprising that a text published over five years ago should warrant a series of reviews in a respected scholarly journal. This occurrence suggests at least two very important outcomes. First, "multimedia"—specifically, the role that music plays within such a context—is finally being given the attention it deserves as a sociologically relevant artifact of contemporary culture, thus worthy of discussion in scholarly music journals. Second, Nicholas Cook's *Analysing Musical Multimedia* has made a significant contribution to this dialogue in its emphasis, as evident from the book's title, upon the musical component of this multimodal experience.

Cook's text does not serve, however, to initiate the analysis of musical multimedia. In fact, the practice of combining music and drama dates back millennia to the Greek dramas of Aeschylus, Euripides, and Sophocles and can be traced throughout the evolution of Western civilization, as represented in the Medieval sacred drama, courtly displays of the Renaissance, Baroque opera, Wagner's *Gesamtkunstwerk*, and the development of sound film in the 20th century. Beginning in the 1950s, with interest intensifying during the most recent two decades, music researchers and psychologists have begun to investigate empirically the relationship between hearing and seeing...sound and image. As I have noted elsewhere,[1] in the field of perceptual psychology, interaction between the aural and visual sensory modalities is well-documented.[2] Empirical studies investigating the intermodal relationship in more ecologically valid contexts was initiated in the middle of the 20th century, but did not begin to attract significant attention until the late 1980s. Using a simple "drop the needle" technique, Tannenbaum discovered that music does influence

---

[1] Lipscomb, in press.

[2] See, for example, McGurk and MacDonald 1976, Radeau and Bertelson 1974, and Staal and Donderi 1983.

verbal ratings collected from participants following a dramatic presentation, whether live on stage, in a studio-taped version, or in a video recording of the live performance.[3] Using an industrial safety film depicting three accidents, Thayer and Levenson found that skin conductance level, a physiological measure, varied significantly between two conditions, one using a series of mildly active major seventh chords ("documentary music") and one using a repetitive figure based on diminished seventh chords incorporating harsh timbres ("horror music").[4] Marshall and Cohen, a study that Cook uses as an empirical basis for his own model of musical multimedia and its cognition, found that the information provided by a musical soundtrack significantly affected judgments of personality attributes assigned by subjects to each of three geometric shapes presented as "characters" in the film.[5] Based on the results of this investigation, the authors proposed a paradigm to explain the interaction of musical sound and geometric shapes in motion entitled the "Congruence-Associationist model." They assumed that, in the perception of a composite A-V presentation, separate judgments were made on each of three semantic dimensions (i.e. Evaluative, Potency, and Activity)[6] for the music and the film, suggesting that these evaluations were then compared for congruence at a higher level of processing. Since the publication of Cook's text in 1998, Annabel Cohen has gone on to expand the model, significantly clarifying the multi-level relationships that occur between sensory modalities.[7]

Cook's text is divided into two parts. The first half of the book provides a foundation for the theoretical framework proposed by the author. The remainder of the text consists of three analytical case studies or exemplars to which this specific framework is applied. As a general outline, this organizational structure is extremely clear and provides the reader a functional and concise method of analysis with examples of its practical application. A

---

[3] Tannenbaum 1956.

[4] Thayer and Levenson 1984.

[5] Marshall and Cohen 1988. For a detailed discussion of these studies and other related work, see Lipscomb 1995 or Lipscomb and Kendall 1994.

[6] Osgood, Suci and Tannenbaum 1957.

[7] Cohen 2001.

detailed analysis of the actual content, naturally, reveals numerous possibilities for further discussion or debate, a continuing process for which Cook has provided an excellent starting point. I have had the opportunity to use *Analysing Musical Multimedia* as a textbook in a graduate-level selected topics course on "Multimedia Cognition" (MUS_THRY 335-0) at Northwestern University. Cook's text, Michel Chion's excellent *Audio-vision*,[8] and a course reader including a variety of theoretical and empirical works related to the multimedia experience provided an excellent triumvirate upon which to build knowledge and facilitate discussion about the multi-modal experience.[9] Given my own background and experience, the present review of Cook's text will represent a dual perspective: that of a music/multimedia researcher and a university professor.

The opening section of Part I introduces many of the concepts that Cook deals with in the following chapters. To demonstrate the manner in which music can influence (or determine) the meaning of a sequence of visual images, several highly creative commercials are deconstructed according their content, both visual and auditory. In my opinion, this is one of the most valuable sections of the text, clearly demonstrating the extremely important role that music plays in this context and hinting at issues of congruence between the audio-visual (A-V) components...an element that will come to play a defining role in Cook's paradigm. I found myself frustrated, at times, reading about the selected examples—completely unfamiliar to me—and wanting desperately to view the commercials so that I could experience the A-V combinations described for myself, affording a basis for critical analysis and debate. Perhaps selecting exemplars that are more readily available would have served the audience better or, ideally, making these commercials available on a DVD, either accompanying the text or available separately as a "companion." Instead, at the outset, the reader is placed in a position where one must simply trust the author's description and analysis of the

[8] Chion 1990.

[9] The complete course syllabus, including a list of literature contained in the course reader, can be found on the "syllabi" page of the present author's web site: http://faculty-web.at.northwestern.edu/music/lipscomb/.

existing interrelationships. Despite this criticism, the clarity of descriptions and select captured still images make the author's intended points effectively.

Particularly important in these introductory pages is Cook's insistence that music be considered a communicative medium, extending beyond mere *effect* into the realm of *meaning*. This is an important distinction, though not novel,[10] since the model of multimedia toward which he is leading the reader will require that the meaning attained by each modality be compared for similarity and/or difference. Equally important is his distinction between connotative and denotative meaning, characteristics clearly differentiated in a musical context within the work of Susanne Langer.[11] Drawing upon information in Joseph Kerman's monograph on opera,[12] Cook suggests that, within this specific musical context, "the identification of word with denotation and music with connotation suggest [a] kind of layered, non-competitive relationship" (119). To accomplish its connotative task, according to Cook:

> Musical styles and genres offer unsurpassed opportunities for communicating complex social or attitudinal messages practically instantaneously; one or two notes in a distinctive musical style are sufficient to target a specific social and demographic group and to associate a whole nexus of social and cultural values with a product (16-17).

With these basic concepts clearly delineated and the multimedia artifact as an object of study, the foundation for Cook's framework has been laid.

At this point in the text, I was confused to find myself thrown into a discourse concerning synaesthesia. Though perhaps a topic worthy of brief mention within a book on multimedia, the amount of verbiage devoted to this phenomenon, affecting such a small percentage of the population, seems to imply an importance that is hugely disproportionate to its actual impact upon the typical multimedia experience. Its relevance might be more marked were there a consistent A-V relationship from one synesthete to another.

---

[10] See, for example, Campbell and Heller 1980 or Kendall and Carterette 1990.

[11] Langer 1942.

[12] Kerman 1956.

This is not the case, however, and the fact that colors perceived in the music listening experience vary greatly between individuals affected by this highly uncommon perceptual anomaly suggests that this is not an appropriate basis upon which to build an overarching theory of multimedia perception. This is, of course, the same conclusion to which Cook comes prior to formulating his own model, making the guided tour through synaesthesia and associated theorists—fascinating as it is at times—seem an unnecessary detour. The many fascinating multimedia works upon which Cook focuses in this section (Messiaen's *Couleurs de la Cité céleste*, Scriabin's *Prometheus*, and Schoenberg's *Die glückliche Hand*) could easily have been made relevant based on aesthetic value, without the need for a long-winded discussion about synaesthesia. It is with the introduction of a metaphor-based model that Cook returns to what will truly become useful in the analysis of cross-modal relationships. The present author found the discussion about "record sleeves" later in this same chapter to be of little relevance to the primary thesis of the text and would have liked to have seen this space allotted to more meaningful and relevant subject matter relating to the true multimedia experience.

In the second chapter, Cook's critical analysis of several important models of cross-modal relationships (Kandinsky, Eisensten, and Eisler) puts the reader right back on track in the process of considering the interrelationship of the auditory and visual perceptual modalities. Supplemented with comments made by esteemed film composer Bernard Herrmann and the results of empirical research into the relationship,[13] the author carefully builds a case for consideration of the multimedia context as a metaphorical relationship, based on "enabling similarity" and the resulting "transfer of attributes" (70).[14] Cook also identifies the presence—and stresses the importance—of "emergent properties" in the multimedia context. Such attributes are said to be negotiated between the interacting media within a specific context and "...cannot be subsumed within a model based on the simple mixing or averaging of the properties of each individual medium"

[13] Primarily that of Marshall and Cohen 1988; discussed previously.

[14] After Marks 1978, and Lakoff and Johnson 1980.

(69). Just prior to the presentation of his own model, Cook summarizes his perspective concisely in the following way:

> ...whatever music's contribution to cross-media interaction, what is involved is a dynamic process: the reciprocal transfer of attributes that gives rise to a meaning *constructed*, not just *reproduced*, by multimedia (97, emphasis added).

This emphasis on an emergent meaning that is constructed as a result of the interaction between the various components of a multimedia work is a significant contribution to the study of multimedia.

Cook's own paradigm ("Models of Multimedia") is carefully delineated in the final chapter of Part I. Here, the author sets out his approach to the study of multimedia, interrelationships between the various component media, and potential source(s) of the resulting meaning(s). At its most fundamental level, this model consists of two steps: a similarity test and a difference test. Space limitations for the present review preclude a full description of the model. Its essence, however, involves determining whether component media are communicating the same basic meaning via different perceptual modalities or whether these constituent elements consist of varying messages. In the latter case, the listener-viewer's cognition is a more complex interpretive process. If the media are considered to be communicating the same message, the relationship is said to be *conformant* (Lakoff and Johnson's "consistent"). At the other extreme, if the media communicate in a manner such that the meaning of each contradicts that of the other, the relationship is said to be one of *contest*. In the middle ground between these two polar extremes of a continuum exists a *complementary* relationship, in which the relationship is neither consistent nor contradictory. In selecting the identifiers used to describe these models, Cook consciously decided to coin a new set of terms instead of using the common terms already used frequently to describe these relationships (consistent, coherent, and contradictory). Upon initial contact with Cook's model, I considered this a major weakness, incorporating—what I considered at the time—an unnecessary level of interpretation and resulting in needless complexity. However, as I continued to use these concepts in a classroom context and to participate in

animated discussions about the roles of the various media in a multimedia context by all involved, I found that having such reserved terms—once their meanings were clearly understood—actually served to facilitate the resulting discussions and enhanced the ability to readily distinguish a variety of meaningful interrelationships.

At this point, with the primary intent of Part I of Cook's text clearly accomplished, it was time to enter the realm of analysis, using specific examples from the vast repertoire available. My own purpose in this review is not to agree or disagree with the specific application of Cook's model to the analyses presented in Part II. Other reviewers in the present volume will take the opportunity to do so. I do, however, wish to take issue in a very general way with the examples selected by Cook for the purpose of demonstrating the appropriateness and functionality of his models. As a musicologist, Cook has chosen explicitly to focus on multimedia examples in which the music plays a primary role (i.e., "musical multimedia"). As a result, each of the three selected examples (the video for Madonna's "Material Girl," the "Rite of Spring sequence" from *Fantasia*, and "Armide" from *Aria*) represents a multimedia context in which the music predates the accompanying visual component and dominates the multi-modal texture. Quite the opposite of instances in which the visual image is autonomous (a situation dealt with by the author at length in Part I of the text), the chosen excerpts focus *solely* on a relationship at the opposite extreme of the spectrum, rather than providing a variety of media types and representative interrelationships. In cinema, arguably the most sociologically significant form of multimedia at present—and, admittedly, the present author's primary area of research interest—the sequence of events involved in production is quite the opposite. Typically, though exceptions to this rule certainly exist, the film composer is given a finished product for which she is asked—within a phenomenally short period of time—to produce a musical score for the purpose of enhancing the dramatic narrative. It would seem appropriate to have included at least one excerpt from a feature film in the set of examples for analysis, given the significance of this artform as evidenced by box office receipts. This is not intended to denigrate the selections of the three very interesting pieces analyzed, each useful in its own right and quite

different one from another. I question only whether—other than the music video—they represent types of multimedia that are exemplary to the extent that the analytical method applied to them can be shown to be appropriate for other similar examples of multimedia that occupy a position of high sociological significance within our culture. I wonder if the selection of such materials doesn't run counter to the author's stated intent to "contribute to the current reformulation of music theory in a manner that loosens the grip on it of the ideology of musical autonomy" (vi). Selecting these specific types of multimedia, intentionally or unintentionally, raises the musical component to the position of most significant feature, upon which all others are based and/or to which they relate specifically. Though perhaps no longer "musically autonomous," in the sense meant by Peter Kivy (according to Cook's own reference), these chosen works represent—at best—music-*centric* multimedia examples. To what extent does a model formulated for the analysis of such specialized examples generalize to multimedia artifacts in which the roles of individual components share more equally in the emergent meaning of the piece?

Despite the minor critiques offered in these paragraphs, I found *Analysing Musical Multimedia* to be a highly informative and stimulating read. The clarity with which Cook expresses his well-informed ideas is exemplary, as is the manner in which he introduces formative concepts that support the basis for his proposed model of analysis. Though this text does not provide the definitive guide—Cook certainly does not presume to make this claim—to understanding or analyzing multimedia, it certainly takes admirable strides in that direction. I found that the book served my educational objectives extremely well in the context of the previously referenced "Multimedia Cognition" course. It provided an interesting and highly useful counterpoint to Chion's *Audio-vision* and the additional selected readings intended to augment understanding of aesthetics in general and inform students regarding the results of empirical research investigating the multimedia context specifically. Students responded well to the manner in which the material was introduced and developed; this communicated to me that Cook's proposed model facilitated their understanding of the interrelationships between various media and

enhanced their ability to communicate about these matters clearly and concisely.

As a music researcher, I find that my primary remaining concern with the text echoes that previously stated by a colleague and friend. In her review of the same text, Annabel Cohen identifies the author's "unwillingness to endorse the cognitive psychology experimental approach."[15] She goes on to state that many of the ideas presented in the text afford a perfect opportunity to be tested experimentally, specifically mentioning issues related to conscious attention, cross-modal figure-ground relationships, the effect of music on perceived synchrony, the effect of synchrony on awareness of the music, and the effect of music on the perceived quality of activity. Many of these topics have already been broached in empirical work investigating the multimedia experience. In agreement, I would argue that experimental research in general and the cognitive approach specifically offer the perfect tools with which to further revise and develop Cook's set of models. Looking to the future, I see Cook's text as a musicological "statement" made to the interdisciplinary academic community at large to which the community of music cognition researchers can respond with an appropriate "answer." If I had but one wish, I would ask that this scholarly "dance" might proceed through numerous iterations, in a way that will afford an opportunity for dialogue and discussion across extant disciplinary boundaries, bringing us closer to an understanding of the processes inherent in the multimedia experience through the systematic investigation of the intriguing relationships proposed by Cook, supplemented by research already carried out, and clarified by research yet to come. After reading this text and formulating its many testable hypotheses, a research agenda could be set that would occupy the next two decades...at least. I hope Cook's text and others like it will stimulate others to join in the search.

[15] Cohen 1999: 258.

References

Campbell, W. and Heller, J. 1980. "An Orientation for Considering Models of Musical Behavior." In *Handbook of Music Psychology.* Edited by D. Hodges, 29-35. Lawrence, KS: National Association for Music Therapy.

Chion, M. 1990. *Audio-vision: Sound on Screen.* Translated by C. Gorbman. New York: Columbia University Press.

Cohen, A.J. 1999. "Musicology Alone?" (a review of Nicholas Cook's *Analysing Musical Multimedia*). *Music Perception* 17: 247-260.

Cohen, A.J. 2001. "Music as a Source of Emotion in Film." In *Music and Emotion: Theory and Research.* Edited by P.N. Juslin and J.A. Sloboda, 249-272. Oxford, UK: Oxford University Press.

Kendall, R.A. and Carterette, E.C. 1990. "The Communication of Musical Expression." *Music Perception* 8/2: 129–164.

Kerman, Joseph. 1956. *Opera as Drama,* 1st ed. New York: Knopf.

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By.* Chicago, IL: University of Chicago Press.

Langer, S.K. 1942. *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art.* Cambridge, MA: Harvard University Press.

Lipscomb, S.D. in press. "The Perception of Audio-visual Composites: Accent Structure Alignment of Simple Stimuli." *Selected Reports in Ethnomusicology* 12.

Lipscomb, S.D. 1995. "Cognition of Musical and Visual Accent Structure Alignment in Film and Animation." Unpublished doctoral dissertation, University of California, Los Angeles.

Lipscomb, S.D. and Kendall, R.A. 1994. "Perceptual Judgment of the Relationship Between Musical and Visual Components in Film." *Psychomusicology* 13/1: 60-98.

Marks, L.E. 1978. *The Unity of the Senses: Interrelations Among the Modalities.* New York: Academic Press.

Marshall, S.K. and Cohen, A.J. 1988. "Effects of Musical Soundtracks on Attitudes Toward Animated Geometric Figures." *Music Perception* 6: 95-112.

McGurk, H. and MacDonald, J. 1976. "Hearing Lips and Seeing Voices." *Nature* 264: 746-748.

Osgood, C.E., Suci, G.J., and Tannenbaum, P.H. 1957. *The Measurement of Meaning.* Urbana: University of Illinois Press.

Radeau, M. and Bertelson, P. 1974. "The After-effects of Ventriloquism." *Quarterly Journal of Experimental Psychology* 26: 63-71.

Staal, H.E. and Donderi, D.C. 1983. "The Effect of Sound on Visual Apparent Movement." *American Journal of Psychology* 96: 95-105.

Tannenbaum, P.H. 1956. "Music Background in the Judgment of Stage and Television Drama." *Audio-Visual Communications Review* 4: 92-101.

Thayer, J.F. and Levenson, R.W. 1984. "Effects of Music on Psychophysiological Responses to a Stressful Film." *Psychomusicology* 3: 44-54.