7

Film Music and the Unfolding Narrative

Annabel J. Cohen

Abstract

This chapter focuses on the role of music in narrative film. Unlike most other sensory information in a film (i.e., the visual scenes, sound effects, dialog, and text), music is typically directed to the audience and not to the characters in the film. Several examples will familiarize the reader with some of the subtleties of film music phenomena. Two aspects of film music are introduced: *congruence*, which focuses on purely structural aspects, and *association*, which focuses on the associative meaning of the music. The nature of and interplay between the emotional experience of the audience (referred to as internal semantics) and the external "reality" of the film (referred to as external semantics) are discussed, and an assessment is made as to where music (in particular, film music) resides with respect to these two domains. Because the two dimensions of structure and association are orthogonal to the internal–external semantic dimensions, they define four quadrants for describing the relation between music (structure and associations) and film narrative's internal and external semantics. Finally, the concept of a *working narrative* (WN) is introduced as the audience's solution to the task of integrating and making sense out of the two sources of information provided in the film situation: sensory information (including the acoustic information of music) as well as information based on experience including a story grammar. The author's *congruence-association model with the working narrative construct* (CAM-WN) accommodates the multimodal context of film, while giving music its place.

The Curious Phenomenon of Film Music: What Is It Doing There?

Music serves many roles in film (Copland 1939; Cohen 1999; Cross 2008; Levinson 1996), from masking extraneous noise in the theater (in the days of silent film, the noise of the film projector), to establishing mood, interpreting ambiguity, providing continuity across discontinuous clips within a montage, furnishing music that would naturally appear in the film's diegesis or story, directing attention, and engaging the viewer. Some of these roles have been submitted to empirical investigation, but most have not.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

During the silent film era, directors hired composers to create original music (e.g., Eisenstein worked with Prokofiev), although more rarely, some directors, like Charles Chaplin, composed their own film music. Creating original music is standard practice in modern film-making (e.g., John Williams composed for *Star Wars*); however, directors often choose to use precomposed music as well. For example, in the film *The Social Network* (Fincher and Spacey 2010), composers Trent Reznor and Atticus Ross adapted Edvard Grieg's *Hall of the Mountain King* for the Henley Race rowing competition scene. Many films, including *The Social Network*, employ a mixture of appropriated and original music.

Sometimes a film score has large segments that are coherent in the absence of visuals and can well stand alone on a soundtrack album. Many film scores employ a musical phrase or repeating motif.[1] Others may sometimes simply employ a long tone, which by itself could have little appeal. Using *The Social Network* as an example (Sony Pictures Digital 2010),[2] the opening titles entail the drone of a low note slowly and unobtrusively embellished by its higher or lower octave over a four-minute period; a simple, otherwise unaccompanied melody unfolds intermittently. In contrast to these uncomplicated configurations and instrumentation, complex orchestral works (e.g., those of John Williams or Michael Kamen) are also used in film. Simple or complex, sound patterns all fall into the category of music—music that serves the film audience, not the characters in the film drama. Imagine a (nonhuman) computer trying to determine what acoustical information is meant for the audience and what is intended as part of the on-screen action. Such a thought experiment reflects the remarkable human ability to integrate mentally the musical information in a film with the other information of the film so as to arrive at an engaging film narrative.

The successful use of original music composed or adapted for a particular film relies, at least in part, on the composer's understanding of the director's goals in conveying a message and engaging the audience. Directors communicate their guidelines to a composer in various ways: they may provide a *temp track* (examples that the director has chosen from songs or compositions familiar to him or her), talk with the composer, or request examples from the composer for consideration. Temp tracks sometimes end up in the film. For example, composer Alex North's complete score for *2001: A Space Odyssey* (1968) was rejected in favor of the classical selections that director Stanley Kubrick had originally chosen as temp tracks (Karlin and Wright 2004:30). A similar fate befell composer Wendy Carlos' work for Kubrick's (1989) *The Shining*. Sometimes however, composers, such as Aaron Copland for the film *Of Mice and Men* (Milestone 1939), are given freer rein. The significance of a

1 A motif is analogous to a meaningful phrase in language—a musical concept distinct from other musical concepts arising from a small number of sounded elements, distinctively different from other sequences of notes.

2 The original score is available online: http://www.thesocialnetwork-movie.com/awards/#/music.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

meeting of minds between the director and composer is underlined by a number of director–composer partnerships: Sergei Eisenstein and Sergei Prokofiev, Alfred Hitchcock and Bernard Hermann, James Cameron and James Horner, Steven Spielberg and John Williams, David Cronenberg and Howard Shore, David Fincher and Trent Reznor. Seasoned composers would perhaps argue that what is essential about film scoring is an innate sense of what is right for the film.

In terms of "what is right for the film," a first thought might be that the music sets the mood of the film. However, music serves many purposes in a film which collectively bridge the gap between the screen and the audience. Decades ago, Aaron Copland (1940:158) quoted composer Virgil Thompson as saying it best:

> The score of a motion picture supplies a bit of human warmth to the black-and-white, two-dimensional figures on the screen, giving them a communicable sympathy that they otherwise would not have, bridging the gap between the screen and the audience. The quickest way to a person's brain is through his eye but even in the movies the quickest way to his heart and feelings is still through the ear.

This implies that music quickly and effectively adds an emotional dimension to film, allowing an audience to relate more fully to it and engage in it, linking the internal world of the audience to the external world represented by two-dimensional moving images and a soundtrack. Emotional information can, of course, be conveyed without music via other visual and auditory media channels; consider, for example, a *visual image* of children playing in a garden on a sunny day, *words* in a subtitle description "happy children at play," the *sound effects* of laughter of children playing, or their fanciful spoken *dialog*. Film music, unlike these other four media channels—scenographic, textual, sound effects, or speech—has a direct route to the audience's "heart and feelings," or so it was thought by Copland and Thompson (see above quotation) and no doubt by many directors and audiences alike. Such a description is, of course, hardly scientific (for a model of various component processes that interact in "direct" feelings of emotion, see Scherer, this volume). Nonetheless, the notion of supplying *warmth* to a two-dimensional figure implies that music makes the film more compelling and engages empathy (though perhaps by overlapping with mirror mechanisms; see Fogassi, this volume). This further implies that when music accompanies film, it prompts emotional associations and responses of listeners and adds emotional dimensions. Some examples are provided by horror films like *Psycho* (Hitchcock 1960) and *Jaws* (Spielberg et al. 1988), where a short musical motif played at particular scary times in the film becomes associated with a particular kind of fear (Cross and Woodruff 2009). Excerpts that take on meaning through their association with events or characters in a drama are referred to as *leitmotifs*, a term applied for similar use in the operas of Wagner (Cooke 2008:80–83). The explanation of the leitmotif

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

rests on developing a conditioned reflex, although understanding what exactly is conditioned, and the time course involved, is not a simple matter (Cohen 1993). Focusing on these leitmotifs alone belies the complexity of music perception and cognition as well as the fact that music engages so much of the brain (Overy and Molnar-Szakacs 2009; Koelsch, this volume; Janata and Parsons, this volume). The perceived unity of a piece of music suggests that music is a singular thing. However, music is comprised of many components: from a single pitch, to relations between pairs of notes and among groups of notes, timbre, loudness and timing patterns, harmony, and patterning on various hierarchical levels, etc.

When discussing the role music in film, I find it useful to distinguish two aspects of music: congruence and association. *Congruence* focuses on the structure of music, which can overlap with structures in other sensory domains. It is exemplified by cartoon music, which matches music to an action. Known somewhat pejoratively as "Mickey Mousing," congruence is readily employed in narrative film. Kim and Iwamiya (2008) studied perception of auditory patterns and moving letter patterns (telops) as might be seen in television commercials. Their studies, which involved rating scales, revealed sensitivity to similar patterns of motion across visual and audio modalities (see also Lipscomb 2005; Kendall 2008). Distinct from congruence (i.e., the structural properties of film music), *association* focuses on the meanings that music can bring to mind. The concept of a leitmotif belongs here as one example.

To date, research on film music has focused primarily on association; that is, how music contributes to the meaning of a film (Cohen 2010). Studies have shown that music influences meaning regardless of whether it precedes or foreshadows (Boltz et al. 1991), accompanies (Boltz 2004; Shevy 2007), or follows a film event (Tan et al. 2008). Several experiments by Thompson, Russo, and Sinclair (1994) indicate that music can evoke a sense of closure. This interpretative or modifying role arises primarily when a scene is ambiguous (Cohen 1993). The direction of the modifier seems to be typically audio to visual, and is consistent with the view that vision is the dominant sense. An unambiguous visual image can, however, enforce interpretation of music that has neutral meaning (Boltz et al. 2009; Spreckelmeyer et al. 2006). Film music has also been shown to affect memory of a film (Boltz 2004; Boltz et al. 1991). However, the complex issue of explaining the role of music considered part of the scene versus the same music intended solely for the audience has only been the focus of two investigations (Fujiyama et al. 2012; Tan et al. 2008).

The Diegetic World of the Film

Whether for a feature film of several hours or for a commercial as short as 15 seconds, twenty-first century audiences typically take the presence of film music for granted (i.e., they find it acceptable and ordinary). Yet, even if all

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

other aspects of the film scene evince utmost realism, most music that accompanies a depicted scene would not belong in the film world's acoustic reality. The life and world of film characters is referred to, in film theory, as the *diegesis*. Information outside the world of the film is referred to as *nondiegetic*. Typically, film music plays no part in the diegesis. Consider a scene from the film *Road House* (Herrington and Silver 1989), where the villain (played by Ben Gazzara) displays his "king of the road" effrontery in a comic light, swerving down the road in his convertible to the accompaniment of the 1950s hit rock and roll song *Sh-Boom*. The music, if noticed at all, seems nondiegetic: its regular rhythm and undulating contour of the melody mirrors the almost sine-wave contour of the car's back-and-forth trajectory across the highway's midline. However, as the scene continues, we discover that Gazzara is singing along to the music (Figure 7.1), thus conferring a diegetic role for the music.

Most film music is, however, nondiegetic; that is, it is not meant to be heard by the film characters, unless for comic effect. In *Blazing Saddles* (Brooks and Hertzberg 1974), for example, the background big band music *April in Paris* is heard nondiegetically as the jazzy sheriff begins his gallant ride on horseback across the desert (Figure 7.2). However, when the Sheriff surprisingly comes across Count Basie and his band performing visibly in the desert, the music immediately becomes diegetic. The jazzy associations of the music, consistent with the super-cool personality of the sheriff, fool the audience into ignoring the music as an audible phenomenon. The appearance of Count Basie and his band in the film summons attention to the music as a performance (e.g., that the sounds are played by musicians, and that one cannot hear this music played any better). It is generally agreed by film theorists that much film music is largely unheard, as captured by the title of Gorbman's (1987) book *Unheard Melodies: Narrative Film Music*; that is, film music is not heard as music. Instead, the audience attends to second-order information provided by the structure of the music or the ideas and emotions associated with the music. This second-order information, in turn, directs attention either to other visual or auditory patterns in the film or to the meanings conveyed by the music which help interpret the meaning of the

Figure 7.1 Scene from the *Road House* (Herrington and Silver 1989): the antagonist (played by Ben Gazzara) careens down the road singing to *Sh-Boom*, demonstrating the diegetic nature of the music. (http://www.youtube.com/watch?v=xLdyuwqik4Q)

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Figure 7.2 In this scene from *Blazing Saddles* (Brooks and Hertzberg 1974), the nondiegetic role of the film music (*April in Paris*) changes abruptly when the main character, Sheriff Bart, encounters Count Basie and his band in the desert playing the very tune being heard by the audience. The joke is on the audience, who most likely assumed (if they assumed anything at all) that what they were hearing was "just film music." (http://www.youtube.com/watch?v=7cLDmgU2Alw&feature=player_detailpage)

film. Characters in the film are unaware of the music, unless, for example, the narrative includes such events as a live concert, or someone practicing saxophone, or a band visibly playing in the desert, as in *Blazing Saddles*. The audience is often unaware of the music as well but, unlike the film characters, it is affected by the music. In a pilot study in which film excerpts were shown without music, Tan et al. (2007:171) report that approximately two-thirds of the participants appeared to believe "that the images were accompanied by a film score even when there was no music at all." Just as people can speak without explicitly knowing grammar, so can they enjoy a film without appreciating all the components of its narrative structure.

In the silent film era, music was primarily nondiegetic, serving to establish moods or interpretive contexts. After 1915, film distributors often included guidelines for music selections. Collections of music suited for various film contexts were specifically developed and published for use in silent film. There were exceptions: the Eisenstein films in Russia had scores composed by Prokofiev (*Alexander Nefsky*, 1938; *Ivan the Terrible*, 1944/1958) and by Edmund Meisel (*Battleship Potemkim*, 1925); shorter avant works, such *Entra'acte* (1924), directed by René Clair in France commissioned music by Camille Saint-Saens. Some of these full scores have survived or been restored. Occasionally live performances of the film and score are mounted in concert halls around the world.

When the talkies arrived in 1926, film directors temporarily retired the nondiegetic use of music, reclaiming it gradually by the mid 1940s. From the late 1920s to the early 1940s, music typically was only included in films when the plot called for it (Prendergast 1992:23). Directors, however, liked to have music in the film, and thus plots were contrived to include it. Eventually, such forced plots were found to be less than effective. Similarly, the capacity for sound recording led to the birth of the movie musical, which became very popular during this time. That the majority of directors, even Hitchcock, were

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

long in rediscovering the nondiegetic function of film music underlines that the presence of music in a realistic film is counterintuitive. Why would film audiences need music to make the film more compelling, when the music is not part of the depicted action and all the actual sound of the scene can be represented through sound recordings?

Based on her comparisons of deaf and hearing audiences as they viewed films from the silent era, Raynauld (2001) emphasizes that sound has always played a role in films, even before the advent of talkies. It was just that silent film audiences had to imagine the sounds. Thus, although films produced between 1895 and 1929 were silent from a technical standpoint, they were not silent "from the vantage point of narrative" (Raynauld 2001:69). Actors and actresses spoke the words in the script even though they could not be heard, and title slides bridged some of the gaps. Consistent with the audience's "hearing" of the speech track of silent films, the first talking pictures often led to audience disappointment when the speech of the characters did not match how the audience had imagined them. This disappointment provided a key plot point in the twenty-first century "silent film" *The Artist* (Hazanavicius and Langmann 2011). Here, the protagonist cannot make the transition to the talkies due to his accent, a point revealed only in the last few minutes of the film. In retrospect, as nondiegetic music had a place in these "silent" films, it might follow that its role would be retained in talking films. However, the 1920s invention of the talking film heralded the demise of the robust silent film music industry, an industry that had supplied work to the vast majority of all performing musicians. In the United Kingdom, for example, 80% of all professional musicians lost their livelihoods at that time (Cooke 2008:46). As mentioned, directors now assumed that music for the talkies must be part of the diegesis; however, their contrived plots or films without music soundtracks were found lacking. Eventually the industry of film music returned under the new guise of the full prerecorded film score to replace the live music performance of the nondiegetic music of the silent film era.

The acceptance of and need for film music in talkies (i.e., the sound film) can be explained by considering the audience member as an active rather than passive receiver of the sensory input of the film (Bordwell 1996:3; see also the approach to narrative by Abbott 2008). While narrative approaches have been applied to the analysis of music on its own (Levinson 1997:167–169), the film as story obviously demands such consideration (Bordwell 2006). Levinson (1996:252) argues for three levels of narrative within a film: (a) the fictional story, (b) the perceptual enabler, who presents "the story's sights and sounds," and (c) the filmmaker, who is not on the same plane as the diegesis. He suggests that film music can be revealing on each of these levels. In his view, the contribution of music to the fiction is fairly obvious, for example, in the case of disambiguating context. The setting of mood in the composed film score may be linked more with the enabler, whereas the establishment of specific association to borrowed music (e.g., known popular music) reveals

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

much about the filmmaker. Levinson challenges the accepted view that film music is unheard (Gorbman 1987) and argues instead that it is by necessity heard so as to help the audience understand the position of the narrator and the filmmaker, although one might still argue that music may be heard more or less unconsciously. He does not believe that the audience is responsible for creating the film, but rather that it is the job of the audience to understand how the narrative, as created by the narrator and filmmaker, is to be understood. It is important to keep Levinson's view in mind while considering the simpler view of the audience as creator of the narrative (and of the higher-order levels of narrator, and filmmaker, should information in the film be sufficient to allow this aspect).

The very presence of nondiegetic music in a film challenges cognitive science, film theory, and musicology for an explanation. I have addressed this issue in previous articles (e.g., Cohen 1990, 2013). The phenomena, however, demand continuing attention and exploration. Robin Stilwell (2007) has referred to the "fantastical gap" between the world of the film and the world of the audience. When we read a book, we construct the story; we do not construct music imagery to accompany the narrative. No music score attaches itself to our day-to-day existence, although some of us may imagine music throughout the day, and such music may be appropriate to our moods. Sometimes such music is merely an earworm that we cannot get out of our heads. When we dream, a music soundtrack is not usually part of our dreams. In contrast to these examples of narrative in our lives, in which music soundtracks play no role, in film, a director uses music to make a scene more compelling. Realistic details down to a car's license plate command a film director's attention, and music is used when it would not normally occur as part of such details. Why, then, is there a place for music? The situation is puzzling. The brain needs music in order for film to work, in order to engage fully in a scene. What gap, then, does music fill? What part of music fills the gap? Is all of the music necessary? Can the brain choose what part it needs and ignore the rest?

Film Score Examples: Opening Runs in *Chariots of Fire*, *The Social Network*, and *Jaws*

Film music does more than create an auditory shortcut to a mental emotion center. The power of music in film can be attributed to its unique *musical* features; that is, the capacity of patterns of notes in melodies and harmonies to connect to emotions via musically induced tension and relief (Meyer 1956), musical implications and realizations (Narmour 1991), and musical motifs and meanings. In addition, music contributes to film through *structural* features shared with or in counterpoint to other aspects of film: visual scenes and objects, sound effects, speech, or even visual text (Cohen 2009; Fahlenbrach 2008). Whereas the contribution of music to film has been appreciated since

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

the time of Copland and even earlier by Hugo Münsterberg (1916/1970)—the first psychologist to direct attention to film—the scientific knowledge acquired in recent years about brain function, music perception, and cognition may help reveal music's contribution in not one but in many different ways to an audio-visual presentation.

Let us consider examples of music from the openings of three well-known films. Each scene involves one or more characters in relentless motion toward a goal that is significant to the film. *Chariots of Fire* (Hudson and Putnam 1981) opens with a group of runners training on a seacoast in England (Figure 7.3). The music begins with quickly repeating (sixteenth) notes in the bass, which establish a reference note (keynote, main note, or tonic) for the entire piece. The persistent repeating note might be regarded as reflecting the determined action of the runners. The higher-pitched soprano line introduces a two-note motif reminiscent of a military bugle call: the second note is higher than the first (the distance from *do* to *sol*, also known as a rising *perfect fifth*). Soon a simple melody that entails triplets (three notes to a beat, as in the word *won-der-ful*) contrasts with the low repeating sixteenth notes (four notes to a beat, as in *mo-tion-pic-ture*) and imparts an additional layer of meaning to the characters and situation. As with the different temporal patterns in the music, there are different temporal patterns in the visual scene: the rolling waves and the runners. For some audience members, the triplet patterns may relate to the rolling waves while the sixteenth notes connect with the runners. The music seems to add to the film but, of course, the music is absent in the diegetic world of the actors: the runners do not hear the music.

Let us compare this to the opening of *The Social Network* (Fincher and Spacey 2010). After being rejected by his girlfriend at the Thirsty Scholar pub, the main character, Mark Zuckerberg, emerges from a subway stop and pensively navigates across the Harvard campus to his dorm (Figure 7.4). The accompanying music signifies that this is no ordinary "walk in the park": an unrelenting low bass note (actually just a semitone away from the keynote of *Chariots of Fire*) is introduced, which segues into its octave (its nearest harmonic neighbor, from the point of view of physical acoustics) and then back

Figure 7.3 Opening scene from the *Chariots of Fire* (Hudson and Putnam 1981).

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Figure 7.4 Opening scene from *The Social Network* (Fincher and Spacey 2010).

again to the bass note. The change might metaphorically signify that something is unstable; something is brewing. This unrelenting bass note line almost becomes attached to the central moving Zuckerberg character: sometimes he is a speck far less visible than other passersby on his route, but we do not lose sight of him. As in the opening of *Chariots of Fire*, a simple higher melody is presented. The path of Zuckerberg can be likened to the audiences' melody line, the figure against the background harmony of the visual scene. His footsteps contribute to the sound effects. Sometimes there are other sound effects: rustling belonging to a passerby, or music played by a street-musician violinist. The violin music, however, sounds like the violinist is tuning up, though visually the violinist appears to be engaged in a demanding, presumably classical, work. The audible violin music and the violinist that is visualized do not fit together in any realistic way, but the audience likely ignores this due to the focus placed on Mark, the main character—a character who will design the social network through his superior quantitative and leadership skills, and who will suffer through a lack of social intelligence. These contrasting aspects of his character are made clear in the first scene at the Thirsty Scholar Pub, and the music, with its contrasting soprano and bass lines, help reinforce this: the simple childlike melody and the brooding baseline of cold pragmatic reality (as described by the director, David Fincher, on the DVD additional tracks). Again, the music in no way distracts from the storyline and is not part of Mark Zuckerberg's world—other than the violinist, whose sound is not faithfully represented.

Now consider the opening scene from *Jaws*: a hippy beach party. Faint realistic music comes from a harmonica here, a guitar player there. The camera focuses on a male eyeing a girl at a distance. The music and sounds of the party are almost inaudible. He approaches the girl; she is delighted and wordlessly beckons him to follow her. She, like the protagonists of the two previous films, begins a determined run along the beach, with the boy following, somewhat in a drunken stupor (Figure 7.5). She gallops ahead, and her silhouette depicts her disrobing piece by piece, until she arrives presumably naked at the water's edge and plunges in. What is different in this example is that there is no music

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Figure 7.5 Opening scene from *Jaws* (Spielberg et al. 1988).

during the determined run. In fact, there is no music at all until she enters the water. At that point, a repeating bass rhythm of the *Jaws* theme begins and develops in conjunction with the screams and thrashing motions of the woman fighting against the shark, depicted by the music.

Each of these films received awards for their film scores. In general, they all open with a repeating low pitch, signifying a protagonist with a driven nature, but there are differences. *Chariots of Fire* and *The Social Network* use music to underscore the act of running; in *The Social Network*, music defines the protagonist whereas in *Chariots of Fire* it depicts a group of runners. In *Jaws*, however, the onset of music foreshadows the entry of the shark, the primary protagonist of the film; the preceding scene of running lacks music altogether. In none of these examples is the music complex, although many films scores do use more complex music from the start, and these three films all use more complex music elsewhere. The simplicity of these backgrounds helps explain the influence of the congruence and associationist aspects of the music on the audience's interpretation of and engagement in the film narrative. Before discussing this further, let us first consider the internal and external worlds of film.

Internal and External Semantics and Unfolding the Narrative of the Film

Film music provides a unique perspective on the semantics of internal and external worlds (see Seifert et al., this volume). By *internal semantics*, I am referring to the emotional perspective of the perceiver: how the audience member feels while watching a film, and how the audience member feels about the characters and events depicted in the film. By *external semantics*, I mean the regularities of nature and cultural conventions of human interaction and human–environment interaction: the rules of language grammar and the pragmatics of discourse, the natural phenomena of climate, geography, and the rise and fall of the sun, the customs of dress and behavior, as well as the systems of social value of various historical times. All of these events can be learned from experience of the physical and social worlds. As the quote earlier from

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Copland and Virgil Thompson implies, film music bridges the internal world of the audience member and the information provided by the screen. Film music thus provides an opportunity to study the relation between internal and external semantics, potentially contributing to the cognitive scientific understanding of how individuals make sense and personal meaning out of the cues that nature and social reality provide. Film music helps with this complex problem because of some essential differences from, as well as some less obvious similarities with, other media channels of the multimodal display. As pointed out earlier, film music, in contrast to the speech and sound effects in a film, is typically to be heard only by the audience, not by the film characters; yet most of the other sounds and sights in the film are audible or visible to both audience and actor.

Consider the audience member in a public or home theater presented for about two hours with a large rectangle of moving images, and sometimes printed text, as well as sounds from audio speakers, typically while sitting in a darkened room, often in the midst of other viewer-listeners. Each viewer-listener faces a two-hour task of organizing the presented information in an engaging and enjoyable manner, justifying the time spent and the financial costs of the movie ticket or, if in the case of home theater, the investment in audiovisual equipment. How does the music facilitate the task? A clue lies in the distinction between internal and external worlds in the film music context because, unlike the other light and acoustic media components of film which depict the external world (i.e., the scenography, the printed text, the sound effects and the spoken dialog), music typically exists in both inner and outer worlds, although, as will be suggested, only a part of the music functions within the internal semantics. As will be shown, music can play on emotions (the internal world) while directing attention to the information in the external world. Music may add unity to the experience of film, because music, even in its simplest form, operates in both amodal structural (i.e., not specific to sound or vision) and associative (meaning) realms, both of which connect to external and internal semantics (see Table 7.1).

Table 7.1 aims to represent how the structural and associationist aspects of various musical elements affect the internal and external meaning of a film. Elements of the film score (left-hand column) are deconstructed into structural congruence and association. The top panel focuses on the structural aspects of these musical elements: how they change over time (e.g., increasing or decreasing intensity or loudness), or how two aspects relate to each other (e.g., the consonant or dissonant relation of tones). The bottom left panel focuses on associations of various musical elements (e.g., what is associated with a low tone as opposed to a high tone, or consonance as opposed to dissonance). The musical elements in the upper and lower panels overlap but their structural and association aspects differ. Thus we find in the list the analysis of the pitch, tempo, and metric relations (temporal patterning), direction of the pitch (whether it ascends or descends), absolute intensity (soft/loud), pattern of intensity change

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Table 7.1 Representation of how music structure and music associations impact the interpretation of the unfolding film narrative with respect to internal and external semantics.

	Internal Semantics of Film	External Semantics of Film
Music Structure (Congruence)	• Cross-domain structural congruencies produce an aesthetic response (e.g., musical triplets to the ocean waves in *Chariots of Fire*) • A low repeating pulsating pitch (e.g., the drone in *The Social Network* or *Chariots of Fire*) engages subcortical processing and higher brain levels • Parallel music/scenographic figure/ground direction of attention (e.g., Zuckerberg figure as a musical stream across the Harvard campus) • Parallel temporal structures may compete for and share syntactic resources (Broca's area) • Bypass acoustics (sounds)	• Congruencies between real (diegetic) music heard by a film character and activities of the character provide information about the character (musicality, human-ness of the character; e.g., Gazzara singing and driving in *Road House*) • Degree of synchrony between music and some other film modality conveys information about the real world of the film (e.g., oscillating path of the car in *Road House*) • Acoustics included in diegetic music • Acoustics audible
Music Meaning (Associations)	• Metaphoric, embodied parallels (e.g., rising pitch) give meaning to events eliciting a mirror neuron system (e.g., motor system to raise pitch of the voice, or raise a limb) • Sounds pulsed at the heart rate engage deep emotions (e.g., *Jaws* semitone motif) • Tension release and increase support personal emotional interpretation and involvement • Bypass acoustics	• Learned associations and conventions (national anthems, funerals, weddings) convey information about era and other context and establish mood • Acoustics may be bypassed if only associations contribute (specific music is not recognized, as it plays no role) • Repeated music-film contingencies create leitmotifs and further plot development • Diegetic music that fits the story is heard as music (e.g., dance music; a performance; Count Basie' s band in *Blazing Saddles*); acoustics enters consciousness

(increasing/decreasing), triad relations (major, minor, diminished, augmented triads), chord analysis, chord progression, and phrase repetition. The list is not meant to be exhaustive; the various dimensions and elements identified are amenable to variation on hierarchical levels not represented here and would connect with various psychomusicological theoretical stances (the generative theory of tonal music by Lerdahl and Jackendoff 1983; see also Lerdahl, this volume; Huron 2006; Narmour 1991). Lipscomb (2005) has identified timing pattern (rhythm) as a structural feature whereas Cohen (2005) focuses on the

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

highness and lowness of pitch and slow or fast tempo for its impact on happy–sad meaning. To the best of my knowledge, no empirical study has varied both structural and association aspects together; such a study begs to be conducted.

In Table 7.1, the two right-hand columns refer to the internal and external semantics of the film. The former focuses on the audience's feelings about the characters or situations and engagement in the plot, whereas the latter refers to acquisition of knowledge portrayed by the film: information about the characters, the time period, and setting.

Below I discuss each quadrant of Table 7.1 in turn so as to explain how musical structure and musically inspired multimodal structural congruence can affect the internal and external semantics in the listening/viewing experience of a film and, likewise, how the associations of, or perhaps to, music can affect the internal and external semantics of a film.

Music Structure (Congruence) and Internal Semantics of Film

Music can create attentional focus to aspects of the scenography with which it shares structure (i.e., temporal patterning). To provide a mechanism for this, I propose that aspects of the structure of the music which match the structure of other nonmusical elements of the film produce similar neural excitation patterns of system-specific (i.e., auditory or visual system) cell networks in the brain. The resonance of these similar excitation patterns brings to attention (i.e., to consciousness) those components in the film that embody these shared characteristics. This view is consistent with a recent review by Kubovy and Yu (2012:254), who argue that "the strongest relation between perceptual organization in vision and perceptual organization in audition is likely to be by way of analogous Gestalt laws." As an example, consider the motion pattern of the runners in *Chariots of Fire*, which matches the low-note rhythm of the music, or the melodic motif of *The Social Network* theme that depicts a moving speck (to which the audience's attention is drawn), which we know to be Mark Zuckerberg wending his way back to his dorm. It can be argued (though yet to be empirically tested) that similar music–visual structure draws attention to the visual information which gives rise to that shared structure. Although Gestalt psychology does not refer to a principle of multimodal similarity, the notion is a simple extension of the rather amodal principle of grouping by similarity within a sensory modality. Also consistent with Gestalt theoretic notions is the aesthetic satisfaction that arises from the registration of good form—an aspect of the internal semantics of the film. Such a response, which engages an audience member from an aesthetic standpoint, would also provide a focus of attention.

Even pitch analysis admits to consideration from the perspectives of structure, though in a completely auditory realm. All melody and harmony require pitch. The study of psychoacoustics tells us that the representation of pitch itself entails complex pattern recognition. The physical basis of pitch is a repeating

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

pattern of air pressure. At a rate of between 30–20,000 times per second, any repeating air pressure pattern will produce the perception of pitch in the normal human auditory system. The sensation of pitch, as distinct from noise, requires the mental registration of approximately six cycles (Boomsliter et al. 1970). Recent evidence provided by Bidelman and Krishnan (2011) shows that all levels of the auditory system follow individual cycles of the waveform beginning with lower brain centers. When two tones are presented simultaneously, simpler ratio relations (like the octave 2:1) are represented most clearly, and the rank ordering of the faithfulness of the representation of these relations matches the ordering of the perceived consonance. While Western tonal music scales generally exploit small ratios or their approximation, certain music exploits the simpler relations more than others. Many so-called good soundtracks (e.g., *Chariots of Fire*, *The Social Network*, and *Jaws*) exploit these simple relations, thus giving processing precedence in both lower and higher brain centers, as compared to music that exploits more complex (more dissonant) musical relations. The engagement of the sensation of pitch (as opposed to noise) may serve to engage the audience member due to the establishment of (possibly large-scale) coherent brain activity entailed by pitch representation.

Low pitches characterize the music of the three award-winning scores highlighted above. My speculation is that low sounds may have special relevance for their ability to command engagement without calling attention to themselves as music per se. Interestingly, low pitch (40–130 Hz) in the audible frequency range is high frequency in brain wave EEG patterns, and it has been suggested that this gamma range of EEG may be significant in the establishment of consciousness (John 2003; Edelman and Tononi 2000). John (2003) has reviewed research that indicates how endogenously (self-) generated rhythms can suppress exogenous (automatically evoked) information. Such a principle may account for the many anecdotes and reports that the acoustic information of film music does not reach consciousness as it normally would outside the film context. Boomsliter et al. (1970) likened John's basic notion of endogenous processing of flicker to the processing of pitch versus noise. It is not suggested that electrical gamma waves would simply arise from physical acoustic energy waves of the same frequency; however, a possible relation between the two and the effect on consciousness necessitates further study (cf. John 2002:5).

A good example of structural congruence can be found in Kubrick's *The Shining*. Toward the climax of the film, the child, Danny, retrieves a butcher knife from the bedside table near his sleeping mother (Figure 7.6). With the knife held up in his left hand, the child's right-hand fingers move up toward the tip of the blade just as an ascending dissonant sequence of three tones sound from Penderecki's *De Natura Sonoris No. 2*; the audience's attention is drawn to the blade, its sharpness and danger. When Danny's fingers descend, the movement is accompanied by two more notes which add further emphasis; as Danny places his fingers lower on the blade, a lower note sounds

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Figure 7.6 Scene from *The Shining* (Kubrick 1989), where Danny senses the sharpness of the knife. His finger motion up and down the blade parallels the direction and timing of the accompanying music—a dissonant melody from Penderecki's *de Natura Sonoris No. 2*. (http://www.youtube.com/watch?v=4iwK2jwqFfk)

and becomes a drone. In addition, immediately prior to and after this short episode, Danny repeats, in a low croaky voice, the word "redrum"—murder spelled backward—which functions as a type of musical ostinato (a repeating melodic fragment), in counterpoint with the discordant instrumental music of Penderecki. This functions much like the background motif of the *Jaws* semitone theme, the repeating rhythm of the *Chariots of Fire* theme, and the repeating bass note of *The Social Network* introduction.

Assuming that watching an action elicits the same neural activity as enacting the action oneself, the engagement of the mirror neuron system is likely to be a major contribution to the internal semantics of the film experience. Watching Danny move his fingers up and down the knife blade may engage an audience member in representing this same activity personally. Because the direction of the melody can also engage the same mirror neuron system (Overy and Molnar-Szakacs 2009), film music can reinforce the empathetic effect. Any ascending pattern of notes will parallel Danny's fingers moving up the knife blade, allowing the audience member more readily to identify directly with the action and empathize with the character. Once attention is directed by visual information, which shares structure with the music, other associative and metaphorical components of the music can connect to the internal semantics, as will next be described. It is important to note, however, that none of the music phenomenon above requires that sounds be heard. The impact of the sound arises from its structure not the sound per se.

Music Meaning (Associations) and Internal Semantics of Film

In addition to the motoric, gestural information described in the example of Danny from *The Shining*, music brings with it emotions that can be ascribed to the experience. Blood et al. (1999) have shown that different regions of the

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

brain are affected by consonant and dissonant music passages. Thus, we would expect the discordant notes of Penderecki in *The Shining* example to increase a personal sense of unpleasantness and danger associated with the sharp object, as compared to the associations that an ascending consonant passage from a Mozart concerto might bring. In the example of Ben Gazzara driving his convertible to the music of *Sh-Boom*, the jaunty melody brings with it carefree associations and old-fashioned coolness. The music helps the audience to become Ben Gazzara, the King of the Road, or to perceive the sharpness of the knife, as Danny's fingers run up and down the knife's blade, like a discordant musical arpeggio. These are no longer cardboard characters; the amodal musical properties of ascending and temporal pitch patterns coupled with embodied motion plus the associations that come with the music enable an audience member to identify more easily with or, in a sense, become the characters on the screen.

Sounds pulsing at the heart beat or breath rate (as in *Jaws*) may bring associations of life and death. The random modulations of the repeating tone in *The Social Network* mirror anxiety and uncertainty—something is about to happen. These associations support personal interpretation and involvement in the drama. As in the structural analysis described earlier, these association effects are independent of actually registering the music as sound. By analogy, readers are typically oblivious to the letters on the page as they extract the meaning of the words.

Music Structure (Congruence): External Semantics of Film

Mental analysis of film music, carried out on various hierarchical levels, may lead to the identification of the music, if it is a known piece, or to the recognition of its style. As discussed, film music may be, or can become, part of the story, the diegesis. In the *Road House* example discussed earlier, the music begins in the seemingly nondiegesis, but as the camera pans in for a close-up of Ben Gazzara, we realize that he is singing along with the music—the music has thus crossed the "fantastical gap" (Stilwell 2007). In this example, the source of music can be attributed to the convertible's radio. Of particular interest to the consideration of music-visual congruence is the synchrony of the music structure and actions of Ben Gazzara, the driver. He sings along with the music. The congruence indicates his musicality. His bass voice is distinct from that of the tenor range of the recording. His voice and mouth are synchronized with the music, as are the movements of his head. He embodies the music, and he imparts that embodiment to his steering wheel and the undulating motion of his vehicle. This music-visual structural congruence contributes information about the external reality of the film narrative, in particular about the Ben Gazzara character. The audience does not give this a second thought, but a very different personality would be portrayed by a driver who sings off pitch, out of synchrony, without embodying the music. Synchrony adds a dimension to

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

this character. It is also a source of humor and drama in the film as the world of abstract structure—that of a dance with a car—clashes with an object that also inhabits the real world; specifically, a driver coming in the opposite direction, quite out of phase with the ongoing congruent audiovisual pattern. This addition to the communication of the external reality of the film world arises by virtue of the structural congruence of the music with other nonmusical structure in the film. The phenomenon requires that music be part of the diegesis and that the music is heard.

Music Meaning (Associations): External Semantics of Film

The style of the music provides associations that contribute to the external semantics. Diegetic music provides information about the world of the film in a very obvious way. It is the way in which directors, in the early days of talkies, thought film music should work. The old rock and roll tune *Sh-Boom* helps represent the character of Ben Gazzara, telling the viewer his tastes in music, characterizing his age, providing a sense of the complexities of his personality, and appearing in a fairly human light in some respects.

Some films incorporate music performances within the drama to showcase a potential hit song for the film. Some background music may be used to set the scene and provide information about the context of the film through association. If new associations are formed between the music and events or characters in the film (i.e., if leitmotifs are created by the pairing of music and film) and these serve to further the plot and provide information about the reality of the film (e.g., a shark is nearby), this can also be considered as a musical contribution to the external semantics of the film. At the same time, the emotional impact of this cue could be regarded as an associative effect with implications on the internal semantics of the film. In some cases, the music contributing to the external semantics of film may be the focus of consciousness, but this is not necessary, as in the case when music associations set the historical period or general setting.

Summary

The relationships between the structural and associationist dimensions and semantic (internal and external) dimensions of a film, outlined in Table 7.1, show how the music from a film score can contribute to the impact of the film, from both personal emotional and external meaning perspectives. This structural–associationist analysis can also be applied to any physical dimension of the film: scenography, the text, the sound effects, the spoken dialog, or music. For example, Danny's repetition of the pseudo-word "redrum" (murder backwards) has a structural component quite apart from any other meaning that it might have: this repeating pattern, spoken in a croaky low pitch, draws attention to Danny and his eerie extrasensory perception.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

The Structure of Film

A few general observations about narrative are in order. Let us first note its roots in our embodied interactions with the physical and social world. It "is found in all activities that involve the representation of events in time" (Abbott 2008:xii). Narrative or story appears in many guises. The most obvious is through written text: fiction books and stories that are a natural part of lives from birth through to old age. Developmental psychologist Katherine Nelson (2005) identifies the development of narrative understanding and generation as the fifth stage in the development of consciousness, and states that it occurs between the ages of three and six years.

Historically, stories have a much earlier origin than texts; stories have been told since the earliest civilizations and are exemplified by myths. Storytelling remains an everyday skill—one that can be developed to an extraordinary level of expertise. In film, stories are told visually through scenography and typically, due to technical advances in cinema technology, with added text and talking voices. Sound effects help to tell stories and so, curiously, does music. Each of these domains makes specific demands on the brain, and it is a puzzle to determine whether additional domains (language, sound effects, and music) make it easier or harder for the brain to process the narrative.

Regardless of the medium of presentation, as in everyday perception and cognition, audiences of a film have two sources of information (Figure 7.7): (a) the lower-order multimodal physical/sensory input (i.e., the patterns of light that hit the retina as well as the patterns of sound waves received at the ear) and (b) their higher-order knowledge, based on their own remembered or stored experience. The latter includes the rules and conventions about social norms, the grammar of story construction, and memory of one's own reactions to such information. Long-term memory includes all of the information that an individual can draw on to make sense of the world based on experience. From this arises the capacity for narrative, developed early in life. Both of these sources might be regarded as coming from the external world, but in different ways. Multimodal physical information (Figure 7.7a), which impacts the sensory receptors of the eye and ear, leads to automatic (exogenous) sensory processing, whereas memory (Figure 7.7b), prompted by fast preprocessing of the sensory information, leads to elicited (endogenous) generation of higher-order hypotheses and expectations about the reality represented at the sensory receptors. The output of the top-down endogenous processes (Figure 7.7b) may or may not find a match in the slower bottom-up exogenous processes (Figure 7.7a). Following the concepts of many theorists (Edelman and Tononi 2000; Grossberg 1995, 2007; John 2003; Kintsch 1998a), the best match leads to consciousness and, in my model for the role of music in film (Cohen 2005, 2009, 2010, 2013), to the *working narrative* (Figure 7.7c).

The effect of music on the interpretation of film has often been shown to be additive; for example, happy music plus neutral scene leads to happier scene

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

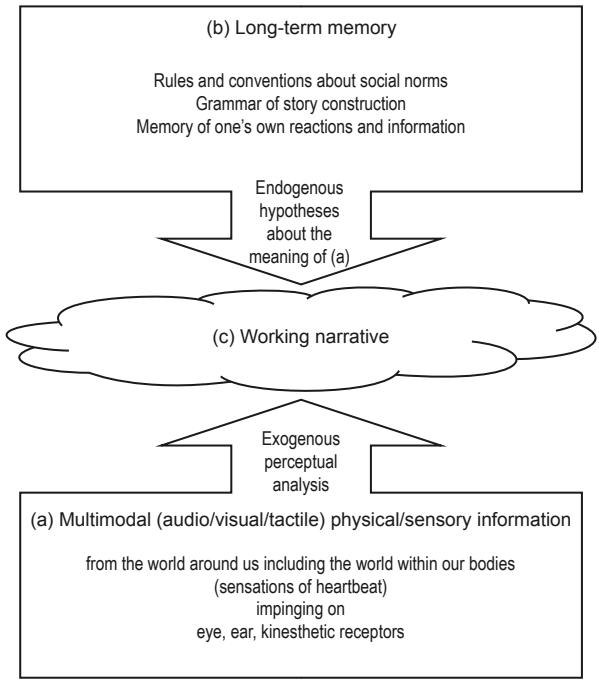

Figure 7.7 Two sources of information are available to the film audience member and lead to the working narrative.

(Cohen 2005). In addition, if camera technique directs attention to a particular visual focus (e.g., to highlight the main character in a film), the associations from the music would also be so focused. However, music itself may offer its own means of directing attention (Table 7.1). The role of congruent music-visual structure on visual attention may modulate an otherwise general associative or semantic effect of film music such that the meaning of the music is ascribed to a *particular* visual focus of attention (Cook 1998; Marshall and Cohen 1988; Cohen 2010). Thus the emotional associations of the music plus the other narrative elements internal to the scene would link together for a cumulative effect. Recall, for example, the scene from *The Shining*, where Danny retrieves the butcher knife that his mother has placed on her bedside stand as she naps. Emphasizing how sharp it is, he glides his fingers up its edge. As his fingers ascend toward the tip, the background music runs through a discordant high pitch arpeggio. The association of discordance and the ascension of the arpeggio add to the ominous nature and danger of the scene. The shared amodal property of ascension, common to the visual action and the music pattern, form a hinge that binds together an audiovisual Gestalt to which now additional associations of the music and the scene can be integrated. The audience member can empathize with this up and down motion along the blade, and the

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

discordance can be ascribed to this focus of attention. This particular motif, or simply the last sustaining low note of the passage, may be linked to the scene or to the knife.

The process described above may be further considered as two stages: (a) focus attention through temporal congruence and (b) ascribe meaning of music to the focus. While there is no evidence that these processes occur in this order, there is evidence for both independently, but not yet in the same study. The anecdotal impression of music and film "going well together" is compelling, as in the example of the knife clip in *The Shining*. The scene from *Road House* previously described provides another. Here the evil city tycoon (Ben Gazzara) drives from his mansion in his convertible singing *Sh-Boom*. As far as he is concerned, he owns not only the city but the highway too, and as the contour of the melody undulates up and down, his vehicle outlines an undulating contour above and below the highway midline. The audio and visual are bound by their amodal similarity, and the jauntiness of the music funnels through to add to the personality of the tycoon. However, the visual and musical trajectories (i.e., their up and down patterns) are, in fact, not exactly in phase, and thus the phenomenon under discussion requires a variety of conditions for testing if we are to understand the relative contributions of the coordinated melodic and visual motion and the ascription of the meaning of the music to the visual focus of attention.

The narrative and multimedia context of film emphasize the ability of the brain to integrate multiple sources of information. This ability, however, is simply what the brain does in most waking minutes. Hasson et al. (2008b) have been exploring the intersubject correlation in brain activity as measured with fMRI across groups of viewers of the same film for more than ten minutes. Their study shows a systematic increase in correlated brain activity as a function of the degree of structure of the film: from a (low-structure) segment of reality video of New York City, to a sophisticated TV comedy *Curb Your Enthusiasm* by Larry David, to *The Good, the Bad and the Ugly* directed by Serge Leone, and finally to a TV thriller directed by Hitchcock, representing increasing levels of aesthetic control. The correlations or cortical activity across audiences (different for the four presentations) range from less than .05 to over .65, respectively. The technique looks promising for application to music variables, provided that the noise of the magnet is not so loud as to disrupt the impact of the music. Hasson et al. (2008b) also review related work, comparing responses to video (silent film) and audio story, which show strong differences in response. In addition, repeated viewings of intact versus scrambled versions of films led to different degrees of correlation.

Film places a boundary on audiovisual stimulus conditions, and this control is helpful in understanding how music and language work within the larger sensory context. With the additional context from other sources of information of visual scenes, text, and sound effects, individual and overlapping contributions of music and speech become apparent. The concept of language and

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

grammar may apply similarly to all five of the domains of visual text, visual scenes, music, sound effects, and speech. By viewing the task of the brain exposed to film as one of integrating information from many sources, it may be suggested that the brain (at least sometimes) does not care where it gets information. It will take it from anywhere as long as it helps to get the story right. Getting the story right is the topic of a larger discussion, but for the present, the criteria would assume Grice's (1975) cooperative principle that information is narrated with the aim of creating meaning and entertainment. However, Grice's principles apply to discourse not to art, and thus additional principles are at stake (e.g., the audience is owed not simply a message but a nice experience, one which might take the scenic rather than the most direct route). There is a host of features on screen at any given time which may reward the viewer's attention, but which may or may not be essential to the story.

It would be parsimonious to apply the same code to information from different sensory sources; information from all source modalities would share the same code. Moreover, there is no reason to tag each piece of information as to its origin; that is, it does not matter whether an association of sadness came from music, or a sound effect, or from text. Instead, it is important to code for external and internal semantics. Structural and associative information apply to both external and personal realms.

Because music adds so much to the film narrative experience (or conversely, because much is lost to film narrative experience in the absence of music), and because how music functions in film is by no means unidimensional, understanding how music operates in film may hold a clue to how all film information is bound together in consciousness, and to how narrative works. Once we figure out how music works in film, it may be easier to understand how multimodal scenes are interpreted in the real world when a music soundtrack is typically absent as well as when, instead, physical and the social information are more strongly integrated (e.g., sensory information is more prominent).

The Congruence-Association Model and Its Extension to Working Narrative

Expanding further on Figure 7.7, my view is that the brain encodes associations and structure from all five channels of sensory information; that is, two visual (scenes and text) and three auditory (music, speech, and sound effects) channels. This five-channel classification was introduced by the film theorist Metz (1974) and promulgated by film scholar Stam (2000; see also Green 2010). In the model presented here, information in these five channels leads first to a passive, automatic, exogenous activity arising from the impact of energy from the real world undergoing electrochemical transduction by sensory neurophysiological processes. Faster and slower bottom-up processes co-occur; the faster processes (or preprocesses) signal to the higher structures in long-term

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

memory, consequently initiating top-down endogenous processes. For a recent discussion of the similar significance of top-down processes, see Gazzaley and Nobre's (2012) article on bridging selective attention and working memory.

Preprocessing of some information primes long-term memory and narrative proclivities for establishing the bottom-meets-top matching process, whereby the best match of top-down inferences to bottom-up sensory input wins out as the *working narrative* in consciousness. In general, diverse processes informed by working memory of recent episodes, as well as by long-term knowledge of the social and physical world, forge a coherent interpretation of (a portion of) the original sensory data. This approach resembles a schema theory (John 2002; Kintsch 1998a; Arbib and Hesse 1986), but some unique features, found only in the *congruence association model with working narrative* (CAM-WN), accommodate film music conundrums (see Figure 7.8). In addition to the five channels previously mentioned, an additional kinesthetic channel is included to accommodate the mirror system activity and extend to the nonfilm (real-world) situation. The nonfilm situation would have a reduced role for the music channel.

The CAM-WN model presents two sources of information available to an audience: *surface information*, which is sensory information in six channels

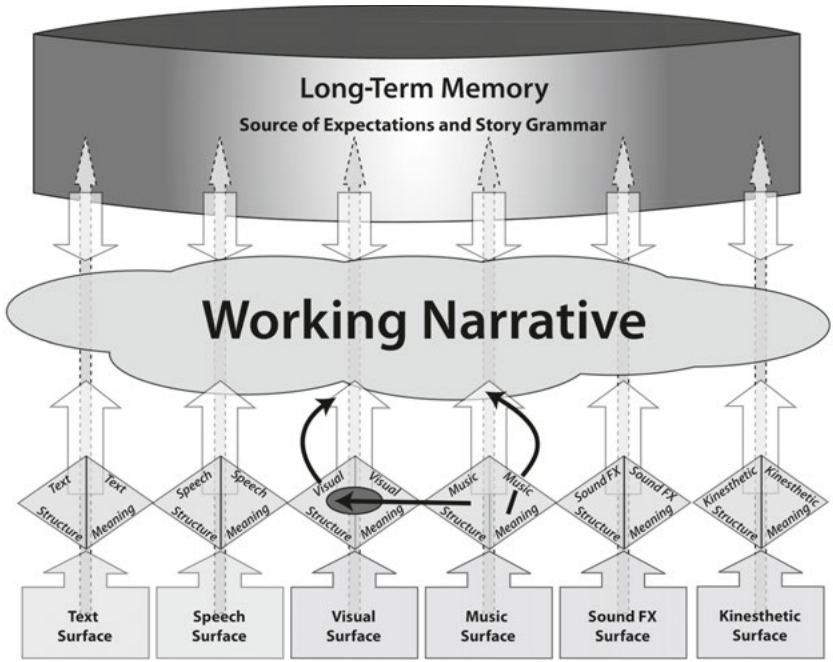

Figure 7.8 Congruence-association model with working narrative (CAM-WN). See text for explanation.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

(text, speech, visual, music, sound effects, and kinesthetic), and *long-term memory*, which includes expectations and story grammar. Surface information entails bottom-up processing—both fast preprocessing (as represented by the dashed lines) and slow, more accurate processing. Long-term memory processes are top down and are elicited by the arrival of activation from the fast preprocessing. The surface information from all six sources is coded similarly into structural features and meaning (associations). The shared code (of structure and meaning) allows for structural congruencies and cumulative meanings, and assists in targeting elements that are relevant to the story. In Figure 7.8, a horizontal arrow from music structure to an oval in the visual channel represents shared music-visual structure in this particular instance. The structural congruence directs attention to the congruent visual pattern which becomes the attended target. These attended elements are the target in a contest to find the best match with top-down processes. The best match enters into consciousness in the working narrative, which is the audience's conscious unfolding narrative. Only matched items achieve consciousness. Thus meanings (associations) of the music and music structures may contribute to the narrative without the actual sounds of the music entering consciousness, because those sounds are not matched by endogenous generation.

The music channel, in part, distinguishes this model from "just a theory of perception" or "just a theory of information processing." The translation of these elements into neural equivalents may well be possible with increasing numbers of studies of brain imaging; however, very few brain imaging studies have focused on music and moving images to date, and more are needed. The CAM-WN framework is helpful because of the focus it gives to music; it also draws attention to the fact that music is not one thing but is comprised of many separate components, each of which can be analyzed in more than one way. Like each of the other channels of film information, music is analyzed for associations and for structure. This approach can inspire studies that isolate particular conditions within and between different modalities (e.g., the effects of different tempos, rhythms, or intensity changes, etc. on meaning within two modalities and the effect of the combination of tempos on meaning when the modalities are combined).

The CAM-WN model is general enough to incorporate other proposals which are relevant to the question at hand but are descriptive and not predictive. Examples at two ends of the spectrum are provided by E. Roy John's neural theory of consciousness (John 2003), as discussed above, and Kathrin Fahlenbrach's insightful approach to explaining the contribution to film meaning through the components of the soundtrack. Fahlenbrach argues that audiovisual scenes refer to presymbolic metaphorical structures in our cognitive system that offer rich material for visual and acoustic concretization of both very complex meanings and the representation of bodily and emotional experience (Fahlenbrach 2008). For Fahlenbrach, audiovisual metaphors function, with reference to emotional metaphors, as emotion markers in empathetic scenes,

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

which represent culminating points of a film. She stresses, however, the need to test her model empirically.

John's focus on the match between exogenous and endogenous processes and Fahlenbrach's focus on the amodal component of media are encompassed by the CAM-WN model. Thus the CAM-WN model allows application of concepts from various levels of discourse within the same framework and specifically incorporates music. Fahlenbrach (2008) does not make reference to specific music features, and John (2003) does not even mention music, although Boomsliter et al. (1970) related his empirical research on exogenously and endogenously produced rhythms to pitch processing. The incorporation of the schema theory proposed by Herrmann et al. (2004) accommodates the matching component necessary for recognition and consciousness and adds the mechanism of the gamma band, consistent with the earlier proposal of John (2003), the function proposed by Fries (2009), and the evidence for frequency following at lower and higher brain levels for pitch analysis by Bidelman and Krishnan (2011).

CAM-WN accounts for the tolerance and prevalence of nondiegetic music that is artificial with respect to the ongoing world of the film (Boltz 2004; Cohen 2010). In CAM-WN, the top-down inference process (from long-term memory) can match the *meanings* of the music without matching the acoustical information (sounds) of the music. CAM-WN proposes something like an illusory conjunction (Treisman and Schmidt 1982) that forms between music meanings and visual object information, leaving behind the acoustic shell or vehicle that is never to appear in the working narrative (the story), though potentially preserved in an independent music memory simultaneously being updated (and without which there would be fewer CD album downloads and sales). John (2003) refers to data that shows how endogenous generation can suppress exogenous information. Endogenous information leads to perception although the exogenous information is in fact registered (without reaching awareness). CAM is descriptive and to some extent predictive, but twenty-first century cognitive science is, or will soon be, able to do better by explaining more specifically how brain mechanisms might carry out and elucidate CAM's functions. Within the last decade there has been progress to this end; however, only one brain imaging study has focused on both the moving image and the music track (Eldar et al. 2007).

Reference was made earlier to the notion of film music contradicting the diegesis. By this I mean that the music would not fit the fictional world depicted: Count Basie and his big band have no place on the desert (*Blazing Saddles*); a violinist happens to be along Mark Zuckerberg's path (*The Social Network*), not three cellos, an electric and a nonelectric piano, and the violinist is playing a piece not tuning up. The acoustics contradict the scene depicted. The emotion conveyed by the music does not contradict the scene, supporting both its internal and external semantics. In CAM-WN the lack of a match between exogenous and endogenous activity associated with the actual sound of the music

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

means that the sound is outside of consciousness. No other current model or framework for narrative deals explicitly with this particular and anecdotally prevalent case.

Using fMRI, Eldar et al. (2007) showed that combining music which evoked emotional meaning with neutral films increased activity in the amygdala, hippocampus, and lateral prefrontal regions, as compared to a condition in which the film was absent. A similar finding was earlier reported by Baumgartner et al. (2006) who, using affective still pictures rather than motion pictures, showed activation in the amygdala and hippocampus only in the presence of music; this led the authors to propose that music engages the emotional system whereas so-called emotional pictures engage a cognitive system (for a discussion of the neural basis of music-evoked emotions, see Koelsch 2010). The work by Eldar et al. and Baumgartner et al. suggest that music, rather than vision alone, provides the link to an emotional feeling state in film. Their data were also supported by subjective responses which showed additivity of emotional meaning from auditory and visual sources. The results concur with musicologist Nicholas Cook's (1998:23) description of the role of music in the multimedia context: "a bundle of generic attributes in search of an object....a structured semantic space, a privileged site for the negotiation of meaning." The level of rigor required or that can be expected is represented by the approach taken by Arbib (2010) and Arbib and Lee (2008) in their attempt to explain how a verbal description arises when information is visually presented. Details remain, however, to be resolved. Arbib and Lee (2008:151) show, for example, how the same visual image will lead to different grammatical interpretations depending on context. Moving this example to the present discussion, such context can be provided by musical structure arising from the first level of CAM analysis, or it can be provided by endogenous generation arising from the flow from long-term memory that has been primed by preliminary processes. The information is available to construct the theory within the CAM-WN framework.

Audience engagement or absorption in a film, particularly when facilitated by music, begs for consideration of the topic of consciousness. Revonsuo (1998) has focused on neural synchrony that leads to binding, which brings about consciousness (see also Crick and Koch 1990). Here I suggest that music input facilitates creation of such synchrony. Luo et al. (2010) obtained magnetoencephalography (MEG) recordings from participants who watched film clips with matched or mismatched audio tracks. Their analysis revealed that a "cortical mechanism, delta-theta modulation, across early sensory areas, plays an important 'active' role in continuously tracking naturalistic audio-video streams, carrying dynamic multi-sensory...information....Continuous cross-modal phase modulation may permit the internal construction of behaviorally relevant stimuli" (Luo et al. 2010:1–2). In particular, the auditory cortex was found to track not only auditory stimulus dynamics but also the visual dynamics as well. The converse was true of the visual cortex. The system favors

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

congruent as opposed to incongruent audiovisual stimuli. "The phase patterns of the ongoing rhythmic activity in early sensory areas help construct a temporal framework that reflects both unimodal information and multimodal context from which the unified multisensory perception is actively constructed" (Luo et al. 2010:10). It might follow that film music produces a source for electrical brain wave patterns that can bind with accent patterns from visual sources (e.g., runners in *Chariots of Fire*, the motion of Mark Zuckerberg's image in *The Social Network* as he traverses the Harvard campus, or Danny's finger motion along the knife blade in *The Shining*) but not necessarily only these visual patterns, perhaps patterns from sound effects (e.g., the sound of motion on the pavement, the sounds of speech patterns) or the visual patterns of articulation.

Hermann et al. (2004:347) describe gamma-band oscillations (30 Hz) that enable "the comparison of memory contents with stimulus-related information and the utilization of signals derived from this comparison." Their model attempts to explain early gamma-band responses in terms of the match between bottom-up and top-down information. Furthermore, it assumes that late gamma-band activity reflects "the readout and utilization of the information resulting from this match." This mechanism, which they describe as the *match and utilization model*, seems like that which is required by the CAM-WN model in preprocessing information from the initial sensory channels, activating relevant parts of long-term memory so that appropriate inferences can be generated to make the best match with the slower and more detailed processing of the initial encoding. Commonalities with the system proposed by John (2003) deserve mention, although John's reference to gamma band begins at 40 Hz, not 30 Hz.

A remaining issue is that of the mirror neuron system. Film is an ideal context for such study because characters in the film, with whom the audience identifies, provide images to be mirrored. In addition, music also leads to mirror activities in terms of dance and other movement (Overy and Molnar-Szakacs 2009; see also Calvo-Merino et al. 2006, 2008). Speech may engage the mirror neuron system (Arbib 2005a; Rizzolatti and Arbib 1998). The prefrontal and parietal location of the mirror neuron system may overlap with a common area of linguistic and music semantic processing. The CAM-WN model accommodates motoric and kinesthetic information by means of the kinesthetic channel. Information in the working narrative normally provided by the kinesthetic surface could be provided by musical information, following Overy and Molnar-Szakacs (2009).

CAM-WN makes sense of Hasson et al.'s (2008b) fMRI correlations arising from viewing film. All audience members received the same four conditions of stimulus input that differed in aesthetic control. Hitchcock's cinematic techniques were found to control processing to a greater extent than any of the techniques used by the other directors in the study. Presumably, Hitchcock is a master of directing audience attention to particular information in the visual scene. He provides sufficient information that passes quickly to

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

long-term memory, engaging narrative processes and hypothesis generation to match or not match information arising from the sensory bottom-up processing. Only some of the parsed and processed information is best matched and integrated into the ongoing working narrative. In contrast to the Hitchcock film excerpt, a video of New York produced few common hypotheses arising from the story grammar and long-term memory. Every viewer's experience or imagined experience of New York differs. Thus, each viewer would create a different working narrative. Adding music to this videography, however, might introduce constraints at several levels and lead to higher correlations among viewers. Future research is needed to test this.

The CAM-WN approach also accommodates Levinson's (1996) challenge of the much accepted view that film music is unheard (Gorbman 1987). He has argued that film music is by necessity heard so as to help the audience understand the position of the narrator and the filmmaker. He believes that the job of the audience is to understand how the narrative, as created by the narrator and filmmaker, is to be understood. CAM-WN certainly accommodates the simpler view of the audience as the creator of the narrative, but it also accommodates the higher-order levels of narrator and filmmaker, should information in the film be sufficient to allow this aspect. This allowance, in part, depends on the audience member's prior knowledge. Clearly an audience member steeped in philosophy, aesthetics, film, and music will have more knowledge to draw on than a member who lacks this experience. To test predictions from CAM-WN, group differences in consciousness of film music, narrator, director, and so on clearly require further psychological research.

With its three auditory channels, the CAM-WN model helps show the subtle balance between speech, sound effects, and music in film. Loud music will mask speech and sound effects, the two primary auditory contributors of diegetic information. The film will then become increasingly ambiguous and more easily influenced by the music because loud music draws attention (still potentially unconscious) to itself and diminishes diegetic information due to masking. Cognitive science and psychoacoustics can provide insight regarding the relative contributions of the sound effects, speech, and music, each of which can convey structural information and semantic meaning. Each of these acoustical modalities is a different kind of vehicle designed to transport some kinds of information better than others. To date, however, no psychological study has investigated the signal-to-noise ratio relations in the soundtrack mix of sound effects, speech, and music. Future studies are needed to address questions about how film music works in film, how the brain handles these multiple sources of acoustic information, and how it integrates them with the visual information.

The inclusion of music in film might be regarded as one of nature's gifts to cognitive science, because, as has been emphasized, the presence of music in film is illogical and paradoxical on several grounds. Other illusions that have dropped in our path, such as Treisman and Schmidt's (1982) illusory

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

conjunctions in vision or Roger Shepard's ascending staircase in the auditory realm, have provided insight into how the mind works. Likewise, the study of film music—from its paradoxes to its most obvious features—may help us unpack the mysterious relations of music, language, and the brain.

Acknowledgments

I wish to thank the Ernst Strüngmann Forum for the opportunity to participate. This chapter benefitted from the insight, inspiration, and encouragement of Michael Arbib, and the conscientious editorial work of Julia Lupp and her team. The Social Sciences and Humanities Research Council of Canada is acknowledged for the generous support of my research in film music for over two decades, and the Canada Foundation for Innovation is acknowledged for providing infrastructure in the CMTC Facility at the University of Prince Edward Island.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.