Sound + Image in Computer-Based Design: Learning from Sound in the Arts

Millicent Cooley

Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
and
IBM Corporation
Austin, Texas, USA

Abstract

Sound is underutilized in software and on the web, in spite of its obvious value to other media, such as film. Many practitioners in computer-based design, particularly those with backgrounds in programming and print design, are simply unfamiliar with the medium of sound. The performing arts has a long history of creating sound which makes a powerful impression on human perception and emotion, and has accumulated a rich body of theories and practical insights for how this is done. These theories and insights should be explored for their usefulness in improving sound design in software. The purpose of the research discussed in this paper has been to learn from the principles and practices of sound design in the performing arts, and to discuss and demonstrate ways in which some of these ideas might be helpful to designers of computer-based media and software.

This research considers performing arts theory with an emphasis on sound, validates some of this theory in the form of a series of interactive multimedia exercises, and describes commentary from performing arts professionals who discuss practical and theoretical issues in sound design from an experienced perspective. Games tend to make better use of sound than other computer-based products and, because of their narrative qualities, occupy a place in design somewhere between traditional performing arts and software. For these reasons, games have lessons to offer to other areas of computer-based design in terms of sound use, and some analysis of game design is included here, as well.

1 Sounds and images together

Discussions about multimedia design sometimes refer to 'the added value of sound'. Although it is true that sound can add important new information to a visual presentation, and strengthen the total experience, it should not be inferred that sound is merely confined to the supporting role behind the image. There are numerous instances of software design incorporating audio in which the sound plays an auxiliary role. However, it is also possible for sound to have importance equal with imagery, and even to carry the majority of the experience in a multimedia design, as we shall see further on. An additive model is useful when it allows us to analyze the independent benefits of sounds and images. But a view of multimedia design in which auditory and visual modes are separate compartments is incomplete [1]. A more mature and integrated model is offered in the performing arts, where sounds and images occur in concert with every other element in the production, and where the whole experience is greater than the sum of its parts.[1]

2 Principles in sound-image interactions: Chion's analysis of film

Michel Chion analyzes the role of sound in films in his book *Audio-Vision*. His thesis is that "...one perception influences the other and transforms it. We never see the same thing when we also hear; we don't hear the same thing when we see, as well" [1, p. xxvi]. He asserts that people possess a natural urge to fuse

[1] Chion says it is not uncommon in the film industry for people to view film sound as an 'add-on' [1].

sounds and images into a cohesive whole as a strategy for making sense of the world, and that this tendency extends to our experiences of viewing films. When this fusion occurs, the meaning perceived by one sense contributes to the meaning perceived by the other.

Chion describes this effect with the term *synchresis* (..."the forging of an immediate and necessary relationship between something one sees and something one hears") which combines *synchronism* (things which occur at the same time) with *synthesis* (perceived relatedness of meaning) [*op. cit.* p.5]. The concept of synchresis explains not only why viewers accept cinematic events that are as logically connected as a banging noise and a slamming door, but also those in which the sounds and images have little or nothing in common with one another. It accounts for how cinematic and theatrical sound can be effective in so many different ways, and suggests the potential for rich and varied use of sound in computer-based media as well.

Chion discusses many uses for sound (he says there are actually hundreds of ways to put images and sounds together) including its use for punctuation, for unification of imagery which would otherwise seem unrelated, and for creating a sense of anticipation about what will follow. Several general ideas important to his thinking include: 1) sound's effect on our perception of time, 2) sound's effect on our perception of space, and 3) the degree to which the sounds and images in an audio-visual event are related, and how this relationship is used to create meaning.

In Chion's discussion on how sound influences our perception of time, including our perception of movement, speed, rhythm, and pacing, he claims that "...the eye is more spatially adept, the ear more temporally adept... the eye perceives more slowly because it has more to do at once; it must explore in space as well as follow along in time" [*op. cit.*, p. 11]. He says that people are more capable of acutely tracking the details of motion with the ear, than with the eye, and that people who possess both sight and hearing can usually understand spoken language faster than they can read.

In film, sound can deliver the illusion of motion which isn't actually visible. Chion gives the example of a scene in *The Empire Strikes Back* where an automatic door opens; first there is a static shot of the door closed, then another of the door open, along with a dramatic whooshing sound. This is enough to make audiences think that they have seen the door move [*op. cit.*]. Computer animators rely on this principle to create an appearance of motion when they want to minimize the use of images.

Motion sometimes occurs in rhythmic patterns, and although rhythm can be established using sequences of visual images, it is more intensely felt using sound. Chion explains several ways that rhythmic sound can control human attention [*op. cit.*]. One way has to do with how a tone is sustained; a tremolo, or fluttering pattern, is more demanding of our attention than an evenly sustained tone. A second way has to do with the predictability of the rhythmic pattern; a regular pattern tends to move into our peripheral awareness as "background", against which other more singular auditory elements stand out, while an irregular pattern calls more attention to itself, and makes listeners more alert. Another way that sound can control attention has to do with frequency levels; sounds containing a greater balance of high frequencies seem closer, and will command more attention [6].[2] Advertisers use this principle when they pump televisions commercials full of high frequency audio that makes characters sound as if they are intruding into viewers' homes.[3]

These sound characteristics for controlling attention could help guide the design of navigation that attracts users to important information, or the design of a set of alert sounds which express varying degrees of alarm, ranging from mild alarm to emergency, in a natural way. There is a particularly effective alarm sound in my laptop for indicating a crash which is serious. It is piercing and intense, like the high screeching crash of breaking metal and glass shards, and distressing to hear-- a sound that doesn't need to be learned.

---

[2] Chion points out that volume is not the only indicator of the distance of a sound source by explaining, "The ear detects depth from such indices as a reduced harmonic spectrum, softened attacks and transitions, a different blend of direct sound and reflected sound, and the presence of reverberation." [1, p. 71]

[3] It is not obvious to an average listener what frequency and rhythm have in common, because most sound frequencies do not have audible rhythmic characteristics, however at very low frequency levels these rhythms do becomes audible [2]. Sound waves themselves form a type of rhythmic pattern, and higher frequency sounds are those patterns where more waves pass a given point within a fixed period of time [3].

3 The use of sound to define physical and non-physical space

An important group of ideas Chion discusses has to do with how sounds define physical as well as non-physical space, and these ideas have potential for use in the sound design of software which supports distance collaboration and communication among workgroups, including email and chat. Peoples' general satisfaction with email and chat forums sometimes seems to suffer from the lack of physical, human presence in a number of ways. Sounds can give a stronger sense of presence to events and objects, and in this context, could help make remote, unseen participants seem more real to each other, as well as provide a number of useful cues about shared and individual activities.

Chion says that images *magnetize* sound [1]; if a movie displayed on a large screen shows feet walking from left to right, the sounds of walking seem to emanate from the image, moving to the right along with the feet, even though in reality the sounds may be coming from one place, the speakers. Sounds often imply a sound source, and sound sources are located somewhere. An exception to this is *non-diegetic sound*, which is sound that is outside the story, such as narrative commentary, voice-over and musical scoring. *Ambient* sounds are generalized; they permeate and surround a scene without attracting attention to any particular source. *Territory* sounds call more attention to a particular area of the surroundings than ambient sounds. *On-screen sound* is linked to a visible sound source, and *off-screen sound* emanates from a sound source which is not visible. Off-screen sound can be *passive* (it doesn't seem important and remains in the background) or *active* (it seems important and attracts attention or curiosity). *Sound-in-the-wings* indicates a sound source that is about to enter the scene or has just left [*op. cit.*].

How could these concepts be applied to the design of a collaborative workspace? I have some colleagues designing communication software with an interesting sound feature; while someone composes you a note an empty message window appears on your screen and you hear typing sounds, until the note is sent and fills the message window. This could be interpreted as a good example of how *off-screen sound* deepens participants' awareness of existing but not-visible physical space, and supports the sense of shared social space. The example could be extended in the following scenario.

Imagine yourself as a member of a group of competitive market analysts working for a large corporation, using collaborative software to share information and make analyses. Clusters of group members are distributed in offices throughout your country and overseas. Your software allows you to conduct group conversations in various topic areas and exchange private messages. At any one time a number of exchanges could occur which may be relevant to the industry areas you follow. The quiet, uneventful activity of distant colleagues conducting work privately on the network could be represented by the *ambient* sounds of smooth machine hums. Irregularity introduced into the pattern of these ambient tones might signify that small conversations have begun to occur. *Territory* sounds might indicate a more lively exchange between members in one topic area; without revealing the substance of the exchange. An *active off-screen* sound sent from one topic area might be an alert sent by a member to notify the group that a significant financial event, such as an important merger, has just occurred; everyone temporarily leaves their other topic areas to find out about the news. Everyone can identify which topic area the alert sound relates to because a graphical indicator next to the topic label on the interface vibrates; this attracts their gaze associates itself with the alert sound, or *magnetizes* it. Another team member who heard about this event down the hall has entered one of the rooms and logged onto his computer in order to join the conversation; while he boots up his software, everyone else who is logged on hears a *sound-in-the-wings* audio signal, indicating that he is about to enter the shared space. When he arrives and reads the news, he presses 'control + !' on his keyboard, expressing his reaction. This sends a lively "boink!!" sound, an audio emoticon, which translates roughly as "How about that one!" During the lively exchange that ensues, you get interrupted by the *non-diegetic sound* of your system agent's voice, reminding you that in order to be on time for an important appointment, you must leave in 10 minutes.

Chion's particular terms may seem awkward for use by programmers, but the concepts they describe could be useful nevertheless, and the labels for the concepts could be modified appropriately for use in the software world. Their meanings are worthwhile because they refer not only to such familiar physical features of space as closer-farther or left-right, but to other features as well, such as intimate-public, seen-unseen, and important-unimportant. These are non-physical spatial distinctions which software designers may be less accustomed to thinking about, particularly with regard to the use of sound.

4 Audio-visual consonance versus dissonance

Another important theme in Chion's theory has to do with the degree of relatedness between combined sounds and images [1]. A consonant or *literal* use of sound and image equates with 'redundancy' in interface design terms, and describes an audio-visual combination where the sound and image each carry the same meaning. In a more dissonant or *cross-literal* audio-visual combination the sound carries a meaning which is different from the image, and the sound and image tend to change each other's meaning. It's useful to recognize when audio-visual combinations should be redundant and when the design would benefit from a more dissonant use.

Redundant auditory cues can help distinguish visual cues which may be otherwise easy to confuse, such as when the visual cues are small or when many of them are placed near each other in the design. Jim McKee, sound designer at Earwax Productions, Inc., pointed out that using sound can make fine motor movements easier by reducing eyestrain that sometimes results from having to focus on small and/or finely detailed visual navigational cues on a computer monitor. Sound can even mimic a sense of texture, which could help make a person using a computer "feel" as if they have made contact with a particular element, without relying as intensely on their eyes, in order to know this [7].

Audio-visual redundancy in the design of a set of cues can offer users the choice for selecting either the auditory or visual mode to represent those cues. This approach would support adaptability to personal preference, individual ability (widening the audience for a product to include people with either hearing or vision impairments), or the needs of a work and/or social situation (allowing an owner to either leave their computer yet still hear the cues, or stay at their desk and turn off the sound, perhaps to avoid irritating nearby colleagues.)

Audio-visual dissonance, on the other hand, allows a design to say more with less, by way of suggestion, which is both engaging and efficient. In Walter Murch's book *In the Blink of an Eye: A Perspective on Film Editing* he points out that suggestion, rather than exposition, engages the participation of an audience [8].[4] Imagine a web site which sells cars; when you click on a visual icon of one of the cars, you hear the low roar of a wildcat ready to pounce, and when you click on an icon of a competitor's car you hear a sputtering sound like an engine running out of gas. More meaning has been conveyed than if a more straightforward car horn beep were used for both images, there is enjoyment in engaging in the suggestion, and the point comes across quickly.

In its most extreme version, audio-visual dissonance occurs when the sound carries one meaning and the image carries an entirely different meaning, and in the combination a third meaning is created which is different from either of others. Walter Murch, who wrote the foreword to *Audio-Vision*, explains Chion's use of the term *sound en creux*, or 'sound in the gap' as a "purposeful and fruitful tension between what is on the screen and what is kindled in the mind of the audience". He likens this concept to a component of depth-perception, where each of our eyes see a separate image, and our mind resolves the disparity between the two to form a third image available to neither one alone, the image of depth (Murch in [1]). Towards the end of the film *Good Morning Vietnam* there is a scene of a field with a path, surrounded by jungle; it is a beautiful day and people walk about. In the field a straw hut, a family's home, goes up in flames, ignored as if it were just another everyday casualty of the war. Added to this scene is music, the rough voice of Louis Armstrong sweetly singing "Oh, What a Wonderful World". The combination of the image and music creates an experience different from and greater than the sum of what they each represent independently-- a sense of bittersweet irony about the pain and beauty that exist alongside one another in this situation.

Use of audio-visual dissonance this pronounced, while common in the arts, is complex to achieve and seldom seen in computer-based design.[5] In multimedia where artistic goals are present, this is a

---

[4] Walter Murch has received numerous awards for his work as an editor, sound designer, director and screenwriter, including from both the British and American Motion Picture Academies. His well-known films include: *The Conversation*, *Apocalypse Now*, *The Godfather*, *Part II* and *Part III*, *The Unbearable Lightness of Being*, *Ghost*, and *The English Patient*.

[5] Chion acknowledges that dissonant audio-visual combinations do not always work. He explains there are many ways that dissonance manifests, that synchresis is inexact, and that it is affected not only by the meanings carried by the particular sounds and images being combined, but by the context of the film story, and the personal and cultural biases of the listener/viewer [1].

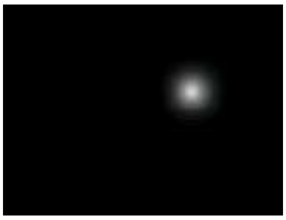
powerful approach to expression that could make designs more deeply engaging and imaginatively compelling.

5 Demonstrating principles in sound-image interactions

In May, 1997 I completed a masters thesis project at Carnegie Mellon University called "Sound + Image in Design", which is a structured series of five interactive multimedia design studies demonstrating some ways that sounds and images affect one another. Every exercise in the series focuses on a different issue, and they are all similarly structured. Four present an animation and offer different options for adding sounds, and one shows different animations, each accompanied by the same (or similar) sound. Each exercise is accompanied by a set of questions for the viewer or audience to consider while experimenting with the different sound-image combinations. Following the completion of the design, a small group of eight people were queried about their responses to the exercises, in order to learn from these responses and to make sure that each exercise successfully communicated its point. The most common response was a sense of surprise that the different sounds could have such strong effects on the images (or in one of the exercises, that the images could alter the perception of a sound).

The first exercise, named "Bumping Squares", demonstrates not only how sound can make images seem more real, but how it can influence a person's perception of a variety of physical characteristics of an object. Whenever one of the sound controls is selected, the squares move toward the center and bump into each other, one time. The user is asked if the various sounds make the squares seem light or heavy, dense or hollow, rough or smooth, large and far away or small and near.

The second exercise, "Walking Triangles", demonstrates how non-vocal sound can help anthropomorphize an inanimate object, and give it a sense of personality. Each time one of the sound controls is selected, a triangle proceeds across the screen from left to right with a wobbling or rocking motion. The user is asked whether the triangle seems like an animate character with personality traits, attitudes and feelings, and how the sound contributes to this impression.



The third exercise, "Boat", demonstrates how we use sound to judge the distance of a sound source. The screen is dark, except for a single shining light, and there is a continuous sound, like water lapping up against a hard surface. The user is requested to think of this image as part of a scenario, where they are in a sailboat at night and hear a warning sound from another boat somewhere in the area; the shining light is on that boat, but they don't know how far away it is. The user is asked about their judgment regarding the distance of the boat, each time they select one of the sound controls.

There are 4 boat horn sounds; beginning with sound #1, each sound seems progressively closer. Sounds #2 and #3 seem as distant from each other as sounds #1 and #2, and as sounds #3 and #4. Yet the actual difference in the volume between sounds #2 and #3 is negligible, and too small to account for the how far apart they seem in distance. The main cause for the perception of distance between sounds #2 and #3 is that #3 contains a greater proportion of high frequency tones. As stated earlier, if volume is held constant, sounds which contain a greater balance of high frequency tones seem closer than sounds at the same volume level, but with a greater proportion of low frequencies.

The fourth exercise, "Petunia", demonstrates that just as a sound can affect our perception of an image, an image can also affect our perception of a sound.[6] There are six different visual animations, and the same sound is put to each of them, with only slightly different variations in the rhythm. Users are asked whether they find each audio-visual combination convincing. There were varied opinions regarding the believability of these combinations, ranging from strongly affirmative to strongly negative, however in general, more responses were affirmative. Given that the sound for all of the images is a pig's oink, and that only one image in the series bears the slightest resemblance to a pig, the amount of affirmative responses regarding the believability of these combinations might support Chion's claim that people's proclivity for fusing images with sounds is strong enough that they will do so, even against the logical evidence [1].

The images in the first two exercises ("Bumping Squares" and "Walking Triangle") were the easiest ones to combine sounds with, that people found convincing. These images are the most abstract ones in the series, which might suggest that it is easier to make a believable audio-visual combination when the image is abstract than when it is more realistic, that abstract images are malleable to suggestions of sound. Respondents made more critical observations about the sounds of staple guns and toilet plungers than of triangles and squares; commments were made such as "my bicycle pump doesn't sound like that", but no one ever said, “a triangle doesn't sound like that.”

The fifth exercise, "Girl's Story", explores how sound and music can shape the way we experience narrative. Kinetic (moving) text is used to tell a little story about a girl who gets lost. The text is designed so that its motion is expressive in a way that is relevant to the story. There are three different sound tracks in the exercise, two with music and one with sound effects. Of the two with music, the mood of one is carefree and breezily happy, and the mood of the other is somber and sad. Users were asked for their interpretations and feelings about the story when they watched the moving text narrative and listened to each sound track.

In the version with somber music, respondents said that the girl's getting lost seems sad, frightening, or dangerous, in contrast with the version containing the carefree music, where her getting lost seems harmless or amusing. In the version with sound effects, the sounds mirror the story in a literal (consonant) way through the majority of the piece, making the experience seem somewhat more real but adding little new information. Near the end of the story the meaning of the sound effects depart on a more

[6] When films are made showing scenes of horrific violence to body parts, the sound effects have to be created from scratch using inanimate objects; the sound of a skull crushing, for instance, has sometimes been made by crushing melons [1].

independent (dissonant) direction from the text. This is the section that seems to engage people the most, and I would speculate that the reason for this is not only the dramatic characteristics of the sound (a bad car crash), and the situation it suggests, but the fact that the sound and image are dissonant and listeners are challenged to imagine what really happens in the scene. Not everyone imagined this scene identically.

6 Dramatic analysis for understanding sound in computer games

The *You Don't Know Jack* computer game series, a marketing success created by Jellyvision, offers an excellent example of multimedia design where the sounds play a more important role than the images. As a member of the genre of games known as trivia games, *Jack* ("the trivia game with attitude") challenges players to win by correctly answering entertaining questions on a variety of ephemeral subjects, at a fast pace, while being distracted with humorous taunts and jokes. *Jack* is unique among computer games because of its emphasis on sound, particularly through its use of dialog, sound effects, and vocal pacing, and because of the minimal and abstract use of imagery. And contrary to the popular adage in application design about the desirability of giving control to users, this game sets the pace and maintains control. For these reasons, it is instructive to analyze how sound works in *Jack*.

There is a classical approach used to analyze films and plays that can be applied successfully to understanding the sound design in *You Don't Know Jack*. It is appropriate to use this method to analyze games because of their inherent narrative qualities. Don Marinelli teaches drama and co-directs the Entertainment Technology Center at Carnegie Mellon University, and is interested in applying dramatic principles to computer media design. He teaches students to analyze computer games such as *You Don't Know Jack* using this classical dramatic approach, which is based on ideas introduced by Aristotle [9]. In *Computers as Theatre*, Brenda Laurel uses classical dramatic theory to discuss human-computer interactions; she states that Aristotle's theory of drama is so "comprehensive and well-integrated" that it has continued to influence dramatic theorists to the present day [10].[7] Aristotle described a model of six elements occurring throughout a play as *character*, *story* (or action), *language* (including pantomime or gesture), *ideas*, *rhythm* (including pacing or dynamics) and *spectacle* [11].[8] The goal of the analysis is to describe the structure of the play, and how each of the six elements contributes to the experience. Having done this, it becomes possible to identify what sounds contribute to each element. It can be seen, for example, what sound contributes to expressing the characters, or to defining the rhythm of the play or computer game.

7 Analyzing sounds and images in You Don't Know Jack

Understanding the use of sound in the opening section in *Jack* is sufficient for understanding the approach to sound design throughout the entire game, and this analysis focuses only on that section, although it is possible to apply to the entire game. The opening section in *Jack*, which is about 50 seconds long, is the equivalent of the exposition in a play. Its function is to introduce the characters and the story; set the tone, and prepare the audience for the experience to follow. The analysis is applied to *Jack* to identify not only what sound contributes to each element of the design, but to compare the relative contributions of sounds and images in *Jack*. The analysis is written to show what someone learns about the game regarding each of the six elements in the model, and what causes them to learn it. The first part describes what is learned about these elements through the sounds, and the second part describes what is learned about them through the images.

SOUNDS: The dialog, vocal textures, accents, and tones of voice (*language*) all convey the idea that the *setting* (spectacle) is a contemporary American television game show about to go on the air, with a main *character* in the role of a dominating, sarcastic game show host, and supporting characters who play

---

[7] Laurel also refers to the contributions of dramatic theorists other than Aristotle, for human-computer design. In *Hamlet on the Holodeck*, Janet Murray explains narrative issues in software and computer media design; in her discussions on theatre, she recognizes the contributions of the Greek classics, but rejects some of Aristotle's conventions [4].

[8] In *Fundamentals of Play Directing*, Alexander Dean and Lawrence Carra provide a more thorough description of the Aristotelian concepts of dramatic structure, and how they are applied in modern theatre [5].

support staff and production crew. Sound effects reinforce the game show setting. The *situation* (story) you face is that you will be aggressively challenged to prove that you know jack (anything at all), and you know this, again, because of the dialog and swaggering, aggressive tone of the host. The machine is in charge and you, the player, are not; the game is quick-paced, there is a sense that you will be rushed along and should try to keep up and prove that you do, in fact, know jack. You feel this pressure because the voice of the host rushes you to sign in, taunting you impatiently at every step (*rhythm*). There is a feeling of excited anticipation about 'going on air' due to the hurried, anxious voices of the supporting characters who interrupt and talk quickly over one another, and because of the introduction of the type of brassy, super-enthused music that tv game shows have come to be associated with. There is a sense that you will be kept off balance, that it will be chaotic and surprising, that there will be a lot to pay attention to, and that it will be fun; the dialog and sound effects indicate a rapid succession of sub-plots where weird things go wrong, the jokes are witty and fast-paced, and you are constantly teased. You know the machine is smart and is paying attention to you, and that it will probably remain interesting, because the jokes are appropriate to your actions and when you repeat scenes you hear new jokes, rather than repeat ones. The material (trivia questions about movies) is ephemeral and irrelevant, and the machine knows this, but it's fun anyway (the *idea* of the game) [9].

IMAGES: The only images in the opening section are the entry field in which to sign your name to start playing, a prop in the *setting* which is an instruction sign for a game show audience (like the ones that tell an audience when to clap), a good deal of kinetic text (*language*), and a few spinning control buttons. All of this takes place against the *spectacle* of a flat black field, leaving the rest to your imagination.

There are two general points to take from this analysis: 1) the confirmation that sounds contributes a great deal more to the total experience of *Jack* than images, and 2) dialog and sound effects are both used, but the dialog and the style of delivery of the dialog play a bigger role than the sound effects.

8 Usefulness of dramatic analysis for sound design in computer software

The approach that has been discussed applies most appropriately to software which has some sort of narrative structure, such as games. It can also apply to other types of software including educational software, software designed for use as a sales tool, or on-line advertising. It seems less useful for products which are clearly non-narrative, such as software tools. However, even these applications have opening (exposition) and closing (resolution) sections, however brief, and using this approach as an exercise might prompt designers to consider these sections in a new light [12]. It might, for example, support a closer examination of system start-up sounds and the expectations that those sounds might create for users. Even when it is difficult to apply this approach to a particular design, it could still be used as a vehicle for learning and communicating about sound design, and lead to sound design ideas which could be applied elsewhere. Nevertheless, although dramatic and narrative theory has begun to be used for understanding non-linear media, it remains to be seen how far the boundaries of its usefulness for all types of software design can legitimately extend, and hopefully those boundaries will continue to be tested.

9 Practical issues in designing sound for You Don't Know Jack

Michelle Gorchow, a director at Jellyvision with a background in film who worked on Jack, points out that the use of sound in Jack resembles radio sound [13]. In film much of the sound emanates from elements in the image, but in radio a soundscape is created purely from sound. Radio scripts are written so that listeners will continually be led to picture images and scenes in their minds based on what they hear. Gorchow also noted the common and irritating problem of overly repetitive computer sounds, and described the trouble they went to in order to keep the experience of re-playing *Jack* seem fresh by writing and building in many different jokes and responses to various user behaviors. And indeed, I have found that with repeated uses of the game, the jokes seldom seem redundant, and when they finally are repeated, enough else has usually happened that they are still fairly funny. Bezerk.com, the web version of *Jack*, allows people to download new jokes every week.

Dave Houghtaling, creative director of *Jack*, explained that emphasizing the use of high quality sound while minimizing imagery proved to be a good engineering strategy for *Jack* [14]. He described that their initial instincts to keep imagery simple and to rely on the strength of the writing, performance and sound effects were reinforced when they realized that it was easier to make a responsive audio-visual environment that contained simple graphics; he stated that high-quality sound files are much easier to load than full-motion video files. Houghtaling noted the example of the game *Myst*, which contains many images that are beautifully rendered but very large; these heavy images cause a perceptible wait time at scene changes while they load. Since a slow, deliberate pace is appropriate for the concept of *Myst* to begin with, some degree of lag may not pose a critical problem. However, since timing is critical to humor, in general, and *Jack* is fast-paced, quick and responsive technical performance is critical to the success of the game.

Contrary to the view held by some developers that rich, detailed images make a design more immersive, *Jack* is immersive without them. Martin Striker, producer from Berkeley Systems on the movie version of *Jack*, expressed his opinion that a literal use of imagery would be distracting and would actually diminish the experience of the game, instead of making it more real [15]. He asserted that the attempt to make images on a computer screen look real, which are obviously not real, just makes users less immersed because they remain conscious of a weak illusion, but that the sense of suggestion offered by convincing sound and engaging dialog combined with visual abstraction is actually more immersive. Striker also stated that half of the budget for the movie version of *Jack* was spent on script writers, which was far more than on sound design and production; this is an interesting commentary for a game in which spoken dialog is so important.

10 Conclusion

In order for sound design in software and computer media to become improved, sound designers need to be integrated into development teams. And in order for programmers and sound designers to communicate well, they need a common language and some shared models. Jim McKee, at Earwax, pointed out the problem of ideating sound design with non-musicians, stating, "you can made a sketch of a house that a non-graphic designer can understand, but non-musicians rarely understand an audio sketch." He said he is able to use language to describe a request for a sound from a musician or sound designer, and the result will be nearly identical to what he asked for [7]. This would indicate the need for programmers and software developers to become fluent using technical audio terminology. But discussions of sound design involve expressive and aesthetic issues, as well as technical ones. Chion's contribution in categorizing, describing and naming many kinds of audio-visual interactions can be helpful in these discussions.

The analytic model discussed in this paper, originally conceived for understanding theater but applied to understanding certain types of multimedia software design, can also support communication between sound designers and programmers; it has stability and integrity due to a long history of use and scrutiny, it is holistic in considering how sound is integrated with the other elements, and it supports the perspective of sound as an agent for engaging users.

References

1. Chion, M. (1994). *Audio-Visio*n. New York: Columbia University Press.

2. Martin, S. (1998). Personal conversation, September 1998, Austin, Texas.

3 Wilentz, J. (1968). *Senses of Man*. New York: Thomas Y. Crowell Company.

4. Murray, J. (1997). *Hamlet on the Holodeck*. New York: The Free Press, div. of Simon & Schuster.

5. Dean, A. and and Carra, L. (1965). *Fundamentals of Play Directing*. Chicago: Holt, Rinehart and Winston, Inc.

6. Cooley, G. (1997). Personal conversation, March 1997, Pittsburgh, Pennsylvania.

7. McKee, J. (1998). Interview, September 1998, San Francisco, California.

8. Murch, W. (1995). *In the Blink of an Eye: A Perspective in Film Editing*. Los Angeles: Silman-James Press.

9. Johnson-Heckbert, B. (1998). Personal conversation, September 1998, Pittsburgh, Pennsylvania.

10. Laurel, B. (1993). *Computers as Theatre*. New York: Addison-Wesley Publishing Company

11. Aristotle (1996). *Poetics*. Translation, introduction and notes by Malcolm Heath. London: Penguin Books.

12. Marinelli, D. (1998). Personal conversation, September 1998, Pittsburgh, Pennsylvania.

13. Gorchow, M. (1998). Interview, September 1998, Chicago, Illinois.

14. Houghtaling, D. (1998). Interview, September 1998, Chicago, Illinois.

15. Striker, M. (1998). Interview, September 1998, Berkeley, California.