The Language of Electroacoustic Music with Moving Images

John Coulter

School of Music, National Institute of Creative Arts and Industries, University of Auckland

In the late 1990's a range of inexpensive digital video cameras and non-linear editing systems became available to the general public. To many Electroacoustic Composers, this new technology offered a multiplicity of creative possibilities, including the potential to advance public appreciation of electroacoustic music (EAM). The artistic merits of the new domain seemed to be supported by the scientific claim that "integrating visual and auditory stimuli serves the purpose of enhancing perceptual clarity[1]" However, nearly ten years later, expert opinion remains divided over the functionality of the language of EAM with moving images. The two opposing viewpoints, succinctly stated, are that adding moving images to electroacoustic music adds to its communicability, and that adding moving images to electroacoustic music subtracts from its communicability. This paper presents both arguments from the perspective of the composer-researcher. Research questions and hypotheses are raised through observations made in creative practice while functional aspects of language are examined using more traditional research methods such as literature review and listening tests. Results suggest that the language of EAM with moving images is governed by general principles of audiovisual attention, and that techniques of audiovisual composition may be categorised according to these principles.

Acknowledgements: Staff and Students of Griffith University and the University of Auckland.

Introduction

The study begins with a definition of the topic under investigation. The term "language" refers to a simple Chomskian model that contains both levels of grammar and levels of meaning (surface and deep structures); the term electroacoustic music is defined by Emmerson and Smalley (2001) and Landy (2000) as encompassing acousmatic and live genres, as well as "innovative forms of popular music, … a good deal of music for film and television and some computer games"; and the term "moving images" refers to video or film that is projected simultaneously during playback of electroacoustic materials. This definition addresses the issue whether EAM with moving images should be considered a style of EAM, or whether there is no novelty beyond that of the language of film. The definition is aligned with the view of Francois Bayle (1997) that "It is a question of the level of description. "Electroacoustic Music" is a generic term that describes a technical means. It does not usually refer to style or philosophy".

Research findings and theories concerning the topic may be drawn from a range of literature including the creative arts and psychology, as well as studies that concern the intersection of these domains. Firstly, there is much to be gleaned from the writings of Chion, Cook et al who directly address the topic from the perspectives of film sound, and multimedia art; secondly there is a body of literature from scientists working in the field of cognitive psychology such as, Doty, Talsma, and Woldorff who provide empirical evidence concerning the human ability to attend to audio, visual and audiovisual materials; and thirdly there is the work of those who straddle these domains such as Bregman, and Riess Jones and Yee. These findings and theories address the topic from a number of standpoints, which support the arguments either for, or against the use of moving images with EAM.

The study makes use of a custom designed practice-led research model based on Winter's 6 rules of action research (1989). In essence, the method relies on the expertise of multiple participants to add rigor to discoveries that were made through engagement in the creative process. As part of the 18-month project, a small research group was established first at Griffith University and later at the University of Auckland consisting of 10 Electroacoustic Composers. Participants contributed to group discussions and undertook a series of listening tests related to the experience of attending audiovisual materials. In ideal circumstances

[1] [Talsma, Doty, and Woldorff 2006] This view is still upheld today by the majority of scientists working in field of cognitive psychology.

this sample would have been larger, and several demographics would have been tested including musicians and members of the general public. However, in the given circumstances the data obtained is conclusive, and as such is included in the study. The general practice-led model is also at the heart of several other papers by the same author, so for reasons of its wide application and bulk, I have taken the unorthodox approach of omitting further details from this paper - rather, I would encourage readers to study the method as a parallel text. For further details please see Coulter 2007, *Practice-Led Research: A Method for Composer-Researchers.*

Due to the nature of action research, results are presented in chronological order, in tandem with methodological details concerning the individual actions undertaken. A summary is also offered at the end of each cycle followed by a restatement of the research question.

Initial Assumptions

Let us begin by investigating an assumption that was made on the part of this Composer-Researcher during the process of realising *Shifting Ground* - a piece for EAM and moving images. I had completed a section of EAM that contained several concrete sounds, and had engineered the phrase so the audience would recognise at least 6-8 sounds. I then added moving images of rockslides to the section. Later in the year, after the completed work had been played several times, I had the opportunity of questioning a respected colleague on his reaction to the materials. To my surprise, he did not recognise any of the referential sounds within the section discussed. Further questioning revealed that almost no one in the audience had recognised the sounds, and that most had heard either “rocks” or “noise”. The assumption that I had made was similar to that of many composers working in the domain in the late 1990s, that adding moving images to EAM would not disrupt the existing sonic language, but would simply add to it. This was not the experience of my colleague, or the majority of the audience members in seeing and hearing the work.

The Audiovisual Contract

I would now turn to literature for explanation. Chion (1994) describes the language of sound with moving images as being primarily founded on visual language, referring to "cinema" as "a place of images, plus sounds" with sound being "that which seeks its place". Here, and in similar settings, he proposes that listeners enter an "audiovisual contract" where it is the function of sound to "add value" to the visual images presented. Chion also describes the ever-present condition of “synchresis” as “when vision affects sound and vice versa” and "acousmatic sound" (audio without images) as that which "draws our attention to sound traits normally hidden from us by the simultaneous sight of the causes - hidden because this sight reinforces the perception of certain elements of the sound and obscures others. The acousmatic truly allows sound to reveal itself in all its dimensions.”

Scientific tests also provide evidence supporting the audiovisual mode proposed by Chion. As Woldorff (2006) states “attending to the visual modality would affect the multisensory integration process differently than attending to the auditory modality”. Furthermore, Talsma, Doty and Woldorff (2006) have found that “audiovisual integration processes appear to associate the visual and auditory stimulus components with each other, even when only the visual component was relevant”, thus supporting the condition of “synchresis” proposed by Chion.

To further investigate the phenomena observed in the playing of *Shifting Ground,* various tests were devised. Two groups of five postgraduate students participated in the trials, which involved written and verbal responses to audiovisual stimulus. Subjects were asked to respond to audio only stimulus with eyes shut, then to respond to audiovisual stimulus with eyes open, and finally, to switch between the two modes using eyes shut and eyes open techniques. The tests were designed to measure “ordinary reactions” to audio, and audiovisual stimuli. In normal circumstances, attending to audio, visual, and audiovisual stimuli is an automatic response that is governed by the requirements of the circumstances. However, individuals are also able to change their experiences of a situation by focussing their attention on one aspect or another. To avoid the unilateral use of selective attention during the testing process, subjects were directed to attend audio only with eyes shut, and audio and video simultaneously with eyes open. Two extracts of work were selected for testing based on their juxtaposed audio and visual contents. The first extract was the section of *Shifting Ground* previously described featuring transformed concrete sounds and

referential moving images. The second extract was a section of *Free Radicals* by Len Lye featuring instrumental music and abstract moving images.

Results were as follows: In 9/10 cases, subjects reported hearing more while attending audio only, irrespective of the style of audio being attended. Comments also indicated that generally it was the quietest components of the sounds that were "hidden … by the simultaneous sight of the causes". In 10/10 cases, subjects reported a change in the quality of audio when attending audiovisual; however, only 4/10 reported this as a positive change. Comments ranged from "the images make it fun" to "I didn't like being told the answer". Two participants referred to mismatches in the perception of "space" brought about by audio and visual materials as the leading cause of disconnection between audio and visual textures. Overall, the results support the existence of the audiovisual contract as proposed by Chion.

To summarise at this point, results from the first action research cycle suggest that audio and audiovisual materials may be classified into two experiential modes: the audio-only mode - a well-suited category for acousmatic EAM, and a general audiovisual category containing film, television, EAM with moving images, and a host of other audiovisual genres. Results also suggest that in the process of adding moving images to EAM listeners are forced to change modes from audio-only to audiovisual irrespective of audiovisual style, and that the audiovisual mode changes the quality of sounds, and hides certain aspects of sounds irrespective of audio style.

These findings offer a satisfactory explanation of the phenomenon observed during the playing of *Shifting Ground;* however, they appear to be in direct opposition to the initial assumption made (that adding moving images to EAM would add to communicability), and to the scientific claim that forms the basis of the assumption, that "integrating visual and auditory stimuli serves the purpose of enhancing perceptual clarity" (Talsma, et al 2006).

To investigate this paradox further, let us first consider a situation in which both rules may apply. If the audiovisual contract is always enforced in the presence of moving images, then "enhancing perceptual clarity" would always be at the cost of hearing less detail in the audio track. This does not seem likely, as it conflicts with a large body of research concerning audiovisual attention (the very research that produced the scientific claim). A more adequate explanation might be that Talsma's statement is valid within the context of specific audiovisual styles. Notwithstanding the constraints of the audiovisual contract, one would assume that in certain circumstances audio and visual media could compliment one another, and that in ideal circumstances the simultaneous sight of the causes could enhance the qualities of the sounds.

This proposal changes the research question from "whether adding moving images to EAM adds to, or subtracts from, its communicability" to "whether adding moving images to EAM subtracts from its communicability in all circumstances". To answer this question, we must first form a classification system of experiential modes of hearing/listening in audio-only settings, and experiential modes of seeing/looking in moving-image-only settings. We must then systematically test all possible combinations of media to determine whether details in the audio materials are obscured in all circumstances.

Modes of Listening

Pierre Schaeffer proposed four modes of listening:
"*information-gathering* where we are occupied with the provenance of the sound and the 'message" it carries;… *passive reception,* where the listener cannot help hearing a sound… *appreciating and responding to attributes of sounds* ignoring any mode one messages they might contain… [and] *responding to a musical language* (Smalley 1996). Chion (1994) provides us with the term "the acousmatic" - a general audio-only category that may be further classified into three modes of listening that are related to Schaeffers modes. The three modes are: *causal listening* which involves recognition of sound sources, *semantic listening*, which involves the decoding of language, and *reduced listening* which involves "focussing on the traits of the sound itself independent of its cause or its meaning". The basis for modal classification is the human ability to attend both data-carrying and form-carrying attributes of any given sound or image – as Smalley (1986) explains: "All sounds possess this dual potential – the abstract and concrete aspects of sound – and all musical structures are balanced somewhere between the two, although exactly how they are found to be balanced can vary greatly among listeners." There is also the difference between

hearing and listening to consider, which equates to receiving information passively, or by focussing attention on either the data-carrying (referential) or the form-carrying (abstract) characteristics of sound. For this category we will propose the term “casual listening” (not to be confused with Chion’s causal listen). This aural “zoom-lens” is in constant operation between the modes of casual listening, referential listening and abstract listening throughout the experience of listening to EAM. For example, in the scenario where listeners are experiencing a work which contains both speech and abstract sounds (played simultaneously), in the presence of speech (the most referential of all sounds), listeners will naturally focus their attention on the data-carrying characteristics of the spoken language, while “hearing” the abstract sounds in the background (casual listening). However, in the absence of spoken materials, listeners will naturally refocus their attention on the abstract qualities of the background sounds illuminating various qualities and bringing them to the psychological foreground. In this way, highly referential sounds, including the sound of speech tend to “hide” certain qualities in accompanying sounds in much the same way as moving images. In the case of a singing voice, it is most likely that attention will be divided between referential and abstract characteristics.

To summarise at this point, we have theorised that referential and abstract qualities of sounds form the basis of the modes of listening as proposed by Schaeffer and Chion, and have suggested that attention may be focussed on one mode or the other (selective attention) or both modes simultaneously (divided attention). We have also proposed that the position from which attention begins its focus, is that of “hearing” (casual listening). This has led to the proposal of three modes of listening, as illustrated in figure 1.

Visual Modes

The visual equivalent of figure 1 presents itself fairly immediately. Hearing and listening become seeing and looking, while the categories of casual, abstract and referential remain. However, various differences should be noted. In particular, the “dual potential” of abstract and referential qualities in visual materials is different than that of audio materials – as Wishart (1986) explains: “Transformation in aural space therefore tends to have a ‘dream-like’ quality removed from the concrete, distanced and often humorous world of visual animation. At the same time, although transformations of abstract forms may be achieved in the visual sphere, these tend not to have the same affective qualities as time processes taking place in the domain of sound.” Notwithstanding these differences, the construct provides us with a classification system for considering the systematic pairing of audio and visual materials.

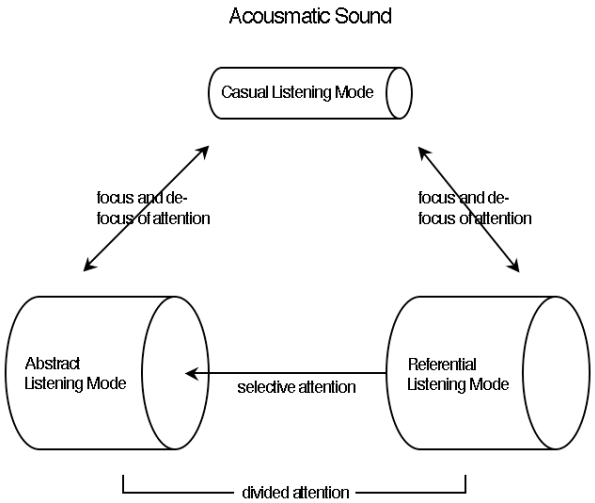

Figure 1: Modes of Listening

Media Pairing Strategies

Media pairing is the practice of selecting audio and visual materials for simultaneous playback – a task perhaps related to concepts of surface-level grammar as described in the introduction. The “dual potential” of sounds and images (Smalley 1986) allows the Composer to pair audio and visual materials based on their referential and abstract characteristics. Media may also be paired based on congruous characteristics (such as the simultaneous sight and sound of a news reader) or incongruous characteristics (such as happy music playing during a sad scene). This provides the Composer with 4 strategies of media pairing: abstract-abstract, abstract- referential, referential-abstract, and referential- referential, with each media pair (audio and visual) evoking either one object (single source) or two objects (dual sources). In this paper we have already discussed the use of referential images with abstract sounds *(Shifting Ground)* and abstract images with referential sounds *(Free Radicals).* I will now examine the remaining categories of abstract images with abstract sounds, and representational images with representational sounds.

Firstly, it appears that a large percentage of Composers working in the domain favour the use of abstract sounds with abstract images, and that audio and visual materials in seminal works often give the appearance of being parametrically related. Indeed, there are those who subscribe to an ideal where images are “driven” by sounds, and where the audiovisual texture evokes a single source (although multiple media pairs may be operating simultaneously). Here, there are many techniques of media pairing employed ranging from envelope following (e.g. amplitude with brightness) to audiovisual granular synthesis and other generative techniques. Perhaps, under these circumstances, where both sounds and images are abstract, and where they compliment one another in evoking a single coherent source, it is feasible that sounds are less affected by the audiovisual contract as proposed by Chion.

To investigate this proposal, I would like to return to the series of tests that were carried out as part of this study. Subjects were again asked to respond to audio only stimulus with eyes shut, then to respond to audiovisual stimulus with eyes open, and finally, to switch between the two modes using eyes shut and eyes open techniques. This time, the work under scrutiny was Gordon Munro’s *Dissonant Particles* – an audiovisual piece in which the sounds and images were generated algorithmically (with one algorithm). From the 5 subjects tested, 4 reported hearing more while attending audio only. Various comments also reported an awareness of an association between the audio and visual materials; however, the majority of subjects perceived the materials as a dual-source audiovisual texture.

Notwithstanding the small number of participants, these results suggest that *Dissonant Particles* (a work where the images are driven by the sounds) is affected by the audiovisual contract as proposed by Chion to almost the same degree as the other audiovisual styles tested. The results also suggest that parametric similarity does not automatically result in the evocation of a single coherent source. By way of explanation I would suggest that, for 4 out of the 5 subjects tested, the “meaning” of the abstract audio and visual materials in the extract may have been different, and that parametric similarity in “grammar” does not automatically result in similar “meanings”. To investigate this hypothesis further, I would once again turn to literature.

Chandler, Cook, Snyder et al present the possibility of an ideal audiovisual language that is founded on common "structural laws", as Cook (1998) wrote: 'We can think only of what is actually "commensurable" i.e. the movement lying at the base of both the structural law of a given piece of music and the structural law of the given pictorial representation' Chandler (2003) provides further detail in his description of “universal features” as Gestalt principles. He states that “According to the Gestalt psychologists there are certain universal features in human visual perception, which in semiotic terms can be seen as constituting a perceptual code”. He then provides a list of Gestalt principles as follows: “proximity, similarity, continuity, closure, smallness, symmetry and surroundness”. Snyder (2000) adds further detail to Chandler’s description in his appraisal of image schemas “Image Schemas are thought to be derived from commonalities in different experiences that seem related; such as, they are believed to form the basis for our conceptual systems, indeed to connect our perceptual experience and concepts.” Snyder then proposes that all experience can be related to combinations of the 5 image schemas: up and down, centrality, motion/linkage/causation, linearity: pathways and goals, and containment: inside and out.

This literature provides us with a general concept that relates to the deep-levels of language (meaning) described in our definition. It also proposes that audio, visual and audiovisual (Gestalt) stimuli are capable of evoking common image schemas in audiences. At first this appears to provide the Composer with a point of commonality or linkage that may be used in the process of media pairing; however, further investigation reveals that these principles are too general to be used as techniques – as Cook (1998) explains, “They mean different things to different people… they may stand for many or even any signifieds; they may mean whatever their interpreters want them to mean”.

To offer further explanation of the functionality of surface-level grammar and deep-level meaning (initially proposed by Chomsky in relation to linguistics), I would now turn to the work of Albert Bregman (1993) who states that “Issues surrounding attending to auditory events that emerge from current theories fall into two categories - those concerned with relatively fast events and lower-order motion properties, and those concerned with relatively slow events and

determinants of sustained attending." Bregman then goes on to explain that sustained attending makes use of "intentional schemes that arise from long-term experience". These short and long-order abstractions may also be related to the autocentric and allocentric perceptual modes as proposed by Schachtel. In the words of Smalley (1996) "The autocentric or subject-centred senses focus on basic responses and feelings of pleasure and displeasure [while] the allocentric perceptual mode is object-centred in that it involves perceiving something independent of the perceiver's needs."

To summarise at this point, we have already determined that abstract audiovisual works (including works where the images are driven by the sounds) are affected by the audiovisual contract as proposed by Chion in the same way as other audiovisual styles tested. We have also hypothesised that congruous media may be at the heart of evoking "perceptual clarity" and that congruousness may be founded on "meaning" as opposed to "grammar". We have so far failed to identify an instance where the communicability of the audio materials remains fully functional in the presence of moving images. To date 3 of the 4 media pairing strategies have been tested. I will now investigate the remaining category of representational sounds and images by returning to the series of listening tests that were carried out as part of this study.

Subjects were once again asked to respond to audio only stimulus with eyes shut, then to respond to audiovisual stimulus with eyes open, and finally, to switch between the two modes using eyes shut and eyes open techniques. Two extracts of representational footage were selected for testing. The first was of a busy city park containing multiple media pairs depicting single audiovisual sources (such as people talking), and dual sources (such as distant buildings with the close sound of water fountains). The second extract was of a person reading from a paper (a single coherent source).

Data from the first test (the city park) showed that from the 5 subjects tested, 4 reported hearing more while attending audio only, while results from the second test (the person reading) showed that from the 5 subjects tested, 2 participants heard more while attending to audio only. Comments from the remaining 3 participants suggested that either the audio materials were unaffected by the presence of moving images, or that they were "enhanced" by the moving images.

I would now turn to science to substantiate these results. Talsma, Doty and Woldorff (2006) state that "integrating visual and auditory stimuli serves the purpose of enhancing perceptual clarity… These results [also] suggest that communication between the visual and auditory brain areas is a highly effective and relatively automatic process". Talsma and Woldorff (2005) have established that "information stemming from multiple senses is not likely to be processed in isolation but will tend to be integrated into a multi-sensory percept under various circumstances". Two super-additive effects also appear to result from the process of attending to audiovisual stimuli simultaneously. The first is in relation to the level of stimulus, as described by Talsma "when audiovisual objects were attended, the P50 to the audiovisual stimuli was larger than the sum of the P50 activity for the auditory and visual stimuli." The second super-additive effect occurs only under certain conditions, and concerns integration time. As Woldorff explains "their results, just as ours suggest that the multisensory integration effect only occurs early in time when both visual and auditory stimulus features can be constructed into a single coherent audiovisual object." The tests that were carried out by Talsma et al made use of single source, or dual source representational audio and visual media only. For example, an image of the letter "T" combined with the sound "Tee".

In the interests of widening the discovery beyond that of the representational, I would now like to recall an unusual phenomenon that occurred during the testing process of the Len Lye work *Free Radicals.* From the 5 subjects tested all 5 experienced the phenomenon at precisely the same time in the work. The isolated moment was confined to a passage where abstract lines appeared to combine with the sound of a singing voice. Comments from 2/5 subjects described the sensation of being inside the experience as "sensational" and outside the experience as "ordinary". Perhaps these comments are related to the super-additive affects proposed by Talsma, Doty and Woldorff,

To review findings at this point, the combination of listening tests carried out as part of this study, together with Talsma's research provides fairly compelling evidence that in settings where referential audio and visual materials compliment one another in evoking single coherent sources, the communicability of the audio materials is less likely to be

compromised by the presence of moving images. This directly answers the main research question, “whether adding moving images to EAM subtracts from its communicability in all circumstances”. The focus of the study now shifts to investigate the process of transition between audio-only and audiovisual modes.

Modal Transitions

The initial research question is “whether there is any transitory state or continuum between the experience of attending audiovisual, and the experience of attending audio-only, or whether the transition is one of immediate change?” Chion (1994) provides a direct answer to this question. He attests to the absence of a continuum between the two general modes, noting that a collapse of "the audiovisual structure" is necessary before a transition to the general audio-only mode (the acousmatic mode) can occur. Tests carried out on postgraduate students, as discussed earlier in the paper also support Chion’s views. In the first series of tests, Subjects were asked to respond to audio only with eyes shut, then to respond to audiovisual with eyes open, and finally, to switch between the two modes using eyes shut and eyes open techniques. Participants recorded that short periods of time (2-5 seconds) were required to adjust to the conditions of the new mode, and reported more difficulty in the process of transition between acousmatic and audiovisual modes than between audiovisual and acousmatic modes. When questioned further, some expressed discomfort in “giving up” the “images” they had formed in the acousmatic mode, but had less discomfort in giving up the images provided by the audiovisual materials.

A new test was then devised to determine whether a “black screen” could evoke the acousmatic mode. The experience of entering the acousmatic mode is easily measured, as it is characterised by de-focussing (physically). Subjects were played a section of *Shifting Ground* that fades to black, while the audio continues. They were then asked to record details concerning any transition of mode they experienced. From the 5 subjects tested, all 5 reported sensations of entering the acousmatic mode soon after they realised that the image was not returning. However, those who remained focused on the projection screen reported a state “close” to that of the acousmatic mode. It is the issue of “closeness” that signals the possibility of a transitory state.

This then raised the question whether the use of minimal images could allow a transitory state between audiovisual and acousmatic modes. From observation this technique is in prolific use in the repertoire of EAM with moving images; however, its effectiveness remains speculative. To answer the question, yet another test was carried out using a section of *Shifting Ground* that faded to white, as opposed to black, while the audio continued. Subjects were again asked to record details concerning any transition of mode they experienced. From the 5 subjects tested, 4/5 reported weak sensations of reduced listening; however, the majority of participants took much longer to enter this mode (10 seconds or more), and subjects noted that the sounds seemed to take on “white” characteristics.

The issue of transition between modes may be further described by the concept of “bottlenecking”, as Riess Jones and Yee (1993) explain “Many approaches to attention build upon an information processing metaphor. An individual is viewed as a communication system through which information flows. Attention is conventionally conceived either in terms of a bottleneck which occurs at certain stages in the flow, or in terms of certain processing limits” This view may help in explaining why, in the process of adding moving images to EAM, we are forced to change modes from acousmatic to audiovisual, and why only minimal images, such as colours, can be tolerated by the acousmatic mode. It is here that I would like to introduce the concept of the “slow switch”. This proposition denies the existence of a continuum between the acousmatic and audiovisual modes, but allows some flexibility of change in terms of time, minimal images (such as colours) and memory residues.

Summary and Conclusions

Results have shown that the experience of attending audio and audiovisual materials may be classified into two general modes: the audio-only mode (the acousmatic mode), and a general audiovisual mode. Results also suggest that in the process of adding moving images to EAM listeners are forced to change modes from audio-only to audiovisual irrespective of audiovisual style, and that the audiovisual mode changes the quality of sounds, and hides certain aspects of sounds irrespective of audio style.

It has been proposed that there are further divisions within the acousmatic mode that are based on the experience of attending either

referential or abstract characteristics of sounds (selective attention) or both characteristics simultaneously (divided attention), and that the position from which attention begins its focus, is that of “hearing” or casual listening. This has led to the proposition of three modes of listening: casual listening, abstract listening and referential listening, as represented in figure 1.

Three visual-only modes have also been proposed based on the same conceptual framework. The modes are: casual viewing (de-focussed attention), abstract viewing (attending form), and referential viewing (attending data) This construct provides a classification system for considering strategies of media pairing based on the abstract and referential characteristics of the media. The four combinations of audio and visual materials are: abstract-abstract, abstract- referential, referential-abstract, and referential- referential. Each media pair (audio and visual) is also able to evoke a single source (congruous media) or dual sources (incongruous media). Please see figure 2 for a graphical illustration of the classification system.

The system also functions to clarify various aspects concerning the nature of each audiovisual category. For example, *Shifting Ground* – a work featuring referential video, and abstract audio would naturally find its place towards the top left, and half way between the single source and dual source planes. This position indicates the halfway point between the composer’s intention and the audience’s reaction to the audiovisual materials. If the intention was to evoke dual sources, then the composer did not provide enough information to signal the need to divide attention between audio and visual materials. The resulting outcome was that, in attempting to integrate the materials into a single source, the audience inadvertently filtered out large amounts of referential data in the audio track. I would speculate that this position in the cube is the most affected by the condition of synchresis proposed by Chion.

In contrast, the video clip of a *man reading from a paper* – featuring referential audio and video materials would be positioned towards the top right of the cube, and right on the single source plane. In this instance it is the data-carrying information in both the audio and visual materials (congruous semantic data) that is being attended, and, according to our results, that is least affected by the condition of synchresis proposed by Chion.

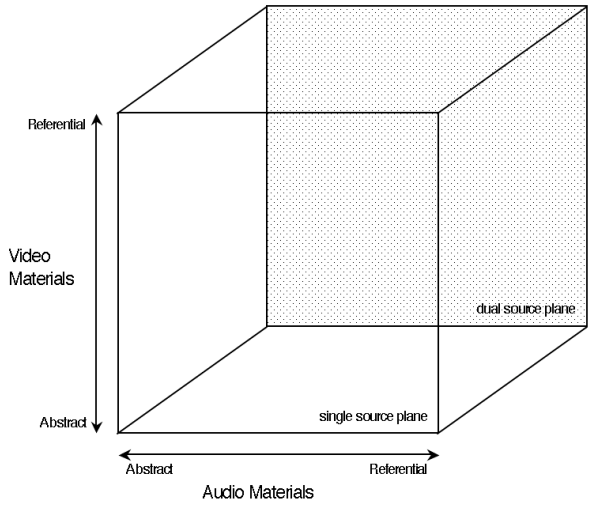

Figure 2: Classification of Media Pairs

The combined results suggest that the scientific claim, that “integrating visual and auditory stimuli serves the purpose of enhancing perceptual clarity” (Talsma, Doty and Woldorff 2006) is relevant only in the instance of attending referential audiovisual data, and that super-additive effects, including greater levels of stimuli and early integration of audio and visual materials are possible only “when both visual and auditory stimulus features can be constructed into a single coherent audiovisual object.” (Talsma et al 2006). In the majority of other instances where different combinations of audio and visual materials are dominant, I would propose that the scientific claim is inaccurate.

By way of explanation of the phenomenon observed in *Free Radicals,* where subjects experienced sensations similar to the super-additive effects described above, I would propose that the abstract characteristics of the sounds and images were autocentrically and allocentrically aligned to an extent where a momentary congruous state exhibiting resonances of a single source representational audiovisual system was evoked. At the basis of this proposal is the assumption that fast integration of single-source representational audio and visual materials was an important faculty in the mind of primitive man, and that the appreciation of congruous audio and visual data is governed by “primitive scene analysis” (Bregman 1993)

In terms of transition between modes, the results suggest that a collapse of "the audiovisual structure" is necessary before a transition to the acousmatic mode can occur, and that entering the acousmatic mode is characterised by the physical function of de-

focussing. Results also show that the projection of a “black screen” can encourage listeners to enter the acousmatic mode, and that minimal images, such as colours can, under limited conditions, provide a transitory state between modes. A major finding of the study is that the transition between the audio-only (acousmatic) and audiovisual modes is not a continuum – it is best described as a “slow switch” that accommodates small amounts of data such as changes in time, minimal images (such as colours) and memory residues.

To conclude, this study has illuminated the limitations of various audiovisual styles, and identified the creative potential of others. As a composer, it has provided me with the framework to consider a free-to-constrained continuum between abstract sound, referential sound, speech (semantic sound) and audiovisual materials. Interestingly this continuum appears to be a simple extrapolation of the abstract-referential continuum proposed by Smalley (1986). In terms of usability, both acousmatic and audiovisual materials possess the capacity to evoke powerful images, yet it is often the relationships between the images that is more significant – as Freud (1947) explains “We learn that what becomes conscious is as a rule only the concrete subject-matter of thought, and that the relations between the various elements of this subject matter, which is what specially characterises thought, cannot be given visual expression”

References

[Bayle 1997] Ed. Sandra Desantos *Acousmatic Morphology: an interview with Francois Bayle,* Computer Music Journal, Massachusetts Institute of Technology 1997

[Bregman 1993] Albert Bregman *Auditory Scene analysis: hearing in complex environments* in Thinking in Sound: The Cognitive Psychology of Human Audition (S. McAdams (Ed.)), Oxford University Press Oxford, 1993

[Chandler 2003] Daniel Chandler. *Semiotics: The Basics.* Routledge: University of Chicago Press, 2003.

[Chion 1994] Michel Chion. *Audio-Vision: Sound On Screen* (Claudia Gorbman Ed.) Columbia University Press, New York 1994

[Cook 1998] Nicholas Cook. *Analysing Musical Multimedia.* Oxford University Press Inc., New York 1998

[Emmerson 1986] Simon Emmerson. The Relation of Language to Materials. In The Language of Electroacoustic Music (S. Emmerson (Ed.)) Basingstoke, UK: The Macmillan Press, 1986

[Emmerson & Smalley 2001] Simon Emmerson and Denis Smalley. *Electroacoustic Music Definition* in *The New Grove Dictionary of Music and Musicians* – Second Edition. Ed. Stanley Sadie (2001).

[Freud 1947] Sigm. Freud *The Ego and the Id* (4th Ed.) The Hogarth Press, London 1947

[Jung 1968] C. G. Jung, *The Archetypes and the Collective Unconscious (Second Ed.)* Bollingen Series X X Princeton University Press, Princeton, USA 1968

[Landy 2000] Leigh Landy. *Co-hear-ence and Electroacoustic Music*. VII Brazilian Symposium on Computer Music Keynote Address Curitaba – July 2000.

[Piaget 1970] Jean Piaget. *Genetic Epistemology,* (Eleanor Duckworth Trans.) Columbia University Press 1970

[Riess Jones and Yee 1993] Mari Riess Jones and William Yee *Attending to auditory events: the role of temporal organisation* in Thinking in Sound: The Cognitive Psychology of Human Audition (S. McAdams (Ed.)), Oxford University Press Oxford, 1993

[Smalley 1986] Denis Smalley. *Spectro-morphology and Structuring Processes.* In *The Language of Electroacoustic Music* (S. Emmerson (Ed.)) Basingstoke, UK: The Macmillan Press, 1986

[Smalley 1996] Denis Smalley *The Listening Imagination: Listening in the Electroacoustic Era* Contemporary Music Review Vol 13 Part 2 1996

[Snyder 2000] Bob Snyder. *Music and Memory – An Introduction.* London: The MIT Press, Cambridge, Massachusetts, 2000.

[Talsma, Doty & Woldorff 2006] Durk Talsma, Tracey J. Doty, and Marty G. Woldorff *Selective Attention and Audiovisual Integration: Is attending to Both Modalities a Prerequisite for Early Integration* Centre for Cognitive Neuroscience, Duke University, Drham, NC, USA 2006

[Winter 1989] Richard Winter *Learning From Experience: Principles and Practice in Action-Research*. Philadelphia: The Falmer Press, 1989

[Wishart 1986] Trevor Wishart. Sound Symbols and Landscapes. In The Language of Electroacoustic Music (S. Emmerson (Ed.)) Basingstoke, UK: The Macmillan Press, 1986