Innovations in Education & Training International

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/riie19

Shifting Attention in Multimedia:
Stochastic Roles, Design Principles and the SSF Model

Bruce L Mann

a Memorial University of Newfoundland, Canada

Published online: 09 Jul 2006.

To cite this article: Bruce L Mann (1997) Shifting Attention in Multimedia: Stochastic Roles, Design Principles and the SSF Model, Innovations in Education & Training International, 34:3, 174-187, DOI: 10.1080/1355800970340303

To link to this article: http://dx.doi.org/10.1080/1355800970340303

Please scroll down for article

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Shifting Attention in Multimedia: Stochastic Roles, Design Principles and the SSF Model

Bruce L Mann, Memorial University of Newfoundland, Canada

SUMMARY

This paper examines the research on design methods for shifting attention between visuals (for example, text, graphics or moving images) and auditory prompts (such as speech, music or effects) in multimedia environments. Stochastic design methods are most prevalent in the literature. Stochastic approaches reflect outdated bottleneck theories of human attention. Results using stochastic design methods have been mostly poor. Studies using the Structured Sound Function (SSF) model, however have shown good immediate results in student retention, and even better results following a latency period. The SSF model relies on the ‘explicitness’ and ‘gist’ requirements inherent in tasks and the interrelatedness of spatial and language representations. It is recommended that designs that incorporate gist and explicit requirements using the SSF model be applied developmentally, because as we get older we get better at extracting gist from sound than details from text.

INTRODUCTION

Consider the following scenario. It’s a sunny day. You are walking along a crowded sidewalk in a major city centre at rush-hour. People hurry past you. Others stand window-shopping. There’s a queue outside a theatre beside you. A hot-dog vendor is there taking orders. Nearby the traffic lights change colour and walking lights begin to flash. Just then you are startled by the honking of dozens of cars and trucks in a traffic jam. A traffic constable blows a whistle and waves the traffic through the intersection.

In this scene, and others like it in everyday life, we know that we become less efficient when we are forced to divide our attention between competing sources; we experience a kind of sensory overload. This overload is what environmental psychologists call ‘cognitive chaos’ (Kaplan and Kaplan, 1982). Cognitive chaos has been associated with interactive technology as well (Astleitner and Leutner, 1995). Knowledge acquisition in hypermedia systems is a problem when learners are required to do several tasks at one learning trial that require the storage, transfer or evaluation of discrete pieces of information (Jonassen and Grabinger, 1990; Schroeder and Grabowski, 1995). The effect constitutes a cognitive overload which requires a high memory demand (Stark, 1990). In brief, we become less efficient as we divide our attention between multiple messages and sources of important information. What can we do to retain our efficiency and reduce cognitive chaos?

HUMAN ATTENTION

Early theories of attention, such as that of Broadbent (1958) emphasized that people are extremely limited in the amount of information they can process at any given time. A common metaphor in these early theories of attention was the bottleneck concept which proposed that there is a narrow passageway in our working memory that limits the quantity of information to which we attend at any one time (Matlin, 1989; Gagné et al., 1993). Neisser (1976) disagreed with this overall limit to our capacity for information proposed by the bottleneck theories. Whereas we can become less efficient as we try to focus on several things at once, our efficiency can be increased with practice (Neisser, 1976) and environment design (Fleming, 1987; Gagné et al., 1988).
Currently accepted theories of attention in the cognitive disciplines (Gagné et al., 1993) state that there are different kinds of attention relative to the difficulty of the assigned tasks and the familiarity of the content under study. On easy divided-attention tasks with familiar items, we use automatic processing (Schneider and Shiffrin, 1977) also known as pre-attentive processing (Treisman, 1986). Preattentive processing of divided-attention tasks can occur in parallel; that is we can handle two or more items at the same time. We implement preattentive processing on divided-attention tasks for easy tasks or with highly familiar items.

On difficult divided-attention tasks with unfamiliar items we use controlled processing (Schneider and Shiffrin, 1977); also known as attention focusing (Treisman, 1986). Attention focusing on divided-attention tasks is serial; only one task is handled at a time. We tend to use attention focusing on difficult or unfamiliar divided-attention tasks. This kind of attention has been described as ‘as the glue that binds the separate features of a stimulus—such as the colour and shape—into a unitary object’ (Matlin, 1989, p. 57). When we pay attention to a visual stimulus, cerebral blood-flow studies shows increased activity in the visual and parietal cortex (located in the top, back region of the brain) (Robinson and Peterson, 1986). When we shift our attention from the visual to the auditory modality our blood flow is increased in our prefrontal cortex (located at the top, front region of the brain) (Matlin, 1989). In this way we gain and hold our own attention, whether it is a response to an emergency message or the beginning stages of a lesson. When our attention is overloaded or distracted, features can be combined inappropriately.

ATTENTION IN EDUCATION

Attention is critical to learning; without it there can be no learning. Designers typically seek to obtain and maintain our attention (Fleming, 1987). Attention sustains our vigilance with sensations, regulating the search and enhancing our sensory acuity that makes us aware of any new cues. Good thinkers can readily apply attentional focus (Pressley and McCormick, 1995), even in the face of distraction (Aks and Cohen, 1990). Kuhl (1985) refers to this ability to insulate goal-directed cognition from distraction as ‘action control’. Good thinkers can inhibit inappropriate responses ( Bjorklund and Harnishfeger, 1990; Dempster, 1992) and are appropriately reflective-thinking through problems rather than reacting impulsively. Even so, they are not reflective to the point that reflection is paralysing (Duemler and Mayer, 1988) and always considering rather than acting (Baron, 1990).

Nevertheless, even good thinkers can not apply attentional focus in every situation. Stress, lack of sleep, personal problems or multiple stimuli can reduce attentional focus. In these situations, software developers can help us to maintain our attention by influencing the display from the educational technology perspective (Fleming, 1987). Whenever we decide to read an assigned history chapter, watch a science demonstration or listen to an explanation of a numerical computation, we self-initiate a system of information processing (Borich and Tombari, 1995). As we attend to these materials, we are holding the unprocessed sights and sounds momentarily before they fade away. These sensations linger long enough for us to associate meaning to them.

Educational technology

Early forms of instructional technology (such as, teaching machines, tape/slide presentations, and so on) usually presented unfamiliar visual and auditory stimuli to learners (Blyth, 1960). Users were required to control their own processing of these separate sensations. At that time, little was known of the effects of auditory manipulations on language learning (Porter 1960). ‘The preoccupation with visual information is a deliberate effort of the communicators on the assumption that the sound will take care of itself. But experimentation shows that sound does not take care of itself’ (Hartman, 1961, p. 256). Consequently, early approaches to learning from auditory-visual delivery technology was found to be ad hoc, improved and required frequent re-examination and clarification (Conway, 1968). Human interactions with technology were often explained as a function of playback system performance instead of how learners divided their attention on difficult and unfamiliar tasks (see Skinner, 1960).

Today, some thirty computer applications permit users to annotate their silent applications and curricular software with as many as 1,024 commands per application. ‘Sound is a great attention-getter, one that can lure us to dig deep into a newly discovered Website’ (Ozer, 1996). In keeping with this trend, developers and users are dividing their important messages between the visual and auditory channels.
of their favourite software application, computer-assisted instruction (CAI) prototype or web service (Gaver, 1989; Brown, 1992; Aamtzen, 1993; Barron and Kysylka, 1993; Hartas and Moseley, 1993; Muraida and Spector, 1993).

'Until now, bandwidth limitations and the large size of audio files has made it impossible to use the World Wide Web to access audio content efficiently and reliably. Newer technologies make more efficient use of the bandwidth available, which in turn makes it possible for you to enjoy the sounds of the web' (Gonzalez, 1996).

Despite this resurgence of interest in audio design, guidelines for its implementation are still explained as a function of playback system performance instead of how learners divide their attention on difficult or unfamiliar tasks. These other recent explanations for adding sound to formatted text, graphics and moving images can be classified into one of two types: stochastic sound roles or the Structured Sound Function (SSF) model. The stochastic sound roles are discussed first.

STOCHASTIC SOUND ROLES

In general, stochastic sound roles usually distract users' attention away from difficult or unfamiliar tasks. The following discussion about stochastic roles for sound is intended to highlight several issues that are frequently obscured by current practices in educational technology research and development.

Stochastic cueing

In stochastic cueing, the user's attention momentarily shifts to announcements, designed to ‘provide forewarning rather than concurrent information’ (Gaver, 1989). Intuitively we believe that cueing with music, effects or words must be an effective way of using sound in educational technology (Jaspers, 1991). However, stochastic sound cueing in educational software is problematic. One problem with it is the inability to ‘activate or deploy attention [as an] expanded instructional event’ (Smith and Ragan, 1993, p 141). In other words, the stochastic sound cues require the user to switch the focus of their attention to the punctuated announcements which in themselves do nothing to actively employ the user. An oral advance organiser, for example, is a form of cueing with sound in educational settings to gain attention. Oral advance organisers, however, have shown disappointing results in educational research (Arwady, 1980; Dougall, 1988; Wolter, 1988). The reason is that oral advance organisers in themselves do nothing to engage the learner. From a user’s viewpoint, this means there will be more sensation to process. From the designers’ viewpoint, it means imprecision and a serious lack of prescriptive information.

A related shortcoming in the design of some software occurs when a user’s attention is divided between mixed messages. This is another example of a stochastic sound cue that requires us to switch the focus of our attention back and forth between conflicting unintended messages. Speech synthesis programs like ‘Speak and Spell’ for example, which offer children practice in sound and letter matching, pronounces words as the child types them. In Daiute’s study (Daiute, 1985) bleeps from the program that were intended to cue the student denoted that either the child’s input was incorrect or that a particular word was not part of the program’s limited vocabulary. Double meanings from single sound cues resulted in several incorrect interpretations by the children.

Surveys of stochastic sound cueing that divides our attention have reported inconclusive findings. Weiner (1989) investigated the differential effects of presentation conditions; visual only and visuals cued with speech, on sight-word learning with 55 handicapped third graders. The results indicated non-significant statistical differences between the presentation conditions. It seems possible that the implementation of a repeated measure by treatment design contaminated the visual-only treatment with memory traces from the speech treatment. Repeated measures in speech research should only be implemented by occasion, and then only after time lapses between occasions.

Two years later, however, Weiner (1991) reported different results with 24 junior-high students using a Listen and Learn Speechware program. The purpose of the study was to investigate the effects of digitized speech components on mildly retarded adolescents’ accuracy to target sight words and to orally read the targeted words; an indefinite role. Unlike Weiner’s (1989) repeated measures study, the visual only group received directions and word presentations without access to the auditory high-quality female spoken cues presented to the visual/auditory group. The results showed that computer-based sound cueing increased mildly retarded adolescents’ accuracy with target sight words and their ability to orally read target sight-words. It seems, then, that the implementation of a repeated measure by treatment design in Weiner’s
(1989) study contaminated the visual-only treatment with memory traces from the audio treatment. It appears that the cueing role for sound remains unproved in terms of its effectiveness in educational software at this time. Cueing without a function for the sound, therefore, should continue to be referred to as simply a stochastic role for sound.

Message redundancy

Message redundancy is the second stochastic role for sound described in the literature (Brown, 1992; Jaspers, 1991). In message redundancy, our attention is divided by sound that restates messages already presented in text or graphics without providing any new information (Shannon and Weaver, 1949; Weiner, 1950). This appears to be an opportunity lost; identical (instead of complementary) information presented from different sources. Usually, when a message is made to be redundant and two communication channels are prescribed, audio takes the redundant role. Redundant sound has been widely-implemented in education (Joyce et al., 1992). The rationale for its use in education is to protect information from errors by increasing the redundancy of the messages (Clemson, 1984; Jaspers, 1991).

Another educational implementation of message redundancy is called audio accompaniment. Computerized audio accompaniment is usually considered to be an asset (Malone, 1981). Simple audio accompaniment is thought to decorate, enhance, reward and represent that which would have otherwise been less effectively communicated as words or numbers. In practice and research, however, redundant sound simply splits the user's attention unnecessarily between two sources that provide identical information. When humans are included as message receivers in educational systems (Ragsdale, 1988; Jaspers, 1991; deHaemer and Wallace, 1992), or in emergency message systems (Buxton, 1987; Huff and Finholt, 1994), an increase in information redundancy has not always increased the control over their behaviour. Cottom (1991) attributed poor results to the novelty of digitized speech. Issak (1988) found that the addition of audio accompaniment resulted in nonsignificant statistical differences between the groups in either reading or listening skills.

Message redundancy tends to portray learners as automatic controllers with the goal of matching their outputs to performance objectives (Blyth, 1960; Porter, 1960; Skinner, 1960; Jaspers, 1991). Comparisons between mechanical and human learning systems have drawn criticism of the message-receiver model (Baggaley, 1980, Hannafin, 1989; Winn, 1991).

‘The learner is no mere receiver of transmitted information as implied in the engineering model of human communication but an active processor of it who interprets the world in characteristic ways determined by his background and personality’ (Baggaley, 1980, p 52).

Moreover, educational software is often criticized for ‘a widespread lack of creativity and innovativeness’ (Hannafin and Rieber, 1990, p 34) in a time when innovation is required to ensure that the computer-based information is understood (Ragsdale, 1988). This criticism includes the tendency to explore better control principles rather than cognitive states, and a predisposition for representing learning in hierarchies of discrete chunks of information, rather than creative or interpretative enterprises (Baggaley, 1980). The physical evidence of interaction with software is considered to be less important than the cognitive activities that the lesson is designed to engender (Jonassen, 1988; Winn, 1991). A richer understanding of the technology's attributes (eg, colour, sound, etc) is also required (Mann, 1992).

‘It is no longer adequate to simply describe interactions in terms of either the input technology employed or the physical characteristics of the responses made. These will certainly change over time. We need a richer understanding of the psychological requirements associated with instructional tasks and responses and a sense of how to extend design science beyond the methods that have evolved through the years’ (Hannafin, 1989, p 178).

An assumption of redundant sound in educational software is that the information channels are attribute-free; that the software is merely a vehicle for conveying the information (Clark, 1987). The opposite position is taken in this paper; that the purpose and place of sound can affect the information conveyed (Grimes, 1990; Mann, 1995a). Computer sound should convey the kind of information about events not visually attended to, and can help to reduce visual clutter (Gaver, 1989; Brown, 1992), making for a synergistic relationship with the visual information. When the auditory-visual relationships are made to be synergistic, the effects can offer learners a better alternative to redundancy. Cooperative action, in this sense, presumes active participation by the listener/viewer. ‘The most interesting questions have no agreed-upon answers. The software should encourage students to explore a variety of possible answers’
Oral reporting

Oral reporting is the third stochastic role for sound described in the literature. Oral reporting divides our attention consistently over the entire duration of interaction with the software. Oral reporting is intended to provide us with continuous synthesized speech with a specific purpose. Brown (1992) used oral reporting 'to replace what could easily be displayed in a visual view to allow users to focus their attention on other visual views' (p 664). Warning systems (Huff and Finholt, 1994), multidimensional audio windows management (Cohen, 1991; Cohen and Ludwig, 1991), environmental sound postcards (Gaggiolo et al., 1990) are other examples of the reporting role for sound. Insofar as the reporting role seems to be an improvement over redundancy, it still requires the users' selective attention and correct subsequent action, sometimes as overlapping tasks (deHaemer and Wallace, 1992).

Overlapping, for example, has become a job skill requirement for all telephone operators and some taxicab drivers. Overlapping for telephone operators involves the instantaneous and simultaneous discrimination, classification and response: (1) from a visual display; (2) continuous computer-based verbal cues or requests; and (3) the constant ‘live’ verbal inquiries from their peers, trainees, visitors and the supervisor—a Point Of View Sound Function (Mann, 1995a). Overlapping, for some cab drivers, requires the additional activity of attending to continuous synthesized speech that reports the whereabouts of other cabs and new fares within designated zones — a Locale Sound Function. Overlapping seems to encourage similar multi-tasking capabilities for users.

However, the simultaneous requirement for human attention to presentation modalities and task requirements has contributed to confusion in some important computer applications. Perhaps the most serious confusion occurred at the Three Mile Island plant where several different warning systems were activated simultaneously (Buxton, 1987; Huff and Finholt, 1994). Something more decisive than an oral reporting role is required to help us to pay attention to several tasks by distributing our attention between more than one competing message or source.

Auditory feedback

Auditory feedback is the fourth stochastic role for sound described in the literature (Blyth, 1960; Porter, 1960; Skinner, 1960; Calvert et al., 1989; Thomas and Clapp, 1989; Wetzel, 1991). Auditory feedback is intended to slow down our interaction with the software by requiring us to shift our attention between new and old program design opportunities. As a consequence the novelty and challenge offered by the new program design opportunities simply replace the controlled processing required to focus on the old. ‘Although aesthetically pleasing, feedback provided in text can go unnoticed by the student’ (Alessi and Trollip, 1991, p 72). A corollary to this concern is that when feedback is provided following errors, it is often more interesting than feedback following correct responses (Ragsdale, 1988; Alessi and Trollip, 1991). Generally, auditory feedback schemes using computerized speech have reported mixed results. The positive results will be discussed first, beginning with Autoskill.

Autoskill (Fiedorowicz and Trites, 1985) is a computer-assisted remedial reading program. During oral reading training with Autoskill, the student responds orally to stimuli on the screen. There are opportunities for teacher remediation. During auditory-visual matching training, students respond with key presses. Student errors are presented and auditory stimulus is repeated without teacher remediation. Finally, during visual matching, students respond with key presses. Student errors are presented without teacher remediation.

Results of studies (Fiedorowicz and Trites, 1985; Thomas and Clapp, 1989; Jordan, 1990) showed improvement in reading in language lab settings for special education learners and for students of English as a second language. Jordan (1990) has also reported subsequent improvements using Autoskill with a wide range of target audiences. Sixty students aged 20 to 35 years of age taking basic skills training at Humber Community College received an average of 27 hours of training with the Autoskill reading program. They gained an average reading grade of 4.7 in word-attack skills. Forty incarcerated males in Dade County, Florida, were placed on the Autoskill remedial reading program for 30 hours over a 90 day period. They achieved an average of 1.8 grade improvement on tests of reading comprehension. Young adult Hispanics in an ESL program in California gained on average 2.7 grades with 17.5 hours using Autoskill reading.
program (Jordan, 1990). Young, single mothers, who were high school drop-outs, enrolled in a United States Federal Job Training gained an average of 3.6 grades in 9 hours of training with the Autoskill remedial reading program (Jordan, 1990).

In another reading program using auditory feedback, Ross (1990) addressed the problem of the poor readers’ inability to decode print adequately. A sample of 32 students from an inner city school were selected. Ross found that speech feedback was associated with improvement in reading of individual words. Subjects ‘with speech feedback’ made statistically significant greater gains in reading voiced words than those ‘without voice feedback’. Moreover, subjects ‘with voice feedback’ made significantly fewer errors than those ‘without voice feedback’.

Similar results were attained elsewhere. Anderson-Inman (1991) gave 58 third- to sixth-grade reading disabled students whole-word, syllable, subsyllable and mixed auditory feedback for unknown or difficult words. On the word recognition tests, students using the auditory feedback system significantly out-performed the control group who did not use the system.

In a different study, however, Zender’s (1990) dissertation research with 41 business communications college students reported statistically non-significant learning effects from varied amounts of verbal reminders in a computer-assisted instruction (CAI) self-monitoring program. Zender attributed these findings, however, to differential prior knowledge between the groups, uncontrolled time on task, lack of a definition for ‘minimum practice’, improperly operationalized reminders, not implementing covariate measures, and students that were not taught the appropriate self-monitoring skills.

Shilling’s (1991) research with 81 kindergarteners reported statistically non-significant learning effects between students using stochastic writing materials and/or computers with and without available synthesized speech feedback over an eight-month period. It seems possible that the implementation of a repeated measure by treatment design, like the Weiner (1989) study, contaminated the visual-only treatment with memory traces from the speech treatment.

It appears that auditory feedback has reported improvements on learning low-level matching tasks, such as auditory-visual, subsyllable, syllable and word matching. However, in studies in which learning was equated to other delivery methods, problems were reported in the design of the research.

In conclusion, it is clear from the above account that stochastic methods do not always focus attention on the competing tasks required in most contemporary interactive programs and online services. Stochastic sound roles that use digitized speech clips with formatted text, graphics and moving images relegate information delivered by the audio channel to novelty or accompaniment status, in which it is inappropriate. Researchers and developers cannot always find reliable information on how to focus attention on the competing tasks required in most contemporary interactive programs and online services. Even when information is found, much time is spent reading irrelevant material. ‘There is a lack of guidelines for the proper use of audio in computer based instruction... Designers must decide if audio should replace, enhance, or mirror the textual information on a CBI screen’ (Barron and Kysilka, 1993, p 277).

Sometimes the guidelines are too general or at least not generic to our needs. Usually we are forced to determine the trade-offs among conflicting guidelines. Guidelines are not sufficiently helpful in supporting decisions about new technology; guidelines lag behind the times. There is no predetermined syntax or structure to limit how skills should be represented.

‘By effective design we can reduce the noise component and increase the information-providing potential of sound. Our ambition is to promote the acquisition of an understanding that will support the design of audio cues that will improve human performance in computer-mediated tasks: designer sound for computer systems’ (Buxton, 1989, p. 1).

A more systemic model is needed for designing sound to improve learning with sound: ‘designer sound for computer systems’ (Buxton 1989, p. 1). One model, the Structured Sound Function model was developed to fill this need; to prescribe principles that compliments formatted text, graphics and moving images and assist users in shifting their attention between the auditory and visual stimuli.

THE SSF MODEL

The Structured Sound Function (SSF) model is a parsimonious approach to applying purposeful sound to a videoclip, graphic or paragraph of formatted text. The model has only two attributes: sound structure and its functions. Unlike the stochastic roles for sound,
the SSF model is derived from design principles, discussed later in the paper. The origins and applications of the SSF model are explained in Mann (1992, 1995a, 1995b, 1996).

**Structures for sound**

Structuring sound means describing or prescribing from among six levels of informational intervention with the image; from either of two roles for an emotional strategy; from a flexible pacing strategy, a continuous or discontinuous rhythm strategy; a spaced, massed, or summarized review strategy; and a convergent or divergent delivery strategy. Figure 1 shows the strategies that structure the sound with the visual presentation.

<table>
<thead>
<tr>
<th>Intent</th>
<th>Visuals</th>
<th>Review</th>
<th>Pacing</th>
<th>Rhythm</th>
<th>Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inform</td>
<td>Visuals</td>
<td>Review</td>
<td>Pacing</td>
<td>Rhythm</td>
<td>Delivery</td>
</tr>
<tr>
<td>Inform</td>
<td>Visuals</td>
<td>Review</td>
<td>Pacing</td>
<td>Rhythm</td>
<td>Delivery</td>
</tr>
<tr>
<td>Motivate</td>
<td>Visuals</td>
<td>Review</td>
<td>Pacing</td>
<td>Rhythm</td>
<td>Delivery</td>
</tr>
<tr>
<td>Motivate</td>
<td>Visuals</td>
<td>Review</td>
<td>Pacing</td>
<td>Rhythm</td>
<td>Delivery</td>
</tr>
</tbody>
</table>

**Figure 1 Structural attributes of the SSF model**

When the primary intent of the audio is to inform, one of the first four audio-visual relationships (that is, cue, counterpoint, dominate or undermine the intended meaning) should be considered. These are substrategies that comprise the Informational Sound Strategies. Cueing, Counterpointing, Dominating, and Undermining can be placed along a relationship-to-image continuum. All four substrategies may be used throughout the sound design in combination with other substrategies. A fifth substrategy is not considered here and not included in the SSF model. It prescribes sound information that supports or merely accompanies the image, making that information unnecessarily redundant.

When the primary intent of the audio is to express emotion or to motivate, one of the last two audio-visual relationships listed in the second column (ie, define visual action, punctuate emotional highlights) should be considered. A choice of Reviewing and Delivery attributes are available for most static or moving visuals. However, Pacing and Rhythm attributes are not usually considered when the visuals are static (ie, formatted text or graphics). Once the structural attributes of the audio have been assigned and the type of audio decided, then a purpose or function can be considered. Decisions about whether to use speech, music or effects should also be addressed at this time.

Pacing has been used effectively in applications of sound to image in artistic, educational and entertainment audio visual environments. Appropriate dialogue cutting points, for example, can improve the smoothness of edits when cutting from one speaker to another in a hypermedium. The two Pacing sound substrategies prescribe how fast, where, and how often the chosen sound function occurs in the program. Sound pacing can be placed along the continuum; fast or slow, occurring in contrast with one another.

Two Rhythm sound strategies prescribe the periodicity for each chosen sound function in a script or sound mix for a desktop video clip. A Continuous rhythm substrategy places uninterrupted sound (Massed review or Summarized review) or interrupted sound at regular intervals (Spaced review) throughout the duration of a node or between nodes of a hypermedia learning environment. A Discontinuous rhythm substrategy places uninterrupted sound or interrupted sound at regular intervals throughout the sequence or entire program.

Massed, Spaced, or Summarized review strategies prescribe the nature of a particular function's re-occurrence in a sound-image node. Contemporary corporate applications often use a Summarized review substrategy to reinforce behaviour modification role-modelling techniques. Some digitized video news clips also use a Summarized review substrategy to recap the main stories. Personal goal areas on video clips or still graphics can be presented all at once and usually out of programme context using a Massed review substrategy. Problem-solving level skills can be effectively presented with a Spaced review substrategy.

Convergent or Divergent delivery substrategies tend to fall along a continuum, their visual counterparts implemented in educational software to prescribe instructional events or learning activities. The application of convergent or divergent delivery methods to sound design is appropriate when applied in this context. In a Convergent delivery substrategy, the designer or learner presents the questions and supplies the answers. In a Divergent delivery substrategy, the designer or learner must supply their own
answers to controversial questions presented by the program.

In sum, structuring skills are relatively common dramatic devices taught in creative writing and some production courses. Combined with one or more sound functions, these substrategies can assist the designer in prescribing how, where, and how often speech, music or effects should be applied within a sound-image relationship. Developers and user should be advised, however, that modifying structured functions of the sound attribute for learning can require as much time, effort and resources as designing the visuals for a hypermedia frame.

Sound Functions
The process of assigning sound functions, like the structuring activity, should be case-based and conceptual. Sound functions may be conceptualized as possible prescriptions for character, place, time, or subject matter in a sound-image relationship. Figure 2 shows the functions that provide a purpose for sound within the visuals.

<table>
<thead>
<tr>
<th>Functional Intent</th>
<th>Functional Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Sound</td>
<td>- A future event (eg, a direction, an instruction)</td>
</tr>
<tr>
<td>Point-of-View Sound</td>
<td>- A past event (eg, feedback, review)</td>
</tr>
<tr>
<td>Atmosphere Sound</td>
<td>- A subjective point-of-view</td>
</tr>
<tr>
<td>Character's Past/Future Sound</td>
<td>- A performer point-of-view</td>
</tr>
<tr>
<td>Locale Sound</td>
<td>- An objective point-of-view</td>
</tr>
<tr>
<td>Character-in-the-Character Sound</td>
<td>- A feeling</td>
</tr>
<tr>
<td></td>
<td>- A mood</td>
</tr>
<tr>
<td></td>
<td>- In their personal past or future</td>
</tr>
<tr>
<td></td>
<td>- In their professional/public past or future</td>
</tr>
<tr>
<td></td>
<td>- In their private past or future</td>
</tr>
</tbody>
</table>

Figure 2 Functional attributes of the SSF model

When the primary intent of audio is to orient learners about a future event or give feedback about a past event between nodes, Temporal sound should be considered. Future and past events require learners to use this knowledge to paraphrase or to make personal references. Point-of-view (POV), Atmosphere and Locale functions are also assignable to most static or moving visuals. Objective, Subjective or Performer POV sound functions can be prescribed as a function of character. A whispered voice-over, for example, may indicate a Subjective POV sound function. A formal narrator voice-over can provide an Objective POV sound function. Cocktail party sound using some combination of subjective and objective voices would be one example of a Performer POV sound function. An Atmosphere/Feeling/Mood sound function can be associated with the visual information presented in a videoclip, graphic or paragraph of formatted text to provide context. It should imply either more or something else about the atmosphere, a feeling, or a particular mood other than what has already been stated or implied about atmosphere, a feeling, or mood in the image. A Locale sound function fulfills an informational role when it is associated with the visual information presented in a video clip, graphic or paragraph of formatted text. Most often, familiar sounds are produced to establish a place. Usually, the Locale sound function is used realistically as appropriate background speech, music, or sound effect in a node.

Character-related functions, namely Character's Past or Future, and The Character-in-the-character sound function, are usually assigned to moving visuals. Three types of Character's Past/Future sound are presented in this model: Personal, private and public. See Mann (1995a) for details. Character-in-the-character sound should be used to depict a certain recurring aspect of the character's behaviour, certain aspects of the character's (moral) character or when his or her peculiar personality (mask) is intentionally prescribed ambiguously (eg, self-effacing music that presents a multi-faceted personality of the character).

Applying a structure and function
A context-based approach seems most appropriate when applying elements of the SSF model to visual information. In the first context, the user or developer already knows the situation in which to add audio before its exact function or purpose. Here, a structure-before-function approach is used for adding audio to the software. In the second context, contriving an audio situation is the most realistic approach. Here, a function-before-structure approach is used to add audio. See Mann (1995a) for details on the two contexts for applying the SSF model with formatted text, graphics and moving images. In general, the
presence of a function and structure for sound is more important than whether the function or the structure is decided first. The systemic implementation of embedding the SSF model within a computer-based instruction (CBI) model requires that the researcher or developer considers the dynamics of the broader context of the instructional situation.

In educational applications, the SSF model becomes a subsystem within a theory-based model of CBI. The systemic requirement of subsuming the SSF model within a CBI model helps to ensure that the outcomes will focus learners’ attention and elicit appropriate mental operations and behaviours. Alessi and Trollip’s (1991) CBI model, for example seems well-suited for this purpose. The SSF model could appear at step five of their storyboarding level, namely: ‘Draw and revise graphic displays, plan other output’ (Alessi and Trollip, 1991, p 319).

‘Other output comprises any presentation that does not appear on the computer screen. The most frequent and important of these is sound… Sound requires special consideration because it is difficult to describe on paper’ (Alessi and Trollip, 1991, pp 325–6).

The SSF model is both structured and flexible enough to accommodate most logistical constraints; particularly when rapid changes must be made without physically including them in the storyboard. In this way, sound can be fully integrated into the design of educational computing activities and not remain just another feature of it.

DESIGN PRINCIPLES

Several design principles account for how we read and listen to multimedia. These principles distinguish the SSF model from stochastic sound roles and have been evident where outcomes have been positive:

- particular model of human cognitive system;
- separate processing streams;
- synergism from the media mix;
- research controls for modality contamination;
- forgetting and reminiscence traces;
- gist is affected by ageing.

The human cognitive system

One design principle that distinguishes the SSF model from the stochastic methods is its application of current theory about the structure of human cognitive systems. The decision to adopt one of these or some other cognitive structure appears to reside in the ‘explicitness’ and ‘gist’ requirements of the tasks and the interrelatedness of spatial and language representations. ‘Unless the fundamental differences of information processing of these modalities is known, a variety of comparisons of different media seems hardly to be valid’ (Hsia, 1968, p 261).

Production system theorists (Newell and Simon, 1972) for example, offer one perspective, namely, that:

1. humans have a unitary mental architecture;
2. the control processes governing the performance of mental tasks are relatively simple, and;
3. the processes governing human memory depend on symbolic rules.

The significance of a particular symbol in a system depends entirely on the rules in which it occurs, and these rules have to be spelled out with an explicit structure within the system. Applications of production system theory can be found in educational technology research on instructional design (Gagné et al., 1988), message design (Salomon and Gardner, 1986) and the ‘situated cognition’ (Winn, 1991).

Connectionists offer a different perspective; see the work on massively parallel computing by McClelland et al. (1986). Instead of production rules set in an explicit mental structure, they postulate holograms distributed over many human experiences which gradually become clearer as we proceed. Applications of connectionism can be found in educational technology research involving generativity (Hannafin, 1989; Higginbotham-Wheat, 1991; Wittrock, 1992) and constructivism (Jonassen, 1991; Winn, 1991).

Dual Coding theorists (Paivio, 1986) present a third view; that human cognitive systems should be conceptualized as having a dual coding structure. According to this view, linguistic (language-like) information and spatial (picture-like) information are processed through functionally independent, interconnected cognitive systems. Applications of Dual Coding theory can be found in educational technology research involving computer graphics (Rieber and Kini, 1991).

The Separate Streams hypothesis

Another design principle that distinguishes the SSF model from stochastic design approaches is that this model reflects current research about differential processing by readers and listeners, as suggested in the literature (Halliday, 1987; Penney, 1989; Higgin-
botham-Wheat, 1991; Mann, 1995a). Listeners attend to the gist (Hildyard and Olson, 1982; Penney, 1989; Reyna, 1992; Brainerd, 1993). Readers attend to details, surface features of the text (Tannen, 1985), verbatim information (Martin and Briggs, 1986; Penney, 1989). Hildyard and Olson (1982) also reported that performances are better for listening to highlights (gist) and reading the details (substance) verbatim. Others (Brainerd and Reyna, 1990; Reyna, 1992) have found behavioural and neurological evidence that indicates distinct developmental paths for verbatim and gist memories.

Media mix

A third design principle that distinguishes the SSF model from stochastic methods is that the SSF model recognizes Computer-Mediated Communication (CMC) as a ‘media mix’ (Mann, 1992, p. 58). The CMC medium, including e-mail and computer conferencing, shares attributes of both oral and textual communication (Johansson, 1991; Lowry et al., 1994; Yates, 1994); ‘it’s talking by writing’ (Coate, 1992, p. 1). Perceptible changes to treatment effects should therefore be expected when movable type is placed within an online learning environment. A study (Adrianson and Hjelmquist, 1993) designed to test Hildyard and Olson’s (1982) assumption about reading and listening differences (discussed above) mistakenly chose Computer-Mediated Communication (CMC) to represent the text condition. One treatment was partially contaminated by the inherently oral attributes of the CMC medium.

Experimental controls

A fourth design principle that distinguishes the SSF model from stochastic approaches is that the SSF model requires adequate controls in both experimental and formative evaluation research. First, the SSF model incorporates modality-specific practice into its studies. Treatment effects are more likely to be detected when the research subjects are provided with adequate instruction and practice within the modality to be tested. Second, experiments and formative evaluations designed to compare memory for information from computer sound versus computer text or computer graphics ensure that the factors are either controlled or sufficiently explained. Researchers (Weiner, 1989; Shilling, 1991; Zender, 1990) reporting negative or equivalent outcomes with computer-based speech with text attributed their findings to research design principles.

Forgetting and reminiscence

A fifth design principle that distinguishes the SSF model from stochastic sound roles is that the SSF model reflects advice about inoculations against forgetting and takes account of reminiscence effects. Working with sound can produce post-test inoculation effects against forgetting (Brainerd, 1993). This can occur when the visual only treatment is permitted to become contaminated with memory traces from the audio treatment. Studies of reminiscence effects indicate that sound groups can show improvement in their retention over their post-test scores following a latency period for private reflection. Interested readers should refer to the psychological and educational technology literature under ‘reminiscence’ (Martin and Briggs, 1986; Draper and Anderson, 1991; Reyna, 1992; Brainerd, 1993).

Ageing

A sixth design principle that distinguishes the SSF model from stochastic sound roles is that the SSF model accommodates the research on ageing; specifically the developmental stages in life as they relate to our biological capacity to extract gist from sound (Reyna, 1992; Brainerd, 1993). ‘Perceptual learning theory highlights interactions between biologically determined potential and the environment’ (Pressley and McCormick, 1995, p. 182). Human memory systems that support the retention of verbatim information begin to deteriorate in early adolescence while gist information continues to improve (Reyna, 1992). However, human memory for gist can compensate for poor verbatim memory, especially because gist suffices in a broad array of circumstances.

CONCLUSION

What can we do to retain our efficiency and reduce cognitive chaos with programs and online services that require us to distribute our attention between more than one competing message or source? Two things, it seems. First and foremost, apply a design scheme like the SSF model to formatted sound, text, graphics and moving images to assist users in shifting their attention between visual and auditory sensations. The SSF model accounts for the recent reports that find that listeners attend to gist while readers attend to detail. It relies on the ‘explicitness’ and ‘gist’ requirements inherent in tasks and the interrelatedness of...
spatial and language representations. This knowledge should be applied developmentally; as we get older we get better at extracting gist from sound than details from text. Second, avoid stochastic approaches. Stochastic approaches are inadequate; some even utilize the outdated bottleneck psychology of human attention. Stochastic designs employ ill-defined or undefined methods that relegate audio to novelty or accompaniment status.

REFERENCES


we get better at extracting gist from sound than details from text. Second, avoid stochastic approaches. Stochastic approaches are inadequate; some even utilize the outdated bottleneck psychology of human attention. Stochastic designs employ ill-defined or undefined methods that relegate audio to novelty or accompaniment status.

REFERENCES


we get better at extracting gist from sound than details from text. Second, avoid stochastic approaches. Stochastic approaches are inadequate; some even utilize the outdated bottleneck psychology of human attention. Stochastic designs employ ill-defined or undefined methods that relegate audio to novelty or accompaniment status.

REFERENCES


we get better at extracting gist from sound than details from text. Second, avoid stochastic approaches. Stochastic approaches are inadequate; some even utilize the outdated bottleneck psychology of human attention. Stochastic designs employ ill-defined or undefined methods that relegate audio to novelty or accompaniment status.
Grimes, T (1990) Audio-video correspondence and its
Dempster, F (1992) The rise and fall of the inhibitory
Draper, S and Anderson, A (1991) The significance of
degaM, G (1988) Effect of single and dual advance
Duemler, D and Mayer, R (1988) Hidden costs of
Fiedorowicz, C and Trites, R (1985) Autoskill:
Hildyard, A and Olson, D R (1982) On the
Higginbotham-Wheat, N (1991) Generative learning strategies in computer-based instruction: Effects of learner-generated visual and verbal elaborations,
Huff, C and Finholt, T (1994) Social Issues in
Jaak, T (1988) Effectiveness of computerised drill and
A V
Huff, C and Finholt, T (1994) Social Issues in
Jaak, T (1988) Effectiveness of computerised drill and
practice and bisensory input in teaching music reading
A V
Jonassen, D (1988) Integrating learning strategies into
courseware to facilitate deeper processing. In Jonassen, D H (ed), Instructional Designs for Microcomputer
Jonassen, D and Grabinger, S (1990) Problems and issues in
Joyce, B, Weil, M and Showers, B (1992) Models of
Teaching (4th edn), Allyn and Bacon, Needham
Heights, MA.
Kaplan, S and Kaplan, R (1982) Cognition and
Tannen, D (1985) Relative focus on involvement in


BIOGRAPHICAL NOTES

Bruce Mann is Assistant Professor of Computers in Education at the Memorial University of Newfoundland in Canada.

Address for correspondence: The Faculty of Education, Memorial University of Newfoundland, St. John’s, Newfoundland, Canada A1B 3X8. Tel: (709)-737-3416; Fax: (709)-737-2345; e-mail: bmann@calvin.stemnet.nf.ca.