Enhancing educational software with audio: assigning structural and functional attributes from the SSF model

Bruce Mann

Bruce Mann currently teaches educational computing at Memorial University of Newfoundland, Canada. Address for correspondence: Dr. Bruce Mann, Faculty of Education, Memorial University of Newfoundland, St. John’s, Newfoundland, Canada, A1B 3X8.

Abstract

Forty-five years of intuitive combinations of audio-visual information have produced only mixed results. This paper proposes a conceptual, case-based method for adding audio to software. This method, unlike the intuitive approach, requires the researcher to assign structural and functional attributes from the SSF Model to specific software activities.

Introduction

During the days of teaching machines, little was known of the effects of auditory manipulations on language learning (Porter, 1960). Programmed instruction research (Blyth, 1960) reflected an intuitive approach to combining visual and auditory stimuli in educational software.

The student may access auditory material, listen to a passage as often as necessary, then transcribe it. The machine then reveals the correct text. He may listen to the passage again to discover the sources of any error (Skinner, 1960, 147).

The recent literature (Aarntzen, 1994, 1993; Barron and Kysilka, 1993; Jaspers, 1991; Muraida Spector, 1993) echoes the intuitive approach to combining audio-visual information. This approach, along with an ambivalence about using audio in educational technology, has been blamed for inconsistent effects (Barron and Kysilka, 1993; Belland, 1992; Buxton, 1989).

Some between-subjects comparisons with visual presentations have shown increased learning (Anderson-Inman, 1991; Badt, 1988; Higginbotham-Wheat, 1991; Loughner, 1987; Mann, 1992b 1993; Salavert, 1988; Wilburn, 1991). Other studies have reported no significant or practical differences (Barron and Kysilka, 1993; Shilling, 1991; Wiener, 1989; Zender, 1990). Oral advance organiser research has been especially disappointing (Arwady, 1980; Dougall, 1988; Wolter, 1988). A systemic model for designing sound was required: "designer sound for computer systems" (Buxton, 1989, p1).
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, 277)

The Structured Sound Function (SSF) Model is proposed to fill this need. The SSF Model can be distinguished from conventional sound roles, namely: redundant sound, oral reporting, sound cueing and auditory feedback (Mann, 1993).

The SSF model: background
The origin of the Structured Sound Function (SSF) Model is Zuckerman's (1949) instructional film sound categories. Alten (1981) modified these categories to reflect a mass communications approach. Mann (1988) complemented these developmental efforts with reviews of the literature and the application of knowledge about using sound in educational technology.

First, a variety of communications studies research was explored: script writing models were explored for their applicability to motion picture research and development (Lee and Misiorowski, 1978; Field, 1982; Root, 1979), film and television sound (Zettl, 1973). Educational television research (Fleming, 1987; Schramm, 1972) and broadcast news (Coldevin, 1981) was also investigated. Second, relevant literature was sought about presentations that used audio with either moving or static visuals to improve the users' understanding (Buxton, 1987; Fleming, 1987; Grimes, 1990; Milhelm and Martin, 1991; Rieber and Kini, 1991). Third, empirical studies were examined in which computer speech was implemented with static text as a delivery attribute (Anderson-Inman, 1991; Badt, 1988; Higginbotham-Wheat, 1991; Loughner, 1987; Salavert, 1988; Wilburn, 1991). Finally, reports were discovered on converting some screen text into digitised speech to offer learners a more effective alternative to reading instructions and feedback (Alessi and Trollip, 1991; Blyth, 1960; Calvert et al, 1989; deHaemer and Wallace, 1992; Gaudino, 1986; Jaspers, 1991; Porter, 1960; Skinner, 1960; Thomas and Clapp, 1989; Wetzel, 1991).

Educational hypermedia computing
The results of these reviews helped to make informed decisions about how to proceed from Alten's (1981) modifications into educational hypermedia computing activities (Mann, 1992a). Program typology distinguished those that were consciously designed as learning systems from the many others that were simply information environments for browsing. Educational programming either supplanted or activated cognitive strategies, aiming to change attitudes or to improve skill-based or problem solving performance. Adding sound to silent computer programs required a new description about what would and would not constitute a sequence between and within hypermedia nodes. Some variables were dropped and others added to account for the integration of movable type, bit-mapped graphics and desktop video samples; greater interactivity within the program, and; more learner/program control decision-making. The psychological distinctions, however, had to be maintained between the images and the sound, between the stated and implied sound, and between text and
subtext; particularly when rapid changes were being made without being physically included in a script. Only in this way, could audio be integrated into the design of educational computing activities and not remain just another feature of it.

The effort yielded a conceptual model for designing sound in educational computing. The new model was named the Structured Sound Functions Model (Mann, 1992a). It was offered in its current form as a message design method for applying audio in communications technology. Conceptual models are influenced equally by practical experience and scientific method (Richey, 1993).

Conceptual models are analytic in nature. They typically describe the relevant events based upon deductive processes of logic and analysis, as well as inferences from observations. Conceptual models, like theories, are generalised in regard to context. One of their most important functions is to provide the foundation for research, which can create support for an inductive theory (Richey, 1986, 19).

The rationale was to provide researchers with a systemic alternative to the intuitive approach of combining audio and visual information. A case-based method is suggested for assigning structural and functional attributes of the SSF Model to visual information. The case-based method, unlike the intuitive approach, requires the researcher to assign structural and functional attributes from the SSF Model to specific software.

Applying the SSF model: two approaches
There appear to be two general approaches to using the SSF Model with formatted text, graphics and moving images. In the first approach, the researcher discovers a structural context for audio. Activities such as audio windows management (Cohen, 1991), event notification (Gaver et al, 1992) and mode-error prevention (Monk, 1986) lend themselves easily to listening to audio as a viable alternative to reading text. According to the second approach to designing sound with the SSF Model, the researcher already knows a specific purpose for adding audio. Activities such as focusing on screen information (Alessi and Trollip, 1991; Calvert et al, 1989; deHaemer and Wallace, 1992; Mann, 1992b 1993; Thomas and Clapp, 1989; Wetzel, 1991) or deciphering warning signals (Huff and Finholt, 1994), require listening for a specific purpose.

Structure-before-function: the first approach
Most often, a structural context for adding audio is known before an exact function or purpose. These situations warrant a structure-before-function approach to adding audio to software.

A sound structure is a combination of strategies working together with one or more functions. A 'strategy', then, is a conceptual scheme for mediating an intended message or expression. The process of assigning structural attributes should be case-based, as stated earlier. Conceptual models like the SSF Model do not provide specific guidelines, as do procedural models. Rather, they facilitate an understanding of those attributes which impinge on the design and its implementation. The strategies
that comprise a sound structure can refer to the plot, the content organisation category, goal-area, code, or story-spine. Sound strategies and substrategies help the designer or user to prescribe where, how, and for which function exactly each component of the message fits into or works in the overall scheme of each scene or sequence of scenes. Structures place appropriate sound functions next to every image sequence to create an expressionistic conceptualisation of reality. Structural attributes may be better understood by reading from left to right across Figure 1.

Structuring sound functions 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 summarised review strategy; and a convergent or divergent delivery strategy.

**The informational sound strategy**

When the primary intent of the audio is to inform, one of the first four audio-visual relationships (ie, 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. Audio segues, headliners, flashforwards or gimmicks are some examples of the Cueing Informational Sound Substrategy where the chosen sound function foreshadows the visual action. The Counterpointing Informational Sound Substrategy has been used to create a visual cast-against-type characters by providing aesthetic meaning. The Dominating Informational Sound Substrategy rarely appears in desktop video clips. If it did occur more often (eg, in many overly-graphical interfaces), it might eliminate audio-visual redundancy afflicting several educational programs. At its rare best, however, this substrategy atypifies a predictable or hackneyed emotional interlude (eg, injects silence where music would bridge two scenes, or where the attention to image is lost briefly). The Undermining Informational Sound Substrategy is considered to be an ironic or sarcastic use of informational sound in that its effect 'sends up' the meaning in the image. Audio 'accompaniment' and 'support' do not constitute appropriate educational assignments for audio with desktop video, graphics and text and therefore do not appear in the list.

The emotional sound strategy
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.

The sound attribute is often prescribed with either of two Emotional Sound Strategies. In desktop video, the visual action may be a gimmick. Strategically placed, Defining Visual Action Intensity Sound Substrategy prescribes how, where, and how often the chosen sound function should punctuate the visual action. A sound or its absence may imply depth by creating suspense or interest with the image or graphic. Restraint through the sparse use of silence or room noise may imply suspense or interest in the story or message. Punctuating an Emotional Highlight Sound Substrategy may also create depth by implying suspense or interest with two or more sound functions in a sound-image relationship. For example, electro-acoustically-produced sound designs may create deeper-felt emotions in the learner.

The pacing strategy
Since the 1930's, pacing has been used effectively in applications of sound to image in artistic, education, 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.

The rhythm strategy
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 Summarised 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.

The review sound strategy
When one or more of the Review Strategies have been chosen for a particular sound function in a hypermedium, then each strategy should show the extent to which the designer or learner wants to manipulate the reality suggested by the image. The decisions should prescribe how, where or how often each sound function will be Massed, Spaced, or Summarised in the node or between nodes. The three review substrategies prescribe the nature of a particular function's re-occurrence in a sound-image node. Contemporary corporate applications often use a Summarised Review Substrategy to reinforce behaviour modification role-modelling techniques. Some digitised video news clips also use a Summarised 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.

The delivery sound strategy
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 dramatic video clip with a convergent delivery, catharsis is reached through the ultimate confrontation of two opposing personalities. In a Divergent Delivery Substrategy, the designer or learner must supply their own answers to controversial questions presented by the program. Two or more diverging points of view emerge but are presented equally for scrutiny. In some cases, it is important to retain this scholastic presentation format wherein no attempt is made to editorialise or show favour.

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.

**Function-before-structure: another approach**

In other situations, a purpose or function for adding audio first becomes evident. In these cases, the learning activities require a function-before-structure approach to adding audio. Functions of sound are characteristics that in themselves prescribe or describe new meaning within the sound-image relationship, instructional method notwithstanding. Functions may be conceptualised as possible prescriptions for character, place, time, or subject matter in a sound-image relationship. Figure 2 shows the six functional attributes that are conceptualised as prescriptions for character, place, time, or subject matter in an auditory-visual relationship.

The process of assigning sound functions, like the structuring activity, should be case-based and conceptual. Functional attributes of audio can be applied by reading from left to right in Figure 2.

**Temporal sound**

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 a personal references. The Temporal Sound Function may be conceptualised as the visual equivalent of a developer's gimmick' such as an instructional epitome or foreshadowing a dramatic event. When the Temporal Sound Function is created, it implies more or something else about a future or past event than what has already been stated or implied about the future or past in the image. Unlike the Character's Past Sound Function which strives to associate personal past and future events, the Temporal Sound Function informs the learner about an occurrence in history unrelated to their own past, like mythical music or an ancient metaphor.

**Point-of-view, atmosphere and locale**

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. When an Objective, Subjective or Performer Points Of View (POV) Sound Function is created, then it must imply another point of view or more about the point of view than what has already been stated or implied about the reference by the image. A whispered voice-over, for example, may indicate a Subjective POV Sound Function. A formal a narrator voice-over can provide a 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.

The 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.

The Locale Sound Function can fulfill 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. When the Locale Sound Function is created, it must imply more or something else about the locale than what has already been stated or implied about it in the image.

**Character-related functions**

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. When the Character's Personal Past or Future Sound Function is created, then the character's professional past implies what the character has been doing for a living: his roles in the corporation, history of relationships with co-workers, and so forth. When the Character's Personal Past or Future Sound Function is created, then the character's personal past or future implies the nature of his/her marital history,
history of educational background, job, and socio-economic history. When the Character's Private Past or Future Sound Function is created, then the character's private past or future may be interactively propelled through a series of events at different nodes based on the learner's preferences and the level of program control. The private past implies the need for fame, money, stability of the marriage, peculiar tendencies, and flaws of character or personality. When the Character's Personal, Private or Public Past/Future Sound Function is created, then it must imply more or something other about the character's past or future than what has already been stated or implied about his/her past in the image. Unlike the Character-In-The-Character Sound Function, this function does not plumb the depths of the character's psyche. Moreover, this sound function tries to answer specific questions related to the message design or plot.

The Character-In-The-Character Sound Function refers to the subtext, story spine or tragic flaw in a character. The 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 his or her peculiar personality (mask) is intentionally prescribed ambiguously (eg, self-effacing music that presents a multi-faceted personality of the character). When the Character-In-The-Character Sound Function is prescribed, the person's character (ie, his or her habits) or their persona (ie, his or her worldly mask) is created. This character or persona must imply more or something else about the character than what has already been stated or implied about him/her in the image. Specific questions in relation to the subtext of the plot are best answered by implication using Character-In-The-Character Sound Function: What does he or she really want? Who/what's really stopping the character from getting what he or she wants? The intention behind prescribing this sound function is to generate a controversy with other sound functions (ie, POV, Locale, Character's Past, etc.) working in the sound design.

**Applying the SSF model**

Educational software development requires that auditory and visual presentation attributes become subsumed within a theory-based system of computer-based instruction. Alessi and Trollip's (1991) model of computer-based instruction (CBI), for example, seems well-suited to this purpose.

The structural and functional attributes of the SSF Model can be nested in their storyboarding stage of software development. This requirement helps to ensure that the sound and visuals will focus learners' attention and elicit appropriate mental operations and behaviours. Applied in this way, then, learning with sound can activate or supplant the learners' cognitive strategies, change attitudes and improve performance. The pertinent instructions, directions, practice and feedback are developed in hypermedia nodes (Grabinger et al, 1992). Hypermedia nodes are information fragments of text, graphics and sound in which learners can generate or access information on the interface. Two approaches have been identified for applying the structural and functional attributes from the SSF Model to formatted
text, bit-mapped graphics and video clips within and between hypermedia nodes: the structure-before-function approach and the function-before-structure approach.

Structure-before-function: recent applications

Recent projects in Britain (Gaver et al., 1992; Hartas and Moseley, 1993) and North America (deHaemer and Wallace, 1992; Booth-Butterfield and Gutowski, 1993; Jordan, 1990; Sutherland, 1993) have investigated the structural attributes of audio. An ongoing project at Rank Xerox in Cambridge, England seems to have captured some of the structure-before-function approach of adding audio to software. The Ravenscroft Audio Video Environment (RAVE) is an audio visual system connected to event notification software on a local area network server (Gaver et al., 1992). The purpose of the project is to promote collaborative work by helping users to maintain an awareness of ongoing events from their distant locations. Several functional attributes of audio in RAVE appear to be well-represented. The RAVE's implementation of audio with visuals satisfies the criteria for Mood, Point Of View, Locale and Temporal Sound Functions. An event daemon's service watches for specified event types (e.g., a RAVE Meeting in the conference room) and produces notification events when they are detected. The personalisation of the event in this way satisfies the criteria for another functional attribute from the model, namely: the Character's Future Sound function.

Similar investigations continue in foreign language learning software. The current technology provides users with new insights about where and how often audio should be used. VoiceCart (Hartas and Moseley, 1993), Dasher (Sutherland, 1993) and Autoskill (Jordan, 1990) offer their users immediate structuring opportunities. The capabilities of recapping, precapping and annotating documents requires a covert skill: user message design. Based on previous research (Mann, 1988, 1992a, 1993) with post-secondary students, even better results can be expected when users assign structural and functional attributes from the SSF Model to the content. This case-based method can also be used to apply speech annotation to word-processed documents and spreadsheets.

Function-before-structure: recent applications

Research in the Netherlands (Aarntzen, 1994; Jaspers, 1991) and Canada (Mann, 1988 1992a 1993) has taken a function-before-structure approach by exploring the functional attributes of audio first. Three studies were conducted with inservice teachers (Mann, 1992b, 1993) using informational temporal speech to solve the problem of users' inattention to critical information in educational software. The nature of the educational problem presented opportunities to assign a functional attribute first: Audio references could be added to the software concerning specific future and past events in the software (i.e., a Temporal function). Unlike oral advance organisers, however, these temporal references contained a constructive requirement. For this reason, a structural attribute (informational, cueing, etc) was considered only after the Temporal function was chosen from the list of functional attributes. Counterpointing (in Figure 2) was then assigned to alternate the dominant and
subordinate roles between the video and audio channels. Next, Summarised Review was assigned to permit learners to use the information to paraphrase and make references before receiving feedback. Summarised Review and Divergent Delivery strategies were then assigned to limit this capability with a requirement to recall personal knowledge to support or contradict visual information. Since the visual information was static (ie, text and graphics), Pacing and Rhythm strategies were not considered. Results of these studies using the function-before-structure approach showed that the immediate effects were greatest when future events were presented in speech and past events in text. Retention effects (after four-weeks) were greatest when both future and past events were presented in speech. In the latter case, opportunities were provided for private reflection and reminisce without further intervention. Although the function-before-structure approach may not be ideal, recent findings using the SSF Model in some educational settings suggest a better-than-average probability of success.

Conclusion
Evolving hardware and software attributes will surely permit users more adaptive and non-linear interactions, and a higher capacity for differentiating audio and visual presentations. This paper introduced the case-based SSF Model for assigning structural and functional attributes of audio into educational computer programs. Although this model can be applied to any didactic or interactive learning environment with audio digitising or synthesising attributes, its application should be selective and fully integrated with current theory and exemplary practice. Although this model may not be ideal for applying sound to visuals, recent findings in some educational settings suggest a better-than-average probability of success.

References
Enhancing educational software with audio


Booth-Butterfield S and Gutowski C (1993) Message modality and source credibility can interact to affect argument processing *Communication Quarterly* 41 1 77–89.


Cohen M (1991) Multidimensional audio windows: Extending user interfaces through the use of spatial auditory information *Dissertation Abstracts International* V52(05) SECB PP2745 (University Microfilms No. ADG91-28969).


Dougall G (1988) Effect of single and dual advance organizers upon learning and retention of self-instructional text *Unpublished Master’s Thesis* University of Saskatchewan, Saskatoon, Canada.


Hartas C and Moseley D (1993) 'Say that again, please': a scheme to boost reading skills using a computer with digitized speech *Support for Learning* 8 1 17–21.


Loughner H (1987) A linear model for a dual coding, Piagetian and information processing theoretic explanation of the role of visuo-spatial learning, verbal and spatial conditions and multiple choice test behaviour information in secondary school chemistry achievement *Dissertation Abstracts International* V48(06) SECA PP1429 (University Microfilms No. ADG87-19306).
Mann B L (1988) Guidelines for the instructional design of motion picture sound Unpublished Masters Thesis Concordia University, Montreal, Canada.
Shilling W (1991) Kindergarten learners constructing knowledge about literacy using conventional materials and computers Dissertations Abstracts International V52(11) SECA PP3876 (University Microfilms No. ADG92-11460).
Wilburn H (1991) An investigation of interaction among learning styles and computer assisted instruction with synthetic speech Unpublished Doctoral Dissertation University of Texas at Austin, USA.


Zuckerman J (1949) *Music in Motion Pictures: A Review of the Literature with Implications for Instructional Films* Instructional Film Research Program, Pennsylvania State College, USA.