

# A Coherence Effect in Multimedia Learning: The Case for Minimizing Irrelevant Sounds in the Design of Multimedia Instructional Messages

Roxana Moreno and Richard E. Mayer  
University of California, Santa Barbara

The authors tested the recommendation that adding bells and whistles (in the form of background music and/or sounds) would improve the quality of a multimedia instructional message. In 2 studies, students received an animation and concurrent narration intended to explain the formation of lightning (Experiment 1) or the operation of hydraulic braking systems (Experiment 2). For some students, the authors added background music (Group NM), sounds (Group NS), both (Group NSM), or neither (Group N). On tests of retention and transfer, Group NSM performed worse than Group N; groups receiving music performed worse than groups not receiving music; and groups receiving sounds performed worse (only in Experiment 2) than groups not receiving sounds. Results were consistent with the idea that auditory adjuncts can overload the learner's auditory working memory, as predicted by a cognitive theory of multimedia learning.

Humans are designed to integrate multimodal stimuli into one meaningful experience, such as when they associate the sound of thunder to the visual image of lightning in the sky. However, when the process of lightning formation is to be taught using a computer, the instructional designer is faced with the need to choose between several alternative presentation formats to promote meaningful learning (Mayer & Moreno, 1998). Within the visual modality, for example, the process of lightning may be shown by a static diagram, an animation, or a video, and it may be described by visually presented text. Within the auditory modality, for example, the process of lightning may be accompanied by sound effects or background music, and it may be described by an auditorily presented narration. The present study examines one aspect of multimedia design, the role of auditory adjuncts such as background music and sound.

## Examples of Auditory Adjuncts in Multimedia Messages

Consider the following multimedia learning scenario. A student sits at a computer station and clicks on an encyclopedia entry for lightning. The computer presents an animation depicting the steps in lightning formation along with a corresponding narration describing the steps in spoken words. This format has been shown to be effective in fostering student understanding as indicated by performance on problem-solving transfer questions (Mayer & Anderson, 1991, 1992; Mayer & Moreno, 1998; Mayer & Sims, 1994; Moreno & Mayer, 1999). However, the difference between

an amateur multimedia program and a professional one often involves the addition of bells and whistles like background music (such as an instrumental music loop) and sounds (such as blowing wind and crackling ice).

Does adding bells and whistles improve multimedia learning? If these additions promote learning, instructional designers can feel confident in using them. In contrast, if these additions actually detract from student learning in some situations, instructional designers may want to restrict their use. To answer this question, we conducted a series of studies examining the cognitive consequences of adding auditory adjuncts to a narrated animation involving a scientific explanation.

## Theoretical Issues in the Design of Multimedia Messages

On the theoretical level, research on auditory adjuncts in multimedia presentations provides a useful venue for testing an important aspect of a dual-processing model of multimedia learning (Mayer, 1997; Mayer & Moreno, 1998; Moreno & Mayer, 1999). The model is based on three major assumptions: (a) learners have at least two different information-processing channels, such as a visual channel and an auditory channel (Baddeley, 1992; Paivio, 1986); (b) each channel (or type of working memory) has a limited capacity (Baddeley, 1992; Chandler & Sweller, 1991); and (c) major steps of cognitive processing within each channel (or each type of working memory) involve selecting relevant material for further processing, organizing the selected material into a coherent representation, and integrating the verbal and visual representations with one another and with relevant material from long-term memory (Mayer & Wittrock, 1996; Paivio, 1986). In this study we focus on the issue of the limited capacity of the auditory channel (or auditory working memory), an issue that has not been previously investigated in the context of multimedia learning. In particular, we examine two competing theories—arousal theory, which

---

Roxana Moreno created the multimedia materials used in the studies. Tricia Mautone assisted in data collection and scoring.

Correspondence concerning this article should be addressed to Roxana Moreno or Richard E. Mayer at the Department of Psychology, University of California, Santa Barbara, California 93106. Electronic mail may be sent to [moreno@psych.ucsb.edu](mailto:moreno@psych.ucsb.edu) or to [mayer@psych.ucsb.edu](mailto:mayer@psych.ucsb.edu).

favors auditory adjuncts, and coherence theory, which rejects auditory adjuncts.

### Evidence Supporting Arousal Theory

On the one hand, arousal theory holds that adding entertaining auditory adjuncts will make the learning task more interesting and thereby increase the learner's overall level of arousal (Weiner, 1990). According to arousal theory, this increase in arousal results in a greater level of attention so that more material is processed by the learner, resulting in improved performance on tests of retention and transfer (Dewey, 1913; Renninger, Hidi, & Krapp, 1992).

In the field of television viewing, a vast series of studies has been conducted to understand the relation between attention and comprehension (Collins, 1982; Huston & Wright, 1983; Kozma, 1991). Although this line of research has been conducted only with children in incidental learning circumstances, a theoretical framework has emerged to explain how viewers' attention can be controlled by the use of formal features—visual or auditory techniques such as zooms, pace, level of action, sound effects, music, and so forth, which are distinct from the content of the presentation and can be used across different messages (Huston & Wright, 1983). Children's attention during TV viewing is discontinuous and periodic, with the audio features of the medium serving as cues to recapture attention (Anderson & Lorch, 1983). Sound effects have been especially associated with increased attention to a presentation in children (Alwitt, Anderson, Lorch, & Levin, 1980; Anderson & Levin, 1976; Calvert & Gersh, 1987; Calvert, Huston, Watkins, & Wright, 1982; Calvert & Scott, 1989). Although the use of music in visual presentations has generally shown little positive effect on learning, there are a few studies of young children where a positive effect was found (Mann, 1979; Wagley, 1978).

In sum, the generalizable framework derived from this line of research is described by Kozma (1991, p. 194): "This research paints a picture of television viewers who monitor a presentation at a low level of engagement, their moment-to-moment visual attention periodically attracted by salient audio cues and maintained by the meaningfulness of the material. This creates a window of cognitive engagement." Therefore, auditory perceptual features not only elicit and maintain attention but aid in the selection of the content materials to be further processed.

### Evidence Supporting Coherence Theory

Previous studies examining how students learn scientific explanations from text and illustrations found that adding extraneous sentences or illustrations, which are called *seductive details*, resulted in poorer retention and transfer performance, even when the added material was intended to be interesting or entertaining (Garner, Gillingham, & White, 1989; Renninger et al., 1992; Harp & Mayer, 1997, 1998; Mayer, Bove, Bryman, Mars, and Tapangco, 1996). We refer to this finding as a coherence effect, and it is the basis of what we call a *coherence principle*: When giving a multime-

dia explanation, use few rather than many extraneous words and pictures (Mayer, 1999).

Coherence theory holds that auditory adjuncts can overload the auditory channel (or auditory working memory). Any additional material (including sound effects and music) that is not necessary to make the lesson intelligible or that is not integrated with the rest of the materials will reduce effective working memory capacity and thereby interfere with the learning of the core material. As less of the core material is selected for further processing, the result will be poorer performance on a retention test. In addition, when learners focus their limited auditory processing capacity on receiving the incoming auditory material, they have less capacity left for building a coherent verbal representation and for connecting it with other representations. The result is poorer performance on a transfer test.

We tested these competing hypotheses about multimedia learning in a set of two experiments, in which we compared learners who received multimedia presentations that included background music (Group NM), sounds (Group NS), both (Group NSM), or neither (N). All students took tests of retention, transfer, and matching. On the basis of the arousal hypothesis, we predicted that students in Group N would perform worse on these tests than students in the other groups, especially Group NSM. On the basis of the coherence hypothesis, we made the opposite prediction that students in Group N would perform better on these tests than students in the other groups, especially Group NSM.

## Experiment 1

The purpose of Experiment 1 was to provide a test of the coherence and arousal hypotheses and thereby contribute to a cognitive theory of multimedia learning. By examining the cognitive consequences of adding auditory adjuncts to a multimedia instructional message, our goal was to contribute to a growing body of research-based principles for the design of multimedia learning environments.

### Method

*Participants and design.* The participants were 75 students recruited from the psychology subject pool at the University of California, Santa Barbara. All participants indicated that they lacked experience in meteorology. There were 19 students in the narration group (Group N), 18 in the narration plus environmental sounds group (Group NS), 19 in the narration plus music group (Group NM), and 19 in the narration plus environmental sounds and music group (Group NSM). The median age was 18, and the overall percentage of women was 35%. Neither age nor gender differed significantly among the groups,  $F(3, 71) = 2.10$ ,  $MSE = 3.79$ ,  $p = .11$ , and  $\chi^2(3, N = 75) = 1.17$ ,  $p = .75$ , for age and gender, respectively.

*Materials and apparatus.* For each participant, the paper-and-pencil materials consisted of a participant questionnaire, a retention test, a matching test, and a 4-page transfer test, with each typed on 8.5 by 11 inch sheets of paper. The participant questionnaire solicited information concerning the participant's SAT scores, gender, and meteorology knowledge. We assessed meteorology knowledge by using a 6-item knowledge checklist and a 5-item self-rating. The checklist consisted of instructions to "Please place

a check mark next to the items that apply to you," followed by a list of six items: "I regularly read the weather maps in the newspaper." "I know what a cold front is." "I can distinguish between cumulus and nimbus clouds." "I know what a low pressure system is." "I can explain what makes the wind blow." "I know what this symbol means: [symbol for cold front]" "I know what this symbol means: [symbol for warm front]". On the 5-item self-rating, students were asked to "Please put a check mark indicating your knowledge of meteorology (weather)" on a 5-point scale ranging from *very little* to *very much*.

The participant questionnaire also included four questions that were intended to assess the acoustic preference of the student. The first question asked, "If you had to study a textbook chapter for an exam, which of the following conditions would you typically choose to do so?" and the participant had to check one of the following options: "Study the chapter while listening to the TV in the background," "Study the chapter while listening to music with lyrics," "Study the chapter while listening to instrumental music," or "Study the chapter in the library or a quiet room." The second question asked, "How many hours did you listen to music last week?" and the participant had to check one of the following options: "10 or more," "More than 5 but less than 10," "More than 1 but less than 5," or "0 to 1." The third question asked, "While walking on the street or hiking outdoors, how aware of the sounds around you do you think you are?" Students had to put a check mark on a 5-point scale ranging from *very much* (4) to *very little* (1). The last question asked, "Which of the following statements apply to you? and the participant had to check as many of the following options as would apply: "I own a musical instrument that I can play," "I sing in a band or choir," "I can play a musical instrument," or "I can read music."

The retention test contained the following instructions at the top of the sheet: "Please write down an explanation of how lightning works." The transfer test consisted of the following four questions, each typed on separate sheets: "What could you do to decrease the intensity of lightning?" "Suppose you see clouds in the sky, but no lightning. Why not?" "What does air temperature have to do with lightning?" "What do electrical charges have to do with lightning?" The matching test presented four frames from the animation along with the following instructions:

Circle cool moist air and write C next to it. Circle the warmer surface and write W next to it. Circle the updraft and write U next to it. Circle the freezing level and write F next to it. Circle the downdraft and write D next to it. Circle the gusts of cool wind and write G next to it. Circle the stepped leader and write S next to it. Circle the return stroke and write R next to it.

The computerized materials consisted of four computer programs for multimedia presentations on how the lightning process works (N, NM, NS, and NSM versions), with each followed by an identical computer program that played a set of environmental sounds while displaying a blank screen. All program versions generated an identical animation depicting air moving from the ocean to the land, water vapor condensing to form a cloud, the rising of the cloud beyond the freezing level, the formation of crystals in the cloud, the movement of updrafts and downdrafts, the building of electrical charges within the cloud, the division of positive and negative charges, the traveling of a negative stepped leader from the cloud to the ground, the traveling of a positive stepped leader from the ground to the cloud, the negative charges following the path to the ground, the meeting of the negative leader with the positive leader, and the positive charges following the path toward the cloud. The N version also included concurrent narration describing each of the major events in words spoken at a slow rate by a male voice. The NM version included the same narration used

in the N version plus an instrumental music loop from a media clip CD ROM, which lasted 20 s and was designed to play in the background of multimedia presentations. The music was synthesized and bland. The NS version included the same narration used in the N version, plus it played the following set of environmental sounds during the respective event: (a) a gentle wind, for the portion of the animation depicting air moving from the ocean to the land; (b) water condensing in a pot, for the portion of the animation depicting water vapor forming a cloud; (c) a clinking sound, for the portion of the animation depicting the formation of crystals in the cloud; (d) strong wind, for the portion of the animation depicting the downdrafts; (e) static sound, for the portion of the animation depicting the building of electrical charges; (f) a crackling sound, for the portion of the animation depicting the traveling of the charges from the cloud to the ground and vice versa; and (g) thunder, for the portion of the animation depicting the flash of lightning. The NSM version was identical to the NS version, but it also included the music loop, which played in the background of the NM presentation. The narration for all versions was clearly audible. In the NS, NM, and NSM versions, where sounds, music, or both were presented in addition to the narration, neither the music nor the sounds masked or made the narrative perceptually less discernible. The four versions had an identical total duration of 180 s and were designed to pause after the animation was over and wait for a mouse click to continue. The second part of the computer program, which played identically for each of the four versions, also had a total duration of 180 s. It started at the click of the computer mouse and played the same set of environmental sounds used in versions NS and NSM, in the same order and duration but while displaying a white blank screen. The multimedia presentations were developed using Director 5 and Soundedit 16 (Macromedia, 1997).

The apparatus consisted of 5 Macintosh IICI computer systems, with each including a 14-inch monitor and Sony headphones.

**Procedure.** Participants were tested in groups of 1 to 5 per session. Each participant was randomly assigned to a treatment group (N, NS, NM or NSM) and was seated in front of an individual computer.

First, participants completed the questionnaire at their own rates. Second, the experimenter presented oral instructions stating that the computer would show an animation of how the process of lightning works and that when the computer was finished the experimenter would have some questions for the participants to answer. All participants were told to put on headphones and to press the space bar to begin the presentation. Third, on pressing the space bar, the respective version of the animation was presented once to all participants. Fourth, when the presentation was finished, the retention sheet was distributed along with instructions to write down an explanation of how lightning works. After 5 min, the participants were asked to stop writing. Fifth, the experimenter replaced the participants' pencils with a colored pen. Participants were asked to put on the headphones and click on the computer mouse to listen to a set of sounds. They were also instructed to use the colored pen to add in their retention sheet any ideas that they might have forgotten but were now remembered by the help of the sound cues. After the sounds had finished playing (for 180 s), the recall sheet was collected. Sixth, the experimenter presented oral instructions for the retention test, stating that there would be a series of question sheets and that for each the participant should keep working until told to stop. Then, the problem-solving sheets were presented one at a time for 3 min each with each sheet collected by the experimenter before the subsequent sheet was handed out. Finally, the matching test was presented and collected after 3 min.

**Scoring.** A scorer who was not aware of the treatment condition of each participant determined the experience score, acoustic score, retention score, cued-retention score, transfer score, and matching score for each participant. A second rater scored a randomly picked subset of 20% of the tests. Agreement between both scorers was 94% on the retention and cued-retention tests, 88% on the transfer tests, and 100% on the matching tests. The experience score was computed from each participant's questionnaire by adding all the check marks from the 6-item meteorology knowledge checklist plus the participant's self-rating score (ranging from 0 points for checking *very little* to 4 points for checking *very much*). Data for students who scored above 5 were eliminated and new students were run in their places ( $n = 6$ ). The acoustic preference score was computed from each participant's questionnaire by adding the scores for the four acoustic-preference questions. For each of the first three questions, a score ranging from 4 points to 1 point was given when the participant had checked the first, second, third, or fourth option, respectively. For the fourth question, we computed the score by adding all the options that the participant had checked.

We computed a retention score for each participant by counting the number of major idea units (out of 19 possible) that the participant produced on the retention test. One point was given for correctly stating each of the following 19 idea units regardless of wording: (a) cool air moves, (b) it becomes heated, (c) it rises, (d) water condenses, (e) the cloud extends beyond the freezing level, (f) crystals form, (g) water and crystals fall, (h) it produces updrafts and downdrafts, (i) people feel the gusts of cool wind before the rain, (j) electrical charges build, (k) negative charges fall to the bottom of the cloud (or positive charges go to the top), (l) a step leader travels down, (m) in a step fashion, (n) the leaders meet, (o) at 165 feet from the ground, (p) negative charges rush down, (q) they produce a light that is not very bright, (r) positive charges rush up, and (s) this produces the bright light people see as a flash of lightning.

We computed a cued-retention score for each participant by counting the number of major idea units that were added with the colored pen to the retention sheet. One point was given for each of the above described 19 idea units that was not included originally in the retention score but was added during the cueing period.

We computed a transfer score for each participant by counting the number of acceptable answers that the participant produced across the four transfer problems. For example, acceptable answers for the first question about how to decrease lightning intensity included removing positive ions from the ground; acceptable answers for the second question about why could there be clouds but no lightning included stating that the tops of the clouds might not be high enough to freeze; acceptable answers for the third question about how is temperature related to lightning included stating that the air must be cooler than the ground; and acceptable answers for the fourth question about how are electrical charges related to lightning included the difference in electrical charges within the cloud. Questions were open ended, so participants could receive as many points per problem as correct answers they gave.

We computed a matching score for each participant by counting the number of correctly labeled elements (out of 8 possible) on the matching test. Participants received 1 point for each part that was circled and labeled with the appropriate letter.

## Results and Discussion

Table 1 summarizes the mean scores (and standard deviations) for the four groups on each of the three tests in Experiment 1. For each of the three dependent measures

Table 1  
*Mean Scores and Standard Deviations for Four Groups on Three Tests—Experiment 1*

Group	Retention test		Transfer test		Matching test	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
N	11.05 <sub>a</sub>	2.78	2.74 <sub>a</sub>	1.15	7.10 <sub>a</sub>	1.66
NS	11.61 <sub>a</sub>	3.88	3.06 <sub>a</sub>	1.83	6.94 <sub>a</sub>	1.16
NM	9.32 <sub>a,b</sub>	3.20	1.95 <sub>a,b</sub>	1.18	6.79 <sub>a</sub>	1.44
NSM	6.26 <sub>b</sub>	3.77	1.10 <sub>b</sub>	1.05	6.21 <sub>a</sub>	1.90

*Note.* Column means not sharing subscripts significantly differ from one another ( $p < .05$ ). Scores ranged from 0 to 19 for the retention test, from 0 to 7 for the transfer test, and from 0 to 8 for the matching test. N = narration; NS = narration plus environmental sounds; NM = narration plus music; NSM = narration plus environmental sounds and music.

(i.e., retention, transfer, and matching), the data were subjected to a two-way analysis of variance (ANOVA) with the between-subjects factors being presence or absence of sounds (NS and NSM vs. N and NM, respectively) and presence or absence of music (NM and NSM vs. N and NS, respectively).

**Coherence effect on verbal recall.** According to the arousal hypothesis, adding entertaining auditory adjuncts should increase the amount of material that students remember from the narration. In contrast, according to the coherence hypothesis, adding sufficient amounts of extraneous auditory material to a narrated animation should reduce the amount of material that students remember from the narration. The first columns of Table 1 present the mean scores and standard deviations of the four groups on the retention test. A two-way ANOVA revealed that students remembered significantly less verbal material when music had been presented ( $M = 7.65$ ,  $SD = 3.73$ ) than when no music had been presented ( $M = 11.37$ ,  $SD = 3.29$ ),  $F(1, 71) = 21.99$ ,  $MSE = 11.61$ ,  $p < .0001$ ; but there was no significant difference between students who had received environmental sounds ( $M = 8.87$ ,  $SD = 3.73$ ) and those who had not received environmental sounds ( $M = 11.37$ ,  $SD = 3.29$ ),  $F(1, 71) = 2.30$ ,  $MSE = 11.61$ ,  $p = 0.13$ . There was a significant interaction between music and sounds,  $F(1, 71) = 4.41$ ,  $MSE = 11.61$ ,  $p < .05$ , in which the combination of music and environmental sounds (i.e., Group NSM) was particularly detrimental to retention performance. Supplemental Newman-Keuls tests (with  $\alpha$  at .05) indicated that Group NSM recalled significantly less information than each of the other groups and that Group NM recalled significantly less information than Groups N and NS, which did not differ from each other.

A one-way ANOVA revealed that the four groups did not differ significantly in their scores on the cued-retention test,  $F(3, 71) = 0.09$ ,  $MSE = .15$ ,  $p = 0.96$ . Similarly, a two-way ANOVA revealed no significant main effect for music,  $F(1, 71) = 0.18$ ,  $MSE = .27$ ,  $p = 0.68$ ; no significant main effect for sounds,  $F(1, 71) = 0.05$ ,  $MSE = .08$ ,  $p = 0.82$ ; and no significant interaction between music and sounds,  $F(1, 71) = .05$ ,  $MSE = .08$ ,  $p = 0.82$ . Listening to the sounds did not prove to help students cue the recall of the idea units.

Overall, these results are inconsistent with the arousal hypothesis and consistent with the coherence hypothesis. Adding a sufficient amount of extraneous auditory material—in the present study in the form of music—tended to hurt students' retention of the information in the narration. Consistent with the coherence hypothesis, we attribute this effect to auditory overload created by adding too much extraneous auditory material. Given that the auditory channel is limited in capacity, there are fewer cognitive resources for processing the narration when learners must also process concurrent music and sounds in the auditory channel. The result of this auditory overload is that less of the narration is processed and eventually stored for later retrieval.

*Coherence effect on problem-solving transfer.* The arousal hypothesis predicts that adding entertaining auditory adjuncts should increase problem-solving transfer, whereas the coherence hypothesis predicts that adding sufficient amounts of extraneous auditory material to a narrated animation should decrease problem-solving transfer. The second columns of Table 1 present the mean scores and standard deviations of the four groups on the problem-solving transfer test. A two-way ANOVA revealed that students generated significantly fewer solutions when music had been presented ( $M = 1.54, SD = 1.19$ ) than when no music had been presented ( $M = 2.84, SD = 1.52$ ),  $F(1, 71) = 17.65, MSE = 1.79, p < .0001$ ; but there was no significant difference between students who had received environmental sounds ( $M = 2.05, SD = 1.76$ ) and those who had not received environmental sounds ( $M = 2.34, SD = 1.21$ ),  $F(1, 71) = 0.62, MSE = 1.79, p = 0.43$ . Finally, there was a significant interaction between music and sounds,  $F(1, 71) = 4.41, MSE = 1.79, p < .05$ , in which the combination of music and environmental sounds (i.e., Group NSM) was particularly detrimental to transfer performance. Supplemental Newman-Keuls tests (with  $\alpha$  at .05) indicated that Group NSM generated significantly fewer solutions than each of the other groups and that Group NM generated significantly fewer solutions than Group N and Group NS, which did not differ from each other.

This pattern is largely identical to that obtained for the retention test and therefore supports the same conclusion in favor of the coherence hypothesis. Adding extraneous auditory material in the form of music tended to hurt students' understanding of the lightning process. Adding relevant and coordinated auditory material in the form of environmental sounds did not hurt students' understanding of the lightning process. Together with the retention results, these findings suggest that auditory overload can be created by adding auditory material that does not contribute to making the lesson intelligible. The results of this auditory overload are that fewer of the relevant words and sounds may enter the learner's cognitive system and fewer cognitive resources can be allocated to building connections among words, images, and sounds.

*Coherence effect on visual-verbal matching.* The mean matching scores (and standard deviations) for the four groups are listed in the third columns of Table 1. A two-way ANOVA revealed no significant main effect for music,  $F(1, 71) = 2.57, MSE = 2.45, p = 0.11$ ; no significant main

effect for environmental sounds,  $F(1, 71) = 0.98, MSE = 2.45, p = 0.33$ ; and no significant interaction between music and environmental sounds,  $F(1, 71) = 0.18, MSE = 2.45, p = 0.67$ . Thus, the matching test failed to produce effects that would allow us to distinguish between the arousal and coherence hypotheses. Unlike the case of retention and problem solving, the matching scores of the groups did not differ significantly from each other. The reason for this discrepancy might reside in the fact that the matching test was not a sensitive enough measure of students' learning processes. As can be seen from the third column of Table 1, all students scored high on this particular measure, suggesting a ceiling effect.

*Coherence effect and individual differences.* For purposes of data analyses, students were classified as low or high in acoustic preference on the basis of a median split. A two-way ANOVA with treatment group and acoustic preference as between-subjects factors failed to reveal significant interactions on the retention test,  $F(3, 67) = .36, MSE = 4.38, p = 0.55$ ; transfer test,  $F(3, 67) = 2.55, MSE = 4.20, p = 0.20$ ; or matching test,  $F(3, 67) = 1.77, MSE = 4.28, p = 0.88$ . Thus, there is no evidence that acoustic preference affected the size of the coherence effect for Experiment 1. Individual differences in acoustic preference, SAT score, and gender were not the main focus of the study, and these factors were not part of the research design.

## Experiment 2

Experiment 1 provided consistent evidence for the coherence hypothesis in which adding auditory adjuncts to multimedia instructional messages can result in overloading the learner's auditory channel. The purpose of Experiment 2 was to provide an additional, independent test of the coherence and arousal hypotheses using different materials and different participants.

### Method

*Participants and design.* The participants were 75 college students recruited from the psychology subject pool at the University of California, Santa Barbara. All participants indicated that they lacked experience in automobile mechanics. There were 20 students in the narration group (Group N), 17 in the narration plus mechanical sounds group (Group NS), 18 in the narration plus music group (Group NM), and 20 in the narration plus mechanical sounds and music group (Group NSM). The median age was 18, and the overall percentage of women was 79%. Neither age nor gender differed significantly among the groups,  $F(3, 71) = 2.17, MSE = 2.98, p = .10$ , and  $\chi^2(3, N = 75) = 0.41, p = .90$  for age and gender, respectively.

*Materials and apparatus.* The paper-and-pencil materials consisted of a participant questionnaire, retention test, transfer test, and matching test like those used by Mayer and Moreno (1998). The participant questionnaire asked participants to indicate their SAT scores, gender, and mechanical experience. Mechanical experience was assessed using a 6-item checklist and a 5-item self-rating scale. The printed instructions for the checklist were to "Please place a check mark next to the things you have done." The list consisted of the following six items: "I have a driver's license," "I have put air into a tire on a car," "I have changed a tire on a car," "I have

changed the oil in a car," "I have changed spark plugs on a car," and "I have replaced brake shoes on a car." The printed instructions for the self-rating were to "Please place a check mark indicating your knowledge of car mechanics and repair." The scale consisted of five items that ranged from *very little* to *very much*.

The retention test had the following instructions typed at the top of the sheet: "Please write down an explanation of how a car's braking system works. Pretend that you are writing an encyclopedia entry for people who are not familiar with brakes." The transfer test consisted of the following four questions, each typed at the top of a separate sheet: "What could be done to make brakes more reliable, that is, to make sure they would not fail?" "What could be done to make brakes more effective, that is, to reduce the distance needed to bring a car to a stop?" "Suppose you press on the brake pedal in your car but the brakes don't work. What could have gone wrong?" and "What happens when you pump the brakes (i.e., press the pedal and release the pedal repeatedly and rapidly)?" The matching test presented a frame from the animation along with the following instructions typed at the top of the sheet:

Circle the brake pedal and write B next to it. Circle the piston in the master cylinder and write P next to it. Circle part of the brake line and write L next to it. Circle the smaller piston in the wheel cylinder and write W next to it. Circle part of the brake shoes and write S next to it. Circle part of the brake drum and write D next to it.

The computer-based materials consisted of four computer programs for multimedia presentations explaining how a car's braking system works. All programs contained the same 45-s animation depicting a foot pressing on a brake pedal, a piston moving forward in a master cylinder, brake fluid being compressed in the brake line, pistons moving forward in the wheel cylinders, pistons pressing against brake shoes, brake shoes pressing against the brake drum, and the wheel slowing down to a stop. All programs also contained the following 76-word narration describing each of the major events in words spoken at a slow rate by a male voice:

When the driver steps on the car's brake pedal, a piston moves forward inside the master cylinder. The piston forces fluid out of the master cylinder, and through the brake lines to the wheel cylinders. In the wheel cylinders the increase in fluid pressure makes a smaller set of pistons move. These smaller pistons activate the brake shoes. When the brake shoes press against the drum, both the drum and the wheel slow down or stop.

The animation and the narration were presented concurrently. The N version of the program contained only the animation and the narration and was identical to the AN treatment in Mayer and Moreno (1998). The NS version was the same as the N version except that the sound tract also contained mechanical sounds corresponding to the movement of the pistons and the grinding of the brake shoes against the brake pad. The NM version was the same as the N version except that the sound tract also contained an instrumental music loop that played in the background as for the NM group in Experiment 1. The NSM version was identical to the N version except that it contained both mechanical sounds and background music on the sound tract. As in Experiment 1, the narration for all versions was clearly audible. In the NS, NM, and NSM versions, where sounds, music, or both were presented in addition to the narration, neither the music nor the sounds masked or made the narrative perceptually less discernible. The multimedia presentations were developed using Director 5 and Soundedit 16 (Macromedia, 1997).

The apparatus consisted of five Macintosh computer systems with 14-inch color monitors and Sony headphones.

**Procedure.** Participants were tested in groups of 1 to 5 per session, were seated in front of individual computer stations, and were randomly assigned to a treatment group.

First, participants completed the participant questionnaire at their own rates. Second, the experimenter presented oral instructions stating that the computer would explain how a car's braking system works and that afterward the experimenter would have some questions for the participant to answer. Participants were told to put on their headphones and press the space bar to start the presentation. Third, the computer presented the multimedia program corresponding to the participant's treatment group (i.e., version N, NS, NM, or NSM). Fourth, when the multimedia presentation was finished for all participants, the experimenter presented oral instructions for the test, stating that the participant would receive a series of question sheets one at a time and should keep working on each sheet until told to stop. Fifth, the experimenter distributed the retention test along with oral instructions to write down an explanation of how a car's braking system works. After 5 min, the experimenter collected the retention test. Then, the four transfer test sheets were distributed one at a time for 2.5 min each, and each sheet was collected before the next one was handed out. Participants were orally instructed to write as many solutions as possible. Finally, the matching test was passed out and collected after 2.5 min.

**Scoring.** A scorer who was not aware of the treatment group determined the experience score, retention score, transfer score, and matching score for each participant, using the same procedure as in Mayer and Moreno (1998). A second rater scored a randomly picked subset of 20% of the tests. Agreement between both scorers was 82% on the retention and cued-retention tests, 93% on the transfer tests, and 100% on the matching tests. We computed the mechanical experience score by adding all the check marks from the 6-item mechanical experience checklist plus the participant's self-rating score (ranging from 0 points for checking *very little* to 4 points for checking *very much*). Data were eliminated for participants scoring above 5 on this 10-point scale, and new participants were used in their places ( $n = 5$ ).

A retention score was computed by counting the number of major idea units (out of eight possible) that the participant produced on the retention. The participant received 1 point for each of the following idea units regardless of wording: (a) driver steps on brake pedal, (b) piston moves forward in master cylinder, (c) piston forces brake fluid out of master cylinder, (d) fluid pressure increases in wheel cylinder, (e) smaller pistons move, (f) smaller pistons activate brake shoes, (g) brake shoes press against drum, and (h) drum and wheel stop or slow down.

A transfer score was computed for each participant by tallying the number of acceptable answers across the four transfer problems. For example, acceptable answers for the first question about redesigning brakes for reliability included adding a backup system or adding a cooling system; acceptable answers for the second question about redesigning brakes for effectiveness included using more friction-sensitive brake shoes or reducing the distance between the brake shoe and the pad; acceptable answers for the third question about troubleshooting a faulty brake system included finding a hole in the brake line or the piston stuck in one position; and acceptable answers for the fourth question about pumping the brake pedal included reducing heat or preventing the pad from becoming worn in one spot. Questions were open ended, so participants could receive as many points per problem as correct answers they gave.

A matching score was determined by counting the number of correctly identified parts (out of six possible) on the matching test, with participants receiving 1 point for each correctly labeled part.

## Results and Discussion

Table 2 summarizes the mean scores (and standard deviations) for the four groups on each of the three tests in Experiment 2. For each of the three dependent measures (i.e., retention, transfer, and matching) the data were subjected to (a) a one-way analysis of variance with group as the between subjects factor, (b) supplemental Tukey tests (with alpha at .05), and (c) a two-way analysis of variance with the between subjects factors being presence or absence of mechanical sounds (NS and NSM versus N and NM, respectively) and presence or absence of music (NM and NSM versus N and NS, respectively).

**Coherence effect on verbal recall.** According to the arousal hypothesis, adding entertaining auditory adjuncts should increase the amount of material that students remember from the narration. In contrast, according to the coherence hypothesis, adding sufficient amounts of extraneous auditory material to a narrated animation should reduce the amount of material that students remember from the narration. The first columns of Table 2 present the mean scores and standard deviations of the four groups on the retention test. A two-way ANOVA revealed that students remembered significantly less verbal material when music had been presented ( $M = 3.57, SD = 1.68$ ) than when no music had been presented ( $M = 2.76, SD = 1.48$ ),  $F(1, 71) = 4.38, MSE = 2.41, p < .05$ ; and when mechanical sounds had been presented ( $M = 3.55, SD = 1.50$ ) than when no mechanical sounds had been presented ( $M = 2.76, SD = 1.66$ ),  $F(1, 71) = 4.31, MSE = 2.41, p < .05$ ; however, there was no significant interaction between music and sounds,  $F(1, 71) = 0.06, MSE = 2.41, p = .81$ . On the basis of supplemental Newman-Keuls tests (with  $\alpha$  at .05), we concluded that students in Group N scored significantly higher than students in the NS, NM, and NSM groups, with students in Group NSM recalling significantly less verbal material than students in the rest of the groups.

Overall, these results are inconsistent with the arousal hypothesis and consistent with the coherence hypothesis as described in Experiment 1. In both Experiment 1 and Experiment 2, students in Group N remembered more from the narration than did students in Group NSM. Thus, there is

consistent evidence that adding a large amount of bells and whistles can be detrimental to learning. In both Experiment 1 and Experiment 2, adding music to a multimedia instructional message was detrimental to learning. However, unlike Experiment 1, adding sounds to a multimedia instructional message also was detrimental to learning.

**Coherence effect on problem-solving transfer.** According to the arousal hypothesis, adding entertaining auditory adjuncts should increase student attention overall and thereby result in increased performance on problem-solving transfer. In contrast, according to the coherence hypothesis, adding sufficient amounts of extraneous auditory material to a narrated animation should reduce problem-solving transfer. The second columns of Table 2 present the mean scores and standard deviations of the four groups on the problem-solving transfer test. A two-way ANOVA revealed that students produced significantly fewer problem solutions when music had been presented ( $M = 4.40, SD = 2.26$ ) than when no music had been presented ( $M = 3.37, SD = 2.07$ ),  $F(1, 71) = 4.09, MSE = 4.01, p < .05$ ; and when mechanical sounds had been presented ( $M = 3.24, SD = 2.16$ ) than when no mechanical sounds had been presented ( $M = 4.50, SD = 2.10$ ),  $F(1, 71) = 6.84, MSE = 4.01, p < .05$ ; in addition, there was a significant interaction between music and sounds,  $F(1, 71) = 7.61, MSE = 4.01, p < .01$ , in which Group N performed better than each of the other groups. On the basis of supplemental Newman-Keuls tests (with  $\alpha$  at .05), we concluded that Group N generated more problem solutions than each of the other groups. Overall, these results on the transfer test mimic those obtained for the retention test, thus providing additional support for the coherence hypothesis.

**Coherence effect on visual-verbal matching.** The mean matching scores (and standard deviations) for the four groups are listed in the third columns of Table 2. A one-way ANOVA revealed that the four groups did not differ significantly in their scores on the matching test,  $F(1, 71) = 0.25, MSE = 2.59, p = .86$ . Similarly, a two-way ANOVA revealed no significant main effect for music,  $F(1, 71) = 0.52, MSE = 2.59, p = .47$ ; no significant main effect for mechanical sounds,  $F(1, 71) = 0.01, MSE = 2.59, p = .93$ ; and no significant interaction between music and mechanical sounds,  $F(1, 71) = 0.57, MSE = 2.59, p = .64$ . As in Experiment 1, the matching test apparently was not a sensitive enough measure of students' learning processes.

Table 2  
Mean Scores and Standard Deviations for Four Groups  
on Three Tests—Experiment 2

Group	Retention test		Transfer test		Matching test	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
N	3.95 <sub>a</sub>	1.57	5.55 <sub>a</sub>	1.88	4.15 <sub>a</sub>	1.53
NS	3.12 <sub>a,b</sub>	1.73	3.06 <sub>b</sub>	1.92	4.29 <sub>a</sub>	1.61
NM	3.11 <sub>a,b</sub>	1.32	3.33 <sub>b</sub>	1.72	4.06 <sub>a</sub>	1.51
NSM	2.45 <sub>b</sub>	1.57	3.40 <sub>b</sub>	2.39	3.85 <sub>a</sub>	1.76

Note. Column means not sharing subscripts significantly differ from one another ( $p < .05$ ). Scores ranged from 0 to 8 for the retention test, from 0 to 8 for the transfer test, and from 0 to 6 for the matching test. N = narration; NS = narration plus environmental sounds; NM = narration plus music; NSM = narration plus environmental sounds and music.

## General Discussion

The major result is that adding sufficient amounts of entertaining but irrelevant auditory material to a multimedia instructional message was detrimental to student learning. This result was obtained across two different dependent measures (i.e., retention and transfer) within two different multimedia learning messages (i.e., an explanation of lighting and an explanation of hydraulic braking systems).

In both experiments, students in Group N generated more problem solutions than did students in Group NSM. Also, in both experiments, adding music to the lesson was detrimental to retention and problem-solving transfer. However,

unlike Experiment 1, in Experiment 2 adding sound effects to the lesson was also detrimental to retention and problem-solving transfer. The different results for sound effects between experiments may be related to differences in the coordination of the sounds and their duration. For the groups that were presented with sounds in Experiment 1, the animation contained a sequence of seven different natural sounds played only once during the whole presentation for a total duration of 180 s. Each sound was coordinated in one-to-one correspondence with a particular event or idea unit in the animation. On the other hand, for the groups that were presented with sounds in Experiment 2, the animation contained a sequence of two mechanical sounds that lasted 9 s, and the same sequence was repeated several times at different points throughout the animation. The repetitious mechanical sounds that were paired with different events throughout the animation in Experiment 2 may have been too intrusive, arbitrary, and ambiguous to associate with the other materials in the lesson.

The different effects for the sounds across our experiments point to the necessity of conducting more studies where the coordination of the sounds is directly manipulated. Our findings suggest that in multimedia lessons, the more relevant and integrated sounds are, the more they will help students' understanding of the materials.

The reported results complement analogous findings concerning the detrimental consequences of extraneous words and pictures on students' learning from print-based media (Harp & Mayer, 1997, 1998; Mayer et al., 1996) and demonstrate a new kind of *coherence effect* involving auditory adjuncts with computer-based media: In multimedia learning environments, students achieve better transfer and retention when extraneous sounds are excluded rather than included. This extension of the coherence effect to auditory adjuncts in computer-based environments has important theoretical and practical implications.

### *Theoretical Implications*

On the theoretical side, these results are consistent with coherence theory—the idea that auditory adjuncts can overload auditory working memory—and inconsistent with arousal theory—the idea that auditory adjuncts motivate the learner to pay more attention to all incoming materials. These findings are consistent with aspects of limited resource models of working memory (Baddeley, 1992), cognitive load theory (Chandler & Sweller, 1991), and dual-code theory (Paivio, 1986). Our interpretation is that auditory adjuncts do their damage by limiting the amount of relevant verbal material the learner selects for processing in working memory and by reducing the learner's resources for building connections between verbal and visual representations of the to-be-learned system. According to Paivio (1986), sounds and music may constitute two nonverbal subsystems separate from the verbal or narration system within the auditory working memory channel. Although these subsystems are capable of functioning independently, they might interfere with each other in the auditory processing if they cannot be

easily associated with each other. We believe that this is the case when music or intrusive sounds (such as the mechanical sounds of Experiment 2) are added to an otherwise comprehensible auditory message.

On the surface, seductive details and auditory adjuncts seem similar—students' retention and transfer performance is diminished when they are added to a lesson. However, the underlying cognitive mechanisms are quite different: Whereas seductive details seem to prime inappropriate schemas into which the incoming information is assimilated (Harp & Mayer, 1998), auditory adjuncts seem to overload auditory working memory.

On the surface, the coherence effect we obtained in this study seems different from the split-attention effect we obtained in a previous study (Mayer & Moreno, 1998)—in one case we hurt students' retention and transfer by adding irrelevant sounds, and in the other case we hurt students' retention and transfer by presenting words as on-screen text rather than narration. However, the underlying theoretical explanation is similar—the coherence effect was attributed to overloading auditory working memory, whereas the split-attention effect was attributed to overloading visual working memory.

### *Practical Implications*

On the practical side, the present studies are directly related to issues in display design and display formatting for multimedia instructional messages. There is a growing need for empirically based principles in the area of instructional design. An example of research in this direction is the work of Park and Hannafin (1993), who have listed a set of empirically based guidelines for the design of interactive multimedia. Two principles from their article read as follows: "Principle 7: Learning improves as the number of complementary stimuli used to represent learning content increases" (p. 72) and "Principle 9: Learning improves as competition for similar cognitive resources decreases and declines as competition for the same resources increases" (p. 74). How should the professional in instructional design apply these principles when deciding about including sounds and music in a multimedia lesson?

On one hand, and according to the first principle, the addition of complementary auditory inputs should not harm but rather work as an extra symbolic system to enhance learning. On the other hand, and according to the second principle, cognitive overload can result from overtaxing the limited cognitive resources, therefore hindering learning. The present studies help to solve the tension between the two seemingly opposing principles by adding what we call the coherence principle: When presenting a multimedia explanation, only include complementary stimuli that are relevant to the content of the lesson. The most straightforward practical implication is that instructional software designers should carefully limit the amount of auditory material in multimedia lessons rather than add auditory materials for reasons of appeal or entertainment.

## References

- Alwit, L. F., Anderson, D. R., Lorch, E. P., & Levin, S. R. (1980). Preschool children's visual attention to attributes of television. *Human Communication Research, 7*, 52-67.
- Anderson, D., & Levin, S. (1976). Young children's attention to "Sesame Street." *Child Development, 47*, 806-811.
- Anderson, D., & Lorch, E. P. (1983). Looking at television: Action or reaction? In J. Bryant & D. R. Anderson (Eds.), *Children's understanding of television: Research on attention and comprehension* (pp. 1-33). New York: Academic Press.
- Baddeley, A. (1992). Working memory. *Science, 255*, 556-559.
- Calvert, S. L., & Gersh, T. L. (1987). The selective use of sound effects and visual inserts for children's comprehension of television content. *Journal of Applied Developmental Psychology, 8*, 363-375.
- Calvert, S. L., Huston, A., Watkins, B., & Wright, J. (1982). The relation between selective attention to television forms and children's comprehension of content. *Child Development, 53*, 601-610.
- Calvert, S. L., & Scott, M. C. (1989). Sound effects for children's temporal integration of fast-paced television content. *Journal of Broadcasting and Electronic Media, 33*, 233-246.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.
- Collins, W. A. (1982). Cognitive processing in television viewing. In *Television and behavior: Ten years of scientific progress and implications for the eighties: Volume II technical reviews*. U. S. Department of Health and Human Services, National Institute of Mental Health, DHHS Publication No. (ADM) 82-1196. Washington, DC: U. S. Government Printing Office.
- Dewey, J. (1913). *Interest and effort in education*. Cambridge, MA: Houghton Mifflin.
- Garner, R., Gillingham, M., & White, C. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*, 41-57.
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology, 89*, 92-102.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*, 414-434.
- Huston, A., & Wright, J. (1983). Children's processing of television: The informative functions of formal features. In J. Bryant & D. R. Anderson (Eds.), *Children's understanding of television: Research on attention and comprehension* (pp. 35-68). New York: Academic Press.
- Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*, 179-211.
- Macromedia. (1997). *Director 5.0* [Computer program]. San Francisco: Author.
- Mann, R. (1979). *The effect of music and sound effects on the listening comprehension of fourth grade students*. Unpublished doctoral dissertation, North Texas State University, Denton.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1-19.
- Mayer, R. E. (1999). Instructional technology. In F. Durso (Ed.), *Handbook of applied cognition* (pp. 551-570). Chichester, England: Wiley.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-dual coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.
- Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*, 444-452.
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*, 64-73.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389-401.
- Mayer, R. E., & Wittrock, M. C. (1996). Problem-solving transfer. In R. Calfee & R. Berliner (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.
- Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia design: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.
- Paivio, A. (1986). *Mental representation: A dual coding approach*. Oxford, England: Oxford University Press.
- Park, I., & Hannafin, M. J. (1993). Empirically-based guidelines for the design of interactive multimedia. *Educational Technology Research and Development, 41*, 63-85.
- Renninger, K. A., Hidi, S., & Krapp, A. (Eds.). (1992). *The role of interest in learning and development*. Hillsdale, NJ: Erlbaum.
- Wagley, M. (1978). *The effect of music on affective and cognitive development of sound-symbol recognition among preschool children*. Unpublished doctoral dissertation, Texas Women's University, Denton.
- Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology, 82*, 616-622.

Received February 5, 1999

Revision received May 17, 1999

Accepted May 17, 1999 ■