35

AUDITORY INSTRUCTION

Ann E. Barron
University of South Florida

35.1 INTRODUCTION

As noted by Winn (1993), "Human speech is the most powerful and expressive medium the designer has available for use in instructional messages" (p. 117). Speech is naturally expressive, and by varying the qualities of loudness, pitch, pace, and tone, designers can use audio to motivate and inform students. Three primary audio elements are used in educational technology—music, speech, and sound effects (Beccue, 2001; Kerr, 1999). Through these elements, audio can deliver information, direct attention, convey emotions, and provide feedback. In fact, "Audio is so integral a part of multimedia that most users would recognize its importance only by its absence" (Lehrman & Tully, 1991).

Although audio in an important instructional tool, it has not been studied as much as other media (Bishop & Cates, 2001; Jaspers, 1995; Thompson, Simonson, & Hargrave, 1996; Wilkinson, 1980). This chapter focuses on published research studies related to audio. Beginning with evaluation research and media comparison studies, the focus is then shifted to auditory memory and multichannel communication. A multimedia section focuses on design guidelines for incorporating audio in multimedia instruction. Finally, an overview of time-compressed speech and a summary are provided.

The chapter is presented in seven parts.

1. Introduction
2. History of Audio Technologies and Research
3. Auditory Memory
4. Multichannel Research Related to Audio
5. Audio in Interactive Multimedia
6. Time-Compressed Speech
7. Recommendations

35.2 SCOPE OF THE AUDITORY INSTRUCTION REVIEW

Because of the enormous amount of information related to sound and audio, this literature review is limited in its scope. The focus is on the use of verbal audio (speech); it does not address the instructional applications of sound effects or music, nor does it include the applications of audio that are specifically designed for students with disabilities.

35.3 HISTORY OF AUDIO TECHNOLOGIES AND RESEARCH

35.3.1 Introduction

Sound, as a physical form, can be described as "vibrations that set into motion longitudinal waves of compression and rarefaction propagated through molecular structures such as gas, liquids, and solids" (Alten, 2002, p. 14). For example, if you hit a drum or a tuning fork, it will vibrate, causing variations in the air pressure around it to reach your ears (Barron, Orwig, Ivers, & Lilavois, 2002). From a physiological aspect, sound is generally described as a phenomenon that is capable of being detected by the organs of hearing—generally between 20 and 20,000 Hz (American Heritage Dictionary, 1982).

Communication and instruction through sound (in the form of speech) have been used throughout time. Especially in the days prior to the printing press, oral instruction was a primary means of education and communication. After the printing press was invented in 1455, it became possible to reach a large audience by disseminating books. From then until the present, university-level instruction has relied primarily on text (Ives, 1992).

Leon Scott is credited with devising the first method of "recording" sound waves, in 1857 (Purcell & Hemphill, 1997). His device, called a phonoautograph, used a large horn, bristle, membrane, and lever to etch a pattern of a sound's frequency. This provided a "picture" of the sound, but it could not be played back. Then, in the late 1800s, the first of many technologies was invented that would allow the storage, duplication, and delivery of audio waves (the phonograph). Audio, in the forms of language, music, nature, and environmental sounds, is now an integral part of our lives and our education—it is perhaps second only to the written word as the distribution medium for instruction (Unwin & McAleese, 1988).

35.3.2 Timeline of Audio Technologies

With technological developments in the late nineteenth century and throughout the twentieth century, additional tools were added to teachers' arsenals of instructional devices. These audio technologies enabled educators to communicate verbally with students located in remote places, bring the voices and music of experts into the classroom, and record and distribute students' audio projects. A time line of approximate dates related to the major technologies is presented in Fig. 35.1 (Access Science, 2003, Fang & Ross,1966; Motion-Picture Technology, 2003).

35.3.3 Audio in Instruction

As indicated in Fig. 35.1, a variety of audio technologies has been developed and has transitioned from analog to digital formats. Although the adoption time frame and degree of penetration for each technology varied (depending on the price of the hardware, the availability of educational software and other factors), they all had an impact on education. Through these technologies, educators have been able to use audio in both synchronous and asynchronous modes.

One of the oldest synchronous technologies (telephone) is still extremely valuable in educational settings. Whether talking to a parent or interacting with distant students, synchronous communications are essential. With the advent of digital technologies, synchronous communications have evolved to include audio conferencing via the Internet, satellite, leased data lines, and wireless connections.

Asynchronous audio technologies began with the phonograph and radio and evolved into audiotape and television. After digital audio formats emerged, audio disc, computers, DVD, HDTV, and other technologies replaced the analog formats.

As the synchronous and asynchronous audio technologies were developed, evaluation studies took place to ascertain their ability to deliver effective instruction. This section provides a review of some of the major studies associated with the various technologies, as well as a synopsis of the media comparison debate.

35.3.4 Telephone

The initial applications of teaching via telephone began in the 1930s and 1940s. Since the 1970s, the trend toward distance education has resulted in expanded use of the telephone in teaching (Olgren, 1977; Olgren & Parker, 1983). "Teaching by telephone may be one of the original distance education media, but it still has an important place in the array of media available for distance learning" (Olgren, 1997, p. 59).

In 1958, Cutler, McKeachie, and McNeil conducted a study concerning the relative effectiveness of teaching via the telephone. Two matched groups of 10 participants were selected. One group was taught elementary psychology in the traditional manner; the other, by telephone alone. No text was used, but a list of suggested readings was furnished. The telephone group was connected to a system in which all participants could speak to each other. Gains in knowledge were found in both groups, and there was no significant difference in the gain between

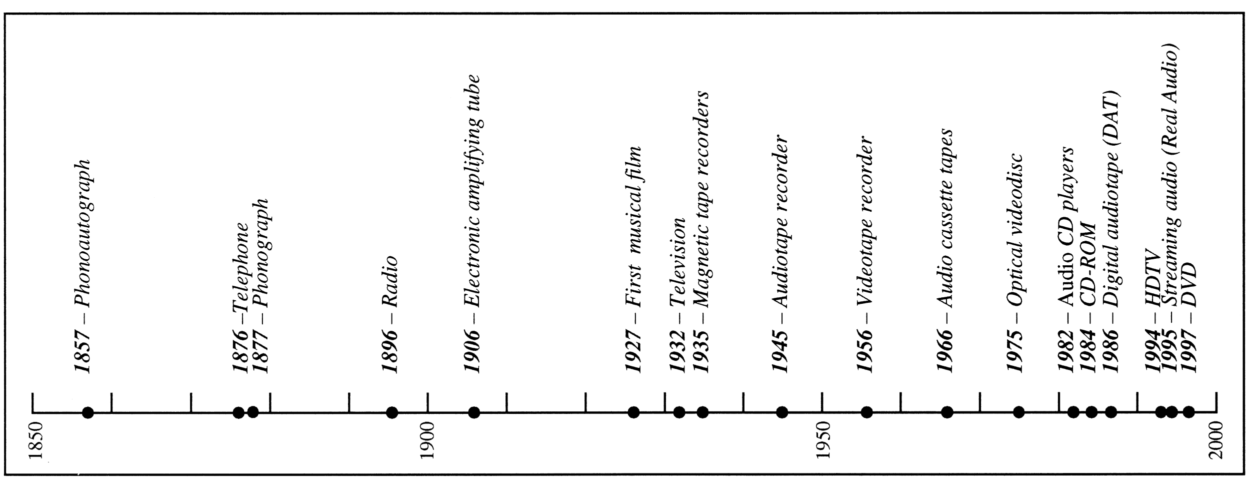

FIGURE 35.1. Approximate dates associated with audio technologies.

the two classes. Although there was evidence of a novelty effect, the method appeared practical. Rao (1977) summarized the limited research on telephone teaching and concluded, "... The research done on the effectiveness of teleteaching indicates that teleteaching is an economical and effective tool" (p. 483).

35.3.5 Phonographic Recordings

Rulon (1943a), using phonographic recordings, conducted an experiment to compare the amount of information gained by students who listened and those who studied the same material in printed form. Time was equalized. A total of 418 students listened to the recordings; 426 students studied the printed material. All students involved took a pretest, a posttest, and a test 1 week after the experiment was completed. Separate *t* tests were used to compare means of the pretest and posttest and also the delayed test. According to Rulon, the study of the printed material was superior to the method employing the recordings. However, a comparison in tests taken after a week showed little difference in methods employed. From this result, he concluded that recordings make more of a lasting impression than printed materials.

Rulon (1943b) later conducted a similar experiment to compare the amount of information gained by students using phonographic recordings with the amount gained by students who studied a unit incorporating the same material presented in a textbook. This experiment probably was closer to an actual classroom situation, although in this case, the textbook was prepared using the recordings as a primary source. Instructional methods using the textbook were not controlled. Testing procedures were similar to the previous study. The results, also using the *t* test, showed that phonographic recordings failed to show any superior effectiveness in teaching the "informational" aspects of the lesson.

In a third study on the effects of phonographic recordings, Rulon (1943c) investigated the motivational values of recordings. Using the same recordings and textbooks prepared for the earlier experiment, two groups of students were given access to supplementary reading materials after one group had heard the recordings and the other had read the material. Motivation was measured by which group used more supplementary reading materials. A total of 193 students used the recording, and 187 used the textbook presentation. Rulon's study showed no difference between the groups in terms of motivation to use supplementary reading material.

35.3.6 Loudspeakers

Educational use of the loudspeaker also attracted the attention of researchers. In 1937, Loder compared the retention of factual materials presented over a loudspeaker system with those presented directly by a speaker. Two groups totaling 449 students were rotated in the experiment. A pretest, a test immediately after the lesson, a test 1 day later, and another test 20 days later were given. One group saw the speaker, and the other group heard him from another room. The direct group performed better, but the means were not significantly different.

Kramer and Lewis (1951) investigated whether there was a difference in memory and comprehension between two groups of students in which one group sees and hears the speaker and another group only hears him. There were 128 students in the visual group and 120 in the audio group. Both groups were located in the same lecture room, separated by a large, heavy curtain. Loudspeakers were used, and the lecture was given simultaneously to both groups. After the lecture, both groups took the same test. Both groups had been told that grades would not be counted. Kramer and Lewis reported that the mean of the visual group was higher than that of the audio group and that the visual group had a wider range. They concluded that the speaker's visible action somehow contributed to the ability to understand and remember the ideas in the lecture.

35.3.7 Radio

Educational radio began at the University of Iowa in 1911 (Wolcott, 1993). The University of Wisconsin followed in 1919, and the Ohio School of the Air was established in 1929. Other "schools of the air" were established at many institutions, but the Wisconsin effort appears to have been the most successful. Various attempts were made to evaluate the effectiveness of radio broadcasts, and Saettler (1990) reports on two important ones. The first, the Ohio Evaluation of School Broadcasts Project, began in 1937 and ended in 1943. The project's goals were to analyze the educational values of radio and to study the social and psychological effects of radio on children. "Without question,the evaluation of school broadcasts project made a significant contribution to educational broadcasting and educational technology. It provided valuable factual evidence and produced helpful aids to the educational broadcaster in the planning and effective use of educational broadcasts" (Saettler, 1990, p. 242).

The second evaluation study, the Wisconsin Research Project, compared radio instruction with conventional instruction in six subject areas (grades 5–12) from 1937 to 1939. The differences were not significant. "The comparisons consistently favored the radio groups only in the field of music, and even there the differences were not large enough to be statistically significant" (Saettler, 1990, p. 242). Although there are not many other studies, those that exist (e.g., Constantine, 1964; Cook, 1964; NHK, 1956) typically showed that radio students performed at least as well as live audiences.

Bates (1983) reported that the British Open University's experience with radio instruction showed that the broadcasts tended to help the weaker students more than the successful ones. It must be remembered that Open University courses were taught primarily through correspondence texts and that the radio broadcasts were intended to supplement the texts. It is not surprising that students who found the texts difficult would welcome the added help afforded by the radio.

Radio is no longer widely used in education in the wealthier countries, but Wolcott (1993) reports that interactive radio is still used where long distances are involved, as in Alaska and Australia. In the poorer nations, radio is used widely because it is cost effective compared to other more sophisticated technologies. Radio is not expensive to produce or receive and can cover long distances using the AM band. The Agency for International Development (no date) reports that from 1974 to 1990 radio instruction programs existed in Nicaragua, Thailand, Kenya, Nepal, the Dominican Republic, Papua New Guinea, Honduras, Bolivia, Lesotho, Costa Rica, Ecuador, Belize, Swaziland, and Guatemala. Reported results were encouraging. For example, in Bolivia effect sizes as large as 0.91 were reported for radio math compared to traditional math. By dividing effect size by cost per pupil, a measure of cost effectiveness was obtained. According to data from the Agency for International Development, interactive radio was generally much more cost effective than textbooks or teacher training. In a related case study of radio-assisted community basic education (Eshgh, Hoxeng, Provenzano, & Casals, 1988) in the Dominican Republic, gains in both math and reading as a result of radio instruction were reported.

Although the history of the use of radio in education is long, there is not a plethora of empirical data concerning its effectiveness. Saettler (1990) gave an institutional history of early educational radio in the United States but did not mention more than a few empirical studies. After World War II, interest in instructional radio declined, so the situation did not improve. As with other media, radio instruction has been found to be at least as effective as conventional instruction. In certain circumstances, such as when conventional instruction is inadequate, the radio can be a cause of improved learning. Given the cost and reach of radio, it appears to be a viable medium in places where other media are too expensive or unavailable. A 1999 ERIC Digest provides information on applications and activities for using radios in the classroom (Ninno, 1999).

35.3.8 Film and Television

The first "talking picture," *The Jazz Singer,* by Warner Brothers, was produced in 1927 (Motion-Picture Technology, 2003). Since then, audio has been an important component of film and television production. When sound films reached academia in the early 1930s, a series of research studies examined their effectiveness. Hoban and Van Ormer (1970) summarized the research from 1918 through 1950 and concluded that films can be equivalent to good instructors and that using films in the classroom can reduce instructional time. In 1966, Campeau published a review of the literature that reported no significant differences between traditional instruction and motion pictures.

Television cameras were invented in 1923, and the first TV broadcast in the United States took place in 1930. By the 1950s, televisions were appearing in schools and instructional settings. In separate literature reviews, both Stickell (1963) and Chu and Schramm (1968) found no significant differences in the majority of the studies that compared face-to-face instruction and television.

35.3.9 Audiotape/Disc

The ability to record audio or to purchase professionally developed audio instruction has had a substantial impact on education and entertainment. Over the years, storage options have evolved from reel-to-reel, cartridges, and cassette to compact disc. These formats have been used to deliver instruction and to provide audio feedback. Because playback units are inexpensive and are readily available in homes and cars, these media offer tremendous versatility for classroom and remote instruction.

Gibson (1958, 1959, 1960) reported a 3-year experimental comparison of a tape-teaching program with conventional instruction at an Omaha, Nebraska, junior high school. Two areas were chosen to study the effectiveness of tape recording: spelling and conversational Spanish. Oral and written tests were used. Results included the assertion that tape instruction was superior to conventional instruction when the criterion was the number of words correctly spelled. Both methods were equally effective for recognition of words misspelled. Spanish classes taught by a non-Spanish teacher using Spanish tapes and classes taught by a Spanish teacher were similar in achievement scores. The following conclusions were made: (a) Tape recording is an effective method for teaching conversational Spanish to seventh graders; (b) regular classroom teachers can effectively teach conversational Spanish by means of tape prepared by Spanish specialists; (c) students can learn to spell as effectively with a tape as with conventional classroom procedures; (d) with proper orientation, large groups can be taught spelling effectively; and (e) teaching with tapes produced no adverse effect on attitudes toward the subject.

Popham (1962) studied the effectiveness of tape-recorded lectures in teaching a college-level education class. Thirty-six students were divided into 18 matched pairs. Chi-square analyses revealed no significant difference among variables. One group was a conventional lecture discussion; the other group was taught by tape-recorded lectures with student-led discussions. This experiment continued over one semester. Pretests in achievement and a test to measure student opinion were given. Both tests were repeated at the completion of the course. Popham reported that both groups had increased performance on the achievement tests; there was no significant difference between them. There was no significant difference on reactions to the courses; however, the opinions of the tape-lecture sections were generally favorable toward the technique.

In a similar study, Menne, Klingenschmidt, and Nord (1969) provided taped lectures, tape recorders, and printed notes to 209 college students. Another 408 students attended regular lectures. Overall, there was no significant difference, but students in the lowest quartile showed an advantage in the tape condition. Also, the dropout rate was lower for the students using tape.

There are few cases of the large-scale, systematic application of audio technology to instruction. One interesting and extensive implementation of audio for instructional purposes is Postlethwaite's "audiotutorial instruction" (ATI). This approach, which has been widely reported (Button, 1991; Postlethwaite, 1970, 1972, 1978, 1980; Svoboda, 1978), is more a complete instructional system than just an application of audio; however, its long history and wide application make it an important source of information.

ATI began almost by accident in 1961 at Purdue University (Postlethwaite, Novak, & Murray, 1972), when Postlethwaite was attempting to provide supplementary materials for weaker students in freshman botany. Simple lecture material was made available on a self-study basis through the audiovisual department. During the semester, these tapes evolved into programmed experiences that directed the students' attention to sections of the textbook, pictures, and diagrams, as well as live plants. Eventually experiments were added, and the entire week's study could be covered without attending any formal sessions. Student reaction was favorable, so in 1961–1962, an experimental section of 36 students was taught entirely by the audiotutorial method. Results on the conventional exam showed that the experimental group performed as well as the regular students. Interviews with the students led to the creation of a completely restructured course.

In designing the new course, Postlethwaite studiously avoided using words like lecture, recitation, and laboratory, which he felt connoted formality and passivity. The new course consisted of independent study sessions (with an audiotape as a tutor and guide to activities), general assembly sessions (for exams and group lectures), and small assembly sessions (which included integrated oral and written quiz sessions). Additional activities were sometimes included.

It appears that Postlethwaite was acutely aware that an audiotape-mediated course might take on the appearance of impersonality. To avoid this, several measures were taken. The tapes were made with an informal, conversational quality. The instructors in the learning center were apparently told to maintain a pleasant personal manner. The senior instructor spent 3 hours per week in the small quiz sessions, meeting with about 48 students. Another 3 hours were spent informally visiting the learning center. He also held a weekly coffee hour to which all students were invited. Finally, he held an open house (for 600 students!) at his home once a semester. Postlethwaite considered this emphasis on personal contact and a well-structured sequence of learning events to be the essential ingredients of ATI.

Joseph Novak (of the previous Postlethwaite group) developed elementary science lessons using ATI. In an extensive longitudinal investigation of the effects of ATI, Novak and Musunda (1991) reported a 12-year study of science concept learning. Twenty-eight of their best science concept ATI lessons were provided to 191 first- and second-grade children. Each lesson required 15 to 25 min to complete. As with the Postlethwaite materials, students were directed by the tape to interact with materials and pictures. Data collection evolved into Piagetian clinical interviews that were then translated into concept maps. Concept maps were then graded for valid and invalid notions. Instructed subjects showed significantly more valid concept understandings and fewer misconceptions. A significant interaction showed that instructed subjects, over the 12 years, had a greater tendency to increase their number of correct concepts and decrease the number of incorrect concepts. This study strongly supports the validity of ATI even under conditions where the instructors are not well trained. Valid concepts were learned from ATI and evidently "scaffolded" more learning throughout the children's 12 years of schooling.

The most comprehensive review of ATI is by Kulik, Kulik, and Cohen (1979). This metanalysis of 48 reports of ATI found a small but significant achievement effect for ATI over conventional instruction. However, ATI had little effect on course evaluations or withdrawal rates. Also, aptitude and achievement correlated highly, indicating that ATI did not have a leveling effect as might have been expected with such a self-directed, self-paced approach.

Several recent studies focus on the efficiency and effectiveness of providing audiotaped feedback for students, as opposed to written feedback. Pearce and Ackley (1995) conducted a review of the literature and summarized the following points (p. 32).

- All studies that reported on the measure found that audiotaping provided more and better feedback than written comments.
- Audiotaping typically required either less or the same amount of time and improved the context, leading to a better understanding of the feedback provided.
- Students liked or were motivated better by taped feedback, although the results on grades and quality were mixed.

In summary, ATI is the most complete and most well-documented method of auditory presentation. It has a general record of success, and although it is not noticeably superior to conventional approaches, it has many valuable applications.

35.3.10 Audioconferencing and Audiographics

In distance education, audio-based technologies are proliferating. Audioconferencing (both audio only and audioconferencing with images or data) is now possible through existing public telephone lines and other connections. Audiographic technologies such as the fax machine, streaming audio, and electronic whiteboards are becoming common. Freeman, Grimes, and Holliday (2000) conducted an experiment in which a 14-week course in statistics was delivered via four treatments: (a) audio-data collaboration,(b) satellite-delivered instructional television, (c) face-to-face instruction in the television studio, and (d) face-to-face instruction in a traditional classroom. Results suggest that "there is no difference in student learning performance between the hybrid audio-data collaboration and instructional television or face-to-face modes" (p. 112). Students in the audiographic group were more satisfied with the technical aspects of the medium than those in the instructional television group. For additional studies related to audioconferencing and other technologies, see *The No Significant Difference Phenomenon as Reported in 355 Research Reports, Summaries, and Papers* (Russell, 1999).

35.3.11 Computer and Web-Based Audio

At its inception, audio on computers was primarily a monotonous feature (such as a beep or a buzz) for inappropriate input by the computer user. As hardware and software improved, digital audio became a viable means of computer-based instruction (Barron & Kysilka, 1993). Several large-scale research studies that focused on computer-based education (with and without audio) have been conducted. In a general sense, these studies have found computer-based, interactive programs to be at least as effective as traditional instruction (Kulik, 1994; Mann, Shakeshaft, Becker, & Kottkamp, 1999; Schacter, 1999; Sivin-Kachala, 1998; Sivin-Kachala & Bialo, 2000). Research related specifically to digital audio in interactive formats is presented under *Audio in Interactive Multimedia* later in this chapter.

Although limited initially by storage space, audio in computer-based education flourished after CD-ROMs were developed. Currently, audio is a very common component of educational software programs. In the initial phases of the Internet and the Web, audio was again a limited commodity because of bandwidth constraints—although audio files could be downloaded and played, the transfer time was excessive. However, in the mid 1990s, technologies that allowed audio to be streamed over the Internet (rather than downloaded and played) were developed. Thanks to streaming technologies and advancements in compression algorithms, audio is becoming a common element in interactive, Web-based training.

Streaming audio is also used to deliver linear "lectures" via the Internet. A 1998 study by Ingebritsen and Flickinger found that achievement grades for an Internet section of a biology course were slightly higher (though not statistically significant) than the scores for a traditional section. Student attitudes toward the course were also very good. A similar study was conducted by Hurlburt (2001) with an introductory statistics course. Although a significant difference was not noted between course grades (of the traditional classroom and the "lectlet" group), the researcher noted that streaming audio effectively presented the course content and allowed a great deal more flexibility for students (who could record them and replay them at any time).

35.3.12 Reading vs. Listening; Print vs. Audio

In addition to the technology-focused studies, there are historical studies that focused on reading vs. listening, without involving a specific technology. Erickson and King (1917) performed one of the earliest studies of the effectiveness of the auditory medium. Four groups of students from third to ninth grade were chosen, and each group was divided in half. One-half received the lesson from silent reading; the other half was given similar material orally by the teacher. The following day, the order was reversed as to which half read and which listened. This procedure was followed two more times, for a total of four lessons. At the third- and fourth-grade level, the median score for the oral group was much higher than the median for the group that read. At the fifth- and sixth-grade levels, the results were inconclusive. At the seventh-, eighth-, and ninth-grade levels, the medians for the oral groups were much higher than for the groups that read the lesson. Needless to say, in these early studies experimental design was not what might be desired. Specific variables, such as teachers' skill and interest, were not taken into account, nor was the subject matter. Also, the only datum that was examined was the median score of the group.

Young (1936) tested comprehension via hearing and reading with 2,000 intermediate students from Iowa and Texas. Four modes were used to present the material: (a) The teacher read aloud to students, (b) the teacher read aloud and students read the selection silently, (c) students read the selection once; and (d) students read for the time allotted the teacher to read orally. At the end of each presentation, a comprehension test was given; a delayed test was given 1 month later. Young reported that the oral presentation was more effective than either of the silent readings, both immediately and after a month. Nevertheless, gains by all groups were poor. He also found that the children who did poorly in comprehending through reading also did poorly in comprehending through hearing.

Larsen and Feder (1940) asked whether psychological abilities differentiate between the processes involved in reading and those involved in listening comprehension. After hearing and reading selected materials that varied in difficulty, 151 students were given both reading and listening comprehension tests. On these tests, there was a superiority of performance on reading comprehension over listening comprehension. However, this superiority depended on the aptitude of the student. Students with lower aptitudes tended to show equal results in listening and reading. The higher-ability groups showed a superiority in the reading comprehension tests.

Taylor (1964) concurred that less competent students retained more from listening than from reading—possibly due to the additional cues provided by the speaker's expressions, gestures, and phrasing. He also noted that younger students prefer listening over reading, and older students prefer reading. "Above grade 7, there is a distinct preference for reading over listening in most learning situations, and better retention results from reading" (p. 17).

Reading and listening seem to demand the same underlying linguistic competence (Mosenthal, 1976–1977). However, comparisons of learning from audio and print have been contradictory. Nugent (1982) and Rohwer and Harris (1975) found no differences for children, and Nasser and McEwen (1976) found no differences for college students. However, in a series of studies involving children and adults, Furnham, Gunter, and others consistently found superiority of print over audio (Furnham & Gunter, 1989; Furnham, Gunter, & Green, 1990; Furnham, Proctor, & Gunter, 1988; Gunter & Furnham, 1986; Gunter, Furnham, & Leese, 1986).

Tripp (1994) tested the differences between audio and print in a direct comparison that attempted to hold other factors constant, including reviewability and novelty, by presenting the same text by computer through either the audio or the video (printed text) medium. As found by Furnham and Gunter, the students who read the text remembered significantly more correct semantic units than the students who heard the passage. Perhaps, as Travers (1970) remarked, "One cannot reasonably ask the general question whether the eye or the ear is more efficient for the transmission of information, since clearly some information is better transmitted by one sensory channel than by another" (p. 85).

35.3.13 Media Comparison Studies: The Great Debate

Over the years, numerous researchers have criticized comparison studies and called for a change in research focus. In 1961, Kumata concluded that, despite numerous research studies that focused on the effect of television (as compared to face-to-face instruction), it did not seem to matter whether or not television was present. He urged that new areas of research, beyond the "comparability" studies, be conceived and "that the emphasis should be shifted to the totality of the teaching-learning process" (Saettler, 1968, p. 340).

In 1969, Gordon noted that the results of measuring technologies (not media) by their impact on academic grades did not help to assess any differences between the media themselves. He stated, "These experiments have shown that the same kind of teaching operates more or less the same way with and without technological aids" (p. 118). A similar sentiment was expressed by Clark (1983, 1994) in his statement, "The best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition" (1983, p. 445). He defended his position by arguing that even where apparent differences are detected, they are rendered dubious by various forms of confounding factors. Some of the forms of confounding variables that Clark cited are the novelty effect, the "John Henry effect" (where participants in a control group exert extra effort because of the sense of competition), unequal instructional strategies, unequal opportunity to learn, and unequal quality of instructional design.

Kozma (1994) responded to Clark, noting that different media possess different attributes and capabilities. He suggested that instead of asking, "Do media influence learning?" we ask "In what ways can we use the capabilities of media to influence learning for particular students, tasks, and situations?" (p. 18). Jonassen, Campbell, and Davidson (1994) then entered the debate, suggesting that we shift from an instruction and media-centered focus to a learner-centered conception of learning. Others, such as Morrison (1994), Reiser (1994), Ross (1994), and Shrock (1994), have voiced opinions about the debate and the pertinent questions related to media.

Salomon (1994) emphasizes that media are not discrete, invariant entities and that "the examination of media, when made from a cognitive-psychological and educational point of view, should not adopt a holistic approach but rather focus on specific critical qualities of media" (p. 26). Salomon defines these critical qualities as the symbol systems—media's ways of structuring and presenting information. Koumi (1994) agreed, urging that we "try to develop and refine the criteria for deploying of media to best effect" (p. 47).

It should be noted, however, that even Clark and Salomon (1986) do not think that all studies that compare or evaluate media are pointless. "The shortcomings of overall media comparisons do not render such studies useless for all purposes. The evaluation of particular products, the weighting of a medium's overall cost effectiveness, and the close monitoring of a medium's employment in practice can all benefit from one or another kind of media comparison"(p. 466).

35.3.14 Summary

Numerous technologies, designed to store and deliver audio, have emerged in the last century. From the ability to speak to someone over a primitive telephone to streaming audio on the Web, education has been enhanced through audio communication. As new technologies appeared in educational settings, there was a natural interest in ascertaining their value to instruction. To that end, a number of studies were conducted, usually comparing instruction delivered through the new technology to instruction delivered in a classroom by a teacher. Most found no significant differences. For instructional designers and teachers, this was good news—their repertoire of instructional classroom tools increased, and they could effectively teach remote students via radio, audiotapes/discs, and computers.

However, it should be noted that many of the predictions of technology-delivered instruction did not bear fruit. As Cuban (1986) points out, radio, film, television, and computers met with only marginal success in schools, and instructional practices in classrooms have changed very little since 1900. The prophesies that films would revolutionize the educational system, that radio would bring the world to the classroom, and that television would relieve crowded classrooms were all doomed to fail (Cuban, 2001; Oppenheimer, 1997; Saettler, 1990). There are many reasons for these failures, few of which have to do with the capabilities of the technology. Technology can be used to "make us smart," but we need to focus on the learner and cognition (Norman, 1993). Hence, we now turn the focus of this chapter to the research related to the perception, attention, and memory of auditory information.

35.4 AUDITORY MEMORY

35.4.1 Introduction

A great deal of research has centered on the question of how we process, store, and retrieve auditory information. Although the quantity and depth of literature related to physiology and psychoacoustics are far beyond the scope of this chapter, an overview of the research that focuses on how verbal information presented via the audio modality is perceived, processed, and remembered is presented. For a more in-depth investigation, see *Thinking in Sound: The Cognitive Psychology of Human Audition* (McAdams & Bigand, 1993) and other publications.

35.4.2 Auditory Processing

Prior to the 1960s, most psychologists did not differentiate between short-term memory (STM) and long-term memory (LTM); instead, they proposed models with a unitary memory structure (Baddeley, 1998). Since then, numerous researchers have postulated two-part or three-part memory systems, and they have investigated potential differences in the processing of visual information and auditory information.

In 1967, Neisser wrote about early-stage, sensory memory systems and labeled them iconic (visual) and echoic (auditory). In 1968, Atkinson and Shiffrin proposed a model that has been nicknamed the "modal model." This model postulates three kinds of memory—sensory memory, STM, and LTM. The model depicts information from the environment initially entering through parallel sensory stores (visual, auditory, and haptic). Most of the information in the sensory store is lost; however, some may enter the limited capacity of the short-term store and may be placed in permanent, long-term store.

Baddeley (1998) also postulates a three-part memory structure, for both visual and auditory information. In particular, he states," Evidence suggests that auditory sensory memory can be split into at least three types, echoic memory extending over a matter of milliseconds, auditory short term memory extending up to perhaps 5 or 10 seconds, and auditory long-term memory" (p. 20).

35.4.3 Echoic Sensory Store

Echoic storage (also referred to as echoic memory) is similar to an echo that lingers after a sound has been heard (similar to the persistence of vision in iconic memory). As a part of the perceptual process, information in the sensory store either fades rapidly or is displaced by subsequent information (Penney, 1975).

Research focusing on echoic memory has investigated its duration and capacity. The duration of echoic memory is generally predicted to be about 200–350 ms (Baddeley, 1998; Cowan, 1984; Efron, 1970; Guttman & Julesz 1963; & Massaro & Loftus, 1996). A few studies, involving pure tones, have resulted in longer durations (Elliott, 1970; Erickson and Johnson 1964; Massaro, 1975).

As far as the capacity, Hawkins and Presson (1986) state that it is "impossible to assign a specific value to the amount of information the echoic memory system can hold" (pp. 26–36). They note that factors that can impact the capacity include the physical similarity of items, the difficulty of discrimination, temporal relations, and the subject's prior knowledge related to the stimuli.

35.4.4 Short-Term Auditory Memory

Whereas echoic sensory memory is passive, STM is active. In STM (which is equated with working memory by some researchers), some of the information in the sensory store is attended to and processed. Selective attention and divided attention are two phenomena that can affect the processing of auditory information.

Selective attention is the ability to focus on or attend to one particular sound or voice. For example, in a crowded cafeteria, the noise level is high, yet it is possible to attend to one conversation while mentally blocking the others. Early research conceptualized a filter, which selected information for further processing (Broadbent, 1958; Cherry, 1953). However, through laboratory experiments, it became clear that even the information that was not directly attended to could sometimes be recalled (Hawkins & Presson, 1986). Many researchers propose a process of resource allocation, wherein the limited resources of working memory can be allocated to various senses or streams of information, based on the task, the stimulus, or its relevance (Kahneman, 1973).

A similar concept is referred to as *divided attention* or *split attention*. In selective attention, the listener seeks to attend to one message or channel among several; in divided attention, "the task is to attend to several simultaneously active input channels or messages, responding to each as needed" (Hawkins & Presson, 1986). A classic example, attributed to Boring (1950), recounts nineteenth-century astronomers who were counting clock beats to measure the time required for a star to traverse a given distance. Unable to obtain agreement among several observers, they concluded that it was impossible to attend to both the clock beats and the star movements simultaneously (Hawkins & Presson, 1986). Although early researchers supported the "law of prior entry"—that attention is indivisible—many now believe that multiple sources of information can be processed in parallel to a certain point and then serially beyond that point (Hawkins & Presson, 1986). Olson (1989) noted that processing multiple streams of information simultaneously can degrade performance, depending on the processing loads and other factors. He concludes: "Concurrent streams of information are processed incompletely, although attention can be switched to sample from one stream to another, often quite rapidly, and inferential processes can attempt to fill in missing information" (p. 57). Experiments involving split attention and its impact on cognitive load and instructional design have recently been conducted by Kalyuga, Chandler, and Sweller (1999) and Mayer and Moreno (1998).

The debate continues over whether the audio and visual modalities share a common resource pool/processor or are allocated different resources/processors (Basil, 1992). Support for the theory of different processing systems derives from findings that audio information is almost always recalled at a higher rate than visual information (Penney, 1975). This short-term "modality effect" has been the focus of numerous research studies and theories. For example, a study by Gelder and Vroomen (1997) tested the immediate, serial recall of spoken words, sounds, written words, and pictures. They found that the recency (being able to remember the last few items in the list of eight terms) was highest for spoken words, intermediate for sounds, and negligible for printed words and pictures. Likewise, Rummer and Engelkamp (2001) found a clear modality effect for audio in tests that involved short-term recall of sentences presented in auditory and visual modes.

There are several theories that seek to explain this recency effect and the advantage of auditory information over visual information in STM (for tasks related to immediate, serial recall). When Murdock and Walker (1969) found evidence of a modality effect, beyond that which could be explained by echoic memory, they promoted the theory of separate STM stores for visual and audio information.

Penney (1989a) hypothesized that auditory and visual information are processed in separate streams, each with different properties and capabilities. She defines three properties of the auditory stream (p. 415):

1. There is a large capacity for storing sensory information. The sensory information persists for periods of up to a minute in the absence of subsequent auditory information.
2. The echo does not decay; rather, it is highly susceptible to interference from subsequent auditory input.

3. There is an automatic generation and maintenance of the acoustic code (referred to as A code).

Whereas speech is encoded in an acoustic and a phonological code, sounds might be encoded in only auditory code, and visual items are encoded in visual and phonological code. The acoustic code is hypothesized to be "rich and very durable relative to a visual sensory code" (p. 399).

Other researchers also suggest the presence of separate channels for processing visual and auditory information. Baddeley (1998) promotes a sensory-based process, which consists of a visual-spatial sketch pad (for visually presented material) and a phonological loop (for verbal material presented in auditory form and for information converted to auditory form through subvocalization). These two processors (referred to as slave systems) operate independently but are governed by a central executive, which regulates retrieving, processing, and storing the information. In this model, verbal input in auditory form (speech) has an advantage in STM because it automatically enters the phonological loop, whereas visual input requires recoding (which takes time and resources) to enter the phonological loop (Gelder & Vroomen, 1997).

Paivio's (1986) dual-coding theory is based primarily on presentation modes—the verbal channel processes spoken or written text, and the visual (i.e., nonverbal) channel processes pictorial materials and nonverbal sounds. In a 1994 study, Thompson and Paivio conducted an experiment with three stimulus lists—pictures, sounds, and picture-sound pairs. The "additive recall of dual-modality stimuli was found to be reliable and robust" (p. 390). They concluded that memory traces could be encoded in separate, modality-specific verbal and nonverbal components.

Mayer (2001) has proposed a sensory-modalities combination approach that recognizes both sensory input (audio vs. visual) and presentation mode (verbal vs. nonverbal). This model is illustrated in Fig. 35.2. Note that spoken words would enter sensory memory through the ears. If some or all of the words are attended to, they will then enter working memory as sounds and be coded as verbal information before being integrated with other information in short-term memory and with prior knowledge (visual and verbal) that was retrieved from long-term memory. In both Paivio's and Mayer's models, there is a semantic cross-connection between verbal and nonverbal memory where one code is accessible by the other.

Estimates of the duration and capacity of STM are generally 2–20 seconds for duration (Baddeley, 1998) and 5–9 chunks for capacity (Miller, 1956). There seems to be considerable disagreement, however, on the distinction between the echoic store and STM, the duration of STM, and the systems involved in processing auditory and visual information (Baddeley, 1998; Engle, 1996;Morra, 2000; Palmer, 2000; Penney, 1989a; Robinson & Molina, 2002). As noted by Crowder (1993), "Issues of perception, coding and immediate memory are notoriously difficult to disentangle" (in Gelder & Vroomen, 1997, p. 100).

35.4.5 Long-Term Auditory Memory

Experimental psychologists define LTM as "information that is stored sufficiently durably to be accessible over a period of anything more than a few seconds" (Baddeley, 1999, p. 16). The capacity and duration of LTM are difficult to assess; most assume that it is extremely large (perhaps unlimited) and has a very long duration (perhaps infinite). Exactly how the information is coded and stored in LTM is not completely understood. Psychologists have yet to agree if LTM is a unitary, dual, or multiple system and whether or not there are distinct audio and visual channels or stores (Baddeley, 1999).

According to Baddeley (1998), a lot of the information that is presented in the audio mode (such as speech) is probably stored in LTM in terms of its meaning, rather than its sound. However, some auditory information, such as music, voices, and sound effects, are assumed to be stored in a strictly audio form. Although there is general agreement about the modality effect in STM, results related to modalities in LTM memory have been more complex and contradictory. To test the possible modality effects in LTM, the subjects are generally presented with a series of words or numbers in either visual or audio format. An interference task is then introduced (to ensure that the memory being tested is long-term rather than short-term). Finally, the subjects are asked to recall as many of the items as possible.

Modality effects in LTM were found by Engle and Mobley (1976) and Glenberg, Eberhardt, and Belden (1987), giving rise

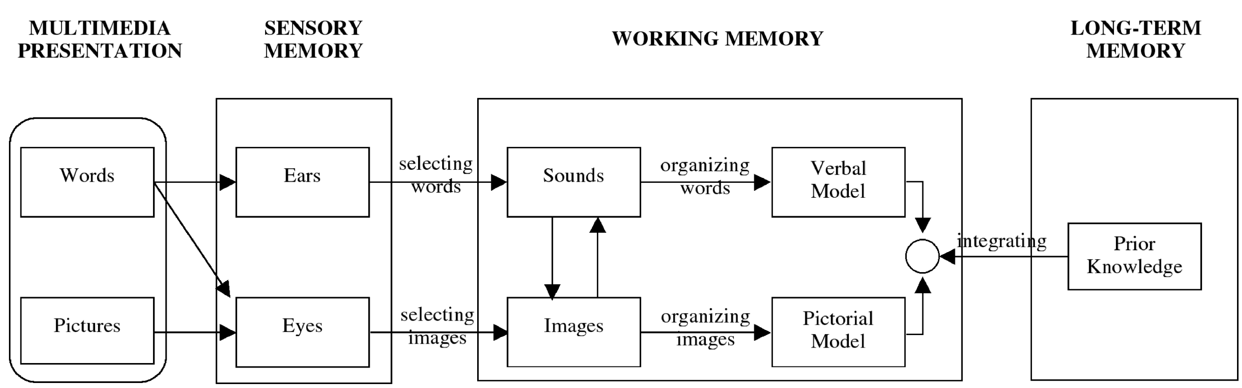

FIGURE 35.2. Mayer's cognition model.

to hypotheses such as auditory items being more distinctive than visual items and temporal information being encoded more precisely. However, in a 1989 study, Penney found a visual superiority for recall and recognition. She concluded that "where modality effects are found in long-term memory, the general effect will not consistently be an auditory superiority, but, rather, as interaction between task requirements and presentation modality" (1989b, pp. 468–469).

In 1994, Duis, Dean, and Derks also investigated the modality effects in STM and LTM. They found an auditory superiority in STM; however, they did not find evidence of a modality effect in LTM. They concluded that previous research designs may have favored audio and stated, "It seems possible that auditory and visual information are processed differently and that an understanding of these differences may lead to the conclusion that auditory superiority may, in fact, be due to the sequential presentation of stimuli that is common to so many experiments investigating the modality effect" (p. 6).

35.4.6 Summary

By examining the differences in people's ability to recall auditory vs. visual information, researchers have sought to learn more about how the human brain processes, stores, and retrieves information. Thus far, the majority of the studies have found a modality effect for STM. That is, audio information (generally a sequence of several words or numbers) is recalled better than the same information presented visually. This finding supports the theory that audio information is processed differently from (and perhaps independently of) visual data in STM. Similar studies investigating modality effects in LTM have produced conflicting results.

As experimental psychologists wrestle with models and theories that focus on differences in modalities, they may help to unravel issues that relate to the design and effectiveness of auditory instruction. As Estes (1989) noted, "Whether information is presented via films, lectures, readings, or other kinds of experiences, the process of adding the information to the stock of knowledge and skill in the mind of the student in usable form is complex and subject to many uncertainties" (p. 3).

35.5 MULTICHANNEL RESEARCH RELATED TO AUDIO

35.5.1 Introduction

Audio instruction is very rarely presented in isolation (Jaspers, 1995)—even with telephone and radio instruction, guidelines dictate having complementary visual materials. This dual approach derives from the fact that many educators believe that more learning will occur if instruction is presented through two sensory channels as opposed to one. This belief was responsible, in part, for the rush to use instructional television and film in the early and middle twentieth century. However, Travers (1970) stated that the assumption was based on "the most flimsy evidence" (p. 104). He disputed many of the early multichannel research studies based on the fact that they were conducted without solid research designs or statistical tests of significance.

Studies investigating channels of transmission have been ongoing for decades and continue in the present. The research typically centers on an auditory or visual channel and compares it with a combined audiovisual presentation. In some cases, the visual channel is further subdivided into a pictorial channel (nonverbal) and a print channel (visual–verbal). Although there are many theories related to multichannel communications, this section concentrates on five that are often cited in research related to auditory instruction.

1. Single-channel processing theory: There is only one channel to higher centers of the brain; therefore, dual-channel transmission can be equal to, but not greater than, the single-channel processing. In fact, if both stimuli arrive at the same time, information jamming may occur and cause the dual-channel effectiveness to be less than that of either of the single channels.
2. Cue summation theory: Dual-channel presentations result in more learning than single-channel presentations because the number of stimuli or cues is increased. "This theory suggests that when pictures are added to the message, the number of cues relevant to the message increases" (Lang, 1995, p. 88).
3. Limited-capacity information processing: If the combined amount of information of two single channels exceeds the upper limit of the central nervous system capacity, then interference may occur, causing equal or less gain to take place. If combined stimuli are less than the capacity, however, then the dual-channel presentation is more efficient and effective.
4. Dual-coding theory: Information is processed in either a verbal or a nonverbal form. By coding semantically connected information in both formats, recall and recognition can be enhanced because information from one code acts as a retrieval cue for information in the other mental store.
5. Cognitive load theory: Working memory is limited; therefore, designers should seek to structure the learning materials to minimize the requirements on STM (by using two modalities, reducing the complexity, organizing the material, etc.).

35.5.2 Single-Channel Processing Theory

Broadbent's (1958) theory of perception was constructed on the assumption that there is only a single channel to the brain's higher centers. By conducting various experiments, Broadbent found that when the limit of the central nervous system has been met, some information is stored in a holding area. Because his system generally handles only one message at a time, it is referred to as the *single-processing theory* or *single-channel theory*. According to this theory, if information (from one or more sources) reaches the brain at a very slow rate, all of the information can be processed. However, when the limit is reached, only one source will be able to enter the system; the others will be excluded.

In 1964, Van Mondfrans and Travers conducted a study that assigned 72 male and female undergraduate students to three modes of presentation. Modes consisted of a series of learning trials with stimulus materials of differing degrees of meaningfulness and redundancy presented in an auditory, a visual, or an audiovisual format. One-fourth of each group received stimuli at the rate of one stimulus every 4 seconds, one-fourth at the rate of one stimulus every 2 seconds, one-fourth at a 1-second rate, and one-fourth at the rate of one stimulus every 0.6 seconds. Results revealed no significant differences between the visual and the audiovisual modes of presentation across all types of stimulus material. Travers (1967) noted, "The use of a single channel for transmitting information seems to be a safe rule except where the information transmission requires the use of more than one sense modality" (p. 145).

Based on a series of research studies and Broadbent's (1958) theoretical model of the perceptual system as a single-channel system, Travers (1970) arrived at the following conclusions regarding multichannel transmission of information.

- The transmission of redundant information through two perceptual systems will not lead to more effective information transmission than the use of a single modality except when the rate of information transmission is very slow (p. 105).
- When non-redundant information is transmitted through two different perceptual systems . . . the two channels together do not result in the retention of greater quantities of information than when one channel is used alone (p. 106).
- The main factor limiting the rate at which information is received and at least temporarily stored depends on events at the highest levels of the nervous system and not on the number of perceptual systems through which information is transmitted (p. 106).
- The bottleneck in the information processing occurs "upstairs," in the brain, and both auditory and visual information seem to encounter either the same bottleneck or bottlenecks of equal size (p. 92).

35.5.3 Cue Summation Theory

Results demonstrating increased learner achievement utilizing two channels of instruction (audiovisual) rather than one channel were realized by several studies. In 1950, the Office of Naval Research sponsored two studies at the Pennsylvania State University with 430 ROTC trainees. Experiments (utilizing film) yielded the following conclusions (Nelson & Moll, 1950, p. 1).

- Significant learning accrues from the presentation of film as a whole than from the presentation of either the audio or the video channel alone.
- Neither channel is consistently better than the other.
- Both channels together are consistently better than either one alone.

An eminent study that compared a single presentation of audio or print with the combined presentation was conducted by Hartman (1961a). This study included 1,184 freshmen at the Pennsylvania State University. The design of the experiment provided subjects with a 5-second exposure to a photograph on motion picture film. For the audio treatment, a professional announcer spoke the person's name in the photograph, paused 1 second and then repeated it; for the print treatment, the name was printed on the lower portion of the frame and appeared simultaneously with the picture throughout the exposure. Hartman's (1961a) conclusion was that "redundant information simultaneously presented by the audio and print channels [was] more effective in producing learning than [was] the same information in either channel alone" (p. 42).

Hartman (1961b) further investigated the relationship to redundant audio/print stimuli in a review of literature and published the following results:

> Of nine studies comparing simultaneous audio-print presentations with audio presentations of the information, four indicated superiority for the combined channels, two favored audio, and three indicated no differences. With regard to the comparisons of simultaneous audio-print with print, seven studies supported the simultaneous presentation, and two found the presentations equivalent.

Hartman concluded, "It is apparent that a simultaneous audio-print presentation is more effective than either audio or print alone when the information simultaneously presented is redundant" (p. 244). He supported his findings with the cue summation theory. This theory predicts that learning of discriminations is increased as the number of available cues or stimuli is increased. Some researchers add the caveat that cues must also be available in a situation wherein discrimination is tested.

In another study, Severin (1968) designed a series of treatment conditions: audio with relevant pictures (cue summation condition), audio only, visual only, audio and print (redundant condition), and audio and unrelated pictures. The sample population consisted of 246 seventh-grade students, and the task was recognition. Results demonstrated that the cue summation combination was significantly superior to the redundant condition, and the visual-only treatment was superior to the audio-only treatment. Although Severin's findings agreed with Hartman's and others, he stressed the distinction between redundant, relevant, and related information. His general conclusions were as follows.

1. Multichannel communications that combine words with closely related and relevant illustrations can provide significant gains because of the summation of cues between channels.
2. Multichannel communications that combine redundant words in two channels (speech and print) do not result in significantly greater gain than single-channel communications because the added channel does not provide additional cues.
3. Multichannel communications with unrelated cues in two channels causes interference between channels and can result in less information gain than if one channel were presented alone.

Severin emphasizes the importance of relevant cues in his interpretation of the cue summation theory. He states that "these results suggest that multi-modality stimulus materials may be capable of producing increments in learning under certain conditions; they also suggest what one of these conditions might be: use of the second channel to carry additional related cues rather than simply redundant cues" (p. 9).

35.5.4 Limited-Capacity Information Processing

A study that was not completely consistent with Broadbent's model was conducted by Hsia (1968b). In this study, 912 students in the seventh grade were randomly assigned to test conditions. The content of treatments was verses of English poetry. Six conditions were set up in differing combinations of channel treatments (audio, visual, and audiovisual) and noise factors (noise and no-noise conditions). The noise factor was either auditory noise (white noise at 25% intensity) or visual noise (random black dots scattered over visual stimuli). After the presentation of each verse, subjects were given just enough time to recall what they had seen on the screen or heard from the tape recorder or both and to write down those recollections. Scoring was based on the aggregate number of words recalled, erred, lost, and presented. Data results for the total amount of recalled information revealed that neither audio nor visual was significantly better. An audiovisual treatment, however, was significantly superior to both the audio and the visual treatments. Hsia concluded that "data showed that both communication efficiency and dependability were higher in the AV channel than in the A and V channels alone" (p. 342).

In the discussion of his experiment and in a follow-up review of research, Hsia (1968b) remarked that Broadbent's single information processing channel was "demonstrably insufficient" (p. 326). He criticized Broadbent's "mechanical model," which expounds that the information system can process only audio or visual information, thereby limiting the amount of information gain in an audiovisual presentation to that of a single channel (Hsia, 1971, p. 52). Hsia based his finding on the capacity limit theorem (Shannon & Weaver, 1949) of the central nervous system. Simply stated, he asserted that

> if the combined amount of information of audio and visual stimuli exceeds the upper limit of the central nervous system capacity, then both selection processes and interference take place; yet so long as neither audio nor visual stimuli reach the limit of the central nervous system, an audio-visual presentation is generally a more efficient method of presenting communication materials. (Hsia, 1968a, p. 253)

Hsia (1971) also pointed out that with the use of between-channel redundancy, the amount of information received from an audiovisual presentation is less than the amount received from the sum of the audio and visual components. In other words, "the total information provided by a stimulus or a message with a number of dimensions cannot be equal to but is always less than the sum of the information in each dimension" (p. 63). In addition, one channel can provide cues and clues for the other, thereby eliminating probable interference or information jamming.

In an attempt to reconcile findings of the cue summation theory and Broadbent's theory, Hsia (1971) hypothesized that man is a multiple-channel organism when input is optimal. In other words, he is capable of processing information through multiple channels, "so long as the inflow is within the limit of his information processing capacity. When input is far beyond this information processing capacity, it is possible that man may act as 'a single communication channel'" (p. 65).

35.5.5 Dual-Coding Theory

The dual-coding theory evolved over a 30-year period from a series of experiments and hypotheses related to imagery (Paivio, 1991). At the core of the theory is the relationship between symbolic systems and specific sensorimotor systems. Verbal systems refer to properties of language and include both text (visual) and speech (audio). Nonverbal systems refer to nonlinguistic items, such as graphics (visual), sound effects (audio), and objects (visual).

Imagery, concreteness, and verbal associative processes are important tenets of dual-coding theory. For example, "lessons containing concrete information and evoking vivid images will be easier to comprehend and remember than lessons that are abstract and not image-arousing" (Clark and Paivio, 1991, p. 173). Although the dual-coding theory has wide-ranging implications for memory, language, and cognition, it is the combination of verbal and nonverbal aspect that is most often cited in multichannel literature. For example, the *Theory Into Practice* website lists the following principle for dual coding: "Recall/recognition is enhanced by presenting information in both visual and verbal form" (Kearsley, 2001, p. 1). A key feature of dual coding is the cross-connection between information in the linguistic and that in the visual–spatial store. Hence, the viability of dual coding depends, to a large degree, on the semantic overlap between both types of information during initial encoding. As noted by Hannafin and Hooper (1993) in *Instructional Message Design,* "Information depicted in each modality must be congruent. Dual coding of text and graphics, for example, is affected by the degree to which the graphics and text reflect redundant information. Nonredundant information increases the processing requirements of the task and may hinder encoding" (p. 196). Based on this principle, many research studies and development projects combine visual (nonverbal) and verbal information, such as graphics with text or speech with graphics. "Dual coding is ineffective when both sources of information employ identical coding mechanisms. Identical presentation of words in sound and text, for example, should be avoided" (p. 196).

Not all theorists agree with the dual-coding theory, and it is often applied in a very broad sense. In the 1996 edition of this handbook, Braden noted that there had not been a direct conflict between the dual-coding and the cue summation theories. Nor had there been any attempts to reconcile or combine them.

35.5.6 Cognitive Load

The cognitive load theory can also provide important theoretical guidance to multichannel instruction. Basically, cognitive load refers to the limitations of working memory. As there is general agreement that STM is severely limited in capacity and duration (see *Auditory Memory*), designers should seek to reduce working memory load. In contrast, LTM is theoretically unlimited and consists of hierarchically organized schemas. "Schemas allow us not only to store learned information in long-term memory but, because multiple elements of information are treated as a single element in working memory, schemas also reduce the burden on working memory" (Kalyuga et al., 1999, p. 351).

Sweller and others differentiate between intrinsic and extraneous cognitive load (Sweller, 1999; Sweller & Chandler, 1994). Intrinsic cognitive load relates to the difficulty of the material as defined by the magnitude of the body of knowledge to be learned and the degree of interrelatedness among its elements. If the information is complex or complicated, the intrinsic cognitive load will be high. For example, the cognitive load involved in learning human anatomy is intrinsically high because of the large number of terms that must be learned. Extraneous cognitive load refers to the way the instruction is designed and organized—if the instruction contains irrelevant information or other material that causes inefficient cognitive processing, the extraneous cognitive load will be high.

With the assumption that there are two separate systems for modalities in STM (audio and visual), "the effective size of working memory may be increased by presenting information in a mixed (auditory and visual mode) rather than in a single mode" (Mousavi, Low, & Sweller, 1995, p. 320). In this manner, there would be an increase in the capacity of working memory, thereby allowing an increase in the amount of information that could be processed (Andres & Petersen, 2001/2002; Goolkasian, 2000; Kalyuga et al., 1999).

Several research studies have tested this theory. In a series of experiments involving high-school students and geometry lessons, Moursavi, et al. (1995) provided instruction in visual/visual (text/graphic) or audio/visual (speech/graphic) formats. The results indicated that the mixed audio/visual format was more effective than the visual/visual format, and they concluded that "when students must split their attention between multiple sources of information that require mental integration, cognitive resources available for learning can be increased by presenting some of the verbal materials in auditory rather than written form" (p. 333).

In a study involving adults, electrical content, and a computer-based instructional program, Kalyuga et al. (1999) employed variations of modality mixes and color coding to reduce cognitive load. Using the cognitive load theory, they provide the following guidance for instructional designers (when dealing with split-source diagrams and text).

1. Textual materials should be presented in auditory rather than written form.
2. Textual materials should not be presented in both auditory and written form.
3. If textual materials must be presented in written form, search for diagrammatic referents should be reduced by using appropriate markers or guides such as colour-coding. (p. 369)

In similar research, Tindall-Ford, Chandler, and Sweller (1997) conducted three experiments where audio text with visual diagrams or tables was presented to some participants and visual text with visual diagrams was presented to other participants. The results (obtained only for instructions with a high intellectual content) showed that the audio/visual treatment resulted in increased performance over the visual/visual treatment. Basing their results on the cognitive load theory, the researchers stated the following.

1. When students are faced with intellectually difficult materials requiring mental integration between multiple sources of information, results suggest that mental integration may be easier if written information is transferred into an auditory form (p. 285).
2. Audio-visual presentations are unlikely to be beneficial if the auditory component is structured in such a way that it overloads working memory—for example, if it is too long in length (p. 283).
3. If the audio component makes excessive memory demands because of its inherent complexity (even if it is short in length), it may overload working memory and render audio-visual instruction ineffective (p. 284)
4. Audio-visual instruction also may not be ideal in situations when the audio component is unnecessary or redundant for understanding (p. 284)

35.5.7 Review of the Literature

In an investigation into some of the apparent contradictions in the multichannel research, Lang (1995) conducted an extensive review of the literature. She analyzed 22 research studies (from the 1960s through the early 1990s) that were related to audio/video redundancy (primarily television). Based on the review, she identified four factors that were contributing to the conflicting results (p. 87).

1. ***Overall theoretical perspective guiding the research.*** Lang notes that although several theories "have been used to attempt to explain why pictures might (or might not) improve memory, they have not been used consistently to guide the conceptualization and operationalization of audio/video redundancy or memory" (p. 88).
2. ***Conceptual definition of audio/video redundancy.*** Variations in definitions may have accounted for a great deal of the conflicting results. In particular, three definitions of redundancy were noted in the studies (p. 90):
   a. the presence of two channels rather than a single channel,
   b. an exact match in content between the audio and the video channels, and
   c. a relationship in semantic meaning between the audio and the video channels.
3. ***Operational definition of audio/video redundancy.*** Lang identified 24 different operational comparisons for testing redundancy, including single words spoken aloud, single words flashed on screens, simultaneous audio/video single words, audio words with visually redundant or conflicting stills, narration with redundant text on screen, narration with redundant video, and narration with conflicting video. For her analysis, Lang grouped all of the operational comparisons into four categories—single-channel presentations, multiple-channel redundant presentations, audio/video conflicting, and talking heads.

4. ***Operational definitions of memory.*** There were multiple definitions of memory, including recognition, cued recall, and free recall.

Using the limited-capacity information processing model, Lang hypothesized that multiple-channel redundant information and talking head video were unlikely to overload the processing system and would thus result in increased processing and memory (over single-channel and multiple-channel conflicting information). By stratifying the studies based on consistent definitions of redundancy and memory measure, Lang concluded that her analysis showed the following effects of audio–video redundancy on memory (pp. 111–112).

- Her research suggests quite strongly that multiple-channel redundant presentations are better than single-channel presentations at every level of processing (e.g., encoding, storage, and retrieval).
- Memory for visual information is not much affected by audio/video redundancy at any stage of processing. But information for audio information is sadly compromised when redundancy between the two channels falls. This appears to be due primarily to a superiority of visual information both at recognition and recall.
- Having pictures—even conflicting pictures—appears to have the largest impact at retrieval.

She also noted, "This review reveals abundant evidence already reported in the literature that visual processing may require less capacity than audio processing and that a lack of audio/video redundancy has a much greater detrimental effect on memory for audio information than it does on memory for video information" (p. 111).

35.5.8 Summary

When examining the research studies and theories presented in this section, one can see many similarities and differences. As noted by Lang (1995), the variations in research results can often be attributed to different definitions of terms (such as redundancy, channel, cue), different philosophical perspectives, and variations in research designs. It seems evident that there are many variables that influence optimal combinations of audio and visual information, including the type and complexity of the information, the attributes of the target audience, and the level of redundancy.

The theories presented in this section serve as examples of theoretical foundations that have been used to analyze the implementation of audio. Other theories, such as interference, schema, and generative learning, are also valuable components in the quest to understand modality-based instruction. Although the controversy over multichannel processing of information is not over, guidelines are beginning to emerge; those that relate to multimedia instruction are presented in the next section, *Audio in Interactive Multimedia*.

35.6 AUDIO IN INTERACTIVE MULTIMEDIA

35.6.1 Introduction

Multimedia instruction can be loosely defined as educational programs integrating media elements (text, graphics, animation, sound, and video) in an interactive environment controlled by a computer or similar processor (Barron et al., 2002). This definition encompasses many other terms from education, commercial, and military environments, such as the following.

- Computer-based education
- Computer-assisted instruction
- Computer-based training
- Computer-assisted learning
- Computer-assisted instruction
- Technology-based training
- Interactive courseware
- Interactive videodisc
- Interactive multimedia instruction
- Multimedia learning
- Web-based training
- Internet-based learning
- e-Learning

Educational materials delivered via a computer date back to the 1960s when PLATO and TICCIT used mainframe computers and "dumb" terminals to disseminate instruction. At that time, audio was not a part of the instruction, other than perhaps a beep for an incorrect answer. When interactive videotape and interactive videodisc appeared, audio, in analog form, became a component of some programs and the related research (for example, the report by Fletcher, 1990, included programs with audio).

Although it was possible to digitize audio and incorporate it into computer programs in the 1980s, it was not practical. The primary limitation was the storage requirements. One minute of digitized audio can consume several megabytes of storage if it is stored in high-quality (44.1-kHz), 16-bit, stereo format (Barron et al., 2002). Finally, in the early 1990s, the advent of large hard drives and compact discs, along with improved, standardized compression techniques and formats, created an environment wherein incorporating audio into multimedia was feasible, easy, and relatively inexpensive (Barron and Varnadoe, 1992).

35.6.2 Applications of Audio in Multimedia Instruction

There are numerous ways to employ digital audio in interactive multimedia, including the following.

- Gaining attention. Audio (in multimedia instruction) can be used to draw attention to a program or direct attention to the most important parts of the screen (Aarntzen, 1993). "In fact, research by Kohlfeld (1971) and Posner, Nissen, and Klein (1976) has confirmed that sounds generally are more effective than images for gaining attention" (Bishop & Cates, 2001, p. 13).

- Temporal sounds. Temporal sounds can be defined as "spoken information provided about future and past events that present highlights and details about static or moving visuals" (Mann, 1995, p. 402). Mann advocates using sound as epitomes (similar to an advanced organizer, except that the key ideas are presented) or verbal summaries to help in the interpretation of the message.
- Assessments. There are several advantages to incorporating sound into computerized assessments, including meeting the needs of nonreaders and visually impaired examinees. In addition, audio can be incorporated into the user interface (spoken directions and interactions) or it can be used to test listening skills (Parshall, 1999).
- Interface. Sounds are commonly used to provide human–computer interactions through interfaces (Gaver, 1997). Operating systems provide audio feedback when a computer boots up or when a dialog box must be closed (Helander, Landauer, & Prabhu, 1997). Multimedia courseware can also be programmed to provide audio cues for navigation, such as menus.
- Instruction. Recorded voice is often used to provide the mainstream of instruction (Aarntzen, 1993). A narrator speaks to the student as text, graphics, and/or animations appear on the screens. Sound effects may also be provided, such as heartbeats for medical programs and sonar signals for the military. "Sounds can communicate information when visual attention is focused elsewhere, when tasks do not require constant visual monitoring, or when the visual channel is overburdened" (Bishop & Cates, 2001, p. 14).
- Prompts and feedback. Some instructional multimedia programs use audio to provide prompts or feedback for the students (Anderson-Inman, 1990; Jongekrijg & Russell, 1999). One approach is to have one voice or character used for instruction and another for hints or feedback.

35.6.3 Research Focusing on Audio in Multimedia

When digital audio became feasible and designers began incorporating it into multimedia courseware, there were few guidelines to follow. Consequently, very little of the design or development was based on instructional theory or research (Barron & Atkins, 1994; Daniels, 1995; Kozma, 1991; Moore, Myers, & Burton, 1996; Moreno & Mayer, 1999). In recent years, several research studies have emerged that focus on specific design options for audio in multimedia courseware. This section presents research related to textual and audio redundancy, temporal contiguity, coherence, audio feedback, aptitudes of learners, and cost/time issues.

35.6.3.1 Textual and Audio Redundancy. Speech is ephemeral, and Winn (1993) noted that "the most serious drawback to using speech (or any sounds for that matter) in instructional messages is that they lack the permanence of text or illustration" (p. 116). There are many ways to counter this drawback, such as repeating the information (as a paraphrase) or providing a button that the student can click to repeat the audio segment (Barron & Kysilka, 1993). Another approach is to provide redundancy of the information through other channels, such as text or images.

There are several options for presenting verbal information through audio/visual combinations in multimedia programs. Assuming that all of the programs contain static or animated graphics, the following combinations of text and audio can be used.

1. Full text mirrored by verbatim audio
2. Full text combined with audio highlights
3. Full text used without spoken audio
4. Partial text combined with full audio
5. No text combined with full audio

The optimal combination of text/audio is a perplexing issue for multimedia developers, and numerous research studies have investigated this matter. Several theories related to the multiple modality of working memory (Baddeley, 1992) and dual coding (Paivio, 1986) promote the premise that combinations of verbal and nonverbal information, using both visual and audio modalities, can increase working memory. Tindall-Ford et al. (1997) explored this hypothesis with three experiments involving adults. Their results, based on the cognitive load theory, were, "When students are faced with intellectually difficult material requiring mental integration between multiple sources of information, results suggest that mental integration may be easier if written information is transferred into an auditory form. Alternatively, when information is not intellectually challenging, the mode of presentation may be of less importance" (p. 285). Their results appear to support option 5 no text combined with full audio in the preceding list. However, the researchers note that many variables must be considered. For example, if the audio segment is long or extremely complex, it may exceed working memory, and a modality effect is not likely to be observed.

Whereas the research by Tindall-Ford et al. contained static graphics and tables, Mayer and Moreno (1998) and Moreno and Mayor (1999) conducted research investigating the combination of animations and audio. They found that in four of four studies, the combination of verbal audio with animations was better than the combination of text and animations. They provided a similar rationale for the findings "When pictures and words are both presented visually (i.e., as animation and text), the visual/pictorial channel can become overloaded but the auditory/verbal channel is unused. When words are presented auditory, they can be processed in the auditory/verbal channel, thereby leaving the visual/pictorial channel to process only the pictures" (Mayer, 2001, p. 134).

A study by Koroghlanian and Klein (2000) is similar to Mayer's research in that it employed both static illustrations and animations, with and without audio. A biology lesson with 109 high-school students was used to determine the effects of instructional mode (audio vs. text), illustration mode (static illustration vs. animation), and spatial ability (high vs. low) on achievement. Contrary to the findings by Mayer and

Tindall-Ford, this research yielded no significant differences, except that the high-spatial ability students achieved more than the low-spatial ability participants. Although the text vs. audio versions did not differ in completion time, the animation treatment took significantly more time than the static illustration treatment.

Kalyuga et al. (1999) have conducted numerous experiments investigating split attention, cognitive load, and redundancy. Some of these experiments included an audio variable.

> According to the cognitive load theory, learning might be inhibited when learners must split their attention between and mentally integrate text and graphics because the integration process might overburden limited working memory capacity. However, when textual information is presented in auditory form, mental integration with a diagram may not overload working memory because working memory may be enhanced by the use of both visual and auditory channels. Such a dual mode of presentation might be used to circumvent cognitive load problems caused by split-attention This phenomenon will be referred to as the instructional modality effect. (p. 353)

One experiment that tested this modality effect involved 34 adults and a training module focused on soldering. The audio-only version was superior to both the text-only and the redundant audio/text versions (Kalyuga et al., 1999). Similar results were reported in 2000—the auditory presentation of text was better than the visual-only presentation and the combined visual and auditory forms. However, in this study, the research team also noted that the results differed, based on the experience level of the learners (Kalyga, Chandler, & Sweller, 2000).

There are other recent studies that did not demonstrate instructional modality effects. For example, Shih and Alessi (1996) compared three treatments (text, voice, text and voice) for 141 college undergraduates. They included two types of information (spatial vs. temporal). The results did not show any significant differences between groups or any significant interactions among treatment groups and type of learning. However, 82% of the students reported that they preferred the combination of text and voice to the other methods of delivery.

Quealy and Langan-Fox (1998) conducted a study that involved 60 adults in three treatments: (a) stills (graphic/text screens), (b) audio (graphic/text screens with verbatim audio), and (c) video (video/text screens with verbatim audio). The results showed no significant group effect for recall.

Rehaag and Szabo (1995) assigned 82 high-school students (low ability and high ability) to either a text-based or a text plus full-audio version of a mathematics program. The program was originally designed as a text-based program, and audio was added to 3 of the 110 lessons. They did not find any significant differences in achievement or attitudes between the two main treatment groups.

Three studies used the amount of text as a variable, using treatments of (a) full text, (b) full text and full audio, and (c) partial text and full audio (Barron & Atkins, 1994; Barron & Kysilka, 1993; Koroghlanian & Sullivan, 2000). All three of the studies found no significant differences, prompting the researchers to conclude that the partial-text/full-audio approach (which is common in industrial and military multimedia programs) is at least as effective as full text/full audio. The partial-text approach is often used to conserve screen real estate in multimedia instruction and allow space for complex simulations or images. The question has been raised as to whether this decrease in visual text would adversely affect learning. As indicated in these studies, the impact was not significant.

A study that investigated audio in a hypertext dictionary environment was conducted by adding sound to the textual definitions. It was hypothesized that the encoding of two modalities would result in better retrieval of vocabulary. The results, however, indicated that the audio resources did not have a significant effect on retention (Tripp & Roby, 1994).

Based on these studies, it seems clear that verbatim audio, along with text does not generally increase achievement over a text-only approach. Some of the studies also indicate that full-audio/partial-text and full-audio/full-text delivery can achieve the same results. However, it is not clear when eliminating the text entirely (and using full audio) will increase achievement. There are obviously additional variables that could be studied. For example, Jeung, Chandler, and Sweller (1997) suggested that dual-mode instruction would be beneficial when mental resources were not required for extensive visual searching. Other variables that could be investigated include the content, duration, and pacing of multimedia instruction. For example, some researchers have noted that many of the studies conducted by Mayer and associates (that favored spoken audio over text) consisted of short segments (less than 3 minutes), focused on technical content, and did not include learner-pacing. In experiments that included different content, length, and pacing, Tabbers, Martens, and van Merrienboer (2001) found, "In the two groups in which the students set the pace of the instruction, no modality effect is found at all. Not only do the students in the visual-user group perform almost equally well on the transfer test, on the retention test they even outperform the students in the audio-user group" (p. 1029). Tabbers (2002) recommends, "Only in situations in which time-on-task is a crucial variable and the instructions are system-paced based on the pace of the narration, should spoken text be first choice in multimedia instruction" (p. 79).

35.6.3.2 Temporal Contiguity. In 1992, Mayer and Anderson proposed a contiguity principle, which recommended that words and pictures should be presented concurrently, rather than sequentially. In a 1999 article, Moreno and Mayer divided the contiguity principle into a temporal-contiguity effect (visual and spoken materials) and a spatial-contiguity effect (printed text and pictures). The rationale behind these effects is that if two stimuli are presented simultaneously, it will be easier for the learner to integrate them in STM. Likewise, if there is only a short time lapse between stimuli (as opposed to a longer one), the learners may be able to build the connections required to integrate the information. Mayer and others tested the principle on a series of experiments that involved low-ability students with high spatial ability. The results in three of five tests measuring retention and eight of eight tests measuring transfer supported the temporal contiguity principle (Mayer, 2001; Mayer

& Anderson, 1991; 1992; Mayer & Sims, 1994; Mayer, Moreno, Boire, & Vagge, 1999; Moreno & Mayer, 1999).

35.6.3.3 Coherence. The term *seductive details* refers to the addition of interesting, entertaining, yet perhaps distracting, information in a lesson or text (Garner, Gillingham, & White, 1989). For example, a textual passage might provide interesting, but irrelevant, information in addition to the main idea (Wade, Schraw, Buxton, & Hayes, 1993). If the learners then recall the extraneous details, to the detriment of the main idea, it is referred to as the seductive detail effect. Centering on research in reading of informative text, a controversy ensued in the early 1990s. Wade (1992) concluded that "the practice of adding anecdotes and seductive details does not facilitate and may even interfere with the learning of important information "(p. 264). Other researchers agreed (Garner, Brown, Sanders, & Menke, 1992; Hidi, 1990).

However, in a review of the seductive detail literature related to reading, Goetz and Sadoski (1995a) concluded that there was insufficient evidence to confirm or deny the construct. Citing methodological problems, they remarked that "it is impossible to determine whether readers were being bewitched by distracting details, bothered by incoherent text, or bewildered by incomprehensible abstraction" (1995b, p. 509).

Recently, several research studies have investigated a similar construct, referred to as the coherence effect, in multimedia. In 11 of 11 experiments, students who received a concise presentation of a multimedia program performed significantly better (in both retention and transfer measures) than those who received a lesson with extraneous material (Harp & Mayer, 1997, 1998; Mayer, Bove, Bryman, Mars, & Tapagco, 1996; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2000). For example, one experiment included four different treatments of an animated lesson that focused on the phenomenon of lightning: concurrent narration (N); narration with sound effects (NS); narration, along with background music (NM); and narration, sound effects, and background music (NSM). The results supported the coherence effect in that the NSM group recalled significantly less information than all of the other groups, the NM group recalled significantly less than the N and NS groups, and groups N and NS did not differ from each other (Moreno & Mayer, 2000). A related study using a different animation and sound effects (which were not as closely related to the content) noted a detrimental impact of the sound effects as well as the music, prompting the researchers to suggest that "in multimedia lessons, the more relevant and integrated sounds are, the more they will help students' understanding of the materials" (p. 124). Based on the compiled evidence of several research studies, Mayer (2001) provides the following coherence principles (p. 133).

- Student learning is hurt when interesting but irrelevant words and pictures are added to a multimedia presentation.
- Student learning is hurt when interesting but irrelevant sounds and music are added to a multimedia presentation
- Student learning is improved when unneeded words are eliminated from a multimedia presentation.

35.6.3.4 Audio Feedback. Sales and Johnston (1993) conducted a research study focused on the use of digitized speech as feedback in multimedia instruction. The subjects were 145 sixth-grade students who received feedback in one of three ways: spoken audio, printed and spoken, or spoken by an animated character. Although the results were not significant, the researchers provided the following recommendations that related to audio.

1. The gender of the speaker may result in differing levels of performance. When a female agent was used, female students outperformed male students.
2. Digitized speech can help overcome reading limitations of students.
3. Students were more on-task when they wore headphones.
4. Clients expect speech enhancements to software, much as they expect color to be used in the lesson.
5. Important symbol systems related to audio include gender of the speaker, tone, mode, pacing, cultural cues, and message. These have different degrees of importance to individual learners.

In another study that focused on feedback, Huang (1995) provided audio only; audio with text; or audio with text and animation feedback for a computer-based lesson about tennis. Sixty-eight adults were categorized as having high or low prior knowledge and randomly assigned to a feedback treatment. Results indicate that there was no interaction between ability and feedback type. However, the elaborate (audio with text and animation) feedback was significantly more effective.

35.6.3.5 Learner Characteristics. It seems obvious that the aptitudes and interests of the learners can and do interact with other variables related to multimedia instruction. Although too numerous to detail, almost every study mentioned in this section provides some information about the learners' characteristics and possible interactions with the treatment. This information is invaluable when findings are generalized to additional audiences. For example, Mayer noted a difference in the effects of various design strategies between low-knowledge and high-knowledge learners and between learners with low spatial and high spatial ability. Observing that design effects were stronger for low-knowledge and high-spatial learners, he concluded that the high-knowledge learners were better able to compensate for less-than-optimal design approaches. Likewise, the low-spatial learners did not benefit as much, possibly because they had less capacity to integrate images and text in working memory (Mayer, 2001). Many of Mayer's studies included only low-knowledge and high-spatial ability students.

In their article, "*Levels of Expertise and Instructional Design,*" Kalyuga, Chandler, and Sweller (1998) emphasize the importance of "knowing one's students" (p. 12). Once individual differences are determined, multimedia designers can adapt the instruction to "accommodate differences in ability, style, or preferences among individuals" (Jonassen & Grabowski, 1993, p. 19). For example, in the article, *"When Using Sound with a*

Text or Picture Is Not Beneficial for Learning," Kalyuga (2000) noted that if a diagram is self-explanatory, experienced learners performed better with the diagram alone—without verbal audio. The researcher offered the following principle: "When presented in auditory form, textual explanations should be easily turned off or otherwise ignored by more experienced learners" (p. 171). One option is to present the instruction in various methods, allowing the learners to select the most appropriate strategy or modality (Mayer, 2001).

35.6.3.6 Impact on Time and Cost. Cost is an important issue that must be considered prior to incorporating audio into multimedia instruction. It could easily cost several thousand dollars per courseware hour to hire a professional narrator, record and edit the audio files, and integrate them with an authoring program. Additional development costs would be incurred for digitizing equipment, editing software, and streaming servers. Given these financial considerations, some researchers have concluded that audio is not worth the price. Based on a study involving 120 adults, Main and Griffiths (1977) concluded that the additional cost and effort involved with the addition of audio "are not warranted until it can be demonstrated that audio and pictorial supplements provide instructional advantages that cannot be duplicated in other ways that are less expensive and easier to implement" (p. 178).

In industrial settings, time often equates to money; therefore, an investigation of audio's impact on time was the focus of several studies. Since audio is time-based, and text is not, it seems likely that programs with full audio could require more time for students to complete. This was, in fact, the case in experiments conducted by Barron and Kysilka (1993). They found that multimedia instruction with full audio required significantly more time to complete than the same program without audio (1993). Likewise Koroghlanian and Sullivan (2000) noted that the full-audio groups required more time. Rehaag and Szabo (1995) conducted a study using two treatments—full audio and full text. They found that the higher-ability students required more time in the full-audio treatment than in the text treatment. However, lower-ability students required equal time for both treatments. They surmised that the higher-ability students were able to scan the text version quickly and thus spend less time. The high-ability, full-audio group also expressed significantly more negative responses to the statement "I felt I could work at my own pace."

35.6.4 Summary

Audio is a major component of many multimedia programs, and instructional designers have many choices to make dealing with when, where, and how the audio elements should be implemented. The design decisions are extremely complex because other media elements, the content, and learners' characteristics must be considered. As Quealy and Langan-Fox (1998) remarked, "The real world (even, when only 'modeled') is awash with uncontrolled variables in a shifting dynamic equilibrium" (p. 275). The good news is that research studies are targeting the use of audio in multimedia design; the bad news is that we have a long road to travel before all of the variables are identified and concise guidelines are available.

35.7 TIME-COMPRESSED SPEECH

35.7.1 Introduction

Unlike text, speech is inherently time-based—it cannot easily or effectively be skimmed. Several research studies have noted that a multimedia program with full narration will take significantly longer for a student to complete than a multimedia program with on-screen text (Barron & Kysilka, 1993; Koroghlanian & Sullivan, 2000). As the use of audio proliferates in multimedia learning and other venues, techniques to minimize students' time commitments are being investigated (He & Gupta, 2001).

An interesting paradox is that people can understand speech faster than narrators can speak—even the fastest talkers reach a physiological barrier at about 215 wpm (words per minute; Beasley & Maki, 1976). Conversation typically takes place at approximately 150 wpm (Benz, 1971; Nichols & Stevens, 1957). Since, in conversation, one is simultaneously listening and composing speech, it was assumed that perhaps another 125 to 150 wpm of unused processing capacity might be available in simple listening. Because the rate for speed reading is 250 to 300 wpm (Taylor, 1965) and the rate for silent reading is 275–300 wpm (Junor, 1992), it was assumed that a similar capacity might be available for listening. In fact, Fairbanks, Guttman, and Miron (1957) found that good intelligibility is possible when speech is compressed up to 50% of its original time length—up to 300 wpm.

35.7.2 Compression Technology

In early studies, time-compressed speech was produced by playing back a recording at a speed faster than the original recording. Although this method is simple to produce, the vocal pitch and intelligibility are affected (the "chipmunk effect"). The limitations of this technique generally rendered research findings questionable.

Miller and Lichlinder (1950) first demonstrated the tape-sampling method accomplished by deleting small segments of the speech signal. A switching device was used that turned off the signal periodically. Garvey (1953) performed further experimentation in compressed speech by editing out segments of the audiotape and splicing the ends of the retained tape together. Although Garvey's technique was successful, it was deemed too tedious except for research purposes. Fairbanks, Everitt, and Jaeger (1954) produced the first electromechanical apparatus that allowed both the expansion and the compression of recorded tape.

A relatively simple technique is the sampling, or Fairbanks method, which consists of removing small portions of the signal at regular time intervals (Arons, 1994). A cross-fade can then be performed between segments to improve the perceived quality. Currently, a common, linear time-compression technique is the synchronized overlap-and-add (SOLA) technique (Roucos & Wilgus, 1985). The SOLA compression method can be completed in real time (on a desktop computer). It improves the quality of speech (over the Fairbanks method) by computing the cross-correlation and locating the optimal match between segments before applying the cross fades (Arons, 1992). Both the Fairbanks and the SOLA techniques are referred to as linear time compression because the algorithms are applied consistently, and all segments are reduced by equal proportions.

Nonlinear time compression (in which the compression rates may vary from point to point) can also be applied. Nonlinear compression techniques are used to reduce redundancies, such as pauses and long vowel sounds. Combinations of linear and nonlinear are common (Arons, 1992). For example, some of the silent segments may be reduced prior to applying the SOLA technique, or the SOLA technique may be applied prior to compressing the vowel sounds.

Advanced algorithms, involving dichotic presentations for the ears or complex, adaptive compression methods, have also been developed (Arons, 1994; Covell, Withgott, & Slaney, 1998; He & Gupta, 2001). These sophisticated algorithms have been able to demonstrate increased compression rates, and at very high compression ratios, users prefer them. However, most advanced algorithms have failed to show a significant advantage in intelligibility or comprehension at the speech rates that are comfortable for listeners because intelligibility reaches a ceiling at about twice the normal speaking rate; this "ceiling" rate can be attained very easily by the linear techniques (Janse, Nooteboom, & Quene, 2001). For example, He and Gupta (2001) tested various linear and nonlinear compression techniques. They found that, regardless of the sophistication of the algorithm, the participants reached a ceiling level. Because the linear techniques, such as SOLA, are easier to implement and are equally effective below the ceiling level, they continue to be the most common (He & Gupta, 2001).

35.7.3 Comprehension

Intelligibility and comprehension are discrete, yet related constructs. Intelligibility refers to being able to identify isolated words, and it has been achieved with compression rates up to 10 times normal speech (Arons, 1997). Comprehension refers to understanding the content of a passage (measured by answering questions related to the content).

Numerous researchers have varied the rate of compression and measured the resulting effect on comprehension. Fairbanks et al. (1957) found little difference in intelligibility of selections compressed to 141, 201, and 282 wpm. Diehl, White, and Burke (1959) determined that listening comprehension was unaffected by changes between 126 and 175 wpm. Foulke and Sticht (1967) tested the intelligibility and comprehension of different rates of speech with 100 college students. Using rates of 225, 275, 325, 375, and 425 wpm, they found intelligibility scores of 93, 91, 89, 85, and 84%, respectively. The comprehension scores were 73, 66, 67, 56, and 53%, respectively. These results demonstrate a 6% loss in comprehension between 225 and 325 wpm and a loss of 14% between 325 and 425 wpm. These and other studies (Boyle, 1969; Carver, 1973; Foulke, 1968; Foulke & Sticht, 1969; Rossiter, 1970; Sticht, 1968; Wasserman & Tedford, 1973; Williams, Moore, & Sewell, 1983–1984) indicated that as word rate is increased beyond about 250 to 300 wpm, there is a decline in comprehension.

However, numerous intervening variables must be considered before a determination of the optimum degree of compression can be made (Duker, 1974). Researchers believe that the ability of subjects to comprehend compressed speech may be dependent on the difficulty of the material. Foulke (1962) determined that the comprehension of a scientific selection was less than the comprehension of a literary work at normal speed. However, at various levels of compression, the comprehension scores of the scientific selection declined less than those for the literary selection. This phenomenon may be because the comprehension scores for the scientific selection were lower at the normal rate; therefore the range in which they could vary was relatively small (Duker, 1974). His data showed that for college students, listening to compressed materials written at the eighth-grade level produced the greatest efficiency. Raising the difficulty level of the materials caused comprehension to drop off abruptly.

Length of presentation may also be a factor in comprehension and memory. Adelson (1975) examined comprehension by a group of college students listening to a 1-hr lecture at 175 wpm, as compared to the same group of college students listening to an equated 1–hr lecture compressed at 275 wpm for 40 min. Compressed materials produced less comprehension than did the normal rate materials. The author concluded that with compressed materials, the length of presentation appeared to be a critical factor, perhaps because of attentional fatigue or other factors.

Learner characteristics may influence the comprehension of compressed speech. Some of these variables include the subject's sex, age, intelligence, and reading ability. Duker (1974) determined that the comprehension scores of male and female subjects revealed no sex-related differences for word rates varying from 174 to 475 wpm. This conclusion is supported by other research studies conducted by Bell (1969), Foulke and Sticht (1967), Klavon (1975), Ludrick (1974), Orr and Friedman (1964), and Ross (1964).

Fergen (1954) and Wood (1966) found that ability to comprehend compressed speech increases with age and grade level of school children. However, beyond age 12, little difference was noted until age 60 or so. Wingfield, Poon, Lombardi, and Lowe (1985) proposed that the decline in comprehension for older people may be the result of a decline in information processing resources and that it becomes more apparent with more complex stimuli. Letowski and Poch (1996) noted a 20% decline in comprehension of time-compressed speech between middle-aged (fifth-decade) and older (seventh-decade) adults. Gordon-Salant and Fitzgibbons (2001) also investigated the relationship among age, content, and comprehension of time-compressed speech. Their research included four compression techniques (linear or selective compression of pauses, vowels, or consonants) and three stimulus forms (sentences, syntactic sets, or random-order words). They reported three major findings related to older listeners.

1. Reduced linguistic cues resulted in reduced comprehension.
2. Acoustic alteration of consonants affected comprehension more than alteration of vowels or pauses.
3. A combination of time compression with reduced contextual cues resulted in a reduction of comprehension.

Aptitude or intelligence may also interact with comprehension. Eckhardt (1970) used a 1-hr multimedia presentation at various rates of compression with Air Force recruits of varying aptitudes. Eckhardt concluded that test differences between the groups were due to aptitude and an aptitude–rate interaction. There was a comprehension loss for lower aptitude subjects at the higher compression levels. Sticht and Glasnap (1972) determined that low-aptitude men learned easier material better than more difficult material as a function of decreased words per minute. High-aptitude men tended to learn material best at 175 wpm, independent of difficulty level. However, other researchers (Sticht, 1968; Watts, 1971; Williams et al., 1983–1984) found that subjects with lower aptitudes or lower reading ability performed as well at higher rates of compression as at normal rates.

Reading ability may also influence comprehension of compressed speech. Breed (1977) tested adult vocational technical school students to determine the differences in listening comprehension when subjects were categorized according to reading ability. The subjects in Breed's study listened to tapes that were time expanded and time compressed, varying in rate from 60 to 240 wpm. Breed indicated that listening comprehension and reading ability appear to be related to verbal skills. The poorest readers exhibited the poorest listening comprehension, and better readers were better listeners as measured by scores on tests of listening comprehension. Goldstein (1940) and Orr, Friedman, and Williams (1965) found a positive correlation between reading rate and ability to comprehend compressed speech. Conversely, both studies further determined that practice in listening to compressed speech resulted in an improved reading rate. Robertson (1977) determined that the comprehension of subjects is not affected when they are presented recorded materials within two reading levels below or three reading levels above their particular grade level. In general, it appears that a relationship between better reading ability and the comprehension of compressed speech can be established, although this may reflect an underlying verbal ability.

Although most of the studies reported in this section are based on limits measured in words per minute, some researchers feel that the limiting factor is compression ratios (and resulting quality) rather than words per minute (Omoigui, He, Gupta, Grudin, & Sanoeki, 1999). Heiman, Leo, Leighbody,& Bowler, (1986) suggest that if a passage is compressed at 50%, all of the redundant information has been removed. Because the original passage may vary in the number of words per minute, they conclude that the compression ratio (50%) is the limiting factor, not the number of words per minute.

35.7.4 Preference

Several studies have been conducted to determine which level of time compression listeners would select if they were presented with several options. Foulke (1969), found that college students preferred a rate of about 207 wpm (approximately 1.38 the normal rate). This rate was less than the 275 wpm speed preferred by blind college students (Foulke, 1965). The researcher surmised that if sighted students had more practice listening to compressed speech, they might prefer the higher rates also.

In a 1995 study, Harrigan offered time-compressed lectures to students at three distinct speeds (1.0, 1.18, and 1.36 the speed of the original lecture). Results showed that 75% of the time, the students preferred the 1.36-rate lecture. Similarly, Omoigui et al. (1999) and Li, Gupta, Senoeki, He, and Yong, (1999) conducted a study that found comfortable speedup rates at approximately 1.4 the rate of normal speech. He and Gupta (2001) concluded that when people were instructed to assume they were in a hurry, rates of 1.6–1.7 the original speed were acceptable.

Wingfield and Ducharme (1999) conducted a study to investigate possible effects of age and passage difficulty on listening-rate preferences. They found that older adults preferred significantly slower rates than did younger adults. Both groups preferred slower rates for difficult passages (as measured by cloze predictability) than for easy passages. The researchers concluded that both age groups were equally effective in their ability to monitor the difficulty and adjust the rate.

There are some interesting affective factors to consider when using compressed speech. Listener attitudes toward the speaker are improved significantly (Maclachlan, 1982). Maclachlan notes that people associate fast, fluent speech with confidence, knowledge, and enthusiasm. Because attitude learning is influenced strongly by feelings toward the speaker, compressed speech may be appropriate in such situations. Also, Boyle (1969) found a preference for listening over reading in young students (under about 14 years old), presumably because of their slow reading rates. Likewise, college students prefer listening to compressed speech over normal recording (Short, 1977) apparently because of the time savings.

35.7.5 Training

It has often been speculated that practice might influence comprehension of compressed speech. Voor and Miller (1965) exposed subjects to five listening sessions at 380 wpm. Test scores indicated that comprehension increased as a function of exposure up to 7 min, and remained constant thereafter. Orr et al. (1965) exposed blind subjects to listening material presented initially at 325 wpm and increased at 25-wpm intervals to a rate of 475 wpm. Subjects were tested for comprehension at 475 wpm and compared to equivalent pretraining test scores. An improvement of 29.3% was noted. Friedman, Orr, Freedle, and Norris (1966) compared the comprehension scores of subjects given 35 hours of massed practice with the test scores of subjects given 14 to 21 hours of distributed practice in listening to compressed speech. The authors concluded that the comprehension of the distributed-practice group was as good as or better than the comprehension demonstrated by the mass-practiced group. Duker (1974) suggested that gradually increasing the words per minute rate might have some benefit on comprehension of compressed speech. Klavon (1975) tested this idea, without effect, in an attempt to provide a controlled transition period. In general, studies have found that although no particular method of training or practice appears to be any more effective than another, even small exposure to compressed speech can improve comprehension (Foulke, Amster, Bixler, & Nolan, 1962; Friedman, Orr, & Norris, 1966).

35.7.6 Applications

Time-compressed speech has application in voice mail systems, training materials for business and industry, and instruction for populations with special needs. Speech compressor/expanders are also used to speed up or slow down foreign languages in language labs and to normalize speech from a voice recognition program. Given the research on the stability of the comprehension levels when speech is presented at an increased pace, the ease with which speech can be compressed, and the amount of speech used in communication and instruction, one might assume that time compression is a common practice. In contrast, Arons (1992) states, "While the utility of time compressing recordings is generally recognized, surprisingly, its use has not become pervasive" (p. 169). Most audio used in education is available at one speed—normal.

One factor that may increase the use of time-compressed audio is the ability to select a playback speed for streaming audio on the web. For example, IBM's Web lecture interface has a slider that allows the user to adjust the speed of the narration. Similar approaches are being implemented for streaming media by Microsoft and Real Networks (He & Gupta, 2001). For example the 2xAV plug-in from Real allows the user to play audio (by moving a slider) at from 0.33 to 2.5 the normal speed (Enounce, 2002); the Microsoft encoder allows users to adjust pause removal and other parameters, which reduces the playback time (Microsoft Corporation, 2002).

The ability to skim or scan audio is another application for time-compressed speech. As more and more companies and universities are making audio-enhanced lectures, webcasts, and digital videos available, there needs to be a way to scan the contents and obtain the pertinent information. When skimming audio, 100% comprehension is not essential; therefore, higher compression rates can be used (Arons, 1997; Harrigan, 2000; He & Gupta, 2001).

35.7.7 Efficiency

The original impetus for speech compression was potential efficiency. However, the instructional implications of using compressed speech for efficiency are somewhat contradictory. When the time saved in compression was used to elaborate certain parts of the text, comprehension for that part of the text increased (Fairbanks et al., 1957). However, as Sticht (1971) pointed out, the time saved in compression was lost in elaboration, and overall comprehension was not improved. Schramm (1972) reported similar results. However, in a recent research study conducted by Omoigui et al. (1999), a task-time savings of 22% was reported when users were allowed to select the most comfortable compression rate; however, comprehension was not measured.

35.7.8 Summary

Research studies have confirmed the hypothesis that listeners can process information at a much higher rate than normal conversational speech. Usually, any short exposure to compressed speech will result in improved comprehension. In general, few differences were detected for ages between about 12 and 50. However, some differences in the ability to comprehend compressed speech may be due to aptitude or verbal ability. Compressed speech may be preferred to normal speech because of the time efficiency. However, the early hopes that time-compressed speech would be fully utilized appear to be unjustified. As interfaces with increased user control are implemented, it is anticipated that time-compression will become more prevalent. For additional information on time-compressed speech (prior to 1972), see the three-volume series, *Time-Compressed Speech: An Anthology and Bibliography,* by Duker (1974).

35.8 RECOMMENDATIONS

35.8.1 Introduction

Sound and speech have existed for thousands of years. However, the technologies that can record, store, and distribute audio have only been available in the last century. Research related to auditory instruction has varied over the years. When audio technologies were first introduced into schools, a series of evaluation studies was conducted to determine whether or not students could learn from a particular technology (Kerr, 1999). Media comparison studies were also quite common—comparing the effectiveness of audio vs. visual presentations in various subject areas and age levels. Another thread of research has focused on the theories that serve as foundations for auditory memory and multichannel communications. More recently, a number of studies have been conducted to determine optimal techniques and strategies for incorporating an audio component in multimedia instruction. This chapter has presented many studies that were conducted in a variety of settings, using different metrics and research designs.

Throughout the literature, several researchers have remarked on the lack of substantiated research related to audio instruction (Barron & Atkins, 1994; Beccue, Vila, & Whitley, 2001; Bishop & Cates, 2001; Thompson et al., 1996). Aarntzen (1993) notes that although multimedia designers and developers are "inclined to use audio (and speech in particular)... hardly anything is known about the use of audio in instruction with respect to the processing of visual and auditive information, learner characteristics, events of instruction, and audio characteristics" (p. 354).

Now that audio components are a feasible, and often expected, component of instruction, researchers in educational technology should continue to provide guidelines for designers that are based on empirical research and cognitive theory.

35.8.2 Recommendations for Future Research

Throughout this chapter, information has been provided about research findings and implications. This section serves as a finale to the chapter by presenting a synopsis of a few "calls" for future research.

35.8.2.1 Auditory Processing and Memory. There are many unanswered questions about how auditory information is perceived, processed, stored, and retrieved (Baddeley, 1998). More definitive answers to issues such as modality effects may lead to better understanding of appropriate applications and designs for audio in instruction. "Future research is needed, concentrating on the exact cognitive mechanisms underlying modality effects in memory performance and in which the findings are extended to more natural stimuli, such as prose passages" (Haan, Appels, & Aleman, 2000, p. 582). Other aspects of modality effects that bear investigation include the differential effects of sequential and simultaneous processing of audio and visual information (Duis et al., 1994) and effects on memory for items generated by the subject instead of the researcher (Penney, 1989b).

McAdams and Bigand (1993) point out that

> one of the areas that has not received much attention, with the exception of some studies in young children, is the understanding of the role of auditory cognitive processes and their integration with those of other sensory and more general cognitive systems in everyday activity.... Research programmes addressing some of these issues are sure to demonstrate a more important role for audition in everyday activity than has been granted to this perceptual system to date". (p. 5)

According to Grace-Martin (2001), "Designing instructional materials with multimedia is a bit like walking a tight rope: the designer tries to provide a rich, informational learning experience for the learner while at the same time not exceeding his or her ability to process and assimilate the information" (p. 407). Additional research into factors that impact cognitive load and appropriate multimodal strategies would be very beneficial (Mayer, 2001; Sweller, 1988; Sweller, Merrienboer, & Paas, 1998).

35.8.2.2 Media Selection. With the popularity of blended solutions in e-learning, emphasis is being placed on determining the optimal combinations of media, technologies, and instructional approaches. Audio, in the form of audio conferencing, multimedia, and streaming audio, is an important component of e-learning. As noted by Koumi (1994), the most common current practice is to design instruction without using a media selection model, "in which case, it is no wonder that allocation of media has been controlled more by practical, economic, and human/political factors than by pedagogic considerations" (p. 56). Mann (1995) agrees and points out, "Forty-five years of intuitive combinations of audio-visual information have produced only mixed results" (p. 16). Structural and functional attributes such as those in Mann's structured sound function (SSF) model may be one approach to producing more consistent, effective implementations of audio. Distinctive attributes and symbols related to audio (as mentioned by Salomon, 1994) need to be defined further so that instructional designers and developers can develop and refine the criteria for the deployment of media to best effect (Koumi, 1994).

The literature does not offer much assistance in media selection. Results related to multichannel communication are complex and contradictory. Based on a review of literature related to learning abstract concepts through audiovisual redundancy, Lai (2000) concluded, "It is clear that the results are inconsistent and there are only few instructional design guidelines available for the optimal relationship between audiovisual redundancy and ability level" (p. 278). Lang (1995) had a similar conclusion about redundancy research when she stated, "Forty years of research has yielded a hodgepodge of contradictory conclusions" (p. 86). Still unanswered are questions such as when, where, and how audio should be integrated with other media.

There are obviously numerous variables that must be considered when deciding whether or not to use audio, with or without additional media (Sutcliffe, 1999). In some cases, the selection of audio as an instructional medium is self-evident. For example, audio is essential to learn a foreign language, study music, or discriminate sound effects such as heartbeats or sonar signals. Audio may also be the optimal (and, in most cases, the only) solution for interacting with distant students, especially in geographically remote or economically disadvantaged areas. In addition, children who have not learned to read and adults who are unable to read at a functioning level can benefit from audio instruction (Beccue, Vila, & Whitley, 2001).

The issue of where and how to use audio in multimedia programs is considerably more complicated. In literature and textbooks that focus on the design and development of multimedia, there is a great deal more attention given to visuals than to audio (Jaspers, 1991). In fact, it is almost impossible to find more than a cursory reference (if any) to design guidelines for incorporating audio into interactive programs in many current books (see, for example, Lee & Owens, 2000; Horton, 2000; Kruse & Keil, 2000; Allen, 2003).

The situation is exacerbated when the few guidelines that exist are often contradictory. For example, Gibbons and Fairweather (1998) advised designers to "avoid echoing text with audio" (p. 375); Clark and Mayer (2003) agree, recommending that designers "avoid e-learning courses that contain redundant on-screen text presented at the same time as on-screen graphics" (p. 100). However, Aarntzen (1991) stated that "when speech is used as the mainstream provider of information it should be accompanied by that same text on the computer's screen" (p. 363). Alessi and Trollip (2001) acknowledge that although including both speech and text helps to meet the needs of students with visual or auditory impairment, the best approach may be to "provide controls that encourage learners to use *either* the text or audio presentation, not both at the same time" (p. 75).

Although the addition of audio in computer-based training has not consistently shown a significant increase in achievement, there are other advantages to audio-enhanced courseware, such as realism and motivation. Some results, such as Mayer's (2001) research on the effectiveness of using audio with animations, offer a good beginning to the types of research-based guidelines that are needed by instructional designers. Mayer, cautions, however, that design principles based on research "must be qualified with respect to different kinds of learners. Additional research is needed to pinpoint the role of individual differences in multimedia learning" (p. 189).

Shih and Alessi (1996) seem to agree with Mayer about the need for qualifications in media research. They stated, "Media is a tool or a learning environment; we educators or researchers must have a better understanding of it before we can wisely use and design it. Therefore, further and multiple dimensions of research on using audio (sound or voice) in education and multimedia are suggested" (p. 217).

35.8.3 Conclusion

A wide range of research is relevant for auditory instruction. As new technologies mature, such as voice recognition, interactive agents, wearable computers, virtual reality, and PDAs, even more tools that can deliver auditory instruction will be available. In his article "Multimedia Learning: Are We Asking the Right Questions?" Mayer (1997) noted, "At this time, the technology for multimedia education is developing at a faster pace than a corresponding science of how people learn in multimedia environments" (p. 4). From the experimental psychologists who are investigating the way in which auditory information is processed and stored to the instructional designers who are seeking to structure optimal environments and strategies for learners, pertinent research areas abound.

Audio technologies offer powerful tools for educators; tools that can entice, motivate, persuade, inform, reinforce, and reward. Sound is a natural part of our lives and environment; it should also be a natural part of our education.

References

Aarntzen, D. (1993). Audio in courseware: Design knowledge issues. *Educational and Training Technology International, 30*(4), 354–356.

Access Science (2003). *McGraw-Hill encyclopedia of science and technology*. Retrieved April 29, 2003, from http://www.accessscience.com

Adelson, L. (1975). Comprehension by college students of time compressed lectures. *Journal of Experimental Education, 44,* 53–59.

Allen, M. (2003). *Guide to e-learning: Building interactive, fun, and effective learning programs for any company*. Hoboken, NJ: John Wiley & Sons, Inc.

Alten, S. R. (2002). *Audio in media*. 6th edition. Belmont, CA: Wadsworth/Thomson Learning.

American Heritage Dictionary. (1982). Boston: Houghton Mifflin Company.

Anderson-Inman, L. (1990). Enhancing the reading-writing connection: Classroom applications. *Writing Notebook, 7*(3), 12–15.

Andres, H. P., & Petersen, C. (2001/2002). Presentation media, information complexity, and learning outcomes. *Journal of Educational Technology Systems, 30*(3), 225–246.

Arons, B. (1992). *Techniques, perception, and applications of time-compressed speech*. Proceedings of 1992 Conference, American Voice I/O Society, Sept., (pp. 169–277). Available from http://xenia.media.mit.edu/~barons/avios92.html

Arons, B. (1994). *Efficient listening with two ears: Dichotic time compression and spatialization*. Proceedings of the Second International Conference on Auditory Display, Santa Fe, NM, (pp. 171–177). Available from http://lucy.media.mit.edu/~barons/icad94.html

Arons, B. (1997). SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction, 4*(1), 3–38.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* Vol. 2 (pp. 89–195). New York: Academic Press.

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556–559.

Baddeley, A. D. (1998). *Human memory: Theory and practice*. Boston: Allyn and Bacon.

Baddeley, A. D. (1999). *Essentials of human memory*. East Sussex, UK: Psychology Press Ltd.

Barron, A., & Varnadoe, S. (1992). Digital audio: A sound design element. *Instruction Delivery Systems, 6*(1), 6–9.

Barron, A. E., & Atkins, D. (1994). Audio instruction in multimedia education: Is textual redundancy important? *Journal of Educational Multimedia and Hypermedia, 3*(3–4), 295–306.

Barron, A. E., & Kysilka, M. L. (1993). The effectiveness of digital audio in computer-based training. *Journal of Research on Computing in Education, 25,* 277–289.

Barron, A. E., Orwig, G. W., Ivers, K. S., & Lilavois, N. (2002). *Technologies for education: A practical guide*. Englewood, CO: Libraries Unlimited.

Basil, M. D. (1992). *Attention to and memory for audio and video information in television scenes*. Paper presented at the Annual Meeting of the International Communication Association, Miami, FL, (pp. 1–52). (ERIC Document ED 347 592)

Bates, A. W. (1983). Adult teaming from educational television: The open university experience. In M. J. A. Howe (Ed.), *Learning from television: psychological and educational research* (213–27). Washington, DC: American Psychological Association.

Beasley, D. S., & Maki, J. E. (1976). Time- and frequency-altered speech. In N. J. Lass Ed.), *Contemporary issues in experimental phonetics* (Ch. 12, pp. 419–458). New York: Academic Press.

Beccue, B., Vila, J., & Whitley, L. K. (2001). The effects of adding audio instructions to a multimedia computer based training environment. *Journal of Educational Multimedia and Hypermedia, 10*(1), 47–67.

Bell, R. (1969). *An analysis of certain elements of an audio-tape approach to instruction*. Unpublished doctoral dissertation, University of Washington.

Benz, C. R. (1971). *Effects of time compressed speech upon the comprehension of a visual oriented television lecture*. Unpublished doctoral dissertation, Wayne State University.

Bishop, M. J., & Cates, W. M. (2001). Theoretical foundations for sound's use in multimedia instruction to enhance learning. *Educational Technology Research and Development, 49*(3), 5–22.

Boring, E. G. (1950). *A history of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

Boyle, V. A. (1969). Visual stimulation and comprehension of compressed speech. *Dissertation Abstracts International, 30,* 5221B.

Braden, R. A. (1996). Visual literacy. In D. Jonassen (Ed.), *The handbook of research for educational communications and technology.* Washington, DC: AECT.

Breed, P. A. (1977). *The relative effect of the controlled reader and the speech compressor on reading rate and comprehension*. Unpublished doctoral dissertation, Northern Illinois University.

Broadbent, D. (1958). *Perception and communication*. New York: Pergamon Press.

Button, G. E. (1991). Audio-tutorial biology, andragogy, and self-esteem: Relationships among independent and dependent variables. *Dissertation Abstracts International, 53*/02, 457. (University Microfilms No. AAC).

Campeau, P. L. (1966). Selective review of the results of research on the use of audio visual media to teach adults. In L. J. Briggs, P. L. Campeau, R. M. Gagne, & M. A. May (Eds.), *Instructional media: A procedure for the design of multi media instruction and suggestions for future research,* Monograph 2 (pp. 99–176). Pittsburgh, PA: American Institutes for Research.

Carver, R. P. (1973). Understanding, information processing, and teaming from prose materials. *Journal of Educational Psychology, 64,* 76–84.

Cherry, C. (1953). Some experiments on the recognition of speech with one and two ears. *Journal of the Acoustical Society of America, 23,* 915–919.

Chu, G. D., & Schramm, W. (1968). *Learning from television: What the research says.* Palo Alto, CA: Institute for Communication Research, Stanford University.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210.

Clark, R. C., & Mayer, R. E. (2003). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco, CA: Jossey-Bass Publishers.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.

Clark, R. E. (1994). Media will never influence learning. *Educational Training Research and Development, 42*(2), 21–29.

Clark, R. E., & Salomon, G. (1986). Media in teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.

Constantine, M. (1964). Radio in the elementary school. *Science Education, 48,* 121–32.

Cook, H. R. (1964). *The effects of learning of structural drills in Spanish broadcast via high frequency AM radio* (NDEA Title VII Project No. 1018). Bloomington, IN. Indiana University.

Covell, M., Withgott, M., & Slaney, M. (1998, May). *Mach1: Nonuniform time-scale modification of speech.* Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Seattle WA. Available from http://www.slaney.org/covell/interval/1997-061/writeup.html

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory and Cognition, 21,* 142–145.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.

Cutler, R. L., McKeachie, W. J. & McNeil, E. B. (1958). Teaching psychology by telephone. *The American Psychologist, 13*(9), 551–52.

Dahlem, W. E. (1985). The effects of supplemental auditory information on the reading comprehension performance of learning disabled, high school students. *Dissertation Abstracts International, 46,* 2396A.

Daniels, L. (1995). *Audio-vision: Audio-visual interaction in desktop multimedia.* Annual Conference of the International Visual Literacy Association. Tempe, AZ, October 12–16.

Diehl, C. F., White, R. D., & Burke, K. (1959). Rate and comprehension. *Speech Monographs, 26,* 229–232.

Duis, S. S., Dean, R. S., & Derks, P. (1994). The modality effect: A result of methodology? *International Journal of Neuroscience, 78,* 1–7.

Duker, S. (1974). *Time compressed speech*. Metuchen, NJ: Scarecrow.

Eckhardt, W. X. (1970). *Learning in multi-media programmed instruction as a function of aptitude and instruction rate controlled by compressed speech*. Unpublished doctoral dissertation, University of Southern California.

Efron, R. (1970). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia, 8,* 37–55.

Elliott, L. L. (1970). Pitch memory for short tones. *Perception and Psychophysics, 8,* 379–384.

Engle, R. (1996). Working memory and retrieval: An inhibition-resource approach. In J. Richards (Ed.), *Working memory and human cognition* (pp. 89–119). New York: Oxford University Press.

Engle, R. W., & Mobley, L. A. (1976). The modality effect: What happens in long-term memory. *Journal of Verbal Learning and Verbal Behavior, 15,* 519–528.

Enounce Inc. (2002). *2xAV Plug-in for RealPlayer*. Available from http://www.enounce.com/products/real/2xav/index.shtml

Erickson, C. I., & King, I. (1917). A comparison of visual and oral presentations of lessons in the case of pupils from third to ninth grade. *School and Society, 6*(136), 146–48.

Erickson, C. W., & Johnson, H. J. (1964). Storage and decay characteristics of nonattended auditory stimuli. *Journal of Experimental Psychology, 8, 28–36.*

Eshgh, R., Hoxeng, J., Provenzano, J. & Casals, B., eds. (1988). *Radio-assisted community basic education (RADECO).* Pittsburgh, PA: Duquesne University Press.

Estes, W. K. (1989). Learning theory. In A. Lesgold & R. Glaser (Eds.), *Foundations for a Psychology of Education*. Hillsdale, NJ: Lawrence A. Erlbaum Associates.

Fairbanks, G., Everitt, W. L., & Jaeger, R. P. (1954). Method for time or frequency compression-expansion of speech. *Transactions of the Institute of Radio Engineers, Professional Groups on Audio, AU 2,* 7–12.

Fairbanks, G., Guttman, N., & Miron, M. S. (1957). Auditory comprehension of repeated high speech messages. *Journal of Speech and Hearing Disorders, 22,* 23–32,

Fang, I., & Ross, K. (1996). *Media history project: Timeline by chronology*. Minneapolis, MN: School of Journalism and Mass Communication. Retrieved April 29, 2003, from http://www.mediahistory.umn.edu/index2.html.

Fergen, G. K. (1954). *Listening comprehension at controlled rates for children in grades iv, v, and vi*. Unpublished doctoral dissertation, University of Missouri.

Fletcher, J. D. (1990). *The effectiveness of interactive videodisc instruction in defense training and education* (IDA paper P-2372).

Arlington, VA: Institute for Defense Analyses, Science and Technology Division.

Foulke, E. (1965). *A survey of the acceptability of rapid speech* (Tech. Rep. Project No. 2430). Cooperative Research Branch, Dept. of HEW, Office of Education. Louisville, KY: Performance Research Laboratory, U. Louisville.

Foulke, E. (1968). Listening comprehension as a function of word rate. *Journal of Communication, 18,* 198–206.

Foulke, E. (1969). *Comprehension of rapid speech by the blind: Part III. Final progress report on cooperative research project no. 2430 covering the period from March 1, 1964 to June 30, 1968*. Louisville, KY: U. Louisville.

Foulke, E., Amster, C. H., Bixler, R. H., & Nolan, C. X. (1962). The comprehension of rapid speech by the blind. *Exceptional Children, 11,* 134–141.

Foulke, E. A. (1962, Mar.). *A comparison of two methods of compressing speech*. Symposium at the Southeastern Psychological Association, Louisville, KY.

Foulke, E. A., & Sticht, T. G. (1967). *The intelligibility and comprehension of accelerated speech*. Proceedings of the Louisville Conference on Time Compressed Speech, Louisville, KY, (pp. 21–28).

Foulke, E. A., & Sticht, T. G. (1969). In *Proceedings of the second Louisville conference on rate and/or frequency controlled speech*. Louisville, KY. University of Louisville. (ERIC Document Reproduction Service No. Ed 61682)

Foulke, E. A. & Sticht, T. G. (1969). Review of research on the intelligibility and comprehension of accelerated speech. *Psychological Bulletin, 72,* 50–62.

Freeman, M. W., Grimes, L. W., & Holliday, J. R. (2000). Increasing access to learning with hybrid audio-data collaboration. *Educational Technology & Society, 3*(3), 112–121.

Friedman, H. L., Orr, D. B., Freedle, R. O., & Norris, C. M. (1966). *Further research on speeded speech as an educational medium* (Report No. AIR-E-50-7-66-TR-3). Silver Spring, MD: American Institute for Research in the Behavioral Sciences. (ED 044 903)

Friedman, H. L., Orr, D., & Norris, C. (1966). *Further research on speeded speech as an educational medium-the use of listening aids* (Report No. 3.). Silver Spring, MD: American Institute for Research in Behavioral Sciences. (ED 044 903)

Furnham, A., & Gunter, B. (1989). The primacy of print: Immediate cued recall of news as a function of the channel of communication. *Journal of General Psychology, 116*(3), 305–310.

Furnham, A., Gunter, B., & Green, A. (1990). Remembering science: The recall of factual information as a function of the presentation mode. *Applied Cognitive Psychology, 4*(3), 203–212.

Furnham, A., Proctor, E., & Gunter, B. (1988). Memory for material presented in the media: The superiority of written communication. *Psychological Reports, 63*(3), 935–938.

Garner, R., Brown, R., Sanders, S., & Menke, D. (1992). "Seductive details" and learning from text. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 239–254). Hillsdale, NJ: Lawrence A. Erlbaum Associates.

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.

Garvey, W. D. (1953). The intelligibility of speeded speech. *Journal of Experimental Psychology, 45,* 102–108.

Gaver, W. W. (1997). Auditory interfaces. In M. G., Landauer, T. K., and Prabhu, P. V. (Eds.), *Handbook of Human-Computer Interaction.* Helander, NY: Elsevier.

Gelder, B., & Vroomen, J. (1997). Modality effects in immediate recall of verbal and non-verbal information. *Psychology Press, 9*(1), 97–110.

Gibson, R. (1958). Tape recordings are used to teach seventh grade students in Westside Junior-Senior High School, Omaha, NE. *Bulletin of the National Association of Secondary Principals, 42,* 81–93.

Gibson, R. (1959). The tape recordings experiment is expanded in Westside Junior-Senior High School, Omaha, NE. *Bulletin of the National Association of Secondary Principals, 43,* 49–72.

Gibson, R. (1960). Final report on the Westside High School teaching by-tape project. *Bulletin of the National Association of Secondary Principals, 44,* 56–62.

Glenberg, A. M., Eberhardt, K. A., & Belden, T. M. (1987). The role of visual interference in producing the long-term modality effect. *Memory and Cognition, 15*(6), 504–510.

Goetz, E. T., & Sadoski, M. (1995a). The perils of seduction revisited: A reply to Wade, Alexander, Schraw, and Kulikowich. *Reading Research Quarterly, 30*(3), 518–519.

Goetz, E. T., & Sadoski, M. (1995b). Commentary: The perils of seduction: Distracting details or incomprehensible abstractions? *Reading Research Quarterly, 30*(3), 500–511.

Goldstein, H. (1940). Reading and listening comprehension at various controlled rates. *Teacher College Contributions to Education,* No. 821.

Goolkasian, P. (2000). Pictures, words, and sounds: From which format are we best able to reason? *Journal of General Psychology, 127*(4), 439–459.

Gordon, G. N. (1969). *The languages of communication*. New York: Hastings House.

Gordon-Salant, S., & Fitzgibbons, P. J. (2001, August). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research, 44*(4), 709–719.

Grace-Martin, M. (2001). How to design educational multimedia: A "loaded" question. *Journal of Educational Multimedia and Hypermedia, 4,*(10) 397–409.

Gunter, B., Furnham, A., & Leese, J. (1986). Memory for information from a party political broadcast as a function of channel of communication. *Social Behaviour, 1,* 135–142.

Gunter, G., & Furnham, A. (1986). Sex and personality differences in recall of violent and non-violent news from three presentational modalities. *Personality and Individual Differences, 7,* 829–837.

Guttman, N., & Julesz, B. (1963). Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America, 35,* 610.

Haan, E. H. F., Appels, G., & Aleman, A. (2000). Inter- and intramodal encoding of auditory and visual presentation of material: Effects on memory performance. *The Psychological Record, 50*(3), 577–586.

Hannafin, M. & Hooper, S. (1993). Learning principles. In Fleming M., & Levie, H. (Eds.). *Instructional message design: Principles from the behavioral and cognitive sciences*. Englewood Cliffs, NJ: Educational Technology Publications.

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.

Harrigan, K. (1995, Spring). The SPECIAL System: Self-paced education with compressed interactive audio listening. *Journal of Research on Computing in Education, 27*(3), 361–370.

Harrigan, K. (2000). The SPECIAL system: Searching time-compressed digital video lectures. *Journal of Research on computing in Education, 33*(1), 77–86.

Hartman, F. R. (1961a, January/February). Recognition learning under multiple channel presentation and testing conditions. *AV Communication Review, 9*(1), 24–43.

Hartman, F. R. (1961b, November/December). Single and multiple channel communications: A review of research and a proposed model. *AV Communication Review, 9*(6), 235–262.

Hawkins, H. and Presson, J. (1986). Auditory information processing. In K. R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance.* New York: Wiley.

He, L., & Gupta, A. (2001). *Exploring benefits of non-linear time compression*. Proceedings of ACM International Conference on Multimedia, Ottawa, Canada, (pp. 382–391).

Heiman, G. W., Leo, R. J., Leighbody, G., & Bowler, K. (1986). Word intelligibility decrements and the comprehension of time-compressed speech. *Perception and Psychophysics, 40*(6). 407–411.

Helander, M. G., Landauer, T. K., & Prabhu, P. F. (1997). *Handbook of human-computer interaction.* Amsterdam: Elsevier.

Hidi, S. (1990). Interest and its contribution as a mental resource for Learning. *Review of Educational Research, 60*(4), 549–571.

Hoban, C. F., & Van Ormer, E. B. (1970). *Instructional film research 1918–1950.* New York: Arno Press.

Horton, W. (2000). *Designing web-based training: How to teach anyone anything anywhere anytime*. New York: John Wiley & Sons, Inc.

Hsia, H. J. (1968a). On channel effectiveness. *AV Communication Review, 16,* 245–267.

Hsia, H. J. (1968b, December). Output, error, equivocation and recalled information in auditory, visual and audiovisual information processing. *Journal of Communication, 18,* 325–353.

Hsia. H. J. (1971, Spring). The information processing capacity of modality and channel performance. *AV Communication Review, 19*(1), 51–75.

Huang, J. C. (1995). *Digitized speech as feedback on cognitive aspects of psychomotor performance during computer-based instruction.* Proceedings of the 1995 Annual National Convention of the Association for Educational Communications and Technology. Anaheim, CA.

Hurlburt, R. T. (2001). "Lectlets" deliver content at a distance; Introductory statistics as a case study. *Teaching of Psychology, 28*(1), 15–20.

Ingebritsen, T. S., & Flickinger, K. A. (1998). *Development and assessment of web courses that use streaming audio and video technologies*. *Distance Learning '98.* Proceedings of the Annual conference on Distance Teaching & Learning, Madison, WI, August 5–7.

Ives, W. (1992). Evaluating new multimedia technologies for self-paced instruction. *Evaluation and Program Planning, 15,* 287–295.

Janse, E., Nooteboom, S., & Quene, H. (2001, August). *Word-level intelligibility of time-compressed speech: Prosodic and segmental factors*. The Netherlands: Utrecht Institute of Linguistics. Available from http://www.let.uu.nl/~Sieb.Nooteboom/personal/timecompressed.pdf

Jasper, F. (1991). The relationship sound-image. *International Journal of Instructional Media, 18*(2), 161–174.

Jaspers, A. (1995). Focus on audio in communication and instruction: An overview. *Educational Media International, 32,* 176–181.

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17,* 329–348.

Jonassen, D. H., Campbell, J. P., & Davidson, M. E. (1994). Learning with media: Restructuring the debate. *Educational Training Research and Development, 42*(2), 45–48.

Jonassen, D. H., & Grabowski, B. L. (1993). *Handbook of individual differences, learning, and instruction*. Hillsdale, NJ: Lawrence A. Erlbaum Associates.

Jongekrijg, T. & Russell, J. D. (1999). Alternative techniques for providing feedback to students and trainees: A literature review with guidelines. *Educational Technology, 39*(6), 54–58.

Junor, L. (1992). Teaching by tape: Some benefits, problems, and solutions. *Distance Education, 13*(1), 93–107.

Kahneman, D. (1973).*Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kalyuga, S. (2000). When using sound with a text or picture is not beneficial to learning. *Australian Journal of Educational Technology, 16*(2), 161–172. Available from http://www.ascilite.org/ajet/ajet16/kalyuga.html

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1–17.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13,* 351–372.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*(1), 126–136.

Kearsley, G. (2001). *Dual coding theory (A. Paivio). Theory into practice.* Retrieved July 28, 2002, from http://tip.psychology.org/paivio.html

Kerr, B. (1999). *Effective use of audio media in multimedia presentation*. Proceedings of the Mid-South Instructional Technology Conference, Murfreesboro, TN, March 28–30, (pp. 1–10).

Klavon, A. J. (1975). *Time-compressed lecture: an alternative for increased teacher learner interaction*. Unpublished doctoral dissertation, University of Maryland.

Kohlfeld, D. L. (1971). Simple reaction time as a function of stimulus intensity in decibels of light and sound. *Journal of Experimental Psychology, 88,* 251–257.

Koroghlanian, C. M., & Klein, J. D. (2000). *The use of audio and animation in computer-based instruction*. Paper presented at the National Convention of the Association for Educational Communications and Technology. Denver CO, October 25–28.

Koroghlanian, C. M., & Sullivan, H. J. (2000). Audio and text density in computer-based instruction. *Journal of Educational Computing Research, 22*(2), 217–230.

Koumi, J. (1994). Media comparison and deployment: A practitioner's view. *British Journal of Educational Technology, 25*(1), 41–57.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179–211.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Training Research and Development, 42*(2), 7–19.

Kramer, E. J., & Lewis, T. R. (1951). Comparison of visual and non-visual listening. *Journal of Communication, 1*(2), 16–20.

Kruse, K., & Keil, J. (2000). *Technology-based training: The art and science of design, development, and delivery*. San Francisco: Jossey-Bass/Pfeiffer.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P.A. (1979). Research on audio-tutorial instruction: a meta-analysis of comparative studies. *Review of Educational Research, 11,* 321–341.

Kumata, H. (1961). *History and progress of instructional television research in the United States*. Report presented at the International Seminar on Instructional Television, Oct. 8–18, Purdue University, Lafayette, IN.

Lai, S. (2000). Increasing associative learning of abstract concepts through audiovisual redundancy. *Journal of Educational Computing Research, 23*(3), 275–289.

Lang, A. (1995). Defining audio/video redundancy from a limited-capacity informaitn processing perspective. *Communications Research, 22*(1) 86–115.

Larsen, R. P., & Feder, D. D. (1940). Common and differential factors in reading and hearing comprehension. *Journal of Educational Psychology, 31,* 241–251.

Lee, D. (1994). *Audio-vision: Audio-visual interaction in desktop multimedia.* Selected Readings from the Annual Conference of the International Visual Literacy Association. Tempe AZ, October 12–16, (pp. 1–8).

Lee, W. W., & Owens, D. L. (2000). *Multimedia-based instructional design*. San Francisco: Jossey-Bass/Pfeiffer.

Lehrman, P. D., & Tully, J. (1991, October). Catch a wave: Digital audio. *MacUser,* 94–103.

Letowski, T., & Poch, N. (1996). Comprehension of time-compressed speech: Effects of age and speech complexity. *Journal of American Academy of Audiology, 7,* 447–457.

Li, F., Gupta, A., Sanocki, E., He, L., & Yong, R. (1999, September). *Browsing digital video*. Proceedings of ACM Conference Computer Human Interaction 2000, pp. 169–176. Available from ftp://ftp.research.microsoft.com/pub/tr/tr-99-67.doc

Loder, L. E. (1937). A study of aural learning with and without the speaker present. *Journal of Experimental Education, 6,* 46–60.

Ludrick, L. A. (1974). *A study of the effects of controlled delivery instruction upon the achievement of college students using compressed speech audio and television pictorials*. Unpublished doctoral dissertation, the University of Oklahoma.

Maclachlan, J. (1982). Listener perception of time-compressed spokespersons for radio commercials. *Journal of Advertising, 22,* 47–51.

Main, R. E., & Griffiths, B. (1977). Evaluation of audio and pictorial instructional supplements. *Educational Communication & Technology, 25*(2), 167–179.

Mann, B. L. (1995, Summer). Focusing attention with temporal sound. *Journal of Research on Computing in Education, 27*(4), 402–424.

Mann, D., Shakeshaft, C., Becker, J., & Kottkamp, R. (1999). *West Virginia's basic skills/computer education program: An analysis of student achievement*. Santa Monica, CA: Milken Family Foundation.

Massaro, D. (1975). Backward recognition masking. *Journal of the Acoustical Society of America, 58,* 1059–1064.

Massaro, D. W., & Loftus, G. R. (1996).*Sensory and perceptual storage.* In E. L. Bjork & R. A. Bjork (Eds.), *Memory,* San Diego, CA: Academic Press.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a 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., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198.

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., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91,* 638–643.

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, 84,* 389–401.

McAdams, S., & Bigand, E. (Eds.) (1993). *Thinking in sound: The cognitive psychology of human audition.* Oxford: Clarendon Press.

Menne, J. W., Klingenschmidt, L. E., & Nord, D. L. (1969, Mar.). *The feasibility of using taped lectures to replace class attendance*. Paper presented at the annual meeting of American Educational Research Association, Los Angeles.

Microsoft Corporation. (2002). *Windows media technologies: Encoder 7.0*. Available from http://www.microsoft.com/windows/windowsmedia/wm7/encoder.asp

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Miller, G. A., & Lichlinder, J. C. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America, 22,* 167–173.

Moore, D. M., Myers, R. J., & Burton, J. K. (1996). Multiple-channel communication: The theoretical and research foundations of multimedia. In Jonassen, D. (Ed.), *Handbook of research on communication and educational technology*. Washington, DC: Association for Educational Communications and Technology.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358–368.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia messages. *Journal of Educational Psychology, 92,* 117–125.

Morra, S. (2000). A new model of verbal short-term memory. *Journal of Experimental Child Psychology, 75,* 191–227.

Morrison, G. R. (1994). The media effects questions: "Unresolvable" or asking the right question. *Educational Training Research and Development, 42*(2), 41–44.

Mosenthal, P. (1976–77). Psycholinguistic properties of aural and visual comprehension as determined by children's abilities to comprehend syllogisms. *Reading Research Quarterly, 12,* 55–92.

Motion-Picture Technology. (2003). *Encyclopedia Britannica*. Retrieved April 29, 2003, from http://search.eb.com/eb/article?eu=119928

Mousivi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87,* 319–334.

Murdock, B. B., Jr., & Walker, K. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 665–676.

Nasser, D. L., & McEwen, W. J. (1976). The impact of alternate media channels: recall and involvement with messages. *AV Communication Review, 24,* 263–272.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Nelson, H. E. and Moll, K. R. (1950) Comparison of audio and video elements of instructional films. Technical Report No. SDC 269–17–18. Instructional Film Program. Naval Training Devices Centre-Office of Naval Research: Port Washington.

NHK Radio-Television Cultural Research Institute. (1956). *The listening effect of radio English classroom*. Tokyo: NHK.

Nichols, R. G., & Stevens, L. A. (1957). *Are you listening?* New York: McGraw-Hill.

Ninno, A. (1999). Radios in the classroom: Curriculum integration and communication skills. *ERIC Digest*. Syracuse, NY: ERIC Clearinghouse on Information and Technology.

Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addison-Wmesley.

Novak, J. D. & Musunda, D. (1991). A twelve-year longitudinal study of science concept learning. *American Educational Research Journal 28,* 117–53.

Nugent, G. C. (1982). Pictures, audio, and print: symbolic representation and effect on learning. *Educational Communication and Technology Journal, 30,* 163–174.

Olgren, C. H. (1997). Teaching by telephone. *New Directions for Teaching and Learning, 71,* 59–66.

Olgren, C. H., & Parker, L. A. (1983). *Teleconferencing technology and applications*. Dedham, MA: Artech House.

Olson, G. M. (1989). Intellectual development. In A. Lesgold & R. Glaser (Eds.), *Foundations for a Psychology of Education.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Omiogui, N., He, L., Gupta, A., Grudin, J., & Sanocki, E. (1999). *Time-compression: Systems concerns, usage, and benefits*. Available from http://www.research.microsoft.com/research/coet/Compression/MM2001/paper.doc

Oppenheimer, T. (July 1997). The computer delusion. *The Atlantic Monthly*. Available from http://www.theatlantic.com/atlantic/issues/97jul/computer.htm

Orr, D. B., & Friedman, H. L. (1964). *Research on speeded speech as an educational medium*. (Progress Report, Grant No. 7–48–7670–203.) Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Orr, D. B., Friedman, H. L., & Williams, J. D. (1965). Trainability of listening comprehension of speeded discourse. *Journal of Educational Psychology, 56,* 148–156.

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford, England: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287.

Palmer, S. (2000). Working memory: A developmental study of phonological recoding. *Memory, 8*(3), 179–193.

Parshall, C. G. (1999). *Audio CBTs: Measuring more through the use of speech and non-speech sound*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April 19–23.

Pearce, C. G., & Ackley, R. J. (1995). Audiotaped feedback in business writing: An exploratory study. *Business Communication Quarterly, 58*(3), 31–34.

Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin, 82,* 68–84.

Penney, C. G. (1989a). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17*(4), 398–422.

Penney, C. G. (1989b). Modality effects in delayed free recall and recognition: Visual is better than auditory. *The Quarterly Journal of Experimental Psychology, 17*(4), 398–422.

Popham, W. J. (1962). Tape recorded lectures in the college classroom 11. *AV Communication Review, 10,* 94–101.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review, 83,* 157–171.

Postlethwaite, S. N. (1970). The audio-tutorial system. *American Biology Teacher, 32,* 31–33.

Postlethwaite, S. N. (1972). The audio-tutorial system: incorporating mini course and mastery. *Educational Technology, 12*(9), 35–37.

Postlethwaite, S. N., Novak, L & Murray, H. T., Jr. (1972). *The audio-tutorial approach to learning,* 3d ed. Minneapolis, MN: Burgess.

Postlethwaite, S. N. (1975). Students are a lot like people! *American Biology Teacher, 37,* 205.

Postlethwaite, S. N. (1978). Principles behind the audio-tutorial system. *NSPI Journal, 17,* 3–4,18.

Postlethwaite, S. N. (1980). Improvement of science teaching. *BioScience, 30,* 601–604.

Purcell, L., & Hemphill, J. (1997). *Interactive audio*. New York: Wiley.

Quealy, J., & Langan-Fox, J. (1998). Attributes of delivery media in computer-assisted instruction. *Ergonomics, 41*(3), 257–279.

Quereshi, S. U. H. (1974). Speech compression by computer. In S. Duker (Ed.), *Time-compressed speech* (pp. 618–623). Metuchen, NJ: Scarecrow.

Rao, P. X. (1977). Telephone and instructional communication. In I. D. S. Pool (Ed.), *The social impact of the telephone* (pp. 473–486). Cambridge, MA: MIT Press.

Rehaag, D., & Szabo, M. (1995). *An experiment on effects of redundant audio in computer based instruction on achievement, attitude, and learning time in 10th grade math*. (ERIC Document ED380123)

Reiser, R. A. (1994). Clark's invitation to the dance: An instructional designer's response. *Educational Training Research and Development, 42*(2), 45–48.

Robertson, E. M. (1977). *The effects of different rates of recorded speech on the listening comprehension of adult remedial readers*. Unpublished doctoral dissertation, University of Georgia.

Robinson, D. H., & Molina, E. (2002). The relative involvement of visual and auditory working memory when studying adjunct displays. *Contemporary Educational Psychology, 27,* 118–131.

Rohwer, W. D., & Harris, W. J. (1975). Media effect on prose learning in two populations of children. *Journal of Educational Psychology, 67,* 651–657.

Ross, R. A. (1964). Look at listeners. *Elementary School Journal, 64*(7), 369–372.

Ross, S. M. (1994). Delivery trucks or groceries? More food for thought on whether media (will, may, can't) influence learning. *Educational Training Research and Development, 42*(2), 5–6.

Rossiter, C. M., Jr. (1970). *The effects of rate of presentation on listening test scores for recall of facts, recall of ideas, and generation of inferences*. Unpublished doctoral dissertation, Ohio University.

Roucos, S., & Wilgus, A. M. (1985). *High quality time-scale modification for speech*. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, IEEE, (pp. 493–496).

Rulon, P. X. (1943a) A comparison of phonographic recordings with printed material in terms of knowledge. *The Harvard Educational Review, 8,* 63–76.

Rulon, P. X. (1943b). A comparison of phonographic recordings with printed material in terms of knowledge gained through their use in a teaching unit. *The Harvard Educational Review, 8,* 163–175.

Rulon, P. X. (1943c). A comparison of phonographic recordings with printed motivation to further study. *The Harvard Educational Review, 8,* 246–255.

Rummer, R., & Engelkamp, J. (2001). Phonological information contributes to short-term recall of auditorily presented sentences. *Journal of Memory and Language, 45,* 451–467.

Russell, T. L. (1999). *The No Significant Difference Phenomenon: As reported in 355 Research Reports, Summaries and Papers* (5th ed.). Raleigh, NC: North Carolina State University.

Saettler, P. (1968). *A history of instructional technology.* NY: McGraw-Hill, Inc.

Saettler, P. (1990). *The evolution of American educational technology*. Englewood, CO: Libraries Unlimited.

Sales, G. C., & Johnston, M. D. (1993). *Digitized speech as feedback in computer-based instruction*. Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology. New Orleans, LA, January 13–17.

Salomon, G. (1994). *Interaction of media, cognition, and learning*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schacter, J. (1999). *The impact of educational technology on student achievement: What the most current research has to say. Milken Exchange on Educational Technology*. Retrieved July 28, 2002, from http://web.mff.org/publications/publications.taf?page=161

Schramm, W. (1972). What the research says. In W. Schramm (Ed.), *Quality instructional television* (pp. 44–79). Honolulu, HI: University Press of Hawaii.

Schramm, W. (1973). *Big media, little media: A report to the Agency for International Development*. Stanford, CA: Institute for Communication Research, Stanford University.

Schrock, S. A. (1994). The media influence debate: Read the fine print, but don't lose sight of the big picture. *Educational Training Research and Development, 42*(2), 49–53.

Severin, W. (1967, Fall). Another look at cue summation. *AV Communication Review, 15*(3), 233–245.

Severin, W. (1968). *Cue summation in multiple-channel communication. Report from the Media and Concept Learning Project* (Tech. Rep. No. 37). Washington DC: Bureau of Research.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Shih, Y., & Alessi, S. M. (1996). Effects of text versus voice on learning in multimedia courseware. *Journal of Educational Multimedia and Hypermedia, 5*(2), 203–218.

Short, H. S. (1977). A comparison of variable time-compressed speech and normal rate speech based on time spent and performance in a course taught by self-instructional methods. *British Journal of Educational Technology, 8,* 146–157.

Sivin-Kachala, J. (1998). *Report of the effectiveness of technology in schools, 1990–1997*. Software Publisher's Association.

Sivin-Kachala, J., & Bialo, E. R. (2000). *Report on the effectiveness of technology in schools* (7th ed.). Washington, DC: Software and Information Industry Association.

Sticht, T. G. (1968). Some relationships of mental aptitude, reading ability, and listening ability using normal and time-compressed speech. *Journal of Communication, 18,* 243–258.

Sticht, T. G. (1969, Summer). Comprehension of repeated time-compressed recordings. *The Journal of Experimental Education, 37*(4), 60–62.

Sticht, T. G. (1971). Failure to increase learning using the time saved by the time compression of speech. *Journal of Educational Psychology, 62,* 55–59.

Sticht, T. G., & Glasnap, D. R. (1972). Effects of speech rate, selection difficulty, association strength, and mental aptitude on learning by listening. *Journal of Communication, 22,* 174–188.

Stickell, D. W. (1963). A critical review of the methodology and results of research comparing televised and face-to-face instruction. *Dissertation Abstracts International, 24*(08), 3239.

Sutcliffe, A. G. (1999). A design method for effective information delivery in multimedia presentations. *The New Review of Hypermedia and Multimedia, 5,* 29–58.

Svoboda, R. G. (1978). *Audio-tutorial courses for college algebra and trigonometry: a progress report*. Fort Wayne, IN: Indiana University. (ED 167125)

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning, *Cognitive Science, 12,* 257–285.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: ACER Press.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12,* 185–233.

Sweller, J., Merrienboer, J. J. B. van, Paas F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Tabbers, H. K., Martens, R. L., & Van Merriënboer, J. J. G. (2001). The modality effect in multimedia instructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1024–1029). Mahwah, NJ: Lawrence Erlbaum Associates.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions: Refining the design guidelines.* Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.

Taylor, S. E. (1964). *Listening: What research says to the teacher*. Washington, DC: National Education Association of the United States.

Taylor, S. E. (1965). Eye movements in reading: facts and fallacies. *American Educational Research Journal, 2,* 187–202.

Thompson, A. D., Simonson, M. R., & Hargrave, C. P. (1996). *Educational Technology: A review of the research*. Washington DC: Association of Educational Communications and Technology.

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology, 48*(3), 380–396.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology, 3*(4), 257–287.

Travers, R. M. W. (1967). *Research and theory related to audiovisual information transmission*(rev. ed.). U.S. Department of Health, Education and Welfare. Kalamazoo, MI: Western Michigan University.

Travers, R. M. W. (1970). *Man's information system: A primer for media specialists and educational technologists*. Scranton, PA: Chandler Publishing Company.

Tripp, S., & Roby, W. (1994). The effects of various information resources on learning from a hypertext bilingual lexicon. *Journal of Research on Computing in Education, 27*(1), 92–102.

Tripp, S. D. (1994, Aug. 5). *Do media affect memory?* Paper presented at the 3d Practical Aspects of Memory Conference, College Park, MD.

Unwin, D., & McAleese, R. (Eds.). (1988). *The Encyclopedia of Educational Media Communications and Technology*. New York: Greenwood Press.

Van Mondfrans, A., & Travers, R. M. W. (1964). Learning of redundant materials presented through two sensory modalities. *Perceptual and Motor Skills, 19,* 743–751.

Voor, J. B., & Miller, J. M. (1965). The effect of practice on the comprehension of time compressed speech. *Speech Monographs, 32,* 452–454.

Wade, W. (1992). How interest affects learning from text. In K.A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 255–277). Hillsdale, NJ: Erlbaum.

Wade, S. E., Schraw, G., Buxton, W. M., & Hayes, M. T. (1993). Seduction of the strategic reader: Effects of interest on strategies and recall. *Reading Research Quarterly, 28,* 92–114.

Wasserman, H. M., & Tedford, W. H. (1973). Recall of temporally compressed auditory and visual information. *Psychological Reports, 32,* 499–502.

Watts, M. W., Jr. (1971). Differences in educational level and subject matter difficulty in the use of compressed speech with adult military students. *Adult Educational Journal, 21,* 27–36.

Wilkinson, G. L. (1980). *Media in instruction: 60 years of research*.

Washington DC: Association of Educational Communications and Technology.

Williams, D. L., Moore, D. M., & Sewell, E. H., Jr. (1983–84). Effects of compressed speech on comprehension of community college students. *Journal of Educational Technology Systems, 12,* 273–284.

Wingfield, A., & Ducharme, J. L. (1999, May). Effects of age and passage difficulty on listening-rate preferences for time-altered speech. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 54B*(3), 199–202.

Wingfield, A., Poon, L. W., Lombardi, L., & Lowe, D. (1985). Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time. *Journal of Gerontology, 40,* 579–585.

Winn, W. (1993). Perception principles. In Fleming M., & Levie, H. (Eds.). *Instructional message design: Principles from the behavioral and cognitive sciences*. Englewood Cliffs, NJ: Educational Technology Publications.

Wolcott, L. L. (1993). Audio tools for distance education. In B. Willis (Ed.), *Distance education: strategies and tools* (pp. 135–164). Washington, DC: American Psychological Association.

Wood, C. D. (1966). *Comprehension of compressed speech by elementary school children*. Indiana University, Bloomington, IN. (ED 003 216)

Young, W. E. (1936). The relation of reading comprehension and retention to hearing comprehension and retention. *Journal of Experimental Education, 5*(1), 30–39.