Audio in Courseware: Design Knowledge Issues

Diana Aarntzen, *University of Twente, The Netherlands*

SUMMARY

Designers and developers of courseware are inclined to use audio (and speech in particular). But they are facing educational problems: 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. In this article information on each of these items is given. Where possible, literature is reviewed. The conclusion of this article is that some research has already been done in this area but in order to obtain more design knowledge on the use of audio in courseware, much more (multi-disciplinary) research is required. Only a minimum of factors related to the use of audio in courseware is mentioned here. Designers should be aware of that. Some other aspects of audio, which are not mentioned here, are the number of voices, gender of speaker and of listener, and listening comprehension. We present research results about aspects of audio, used in courseware, that are available. The design principles that can be extracted from the research reviewed will be summarized in the concluding section.

INTRODUCTION

Audio in courseware is not a new phenomenon, but audio in courseware engineering research is. Compared to the amount of research on the visual aspects of instruction hardly any research has been done to examine the auditive aspects. Therefore not much is known about the use of audio in courseware. Only recently have designers of courseware become interested in the use of audio. They did already use audio but only when they had to or when its design did not need a lot of attention. Sound effects like beeps were used to point out to the student that the wrong button had been pressed or the wrong answer to a question had been given. In language and speech training audio was also a common feature. The students could practice their pronunciation by repeating words or sentences that were played back by a computer-directed tape. Their own pronunciation was recorded as well. Afterwards the computer would play back a word or sentence from the mastertape followed by the student's

pronunciation. The student could then compare the two.

Some recent studies on the use of audio in second language learning can be found in Janssen and Van Loon (1991). These (the use of beeps and the use of audio in language learning) are still the most common uses of audio in courseware. Designers now want to extend the use of audio. To enable a justified use of audio they need design principles; but currently there hardly are any proven design principles for audio in courseware. This article does not pretend to be the solution to that problem; it can only give a review of literature on this subject and try to list the aspects of audio in courseware that should be considered by designers.

FUNCTIONS OF AUDIO IN COURSEWARE

Before we can discuss other aspects of audio in

courseware we first discuss the possible functions of audio. One such function is the purpose it serves for presenting computer-controlled information to the learner. Audio can draw and hold the attention, complement the visual information on the screen, support the student reading the text on the screen, minimize the amount of information that it is necessary to present on the screen, announce some event, and motivate the student. These are only a few examples of functions of audio; others will be mentioned later in this section (indicated in *italic*). Since the personal computer is often part of a multimedia environment which includes audio, audio will not usually be used in isolation. The functions of audio in courseware must be considered in the context of the presentation of information through multimedia. Audio can be presented as the mainstream of information but it can also be used to complement the visual information (Unwin and McAleese, 1988, call these the predominant and the secondary, but important, roles of audio). The source of audio must be chosen depending on the function the audio has in the courseware. In general the following audio sources are distinguished: (a) voice or speech, (b) music, and (c) sound effects or natural effects (Brewer, 1986; McPherson and Timms, 1988; Taylor, 1988). Therefore the functions of audio will be discussed in relation to how it is presented (as the mainstream of information or as a complement to visual information) and what source of audio (voice/speech, music, or sound/natural effects) is used.

Audio as the mainstream of information

Voice/speech: when audio forms the mainstream of information, voices or speech seem to be most appropriate. A voice can *give the instructions*. Still or moving pictures can be used to illustrate the given information. Text on screen can complement or summarize the information. Text on screen can also be used to highlight the most important information. An important aspect that should be considered here is redundancy. Redundancy means the simultaneous presentation of semantically identical information that will be processed by the receiver through multiple sensory channels (Heuvelman, 1985, p. 12). Redundancy is important here because a simultaneous presentation of the same message auditively as well as visually might enhance the

processing and storage of the information presented. This will be discussed in a later section.

Music: music can also be the mainstream of information. In music education the computer can be used to allow the students to experiment with music. An example of such use is 'Touching Sound', a project concerned with the learning process, music education and new technology (Ellis, 1989). Touching Sound is a program with which timbre, tone colour, texture, volume, and pitch can all be accessed immediately, using joy-sticks or touch screen and later can be played using a piano-type keyboard attached to the computer. Sounds can be *discovered, explored and created*. Plenckers (1991) mentions three ways in which the computer can be used in the learning of the musical representation ability: (a) learning musical terminology; (b) learning by hearing; (c) learning by doing. With respect to (a), learning musical terminology, Plenckers explains that in Western music a vast theoretical system of terms has been developed *to distinguish different sorts of tone combinations*, based on the concept of tonal harmony. The acquisition of the harmonic terminology is done by drill and practice. According to Plenckers it is therefore suitable for Computer-Based Education (CBE). With respect to (b), learning by hearing, a computer and its internal sound generator or the Midi-synthesizer extension are considered to be adequate tools for playing musical sounds together as a chord, or one after another as melodies; consequently it is excellent equipment *to train a student's hearing abilities*. With respect to (c), learning by doing, Plenckers refers to, for example, *singing a melody that is written out*, exercises in harmony or counterpoint which are exercises *to write a second voice or a chord accompaniment to a melody*. 'Learning by doing' can be considered as a situation where music is the mainstream of information.

Sound/natural effects: Just like music, sound effects or natural effects are also not often used as the mainstream of information. They are usually used *to draw or maintain attention*. In concept learning sometimes a natural effect can be of use; for example, to be able to recognize the song of a specific bird (*to recognize the sound related to a concept*). Also, in vocational training sound effects can be used; for example, the sound of the correct use of a drill. Mechanics can learn how to

recognize what is wrong with an engine by the sound of it (*to recognize the sound of a certain action*).

These were several examples of how audio can form the main information in CBE. But speech, music, and effects can also be used in courseware in which visual information is most important. This will be described in the next section.

Audio as complement to visual information

If text on screen and still or moving pictures provide the learner with most of the necessary information, audio can still fulfil an important function. Audio can *direct attention* to the most important parts of the screen, *draw and maintain the attention* of the learner, also *create a mood to motivate* the learner. A conclusion that can be drawn from studies in the literature reviews on the effects of music by Hlinka (1980) and by Seidman (1981) is that music *affects emotional states* and *creates a mental set that influences the perception* of the message, but that effects on cognitive and attitudinal learning results are dubious (Jaspers, 1990).

Brewer (1986; pp. 282–3) mentions a list of some fundamental audio functions that could play a part in interactive audio applications. The functions will be listed using the verb that describes the action fulfilled by the function. These verbs are: (a) *greet, welcome*; (b) *explain, instruct*; (c) *prompt, request, demand*; (d) *remind, alert, warn*; (e) *help, assist, hint*; (f) *reinforce, praise, flatter*; (g) *encourage, reassure*; (h) *thank, acknowledge*; (i) *answer, confirm*; (j) *advise, suggest, guide*; (k) *lament, regret*; (l) *confuse, divert*; (m) *scare*; (n) *excite, arouse*.

For every function one specific source of audio will be the most suitable to realize it. For example, a sound effect could be used to alert or warn the learner that something is wrong. Music could be used to excite or to reinforce the learner. A voice can be used to explain and to instruct.

Voice/speech: Brewer also mentions two functions of narration (speech or voices) in combination with visual presentation: 'Here the computer-controlled audio CD can *provide high-quality narration for still sequences*, as well as *provide narration in multiple languages for the same visual program*'. (p. 278).

Music: Taylor (1988; p. 107) mentions some functions that music can have in pictures. Music might have these functions in CBE. She notes that music can be used to:

- *create atmosphere and mood, tension and relaxation*;
- *underscore the mood*;
- *change the pace of the program and maintain arousal*;
- *introduce a new phase of the story*;
- *unify a series of ideas*; and
- *separate and emphasize the points to be made*.

An example of how music could be used in courseware is to use calm music (which should not be too well known or too uncommon so it might distract the learner's attention) to create a relaxing atmosphere. Turning down the volume at the end of a section might bring more structure in the courseware. Pre-structuring lesson materials is especially important for poor learners.

Taylor (p. 107) also points out that music should not be used indiscriminately:

Decide whether it will add a new dimension to your programme and how important that dimension is. If it will be an unnecessary distraction or if you run the risk of destroying what you have already recorded, play safe and do not attempt it.

Sound/natural effects: Buxton (1989) focuses on non-speech audio. He states that non-speech audio messages can provide three distinctive general types of information: (a) alarms and warnings; (b) status and monitoring indicators; and (c) encoded messages. *Alarm and warning messages* (a) are usually loud, easily identifiable sounds with sharp transients. Status and monitoring messages (b) *provide information about some ongoing task*. The nature of such cues depends very much on the type of task being monitored. 'The design of this type of message exploits the fact that the human perceptual system does not remain conscious of steady-state sounds. In contrast, it is very sensitive to change' (p. 5). Encoded messages (c) are used *to present numerical (or quantitative) data*, such as statistical information, in patterns of sound. Lunney *et al.* (1983) used audio motives to present spectral information of various chemicals to blind students. The complex and varying sounds used here contrast with the penetrating one or two sounds

used with alarms or with the steady-state tones or patterns used in status monitoring. According to Buxton, encoded messages often exploit our capabilities of pattern matching and recognition.

Again redundancy of information is an important aspect to consider when audio is used to complement visual information. Therefore the next section will be about the relations that can exist between visual and auditory information that are presented simultaneously.

THE RELATIONSHIP BETWEEN AUDITIVE AND VISUAL INFORMATION

Having discussed functions of audio, we are now in a position to discuss other aspects of audio in courseware. Several theories exist about how the information is processed through different sensory channels and what the consequences might be for designing mediated content materials. Unfortunately, there is no agreement on which theory best reflects the actual processes. One of the theories is Paivio's Dual-Coding Hypothesis (1979). The dual-coding hypothesis assumes that imaginal and verbal processes are differentially available as memory codes for abstract words, concrete words, and pictures. The image code increases in availability uniformly over the three levels, whereas the verbal code is highly available as a representational response to words but somewhat less available as a verbal referential (labelling) response to pictures. Paivio (p. 240) makes a functional distinction between verbal and non-verbal symbolic modalities rather than between auditory and visual sensory modalities. At the same time, however, he recognizes that symbolic and sensory modes are partially correlated (for example, speech is verbal and auditory and pictures are non-verbal and visual). Heuvelman (1985) describes the same distinction. According to him (p. 12) the nature of the relation between spoken and written language can be redundant but the relation between visualisations (images) and spoken text can not be redundant. The reason for this is that the meanings of the different signals are not the same.

The meanings of linguistic utterances, spoken as well as written, are conventionally expressed. This does not apply to 'pictorial language' (ie, the content and the structure of iconic symbols).

Images and texts may not be redundant, they can 'fit' better or worse to each other as far as meaning is concerned. Images and texts can be more or less congruent to each other with regard to their meanings.

These considerations are important to designers. It means that a picture cannot be replaced by a spoken or written text. It also means that a written text could be replaced by a spoken text. In this context redundancy can only occur when the same message is processed through two different sensory channels and the information of one channel could be replaced by the information of the other channel.

Redundancy

With respect to redundancy, Paivio (1979, p. 180) poses the following 'coding redundancy hypothesis': 'Memory increases directly with the number of alternative memory codes available for an item'. This would mean that coding an item in two ways will enhance memory more than coding an item in one way only. So, presenting a stimulus auditively as well as visually would lead to a better memory performance than presenting it only visually or only auditively.

Heuvelman (1985) states that the processing of redundant auditory and visual signals will be faster than the processing of one-channel signals. He claims that redundant signals reinforce each other mutually. This effect of redundancy only takes place when there is no or hardly any background-noise (ie, other information that is not redundant with the auditory signal) on the screen. When this condition is met the effect of redundancy will be so strong that information processing by the viewer will significantly improve.

Designers of courseware should consider the question of how the combination of visual and auditory information should be. In the above-mentioned hypotheses about redundancy it is stated that both the visual and the auditory channel should bring the same information simultaneously to improve information processing. But, depending on the function audio has in the courseware, the auditory information does not always have to be semantically the same as the visual information. An example of such an exception is when audio has an anticipating function. Using just a short auditory signal to

prepare the learner for a change of subject does not necessarily have to be complemented by a visual signal as well. The auditive signal is not used to enhance memory, but only to help the learner structure the learning material. Redundancy is only a matter to consider when audio is used to convey the educational objective.

Dominance of vision over audio

Reese (1982) demonstrated that when documents were summarized verbally and were simultaneously shown verbatim, viewers inevitably read the document and ignored the verbal summary. This appears to be an involuntary response. Subjects reported they 'can't help' but read the accompanying document. The same was found by van Parreren (1988). Only exact duplicates on the screen of what is being said in the narration appear to lead to unimpeded processing of the intended message (Grimes, 1990). This is also important with respect to the latter section on redundancy. When a voice is paraphrasing the text on the screen it is quite possible that only the text on the screen will be read but that the information given by the voice will be ignored. This leads to the conclusion that the auditive information should be an exact duplicate of the visual information when it is supposed to facilitate information processing. When the two messages are not duplicates they might even hinder information processing. This applies to spoken and written text. The following can be said about combining other sounds and visual information.

Keuss, van der Zee, and van den Bree (1990) performed a series of experiments on intersensory facilitation. They demonstrated that non-informative sound of low to moderate intensity (30 to 80 dB) facilitates the reaction to a visual stimulus that is offered at the same time. It appears that auditory intensity reduces choice reaction time independently from the influence of the intensity and duration of the visual signal. Audio intensity interacts with the effect of stimulus degradation. Degraded stimuli demand more attention to be able to interpret them than intact ones. One could add audio to a degraded stimulus to counteract this phenomenon. An example of how this could work out in a very common situation is the following. Traffic signs such as arrows, barely visible in bad weather conditions,

may be better observed or identified when sound or noise of non-informative character is added to the field situation. In courseware the learning process could be improved by using non-informative sounds to draw the learner's attention to new and/or important information.

LEARNER CHARACTERISTICS IN RELATION TO AUDIO

Learner characteristics form a factor that is important to consider when audio is used. The mere fact that audio can not be used for deaf persons illustrates this point. But other learner characteristics also seem to be important, as the following literature review will show.

Reiser and Gagné (1983) mention the following learner characteristics that could be important in media selection:

- reading ability;
- age (sometimes noted as 'amount of experience as a learner'); and
- the kinds of characteristics that give students particular affinities for learning from certain media.

These and other learner characteristics will be discussed in the following sections.

Reading ability

According to Reiser and Gagné (1983) it is important to consider whether a student can, with reasonable efficiency, gain information from printed text. The identification of students as readers or non-readers serves as one of the major divisions in their media selection model. They assume poor readers to benefit more from spoken words than from written words. With respect to good or proficient readers Reiser and Gagné seem to prefer print materials. According to them, print materials provide greater pacing flexibility and may therefore offer a learning advantage over audio presentations. Another feature of print materials is that they allow the students to backtrack easily in order to review points that were previously missed. This is not possible with audio presentations. Especially for conveying verbal information they consider the most efficient medium for proficient readers to be the printed text.

Anderson-Inman *et al.*, (1990) also mention some assumptions with respect to the ability to read. They assume that speech, which provides orientation messages, programme directions, and instructional feedback, enables beginning readers to work at the computer independently. The use of speech can especially be helpful when visual programme instructions and feedback are beyond the reading level of the intended user. Speech should be exceptionally clear and preferably augmented with graphics for poor and non-readers.

With respect to 'more literate readers' Anderson-Inman *et al.*, (1990) assume that speech can provide an additional information channel which can enhance learning. Here they do not seem to agree with Reiser and Gagné (1983) who assume that printed text should be preferred to speech for good readers.

Wiener (1991) deals with computerized-speech output to enhance reading skills in special education. According to Wiener, research suggests that for handicapped students a combined visual and auditive presentation (Wiener calls this a graphics-plus-voice presentation) would be superior to a visual presentation to enable the student to read individual words. A graphics-plus-voice presentation could be helpful in diminishing reading obstacles, thus enhancing students' ability to focus on the educational objective. The findings of Wiener's pilot study support the theory that software with digitized-speech components increases mildly retarded adolescents' accuracy and ability to orally read target sight words. Thus, speech seems a necessity for learning and reading impaired learners.

Experience as a learner

Vossen and Ruess (1989) reason that because of the serial character of speech and the resulting sacrifice of time, it appears especially important to provide differently experienced users with ways of interaction that are appropriate to their different needs. On the one hand, the novice will need a lot of help and therefore the learning material should be divided into small steps. On the other hand, the expert does not want to experience inadmissible waiting and therefore wants to skip parts of the lesson material and wants to go through the programme in much bigger steps than the

inexperienced learner. This has consequences for the design of the courseware. In courseware for inexperienced learners there will be more but smaller audio-files than in courseware for experienced learners. For inexperienced learners the possibility to repeat a step (auditively and/or visually) should be available after each step, while for experienced learners this might not be necessary.

Preferences for certain media

Sometimes the term 'modality preference' is used. Not much is known about this feature. Instinctively many people would acknowledge such a preference. Reiser and Gagné (1983) assume that as a result of specialized practice with motion picture film and TV presentations a set of 'mental skills' may be developed in learners. These mental skills could become stable characteristics of the learner. Such learner characteristics may be developed by appropriate design of media-based training, they say. Romiszowski (1988) also notes that learners differ in their preference for learning by observing or by listening.

Romiszowski calls the different types of learners 'visual learners' and 'aural learners'. Barbe, Swassing and Milone (1979) define a modality as any of the sensory channels through which an individual receives and retains information. According to Barbe, Swassing and Milone there are three views of modality: (a) modality as a fixed neurological characteristic; (b) modality as a preference; and (c) modality as a measurable behaviour. They acknowledge that the manner in which an individual is most comfortable receiving information is not always consonant with the way in which information is most efficiently received and processed. To optimize the learning process it is not enough to find out what the preference of the learner is. Designing courseware according to modality preferences of learners might motivate those learners but it does not automatically mean that the information processing will improve. Another consideration Barbe, Swassing and Milone mention is that a modality may come in different grades: (a) dominant modality; (b) secondary modality; and (c) mixed modality. An individual's dominant modality is that channel through which information is processed most

efficiently. Beside a dominant modality, many people evidence a secondary modality upon which they can rely. A secondary modality is not so efficient as its dominant counterpart. Mixed modalities occur when no single modality is clearly dominant. According to Barbe, Swassing and Milone the proportion of persons with mixed modalities is larger among adults than children. They suspect that cognitive maturity and the opportunity of practice in all three modalities (visual, auditive, and kinesthetic) are the principal reasons that mixed modalities occur.

Barbe, Swassing and Milone (1979) developed a test to assess modality strengths of children: the SBMI (Swassing-Barbe Modality Index). This test uses shapes to assess modality strengths. First the teacher or examiner shows the student some shapes that he or she forms with blocks. After the removal of a shape the student has to form that same shape. When more than one shape is involved the simple matching of the shapes is not enough. They must also be in the same order in which they were presented. The next part of the test consists of the teacher or examiner reading the names of the shapes out loud. The student should not start reproducing the pattern until the entire sequence is read. The student has to form the shapes with the blocks. The last part of the test is the kinesthetic subtest. The student can feel the complete sequence of the figures once. Then the student must reconstruct the pattern. The student's score for each item is the number of shapes that are correctly placed in sequence. To identify the relative strength of each modality, it is necessary to compute the percentage of the total score each modality represents. The student's results on the SBMI can be used to determine the appropriate use of visual and auditive modalities in courseware. A courseware designer should make it possible for the teacher to switch the audio or the video on or off for any student, depending on the student's results on the SBMI.

Until now, the distinction between visual, auditive, and kinesthetic sensory channels has been discussed. But in literature the term 'modality preference' is also used for the distinction between pictorial and linguistic information. Here text on screen and spoken text are in the same dimension: linguistic information. Clark and Chase (1972) developed the 'star-plus' test to be able to distinguish between learners who

prefer pictorial information and learners who prefer linguistic information (Pezdek *et al.* 1987). Because the term 'modality preference' can have different meanings it is not always possible to detect what an author, who used the term, meant by it. This applies especially when the author only mentions that he or she assumes that modality preference is important to consider when media for education have to be selected but does not review literature or mention results of research. Unfortunately, this is the case for most literature on preferences for: (a) receiving and processing information through a specific sensory channel (auditive, visual, or kinesthetic); (b) receiving and processing linguistic or pictorial information. Usually no literature or results of research, on which the author's assumption that modality preference is important for media selection is based, are mentioned.

Physical disabilities of learners

Apart from deafness, other physical disabilities that might be important for the use of audio are mentioned by Romiszowski (1988), such as, poor vision, dyslexia, and colour blindness. With respect to the visually impaired Buxton (1989) made the following remark: 'There is one very clear use of the audio channel: providing interfaces for people with visual disabilities' (p. 3). Anderson-Inman *et al.* (1990) also mention the visually impaired. According to them, for this group of learners speech quality is less important than speed, which can and should be much faster than normal speech. The reason for this is supposed to be that the visually impaired are far more experienced at receiving input through their auditory channel and are often frustrated with the slowness of speech-based programs designed for readers who are not visually impaired.

The question, dealt with by Krause (1990), is: can the information produced on a computer display really be transformed into acoustical events, preferably into speech? The verbal description must evoke concepts in the mind of the blind person. Krause mentions the objects presented on computer displays which should be transformed into acoustical events for blind people: (a) normal text; (b) data organized in tables with perhaps many columns; (c) data from tables transformed into bar or pie diagrams; and (d) two- or three-

dimensional diagrams with curves. A transformation into speech by text to speech systems seems to be the best solution for normal text on screen. For the data organized in tables a transformation of the 'gestalt' of the tables by verbal description of the organization of the tables and the content of columns and lines could be helpful. The data in bar or pie diagrams are taken from tables and therefore they can be verbalized in a similar manner. Data in two- or three-dimensional diagrams with curves are also table-oriented but often from a nearly continuous data set. Interesting parts of curves are maxima and minima, and the trend. These geometrically outstanding properties could be extracted and transformed into a verbal description.

EVENTS OF INSTRUCTION

Gagné (1977) mentions the following 'events of instruction':

- gaining attention;
- informing the learner of the objective;
- stimulating recall of prerequisite learning;
- presenting the stimulus material;
- providing 'learning guidance';
- eliciting the performance;
- providing corrective feedback;
- assessing the performance; and
- enhancing retention and transfer.

This section is related to the section on functions of audio in courseware. For every event of instruction another function of audio is required. This influences what audio source can be used. Reigeluth (1987) mentions raising the voice as one of the possibilities to gain attention (1). A voice could also inform the learner of the lesson objective(s) (2) and stimulate recall of prior learning (3), depending on the different learned capabilities desired as outcomes. With regard to presenting the stimulus material (4) Reigeluth (1987) states that any stimulus must have its distinctive features clearly delineated. The use of variety of tone, pitch, and volume in speech can be of help here. Providing learning guidance (5) involves making the stimulus as meaningful as possible. This means that the learner has to be able to link the new information in many ways to already present information. One way of helping the learner with this is to provide the same

information through different sensory channels (see the section on redundancy). For certain learning outcomes it is possible to elicit the performance (6) by asking questions. These questions could be asked visually and/or audiotively. Feedback (7) could also be provided audiotively (by beeps or by a voice telling the learner whether or not the answer was correct, giving a hint or giving the correct answer). Examples of how 'assessing performance' (8) and 'enhancing retention and transfer' (9) can be performed audiotively could not be found in literature.

Audio could be used for every event of instruction, but is it wise to do so? In literature, audio was mentioned as being useful for providing aural directions, help and reinforcement (Baker and Bixler, 1991). In feedback verbal language and beeps can be used (Schimmel, 1988).

Ellington (1987) mentions four functions of audio, depending on the nature of the (as he calls it) multimedia multi-activity study unit or module: (a) providing any form of information; (b) guiding the learner through different learning activities; (c) providing the learner with questions the learner should answer and giving feedback with regard to the answers; (d) providing extension material (material that builds on what was learned from other materials and activities in the unit; for example, a profound discussion of important points). According to Anderson-Inman *et al*, (1990) speech can provide orientation messages, program directions, and instructional feedback to beginning readers. For more literate readers speech can be used to enhance learning by providing an additional information channel. Brewer (1986) mentions that various audio sources can be used to play different characters. One character could be the 'teacher', dispensing information and asking questions; the other character could be someone who gives hints and reinforcement. Bridgman (1990) investigated the value of, and developed methodologies for, the use of digitized interactive voice within CBE. Bridgman used voice error-messaging to draw the trainees' attention to the fact that they had made an error. Repeat prompting was also designed as a voice operation. If a mistake was repeated, the error and repeat prompt message could change to give extra information and/or encouragement with verbal inflection to reinforce the message. After

three repeated errors on the same question, the program inserted the correct response using voice to inform the trainee of the correct answer and the program's action. Vossen and Ruess (1989) investigated the usability of auditive menus. Audio could be used for the presentation of a menu, but because of the transient and serial nature of audio there are limitations: the options that are provided by the menu have to be remembered by the learner since they are gone after they were heard. A long and broad menu would lead to memory-problems because of the limitations of human memory capacity. And even if memory capacity would be sufficient, the serial character of audio would affect the time that is necessary to make a choice since the learner can not just 'scan' the options but they must be repeated one after another. Vossen and Ruess therefore advise limiting auditive menus to three to five options.

AUDIO CHARACTERISTICS

Audio has specific characteristics. One of these is that its nature is transient and sequential. Audio can also be used interactively. This means, for example, that the computer and the learner could speak to each other. There also are different techniques that can be used to add speech to educational software. These aspects will be discussed in this section.

The transient and sequential nature of audio

A visual cue can remain on the screen of the computer for as long as is necessary. The learner can watch or read it until it is removed. An auditive cue, however, disappears immediately after its production (ie, it is transient). This is a very important characteristic of audio. In order to hear the information again, it must be repeated. The designer of courseware must consider whether to make this repetition of audio computer- or learner-driven. One learner might need to repeat a long and complicated sentence five times while another learner needs to repeat that same sentence only once.

Another characteristic of audio is its sequential nature. This means that a person cannot listen back to an earlier part of the sentence while he or she is listening to that sentence. You have to

rewind the tape or start playing it back from the beginning in order to be able to hear the interesting or missed part again. A designer of courseware should be aware of this because it could be very time-consuming. A long piece of audio should be divided into many small parts and stored as such. When the learner wants to repeat only one segment it will not take as much time as repeating the whole piece of audio.

Audio that is truly interactive

Interaction between the computer and a learner can occur in many different ways. The learner can use a keyboard, a mouse, or touch-screen to communicate with the computer. The computer can use text, still and moving pictures on the screen; or auditive text, music, or effects to communicate with the learner. Another way might be interactive audio. The computer talks and the learner talks back. The talking computer is nothing new. The next section will discuss the different techniques to make the computer speak. But what about the learner speaking to the computer and the computer being able to 'understand' it and act according to the learner's response? It is not possible yet for the learner to use natural language when he or she is talking to the computer. In order to make this possible there are several important directions in research. IBM (1991) mentions (a) the Hidden Markov Models; (b) the neural networks; and (c) dynamic programming as promising techniques. For now the learner will have to use restricted language. Gilbert, Wooffitt and Fraser (1990) think that the computer will be able to understand connected word speech from learners using a moderately large vocabulary (about 1000 words) within one to two years. Until then it will not be easy to use complete interactive audio. With regard to the future, it is assumed that natural language will become an important option for the learner to communicate with the computer. Whether the computer will be able to respond in natural language is another question. In that case the computer should be able to recognize whether the words spoken by the learner were meant seriously, cynically, sceptically, etc. in order to respond correctly. When the courseware limits the context of the conversation (and no remarks that are made by the learner outside of that context are accepted by the computer) the computer might be able to

respond in natural language.

Speech technology

The three most common technologies for adding speech to educational software are (a) digitized speech, (b) encoded speech, and (c) synthesized speech (Anderson-Inman *et al*, 1990). Wiener (1991) uses different names for the same three technologies for adding speech: (a) compressed, digitized human speech; (b) linear predictive coding (LPC); and (c) text-to-speech. The highest quality computer speech is derived from human voice that is recorded, digitized, compressed, and then stored in the computer's memory (a). Unfortunately, the amount of digitized voice that can be stored and processed by a microcomputer is severely limited unless a mass-storage device is used. Encoding speech (b) is less (external) memory intensive, and it still produces very intelligible speech. With this method, human speech is recorded, digitized, and broken into frames that are approximately 25 milliseconds long. A portion of each frame is electronically extracted and the results are then 'hand' edited to eliminate errors introduced during the process. Using a special speech synthesis chip the resulting units are then compared to a computerized mathematical model in order to simulate the original speech signal as closely as possible. Encoded speech is, according to Anderson-Inman *et al*, extremely memory efficient. Synthesized speech (c) is appropriate when 'unlimited' speech is required. In the case of unlimited speech the speech segments are relatively large (lasting longer than a few sentences), making digitized and encoded speech – which would impose too heavy memory requirements – not feasible. One use of synthesized speech is text-to-speech translators. The more complex the text-to-speech algorithm, the more (internal) memory is required. Designers of courseware face a trade off between intelligibility and the amount of memory that is required for speech.

CONCLUSIONS

The following design principles can be extracted from this article:

- Audio can be used as the mainstream provider of information. For this purpose voices (speech)

should be used. Music and effects as the mainstream of information can only be used when the subject involves music or effects.

- When speech is used in courseware, it is usually accompanied by still or moving pictures and text (since the personal computer is often part of a multimedial environment). The text-on-screen should not be different from the text that is simultaneously presented auditively, since vision dominates hearing. Therefore the learner is inclined to pay attention only to what is presented on the screen instead of also paying attention to what is presented auditively. (The audio presentation will in this case only impair instead of enhance the visual presentation.)
- When speech is used as the mainstream provider of information it should be accompanied by that same text on the computer's screen. According to redundancy theories, this will enhance learning.
- Audio can have many different functions. During this article these functions were mentioned. For example: in the sub-section 'Audio as complement to visual information' a list of functions indicated by verbs, that Brewer (1986) mentioned, is referred to. For every function one of the audio sources usually is more appropriate than the others. Examples that were mentioned in the text are: (a) a sound effect could be used to alert or warn the learner that something is wrong; (b) music could be used to excite or to reinforce the learner; and (c) a voice can be used to explain and to instruct. The section 'Functions of audio in courseware' will be helpful for the designer to decide which source of audio should be used for the function the designer has in mind for audio.
- Learner characteristics like reading ability, experience as a learner, modality preference, and physical disabilities of learners affect the effectiveness of the use of audio in courseware. Poor readers really need spoken text to accompany or to replace the written text, since their reading ability is not sufficient to draw the meaning from the written text. Inexperienced learners take small steps and need a lot of rehearsal which means that the audio-files should be small (to make it easier to repeat and to make it quicker, because only the part of interest is repeated). 'Modality preference' is the preference of the learner for a certain medium (be it auditive, visual or kinesthetic).

Information should be presented by way of the sensory channel that the learner prefers to enhance learning or that the particular learner can use most effectively. Physical disabilities of the learner can force the user either to use or not to use audio. For blind people braille or audio can be used to submit information. For deaf people audio should not be used, as might be obvious.

- As was true for the functions of audio, every event of instruction can be provided by a source of audio. Some examples were mentioned in the section on 'Events of instruction'. A designer should be careful what source of audio he or she uses for a specific event of instruction. Also some parts of instruction should not be presented auditively. An example is the presentation of a menu or other sort of listing consisting of more than five items. This exceeds the capacity of the learner's memory.
- Because of the transient and serial nature of audio the audio should be divided into many small parts and stored as such. When the learner wants to repeat only one segment it will not take as much time as repeating a longer piece of audio in which the smaller part, that is of interest to the learner, is included.
- It is already possible to let the learner talk to the computer and to let the computer respond to his or her voice, but until it is possible to use natural language as a means of communication between learner and computer, the learner will have to use a restricted vocabulary. Designers of courseware have to adjust their courseware to this limitation of the computer's capacity.
- Depending on the quality of speech that is required (blind people can do with less quality than poor readers) and depending on the amount of external memory (eg on hard disk) and internal memory available, one of three types of speech technology should be chosen. When a good quality of speech is necessary and a lot of memory space is available, then digitized speech should be used. When speech should be intelligible but the amount of memory space is not sufficient to use digitized speech, then encoded speech will do. The quality of encoded speech is not much less than the quality of digitized speech but it requires more effort and time from the designer of audio. When unlimited speech is necessary and the quality of

speech is not very important, then synthesized speech should be used.

REFERENCES

- Anderson-Inman, L., Adler, W., Cron, M., Hillinger, M., Olson, R. and Prohaska, B. (1990) Speech: The third dimension. *The Computing Teacher*, April 1990, 35-40, 53.
- Baker, G. and Bixler, B. (1991) Computer-Aided Instruction design techniques for the low-literate adult. *Computers in Adult Education*, 2, 1, 18-27.
- Barbe, W.B., Swassing, R. and Milone, M.N. (1979) *Teaching through modality strengths: Concepts and practices*. Zaner-Bloser, Columbus, OH.
- Brewer, B. (1986) Compact disc interactive audio. In Lambert, S. and Ropiequet, S. (eds), *CD-ROM: The new papyrus* (pp. 273-90). Microsoft Press, Redmond, VA.
- Bridgman, T.F. (1990) Applied interactive voice in CBT: A pilot project. *Computers and Education*, 15(1-3), 173-81.
- Buxton, W. (1989) Introduction to this special issue on nonspeech audio. *Human-Computer Interaction*, 4, 1-9.
- Clark, H.H. and Chase, W.G. (1972) On the process of comparing sentences against pictures. *Cognitive Psychology*, 3, 472-517.
- Ellington, H. (1987) *How to produce audio materials: Teaching and learning in higher education*, 15. Scottish Central Institutions Committee for Educational Development, ERIC ED 289 498, Aberdeen, Scotland.
- Ellis, P. (1989) Touching Sound-Connections on a creative spiral. *Education and Computing*, 5, 1, 2, 127-32.
- Gagné, R.M. (1977) *The conditions of learning* (3rd edn.). Holt, Rinehart, and Winston, New York.
- Gilbert, G.N., Wooffitt, R.C. and Fraser, N. (1990) Organising computer talk. In Luff, P., Gilbert, N. and Frohlich, D. (eds) *Computers and conversation*. Academic Press, Harcourt Brace Jovanovich, London.
- Grimes, T. (1990) Audio-video correspondence

- and its role in attention and memory. *Educational Technology Research and Development*, **38**(3), 15-25.
- Heuvelman, A. (1985) *Invloed van Visualisatie Wijze op het Leren van Televisie* [Influence of way of visualizing on learning from television]. Technische Hogeschool Twente, Enschede.
- Hlinka, D. (1980, May) *Film music: Implications for instructional films and television*. Paper presented at the annual conference of the Australian Society for Educational Technology. ERIC ED 244.602.
- IBM (1991) De computer leert luisteren [The computer learns to listen]. In IBM, *Berichten, Speciaal Universiteiten* (pp. 34-5). RUG, Groningen.
- Janssen, J. and Loon, H. van (1991) *New Media in the Humanities*. Selected papers from the International Conference on interactive audio, interactive video and open learning, 24 and 25 October 1990. University of Amsterdam, Faculty of Arts, Amsterdam.
- Jaspers, F. (1990) *Music in instructional materials. Internal report*. Enschede: University of Twente, Department of Education, Division of Educational Instrumentation.
- Keuss, P.J.G., Zee, F. van der and Bree, M.B.M. van den (1990) Auditory accessory effects on visual processing. *Acta Psychologica*, **75**, 41-54.
- Krause, M. (1990) Information capacity of the visual channel and its compensation by other channels. In Fellbaum, K. (ed.), *Access to visual computer information by blind persons: State of the art and proposals for projects*. Technical University of Berlin, Institute of Telecommunications, Berlin.
- Lunney, D., Morrison, R.C., Cetera, M.M., Hartness, R.V., Mills, R.T., Salt, A.D. and Sowell, D.C. (1983) A microcomputer-based laboratory aid for visually impaired students. *IEEE Micro*, **3**, 4, 19-31.
- McPherson, A. and Timms, H. (1988) *The Audio-Visual Handbook*. Pelham Books, London.
- Paivio, A. (1979) *Imagery and Verbal Processes*. Lawrence Erlbaum, Hillsdale, NJ.
- Parreren, C.F. van (1988) *Ontwikkelen Onderwijs* [Education under development]. ACCO, Leuven/Amersfoort.
- Pezdek, K., Simon, S., Stoeckert, J. and Kiely, J. (1987) Individual differences in television comprehension. *Memory and Cognition*, **15**, 5, 428-35.
- Plenckers, L.J. (1991) Some reflections on the use of computers for education in music theory. In Janssen, J. and van Loon, H. (eds), *New Media in the Humanities* (pp. 93-9). Selected papers from the International Conference on interactive audio, interactive video and open learning, 24 and 25 October 1990. University of Amsterdam, Faculty of Arts, Amsterdam.
- Reese, S.D. (1982) *Multi-channel redundancy effects on television news learning*. Unpublished doctoral dissertation, University of Wisconsin.
- Reigeluth, C.M. (1987) *Instructional theories in action: Lessons illustrating selected theories and models*. Lawrence Erlbaum, Hillsdale, NJ.
- Reiser, R.A. and Gagné, R.M. (1983) *Selecting media for instruction*. Educational Technology Publications, Englewood Cliffs, NJ.
- Romisowski, A.J. (1988) *The selection and use of instructional media: For improved classroom teaching and for interactive individualized instruction* (2nd edn). Kogan Page, London.
- Schimmel, B.J. (1988) Providing meaningful feedback in courseware. In Jonassen, D.H. (ed.), *Instructional designs for microcomputer courseware* (pp. 183-95). Lawrence Erlbaum, Hillsdale, NJ.
- Seidman, S.A. (1981) On the contributions of music to media productions. *Educational Communications and Training Journal*, **29**, 49-61.
- Taylor, M.H. (1988) *Planning for video: A guide to making effective training videotapes*. Kogan Page, London; Nichols, New York.
- Unwin, D. and McAleese, R. (1988) *The encyclopaedia of education media communications and technology* (2nd edn.). Greenwood Press, New York.
- Vossen, P.H. and Ruess, M. (1989) *Auditive Menüs* [Auditive menus]. Projektbericht SUNSTAR 1989.08.25. Psychologisches Institut der Universität Tübingen, Tübingen, Germany.
- Wiener, R. (1991) Computerized speech: A study of its effect on learning. *Technological Horizons in Education*, **18**, 7, 100-102.

BIOGRAPHICAL NOTES

Diana Aarntzen is a postgraduate student (PhD student) at the Division of Educational Instrumentation Technology at the University of Twente. Her research area consists of audio in courseware. The goals of her research are (a) to map as many aspects of audio that could influence

the use of audio in courseware; (b) to create a research model in which these aspects can be linked; and (c) to gain data experimentally on some of these aspects.

Address for correspondence: University of Twente, Department of Education, P.O. Box 217, 7500 AE Enschede, The Netherlands.