Multimodal Slide Shows as Asynchronous Presentation Reviews

Roger J. Chapman[1]
Computer Science Department, University of Hawaii at Hilo
Hilo, HI, 96720, USA

Abstract

This paper describes the design of, and initial results from using, a software application for recording multimodal slide show presentations that was used to create pre-examination reviews of course material in a traditional computer programming class. The results suggest these students found the reviews and software to be useful, and particularly valued well-synchronized speech and pointing when it helped focus attention, but they also found unnecessary pointing to be distracting. More generally, the results suggest that with appropriately designed software, faculty, often already in the habit of duplicating presented material for students, can recreate a more natural, significant part of the classroom experience, without having to spend a lot of time working with relatively complicated authoring systems.

Keywords: Multimodal interaction, educational software; presentation systems; courseware; asynchronous learning; human-computer interaction; deictic gesturing.

While lecturing is not usually embraced from a pedagogical perspective, because it is not considered to be student-centered, the fact is that a very large number of, if not most, instructors spend a lot of time teaching this way. This environment may be described as a *multimodal* learning environment, because it requires more than one sensory channel to interpret what is presented as both speech and visible activities occur. These activities normally don't occur as independent events either, as what is said often relates to what the student is assumed to be looking at, including the gestures of the instructor. One type of gesture frequently used by instructors is deictic gesturing (McNeill, 1992), where a person points to something referenced in a simultaneously spoken sentence to gesture "this", "that", or "there", etc., so that a particular object or location is more precisely referenced. Instructors will also often *talk through* a diagram, making a series of linking deictic gestures (Ware, 2000).

Today, presentations often take the form of slide shows using software such as Microsoft PowerPoint. The potential advantages of using slide show based presentation software include:

1. The slide show serves as a structuring and memory aid to the instructor that can be referenced while still communicating with the class;
2. The slides can quickly be made visible to both the instructor and students, and they then share a common view of that material;
3. The slides can easily be stored, edited, reused, and shared electronically.

The potential disadvantages of using slide show based presentation software include:

1. Instructors can go through the slides too quickly if students don't have a hardcopy of the slides, and are attempting to reproduce the slides on paper;

[1] roger.chapman@acm.org

2. Instructors can put so much of the targeted knowledge on the slides, that there is little value to be added as the slide show is actually presented, leading to instructors simply stating the obvious or students not attending classes;
3. Instructors can lose the advantage of having a previously created "script" if questions or open discussion causes that static script to be far less useful as the dynamics of the actual classroom situation have deviated significantly from what was predicted.

In this paper a study is described that involved using a software application, designed to capture speech and deictic gesturing to objects and locations on a slide, to create reviews for material in a data structures and program design class. This study suggests that students can benefit from this type of software being used to create reviews of material covered in class, but it also reveals that deictic gesturing must be supported with care.

1. MULTIMODAL COMMUNICATION SYSTEMS

The research on multimodal systems can be categorized in a number of different ways. One way to divide the research is based on the type of task being performed, such as presenting, problem solving, scheduling, or authoring. Another is based on who the communicating agents are, such as human-to-human communication, human-machine communication, and Computer Supported Cooperative Work (CSCW) between humans. There are lessons to be learned from research in all these areas, but to expand on them all here would be beyond the scope of this paper. A sample of the research relevant when considering how to create computer-supported presentations is therefore given.

Chapanis (1975) conducted a much-referenced early study involving human cooperative problem solving and found that speech based communications involved many more words per minute than text based and that tasks were solved faster when speech was used. Research has also shown that people point naturally when working in small groups and they are involved in a design task (Bly 1988; Tang, 1991).

The research on multimodal interaction between humans and computers attempts to support human input that is expressive and natural, in combination with multimedia output (Oviatt, 1999). Today's research in this area includes capturing more than mouse pointer based deictic gestures input to the computer, including capturing written input, manual gesturing and facial expression, but these require special equipment. The systems that capture deictic gestures are also attempting to unambiguously determine what the command was to the computer rather than just record deictic gestures and leave the interpretation to the human viewer. Thus, many of the issues in this research area are not relevant here, but general observations such as the fact that in one study 95% to 100% of users preferred to interact multimodally when they were free to use either speech or pen input in a spatial domain, but that users typically intermix unimodal and multimodal expressions (Oviatt, 1999), are interesting assuming they translate to computer supported multimodal human to human communication.

In eye tracking and recall testing studies, Faraday and Sutcliffe (1997) investigated attention and comprehension by users interacting with multimedia presentations. From this study a set of guidelines for controlling attention in multimedia presentations was produced, including the following:

- *Use object motion to control attention and viewing order;*

  (Participants' attention was drawn to motion and fixations tracked the moving objects path)
- *Use animation with care;*

  (Multiple simultaneous animations from moving objects or revealed objects and labels, or too rapid motion, sometimes caused attention to unintended areas)
- *Reveal important information to emphasize it;*

  (Static objects and labels received less attention than those revealed or in motion)
- *Use symbols to direct attention to specific objects and locations;*

  (The arrow symbol shifted fixations to that which the arrow pointed)
- *Speech information should reinforce image;*

  (Propositions given only in the image or animation without speech cueing were poorly recalled)
- *Captions or labels may be useful in re-inforcing the speech track;*

  (Propositions given only in the speech track were generally poorly recalled)
- *Cue animations with speech;*

  (Animated objects, which were cued by the speech track, were well recalled)

Research with the goal of developing better computer systems for collaborative work has also compared different communication modes. For instance, Neuwirth et al. (1994) compared the nature and quantity of voice and written comments produced in each mode, when reviewers gave feedback to writers. They found: (1)

reviewers used more words in voice than text mode during the same time period, but that the same number of annotations was made on average. The additional words were attributed in part to providing more reasons why the reviewers thought something was a problem and for polite language that mitigated the problem; (2) evaluations of reviewers were less positive when reviewers produced written annotations than spoken; and (3) comments about low-level mechanics were preferred in text. Daly-Jones et al. (1997) conducted a study where 'manager-secretary' pairs were asked to complete an asynchronous appointment-scheduling task and an equipment-booking task in three conditions: *Fax-only* involved using Microsoft Paintbrush; *Voicefax* involved using Lotus Screencam (an application that allows synchronized voice and pointing to be recorded by creating a 'movie' from the output on the user's computer display over time while recording and synchronizing any audio input) with a Paintbrush image; *Voice-only* involved just audio. For both sending and receiving, voicefax was rated most useful, then fax-only, then voice-only. Subjects took the same amount of time to complete the tasks in each condition, but fewer messages were sent with voicefax. The results of this study are consistant with Ware's (2000) discussion on computer supported communications where he reports that voice communications and shared cursors are the critical components in maintaining dialog and adds that it is generally thought to be much less important to transmit the image of the person speaking.

Multimodal messaging systems integrated with E-mail and Newsgroup systems, have been developed and been popular with their users. The Collaborative Slide Annotation Tool (CSLANT) supports the asynchronous exchange and tracking of annotated slides with traditional annotation marking and deictic gestures via synchronized voice and mouse pointer recording (Chapman et al., 2000a). Its uses have included supporting multimodal communications between airlines and the Federal Aviation Administration (FAA) to discuss the reasons for poor flight performance, and both dispatchers and traffic managers considered it useful and useable (Chapman et al., 2000b). The Microsoft Research Annotation System (MRAS) is another multimodal messaging system being used for "on-demand training" (Bergeron et. al., 1999, 2001). It supports streaming video with personalized and sharable student annotations tied to specific portions of the video presentations. This Web-based system supports messages organized in a bulletin board structure to implement the sharing of annotations. WebCT (http://www.webct.com/) is a commercially available system for creating web sites that also supports standard electronic text based communications, such as chat, E-mail, and bulletin board, although it doesn't have rich support for multimodal communications and annotations.

This research suggests that there are benefits from multimodal communication in a variety of situations. In the study presented here empirical evidence is gathered to determine how difficult it is to create and how effective asynchronously presented multimodal slide shows (with mouse pointing deictic gestures) are in this domain when used as presentation reviews.

2. CREATING POINT 'N' TALK RECORDINGS

Point 'n' Talk[2], a Microsoft Windows application, was developed with the goal of constructing a simple presentation system capable of supporting synchronized pointing and speech over a graphics image, and able to run in either a record or play mode, or a play-only mode. A snapshot of the interface to Point 'n' Talk is shown in figure 1. The version given to students was only capable of running in the play-only mode. This is because the intention was not to replace in-class questions with Point 'n' Talk questions. The software and recordings were made available to students on a web server, but the recordings were not streamed, so students had to completely download each recording before playing could begin. Eight recordings were made to review topics from one chapter of Kruse and Ryba's textbook (1998), "Data Structures and Program Design in C++". Recordings were made over figures from the textbook. The topics covered were: An overview (2 mins 12 secs; 1.11Mb); The call stack and recursion trees (1 min 8 secs; 673 Kb); The factorial function as an example of recursion (4 mins 30 secs; 2.33 Mb); The Fibonacci function as an example of recursion (1 min 15 secs; 834 Kb); The tower of Hanoi function as an example of recursion (3 mins 39 secs; 1.87 Mb); The eight queens problem as an example of recursion (8 mins 25 secs; 4.33 Mb); Game trees and the Minimax algorithm (5 mins 2 secs; 2.52 Mb); The game of eight as an example of a game tree (3 mins 20 secs; 1.73 Mb). The goal was to make the recording process no harder than speaking over the top of slides in the classroom setting, but during the process of creating the recordings there were two clear differences: (1) when there were many points to be made regarding one slide it was difficult to make the recording in "one take"; and (2) it wasn't always necessary to point, but the pointer's position was constantly recorded for playback, and this caused the instructor to move it into a "white-space" area when speaking but not referring to a point or object on the image for extended time periods. The fact that short recordings were being made is perhaps the reason for greater instructor sensitivity to the words used and pauses between sentences. This became less of a problem when a modest "top-down design" strategy was taken with the recordings, where the main topics to be made were identified and then recorded sequentially, stopping if necessary to have more time to compose.

[2]Also written as *PointnTalk*.

The second difference is perhaps a function of the fact that speech was being recorded with limited "presence" of the instructor. In a real classroom situation the instructor can make eye contact or walk away from the projection system to communicate that attention is no longer intended to that pointed to by the mouse pointer.

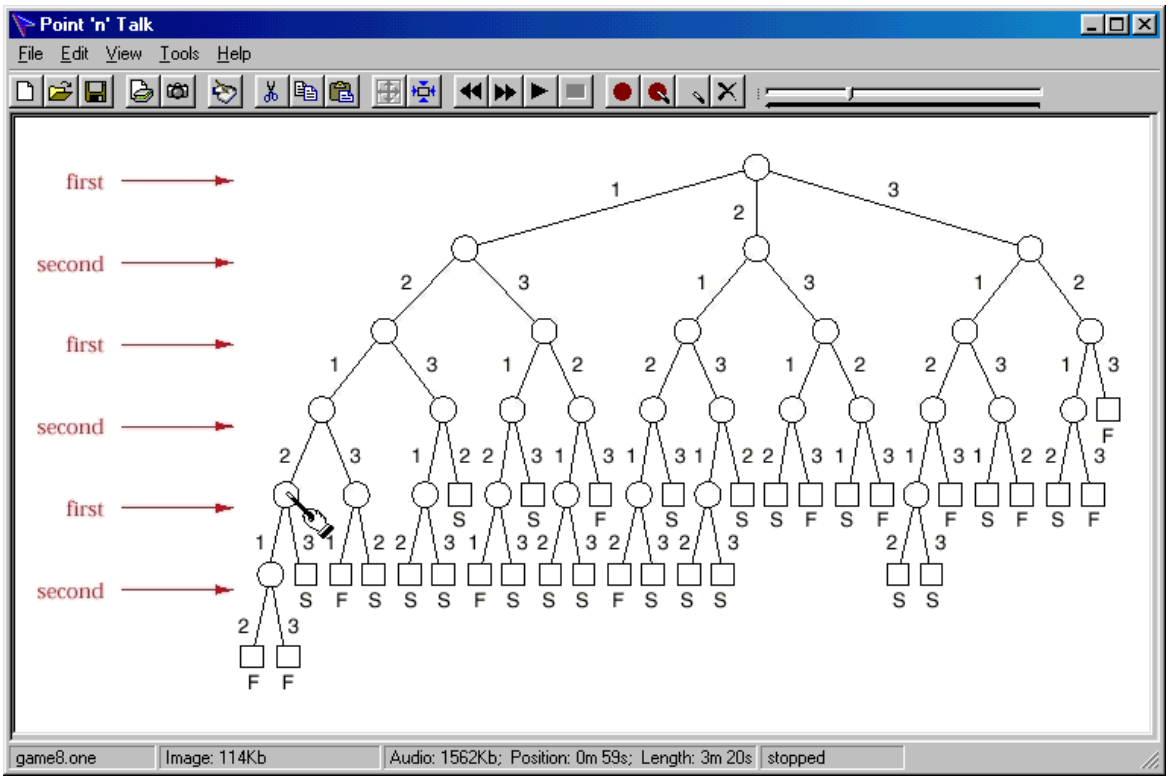

Figure 1. Screen Capture from a Point 'n' Talk Recording for the Game Eight

Student Responses														
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	Avg.
Q1	7	7	6	7	5	5	6	7	6	6	4	6	5	5.92
Q2	7	6	7	7	7	5	6	5	6	5	6	7	6	6.15
Q3	1	2	3	2	3	2	3	2	2	1	2	1	2	2
R1	8	7	8	8	5	3	8	8	5	4	3	8	8	6.38
Q1: *On a scale of 1 to 7, where 1 means not very useful and 7 means very useful, how would you rate the usefulness of Point 'n' Talk to provide review information to students in CS151 classes?*														
Q2: *In a scale of 1 to 7, where 1 means very difficult to use and 7 means very easy to use, how would you rate Point 'n' Talk?*														
Q3: *How many times do you think you tended to play each review?*														
R1: The entries in this row are the number of reviews the students downloaded.														

Table 1 Student survey responses

3. STUDENT REACTION TO THE REVIEWS

The students in this course had three written examinations. A "review sheet" was created before each one, in the form of a Microsoft Word document. These were not an attempt to directly teach, but merely contained a list of the types of questions that might be asked and a more detailed list of the topics that had been covered than was on the syllabus. The Point 'n' Talk reviews were quite different from these in that they repeated material covered in class, and thus were an attempt at more direct teaching. They were however more succinct as a priority was placed on trying to keep the recordings reasonably short in order to help reduce the file size and maintain the students' attention.

This was a small class of 17 students. Thirteen completed the online questionnaire, which contained the following initial statement to encourage students to be objective in their assessment: "The following questionnaire is to gather information regarding the potential for Point 'n' Talk to help provide reviews of course material in future computer science classes. Your input is very important, but will not be viewed until after grades are assigned for this class."

The answers to the questions shown in Table 1 suggest that this group of students considered the software both useful and easy to use. It is not clear why some students didn't play all the recordings. Some may have decided they didn't need a review on all the topics, but based on their answers to the open ended questions discussed below, some may not have wanted to take the time to download those recordings.

In open-ended questions, students were asked if there was something they particularly liked about the software, and if there was something they particularly disliked. Seven students specifically mentioned liking the synchronized speech and pointing. Several students also mentioned how easy they found the software to use and the concept of the software itself as an application for playing reviews in the student's own time. One student stated, "I thought it was a nice way to have the classroom at home, where you have an instructor and there is a visual aid plus a "ruler", which you can follow when hearing the lecture." (This limited "telepresence" is perhaps created the most by the combination of a human voice and "life" that is also captured in the recorded pointing.) As a dislike, some students mentioned that they thought it took too long to download the recordings with a modem and therefore suggested compressing the files further. HCI related suggestions were to support resizing the images depending upon the current window size and a mouse wheel for scrolling. To further illicit feedback without making the students feel they were criticizing their instructor, students were asked what characteristics they thought an effective presentation would have and what characteristics an ineffective presentation would have. This resulted in students indicating that the following were considered important: a clear, enthusiastic voice; concise presentations; informative images; informative comments; and only moving the pointer when necessary. Five students made the latter point (three of whom also listed the deictic gesturing as something they particularly liked about the software).

The feedback from these students demonstrates an important basic point. There must be value added by the comments and gestures the instructor makes. To try to make the in-class lectures more interesting instructors add value at presentation time. However, this means the slide shows alone are less informative. The instructor adds value with verbal comments, but the instructor also wants to point at times to add value. As this study shows, it is important for software capturing deictic gestures to only capture pointing that was intended, or the attention focusing effect of a moving pointer works negatively.

In response to the experiences the instructor had while creating these presentations and the feedback from the students in this class the following enhancements were later made to Point 'n' Talk:

1. To support situations where more careful wording is needed the ability to display a script in the same window as the image was added to Point 'n' Talk. This works better than a script far from the image being described, but speaking, reading a script and pointing somewhere else at the same time is obviously very difficult! A more reasonable strategy is to read a portion of the script, committing it to memory, then make a recording for that portion, and repeat this process.
2. To support unimodel and multimodal comments additional recording modes were added so that the user has three options for any portion of the presentation:
   (a) Voice only, so the pointer for deictic gestures is not seen;
   (b) Voice and pointing;
   (c) Pointing over previously created voice recording to allow the user to separate these tasks if that is cognitively less demanding. (Oviatt et al. (1997) found that pointing generally precedes speech, which explains why the results can sometimes appear slightly unnatural when pointing is added as a second phase.)

   Further, function keys and mouse click commands were introduced as options to control

the recording and thereby avoid recording the path to and from buttons or pull-down menu options. These changes made the interface a little more complex and less natural, but the utility gained was considered to outweigh that cost.

3. Actual size and best fits options were added for displaying the image in a window.

How useful software like this is for instructors partly depends on the extent to which they already use slide shows and how they tend to use them. A brief questionnaire was sent to the ACM's (Association for Computing Machinery) SIGCSE (Special Interest Group on Computer Science Education) list server, and 58 instructors responded. When asked, "Do you ever use a slide show to structure material to be covered in your classes (e.g. slides on transparencies, PowerPoint slides, Adobe Acrobat pages, a pre-planned sequence of Web pages, etc.)" 46 (79%) indicated they do and 12 (21%) indicated they do not. When the 46 who responded that they do use slides were asked, "On average, what percentage of the time in your classes is conducted using slides? (Do not include lab time)" there was great diversity in the answers given, but the result still represents a significant amount of time spent using slides: 13/46 (28%) said between 1% and 20%; 5/46 (11%) said between 21% and 40%; 12/46 (26%) said between 41% and 60%; 6/46 (13%) said between 61% and 80%; and 10/100 (10%) said between 81% and 100%. When asked, "Do students have access to the same set of slides?" 44/46 (96%) indicated they do and 2/46 (4%) indicated they do not. When asked, "Do you use annotated copies of your slides as part of your preparation for teaching your classes?" 20/43 (47%) indicated they do and 23/43 (53%) indicated they do not (3 gave no answer). These answers suggest that a lot of educational material is being created in the form of slide shows and faculty are normally providing copies of those slides to their students. It would therefore appear that there is a large number of faculty who might be able to use multimodal slide show creation software to enhance these slide shows. What effect using multimodal slide shows as part of the class preparation process would have is also a question for further research, as there could be some benefits from a more authentic representation of the portions of the class that are intended to take place in a lecture mode.

4. CONCLUSIONS

This paper is not claiming that slide show presentations outside the classroom are pedagogically preferable to other ways of teaching and learning. However, the fact is that a lot of instructors use slide shows and make them available to students, and the question is simply raised here how the value of those slide shows can be increased when they are used as a supplement in a traditional course, with modest additional workload demands on the instructor. In the study presented, multimodal slide shows were used for reviews partly because the instructor did not want to make them available until after the material had actually been covered in the classroom. In this case, the goal was clearly not to turn the course into a distance education class, but to assist in the review process, when the proximity of an examination seems to cause many students to be more motivated to learn than at other times during the term. At the same time, the results here are relevant for designing distance education courses where multimodal communication is being considered.

The reviews created with Point 'n' Talk were considered valuable by the students in this study, but a practical concern was the file size for students downloading them using modems. Creating a streaming version of the software or storing the slide shows on a compact disc would seem to be the obvious solution to this problem. A major attraction for these students was the multimodal nature of the reviews created, and particularly the deictic gesturing. This study also demonstrated however that this feature can be distracting from what was intended to be the focus of attention if used poorly. This is an important lesson for both software developers and instructors using this type of software.

5. REFERENCES

Bargeron, D., Gupta, A., Grudin J.& Sanocki, E. (1999). Annotations for streaming video on the web: System design and usage studies. *Proceedings of the Eighth International World Wide Web Conference.* Toronto, Canada.

Bargeron, D., Guputa, A., Grudin, J., Sanocki, E. and Li, F. (2001). Asynchronous collaboration around multimedia and its application to on-demand training. *Proceedings of the 34th Hawaii International Conference on System Sciences*, 1-10. Maui, HI.

Bly, S. A. (1988). A use of drawing surfaces in different collaborative settings. *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW88)*, 250-256. Portland, OR.

Chapanis, A. (1975). Interactive human communication. *Scientific American*. 232, 34-42.

Chapman, R. J., Smith, P.J., Klopfenstein, M., Jezerinac, J., Obradovich, J., and McCoy, C.E. (2000a). CSLANT: An asynchronous communication tool to support distributed work in the National Airspace System. *Proceedings of the 2000 Annual Meeting of the IEEE Society on Systems, Man and Cybernetics*, 1069-1074. Nashville, TN.

Chapman, R. J., Smith, P.J., Klopfenstein, M., Jezerinac, J., Obradovich, J., and McCoy, C.E. (2000b). Supporting collaboration in the National Airspace System with multimodal, asynchronous communications. *Proceedings of the 2000 Conference on Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*. 324-329. Savannah, GA.

Daly-Jones, O., Monk, A., Frohlich, D.M., Geelhoed E. & Loughran S. (1997) Multimodal messages: The pen and voice opportunity. *Interacting with Computers* 9, 1-25.

Faraday, P. M. and Sutcliffe, A. G. (1997). Designing effective multimedia presentations. *Proceedings of CHI '97, ACM, 272-279*. Atlanta, GA.

Kruse, R. L., and Ryba, A. J. (1998). *Data Structures and Program Design in C++*. Upper Saddle River, NJ: Prentice Hall.

McNeill, D. (1992) *Hand and Mind: What Gestures Reveal about Thought.* Chicago, IL: University of Chicago Press.

Neuwirth, C. M., Chandhok, R., Charney, D., Wojahn P., and Kim L. (1994). Distributed collaborative writing: A comparison of spoken and written modalities for reviewing and revising documents. *Human Factors in Computing Systems*. April 24-28, 51-57.

Oviatt, S., DeAngeli, A., and Kuhn K. (1997). Integration and synchronization of input modes during multimodal human-computer interaction. *Proceedings of CHI '97, ACM, 415-422*. Atlanta, GA.

Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*. 42(11), 74-81.

Tang, J. C. (1991). Findings from observational studies of collaborative work. *Int J Man Machine Studies,* 34(2), 143-160.

Ware, C. (2000). *Information visualization: perception for design*. San Francisco, CA: Morgan Kaufmann Publishers.