The Design Amanuensis

An Instrument for Multimodal Design Capture and Playback

Mark D. Gross, Ellen Yi-Luen Do, and Brian R. Johnson
Design Machine Group, University of Washington

Key words: Protocol analysis, recording, design process research

Abstract: The Design Amanuensis* supports design protocol analysis by capturing designers' spoken and drawing actions and converting speech to text to construct a machine-readable multimedia document that can be replayed and searched for words spoken during the design session or for graphical configurations.

1. INTRODUCTION

Design researchers have used think-aloud, or protocol analysis, studies of designers in action to better understand their design processes. Protocol analysis studies are used across the domains in which cognitive scientists seek to understand human problem-solving. Protocol analysis (Newell, 1968) was brought to the attention of the emerging fields of cognitive science and artificial intelligence in 1972 by Herbert Simon and Allen Newell's important book, "Human Problem Solving" (Newell, 1972). Newell and Simon were among Carnegie Mellon's most prominent scientists and this method of protocol analysis was employed in some of the first empirical studies in design research by Carnegie Mellon researchers Eastman (Eastman, 1968) and later Akin (Akin, 1978). Donald Schön based his analysis of design as an example of reflection-in-action on an analysis of protocols taken in architecture design studios at MIT (Schön, 1985). Many

* aman·u·**en**·sis: Latin, from *a manu* slave with secretarial duties; one employed to write from dictation or to copy manuscript (Merriam Webster Collegiate Dictionary)

subsequent research studies (e.g., Goel's Sketches of Thought (Goel, 1995); Goldschmidt's studies of visual reasoning in design (Goldschmidt, 1992)) have employed the study of think-aloud protocols, including a recent workshop (Cross, Christiaans and Dorst, 1996) that brought together a diverse collection of research teams who studied the same design protocol to compare their analysis and conclusions. Think-aloud protocol analysis has some reported shortcomings, notably that talking aloud interferes with performing the task at hand (Wilson, 1994); nonetheless it is widely accepted as a useful method of design research.

In a typical think-aloud study designers are given a problem and then they are observed, audio-taped, and often videotaped as they work the problem. Later, the design researcher transcribes the taped record of the design session and using his or her notes constructs a time labelled record of the events in the session. The study may examine a single designer working alone or a team of designers working together. The record typically consists of the drawings made and the designers' comments; in addition it may indicate gestures such as pointing to portions of the drawing or body movement. This transcript is typically the main corpus of research material used for further analysis.

Many researchers proceed from the transcript to make time-based representations of the design history, from which they reason further about the design process. These 'design scores' take various forms, for example Goldschmidt's Link-O-Gram diagrams (Goldschmidt, 1990). These more abstract representations of the design process may be more directly useful in analysing design behaviour, but they also are abstractions or translations of the raw protocol data captured during the design session.

Valuable as think-aloud studies are in design research, they also require a great deal of tedious effort on the part of the researcher. One of the more laborious tasks in conducting a protocol study is transcribing the design session after it has been recorded on tape. Potentially relevant events in a design session often happen every few seconds, and co-ordinating the times of events in different modalities is important. For example, based on a protocol analysis study Akin and Lin report that novel design decisions usually occurred when the designer was in a "triple mode period": drawing, thinking, and examining (Akin and Lin, 1995). This type observation can only be made based using a time-accurate record of the design session.

Depending on the level of detail that the researcher is interested in, a one-hour design session can demand tens of hours of transcription time. The researcher must carefully watch a video tape frame-by-frame to locate start and end times for events and transcribe the words spoken during the session. A single design session may result in a transcript that is tens of pages long.

We (Gross and Do) first realised the need for a design recording instrument when Do was capturing data for her doctoral dissertation (Do, 1998), which involved transcribing and analysing videotapes of four one-hour design sessions. A great deal of effort was spent locating the precise times of spoken comments, drawing marks, and associating them with drawings made on paper, or images of drawings recorded on video tape. Although several research efforts have tried to support recording and analysis of multimodal conversations, no off-the-shelf tools were readily available for our purpose.

The Design Amanuensis project aims to facilitate this transcription, by helping capture the spoken and drawn events of a design session, and by constructing a machine-searchable transcription that serves as a pointer into the source data captured during the original design session. We have built a first working version of this system that captures the designer's drawings using a digitising tablet and pen as well as the spoken think-aloud protocol. The captured audio is run through an off-the-shelf speech recogniser to generate a text transcription of the design session. The three components: drawing, audio, and text transcript, are arrayed in an interactive multimedia document. The design researcher can then review, correct, and annotate this document to construct a transcript of the design session.

As current speech-recognition software is somewhat unreliable, it is important to allow the design researcher to repair the machine-made text transcript; nevertheless, starting with an initial transcription is a significant improvement over starting from scratch. The speech recognition software associates the text transcript with the corresponding recorded audio so these repairs are easier to make than working with an ordinary audio or video recording.

The paper reports on the design and implementation of the Design Amanuensis. We review some of the specific challenges in constructing this system and how we addressed them, as well as our plans for enhancing this system in the future. Finally, we discuss other applications for the Amanuensis, including its application as a note-taker for distributed design meetings that take place over the Internet.

2. RELATED RESEARCH

Many efforts to build recording and indexing systems for mutli-modal information (e.g (He, Sanocki, Gupta et al., 1999)) are designed for automatically indexing streamed video, for example newscasts and lectures. Others are designed to "salvage" meetings; to create a useful, searchable record of a multimodal conversation taking place at a whiteboard or around a

table. Still other "personal notebook" efforts focus on studying or supporting the individual note-taker engaged in a specific task, for example, listening to a lecture.

In light of the relatively small number, world wide, of design researchers, it is unsurprising that we could find only one project specifically designed to record and index design protocols. With much the same goal as we have in mind, McFadzean, Cross, and Johnson (McFadzean, 1999; McFadzean, Cross and Johnson, 1999), have developed an experimental apparatus called the Computational Sketch Analyser (CSA), which they use for capturing design protocols. The CSA records drawing marks on a digitising tablet, classifies, time stamps, and co-ordinates them with a videotaped record of the drawing activity. However, the CSA does not transcribe the think-aloud protocol, leaving this task to the researcher, nor does it enable the researcher to search the record. In reviewing a design protocol that may extend to fifty or more transcribed pages, it is useful to be able to search the record for specific words or phrases, as well as specific drawing marks.

Oviatt and Cohen's Quickset program aims to support multimodal conversations, for example, military mission planning (Oviatt and Cohen, 2000). Quickset's important contribution to Human-Computer Interface research is using context from one mode (e.g., speech) to aid recognition in another mode (drawing). However, the Quickset system does not record speech continuously and independently of drawing acts. Rather, each speech event is associated directly with a drawing event: audio recording for that event begins when the pen goes down. Continuous and independent speech and drawing are typical in design protocols, and so Quickset would not serve our purpose.

Stifelman's Audio Notebook (Stifelman, 1996) was designed as a note-taking aid for a student attending a lecture. The student's handwritten notes were recorded using a digitising tablet and the lecture was simultaneously recorded in an audio file. The audio track and hand written notes were co-ordinated using the time stamps, so that a student could review the lecture by pointing to the notes he or she had taken, which served as an index into the audio track. The Audio Notebook did not perform speech recognition, so the record was indexed by marks made on the pad, but it could not be searched.

The Classroom 2000 project (Abowd, Atkeson, Feinstein et al., 1996; Abowd, 1999) used ubiquitous computing concepts to capture lectures, instructor's whiteboard notes, and student notes on personal handheld devices, and to later enable students to access these records of a class to enhance learning. Initially focused on capture and replay of lectures and notes, the project has recently explored using commercial speech recognition software to transcribe spoken lecture material.

Moran's Tivoli project has explored capture and indexing of spoken audio and whiteboard drawing to support 'salvaging'—an activity involving "replaying, extracting, organizing, and writing"—the records of meetings. (Moran, Palen, Harrison et al., 1997). This project also did not attempt to transcribe the spoken audio, but provided a case study of how the system was used over an extended period of time.

3. DESIGN PROTOCOLS

3.1 Example of a design protocol

Figure 1 shows a typical excerpt from a design protocol taken without the aid of the Design Amanuensis (Do, 1998). Entries in the leftmost column contain frames clipped from a videotape made during the session, in which the content was aimed at the designer's work area. The second column indicates the elapsed time in minutes and seconds since the session began. The third column records the designer's spoken comments (e.g., "ummm... what if we put the main entry here..." as well as the designer's physical actions (e.g., "picks up pencil and begins to draw rectangles"). The fourth column contains the researcher's annotations, and the rightmost column contains scanned excerpts of the design drawing made at this point in the protocol.

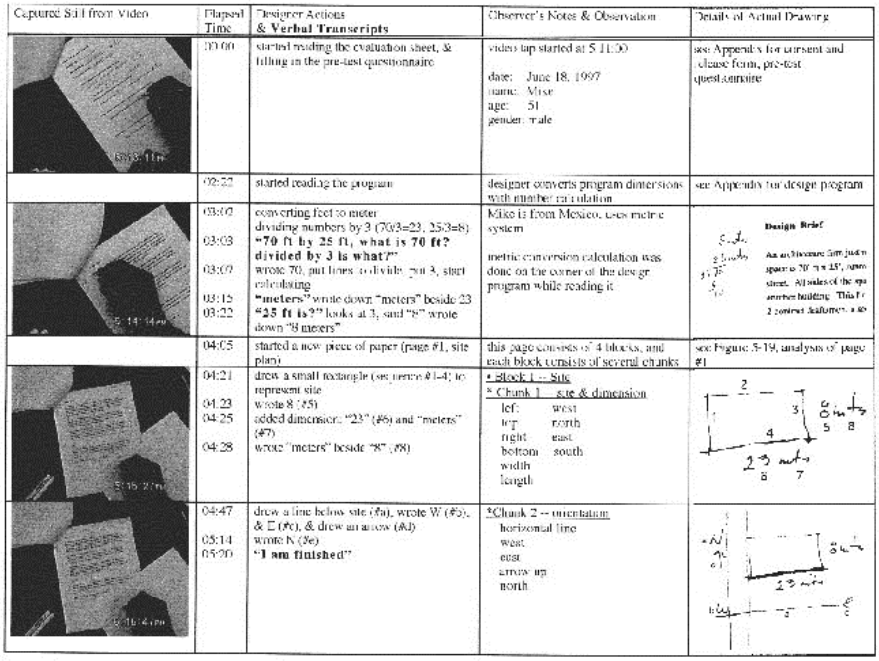

Figure 1. Excerpt from a design transcript showing frames from videotape, transcribed think-aloud protocol, and design actions

3.2 Recording a design protocol

The Design Amanuensis helps create this type document by recording all drawing and audio data as it is produced. The designer draws on an LCD digitising tablet and all spoken remarks are recorded as digital audio. Although the digitising pen and tablet are quite comfortable and natural to use, some time is required for the designer to become familiar with this mode of drawing. Similarly, the speech-to-text software works best when trained for the individual speaker. Once the designer is familiar with the drawing environment and the speech recognition software has been trained, the design recording begins. As the designer draws and talks, the program records these actions into files.

3.3 Reviewing the design history

When the design session is finished, the Design Amanuensis processes the files and makes them ready for replay. Then, the researcher can play back the entire design session from the beginning, or any portions of it. The

researcher can move forward and backward through the session (using a tape-recorder style control panel), or search for words in the text transcript or for drawn figures. The researcher can also point to a specific element of the drawing to review the corresponding portion of the design record. As the researcher reviews the record, he or she can correct erroneous text transcriptions by simply selecting and retyping them as in a text editor, as well as add annotations to the record.

Figure 2 shows the simple player interface for browsing the design record. Interface buttons enable "rewinding" to the beginning of the session, moving backward or forward one (speech or drawing) event at a time, selecting an element from the drawing to set the playback session clock as well as loading new session data and establishing a time mark. The three numbers display the start timestamps of the most recent previous event (left), the current event (centre) and the next event (right).

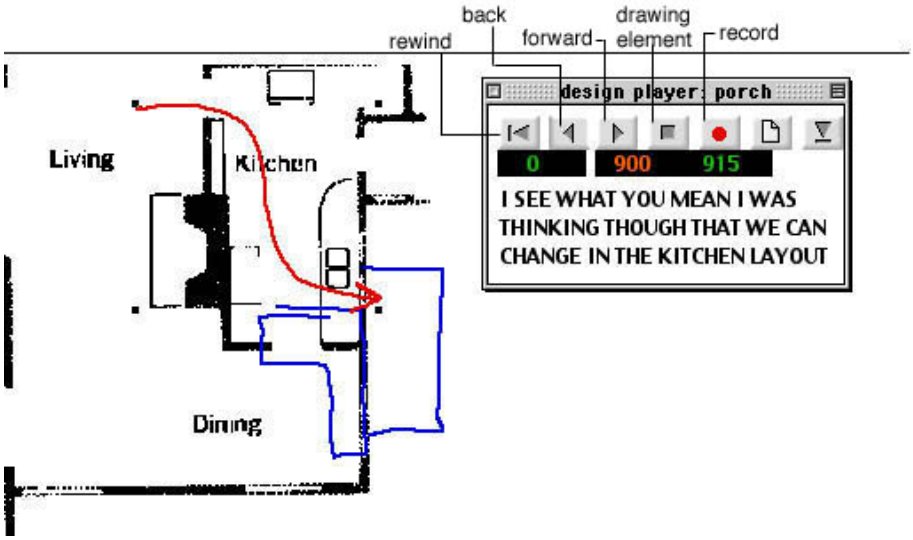

Figure 2. Playing back part of a design conversation

4. IMPLEMENTATION

The Design Amanuensis is implemented on Macintosh computers using a combination of software components. Drawing capture and management is handled by an extension of an existing research software program, the Electronic Cocktail Napkin (Gross and Do, 2000). Speech capture and recognition is handled by an off the shelf commercial application, IBM's

ViaVoice™ for Macintosh. Audio file format translation is handled by Norman Franke's excellent freeeware utility, SoundApp PPC and inter-application control is handled by Apple's AppleScript language. All these components are co-ordinated by code written especially for this purpose in Macintosh Common Lisp.

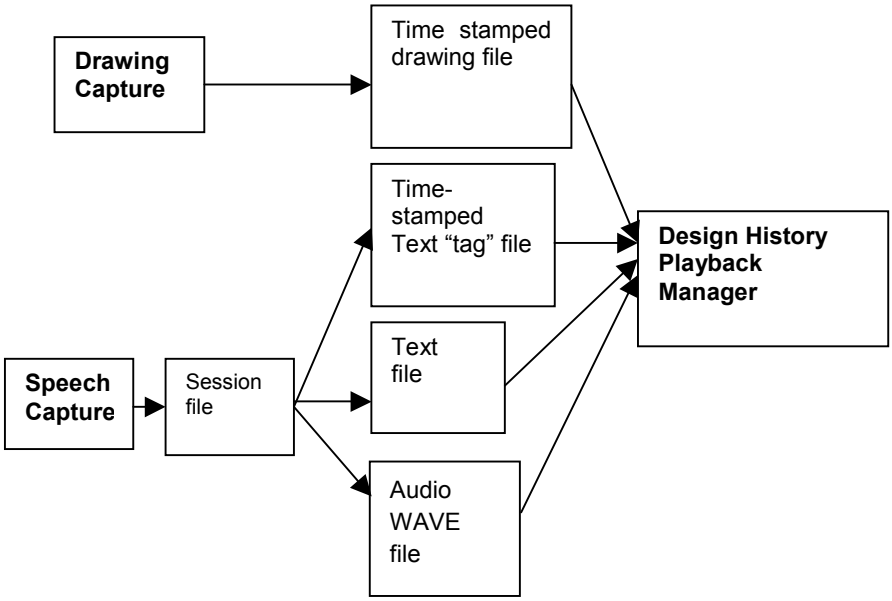

Figure 3. System diagram of the Design Amanuensis showing the two capture modules (drawing and speech capture), the four main files these modules create, and the design history playback module.

4.1 Drawing capture

The drawing capture component of the Amanuensis is an extension of the Electronic Cocktail Napkin. This program records each drawing mark as it is made, and records with each mark the time when it was initiated. In addition to recording the (x, y) co-ordinate stroke information for each drawing mark, the Electronic Cocktail Napkin attempts to recognise the marks as they are drawn. The Napkin program produces a drawing file that contains the entire set of time stamped drawing marks.

4.2 Speech capture and transcription

The Amanuensis uses IBM's commercial product, ViaVoice™ for Macintosh, for speech capture and transcription. This program records audio and transcribes the audio record to text. ViaVoice produces a single "session file" that contains the recorded audio, the text transcription, and tags that indicate the starting and ending times for each identified word and phrase in the audio, as well as alternative transcriptions of each word. The Amanuensis parses this single large file into several smaller component files (Figure 3). It also uses a helper application (SoundApp PPC) to convert the audio from the WAVE format that ViaVoice™ produces to the Mac native AIFF format.

4.3 Initiating a design session recording

Initiating a design session recording is straightforward: The program must simultaneously start recording audio and zero the clock for the Cocktail Napkin drawing program. In our current version these applications run in parallel on separate machines, because the speech recording software requires a dedicated processor. An AppleScript™ program launched from the Common Lisp environment tells ViaVoice™ to start recording and the Napkin program zeroes its clock. When the session is over, a similar AppleScript is launched to tell ViaVoice™ to stop recording and to save the session in a file. Once the file is saved, the Amaneuensis co-ordinating software parses ViaVoice™'s session file into its several components, again using AppleScript to call on SoundApp PC to convert the WAVE format audio to the Mac native AIFF. Once this is accomplished, the design session is ready for playback.

4.4 Design history playback

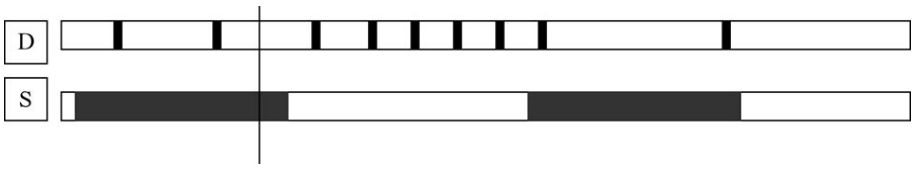

Figure 4. A diagram of the drawing (D) and speech (S) tracks of a design history. Several relationships are possible between the start and end of a speech event and the previous and next drawing events.

Figure 4 illustrates the two tracks of the design history. The drawing track consists of discrete events that have a single time-stamp, the time of the pen-down event that initiated that drawing mark (indicated by vertical lines in the "D" track). The speech ("S") track consists of continuous segments of speech alternating with silence (indicated by blocks in the track). A thin vertical line crossing the two tracks indicates the 'current time' being replayed on the recorder.

Events on the two tracks may have several different relationships: The designer may have been drawing silently, speaking without drawing, or drawing while speaking. Replaying the record from the beginning of the session (or any given time) is straightforward: set the clock to the desired time, find the nearest events in the speech and drawing record, and begin replaying the two tracks in time sequence. To ensure that the speech track remains intelligible, the speech track is always replayed starting at the beginning of the word that was being spoken.

For careful study of the record it is useful to 'single step' through the design history, moving forward one event at a time. When the designer presses the forward button on the playback console, the Design Amanuensis examines the audio and drawing tracks to determine which contains the next event. It then plays that event, displaying the text and playing the audio if the next event is on the speech track, or revealing the next drawing mark if the event is on the drawing track.

Specifically, if the next event is a drawing event, the Amanuensis will play (or replay) any speech event whose start and end times bracket the drawing event time. If the next event is a speech event, the Amanuensis will play any and all drawing events that occurred between the start and end times of the speech event.

It is also useful to be able to point at a mark on the drawing, and review the conversation that was taking place at that time (a common feature in audio indexed notebook systems). When the reviewer selects a drawing mark and requests to hear the audio, the system sets the clock to the timestamp associated with the drawing mark and plays a single event at that time.

Finally, the system supports both text and graphical search. The text search feature enables the reviewer to search for a word or phrase in the transcribed record, and set the playback clock to the time associated with that phrase. For example, the reviewer could request to review the design record associated with the phrase "light from the west." The system would find the point in the design session where these words were spoken and set the playback clock to the appropriate time. Similarly, the reviewer can use a graphical search to find parts of the record where a particular drawing was made. The Electronic Cocktail Napkin's diagram similarity matching is

employed to find parts of the drawing that are similar to the search sample the reviewer provides, and then the playback clock is set to the time stamp associated with the part of the diagram found.

5. FUTURE WORK

Our main goal is to use the Design Amanuensis as a tool for design protocol analysis in empirical studies of designers in action. We plan to develop and refine the capabilities of the program for this purpose, as we have found design protocol studies useful in understanding what designers are doing, but we have been inhibited in conducting such studies by the sheer amount of data to be managed and the lack of appropriate tools to manage this data.

We do, however, see a wider potential application of this work. An application of the Design Amanuensis is to record design conversations, both where the designers are co-located and where several designers who are geographically distributed collaborate via the Internet, drawing, writing, and talking. Our research laboratory space was recently renovated, and over the course of six months we participated in a number of design meetings at which drawings were reviewed, changes suggested, and decisions taken. On several occasions we wished to review the meetings (some of which lasted two hours), and specifically to recall what had been said about a particular topic or physical decision. A searchable digital record would have been useful.

An obvious addition to the current system is co-ordinating a video record of the design session. This would enable design researchers to include in the record gestures made over the drawing—which often include references to drawn elements—as well as body language that the current system does not capture. Time-stamped video would also be valuable for the distributed meeting support application of the system.

Finally, going beyond the immediate applications of protocol analysis and distributed meeting support, the system suggests an approach to constructing informative—yet informally structured—hyperdocuments that include mark-up and annotation on traditional design drawings, that record spoken and transcribed design rationale, in a way that can be browsed by people and searched by machine.

6. ACKNOWLEDGEMENTS

This research was supported in part by the National Science Foundation under Grant No. IIS-96-19856 and IIS-00-96138. The views contained in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

7. REFERENCES

Abowd, G. D., 1999, “Classroom 2000:An experiment with the instrumentation of a living educational environment.” *IBM Systems Journal* 38(4), p. 508-530.

Abowd, G. D., C. G. Atkeson, A. Feinstein, et al., 1996, "Teaching and Learning as Multimedia Authoring: The Classroom 2000 Project", In *ACM Multimedia '96*, ACM, Boston, p. 187-198.

Akin, O., 1978, "How Do Architects Design", In E. J.-C. Latombe (Eds.), *Artificial Intelligence and Pattern Recognition in Computer Aided Design, IFIP*, North-Holland Publishing, New York, p. 65-104.

Akin, O. and C. Lin, 1995, “Design Protocol data and novel design decisions.” *Design Studies* 16(2), p. 211-236.

Cross, N., H. Christiaans and K. Dorst, eds., 1996, *Analyzing Design Activity*, John Wiley & Sons, New York.

Do, E. Y.-L., 1998, The Right Tool at the Right Time: Investigation of Freehand Drawing as an Interface to Knowledge Based Design Tools, Georgia Institute of Technology, .

Eastman, C. M., 1968, "On the Analysis of Intuitive Design", In G. T. Moore (Eds.), *Emerging Methods in Environmental Design and Planning*, MIT Press, Cambridge, p. 21-37.

Goel, V., 1995, *Sketches of Thought*, MIT Press, Cambridge MA.

Goldschmidt, G., 1990, “Linkography: Assessing Design Productivity.” *Tenth European Meeting on Cybernetics and Systems Research.*

Goldschmidt, G., 1992, “Serial Sketching: Visual Problem Solving in Designing.” *Cybernetics and Systems: An International Journal* 23, p. 191-219.

Gross, M. D. and E. Y.-L. Do, 2000, “Drawing on the Back of an Envelope: a framework for interacting with application programs by freehand drawing.” *Computers and Graphics* 24(6), p. 835-849.

He, L., E. Sanocki, A. Gupta, et al., 1999, "Auto-summarization of audio-video presentations", In *ACM Multimedia 1999*, ACM SIGCHI, p. 489-498.

McFadzean, J., 1999, "Computational Sketch Analyser (CSA): Extending the Boundaries of Knowlwedge in CAAD", In *eCAADe'99: 17th International Conference on education in Computer Aided Architectural Design Europe*, The University of Liverpool, UK., p. 503-510.

McFadzean, J., N. Cross and J. Johnson, 1999, *Notation and Cognition in Conceptual Sketching*. (VR'99) Visual and Spatial Reasoning in Design, MIT, Cambridge, USA.

Moran, T. P., L. Palen, S. Harrison, et al., 1997, "I'll get that off the audio": A case study of salvaging multimedia meeting records", In *CHI'97 Conference on Human Factors in Computer Systems*, ACM, Atlanta,GA, p. 202-209.

Newell, A., 1968, "On the analysis of human problem solving protocols", In J. C. Gardin and B. Jaulin (Eds.), *Calcul et formalisation dans les sciences de l'homme*, Centre National de la Recherche Scientific, Paris, p. 146-185.

Newell, A., & Simon, H.A., 1972, *Human problem solving*, Prentice-Hall., Englewood Cliffs, NJ.

Oviatt, S. and P. Cohen, 2000, "Multimodal Interfaces That Process What Comes Naturally." *Communications of the ACM* 43(3), p. 45-53.

Schön, D., 1985, *the Design Studio*, RIBA, London.

Stifelman, L. J., 1996, "Augmenting Real-World Objects: A Paper-Based Audio Notebook", In *Human Factors in Computing (CHI) '96*, ACM, p. 199-200.

Wilson, T. D., 1994, "The Proper Protocol: Validity and Completeness of Verbal Reports." *Psychological Science: a journal of the American Psychological Society / APS.* 5(5 (Sep 01)), p. 249-252.