

Multimedia Chronicles for Business Communication

Marko Balabanović
Ricoh Silicon Valley
2882 Sand Hill Road, Suite 115
Menlo Park CA 94025, USA
marko@rsv.ricoh.com

Abstract

Email is increasingly popular for business communication, often as a means of transmitting, commenting on or collectively creating documents. Although text email is easy to author, it lacks the rich expressiveness of oral media. Multimedia chronicles are introduced as conversations that weave through a fabric of documents, URLs, photographs and other resources. They consist of audio narrations with references to existing digital objects. Over time, an accumulated collection of intersecting chronicles forms a valuable organizational memory.

This paper describes a simple point-and-talk interface for composing and viewing multimedia chronicle messages. The design incorporates the semi-automatic creation of a visual representation, in order to improve usability and comprehension for a viewer. In addition, a prototype interface for accessing a collection of such messages is described.

1. Introduction

Researchers of communications media have noted that there is often a tradeoff between rich, oral media and less rich, written media [9, 11]. Oral media enable negotiation, clarification, explanation and exchange of subjective views. Written media enable the exchange of large amounts of accurate, objective or numeric data. This kind of split has been observed, for instance, with annotations of documents: Chalfonte *et al.* [8] have shown that users prefer to use voice for commenting on higher-level, global issues in a document, and prefer to use text for commenting on lower-level, local issues such as spelling and grammar. In the electronic realm, simple textual email messages, although easy to author, do not allow rich, expressive communication as is sometimes required. On the other hand, tools for creating richer, more expressive messages, for instance multimedia

presentations, are too complex and time-consuming for casual or day-to-day use.

Existing text-based messaging systems can successfully be used to participate in and record conversations. For instance, it is common for email users to create conversations consisting of “mosaic” messages that include the text of previous relevant messages in the conversation [22]. Often each such mosaic message includes a record of the entire conversation up to that point. This kind of structure is also used by many chat rooms, bulletin boards and persistent conversation systems such as *Babble* [13]. For multimedia messages, however, there is not a generally accepted mechanism to represent these kinds of persistent conversations.

This paper describes interaction designs for a communication system that extends some of the affordances available with text-based persistent conversations to the multimedia realm. The system has four components:

- The *multimedia chronicle* data structure, a restricted kind of hypermedia structure, with an associated XML representation [6];
- The *TouchVerse* interface for composing and viewing multimedia chronicles;
- An underlying infrastructure for storage, access control and transmission of multimedia chronicles, utilizing email and Web standards;
- An HTML interface to a database of multimedia chronicles, and to individual multimedia chronicles when the specialized *TouchVerse* application is not available.

A working prototype of this system has been deployed to 13 test users.

The remainder of this introduction elaborates on the tradeoffs between different communications media. These

lead to design goals for the multimedia chronicle system, first with respect to the composition and viewing of individual messages, and second with respect to ongoing and archived conversations. Following the introduction, the multimedia chronicle structure is described in detail. Next, both the TouchVerse interface and the underlying infrastructure are explained. The paper concludes with some preliminary usage and usability findings, related work and a summary.

1.1. Tradeoffs among communications media

The *richness* of a communications medium has been mentioned but not defined. Daft and Lengel [9] define *information richness* as the ability of information to change understanding within a time interval. Their reasons for differences in the richness of media are fourfold:

1. Timely feedback;
2. Cues, such as tone of voice and gestures; utilization of parallel channels;
3. Personalization, the ability to convey personal feeling and emotion;
4. Language variety.

Thus face-to-face communication is the richest medium, followed by phone calls, personal written communications including email, fax and letters, and finally impersonal or numeric documents.

In the *media richness theory* formulation, two forces influence the choice of medium for a message: uncertainty (the absence of information) and equivocality (ambiguity; multiple conflicting interpretations). Rich media are well suited to reduce equivocality, while less rich media might be better for the reduction of uncertainty. Section 5 will illustrate how test users choose among communications media for situations of uncertainty versus equivocality. Note that this formulation considers individual messages, whereas a persistent conversation could exhibit a cumulatively greater richness.

In contrast to media richness theory, the *social network perspective* views the type and intimacy of the relationship between correspondents as the main factor determining the choice of media. A study has shown that the frequency of information exchange between two people is correlated with the number of different media used [17], suggesting that social factors are as important a consideration as message content. Indeed, richness is not the only salient difference between media types. Email, steadily increasing in popularity for business communication, provides a unique set of capabilities:

- It enables asynchronous communication, like other forms of written communication;
- Transmission time and cost are minimal;
- Recipients can be specified individually, unlike when sending a fax to a machine shared among a workgroup or office;
- Multiple recipients can easily be specified at no additional time or cost;
- Digital objects can easily be attached to a message;
- All commonly used email client software includes the ability to store, sort and organize past messages, both sent and received. Although there is no reason in principle why voice mail could not be used the same way, the telephone interface discourages it, and as a result users tend to perceive voice mail messages as ephemeral [11].

This combination of features can cause email to be chosen even when a richer medium may be more appropriate to the message in question [12].

Of course, within the email standard it is possible to include multimedia attachments, allowing for an increased richness. However, this kind of multimedia communication has been slow to be adopted. We argue that there are two main reasons: authoring complexity and viewing difficulty. Tools for creating multimedia presentations tend to be complex and feature-rich, aimed at professional multimedia producers more than end-users. It is relatively easy to record and send simple audio or video segments, but these linear streams are time-consuming and inefficient to view. When reading text documents, users scan over sections, move around a document and judge the length and content without reading fully [4]. None of these affordances are readily available with basic audio and video players.

In a business setting, messages and conversations are often about documents—an interesting Web page is forwarded to a colleague, a report is written collaboratively, comments are invited on a specification, there are questions about a form to be filled in, and so on. In written media, such as text email, although it is easy to refer to a document (by including a URL or an attachment), it is cumbersome to refer to regions within it. Unlike in a face-to-face interaction, the relevant part cannot be pointed to.

1.2. Design goals

The arguments of the previous section suggest that, despite the arsenal of communications tools at our disposal, there is still an unmet need. A tool is required that allows users to compose rich messages, easily include references to

documents and parts of documents, with otherwise equivalent capabilities to text email. One set of design goals for such a tool relates to the composition and viewing of individual messages:

- Authoring complexity should be minimal, with message creation no more daunting than text email composition;
- The design should enable effective viewing—a recipient needs an overview as well as the ability to jump around within a message;
- All key features of text email should be supported.

A further set of goals relates to conversations and collections of conversations. As has already been mentioned, conversations over email are often conducted in the form of mosaic messages that grow as they are sent back and forth. In addition, every participant has the opportunity to store individual messages. When retrieving a document, the conversations about it, if available, provide valuable context and history. Furthermore, the metadata or content of such conversations are cues that may have enabled the retrieval in the first place. However, it is often the case today that organizations store messages entirely separately to documents, especially where groupware or knowledge management technologies have not been deployed. These observations lead to the second set of design goals:

- Users should be able to contribute to and store mosaic messages;
- The system should support the storage, browsing and retrieval of both conversations about documents and documents referenced in conversations.

The utility of such a conversation database cannot be estimated without a long period of usage, allowing sufficient documents and conversations to accumulate. This paper describes a design and implementation, but usage results are reported only for users' shorter-term personal communications needs.

2. Multimedia chronicles

The communications system to be described relies on a particular type of hypermedia structure, a multimedia chronicle. Concentrating on a restricted class of structures allows a simpler authoring interface to be designed (coming up in section 3). The intent is to capture some natural and useful forms of communication, not to cover all possible situations.

A multimedia chronicle (Figure 1) consists of one or more *sub-chronicles*, where each sub-chronicle in turn consists of a single *narration* thread and multiple *references* to

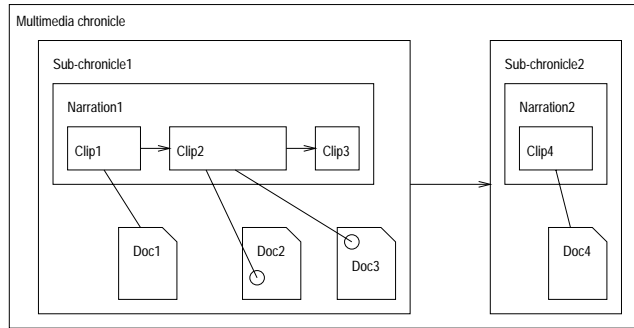


Figure 1. Multimedia chronicle structure.

documents. As shown in the figure, the narration thread is segmented into an ordered list of separate *clips*. The references optionally include a specific point within a graphical representation of the document¹. A linear ordering over the sub-chronicles can be specified, allowing a default playback order. In general we refer to this as a “narration + references” structure.

In the implementation to be discussed, the narration takes the form of audio (recorded voice). Each reference comprises a timestamp (relative to the start of the narration) and the URL of the referenced document. Each narration is further timestamped with its time of recording, and these timestamps are used to order the sub-chronicles. Extensions where the narration is a text stream can easily be imagined. However, since part of our motivation is to investigate the utility of rich oral messages, the system described here does not allow users to add typed text.

Consider an example application of multimedia chronicles: capturing narratives that relate to a collection of digital photographs. In this case a user talks about the photos, and points to them as they are referred to. The referent of a pointing gesture can be a whole photo (“*this picture is our family*”), a point within it (“*this is Great Uncle Alex*”), or a more complex region (“*from here to here is where we went*”). In a straightforward implementation, a chronicle plays back as a slide show, with the photos appearing as they are first referred to.

A set of XML tags are used to represent a multimedia chronicle (Figure 2). This allows both for easy transportability (XML input and output libraries are available for many popular programming languages and database systems) and for extensibility (extra tags or attributes added by later versions will be ignored by present systems).

¹The coordinates of such a reference point have a different interpretation depending on the type of the document. For instance, if the document is an image, they can be absolute pixel coordinates. If the document is a Web page, they can be a character position within an HTML file.

```
<CHRONICLE SAVETIME="927336993806" AUTHORID="marko">
<SUBCHRONICLE AUTHORID="marko" STARTTIME="927336940440">
<CLIP><AUDIOREFS LENGTH="3240">
<AUDIO SRC="http://Server/Clip1.au"/>
<CLIPREF TIME="1753" REF="http://Server/Doc1"/>
</AUDIOREFS></CLIP>
<CLIP><AUDIOREFS LENGTH="11600">
<AUDIO SRC="http://Server/Clip2.au"/>
<CLIPREF TIME="2293" POINTX="250" POINTY="750"
REF="http://Server/Doc2"/>
<CLIPREF TIME="6259" POINTX="330" POINTY="250"
REF="http://Server/Doc3"/>
</AUDIOREFS></CLIP>
<CLIP><AUDIOREFS LENGTH="1920">
<AUDIO SRC="http://Server/Clip3.au"/>
</AUDIOREFS></CLIP>
</SUBCHRONICLE>
<SUBCHRONICLE AUTHORID="marko" STARTTIME="927336977200">
<CLIP><AUDIOREFS LENGTH="6560">
<AUDIO SRC="http://Server/Clip4.au"/>
<CLIPREF TIME="4086" REF="http://Server/Doc4"/>
</AUDIOREFS></CLIP>
</SUBCHRONICLE>
</CHRONICLE>
```

Figure 2. XML representation for the multimedia chronicle of Figure 1 (tags relating to graphical layout omitted).

3. TouchVerse user interface

This section describes the TouchVerse application, which serves as both a composition and a viewing tool for multimedia chronicles. A natural fit exists between “narration + references” data structures and a “talk + point” interaction, as in the photo application already described. Therefore to use TouchVerse you talk (recorded as the narrative) and point at documents or points within documents (recorded as references). An important goal was to keep interaction as simple as that suggested by the “talk + point” slogan.

The application runs in a Microsoft Windows desktop environment, and assumes the user has access to all the usual tools such as a Web browser, a file hierarchy viewer, and so on. The application was built using Java 2, with additional native libraries for sound recording and playback.

3.1. Composition

Initially TouchVerse presents the user with a blank screen. Clicking anywhere causes recording to begin, indicated by the appearance (at the clicked point) of a colored bar that grows as recording continues. The bar represents the audio being recorded: its length is proportional to the length of the audio. The TouchVerse window thus provides a freeform surface on which audio clips can be arranged, mapping inherently sequential and temporal speech streams onto the spatial domain. Clicking elsewhere on the screen stops the first bar growing and begins a second bar at the

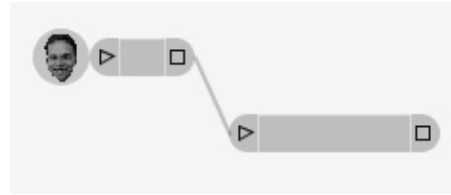


Figure 3. Two recorded audio clips, denoted by bars, as shown in the TouchVerse interface. Each has associated “play” and “stop” buttons; the first is preceded by an icon of the narrator’s face.

new clicked point (Figure 3). The two bars are connected to indicate the recording order. Each bar has “play” and “stop” buttons at either end; in addition, the first bar is preceded by a miniature image of the narrator’s face.

Any document can be dragged into the TouchVerse window from the Windows desktop or another application. The default representation for a document is an icon based on its inferred type. Some document types have associated handlers that can infer or fetch thumbnail representations. These include:

- Images (JPEG and CompuServe GIF) can be displayed directly, allowing, for instance, photographs to be included as in the earlier example.
- Documents from an internal “IM³” scanned-image database [18] (dragged in as URLs from a Web browser) have associated thumbnails that can be fetched. This database automatically captures all printed and copied documents at the test location, so it covers a large proportion of the documents people need to discuss.

The interface allows the author to choose among three different icon/thumbail sizes—holding the mouse down over a document icon causes it to cycle between available sizes.

Pointing to a document during recording creates a reference, represented by a line—Figure 4 shows four such references. In keeping with the freeform nature of the display, all objects can be dragged around and rearranged at will. Bars and lines are rendered semi-transparently, allowing them to be overlaid on document images.

No *a priori* meaning has been assigned to the placement of audio clips or the segmentation of a narration into multiple clips. It will be interesting to see if conventions arise. For instance, starting a new clip may become akin to starting a new paragraph when composing a text email.

An open question for future research is whether users would find additional visual tools useful, for instance the

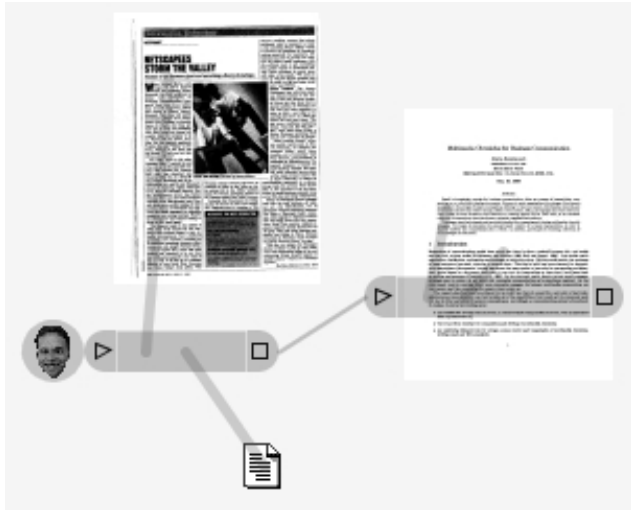


Figure 4. This multimedia chronicle has two audio clips. The first (leftmost) references a photocopied news article (represented by a thumbnail) and a text document (represented by an icon). The second references two points within a paper under development (space does not permit showing the much larger thumbnail that was actually used in this case).

ability to specify emotion or intention. As opposed to the usual *glottographic* writing systems, such as that used for English, where symbols represent the sounds of human speech, *semasiographic* writing systems have symbols that directly represent meaning [26], for instance the “picture writing” of some Native American tribes. The “smiley” emoticon :-) is a semasiographic symbol commonly seen in email; it is not a representation of a spoken utterance. Similarly Tonfoni [27] suggests a number of icons to express author intentions such as “describe” or “define.” Although the audio clips of a multimedia chronicle represent human speech at a much lower and more direct level than a writing system, their arrangement on the screen could form the beginnings of a specialized semasiographic visual language.

3.2. Playback

As seen in Figure 3, every audio bar has associated play and stop buttons to control playback. Since the original recording order is preserved, at the end of one clip the next is automatically played. A received message, when opened, will automatically start playing the most recent sub-chronicle. Clicking on a document icon opens that

document in the user’s Web browser or other appropriate viewer.

In this way the same interface covers playback as well as composition of messages—a user can freely intersperse recording and playback of audio clips.

3.3. From messages to conversations

Now that the means for composing and viewing individual messages have been described, this section explains how replies and eventually conversations are created.

Email software usually allows the option to include the original message in a reply, creating a mosaic message. The reply can either be interspersed among the lines of the original message, or it can be prepended or appended to the original message as a single block. Since TouchVerse allows a freeform layout, the user can create any of these structures.

Elements of the original message (that may in turn contain previous messages) may be moved and rearranged within the composition window, but neither the audio nor the references can be edited. To distinguish audio clips recorded by different people, different colors are used.

Figure 5 shows an example of a third-generation reply to a multimedia chronicle, with multiple document references. It can be seen that a variety of references have been employed: both to newly introduced documents and to existing documents from previous messages, both to whole documents and to particular points within them.

A great deal of the structure of the message and the content of the individual audio clips can be inferred purely from this visual representation. For instance, a user can easily find and replay the part of the message that discusses the meaning of the different columns of the form *D*. Conversely, Figure 5 also shows the visual clutter that can result as multiple replies accrete on a chronicle. This clutter could be reduced by giving users control over the visible time intervals from a conversation’s history. For now the only option, as with long-running email threads, is to occasionally begin a fresh chronicle.

4. Communications system design

The Web is used to distribute multimedia chronicles. Prior to transmission, chronicles are uploaded to a Web server. Then just the URL is sent by regular email. For situations where the network is not accessible, chronicles are temporarily stored locally. In addition to the XML chronicle representation, the Web server stores copies of all referenced documents. These are uploaded as they are dragged and dropped into a composition window.

Access control mirrors the implicit model underlying email: a user has access to a chronicle, or a document referenced from a chronicle, if he or she is the author of the

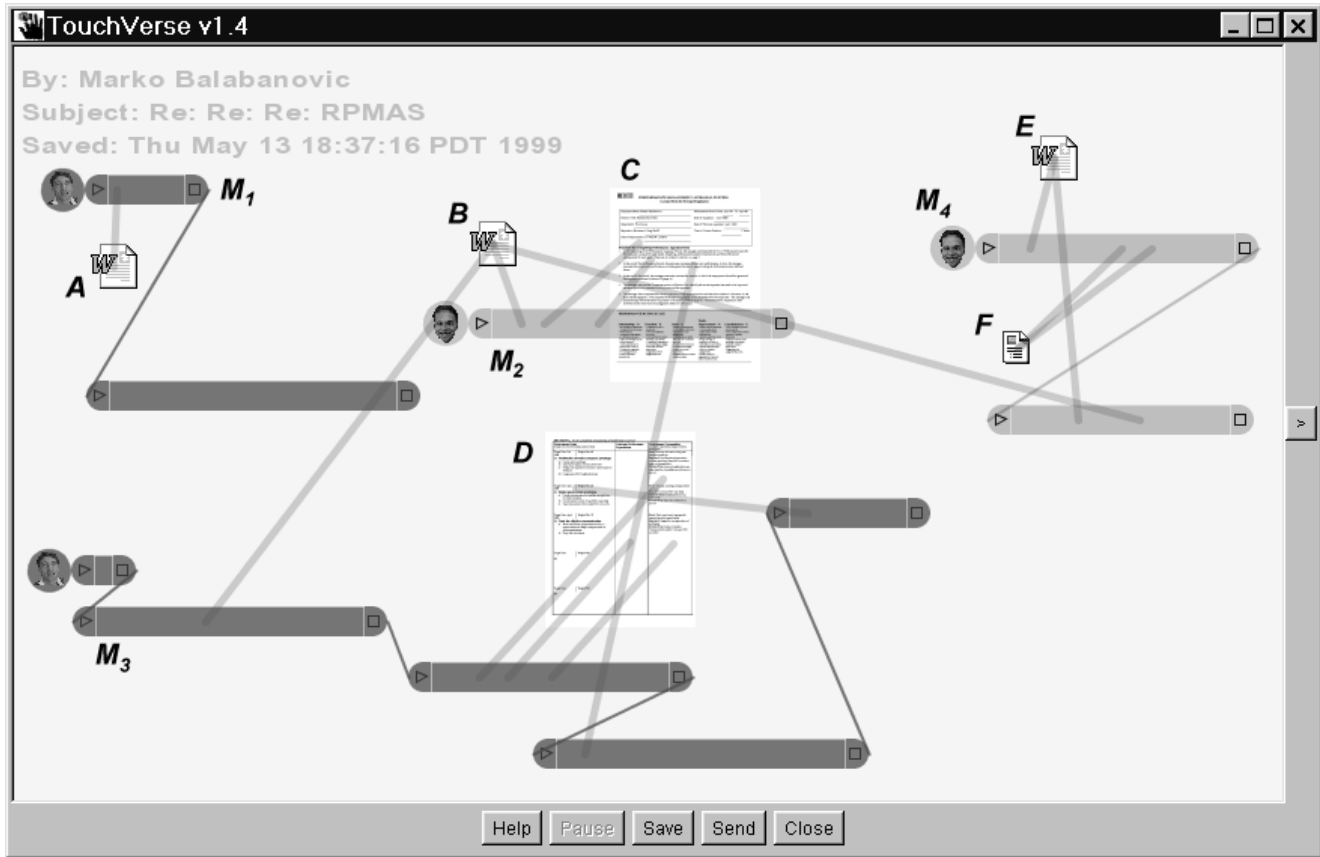


Figure 5. An actual TouchVerse conversation, about an upcoming performance review. The original message M_1 references Microsoft Word document A , a template “performance management appraisal form.” The reply M_2 references Word document B that is a filled out version of the form A , and then has some questions about specific points. These reference C (a printed version of the same form, retrieved from the IM³ scanned image database). The reply to the reply is M_3 . It further references D , the second page of the printed form, where a mistake is pointed out: some entries in rightmost column should have been in the center column. The final message M_4 references E , a corrected version of the Word document, as well as F , an unrelated Web page (by this point the conversation has drifted onto a different topic). The total length of the audio here is around 3.5 minutes.

chronicle, or has been directly sent the chronicle at any time. The current implementation does not expand group mailing lists (i.e., to be granted access to a stored chronicle that was sent to you, you need to have been addressed individually, not as a member of a mailing list).

4.1. Message management

Apart from receiving a multimedia chronicle URL by email, there are two ways users can access messages:

- An HTML interface allows access to all the multimedia chronicles that a user has permission to see. They can be sorted by type (sent, received or unsent), au-

thor, subject, recipient or date. Threads of messages and their replies can be followed. This interface provides essentially equivalent functionality to a Web-based email access system.

- Given a document in the IM³ database, the system can list all of the messages that reference it. Consider, for example, the discussion of Figure 5 about form A . At some later time, a user finding this form in the document database would also be able to retrieve the related multimedia chronicles. In this case, there would be four messages, culminating in the mosaic message containing them all shown in Figure 5.

This centralized storage complements users' existing email systems. The resulting collection of chronicles and referenced documents forms a valuable organizational memory: documents and the conversations about them are automatically captured; given a document, the surrounding chronicles provide context and history.

The server can additionally provide alternative views of individual chronicles. A chronicle can be viewed in two forms:

- Using the TouchVerse interface as described;
- As a Web page—the server can generate a simple HTML representation for an individual chronicle, including links to the referenced documents and audio. Users without the TouchVerse application can thus access multimedia chronicles, albeit with limited interactivity.

5. Initial usage

The TouchVerse system was used for a one month trial period by 13 employees at the Ricoh Silicon Valley California Research Center—a mix of research scientists, managerial and administrative staff. All are regular email users with their own PCs, voice mail accounts and access to a shared fax machine. This is a somewhat unusual test environment. All the users have offices at the same location, so there are fewer barriers to impromptu face-to-face interactions than would be the case for a more distributed test group. Nevertheless, flexible working hours, regular conference and business trips, as well as users occasionally working at home, mean that face-to-face communication cannot always be used when a rich interaction is desired. Since the initial trial, several users have continued to use TouchVerse, albeit less frequently.

5.1. Communications media survey

The motivation of this preliminary study has been primarily to gain insight into the kinds of communication situations for which multimedia chronicles might be appropriate. One month is not long enough for regular communications patterns to be altered significantly, particularly when an initial learning and experimentation period is taken into account. However, it is long enough for users to develop an understanding of the capabilities of TouchVerse, especially in relation to email, voice mail and fax, the other commonly used asynchronous media. A survey was conducted to elicit users' preferences among different communications media for each of 21 hypothetical communication situations, using a simple questionnaire.

Daft *et al.* [10] used this kind of questionnaire to determine preference between six different non-electronic media

in 60 communications situations rated for equivocality. In their study, richer media were preferred for more equivocal communications. El-Shinnawy and Markus [11] subsequently used 7 of the highest-equivocality situations, as well as 7 new situations believed to represent high uncertainty, in order to measure preference between email and voice mail. They found that email was preferred even for equivocal communications, mainly due to email's better affordances for persistence and mailbox management.

For our study these 14 situations have been re-used, along with 7 new situations involving discussions of specific documents. Table 1 lists three example situations from the survey. Of course, given the early stage of the testing and the small test group, statistically significant results are not expected. Rather, the questionnaire provides a vehicle for further qualitative discussion about the overall utility of the multimedia chronicle concept. In addition, regular, informal usability interviews (as espoused by Nielsen [24]) have been conducted with users to elicit more detailed feedback about the interface design; the next section presents the findings.

Two trends were expected from this informal testing:

1. That multimedia chronicles would be preferred to other media for communications situations that involved discussing specific documents or regions within them.
2. That multimedia chronicles would be more useful for communications situations high in equivocality than for those high in uncertainty.

The constant-sum method was used to determine preferences among communications media for each situation. Users were asked to indicate how many times each medium would be used over 10 occurrences of the situation. The choices were email, voice mail, fax and TouchVerse; synchronous media were excluded. Thus for each situation, 4 numbers summing to 10 were required.

The 7 most regular users completed the questionnaire. The results show that multimedia chronicles were the top-ranked communications medium for 83% of the document discussion situations (averaging across all of the respondents). In one third of these cases multimedia chronicles were ranked equally with email; in the remaining two thirds of the cases they were ranked first outright. Of the remaining situations, TouchVerse was top-ranked for 33% of the high equivocality situations and 19% of the high uncertainty situations. Follow-up questioning confirmed that the expected trends did indeed hold true among this small number of initial respondents.

Almost all of the respondents reported unease with any medium other than face-to-face meetings for some of the high equivocality situations (especially "To explain to a new

Table 1. Sample situations used in a questionnaire to determine preferences among communications media; there were 21 situations in total.

| | |
|---------------------|--|
| High equivocality | To work out a personality problem that has affected the working relationship between you and your boss. |
| High uncertainty | To receive lengthy information about a two-day management seminar |
| Document discussion | To communicate examples of a new corporate identity with presentation templates and business card designs. |

rather sensitive employee that she mishandled a personnel conflict in your work group”, survey item 25 from [11]), usually due to the desire for synchronous communication, but in some cases also to prevent any record of the conversation.

5.2. Usability findings

From informal discussion with the test users, as well as monitoring actual usage of the system, three main reasons for choosing to send a multimedia chronicle have emerged:

1. The message requires talking about and pointing to specific parts of a document. As one of the requirements driving the design, this reason came as no surprise.
2. The message requires referring to a large number of documents. With hindsight it is clear that current email systems have few affordances for messages with several attachments. Typically a composition window allows for a single text component and a separate list of attached documents, but there is no way to comment on the documents individually.
3. The message consists of a brief comment and a simple reference. This seems to stem from the fact that talking is quicker than typing, and thus more convenient for the sender of a message, although not necessarily for the recipient [16].

Some examples of messages that users felt were best expressed as multimedia chronicles included: a trip report referring to 22 separate papers and Web pages, a discussion about a ceramics class that included photographs of different pots, a series of exchanges about a patent application, including references to discovered related work, a discussion about choosing among different hardware configurations for an upcoming purchase.

Figure 5 shows some common strategies for placement of audio clips: they are close to or overlap the documents they reference; in addition, most clips begin near the endpoints of their chronological predecessors. As can be seen

in this example, some users extend a chronicle downward, as is done when composing text, whereas other users extend a chronicle to the right, in the direction that the audio clip representation grows.

One of the most frequently requested improvements was the ability to see representations of and point to sections of *any* document. Currently the system only has thumbnail images for documents that have been automatically captured by the IM³ system; Web pages and source code would be useful additions. Users also wanted to type or paste text into messages, to edit and especially to delete audio clips. Some ingenious workarounds appeared to address the latter problem. One user dragged all unwanted audio clips into a pile in one corner of the composition window, and then instructed the message recipient to ignore any clips from down there.

For over half of the test users this was the first use of their PC’s multimedia capabilities, and accordingly there was a period of experimentation with microphone and speaker placement, loudness of voice while recording, etc. Recording problems were exacerbated by the lack of sound level feedback.

As would be expected from an initial deployment of prototype software, much of the user feedback took the form of bug reports. The TouchVerse software was designed to automatically upgrade itself to the latest available version upon starting. This proved invaluable, allowing bug-fixes and improvements to be distributed quickly and without user effort.

Finally, some general differences between oral and written media were observed, as have been reported in other studies (e.g., [12]):

- Users whose native language is not English reported greater comfort with composing text rather than voice messages;
- In some situations users preferred text for a more “official” record of a conversation;
- Users worried about how a voice recording might be used in the future, seeing this as more of a “personal” medium than text.

Following the initial test period, usage of the TouchVerse system declined. This is partly due to the prototype nature of the software. An additional problem is that the current test group cuts across existing work and social groups. There are few clusters of people who communicate frequently where every member is also a TouchVerse test user. Longer-term tests will focus on larger groups that are more tightly knit in terms of work or personal relationships.

6. Related work

The notion of capturing pointing while talking was capitalized upon by the early *Put-That-There* system [5], and since that time a number of systems have investigated the use of pen + speech input. *Freestyle* [21] is closest in spirit to our work—a communications system where the messages contain synchronized voice and stylus-drawn annotations over document images. Unfortunately it was deployed before Web protocols made it easy to transmit document references, and in environments where multimedia hardware, high-resolution monitors and even PCs could not be taken for granted [15]. Another common application is recording meetings or talks. The *NoTime* system [20] records pen strokes synchronized with audio or video; various extensions to this kind of system have been proposed in [28, 29]. One of the most thorough applications of these technologies is the *Classroom 2000* project [1]. It posits that multimedia authoring is the central activity of both teaching and learning. Both audio and video are recorded during a class, and correlated with teacher-supplied materials and students' notes. The resulting multimedia structure is provided via the Web for students to use for review and study.

Screen recorders form a related class of applications (e.g., *Lotus ScreenCam*). They capture audio at the same time as recording activity on a computer screen, in effect digitally videotaping a computer session.

All of the above systems share two disadvantages:

- It is not easy to include references to existing documents;
- No visual representation is constructed for the audio stream.

A number of projects have developed audio interfaces that do not use visual displays. Navigation through hyperlinked audio segments was demonstrated by the *HyperPhone* system [23], and Arons created more sophisticated interfaces allowing, for instance, quick “skimming” [2, 3]. *HyperVoice* [25] allowed the construction of specialized bulletin boards such as event calendars, contributed to and accessed by telephone.

Annotations are an important ongoing research topic. The *Etherphone* project [30] developed an extensive infrastructure enabling creation and editing of voice annotations

for documents and email messages. Collaborative writing tools such as *Quilt* [14] allow voice and text annotations linked to portions of text documents, supporting both suggested revisions and explicit communication between collaborators. Indeed audio annotation is an integral part of many modern word processors. In a sense multimedia chronicles have the inverse structure—a narrative referring to many documents as opposed to a document with many annotations.

A final comparison should be drawn with hypermedia authoring tools, e.g., [19, 7]. These typically allow construction of structures that are much more complex than multimedia chronicles, and the authoring process correspondingly requires more skill, training and time. Today's multimedia documents are often intended for mass publication or distribution, and are therefore worthy of a large investment in authoring time. In contrast, multimedia chronicle messages may be read just once by one person. A simple HTML authoring package with integrated email tools (e.g., *Netscape Composer*) comes closer to the system described here, but still requires the user to think about naming of files and locations of servers, and provides no affordances for time-based media such as audio.

7. Summary

This paper has introduced a communications system comprising:

- The multimedia chronicle data structure, capable of encoding rich audio conversations with references to documents;
- The TouchVerse interface, designed to make authoring multimedia chronicles as easy as talking and pointing, and to allow recipients of multimedia chronicles to easily access their content;
- A server that stores and provides access to multimedia chronicles, allowing retrieval of conversations about documents as well as documents referenced in conversations.

Preliminary usage among 13 test users for a month has shown that multimedia chronicles fill a gap among presently available communications media. They are especially appropriate for communications situations where it is necessary to point to documents or parts of documents, or to refer to many documents.

This paper describes initial usage experiences. Future reports will document longer-term usage, gauge the utility of the database of collected chronicles and referenced documents, and investigate techniques for managing tangled structures of intersecting chronicles.

Acknowledgments

The design of the multimedia chronicle system benefited from many discussions with Greg Wolff, and from suggestions and feedback from Jed Harris and the group of the test users.

References

- [1] G. D. Abowd, C. G. Atkeson, A. Feinstein, C. Hmelo, R. Kooper, S. Long, N. Sawhney, and M. Tani. Teaching and learning as multimedia authoring: the Classroom 2000 project. In *Proceedings of the ACM Multimedia Conference*, pages 187–198, 1996.
- [2] B. Arons. Hyperspeech: Navigating in speech-only hypermedia. In *Proceedings of the ACM Hypertext Conference*, pages 133–146, December 1991.
- [3] B. Arons. SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction*, 4(1):3–38, March 1997.
- [4] S. Birkerts. *The Gutenberg Elegies: The Fate of Reading in an Electronic Age*. Fawcett Books, November 1995.
- [5] R. A. Bolt. Put-That-There: Voice and gesture at the graphics interface. *Computer Graphics*, 14(3):262–270, 1980.
- [6] T. Bray, J. Paoli, and C. M. Sperberg-McQueen. Extensible markup language (XML) 1.0. Recommendation, World Wide Web Consortium, February 1998.
- [7] A. Caloini, D. Taguchi, K. Yanoo, and E. Tanaka. Script-free scenario authoring in MediaDesc. In *Proceedings of the ACM Multimedia Conference*, pages 273–278, Bristol, UK, September 1998.
- [8] B. L. Chalfonte, R. S. Fish, and R. E. Kraut. Expressive richness: A comparison of speech and text as media for revision. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '91)*, pages 21–26, New Orleans, LA, April 1991.
- [9] R. L. Daft and R. H. Lengel. Organizational information requirements, media richness and structural design. *Management Science*, 32(5):554–571, May 1986.
- [10] R. L. Daft, R. H. Lengel, and L. K. Trevino. Message equivocality, media selection and manager performance: Implications for information systems. *MIS Quarterly*, pages 355–366, September 1987.
- [11] M. M. El-Shinnawy and M. L. Markus. Media richness theory and new electronic communication media: A study of voice mail and electronic mail. In *Proceedings of the 13th International Conference on Information Systems*, pages 91–105, 1992.
- [12] M. M. El-Shinnawy and M. L. Markus. The poverty of media richness theory: explaining people's choice of electronic mail vs. voice mail. *International Journal of Human-Computer Studies*, 46:443–467, 1997.
- [13] T. Erickson, D. N. Smith, W. A. Kellogg, M. Laff, J. T. Richards, and E. Bradner. Socially translucent systems: Social proxies, persistent conversation and the design of “Babble”. In *Conference on Human Factors in Computing Systems (CHI '99)*, pages 72–79, May 1999.
- [14] R. S. Fish, R. E. Kraut, M. D. P. Leland, and M. Cohen. Quilt: A collaborative tool for cooperative writing. In *Conference on Office Information Systems*, pages 30–37, Palo Alto, CA, March 1988.
- [15] E. Francik, S. F. E. Rudman, D. Cooper, and S. R. Levine. Putting innovation to work: Adoption strategies for multimedia communication systems. *Communications of the ACM*, 34(12):53–63, December 1991.
- [16] J. Grudin. Why CSCW applications fail: Problems in the design and evaluation of organizational interfaces. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 85–93, Portland, OR, September 1988.
- [17] C. Haythornwaite and B. Wellman. Work, friendship, and media use for information exchange in a networked organization. *Journal of the American Society of Information Science*, 49(12):1101–1114, 1998.
- [18] J. J. Hull and P. Hart. The infinite memory multifunction machine (IM³). In *Proceedings of the 3rd IAPR Workshop on Document Analysis Systems*, Nagano, Japan, November 1998.
- [19] M. Y. Kim. Creative multimedia for children: Isis story builder. In *Conference on Human Factors in Computing Systems (CHI '95)*, pages 37–38, Denver, CO, May 1995.
- [20] M. Lamming and W. Newman. Activity-based information technology in support of personal memory. Technical Report EPC-1991-103, Rank Xerox EuroPARC, 1991.
- [21] S. R. Levine and S. F. Ehrlich. The Freestyle system: A design perspective. In A. Klinger, editor, *Human-Machine Interactive Systems*, pages 3–21. Plenum Press, 1991.
- [22] M. L. Markus. Electronic mail as the medium of managerial choice. *Organization Science*, 5(4), 1994.
- [23] M. J. Muller and J. E. Daniel. Toward a definition of voice documents. In *Proceedings of the Conference on Office Information Systems*, pages 174–183, 1990.
- [24] J. Nielsen. Guerilla HCI: Using discount usability engineering to penetrate the intimidation barrier. In R. G. Bias and D. J. Mayhew, editors, *Cost-Justifying Usability*, pages 245–272. Academic Press, Boston, MA, 1994.
- [25] P. Resnick. Hypervoice: A phone-based CSCW platform. In *Proceedings of the Conference on Computer-Supported Cooperative Work*, pages 218–225, November 1992.
- [26] G. Sampson. *Writing systems: A linguistic introduction*. Stanford University Press, Stanford, CA, 1985.
- [27] G. Tonfoni. *Writing as a visual art*. Intellect Books, Oxford, UK, 1994.
- [28] S. Whittaker, P. Hyland, and M. Wiley. Filochat: Handwritten notes provide access to recorded conversations. In *Conference on Human Factors in Computing Systems (CHI '94)*, Boston, MA, 1994.
- [29] L. D. Wilcox, B. N. Schilit, and N. Sawhney. Dynamite: A dynamically organized ink and audio notebook. In *Conference on Human Factors in Computing Systems (CHI '97)*, pages 186–193, Atlanta, GA, March 1997.
- [30] P. T. Zellweger, D. B. Terry, and D. C. Swinehart. An overview of the Etherphone system. In *Proceedings of the 2nd IEEE Conference on Computer Workstations*, pages 160–168, 1988.