Jotmail: a voicemail interface that enables you to see what was said

Steve Whittaker, Richard Davis*, Julia Hirschberg, Urs Muller

ATT Labs-Research
180 Park Avenue
Florham Park, NJ 07932, USA
+1 (973) 360 8339
stevew/julia/urs@research.att.com

*Virtual Ink Corporation
56 Roland St. Suite 306
Boston, MA 02129, USA
+1 (617) 654 0126
richard.davis@virtual-ink.com

ABSTRACT

Voicemail is a pervasive, but under-researched tool for workplace communication. Despite potential advantages of voicemail over email, current phone-based voicemail UIs are highly problematic for users. We present a novel, Web-based, voicemail interface, Jotmail. The design was based on data from several studies of voicemail tasks and user strategies. The GUI has two main elements: (a) *personal annotations* that serve as a visual analogue to underlying speech; (b) automatically derived message header information. We evaluated Jotmail in an 8-week field trial, where people used it as their only means for accessing voicemail. Jotmail was successful in supporting most key voicemail tasks, although users' electronic annotation and archiving behaviors were different from our initial predictions. Our results argue for the utility of a combination of annotation based indexing and automatically derived information, as a general technique for accessing speech archives.

Keywords

Voicemail, annotation, speech access, note-taking, asynchronous communication, "speech as data", empirical evaluation.

INTRODUCTION

Voicemail is a pervasive but under-researched workplace communication technology, with an estimated 68 million users worldwide. Many organizations rely heavily on voicemail for conducting everyday work, and voicemail is often preferred to email [10]. The advantages of voicemail over email are: speech is expressive, easy to produce and critical in many workplace tasks [3,4]. Voicemail is also ubiquitous - any phone acts as an access device. It is also a common feature of most new cellular phones.

In the past, the phone was the only universal access device. As a result, voicemail interfaces were either touchtone or speech-based. However, the Web and PDAs will soon make *graphical* UI methods more widely available for accessing voicemail. Graphical access may have significant advantages: *visual indices* have been used successfully as a general technique to access other types of speech archives [1,2,5,6,7,9,11,12,15,17]. Visual representation of speech structure allows random access to an inherently serial medium. The aim of this paper is to explore how these new visual indexing techniques can be applied to voicemail access, in particular to address documented problems with current touchtone UIs [13]. We also wanted to evaluate our system with real users: much prior research on speech access has focused on new techniques and not on their evaluation.

The structure of the paper is the following. We present an extended analysis of a previous study of voicemail usage [13], identifying four key user problems: *message scanning, information extraction, status tracking* and *archiving.* A central user strategy for voicemail processing relies on *message indexing by note-taking*. We implement a novel Web-based voicemail GUI that supports annotation for indexing. The UI allows users to take temporally indexed notes associated with individual messages. These notes serve as a *visual analogue* to the underlying speech in the message, allowing straightforward access to message contents, message scanning and status tracking. We also provide people with automatically derived header information for each message. We evaluated Jotmail in an 8-week field trial, where people used it as their only means for accessing voicemail. Jotmail was highly successful in supporting most key voicemail tasks, although users' electronic annotation and archiving behaviors differed from our predictions. Our results argue for the utility of a combination of annotation based indexing and automatically derived header information, as a general method for accessing speech archives.

VOICEMAIL TASKS AND PROCESSING STRATEGIES

We collected qualitative and quantitative data to identify users' key tasks and strategies for processing voicemail, for a typical voicemail system, Audix™, including: (a) server logs from 782 active users; (b) surveys from 133 high volume users (receiving more than 10 messages/day); (c) interviews with 15 high volume users.

The server data showed intensive voicemail use: people accessed the system a mean of 7.1 times each working day, receiving 8.7 messages, and storing 4.8 mins. of messages overnight. Voicemail messages also contained significant amounts of information: about half those surveyed reported average message lengths of between 30-60 secs. and about half reported lengths of 1-2 mins. Our interviews also

indicate that voicemail messages contain complex information, not simple *"call me back"* requests: *"[a voicemail message] is really like a whole memo, or a huge email message worth of information."* Furthermore, voicemail often substitutes for a series of face-to-face meetings: *"entire transactions or entire tasks are accomplished by exchanging [voicemail] messages. That is, you will never talk to the person in real time."* Finally people stressed that a key value of voicemail is ubiquity: *"the most important feature for voicemail as opposed to e-mail is that it is easily accessible from any telephone. People tend to respond quicker to voicemail than they do e-mail, because you can access voicemail from anywhere."*

Users report four main tasks when processing voicemail: *scanning* the mailbox to identify important messages; *extracting information* from individual messages; *tracking the status* of current messages; and *managing* their archive of stored messages.

Scanning

Scanning is used for *prioritizing* incoming new messages, and for *locating* valuable saved messages. *Prioritization* is critical for users who have to identify urgent incoming messages, while accessing the mailbox under time constraints (e.g. during a meeting break). These users have to rapidly determine which new messages require their immediate attention. *Message location* occurs when users search for saved messages containing valuable information.

Users' current scanning strategy is to sample all messages in sequence to determine location and status. For prioritization, only 24% of people we surveyed use voicemail message headers to identify urgent messages, reporting they are too slow. Instead they listen to the first few seconds of each message, to the speaker's intonation, to determine whether a message requires immediate action. *"Typically if I am at a meeting, I am on a short break and I find I have 10 or 12 messages … I only sort of skim them, but listening to the first couple of seconds, as to who it is and what the issue is, to see whether it has to be dealt with immediately. If not, I will just save the message and go on, so I can pick up any priority ones."*

In *locating* stored messages, most users do not retain a detailed model of their archive and 76% of those surveyed report that "listening to each message in sequence" is their standard procedure for finding archived messages. However, the linear nature of mailbox search makes location onerous when more than a few messages are stored: *"if I've got 20 messages stored … and I want that last message, it's a real pain to get to that last message. And … most of the time I don't even know what message I want to get to"*.

Information extraction

When a relevant message is identified, users have to *extract* critical information from it. This is often a laborious process involving repeatedly listening to the same message for verbatim facts such as caller's name and phone number. Multiple listens are also necessary with vague or highly detailed messages. *"Often you get messages that are somewhat vague… you need to listen to them several times to understand exactly what the agenda is"*. Of those surveyed 46% report that they relisten to messages "about half the time".

To reduce repetitive processing, 72% of our survey users report "almost always" taking *written notes*. Users employ two different note-taking strategies. The first strategy is *full transcription*: here users attempt to produce a written transcript of the target message, so as to reduce the need for future access. *"Those notes are like a memory, there are sort of a paraphrase of the message … sort of a synopsis of what I felt the conversation was about"* The second strategy is to take notes as *indices*. According to our users, voicemail messages have a predictable structure, and the object of this strategy is to abstract the key points of the message (such as caller name, caller number, reason for calling, important dates/times and action items). In most cases, users keep the original voice message as a backup for these incomplete and sometimes sketchy notes. *"I will write a word or two … something that will jog my mind about the message … to get the details, I will go back and listen to the voicemail."* These notes are either kept on scraps of paper (75% of users) or in a dedicated note-pad (25% of users), and people refer to them when searching their voicemail archive to locate particular messages.

Status tracking

Workplace tasks are often delegated through voicemail, and a common user problem is *tracking message status.* Status tracking is a prevalent problem for users accessing voicemail under time pressure. They often defer processing a significant number of incoming messages. When accessing voicemail later, they are often unclear about which messages they have dealt with. *"I access my messages in the car, so I don't take any notes and I can't remember which messages I responded to. That makes people very cross"*.

There are two main techniques for status tracking. In the first, people use notes taken during *information extraction* as reminders. These notes, taken on scraps of paper, are left around the user's work area to remind them about what needs to be done. *"I sit here with those little stickies and I go, I got to call this person, this person, this person"*. One problem with this note-taking strategy is that people sometimes lose these notes, especially when voicemail was originally accessed in a remote location such as a meeting room. Losing notes is less of a problem when people use a dedicated logbook for recording message details.

With the second status tracking strategy, users take no notes but leave undischarged messages in their voicemail mailbox. Reminding takes place when users next scan their archive. *"I will call back to the voicemail, even if the light isn't on, knowing that I have got some messages that I need to respond to"*. In the course of scanning they are reminded

of outstanding undischarged messages. The weakness of this second strategy is that there is no visible reminding cue, so that if people do not access the voicemail archive they are unaware of the presence of unresolved items.

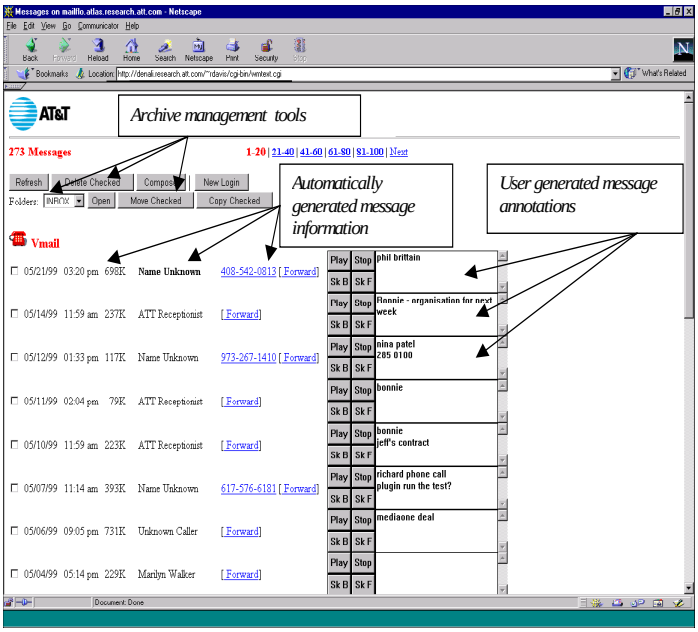

Figure 1 – Jotmail User Interface

Archive management

People also have to *manage their archives*. Given their access strategies, most users' archives consist of a backlog of undischarged messages as well as a store of saved valuable messages. They therefore engage in periodic "clean-ups": accessing each message in sequence to determine whether it should be preserved. *"I do try to schedule at least a half an hour a day where it's not meeting oriented, so I can clean up my messages"*. By removing superfluous messages, users also make it easier both to scan for existing valuable messages, and monitor reminder messages. Those who do not engage in "clean-ups" report being surprised by the extent to which they are accumulating irrelevant messages. *"I will go back and start listening to everything [in the archive] and by and large what I find out is that 90% of the time I delete most of the messages"*.

JOTMAIL USER INTERFACE

We devised a novel Web-based GUI, Jotmail, with the goal of supporting *scanning*, *information extraction*, *status tracking* and *archive management* tasks, in response to these findings, along with those obtained from controlled laboratory studies of voicemail access [14]. The design was based around the strategies we had observed being used for processing voicemail, paying particular attention to the critical role of note-taking. A key benefit of voicemail reported by our users is ubiquity. By developing a Web-based UI, we provided access to voicemail in any location where there is a computer with an Internet browser. The UI is shown in Figure 1. The center right of the screen shows text boxes for user generated message annotations, with play control buttons located to the left of them. On the left of the interface is more general header information about each message that has been derived automatically. The upper left of the screen shows archive management tools for creating and managing voicemail folder structures. The design was finalized after several iterations with trial users.

Information extraction using annotations

A key strategy for addressing *information extraction* was the use of personal notes. A central, novel, feature of the UI is therefore support for user annotations. Users can record personal notes, (e.g. "phil brittain", "Bonnie - organisation for next week") in the scrollable text box associated with a given message. One use of annotations is for message summaries. In all our studies users reported the need for repeated replayings of the message to extract critical information. We therefore wanted to provide ways to rapidly identify and replay only the most relevant parts of the message, without having to listen to the entire message. User notes are therefore also *time-indexed* [9,12,15]. The motivation for this came from the observation that handwritten notes serve as an index into the underlying structure of the original voicemail message: *"my notes trigger things - they are … meant to just give me place holders while I am browsing. Then I have to go back and listen to stuff"*.

Time-based indexing works as follows (see Fig. 2): users take notes as the message is played, and each note is co-indexed with the speech currently being played. If users later click on a given word in their notes, they automatically access the speech that was being played when the note was taken. In this way, notes provide reasonably precise access into the underlying speech, allowing users to focus on areas of specific relevance[1]. To further help information extraction, we provide general play controls for navigating within the message without recourse to notes. These allow people to *play* and *stop* a given message as well as *skip forward* and *skip back* 2 secs. within a given message (Sk B and Sk F buttons).

Scanning using message overview information

Our user studies also revealed the requirement for *scanning* to prioritize and locate important messages. In addition to user generated annotations, an important set of cues to aid scanning is message header information [11]. The UI therefore displays the following information about each message: *date*, *time*, *size* in Kbytes, *caller-ID number* (when it is available) and *caller name* (for internal calls

[1] There is a delay between the time at which users hear the relevant information and when they enter their related note. In later versions of the UI, we therefore introduced an indexing off-set, so that notes are indexed to material being played two seconds previously. This off-set was determined after iterations with several early users. In later versions we plan to make this interval user configurable. Figure 2 does not show the off-set for ease of exposition.

only, derived by looking up the name for the caller number in the corporate directory). In Fig. 1, the first message in the mailbox was from 408 542 0813, on 5/21/99 at 3.20pm. The message length was 698K, but because the message was from outside the local PBX the system was unable to infer the caller name. By depicting this general information we enable users to visually scan and randomly access messages. They no longer have to access messages in sequence to identify specific messages.

As can be seen from the example of the first message, users manually supplement automatically generated overview information with their own notes. For the first message, the user has added the caller name ("phil brittain"), because the system was unable to infer this, and the caller-ID number was unfamiliar. In this way, annotations were used to support *scanning* as well as *information extraction.*

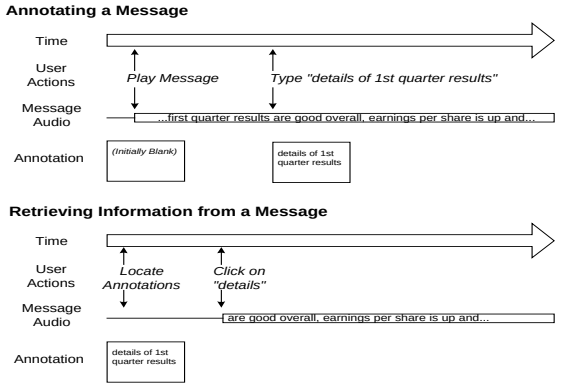

Figure. 2: Time-based indexing

Status tracking using annotations and overview information

Users also reported problems in trying to remember what outstanding actions were required for a given message. The user interface was designed to support status tracking in two ways – again by analogy with people's paper based strategies of leaving themselves visual reminders. Annotations could be used to *explicitly* record the actions necessary for each message. So for example "*Richard phone call plugin run the test"* states the action that was requested in the message, namely to run the relevant test. More *implicitly*, we hoped that the mere fact of having a visual representation of each message visible in the mailbox would serve to remind people of the necessary action whenever they access Jotmail. For example seeing a message from Marilyn Walker (last message in the inbox) might remind me of the action that message requires. A final cue to message status is that unaccessed messages are depicted in **bold** (the first message in Fig.1). Once accessed, their status changes.

Archive management

Users also reported problems in remembering the contents of their archive, and in preventing the build up of irrelevant messages. The Jotmail interface provided them with a set of tools for organizing, managing and deleting voicemail data. Labeled buttons allowed them to create new folders, as well as move, delete and copy information to those folders. More implicit support for archive management is provided by the visibility of messages, enabling the archive to be quickly scanned to identify important messages and filter out superfluous ones.

Implementation

Jotmail is built on top of Webmail, a research system that supports email and voicemail access. Webmail is implemented as a CGI script that connects to a standard mail server. When the script is run, it produces HTML pages with interfaces for viewing, browsing, and archiving messages. Voicemail messages are retrieved from the voicemail system and stored as email messages with special headers and data.

The requirement for broad access influenced our choice of platform. The annotation system in Jotmail was implemented as a web browser plug-in. Our HTML plugin will work on most browsers, but at the same time using a plugin restricts the complexity of possible UI implementations when compared with what could be implemented in other languages such as Java. Webmail was modified to store annotation files alongside voicemail messages, and to display small annotation plug-in windows next to each message. The plug-in application downloads the annotation file and the audio file for each message and allows the user to play the message and type text in the annotation text box. If the message is playing when the user is typing, then the current time index in the message is stored with each word. By holding "Control" and clicking on a word, the user can play the message from that time index. When the plug-in closes, modified annotations are sent back to Webmail so they will be displayed the next time the page is loaded. This gives Jotmail users persistent annotations for information extraction and status tracking.

EVALUATION DESIGN

A major goal of our evaluation was to investigate the experimental system being used by people for their everyday work. We designed the trial to collect the following data to investigate how effectively the system supported the tasks we had identified. We were also concerned with how well the system supported annotation behaviors, given the centrality of this strategy.

Preinstallation survey: Before installing the new system we administered a user questionnaire addressing use of the standard touchtone (TT) voicemail system. We asked people about the number of voicemail messages they currently received and sent, and how they processed these. We also gathered data about *scanning, information extraction, status tracking* and *archiving* tasks, as well as note-taking strategies. We surveyed people about the success of TT features (header information and message operations) in supporting these core tasks.

HYPOTHESIS	MEASURE	TOUCHTONE SYSTEM (MEAN)	JOTMAIL (MEAN)	STATISTICAL DIFFERENCE & HYPOTHESIS CONFIRMED?
O1	Overall ease of processing voicemail (5 = "very easy", 1 = "very hard")	2.6	4.6	$t_{(6)}$= 4.10, p<0.01 confirmed
O2	Ease processing voicemail compared with email (5 = "much easier", 1 = "much harder", 3 = "about the same")	1.3	2.7	$t_{(6)}$=4.80, p<0.005 confirmed
S1	Scanning (5 = " very easy", 1 = " very hard")	1.7	4.7	$t_{(6)}$=6.22, p<0.0001 confirmed
S2	Ease of locating a specific message (5 = " very easy", 1 = "very hard")	1.8	5.0	$t_{(6)}$=11.50, p<0.0001 confirmed
S3	Preferred method for locating a message	All users listened to the first few secs.	All users employed visual scanning of the mailbox	Cochran's $Q_{(1)}$= 7.00, p < 0.01, confirmed
I1	Information extraction (5 = "very easy", 1 = "very hard")	2.1	4.0	$t_{(6)}$=5.29, p<0.002 confirmed
I2	Replay frequency (1 = " very frequently", 5 = "never")	3.3	3.3	$t_{(6)}$=0, ns no effect
I3	Note-taking frequency (1 = " very frequently", 5 = "never")	2.0	3.0	$t_{(6)}$=3.57, p<0.02 disconfirmed
ST1	Ease tracking message status (5 = "easy", 1 = "hard")	3.0	3.9	$t_{(6)}$=2.34, p=0.05 confirmed
ST2	Frequency of losing notes (6 = "never", 1 = "frequently")	3.4	5.3	$t_{(6)}$= 5.46, p<0.002 confirmed
A1	Archival behavior (5 = "never archive", 1 = "usually archive")	3.4	2.3	$t_{(6)}$=1.80, ns. no effect

Table 1: Comparison of Jotmail and touchtone user interfaces

Jotmail logs: We logged usage data for 8 weeks. We collected data about: number and duration of Jotmail sessions, messages stored and accessed, operations on messages (stop, play, skip) as well as information about what notes people took and when they used these to replay messages. This data was used to identify the main types and functions of user annotations.

Post-installation survey: After 8 weeks, we took the system away and administered an extended version of the original questionnaire containing additional questions about the basic features of Jotmail, note-taking, and the use of notes for playback.

Interviews: We also carried out two semi-structured interviews with users, one while the system was installed and one after it was taken away. These probed the same issues as the surveys, but we tried to elicit fuller descriptions of the way that the system was being used as well as its main benefits and drawbacks. We also tried to find examples of novel or unexpected uses of the system. We supplemented these interviews with some observations of people using the system.

RESEARCH HYPOTHESES AND RESULTS

We installed the system and collected logs for 9 users for a total of 184 sessions over 935 hours. Our users were researchers and secretarial support staff at AT&T. Of these, 7 completed all questionnaires and interviews. Our logs show frequent system use. People used the system for an average of 20.4 sessions, of mean length 5.1 hours. Jotmail users tended to keep the application constantly running, unlike TT where short sessions were the rule. In each session they would typically access 4.1 messages, and the mean distribution of play operations per session was 2.1 plays, 1.8 skips and 0.3 annotation-based plays. Overall users replayed 36% of messages. On 30% of occasions they listened to the same message two or more times in sequence. Users would typically save 11.0 messages at the end of each session.

Our hypotheses and results were as follows (see Table 1):

Overall preferences

O1: We predicted that people should rate it easier overall to process their messages with Jotmail, given the support it provides for scanning, information extraction, status tracking and archiving.

O2: For the same reasons, people should rate Jotmail processing as closer to email than TT access.

Both these hypotheses were confirmed (see Table 1). User comments also bear out the overall superiority of Jotmail. *"The whole process of dealing with voicemail is that much easier. In the past [with TT] when I came into the office I used to put off dealing with voicemail and always look at my email first, but now I often look at my voicemail first."* Contrast this with comments about TT voicemail access: *"I hate managing voicemail with so little information. What I dislike most is that you cannot "see" messages, whether there are new messages, if so how many, or saved messages.* Users were all highly displeased when we took the system away after 8 weeks.

One repeated observation was about the greater efficiency of Jotmail. People complained that TT required them to execute too many operations to access their voicemail. *"There's too many key presses to get into TT - with Jotmail there's no keys - I like its easy accessibility compared with the time consuming process of voicemail retrieval through the phone."* Another user commented about TT: *"it took me 27 seconds and 16 key presses to access a 5 second message"*.

Scanning

S1: Scanning should be rated as easier with Jotmail than TT access, given the visual representation of Jotmail

messages, automatically generated information and annotations.

S2: Users should find it easier to locate specific messages with Jotmail than TT access, because of the increased ability to scan.

S3: Users should switch from scanning by sequentially sampling the first few seconds of each message, to using Jotmail to look through headers and notes.

All 3 hypotheses were confirmed. The user logs also showed evidence of using the visual representation for random access: users accessed 36% of messages in an order different from that in which they were received and stored.

By providing the ability to scan messages, Jotmail allows users to prioritize their access. One user said of TT: *"With the old system I never knew which messages were there [in the mailbox] in what order so I couldn't selectively pick the ones I wanted to deal with. I used to put off listening to voicemail until I had the time to go through it all"*. Scanning also allowed rapid access to old messages: *"I can see what's in my mailbox at a glance… the fact that messages are visible means that I can find old ones easily"*.

Information extraction

I1: Users should rate information extraction as easier with Jotmail than TT access, because of the presence of annotations, and the ability to do time-based playback.

I2: Users should replay messages less often with Jotmail - both because they have more information automatically available about each message, and their own notes provide reminders about message contents.

I3: People should take more notes with Jotmail because of the utility of notes as indices.

Table 1 shows that the I1 was confirmed. However there was no difference in the number of replays with the two systems (I2). It may be that greater control over access provided by Jotmail means that users are happy to replay messages. Contrary to our predictions, people took *fewer* notes with Jotmail. The decreased amount of note-taking with Jotmail (I3) may occur because it *automatically* logs relevant information, obviating the need for some manual notes: *"You can take fewer notes with Jotmail because the name of the caller time and date for inside callers is shown on the screen."*

Notes were still useful in Jotmail, however. Their utility was demonstrated by the fact that every user reports taking notes on the system. There was also no reported difference in the types of notes taken with the two systems – a few key words such as name, phone number and action. Nor is there a difference in the reason for taking notes: the majority (86%) of users in both cases said that they take notes as a reminder either about message contents, or about the action a message requires. The fact that Jotmail notes are similar to their paper analogues offers good support for the naturalness of this aspect of the user interface.

Status tracking

ST1: Users should find it easier to track messages with Jotmail, because of the ability to scan outstanding messages at a glance.

ST2: People should be less likely to lose notes with Jotmail than TT voicemail because Jotmail notes are stored at a single on-line location.

Both hypotheses were confirmed, and multiple user comments indicated that status tracking was a critical perceived benefit of Jotmail, especially in the face of constant interruptions: *"I am always being side-tracked and interrupted in my job. Jotmail is like a tickler file. It provides a constant reminder of the things that I have to do"*.

Some users exploited the note-taking features of Jotmail to explicitly add multiple successive comments to the original message "*Fax sent*", to track progress after each action taken in responding to the call. They also commented on the benefits of having their notes on-line in close association with the original message. *"I used to save post-its as a record of what I was doing, but this way (taking on-line notes), I don't have loose pieces of paper that can get lost"*.

Archiving

A1: Given the increased ease in managing the archive, we expected that users would archive more messages with Jotmail versus TT.

This hypothesis was not confirmed. On the one hand, it was clear that Jotmail made it easier to store and access valued messages. On the other hand, it seemed that the visibility of the archive meant that users were better able to clean up and hence prevent the inadvertent build up of superfluous forgotten messages: *"When I access voicemail over the phone I don't usually access old messages whereas Jotmail provides reminders that I have old messages"*.

We also examined voicemail filing techniques. Somewhat to our surprise, although some users saved many messages, no-one categorized messages. They kept all their messages in the inbox, rather than creating task-specific folders. One user explained this as follows: *"I use the system to track things I have to do. If I haven't done them, I want to keep them in the inbox to remind me that they need attention. If they're done then I delete them. It's just making extra work to file them and have to remember where they are."* The emphasis therefore seems to be on status tracking rather than the construction of a complex archive. Better information extraction in Jotmail may also reduce archiving. One user pointed out the difficulty of information extraction with TT. She didn't keep as many

messages with Jotmail because information extraction was more straightforward, and she ended up with paper-based summaries of messages.

Annotations analysis

We also investigated a number of questions concerning annotations, given the centrality of users note-taking strategies in our initial studies:

Types of Jotmail annotations: These tended to be relatively brief (mean of 6.3 words), falling into 6 main categories: caller name, message topic, caller number, time, date and location, with frequencies per message being respectively: caller name (0.75), message topic (0.53), caller number (0.27), time (0.09), date (0.03) and location (0.03). The average message has two of these annotation types and the most frequent combinations of annotation types are: name and topic, name and caller, number and topic. Finally adding further annotations to a previously annotated message occurred relatively frequently, with 44% of annotations being additions to a previously annotated message. According to users, many of these re-annotations were being used to track the status of previously annotated messages.

Functions of Jotmail annotations: Overall, 29% of messages were annotated, with each user annotating 21.8 messages. Annotations were usually associated with messages that were important to the user; annotated messages were played more often than unannotated ones (respective means: 2.79, 0.92, $t_{(458)}=5.08$, $p<0.0001$). Annotations were not widely used to control playback: analysis of play operations indicated that time-indexed playback accounted for only 7% of play operations compared with 51% "play from start of message" operations and 43% skip based plays. 67% of users exploited the time-based indexing feature, but the remainder never did. Non-users argued that they received mainly short messages, reducing the need for controlled access to message contents. With short messages there is little cost to replaying an entire message to extract a single piece of information. This is supported by the fact that people tended to annotate longer rather than shorter messages (respective means: 422.7 and 356.1 Kbytes, $t_{(458)}=2.25$, $p<0.025$). However, it turned out that messages accessed by time-based playback were no longer than messages accessed using "play from start" only ($t_{(200)}=0.28$,ns)

Reasons for annotating specific messages People were more likely to annotate messages from unfamiliar callers (defined as those from outside their immediate workplace), $\chi_{(1)}=6.04$, $p<0.025$. They also made annotations more frequently with messages that had less automatically generated information $\chi_{(1)}=5.61$, $p<0.025$.

Unanticipated uses of Jotmail

Jotmail also led to more call screening. One user pointed out that the ease of accessing messages with Jotmail meant that he fundamentally altered his handling of incoming calls. With Jotmail he was more likely to screen calls by letting them go through to voicemail. *"With Jotmail I let live calls go through to voicemail because I knew I could easily get them later. With [TT] I don't do that because its so time-consuming to go and get them back"*. Another unanticipated use of Jotmail was for playback to a live audience. Two users reported replaying Jotmail messages to others (either face-to-face or over the phone). Again this was facilitated by the greater ease of message access: *"The only confirmation I had of S's promotion was a voicemail message from R., so I replayed that to him. I can't imagine being able to find that message using [TT]"*.

CONCLUSIONS

We built a novel Web-based UI, to voicemail centered on the notion of note-taking, that also provided automatically generated message information and archiving tools. The design was based on requirements data from interviews, experiments, surveys and user logs identifying key user tasks and strategies for voicemail processing. Data from an 8-week field trial showed that Jotmail was much preferred to a TT UI. As predicted, Jotmail improved scanning, information extraction and status tracking tasks.

User archiving and note-taking behaviors were not as expected however. Archive size did not increase with Jotmail, although this may follow from the superior scanning capabilities of Jotmail, preventing the unintentional accumulation of irrelevant messages that often occurred with TT. Removal of superfluous messages may therefore have counterbalanced increases in intentionally archived messages. Archiving behavior may also be influenced by users' prior experience with a previous TT system, which deleted messages after 14 days (a common feature of many such systems). This may have led users to view voicemail data as inherently ephemeral, despite user assertions that some messages had long-term value. A longer field trial might produce more instances of archiving activity once users habituate to the idea of message permanence. Users also failed to exploit the archiving tools provided, preferring to leave all messages in the inbox. Their comments suggested that systematic filing is onerous. Filing may also compromise the ability to track message status [16]: once filed, messages are no longer visible in the inbox, leading them to be forgotten. Given that voicemail volumes are lower than email, it may be possible to keep all current important messages visible in the inbox. Lower message volume in voicemail may reduce the pressure to file that has been reported in email studies [8,16].

Annotation behavior was also not completely as predicted. Users were very positive about their ability to annotate messages and use time-indexed playback. Annotations were used for reminding, status updates and their on-line location meant that they were not mislaid like paper notes.

Screen-based notes were similar to paper ones. As predicted, they were also associated with longer messages that were accessed more often. Despite this, users reported taking fewer notes with Jotmail. Reduced note-taking may have occurred because we automatically generated message header information, and indeed fewer notes were taken for messages with more such information. Time-based indexing was also used infrequently. There are several possible explanations: (a) messages were short enough to replay without undue cost, reducing the need for precise control during information extraction; (b) users found it hard to anticipate what notes would be useful for future retrieval; (c) sparse notes may be sufficient to *remind* users of the contents of the message, without the need for reaccessing the underlying speech. Other work is consistent with the reminding explanation, arguing there are important trade-offs between the *efficiency* of relying solely on hand-written notes as (imperfect) *reminders* versus the *accuracy* of accessing the verbatim speech record itself [9,15].

There are also important system extensions we are currently investigating. These include using automatic speech recognition to produce *transcripts* of voicemail messages. Although the transcripts are errorful, they nevertheless provide a browsable text for each message, allowing users to *read* rather than listen to voicemail. Like user annotations, they also serve as a visual analogue to each voicemail message. We are also exploring techniques for automatically extracting significant information such as names, dates, times and telephone numbers from these transcripts. These new automatic techniques should provide further support for *information extraction* and *scanning tasks.*

Finally, there are both practical and theoretical implications to our results. First our tool successfully addresses a significant problem for many users - namely *efficient* voicemail retrieval at any location where there is Web access. It seems to address many of the problems that users currently experience with TT voicemail. Our data also contribute to a growing body of research on general methods for speech access. We present data showing that, consistent with the claims of prior work, providing a visual analogue as an index into underlying speech structure is important for supporting browsing and retrieval [1,2,5,6,7,9,11,12,15,17]. As with other approaches [6,7], our results suggest that for personal data such as voicemail, a combination of automatically generated data and personal annotations provides a general technique for accessing complex information in speech.

REFERENCES

1. Abowd, G., Atkeson, C., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N., Tani, M. Teaching and learning as multimedia authoring in the classroom 2000 project, *Multimedia96*, 187.
2. Arons, B. Interactively skimming speech. Unpublished PhD thesis, MIT Media Lab, 1994.
3. Chalfonte, B., Fish, R., and Kraut, R. Expressive richness. In *CHI91*, 21-26, 1991.
4. Chapanis, A., Ochsman, R., Parrish, R. and Weeks, G. Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors*, 14, 487-509, 1972.
5. Degen, L., Mander, R., and Salomon, G. Working with audio. In *CHI92*, 413-418, 1992.
6. Hindus, D., Schmandt, C., and Horner, C. Capturing, structuring and representing ubiquitous audio. *ACM Transactions on Information Systems*, 11, 1993.
7. Kazman, R., Al-Halimi, R., Hunt, W., and Mantei, M. Four paradigms for indexing video conferences. In *IEEE Multimedia*, 3(1), 63-73, 1996.
8. Mackay, W. More than just a communication system: diversity in the use of electronic mail. In *CSCW86*, 344-353, 1986.
9. Moran, T., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., van Melle, W., and Zellweger, P. "I'll get that off the audio": salvaging in a multimedia meeting. In *CHI97*, 202-209, 1997.
10. Rice R. and Shook, D. Voice messaging coordination and communication. In C. Egido, J. Galegher and R. Kraut, eds., *Intellectual Teamwork*, Lawrence Erlbaum, NJ, 1990.
11. Schmandt, C. and Arons, B. Phoneslave: a graphical telecommunications system, *SID International Symposium*, 25, 146-149, 1984.
12. Stifelman, L. Augmenting real-world objects: a paper-based audio notebook. In *CHI96*, 199-200, 1996.
13. Whittaker, S., Hirschberg, J. and Nakatani, C. All talk and all action. In *CHI98*, 249-250,1998.
14. Whittaker, S., Hirschberg, J. and Nakatani, C. Play it again: a study of the factors underlying speech browsing behaviour. In *CHI98*, 247-248,1998.
15. Whittaker, S., Hyland, P, and Wiley, M. Filochat: handwritten notes provide access to recorded conversations. In *CHI94*, 271-277, 1994.
16. Whittaker, S. and Sidner, C. Email overload: exploring personal information management of email. In *CHI96*, 276-283, 1996.
17. Wilcox, L., Schilit, W., and Sawhney. Dynomite: a dynamically organized ink and audio notebook. In *CHI97*, 186-193, 1997.