The Audio Notebook
Paper and Pen Interaction with Structured Speech

Lisa Stifelman[*]   **Barry Arons**[*]   **Chris Schmandt**

MIT Media Laboratory
20 Ames Street, Cambridge MA, 02139 USA
{lisa | barons | geek}@media.mit.edu

ABSTRACT

This paper addresses the problem that a listener experiences when attempting to capture information presented during a lecture, meeting, or interview. Listeners must divide their attention between the talker and their notetaking activity. We propose a new device—the Audio Notebook—for taking notes and interacting with a speech recording. The Audio Notebook is a combination of a digital audio recorder and paper notebook, all in one device. Audio recordings are structured using two techniques: *user structuring* based on notetaking activity, and *acoustic structuring* based on a talker's changes in pitch, pausing, and energy. A field study showed that the interaction techniques enabled a range of usage styles, from detailed review to high speed skimming. The study motivated the addition of phrase detection and topic suggestions to improve access to the audio recordings. Through these audio interaction techniques, the Audio Notebook defines a new approach for navigation in the audio domain.

Keywords

Audio, speech, pen interaction, paper, speech as data, user structuring, acoustic structuring, speech interfaces.

INTRODUCTION

In our everyday lives we are presented with many situations when it is desirable to capture a detailed account of a presentation or conversation. For example:

> A student attempts to write down what the professor is saying as well as information written on the blackboard. Students must divide their attention between listening to the professor, trying to comprehend the material, and taking notes for later review.

> When a reporter performs an interview, he/she is trying to take notes, listen to the response, and think of the next question. In addition, notetaking can be distracting to some interviewees. Thus, reporters' notes often have gaps that need to be filled to enable direct quotes for the story.

In these scenarios, the listener could have made a tape recording. A tape can capture exactly what and how things are said, and contains information that cannot easily be described in a transcript or handwritten notes. The emotion of speech, the quality of voice, and subtleties of accents and pauses are lost in a transcript. It is time consuming and often frustrating to find information on a tape. A user has to shuttle between fast forward and rewind to find the portions of interest on the tape, and it is difficult to skim through the recording or correlate it with handwritten notes.

This paper addresses the problem of trying to capture and later review information presented during a lecture, meeting, interview, or conversation. We propose a new device—the Audio Notebook—for taking notes and interacting with a speech recording. The Audio Notebook is a combination of a digital audio recorder and paper notebook, all in one device.

The Audio Notebook combines the familiarity of paper and pen with the advantages of an audio recording. The goal is to retain the original audio while allowing a listener to quickly and easily access portions of interest. The Audio Notebook synchronizes the user's handwritten notes with a digital audio recording. The user's natural activity—writing and page turns—implicitly indexes the audio for later retrieval. This is referred to as *user-structured* audio. *Time* is mapped to *space*—the spatial layout of writing in a physical notebook enables rapid navigation in the audio domain. Familiar objects like paper and pen are used for interacting with the audio rather than artifacts left over from analog devices, such as fast forward and rewind controls. Users can *skim* their handwritten notes and the audio recording simultaneously.

Why Paper?

Previous systems have correlated an audio recording with notes written on an LCD display [4][13]. This work takes a different approach. Rather than writing on a display, the user writes in an ordinary paper notebook. The goal of the Audio Notebook is to *augment* rather than replace familiar objects like paper and pen [12].

The DigitalDesk takes the opposite approach of conventional computer desktop interfaces [6][11]. Rather than trying to replicate a physical desktop on a computer screen, the DigitalDesk integrates the computer into an actual desktop. A computer display is projected onto paper documents that users interact with directly.

Paper and pen provide a portable, tangible, and flexible way of capturing information [6], and are still widely used. Paper documents have many advantages over digital ones—a sheet of paper can be quickly torn from a notebook, stuffed in one's pocket for easy access, or handed to a friend. Ideas can be quickly scribbled down on paper. While it is possi-

---

* Currently at Tellme Networks, Inc.

ble to write on the screen of a pen-based computer, many people find writing on paper faster, more accurate, and more familiar.

Prior Work in Indexing Audio

The AIR project (Activity-based Information Retrieval) proposed employing user activity (e.g., notetaking, writing on whiteboards, user location) to index multimedia data [4]. Filochat also linked writing on an LCD tablet to audio recordings of business meetings [13]. In contrast, the Audio Notebook links audio recording to notes taken on paper and provides several techniques to access the audio.

Marquee used digital pen strokes for real-time video logging [9]. However, the logging activity is very explicit. For example, users must draw lines across a notetaking area to indicate a "time zone." All notes in a time zone are indexed to a single time point in the video. The Audio Notebook is fine grained, implicitly linking every pen stroke to a different part of the audio recording.

Dynomite indexes audio with notetaking activity on a pen-based computer [10]. Users can manually assign keywords to pages of notes. Although audio is recorded continuously, "only portions of audio highlighted by the user are permanently stored." The Audio Notebook captures the entire audio recording.

Coral is a "confederation" of tools for capturing, indexing, and salvaging collaborative activities [5]. Here the focus is on group rather than personal use. The audio and video are indexed by all notes written on, or beamed to, the Live-Board, and by page changes on the LiveBoard.

Classroom 2000 [1] captures audio, video, and slides, and links it to notes taken on tablet computers and electronic whiteboards. Access to the captured material is through an HTML-based web browser. NotePals [3] allows sharing of handwritten notes taken on LCD-based PDAs and paper-based digital tablets. Notes are viewed using a web browser; however, there is no audio component.

USER-STRUCTURED AUDIO: THE AUDIO NOTEBOOK

Audio Notebook Version 1

An early version of the Audio Notebook [7] demonstrated the basic concept of linking notes on paper with an audio recording. This early prototype showed the concept of automatic page number detection, and selecting on the page to begin audio playback (Fig. 1).

As the user writes in an ordinary paper notebook, the audio of a lecture or meeting is recorded digitally. Button controls for starting and stopping recording are activated by dipping the pen inside them. After recording, audio can be accessed by space or by time. Playback can be started by pointing to a location in the notebook with the pen. Playback begins at the point in the audio recording that corresponds to when the note was originally written (for more information see the section on Audio Snap-to-Grid using Phrase Detection). Dragging the pen along an audio scrollbar navigates a timeline of the audio associated with each page.

The paper notebook is placed on top of a digitizing tablet; a U-shaped cover creates a slot for the notebook over the tab-

let's active area. The user takes notes using a cordless digitizing pen with an ink cartridge. The prototype can sense through a notebook of about 60 pages. Playback is triggered by the location and pressure of the pen. The current page number and state of interaction (e.g., playback time) are shown on a small LCD display. An LED is used to indicate when the device is recording.
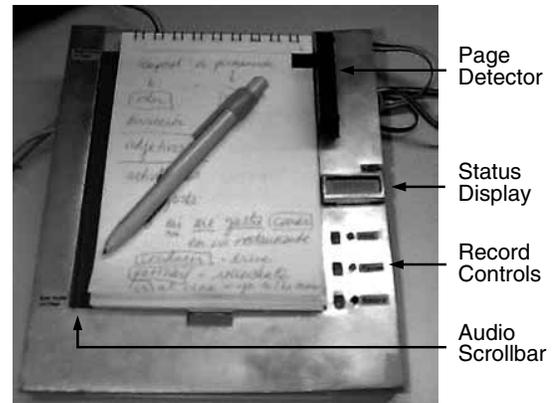

Fig. 1: Audio Notebook version 1. A user can randomly access parts of the audio recording by pointing to a location in his/her notes.

The scrollbar in this version didn't show the user's place in the recording, and it was difficult to restart playback where it left off. A page number detector was in the user's way, and could be shadowed by the pen or the user's hand.

Audio Notebook Version 2

In designing a second version of the Audio Notebook (Figures 2 and 3), one goal was to build a more robust prototype for field-testing. The prototype needed to run reliably and hold up under the stress placed on it by multiple users over several months.

The overall architecture of the second Audio Notebook is similar to the first, with most design changes focused on the front-end of the prototype (Fig. 3). The front-end design evolved based on experience with early users. One important change to the back-end design, however, is the addition of a removable drive for audio storage. A one-gigabyte Iomega Jaz cartridge is used for each notepad, storing up to 12 hours of audio.

Page Codes

In version 1 of the Audio Notebook, only the page numbers were coded. In version 2, both the notepad and page numbers are coded on the pages. In this new design, there are 4 bits allotted for coding the notepad, 6 bits for coding pages, and 1 bit for error checking (Fig. 4), allowing the device to uniquely identify 16 notepads of up to 64 pages.

Audio Scrollbar with *Audio Cursor*

In version 1 of the Audio Notebook, the audio scrollbar was only an input device. Once users removed their pen from the scrollbar, there was no indication of where they left off. In version 2, an *audio cursor* lights up to show the user's current position in the audio scrollbar. The scrollbar is made out of a series of 80 tri-color LEDs. The audio cursor is indicated by a green LED that moves along the scrollbar

as the audio plays. The scrollbar represents the audio associated with each page of notes. This new audio scrollbar acts as both a control and a display, serving several important functions. First, when a user selects somewhere on a page to begin playback, the audio cursor lights up showing the corresponding location in the timeline. The user can then fine-tune the starting point of playback by moving the cursor forward or backward in the timeline (Fig. 5). For example, a user might backup in the audio to get more context. This also gives the user a space-to-time correspondence—users can select on one note and see when it was written in relation to another. Second, the audio scrollbar provides the user with fine-grained navigational control. Lastly, the scrollbar can be used to display other information about the audio. The scrollbar LEDs can each be green, red, or orange.
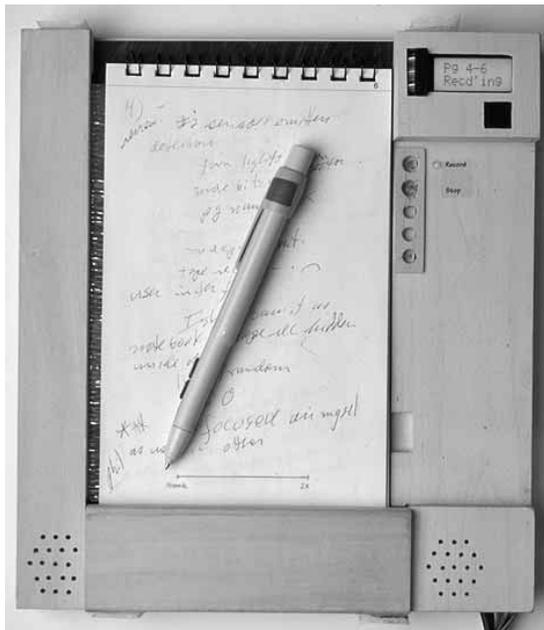

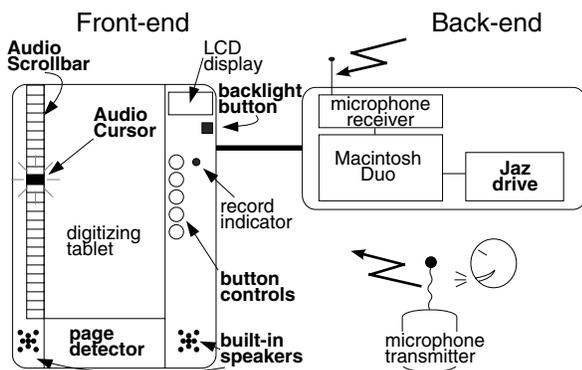Fig. 2: The Audio Notebook version 2.

Fig. 3: Components of the version 2 prototype. Design changes from version 1 are shown in bold.

LONGITUDINAL FIELD STUDY

Several students and reporters were observed using the Audio Notebook over a five-month period. Users were observed during both the capture and review of their notes and audio recordings. The students and reporters used the Audio Notebook for real tasks (e.g., to write a story for publication, to study for an exam), not artificial tasks performed in a laboratory. Rather than observing users during a single session, this study aimed to observe a small number of users in depth over time. The goal of the study was to observe how the subjects made use of the audio recordings given the audio interaction techniques provided by the Audio Notebook, and to learn what other interaction techniques are needed to navigate in the audio recordings.
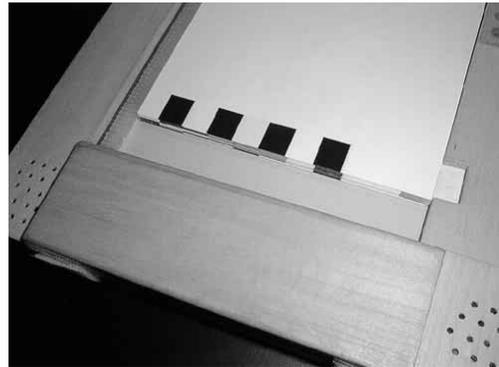

Fig. 4: Both notepad and page numbers are coded on each page. Sensors underneath a hand rest read the code.

Fig. 5: A user fine-tunes the starting point of playback using the audio cursor and scrollbar.

The Audio Notebook was used by four students and two reporters. Two students used the Audio Notebook regularly in their classes throughout one semester. The students were instructed to try the Audio Notebook once or twice and then to decide whether or not to continue using it. Both students requested to continue using the device throughout the semester. Two additional students in one of the classes used the Audio Notebook to review classes that they were unable to attend. Lastly, two reporters each used the device to interview a subject and write a story for publication.

Student 1: Rapid Skimming

S1 skimmed quickly through her notes during each session, mainly reviewing material that was unclear during the lecture or not clearly recalled afterwards. This student's total listening time averaged approximately one third of the original recording. S1 skipped around in the audio using spatial navigation and the audio scrollbar, listening for potential sections of the lecture to review. She often used the

audio scrollbar as a skimming control by selecting on every few elements of the display.

Review Sessions
S1 used the Audio Notebook to review classes nine times during the semester. She used two notepads containing a total of 24 hours of audio.

One or two classes were reviewed in each session, typically within one week of the class. During the review sessions, the student reviewed pages of notes and audio associated with 13 of the 16 classes recorded.

The audio scrollbar was used 1.7 times as often as selection on the page (134 vs. 79). S1 used the audio scrollbar to fine-tune the starting point of an audio selection; she selected in her notes and then adjusted the starting point forward or backward using the scrollbar. This student and other subjects in the study discovered this use of the scrollbar through exploration of the interface, without explicit instruction.

Use of the Audio Recordings
S1 used the Audio Notebook to review portions of the lectures that were unclear or missing from her notes. During each review session, she skimmed through the notes and audio by selecting on different areas in her notes and using the audio scrollbar. Once she found a place of interest in the lecture, she often listened to a large amount of the audio without jumping around. She reviewed information that she did not remember clearly, and skipped everything else.

Many times S1 drew a diagram but did not recall the meaning of it, or did not understand it originally. Fig. 6 shows a page from her notes containing a diagram. During the review, she selected on the three vertical lines drawn on the bottom right corner of the page to hear the corresponding audio from the lecture. She said "I totally forgot what he was saying in class and if I looked at my notes alone, I wouldn't have known what that meant."
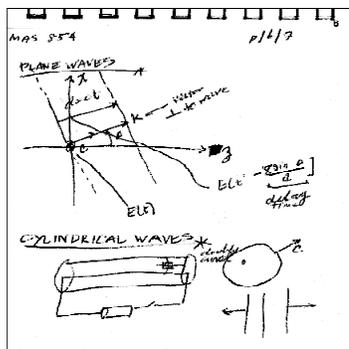

Fig. 6: S1 used the Audio Notebook to determine the meaning of the three vertical lines at the page bottom.

In other instances, a particular phrase written in the notes was unfamiliar afterward. By selecting on the words or graphic and hearing the corresponding audio, she was able to recall their meaning. In another example, the student marked her notes with a "?" explicitly indicating missing information for later review. By touching on the question marks with the pen, the user is able to review this informa-

tion. These examples demonstrate how the Audio Notebook was used to retrieve information that was unclear in the notes, or explicitly marked for review. Note that there were other cases in which missed information was serendipitously discovered upon review of the audio.

Skimming Using the Audio Scrollbar and Speed Control
S1 often used the audio scrollbar to skim through the audio associated with a page of notes. For example, she often started playback by selecting somewhere in the notes, and then jumping ahead by equal amounts in the audio scrollbar. However, she wanted to be able to listen quickly without having to skip over portions.

Starting with the seventh review session, a speed control was added, allowing the recording to be heard at up to twice the normal rate without changing the pitch. The first time S1 used the speed control, she immediately started listening to the speech at twice the normal speed without slowing down at any time. She was able to comprehend the speech at this fast rate even though she had no previous experience listening to time-compressed speech. This is most likely because the professor's speech was familiar to her and she heard the material before. She said "I really like it because you don't lose anything and it's easier to listen to this way… when the speech is slow, your mind tends to wander… this way you concentrate on it more." In subsequent review sessions, S1 used the speed control and the audio scrollbar in combination for even faster audio skimming [2].

Audio Notebook versus Tape Recorder
S1 said that backing up and repeating a portion of a recording, and skipping around was easier using the Audio Notebook than with a tape recorder. She also liked the fact that the Audio Notebook allowed her to look at her notes and listen to the lecture at the same time. This appears to be an important advantage of controlling audio playback directly from the notebook—since users can maintain visual focus on their notes during playback, they do not lose their place.

Student 2—Detailed Review
Student 2 (S2) used the Audio Notebook very differently from S1. For S2, the Audio Notebook provided the opportunity to re-listen to the lecture and replay particular portions to achieve a better understanding of the material. While S1 did not significantly change her notetaking habits when using the Audio Notebook, S2 began taking fewer detailed notes, and outlining key points from the lectures instead. Upon review, S2 added detailed information not written down while originally attending the lecture. In some ways, S2 relied more heavily on the audio than S1, because she did not take all the information down in class.

The same audio navigation tools—spatial navigation, audio scrollbar, and speed control—were used by each student in different ways. S1 used these tools to jump around in the audio, skip over parts not of interest, and locate specific portions for review. These controls allowed S1 to *skim* quickly through the audio recording. S2 used these same audio navigation controls as a mechanism for detailed review—to replay portions of audio, and transcribe quotes of interest.

Taking Notes

S2 altered her notetaking habits when using the Audio Notebook. The student began writing notes that outlined "key points or concepts" presented in class. She said these notes were more structured than the notes she was taking previously. During class she would create an outline of important information, with bullet items for various concepts. She relied on the Audio Notebook to be able to go back and review this information after the lecture and fill in the detail where needed. Once S2 began using the Audio Notebook, she felt secure that the material was being captured, and there was no need to write down what the professor said verbatim.

Review Sessions

S2 reviewed seven of the ten lectures recorded using the Audio Notebook. Two classes were also recorded onto DAT tape when the Audio Notebook was not available. It is interesting to note that these recordings were never reviewed.

During the first review session, S2 mainly listened to the lecture from start to finish, without skipping portions. The student only made a few selections in the notes. However, she used the scrollbar frequently to back up and replay portions of the recording.

S2 usually spent more time reviewing the recording than in attending the original lecture. This is in sharp contrast to the type of high-speed skimming used by S1. On average, S1 spent 31% of the original listening time reviewing the audio, whereas S2 spent 104% on average. The review time exceeded the duration of the original audio, because S2 often replayed portions of each recording.

Use of the Audio Recordings

S2 used the Audio Notebook as a "study tool." Reviewing the notes was not just a matter of reviewing missing or unclear information. It was a chance to go over the information a second time to obtain a better understanding. During each review session, the student added information to her notes, sometimes transcribing quotes from the professor.

Additional notes were added in pencil so "high level" information could be distinguished from this added detail. Fig. 7 shows an example from one of the notebooks. The writing in gray are the notes added during review. The student sometimes specially marked something during class as a reminder to review this information later. For example, the student wrote "Key things to observe" intending to fill these in upon review. A large space was left in the notes after this phrase for the additional information. During review, more detail was added in this area of the notes. When adding notes, the student sometimes used two hands—simultaneously controlling the audio scrollbar with one hand, and adding notes with the other.

Building a Story around Quotes

Reporter 1 (R1) used the Audio Notebook to record an interview for publication. R1's use of the Audio Notebook provides another perspective on how audio captured by the notebook will be used. R1 used the Audio Notebook primarily to locate and transcribe quotes for his story. In R1's case the search was very directed. He was listening for only

the most important "sound bites" to include in his story. R1's notes were also much sparser than the students' notes. For a reporter performing an interview, it is important to maintain eye contact with the subject, so fewer notes may be taken. His notes were indices of topics and quotes, each line of notes containing only a few words.
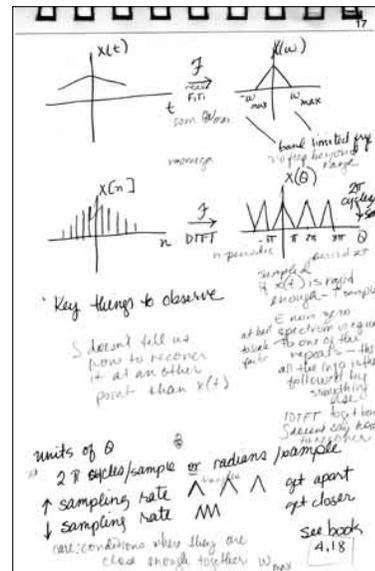

Fig. 7: The student wrote "key things to observe" and filled in the notes (in gray) during a review session.

Review Session

Prior to the review session, R1 went through his notes and made a few annotations. He put star symbols next to important areas, and marked things to review. While reviewing the audio, he made a few additional annotations using a red pen. R1 listened to the interview starting from the top of his first page of notes. Once he heard a quote of interest, he stopped the audio, and typed it into his laptop computer. Then he used the audio scrollbar to backup and replay the quote to check the accuracy of his transcription.

It was interesting that the quotes he selected often were not indexed in his notes. This indicates that navigation using the page alone is not enough. If a portion of the audio is not indexed on the page, it would be lost to the user without the scrollbar. Something that does not seem significant at the time of the original recording may become more important later.

R1 used the scrollbar 3.3 times as often for navigating through the audio then selecting on the page (124 vs. 38). At first, R1 preferred the scrollbar to page selection because when he started playback from the page, it often started in the middle of a quote. He discovered that he could save time by first selecting in his notes to begin playback and then adjusting the starting point using the scrollbar.

Filling in a Story Outline

Reporter 2 (R2) was very skeptical about using the Audio Notebook. He asked several times why he needed the Audio Notebook, and seemed reluctant to use it for the interview. R2 said that he used a tape recorder in the past and found it

"burdensome." He said that when you take notes, you only write down the parts of interest, whereas a tape recorder captures everything and is difficult to find the parts of interest afterwards. R2 took very detailed notes during his interview. He did not rely on the audio recording to capture the information.

Taking Notes

R2 was instructed that he could take notes as he normally would during an interview and that he could determine after the interview whether or not the Audio Notebook was of any use in writing the story. This is another instance that shows the advantage of using actual paper and pen. R2 was not forced to conform to a new way of taking notes.

Review Session

Several weeks after the interview, R2 asked to review the material using the Audio Notebook. Prior to the meeting, he had typed in an outline of the story based on his handwritten notes. His goal was to review the audio to clarify several parts of the story. For example, he wanted to verify the timeline of events that were presented, and look up some specific names and dates.

The total review session lasted 39 minutes, less than two-thirds the time of the original interview. R2 spent approximately 25 minutes updating his story, and the remainder of the time exploring the Audio Notebook. R2 used selections on the pages to jump around in the interview. He located an area of his notes corresponding to the information of interest, and started playback from the beginning of a line of notes. He used the scrollbar to jump forward or backward in the audio 22 times, but used selection on the page approximately twice as often (41 times). He found the scrollbar to be more "hit and miss" than selection from the page itself.

Audio Notebook versus Tape Recorder

Several months later, R2 was interviewing another subject. R2 tape-recorded the interview because he thought it would be important to capture the way things were said. This time, R2 did not have access to the Audio Notebook and used an analog tape recorder instead. After a frustrating time of trying to locate the desired information on the tape, he gave up. He decided the only way to manage the audio would be to transcribe the entire interview. It took him a total of 6 hours to transcribe the approximately 1.5 hour interview. In contrast, using the Audio Notebook, it only took him 39 minutes to review the earlier 68-minute interview, quickly locating and taking notes about the portions of interest.

COMBINING USER AND ACOUSTIC STRUCTURE

The field study enabled us to learn how subjects used the audio recordings given structure provided by the their notetaking activity. However, the study also pointed out areas where this user-structuring of the audio was not enough. Additional information was needed to improve the correlation between the user's notes and the audio, and to provide structure where little or none was generated by the users activity.

After the field study, acoustic processing techniques were developed to further structure the audio for improved access. An acoustic study was performed [8] with two resulting techniques for acoustically structuring a speech recording: phrase detection, and prediction of new topic introductions. This section describes how these techniques are incorporated into the Audio Notebook user interface.

Audio Snap-to-Grid using Phrase Detection

In the field study, users noted that playback often started in the middle of a phrase when they selected on the page. Knowledge of phrase beginnings can be exploited so that when users make selections in their notes, playback starts from the beginning of the nearest phrase. Starting playback in the middle of a phrase is partly caused by the delay between hearing something and writing it down—a listener must first assimilate what was said before writing notes. When users select in their notes to begin playback, the system first determines the closest X-Y location in the stored pen data and the associated time point in the audio recording. Next, to account for the delay between listening and writing, the Audio Notebook subtracts a *listening-to-writing offset* from the playback start time. However, when backing up by a fixed amount, there is no guarantee of finding a coherent starting point in the speech.

In many graphics programs users can turn on a grid so when they are drawing an object, the mouse "snaps" to the nearest grid point. Given knowledge of major phrase break locations, the Audio Notebook can "snap back" to the nearest phrase beginning when a selection is made—we refer to this as *audio snap-to-grid*. A phrase detection algorithm [8] is used to process Audio Notebook recordings. For each recording, the system predicts phrase starting and ending times. Now when users select in their notes to begin playback, the system first backs up by the listening-to-writing offset, and then snaps back to the nearest phrase beginning.

Fig. 8 shows some examples of the correlation between R1's notes and the audio with and without phrase detection. Notice that without phrase detection, the beginning of a phrase was often cut off, or playback started in the middle.

Selection in notes	Playback–no phrase detection	Playback with phrase detection
Speech as interface	"handheld devices"	"the idea was to look at speech as an interface to handheld devices"
faster ~ fedex	"–an the federal express, the guy in the federal express commercials"	"faster than the federal express, the guy in the federal express commercials"
We brain-stormed	"–stormed about it a little bit"	"and we brainstormed about it a little bit"

Fig. 8: Correlation between notes and audio with and without phrase detection.

Topic Suggestions

Additional structural information is also useful for providing navigational landmarks in the audio timeline. For example, when student 1 was trying to skim quickly through a page of notes, she randomly selected on every few ele-

ments in the audio scrollbar. Given knowledge of where new topics begin, the system could suggest places to navigate in the recording. Rather than blindly jumping from one location to another, topic suggestions can provide a guide for the user's skimming activity. Based on the acoustic study, a processing technique was developed for predicting topic introductions using pitch, pausing, and energy. Each suggestion represents a potential topic change location, and is displayed along the audio scrollbar.

The audio scrollbar was designed to be both a control and a display. Prior to the integration of the topic suggestions, the scrollbar was only used to display the audio cursor indicating the users' current position in the timeline for a page. Now topic suggestions are displayed in red along the audio scrollbar. This kind of structural information provides navigational landmarks in the timeline for more intelligent navigational control. Rather than simply jumping to the beginning, middle, or end of the timeline, the user can now jump from one topic suggestion to another. When the user selects on one of the red LEDs, the audio begins to play from the start of the topic prediction (Fig. 9).

Fig. 9: Topic suggestions displayed for student 1's notepad. The user is selecting the first suggestion.

There are many important uses of the topic suggestions. First, topic suggestions index the audio when there is a lack of user activity. Fig. 9 shows some suggestions for one page of S1's notebook. In this example, a large portion of the audio is not indexed because the user did not take notes. Perhaps the user was listening intently to the lecturer, did not believe the information to be important at the time, or had even fallen asleep. The topic suggestions provide structure where none was generated by the user's activity. This is important because the Audio Notebook should free the listener to devote more attention to the talker and not force the user to take verbatim notes.

Secondly, topic suggestions further enhance the correlation between the user's notes and the audio recording. Just as the system "snaps back" to the nearest phrase beginning when the user makes a selection in the notes, the user can also back up to the nearest topic suggestion. In the example shown in Fig. 10, the user has marked down a topic heading in the notes, but it does not correspond exactly to the

beginning of the topic in the audio recording. The user can select on the heading in the notes and then back up to the nearest suggested topic beginning. In this case, when the user selects on the heading titled *4. Metrology and Microscopy*, the audio begins "uh really neat that's called phase conjugation…"; but when the previous topic suggestion is selected, playback begins "Okay number 4 is going to be Metrology and Microscopy, and uh as an example of this… that's going to be uh really neat that's called phase conjugation…"
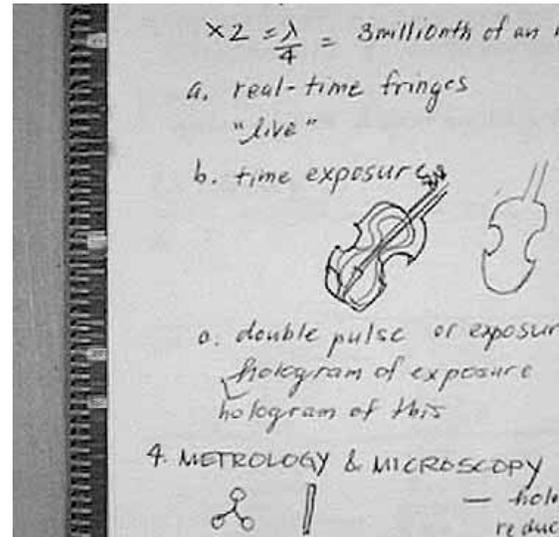

Fig. 10: Topic suggestions (lighted elements in the scrollbar) are displayed for a page of S1's notepad.

In the field study, S1 randomly selected on every few LEDs in the scrollbar, without knowledge of the content. The topic suggestions displayed along the scrollbar now provide a guided way of skimming more efficiently through the recording, allowing the user to jump from one topic suggestion to the next.

Adjusting the Number of Suggestions

The topic predictor will make errors, sometimes missing a topic beginning, other times predicting a topic change where none exists. Tradeoffs must be made between identification and accuracy—the more topics predicted, the more potential for false alarms. Interaction strategies were therefore needed to allow the user to take advantage of correct predictions while coping with potential errors.

The Audio Notebook allows the user to trade off between identification and accuracy. The user is given control over the number of suggestions. The number of suggestions is dependent on the data, so the exact number will vary, falling into one of the following ranges: 5–9 (min), 10–14 (medium), and 15–20 (max). The user can select between these three levels of suggestions using button controls (Fig 11). The fewer the number of suggestions, the higher the accuracy.

Fig. 9 shows the audio scrollbar display of topic suggestions for a page from a S1's notebook. In this photo, the number of suggestions is set to the minimum number. These five topic suggestions begin as follows:

1. "Now there are lots– we're going to work our way through the history of this uh definition number one…"
2. "So this is the stuff we're going to be talking about…"
3. "But it turns out that in China for example they're making counterfeit holograms…"
4. "So there are two issues in security…"
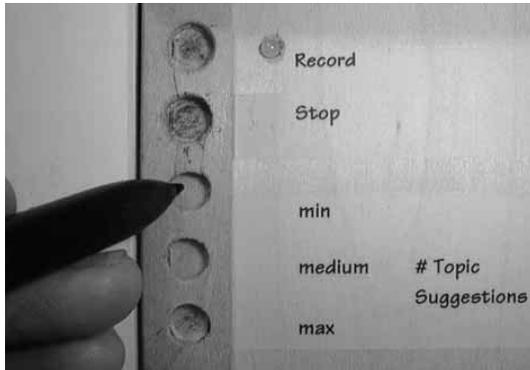5. "Okay okay so that's finishing up chapter one…"

Fig. 11: The user can select between three levels of topic suggestions.

Each of these topic suggestions begins with one of the cue words—"now", "so", "but", or "okay." A cue word at the start of a phrase may signal a topic change. Note that the algorithm selected these phrases using only acoustic analysis, without knowledge of the lexical content of the audio recording.

CONCLUSION

The Audio Notebook takes advantage of the user's natural activity to index an audio recording for later retrieval. Handwritten notes and page turns serve as indices into an audio recording. It is important to emphasize that no *explicit* user segmentation of audio is required (e.g., button pushes to mark important parts). Additional activities during recording can burden and distract listeners from their primary task.

A goal of the Audio Notebook is to free listeners to devote more attention to the talker, so they are not always required to take detailed notes. By providing more accurate and additional structural indices into the audio recording, the system augments the user's activity, helping the user to find the desired portions of audio.

This work has presented a new approach for rapid navigation and skimming in the audio domain. The approach combines user activity and acoustic cues for structuring audio recordings. This structure makes the recordings more accessible and manageable than they have been traditionally, so users can quickly and easily locate portions of interest.

The Audio Notebook provided reassurance that key information would be available for later review yet did not interfere with regular notetaking activity or interactions with others. Rather than replacing real-world objects like paper and pen, we can successfully augment them, combining the advantages of the physical world with the capabilities of digital technology.

REFERENCES

1. G. Abowd. Classroom 2000: An experiment with the instrumentation of a living educational environment. IBM Systems J. 38(4): 508–530, 1999.

2. B. Arons. SpeechSkimmer: A System for Interactively Skimming Recorded Speech. ACM Trans. on Computer-Human Interaction, 4(1):3–38, 1997.

3. J. Landay and R. Davis. Making sharing pervasive: Ubiquitous computing for shared note taking. IBM Systems J. 38(4): 531–550, 1999.

4. M. Lamming. Towards a Human Memory Prosthesis. Tech. Rpt. #EPC-91-116 Rank Xerox EuroPARC, 1991.

5. S. Minneman, et al. A Confederation of Tools for Capturing and Accessing Collaborative Activity. Proc. of Multimedia '95, p. 523–534. ACM, 1995.

6. W. Newman and P. Wellner. A Desk Supporting Computer-Based Interaction with Paper Documents. Proc. of CHI '92, p. 587–592. ACM, 1992.

7. L. Stifelman. Augmenting Real-world Objects: a Paper-based Audio Notebook. CHI Conference Companion. p. 199–200. 1996.

8. L. Stifelman. The Audio Notebook: Paper and Pen Interaction with Structured Speech. Ph.D. Thesis. MIT, 1997.

9. K. Weber and A. Poon. Marquee: A Tool for Real-Time Video Logging. Proc. of CHI '94, p. 58–64. ACM, 1994.

10. L. Wilcox, B. Schilit and N. Sawhney. Dynomite: A Dynamically Organized Ink and Audio Notebook. Proc. of CHI '97, p. 186–193. ACM, 1997.

11. P. Wellner. Interacting with Paper on the DigitalDesk. *Comm. of the ACM*, 36(7):87–96, 1993.

12. P. Wellner, W. Mackay and R. Gold. Computer-Augmented Environments: Back to the Real World. *Comm. of the ACM*, 36(7):24–26, 1993.

13. S. Whittaker, P. Hyland and M. Wiley. Filochat: Handwritten Notes Provide Access to Recorded Conversations. Proc. of CHI '94, p. 271–277. ACM, 1994.