Meeting Capture in a Media Enriched Conference Room

Patrick Chiu, Ashutosh Kapuskar, Lynn Wilcox

FX Palo Alto Laboratory
3400 Hillview Ave, Bldg 4, Palo Alto CA 94304, USA
Email: {*lastname*}@pal.xerox.com

Sarah Reitmeier

University of Michigan
School of Information
Ann Arbor, MI 48109, USA
Email: sreitmei@umich.edu

Abstract. We describe a media enriched conference room designed for capturing meetings. Our goal is to do this in a flexible, seamless, and unobtrusive manner in a public conference room that is used for everyday work. Room activity is captured by computer controllable video cameras, video conference cameras, and ceiling microphones. Presentation material displayed on a large screen rear video projector is captured by a smart video source management component that automatically locates the highest fidelity image source. Wireless pen-based notebook computers are used to take notes, which provide indexes to the captured meeting. Images can be interactively and automatically incorporated into the notes. Captured meetings may be browsed on the Web with links to recorded video.

Keywords. meeting capture, note taking, roomware, cooperative buildings, multimedia applications, video applications

1 Introduction

Public conference rooms are sites of meetings and organizational activities that contain a wealth of visual and verbal information. Meetings span a broad spectrum of informational and collaborative activities; examples are staff meetings, design discussions, project reviews, video conferences, presentations and classes. It is often important to have a record of the meeting. This is usually done with handwritten notes, augmented with presentation material that is either hand copied or obtained from the speaker. In some cases, more detail is needed and the meeting is recorded on audio or video. A meeting record allows people who were at the meeting along with those who were absent to review the meeting. The tasks performed during review can be

simple retrieval of facts and details, or more involved activities such as studying, preparing reports, and creating meeting summaries.

Multimedia is a promising technology for supporting meeting capture and note taking. It can capture activity in the meeting room as well as the presentation material. Digital video has been used for meeting capture in systems such as STREAMS (Cruz and Hill, 1994), but it uses a room camera to take images of the presentation material and is subject to poor image quality and interference when people or objects obscure the display. Other systems like Tivoli (Moran et al., 1996, 1997) and Classroom 2000 (Abowd et al., 1996, 1998) uses LiveBoard electronic whiteboards (Elrod et al.,1992, Pedersen et al., 1993) to capture visual material indexed to an audio recording. Classroom 2000 supports note taking on PDA devices with pre-loaded presentation slides and lacks the flexibility of real time slide capture. The Coral system (Minneman et al. 1995) is a confederation of tools that support multimedia recording of meetings. Coral also provides infrastructure for synchronization of video to digital ink notes taken with Marquee (Weber and Poon, 1994). With Marquee, images cannot be incorporated into the notes. Forum (Isaacs et al., 1994) is a workstation-based system that uses video for distributing live presentations and allows users to annotate slides with keyboard and mouse.

At the FX Palo Alto Laboratory, we have a media enriched conference room equipped for meeting capture with room cameras and microphones, video conference cameras, and a large display rear video projector. A variety of *roomware* (Streitz et al., 1998) facilitates the capture, display, and transfer of multimedia information. Meeting capture at its most basic level is supported by recording the video and audio streams, and by taking notes on wireless pen notebook computers. The images of the room activity and the presentation material can be interactively incorporated into the meeting notes. High quality images of the presentation material are captured by a smart video source management component. Captured meetings and notes with links to recorded video may be reviewed on the Web.

This paper is organized as follows: Section 2 describes the media enriched conference room, Section 3 discusses how meeting capture and note taking is performed, Section 4 shows accessing and browsing a captured meeting, Section 5 explains the media management and system architecture, and Section 6 is on user experience.

2 A Media Enriched Conference Room

The conference room at our lab is designed to support multimedia meeting capture and note taking in a flexible, seamless, and unobtrusive manner in a public conference room that is used for everyday work. A blueprint of the room is shown in Fig. 1, and a photo in Fig. 2. The center area of the room has the typical and familiar conference room furniture with standard tables and chairs in a U-shaped arrangement. As encountered in a field study by Covi et al. (1998), most shared meeting rooms have only tables and chairs, and it is useful to be able to work in our conference room in this familiar setting. For interacting with the digital world, wireless pen-based notebook computers, which may be freely positioned and moved around the room, serve as unobtrusive devices for meeting capture.
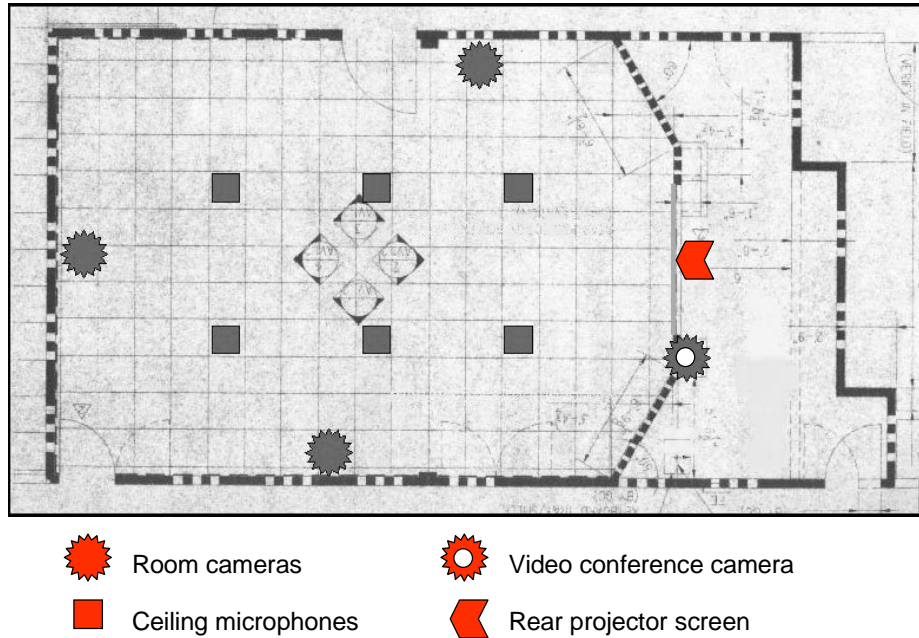
Fig. 1. Blueprint of conference room.

Fig. 2. Picture of conference room.

PC display, keyboard
and mouse

Media selection
and controls

Fig. 3. Podium and room viewed from the front.

On the front wall of the room is a flush-mounted large 120-inch screen rear video projector for displaying presentation material. Video of presentation material is fed into the rear projector from any of the following: a PC workstation, a document camera, a VCR, or a DVD player. A control room houses all of this hardware out of sight, and is walled off from the conference room with its own entrance in the hallway (see the right of Fig. 1). The document camera folds up and retracts into a podium drawer. A user may bring a laptop and plug it into a connector at the podium. The podium has controls that allow the presenter to select a source for the rear projector (see Fig. 3). It also has a thin LCD display, a keyboard and mouse hooked up to the PC in the control room.

There are three computer controllable cameras in the room plus a video conference camera for capturing and transmitting room activity. A room camera can be used to obtain an image of the whiteboard. Audio is handled by six ceiling microphones, combined into a single audio stream and mixed together with the video. Network connectivity is provided by a 1Mb wireless system. A small ink jet printer is available to produce hardcopies of notes or presentation material.

The room cameras may be tilted, panned, and zoomed from the control room. We have presets programmed for different types of meetings. For example, in a presentation meeting, one of the side cameras is aimed at the speaker at the podium, the other side camera at the participants around the table, and the back wall camera is set for a wide-angle shot of the whole room. When higher quality production is required, a person sits in the control room and directs the cameras.

With this setup, the underlying medium for capturing all types of visual images is video. The room video cameras provide images of the room activity and the scribbles on the whiteboard, the video conferencing system provides images of a remotely connected room, and the rear video projector provides images of the presentation material. Thus, video gives a seamless and flexible way to capture a variety of visual information from a meeting. There is a tradeoff between versatility and fidelity, which we will discuss in a later section. Before doing that, we describe how meeting capture is performed.

3 Meeting Capture

A meeting in the conference room is captured by recording the video streams from the room cameras, video conference sources, and rear video projector. The audio is captured by the ceiling microphones and mixed into the video streams from the room cameras. For later browsing and access, indexes for the video recordings are extremely helpful. A natural way to obtain indexes is to make use of notes taken by meeting participants. For this purpose, we have designed and built a client-server application called NoteLook. The standard technique of time-stamping notes and correlating them to multimedia data for retrieval was pioneered by Lamming and Newman (1991), and may be found in systems such as We-Met (Wolf et al., 1992), Filochat (Whittaker et al., 1994), Tivoli (Moran et al., 1997), Classroom 2000 (Abowd et al., 1996), Dynomite (Wilcox et al., 1997), and Audio Notebook (Stifelman, 1997).

NoteLook allows the user to take handwritten notes and interactively incorporate images from the room cameras, video conference cameras, and rear projector into the note pages. The client application runs on wireless pen-based notebook computers in the room (see Fig. 2). Users can write annotations and freeform notes with digital ink. A screen shot is shown in Fig. 4.
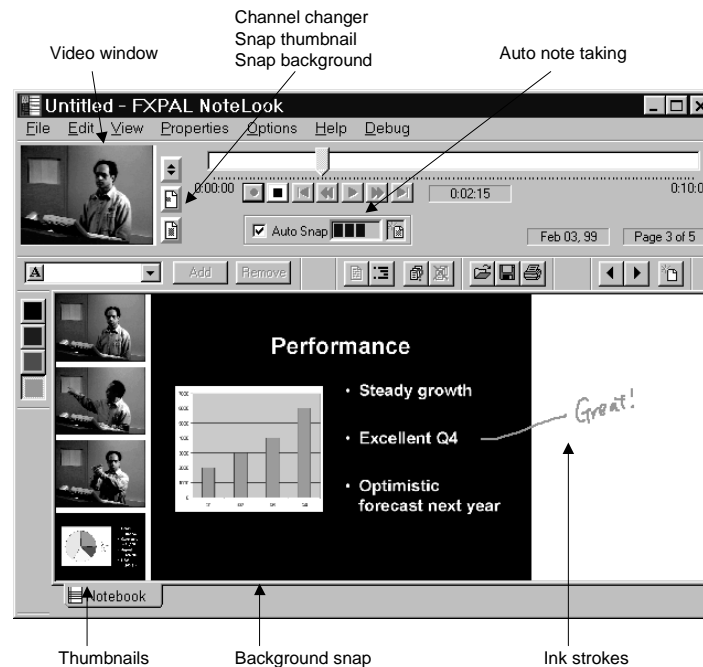

Fig. 4. NoteLook client application screen shot.

The user can view live video in the small window on the upper left corner. Next to the video window are three buttons for interacting with the video. The top button changes the video channels. We currently support two channels: one for the room activity from a pre-selected room or video conference camera, and one for presentation material shown on the rear projector. Usually, the pre-selected camera is a room camera pointed at the speaker at the podium. The middle button snaps the image in the video window as a thumbnail into the margin of the note page. When a sequence of thumbnails is snapped, they are automatically placed one below another. The bottom button snaps in a large background image. A newly snapped background image overwrites an existing background image on a page.

The interaction technique in NoteLook is *YCAGWYS (You Can Always Get What You See)*. Images of the room activity and the presentation material can be captured in real time as the user sees them. By using video as the underlying medium, this is accomplished by NoteLook in a seamless manner. To transfer information at a finer granularity between the shared display and pen-based notebooks, it is possible to employ techniques such as Pick-and-Drop (Rekimoto, 1998).

NoteLook has a set of standard VCR-type controls for recording and playback. Pressing the RECORD button makes a connection to the NoteLook server and initiates video recording and transmission to the clients. The video window displays the live video during note taking and the recorded video during playback. Above the VCR controls is a timeline with a pointer indicating the current video time position.

At the top right corner of the note page are buttons for previous page, next page, and new page. On the left is a palette of four pen colors for writing notes and annotations. Underneath the video window is a list box for entering keywords, and adjacent to the right is a set of four buttons for query and retrieval. The query and retrieval features are inherited from its predecessor Dynomite (Wilcox, 1997), which is a stand-alone note-taking application with audio and ink.

Furthermore, NoteLook has a facility for automatic note taking. In this mode, when the presenter puts up a new slide on the rear projector, it is automatically detected and snapped in as a background of a new page, and this page is appended to the stack of note pages. Also, a sequence of thumbnails from the room cameras is placed in the margin of that page. When the user turns to that page, she can annotate the images with ink. This feature relieves the user of the repetitive task of snapping in many slides during a presentation. In our experience, it is common to see 20 slides in a presentation and we occasionally have talks with over 50 slides.

4 Accessing and Browsing Captured Meetings

Captured meetings that have been indexed with NoteLook notes may be browsed on the Web. A sample is shown in Fig. 5. The NoteLook application has a menu command to generate HTML pages. On the Web pages, the thumbnails, background snaps, and ink strokes have links to the recorded video. These objects are all time-stamped during note taking, and the video playback is correlated to those times. The video is played back in a separate application window. We have integrated NoteLook Web pages with a video playback application developed at our lab called the Metadata Media Player (Girgensohn et al., 1999).
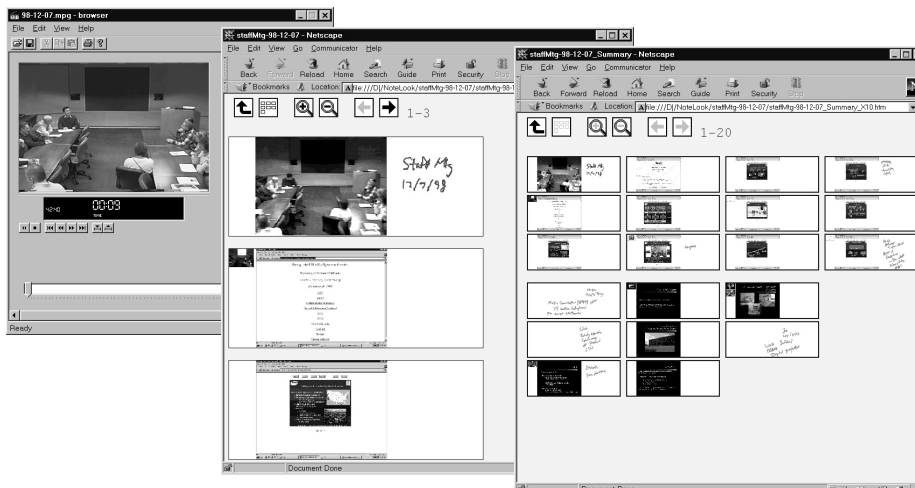
Fig. 5. Web access to captured meetings and notes. On the right is a table of contents page, in the center is a NoteLook Web page, and on the left is the Metadata Media Player.

Additionally, there are several standard navigational features on the Web pages. These are straightforward and we will only give a brief description. A top level page lists available NoteLook notes. For each session, a single table of contents page shows reduced images of all the pages. Clicking on a reduced image of a page brings up that page. Each page may be zoomed in or out with a range of five different magnification levels.

For private notes, users can store and playback NoteLook files on pen computers like the ones used in the conference room for meeting capture. The thumbnails, backgrounds, and ink strokes can be "played" by selecting them and pressing the PLAY button on the VCR controls. The video plays back in the NoteLook video window (see Fig. 4).

5 Media Management and System Architecture

The NoteLook client application is designed to be lightweight and flexible. However, digital video is a heavyweight medium because a substantial infrastructure is required to obtain adequate quality images of the room activity and presentation material. To deal with this tradeoff, we off-load most of the video processing and media management to the NoteLook servers and switchers. While the space in the conference room is relatively clutter free (as shown in Fig. 2), there are many pieces of the system outside the room hidden away from the users. The various components of the system are shown in Fig. 6. The key pieces are the NoteLook clients, servers, and switchers for video source management. We describe the interplay of these along with other components in more detail below.
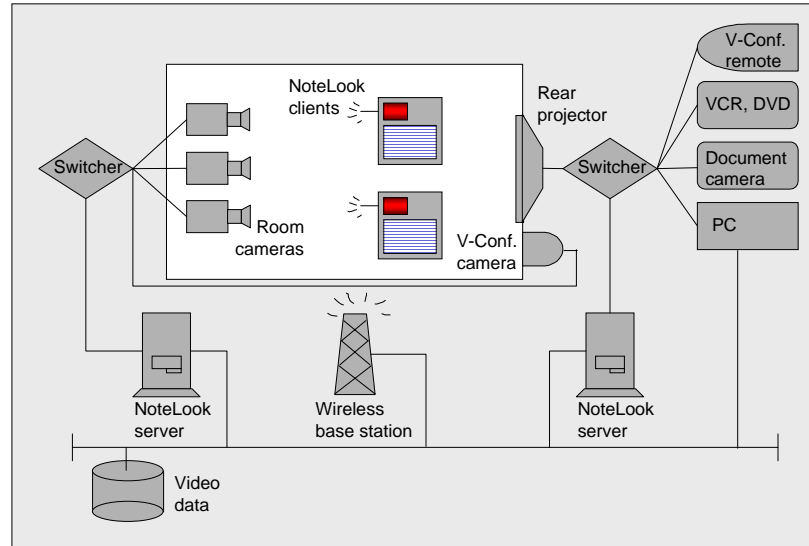
Fig. 6. NoteLook system architecture.

The NoteLook system is auto-configurable, extensible, and scalable. The clients and servers configure themselves automatically using resource discovery techniques. Adding and removing servers or channels does not require modifying existing clients, and multiple clients are supported by multicasting.

Each video channel corresponds to a server, which is associated to a set of sources. Currently, we support two channels: one for the room activity given by the room cameras and video conference cameras, and one for the presentation material given by the set of sources that feed into the rear video projector. The switchers are used to manually and automatically select the desired source.

A smart source management component addresses the *versatility/fidelity* tradeoff. Video provides a versatile way to capture room activity and presentation material. The images of the presentation material can come in a variety of forms: PowerPoint slides or Web pages from a computer, paper or plastic transparency overhead slides via the document camera, whiteboard via a room camera, video clips from VCR or DVD, etc. While the rear projector video feed is versatile enough to capture images of any type of presentation material, it does not always provide the highest quality images. For example, by the time an image of a PowerPoint slide travels from a PC's video output through the plumbing (which may contain various splitters and scan converters) and reaches the rear projector, the captured image is degraded to a level that sometimes makes it difficult to read the text on the slide.

The source management component deals with this problem by identifying the highest fidelity source available for capturing images. In the previous example, when the rear projector displays PowerPoint slides running from the PC workstation, the source management component directs the server to get the images from the PC by screen snap (i.e. the PC's screen bitmap, not the PC monitor video signal, not the rear projector video signal). In the case when a speaker supplies her own laptop, the server must gets its images further downstream from the video signal of the rear projector with some unavoidable degradation in fidelity. The source management component

operates automatically in real time and interfaces with the switchers and a commercial AMX room control system. The result is that the best obtainable images are always captured while video source management is hidden from the user.

The NoteLook servers take video and audio inputs, process them, transmit the output to the NoteLook clients, and store the data for later retrieval. When a user initiates a session by pressing the RECORD button on the client application, it broadcasts a request for service, the servers respond and identify themselves, and connections are established. The video is transmitted to the clients at a highly reduced frame rate (1 per 2 seconds) to conserve wireless bandwidth. Meeting participants do not necessarily need full motion video for note taking since they are present in the room watching the live action. Automatic note taking is handled by a software component that runs on the servers and analyzes the video data. When the speaker puts up a fresh slide, it is detected and packaged along with a sequence of thumbnails of room images, and these are sent to the client for creating a new note page.

6 User Experience

We have conducted a user study over a six-week period with 13 meetings (Reitmeier et al., 1998). These meetings were presentations, staff meetings, and Japanese classes. We found that the system performed successfully for meeting capture and note taking. It supported seamless capture of room activity and a variety of presentation material. From interviews, we found the system to be minimally intrusive to the speaker and the participants in the room. The user study provided insights that resulted in several refinements to the system, notably it led us to develop the video source management component and automatic note taking feature.

We are currently using the meeting capture capabilities of our media enriched conference room in many of our meetings. Over the long term, we plan to gain more usage experience, continue to refine the system design, and observe how it co-evolves with the meeting work practice.

Acknowlegements

References

1. Abowd, G. D., Atkeson, C. G., Brotherton, J., Enqvist, T., Gulley, P., and LeMon, J. (1998). Investigating the capture, integration and access problem of ubiquitous computing in an educational setting. *Proceedings of the CHI '98 Conference*. ACM Press, pp. 440-447.
2. Abowd, G. D., Atkeson, C. G., Feinstein, A., Hmelo, C., Kooper, R., Long, S., Sawhney, N., and Tani, M. (1996). Teaching and learning as multimedia authoring: the classroom 2000 project. *Proceedings of the ACM Multimedia '96 Conference*. ACM Press, pp. 187-198.

3. Covi, L., Olson, J., Rocco, E., Miller, W., Allie, P. (1998). A room of your own: What do we learn about support of teamwork from assessing teams in dedicated project rooms? *Proceedings of CoBuild '98.* LNCS 1370. Springer - Verlag, Heidelberg, pp. 53-65.

4. Cruz, G. and Hill, R. (1994). Capturing and playing multimedia events with STREAMS. *Proceedings of the ACM Multimedia '94 Conference*. ACM Press, pp. 193-200.

5. Elrod, S., Bruce, R., Gold, R., Goldberg, D., Halasz, F., Janssen, W., Lee, D., McCall, K., Pedersen, E., Pier, K., Tang, J., Welch, B. (1992). LiveBoard: A large interactive display supporting group meetings, presentations and remote collaboration. *Proceedings of the CHI '92 Conference*. ACM Press, pp. 599-607.

6. Girgensohn, A., Boreczky, J., Wilcox, L., Foote J. (1999). Facilitating video access by visualizing automatic analysis. *Proceedings of Interact '99, to appear.*

7. Isaacs, E. A., Morris, T., and Rodriguez, T.K. (1994). A forum for supporting interactive presentations to distributed audiences. *Proceedings of CSCW '94*. ACM Press, pp. 405-416.

8. Lamming, M. and Newman, W. (1991). Activity-based information technology in support of personal memory. Technical Report EPC-1991-103, Rank Xerox, EuroPARC, 1991.

9. Minneman, S., Harrison, S., Janssen, B., Kurtenbach, G., Moran, T., Smith, I., and van Melle, B. (1995). A confederation of tools for capturing and accessing collaborative activity. *Proceedings of the ACM Multimedia '95 Conference*. ACM Press, pp.523-534.

10. Moran, T. P., Chiu, P., Harrison, S., Kurtenbach, G., Minneman, S., and van Melle, W. (1996). Evolutionary engagement in an ongoing collaborative work process: a case study. *Proceedings of CSCW '96*. ACM Press, pp. 150-159.

11. Moran, T. P., Palen, L., Harrison, S., Chiu, P., Kimber, D., Minneman, S., van Melle, W., and Zellweger, P. (1997). "I'll get that off the audio": a case study of salvaging multimedia meeting records. *Proceedings of CHI '97*. ACM Press, pp. 202-209.

12. Pedersen, E. R., McCall, K., Moran, T. P., and Halasz, F. G. (1993). Tivoli: An electronic whiteboard for informal workgroup meetings. *Proceedings of INTERCHI '93*. ACM Press, pp. 391-398.

13. Reitmeier, S., Chiu, P., Bly, S., Kapuskar, A., Wilcox, L. (1998). NoteLook User Study. FXPAL Technical Report TR98-039, FX Palo Alto Laboratory.

14. Rekimoto, J. (1998). Multiple-Computer Interfaces: A cooperative environment consisting of multiple digital devices. *Proceedings of CoBuild'98*. LNCS 1370. Springer - Verlag, Heidelberg, pp. 33-40.

15. Stifelman, L. (1997). *The Audio Notebook: Paper and Pen Interaction with Structured Speech.* PhD Thesis. MIT, 1997.

16. Streitz, N., Geißler, J., Holmer, T. (1998). Roomware for cooperative buildings: Integrated design of architectural spaces and information spaces. *Proceedings of CoBuild '98*. LNCS 1370. Springer - Verlag, Heidelberg, pp. 4-21.

17. Weber, K. and Poon, A. (1994). Marquee: A tool for real-time video logging. *Proceedings of CHI '94*. ACM Press, pp.58-64.

18. Whittaker, S., Hyland, P., and Wiley, M. (1994). Filochat: handwritten notes provide access to recorded conversations. *Proceedings of CHI '94*. ACM Press, pp. 271-276.

19. Wilcox, L. D., Schilit, B. N., and Sawhney, N. (1997). Dynomite: A Dynamically Organized Ink and Audio Notebook. *Proceedings of CHI '97*. ACM Press, pp. 186-193.

20. Wolf, C., Rhyne, J., and Briggs, L. (1992). Communication and information retrieval with a pen-based meeting support tool. *Proceedings of CSCW '92*. ACM Press, pp. 322-329.