Multiple View Perspectives: Improving Inclusiveness and Video Compression in Mainstream Classroom Recordings

Raja S. Kushalnagar*, Anna C. Cavender†, Jehan-François Pâris‡

Information and Computing Studies
Rochester Institute of Technology
Rochester, NY 14623 USA
rskics@rit.edu

†Computer Science & Engineering
University of Washington
Seattle, WA 98195 USA
cavender@cs.washington.edu

‡Computer Science
University of Houston
Houston, TX 77204 USA
paris@cs.uh.edu

ABSTRACT

Multiple View Perspectives (MVP) enables deaf and hard of hearing students to view and record multiple video views of a classroom presentation using a stand-alone solution. We show that deaf and hard of hearing students prefer multiple, focused videos over a single, high-quality video and that a compacted layout of only the most important views is preferred. We also show that this approach empowers deaf and hard of hearing students by virtue of its low cost, flexibility, and ease of use in the classroom.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems; K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities

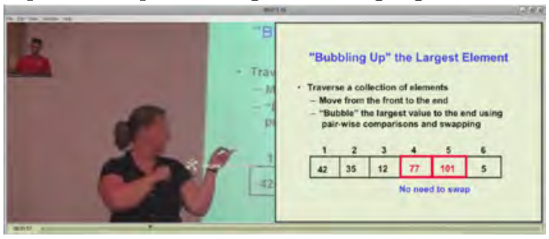
General Terms

Design, Experimentation, Human Factors

Keywords

Accessible Technology, Deaf and Hard of Hearing Users

(a) A mainstream classroom with interpreter and captions. Important regions are highlighted.



(b) MVP with only the most important information presented visually closer together.

Figure 1: Multiple View Perspectives (MVP) brings important aspects of the classroom visually closer together for deaf and hard of hearing students.

1. INTRODUCTION

Deaf and hard of hearing students often utilize sign language interpreters or real-time captioners to access mainstream classroom lectures. Students have to divide their visual attention between these and other information sources such as presentation materials, instructor, personal notes, and other classmates, which are often spatially distributed around the classroom. Consequently, students can easily miss information because they can only attend to one of these information sources at any given time. Capturing the lecture with cameras, whether in real-time or for post-lecture viewing, and presenting those views on a single laptop screen may help by bringing the many sources of information visually closer together [1, 7]. Figure 1 illustrates a typical classroom (a) compared to our Multiple View Perspective (MVP) interface designed to optimize viewing (b).

Traditional lecture capture systems seem like promising solutions, but typically use only one camera to capture video, and, unless dedicated staff are on hand to manage the system, the view from the camera does not change and results in a video that is boring to watch [9]. With only a single view, users may lack the visual information required for adequate context [14]. Dedicated video production staff can improve video context and interest, but at a significant cost. Moreover, a camera operator operating a single video camera or even multiple video cameras may not be able to predict the learning needs of a deaf or hard of hearing student, let alone several deaf and hard of hearing students.

MVP is a collection of video perspectives where a video

*This work was done while the author was with the Computer Science Department at University of Houston.

ASSETS'10, October 25–27, 2010, Orlando, Florida USA.

perspective is a context-aware video view of a specific classroom region. MVP overcomes the limitations of traditional lecture capture systems and empowers deaf and hard of hearing students to independently bring regions of the classroom together onto one screen. This becomes useful in modern classrooms that often contain multiple visual sources of information, such as overhead slides, demonstrations, white boards, and other students. It also provides a mechanism for deaf and hard of hearing users to easily customize their learning focus in the classroom for their own needs, by enabling them to select, focus and customize their view.

In this paper, we analyze the effectiveness of Multiple View Perspectives for improving visual access for deaf and hard of hearing students in viewing and recording accessible mainstream lectures. We show that for deaf and hard of hearing students viewing and recording a classroom presentation via the student's laptop and multiple mobile cameras, the MVP approach can reduce information loss and positively affect cognitive load.

2. BACKGROUND AND RELATED WORK

The basis for our Multiple View Perspective (MVP) work is rooted in a long history of cognitive psychology, literature on the visual needs of deaf and hard of hearing students in learning environments, and our own experiences and observations. The following outlines this prior work and ends with design guidelines we have developed based on this work.

2.1 Visual Perception and Deaf Students

Cognitive load refers to the extent to which cognitive resources, especially working memory, are utilized during learning, problem solving, thinking, and reasoning. Research on human memory has demonstrated that, in contrast to long term memory, which is practically unlimited, working memory has a limited information storage capacity and is fragile, that is, distractions can easily cause forgetting. A high working memory load may be caused by new information (extraneous cognitive load) and the complexity of information (intrinsic cognitive load), and it may interfere with other cognitive processes and performance. Cognitive psychology distinguishes between focused attention, which is the processing of a single input, and divided attention, which is the simultaneous processing of multiple signals [16].

Students learn, think and reason through a teacher's presentation. The presentation information passes through the student's working memory, which has a limited information storage capacity. Hearing people can effectively receive and process information simultaneously through listening and reading in the classroom as they use separate working memory buffers (auditory and visual respectively). But in the case of multiple visual information streams, hearing and deaf alike may have a harder time as the multiple visual streams compete for the same visual working memory buffer.

Accessible multimedia that includes visual representation of the audio stream (i.e. sign language interpreters or captions) may result in visual dispersion and cognitive overload. Visual dispersion is defined as the juggling of multiple concurrent visual streams, and is often cited as a major reason why deaf and hard of hearing students get less out of classroom lectures than their hearing peers [8]. Therefore, accessible multimedia that includes visual representation of the audio stream must be presented in a way that reduces the effects of visual dispersion and cognitive overload [1].

Previous research shows that hearing students benefit from combined visual and auditory materials [12] and multi-modal classrooms are now becoming the norm. Furthermore, while hearing students can simultaneously view a region of interest and listen to the audio using separate modal senses and working memory, this processing strategy is not available to deaf and hard of hearing students receiving accessible presentations. Instead, the classroom auditory information is translated into a visual representation that is simultaneously presented with the classroom video - a simultaneous multiple visual view. Therefore deaf students have to multiplex their single visual channel attention between the multiple visual streams, especially the visual representation of the auditory channel and the instructor's current visual focus, usually the slides or white board [10, 11].

Although multiple perspectives aid in learning and retention, it is difficult to capture and show all perspectives simultaneously with a single video camera due to the single angle and minimum resolution. This raises a significant problem for deaf participants, who acquire information primarily through vision; for them, rapidly switching and simultaneously processing information among multiple perspectives can rapidly exhaust the attention resources, more so when these perspectives are dispersed.

MVP can improve access for deaf and hard of hearing consumers by reducing their cognitive load while viewing multimedia content using accessible technology solutions, especially in classroom lectures.

2.2 Technological Solutions

Studies of classroom presentation tradeoffs indicate that capture, broadcasting and viewing solutions should be portable, passive, require no pre- or post-processing, capture synchronized high resolution visual aids, and capture audio and video of the lecture automatically [20]. One major line of research in automating classroom capture focuses on harnessing advances in motorized cameras in improving single video capture of meetings by automating zoom techniques [14]. This set up is not flexible or easy to deploy in the classroom, and also cannot handle multiple simultaneous views. Another research line used an array of board cameras to capture a wide angle field [4] or fish-eye lenses for low cost [19]. This approach captures more data than needed for classroom recordings, and could perhaps be combined with our MVP approach. That line of work has been extended by building in software tracking of the speaker within a panoramic camera view [3]. While this work has contributed to the ease and inexpensiveness of classroom recording and a single camera with region-of-interest capability is physically simpler to deploy in the classroom than multiple cameras, the main disadvantage of this region-of-interest approach is that does not eliminate bandwidth costs of high definition video recordings, and could not easily accommodate multiple students with differing interests.

With the advent of inexpensive camera-equipped smart phones or netbooks that cost much less than traditional video cameras, it is now possible to process and distribute automated multiple video feeds cheaply. Separate capture and transmission of distinct multiple perspectives of a presentation to participants offers several advantages. First, separate video feeds for each focus enables participants to simultaneously see clearly and select from an array of presentation foci with a preset zoom, angle and resolution. This eliminates the need for a camera operator. Second, participants take control of their learning process by prioritizing their views of the multiple video streams to accommodate their visual learning preferences, rather than depending on the camera operator's preferences. Third, grouping the foci on their computer reduces their visual dispersion demands, especially for deaf participants dependent on visual translations of audio information. Fourth, participants have the option to rewind real-time recordings of each perspective to review any missed information while still keeping up with the presentations in other perspectives. The MVP approach is more economical in that it uses lower cost webcams or camera equipped mobile phones, and puts more of the control in the hands of the student. It also empowers students to pick which configurations will work best for them and gives the control that no automated technique could predict [1].

2.3 Design Criteria

Based on prior work as well our own observations and experiences, we have developed the following design criteria for effective presentation of video for deaf and hard of hearing students:

1. The presentation must have enough visual information to be understood by deaf and hard of hearing viewers. Capture of the visual streams can be challenging in that the presentation may involve simultaneous views of the presenter, the presentation and the visual translations of the audio of the presentation.
2. The presentation views must have enough resolution to enable the deaf and hard of hearing viewers to comfortably understand the information and dialog presented.
3. Setting up the presentation view must not require substantial effort.
4. The hardware and software requirements must not minimize cost and intrusiveness.

3. MVP CAPTURE

MVP capture is a context-aware video encoding and delivery technique that reduces cognitive load and bandwidth demand by capturing and delivering multiple non-overlapping small targeted video streams, as opposed to a single large video stream. This approach is possible and effective because presentations normally have multiple non-overlapping regions of interest that have different sizes, resolutions and angles as seen from the viewer's perspective. The MVP approach discards less-interesting regions for further space and bandwidth savings.

The MVP capture approach can easily capture multiple simultaneous regions of interest either through a single high definition camera or through multiple cheap, off-the shelf mobile camera-equipped devices. Either approach avoids the expense and inflexibility of a dedicated camera operator, although the multiple camera approach has an advantage in capturing any region of interest in the classroom, not merely a single perspective (an example MVP set up is shown in Figure 2). MVP enhances usability and reduces technological and economic demands through the following process optimizations: regions of interest are extracted to reduce field of view, view context awareness enhances video

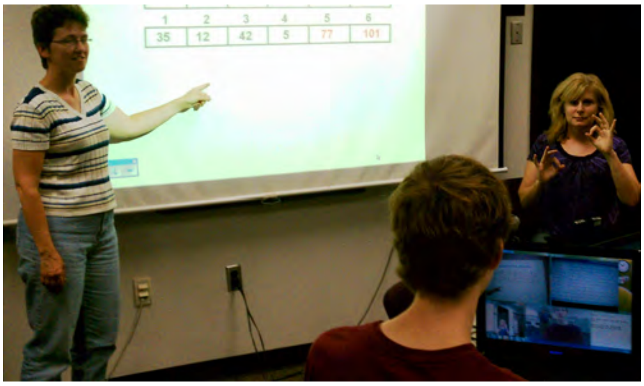


Figure 2: An example MVP set up.

compression, and presented views are prioritized. Collectively these optimizations create a very scalable and robust system that is a standalone, portable system for deaf and hard of hearing students.

Moreover, since deaf and hard of hearing students cannot always rely on the presenter to provide technological accommodations, it is better to empower them by providing a portable, personal technological solution. This approach empowers Deaf students in many subtle ways. For example, their seating choices are typically limited; in order to minimize information loss, interpreters and captioners sit obtrusively next to the slides and instructor and deaf and hard of hearing students must sit up near the front. If they sit in the middle or rear, they are more likely to have poor, blocked or widely dispersed views. Mobile ubiquitous devices can unobtrusively be positioned and brought into MVP to allow the student more freedom in the classroom.

This system can also be used by instructors to distribute multiple smaller server-hosted video streams and let individual students select which streams they will subscribe to. This combination allows each student to fit received streams within their available bandwidth, thus maintaining a reasonable quality of service and an effective learning experience. They can see multiple perspectives simultaneously without losing much information, by placing the views side by side. it can also be viewed as an universal access approach that benefits other classroom participants including the presenter, other students and access staff such as interpreters or captioners.

As an example, consider Jane and Joe, two first-year Deaf students enrolled in an accessible introductory computer programming class at their university. Access to classroom audio is provided via a sign language interpreter and a captioner. Jane prefers captions, but also likes to watch interpreters. Joe prefers interpreters only. A high definition recording of the classroom can adequately capture most of the regions of interest common to both students, namely the instructor, captions, sign language interpreter and overhead slide views, although it cannot capture the students' questions and answers in the lecture without the aid of a camera operator. However, because their laptop screen neither has the size nor resolution of a high definition television set, their view of the slides, interpreter and captions is not optimal. By utilizing the fact that both Jane and Joe are not interested in the areas that fall outside the regions of interest, they need not capture nor record these areas. The collection of views has equal resolution and quality as the original high definition view, but consumes less screen space, video file size and draws closer the views.

3.1 Video Compression and Scalability

Video compression discards information that is either redundant or noticeably unimportant to human viewers. Stated another way, compression looks for similar information and collapses them together. MVP uses the H.264 video compression standard that eliminates spatial redundancy in individual picture frames, which yields a compression ratio of up to 25:1. It then eliminates temporal redundancy, i.e., the similarities between series of pictures in the video, which yields a compression ratio of up to 200:1.

The crux of the MVP approach is to then apply different compression profiles to different regions of interest (views) in presentations. In a presentation, each view has different characteristics and can be treated separately. This approach avoids the global constraint problem. For example, a minimum frame capture rate depends on how fast the content changes. A global view recording must set a minimum frame capture rate to match the region with the most rapidly changing content, regardless of the rate of change in other regions. On the other hand, local view recordings can set more optimal minimum frame capture rates for each region of interest depending on its individual rate of change. For example, to accurately capture a fast moving sign language interpreter, a view with rapidly changing content, the minimum frame capture rate cannot be lowered from the full recording rate. Conversely, to accurately capture slides, white boards or real-time transcripts, which are all views with slowly changing content, lower minimum frame capture rates can be used to yield a compression ratio of up to 30:1. This frame rate capture scaling enables us to provide meaningful quality of service for end-users who are viewing presentations with multiple regions of interest, even if they have limited bandwidth and screen resolution, as it is often the case for smart phones and notebooks.

Modern video viewing clients vary widely in download bandwidth and video resolution viewing capacity: connection speed can vary from gigabit to 4G speeds, and video resolution can vary from high definition resolution on home theater systems to standard definition or smaller on handheld mobile peers. Supporting receivers at a single video streaming rate or resolution is not appropriate, as it can either overwhelm slower receivers, or provide insufficient quality to powerful receivers. An elegant solution is to divide the video stream into sub-parts that can be independently viewed, and combined to reconstitute the original. The substreams can be divided in terms of temporal, spatial or resolution properties.

The MVP video partitioning scheme is orthogonal to, and combines well with other coding approaches such as Multiple Description Coding (MDC) [5] and Scalable Video Coding (SVC) [15]. Previous experiments have shown that incorporating multiple description codes and MVP at the video capture and encoding stage enables viewers to use a wider range of heterogenous resolution devices such as desktops, laptops, netbooks, and mobile devices and phones [7].

MVP offers the following technological advantages:

1. MVP captures a wider field of view in the classroom for the same resolution and size. In addition, each camera can be placed for optimal viewing in terms of distance, angle and lighting.
2. MVP ignores less important areas of the presentation.
3. MVP captures multiple views in parallel, which enables parallelization approaches to be used for capture, processing and distribution.
4. MVP improves video compression by taking advantage of temporal characteristics of each view:
    - Overhead slides change slowly, and rarely need more than 1 frame per second refresh rate.
    - Real-time captioning changes slowly, and rarely need more than 10 frames per second.
    - Depending on the lecturer, the view of the lecturer and whiteboard may not need more than 15 frames per second.
    - Sign language interpreters or movies may need 15-30 frames per second.
5. MVP reduces the user's field of view of the classroom, by bringing together important classroom views.

The video compression and scalability reduces and gives the ability to prioritize the total information captured, which in turn reduces processing power and battery demand. This reduction enables limited resolution or bandwidth devices to be used as accessible technology devices, including cheaper and more consumer friendly mobile devices.

4. EVALUATION

To evaluate the efficacy of Multiple View Perspectives (MVP), we compared a Single Video Perspective (SVP) layout (shown in Figure 3) with an MVP layout of a lecture (shown in Figure 4). We recorded a single lecture on BubbleSort, a simple computer science concept, and created both MVP and SVP versions of the lecture. We chose this BubbleSort concept because it is a non-trival problem that is not very technical and highly visual. We divided the 9-minute long lecture into two equal segments so that we could show the MVP and SVP conditions in a balanced, repeated measures design without modifying the actual order of the lecture. Participants were also shown a short 30 second introductory video to familiarize themselves with both kinds of views. The lectures were made accessible by a University of Houston staff sign language interpreter and captioner. Both have extensive experience with providing access for university level classes for deaf and hard of hearing students.

4.1 Procedure

We recorded a 8 minute 58 second long classroom lecture, using a single high definition camera (Sony HDR-XR200) recording at the best quality HD resolution of 1920x1080 pixels at 30 frames per second with a bit-rate of 17 Mbps. Because the interpreter was partly blocking the view of the power point slides, we also recorded the classroom powerpoint lecture using a standard definition camera (JVC CyberCam) recording at the best quality mode at 640x480 pixels at 30 frames per second. We then encoded the high definition Single View Perspective (SVP) H.264/AVC video format using QuickTime Pro 10 for Windows at the recommended QuickTime HD bitrate of 9 Mbps [13], and the podcast video file size was 602 MB.

Figure 3: SVP (Single Video Perspective) layout: instructor, interpreter, captions and slides

Figure 4: MVP (Multiple Video Perspective) 4-view layout: instructor, slides, captions and interpreter

For the sake of minimizing variables in terms of lighting, zoom, angle and compression quality, we adopted a regions of interest extraction approach for the instructor, interpreter and caption views. This approach minimizes differences in the inherent video quality between the single view and multiple views videos. We extracted four regions of interest (ROI) from the SVP HD video, and combined them into predefined MVP video files. Each ROI was manually selected to focus on the instructor, real-time captions and sign language interpreter. Each ROI's window size was equal to the standard SD resolution of 640x480 pixels. The combined size of each of the three ROI views, plus separate camera view of the slides was one-sixth that of the HD view. Upon combining all the views, the total size was about half the size of the HD view - a savings of about 50% off the SVP file size. As expected, each perspective of the MVP had a file size of 90MB. Then contextual temporal re-compression was applied to each perspective video. The slides perspective video was recompressed with 1 frame per second (fps), reducing the file size by over 95% to 3 MB. Similarly, the real-time captions perspective was re-compressed with 10 fps, which reduced the file size by 66% to 30 MB. The instructor perspective video was recompressed with 15 fps, reducing the file size by 50% to 45 MB. Finally, we did not apply any temporal compression for the interpreter perspective. The four compressed files were combined into a single file for viewing as a Multiple Video Perspective file, and the file size was 168 MB, which is a total reduction of 75% from the Single View Perspective file size. The final step was to partition each version of the nearly 9 minute long lecture files into two halves of nearly 4.5 minutes each, so as to balance participant views of the lecture perspectives.

4.2 Evaluations of MVP and SVP

We recruited 19 deaf and hard of hearing participants ages 20-45 (12 female, 7 male). All participants typically request accommodations such as sign language interpreters or captioners for classes or presentations. After completing a short demographic questionnaire to determine eligibility for the test, participants downloaded and watched the video lectures which consisted of two videos segments, one for the first half (four and a half minutes) and one for the second half of the lecture. Half of the participants watched the first half of the lecture presented as a single high definition view (SVP) and the second half of the lecture presented as 4 equal sized views of the instructor, slides, captions and interpreter (MVP). The other half watched the first half of the lecture presented as MVP and the second half presented as SVP. The total time for the study was about 15 minutes. We measured user preferences using Likert scale and open-ended questions (see Figure 7) after each segment, and evaluated whether students perceived the multiple views as helpful.

4.3 MVP and SVP Results

Participants rated the MVP and SVP classroom views by answering the questions "What is your rating of the [single/multiple] view lecture?" with answers ranging from 1-"Didn't like it at all" to 5-"Liked it a lot." and "Did the [single/multiple] view help during the lecture?" "Is [single/multiple] view easy to use?" "I am confident in using [single/multiple] view in class." "I would recommend [single/multiple] view for use by other deaf or hard of hearing students." all with answers ranging from 1-"Not at all" to 5-"Very much."

While the overall responses were on average slightly higher for the Multiple View Perspectives (MVP) at 3.8 (SD=0.8) than for the Single View Perspective (SVP) at 3.6 (SD=0.9), the differences were not statistically significant ($\chi^2 = 19.87$, $N = 19$, $df = 2$, $p = .019$). However, the students were excited about the potential improvements of creatively dividing a single view into only the most relevant and salient aspects of the classroom.

Specifically, their feedback is as follows:

1. Students using accomodations do not need much resolution for the instructor. That video can be reduced to a small sub-picture within another picture, preferably in the interpreter window.

2. Whiteboard or other occasionally used regions of classroom could be put in a sub-picture as well. Allow the student to switch sizes with the main window when the white board or demo is actually in use.

The feedback we were able to solicit from this trial was then incorporated into an improved version of MVP which we call Compact-MVP or C-MVP.

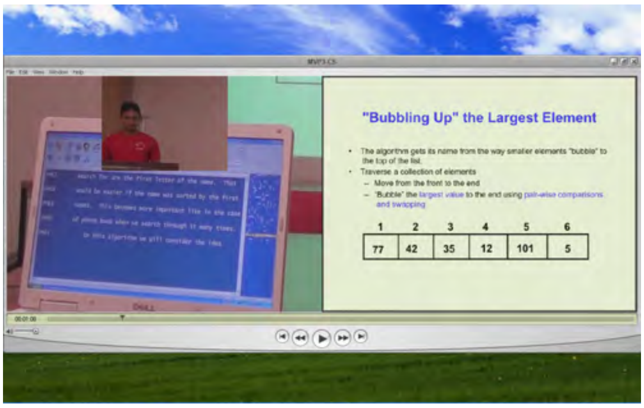


Figure 5: C-MVP 3-view layout: captions, slides and instructor inset

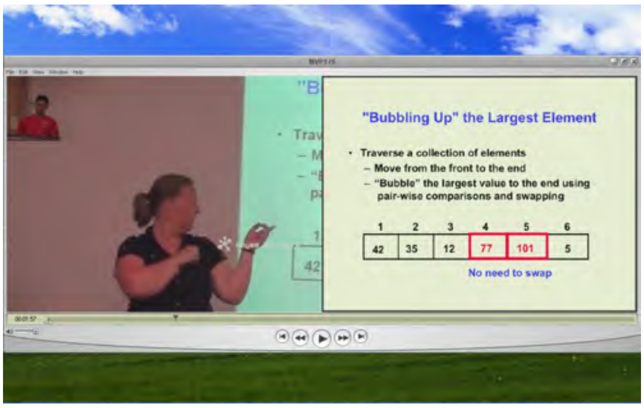


Figure 6: C-MVP 3-view layout: interpreter, slides and instructor inset

4.4 Evaluation of C-MVP

During the first evaluation comparing MVP to SVP, participants commented that the 4 screen view was too distracting. Specifically, many participants commented that they did not need to see both the caption and interpreter screens on a laptop screen that had limited screen size, measuring 14 inches diagonally. Many participants also commented that the instructor view was too large. So for the second iteration, the 4-screen view was revised, with the instructor view reduced to a picture-in-picture view. A 3-screen view was also created and offered, which contained only the slides view, and either a caption or interpreter view, with the instructor view embedded as a picture-in-picture view. We call this new layout Compact-Multiple View Perspective, or C-MVP (see Figures 5 and 6).

By using these techniques, the video display was reduced to three windows and to one or more picture-in-picture windows. As a result, distraction is reduced, and the video bitrate, bandwidth and file size get reduced by an order of magnitude.

After these iterative revisions, a second phase of evaluation was conducted with the same participants, in which multiple versions of MVP were prepared with either interpreter or captions and slides to fit their preferences and reduce distractability and information overload.

4.5 C-MVP Results

The second iteration of trials with C-MVP revealed improved preference scores for each of the five questions asked (for a complete list, see Figure 7). Participants rated C-MVP higher (Q1): the average rating for C-MVP was 4.5 (SD=0.5) whereas MVP and SVP were only 3.7 (SD= 0.9) and 3.5 (SD= 0.8) respectively. Similarly, participants felt that C-MVP was more helpful (Q2) and that it was easier to use (Q3). For example, one participant noted, *"[In MVP,] all information is separated for easy display in a static location. [There is] less 'empty space.'"* Participants also indicated that they felt more confident using C-MVP (Q4) and that they would be more likely to recommend it to other deaf and hard of hearing students (Q5).

Although the C-MVP 3-screen views had the same lighting, resolution and zoom as the old 4-screen view (MVP) and the single view (SVP), viewing quality appeared to improve, due to the fact that the views were bigger on the limited laptop screens. As the video was scaled to fit into the display, laptop screen size impacts satisfaction rate for all views.

5. CONCLUSIONS

We propose a new Multiple View Perspective approach that uses multiple cameras to record and view multiple areas of a classroom presentation. We show that deaf and hard of hearing students prefer this approach over traditional video capture. We also show that it is less expensive and is more inclusive, flexible, easier to deploy. In summary, MVP offers the following advantages for accessible viewing and recording. First, the multiple views can be combined on a screen side by side or even overlapping so as to reduce the field of view, reducing the likelihood of information loss. Second, MVP captures a wider field of view than a single camera and can avoid capturing areas of interest at an awkward angle. Third, MVP captures a collection of areas of interest, and does not capture non-interesting areas of interest, which reduces the video stream bitrate without loss of essential information. Fourth, MVP partitions video into parallel chunks that P2P approaches can efficiently exploit. Fifth, MVP's video partitioning combines well with other video partitioning schemes such as Multiple Description Encoding, for more bitrate scalability. Sixth, MVP scales in terms of video resolution and bitrate, which enables efficient retrieval and viewing over a wide range of consumer devices. Seventh and finally, MVP can be viewed as a new way of transparently multiplexing alternate video or audio presentations and this aids in accessibility and indexing.

The MVP approach of extracting individual regions of interest from a classroom enables context-aware processing of each video so as to enable application of individually optimized recording profiles for improved bandwidth, power consumption and views. Another benefit in bit rate and bandwidth savings is obtained by selecting interesting regions of interest and discarding non-interesting regions of interest. This approach reduced file size and bandwidth demand by 75% for the files used in this study. These two approaches combined yield close to an order of magnitude improvement in the substantial savings in bit rate and bandwidth usage. Together, these two approaches enable a presentation to be streamed effectively to a wide variety of devices in different

Statement and rating options:	SVP	MVP	C-MVP	Sig. N=19, *df*=2
1. What is your rating for single/multiple view. 1 = "Didn't like it at all" ... 5 = "I liked it a lot"	3.5 (0.8)	3.7 (0.9)	4.5 (0.5)	$\chi^2=16.85$, $p < .001$
2. Did the single/multiple view help during the lecture? 1 = "Not at all" ... 5 = "Very much"	3.6 (1.0)	4.0 (0.8)	4.7 (0.5)	$\chi^2=12.72$, $p = .002$
3. Is single/multiple view easy to use? 1 = "Not at all" ... 5 = "Very much"	3.8 (0.9)	3.5 (0.9)	4.4 (0.5)	$\chi^2=10.43$, $p = .005$
4. I am confident in using single/multiple view in class. 1 = "Not at all" ... 5 = "Very much"	3.6 (1.1)	3.9 (0.6)	4.6 (0.5)	$\chi^2=14.29$, $p < .001$
5. I would recommend single/multiple view for use by other deaf or hard of hearing students. 1 = "Not at all" ... 5 = "Very much"	3.6 (1.0)	4.2 (0.7)	4.8 (0.4)	$\chi^2=15.87$, $p < .001$

Figure 7: Participant responses to questions asked after they watched a lecture using a Single View Perspective (SVP), a Multiple View Perspective (MVP), and a Compact Multiple View Perspective (C-MVP). Participants rated the Compact view higher, felt it was helpful and easy to use, felt confident in its use, and were more likely to recommend it to other deaf and hard of hearing students.

situations, such as laptops with wireless connections, and smart phones with mobile data connections.

This paper identifies challenges faced by DHH students in mainstream academics. Our survey showed that DHH students had a clear preference for a compact version of our Multiple View Perspective appraoch (C-MVP). In addition, we show that the approach reduces visual dispersion, reduces capture and bandwidth costs, empowers students, enables missed content review and alternate representations. Finally and most importantly, it creates a more inclusive, more versatile classroom environment. Collectively, these approaches yield a degree of video scaling and information hiding that enables provision of meaningful quality of service for end-users viewing presentations with multiple regions of interest.

6. FUTURE WORK

There is great potential for further work using the MVP paradigm. In terms of the physical set up, one participant suggested for better lighting on the interpreter, *"[You could add an] optional light attached to camera during low light or possible no light presentations when lights are dimmed for slide shows."*

In terms of improving video control and customization, several participants requested a way to zoom in or manipulate the video to better suit their needs. Previous research efforts have successfully implemented real-time region-of-interest (ROI) cropping procedures [3], which could be extended to utilize context to detect borders automatically. This is feasable as the locations of the interpreter, power point display and captions do not normally change over the duration of the presentation. OCR applications like SubRip [18] can be used to convert caption video to text. This approach can yield great compression gains and enhance the accessibility of the audio stream. OCR can also be used to extract text from slides and other views with text, for better low vision and blind participant accessibility. The ability to extract text from views enables indexing by views, such as captions, power point slides, or defined interpreter gestures. Overhead slide indexing has been shown to be very useful for podcast reviews [17]. Automatic camera tracking techniques could be used in the future to adjust to changing classroom situations, like teachers who walk around a lot [14]. Video perspective view aspect correction could be added to enable MVP users to improve an awkward or blocked view by correcting viewing angles [6]. To improve bandwidth performance and power consumption, video recording speed could be dynamically adjusted based on the signing activity by interpreters or students [2].

The use of multiple video cameras and view perspectives to improve mainstream classroom visual accessibility for deaf and hard of hearing students is an exciting area of research. Our iterative design of MVP, a context-aware collection of multiple view perspectives, shows that limited screen space and bandwidth is better utilized by condensing the view of the classroom to only the most important, salient views and that this approach is beneficial and preferred by deaf and hard of hearing students.

7. ACKNOWLEDGMENTS

We thank our participants for their time and feedback. We also thank Anshul Verma for volunteering to give the Bubble Sort lecture used in the evaluations, and the real-time captioner and sign language interpreter for their work in making the lecture accessible to deaf and hard of hearing students.

8. REFERENCES

[1] A. C. Cavender, J. P. Bigham, and R. E. Ladner. ClassInFocus: Enabling Improved Visual Attention Strategies for Deaf and Hard of Hearing Students. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009)*, pages 67–74, Pittsburgh, PA, USA, 2009.

[2] N. Cherniavsky, A. Cavender, R. Ladner, and E. Riskin. Variable frame rate for low power mobile sign language communication. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, page 170. ACM, 2007.

[3] J. Foote, D. Kimber, and B. Manjunath. Region of interest extraction and virtual camera control based on panoramic video capturing. *IEEE Transactions on Multimedia*, 7(5):981–990, October 2005.

[4] J. Gluckman and S. Nayar. Catadioptric Stereo Using Planar Mirrors. *International Journal on Computer Vision*, 44(1):65–79, Aug 2001.

[5] V. Goyal. Multiple description coding: compression meets the network. *IEEE Signal Processing Magazine*, 18(5):74–93, 2001.

[6] G. Hughes and P. Robinson. Photonote evaluation: aiding students with disabilities in a lecture environment. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, page 106. ACM, 2007.

[7] R. S. Kushalnagar. A comparison of p2p approaches in handling multiple video perspective streams by mobile devices to support a more inclusive learning environment. In *Proceedings of the Eighth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom 2010)*, Mannheim, Germany, 2010.

[8] H. G. Lang. Higher education for deaf students: Research priorities in the new millennium. *Journal of Deaf Studies and Deaf Education*, 7(4):267–280, 2002.

[9] Q. Liu, Y. Rui, A. Gupta, and J. J. Cadiz. Automating camera management for lecture room environments. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*, pages 442–449, 2001.

[10] M. Marschark, J. B. Pelz, C. Convertino, P. Sapere, M. E. Arndt, and R. Seewagen. Classroom interpreting and visual information processing in mainstream education for deaf students. *American Educational Research Journal*, 42(4):727–761, 2005.

[11] R. Mayer, J. Heiser, and S. Lonn. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Jrnl of Educational Psychology*, 93(1):187–198, March 2001.

[12] R. Mayer and R. Moreno. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Jrnl of Educational Psychology*, 90(2):312–320, June 1998.

[13] QT H.264. Apple - QuickTime - Technologies - H.264. `http://www.apple.com/quicktime/technologies/h264/`.

[14] A. Ranjan, J. Birnholtz, and R. Balakrishnan. Improving meeting capture by applying television production principles with audio and motion detection. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 227, 2008.

[15] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1103–1120, September 2007.

[16] A. Shachak, M. Hadas-Dayagi, A. Ziv, and S. Reis. Primary care physicians' use of an electronic medical record system: a cognitive task analysis. *Journal of general internal medicine*, 24(3):341–8, March 2009.

[17] J. Subhlok, O. Johnson, V. Subramaniam, R. Vilalta, and C. Yun. Tablet PC video based hybrid coursework in computer science: report from a pilot project. In *Proceedings of the 38th SIGCSE technical symposium on Computer science education*, page 78. ACM, 2007.

[18] SubRip. SubRip on Wikipedia. `http://en.wikipedia.org/wiki/SubRip`.

[19] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical Report CUCS-061-08, Department of Computer Science, Columbia University, Nov 2008.

[20] C. Zhang, J. Crawford, Y. Rui, and L. He. An automated end-to-end lecture capturing and broadcasting system. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 4(1):1–6, 2005.