Replicating Two TelePresence Camera Depth-of-Field Settings in One User Experience Study

Jennifer Lee Carlson
Sr. User Experience Researcher
Cisco Systems, Inc.
170 W Tasman Drive
San Jose, CA 95134 USA
jennicar@cisco.com

Mike Paget
Sr. Technical Marketing Manager
Cisco Systems, Inc.
170 W Tasman Drive
San Jose, CA 95134 USA
mpaget@cisco.com

Tim McCollum
Sr. Design Manager
Cisco Systems, Inc.
170 W Tasman Drive
San Jose, CA 95134 USA
tmccollu@cisco.com

Abstract
This paper describes an experience study to understand the user perceptions on two camera focus settings in a TelePresence room: limited- and infinite-Depth-of-Field. The results influence future TelePresence experience design.

Author Keywords
User experience; comparative study; TelePresence; Depth–of–Field; video conferencing codec; macroblocks; network bandwidth

ACM Classification Keywords
H.4.3. Information Systems: INFORMATION SYSTEMS APPLICATIONS: Communications Applications

General Terms
Human Factors, Experimentation, Design

Introduction
Depth-of-Field is a description of the focal characteristics within a captured image. It describes the sharpness of the image from the foremost to farthest areas on the z-axis within the cameras field of view. The cameras Depth-of-Field is determined by four key factors which were related to works in this study:

1. The proximity of the two lenses to the camera sensor and the cameras overall proximity to the subject, otherwise known as focal length.
2. The amount of light that is allowed to reach the sensor controlled by the aperture setting.
3. The duration at which light is allowed to pass through the aperture, which is called shutter speed.
4. Camera gain setting, which can increase perceived brightness in the image.

Two very common approaches to image capture produce very different resulting images under the same environmental conditions. The approach of limited Depth-of-Field is to limit the amount of focal area within the image to achieve controlled focal points. Generally this is an artistic decision for a particular aesthetic style. This has also been used in video applications such as cinema for the same artistic purpose. However, in video applications such as conferencing, the same image characteristics have been used for a completely different purpose [1]. Current video codecs used in conferencing systems apply an algorithm that defines what information is sent based on changed events rather than sending the entire image. The algorithm groups areas of information together in macroblocks and sends these chunked updates when an area of the macroblock has changed. This approach requires tedious preparation of the environmental conditions and the camera settings. In the case of the Cisco TelePresence System 3000-series (CTS-3xxx) system designs, they were purposely built for a dedicated room that was optimized for very high quality at a low network bandwidth. Therefore they followed the model of camera settings that provided limited Depth-of-Field.

On the other hand, the approach of infinite Depth-of-Field, where a controlled focal point is not established, is also common in both still image and video image capture. This approach captures more detail within the resulting image and requires the viewer (or end user) to determine their own focal points as they view and process the image. In a conferencing system this approach will capture objects within the camera's field of view, as they exist without the need to adjust the amount of sharpness. Such a system could require additional processing power and bandwidth requirements but it doesn't require the same tedious attention to detail of the environmental conditions or the camera settings. Therefore this approach offers a more flexible deployment model for a wider range of conditions. In the case of the systems that utilized the Precision HD camera, such as the Cisco TelePresence 3-series (T3) system, they shared the same camera for both dedicated and multipurpose room systems. Therefore use of an infinite Depth-of-Field configuration was preferable to allow greatest amount of flexibility.

The two depth-of-field applications were largely based on technical and business reasons. What are the user experience impacts, if any, from the two camera settings in a TelePresence room? Our usability study was to answer the following questions:

1. Are users aware of the difference in the two camera focus approaches? If so, how do they differ?
2. Which approach feels more life-like to users? What made it more life-like?
3. Which approach do users prefer and why? Are there other considerations besides being life-like?

Methodology

In August 2011 the Cisco TelePresence User Experience team conducted a formal usability study in an immersive TelePresence room (see Figure 1).

The study replicated the two camera settings in the same TelePresence room to evaluate the user experience in the context of a meeting. During the session, users focused on the moderator, no documents were shared, and the room had sufficient depth and background to identify the moderator's unique location.

We conducted a total of 27 within-subject comparative usability study [2] sessions, with each session lasting approximately 15 minutes. All participants have experience with TelePresence.

Figure 1. Cisco TelePresence CTS-3000 System

Participants entered the TelePresence room containing three side-by-side HD screens. The middle screen was turned off during the entire study. The participant was seated in the middle of the room so that the left and right screens were the same distance from their seat. The left screen displayed an infinite Depth-of-Field, where both moderator and background were in focus. The right screen displayed a limited Depth-of-Field, where the moderator was in focus but the background was blurred at a noticeable level. After the second day of sessions (completed 15 participants) the background objects were switched completely to counter-balance any effects due to the background objects.

In this study, in order to evaluate the user experience impact from the two depth-of-field settings, it was critical to make the other aspects of the images as similar as possible, such as the field of view and the subject matter within the frame. The two cameras (used for the CTS-3xxx and T3 systems) for which we wanted to test had very different physical characteristics. But it was important that users couldn't tell the different cameras by their physical appearances from the room. It was not possible to house both types of cameras within the same system. Therefore, one camera type was selected that fit the appropriate physical characteristics as well as possessed settings that could achieve both a limited depth-of-field and infinite depth-of-field.

Replicating the Depth-of-Field technique is relatively easy in some areas and difficult in others. The lighting and camera settings (hard and soft) can be replicated easily with this controlled environment. The actual focal settings are more challenging because we didn't actually measure the depth of field with any equipment. It was assumed based on camera and light settings, and by looking at the two set-ups subjectively. However, if we were to focus more effort on measuring the depth of field as to define the distance and amount of sharpness or blurriness, it could be more easily replicated. The other area that was challenging to replicate are the objects in the background. We setup similar background based on props we had available. We could define more parameters on those props to better replicate the testing.

Procedures

Participants were told they would have a conversation with a moderator via TelePresence to discuss their experience with TelePresence, provide feedback and rate their experience. Participants were not informed of

the difference in camera approaches until after they had separately provided feedback and rated both views. They looked at one view at a time until the very end of the session when they compared the views side by side.

The study itself was comprised of three separate elements:

Camera Setting 1
Participants were first presented with a view (segment) of the moderator on either the right or left screen (the order was reversed for every other participant to avoid potential order effects). After answering TelePresence-related questions for several minutes, participants were asked to rate the TelePresence session in terms of video quality and how lifelike it appeared.

Camera Setting 2
Then the view was switched to the opposite side of the room and the moderator moved to the displayed view to interact with the participant. After several minutes of additional conversation, the participants were again asked to rate the video quality and lifelike appearance of the view.

Comparisons
Participants were asked if they could tell any differences between the two views they just looked at. If there were any differences, how the two views appeared differently. Then they were shown both views - one at a time - and asked if they noticed any difference, or if they have noticed any other differences. At the end, participants were shown both views simultaneously so that they could make direct comparisons. Participants were asked to describe any differences they observed. If the participant could not discern a difference in background clarity, the moderator explained the differences between the infinite and limited Depth-of-Field camera approaches. With this knowledge, participants then rated how appealing each view was, which view they preferred and why.

A 7-point scale rating scale was used for all rating questions, where 1 represented the 'worst' rating and 7 represented the 'best' rating.

Findings

The study has identified the following key findings based on participant behavior, feedback and preference ratings:

1. Approximately 93% (25 of 27) participants were unable to distinguish the camera focus approaches on their own without viewing the images side by side. Even after viewing the images side by side, only 37% (10 of 27) of participants were able to discern the difference in background clarity between the two views.

2. Between the two camera focus approaches, on average there were very minimal perceived differences in terms of being lifelike (5.93 for infinite Depth-of-Field vs. 5.86 for limited Depth-of-Field) and video quality (6.32 for infinite Depth-of-Field vs. 6.29 for limited Depth-of-Field.)

3. After understanding the camera focus difference: More participants (11 of 27 or 40%) preferred the infinite Depth-of-Field approach. Fewer participants (8 of 27 or 30%) preferred the limited Depth-of-Field approach. Almost one-third (8 of 27 or 30%) participants did not have a preference between the two approaches. On average the infinite Depth-of-Field view was rated slightly more appealing (5.93 for infinite Depth-of-Field vs. 5.52 for limited Depth-of-Field).

Conclusion

Camera's Depth-of-Field setting is not a significant experience differentiator for an immersive TelePresence room. Infinite Depth-of-Field could potentially provide a more lifelike experience and perceived as better quality.

Potential Future Work

This study was meant to be the first of a series of studies. We want to find out what degree of camera focus difference will be perceivable by most users. We also want to study and analyze how user's preferences for camera focus relate to the different types of meetings: such as an interactive brainstorming session, a round-table team meeting, a single-speaker presentation, or other types of meetings.

One hypothesis was that users who are more technical or goal oriented might show a stronger preference for limited Depth-of-Field because they might focus more on the people than their environment; users who are more artistic or context sensitive might show a stronger preference for infinite Depth-of-Field because they care more about the surroundings of whom they meet with. There wasn't any analysis on how the Depth-of-Field preferences relate to participants' job roles or personalities.

Acknowledgements

The authors thank all participants in Cisco Systems who participated in the study described here. We also thank Laura Borns of Cambridge Consultants for her note-taking and analysis assistance for the study; Kevin Nguyen and Rick AtKisson of Cisco Systems for their support in TelePresence room set-up for the study, and Chris Dunn of Cisco Systems, who initiated this research study and reviewed this submission.

References

[1] O'hara, K., Kjeldskov, J., Paay, J., Blended interaction spaces for distributed team collaboration. In ACM Transactions on Computer-Human Interaction (TOCHI) TOCHI Homepage archive, Volume 18 Issue 1, April 2011, Article No. 3.

[2] Sauro, J., Lewis, J.R. Quantifying the User Experience: Practical Statistics for User Research (2012, ISBN-10: 0123849683 | ISBN-13: 978-0123849687), 10-11.