Ad hoc adaptability in video-calling

Sean Rintel
Kenton O'Hara
Behnaz Rostami Yeganeh
Roman Rädle
Microsoft Research
Cambridge, UK
serintel@microsoft.com
keohar@microsoft.com
t-behros@microsoft.com
t-rorad@microsoft.com

Abstract
In this position paper we explore ad hoc adaptability across devices in video-calling. We note the current difficulty of even simple combinations, discuss design issues, briefly report on a study of ad hoc screen mirroring, and note future directions.

Author Keywords
Video-mediated communication; multi-device interaction; adaptability; intelligibility; collaboration

ACM Classification Keywords
H.4.3 [Communications Applications]: Computer conferencing, teleconferencing, and videoconferencing

Introduction
Many opportunities for sharing resources in a conversation or meeting cannot be planned. As Suchman [10] noted, even when tasks themselves are planned, the achievement of those tasks is often a series of situated actions, the nature and needs of which change with the dual retrospective-prospective view that we take of successive actions. Many tasks are achieved through a combination of prepared resources which are brought to bear as situationally appropriate, discovering the need for unexpected resources, and developing resources within the task itself. Ad hoc sharing, then, is crucial to task fulfillment.

Despite a long history of research into the needs and methods of supporting ad hoc shared access to multiple

information resources across domestic and institutional video-mediated communication [2,3,5], the dominant free and commercial systems on the market are still largely funneled and siloed. Ad hoc-ness is limited or in some cases impossible.

These limitations are the result of the 'one username per device per call' model that pervades most video-calling architectures. Even simple combinations are cumbersome. A collocated group will tend to cluster around a single endpoint, funneling participation into serial sequences of displays unless everyone locally joins the call, which is unlikely and unwieldy. Even a single person at one endpoint cannot expect all their devices to be aware of one another's' presence, state, and capabilities, let alone make use of those capabilities in parallel. As such opportunities for enabling rich conversations (e.g. Figure 1 and Figure 2) and work (e.g. Figure 3) are lost because the socio-technical transaction costs of combining the capabilities of devices is too high.

Figure 1: Personal combinations: Using a smartphone for audio and laptop for video in a noisy environment.

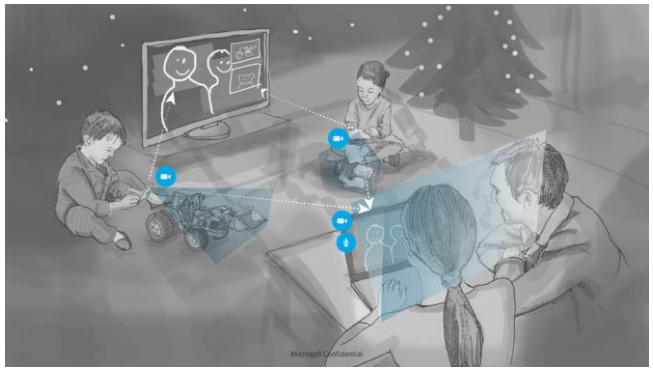

Figure 2: Domestic combinations: Using additional personal or shared devices to allow children to play in their own space while parents converse with grandparents from a laptop.

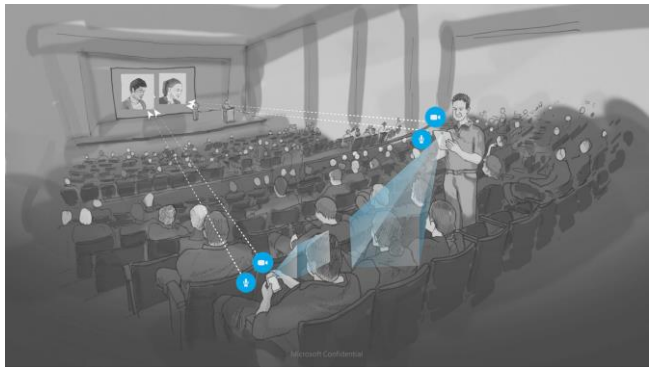

Figure 3: Work combinations: Using additional personal devices to be seen and heard while asking questions in an auditorium.

Designing for ad hoc adaptability

The issues of designing for the kind of ad hoc video-calling scenarios such as those above have been well articulated, especially in the work of Edwards and colleagues Speakeasy system [1] and, in the video-mediated communication field specifically, Neustaedter

Figure 4: Pointing a smartphone at the QR code provided a seamful way for participants to understand that one participant would soon be mirroring.

Figure 5: Local participants could acquire the QR code in parallel and then negotiate serial access to the displayed mirror.

and colleagues' Peek-a-Boo system [8]. Any 'recombinant' system needs to design for awareness of opportunities, intelligible system status, accountable control of status, recoverability and history, flexible and context aware combinations, feedback, simplicity of use and learning, and security and privacy.

Of these, accountability and intelligibility are the paramount drivers of designing for ad hoc adaptability. Accountability, in the ethnomethodological sense, refers to the way in which social order is achieved in the moment through treating social reasoning is observable and reportable in the actions of oneself and others. System accountability should operate in the same way, manifested through intelligibility, which is an ongoing awareness of the state of the system. The actions of the system and its users are thus holistically subject to practical moral reasoning. The other design principles noted above then feed into such reasoning. Further, as with the ad hoc inclusion of capabilities themselves, this moral reasoning will often need to be accomplished on the fly, even if there are pre-established and persistent policies for certain combinations.

As has been argued for home networking [4], we would also argue that intelligibility relies on providing users with access to both simple and detailed depictions of the connections and policies currently invoked in an ad hoc system. To a certain extent, then, we disagree with the 'it just works' market trend promoting the value of invisible, magical, seamless connection. Connections should certainly be easy to accomplish, but there is value, too, in visible 'seamfulness', such that users are not confused or surprised by any given connection. Seamfulness should not be a barrier to action – endless notifications and requests or convoluted specification of all the steps involved in connection – rather it should provide for expectable experiences.

We have explored some of these issues in a small-scale study of screen mirroring [9]. Screen mirroring is limited in many current setups. Access to screen mirroring tends to be restricted to the person driving the host computer. Swapping control is cumbersome and bringing to bear materials from a broader ecosystem of mobile devices and physical information surfaces even more so. The work-around of joining the video-call as an additional participant still involves social negotiations about taking the (displayed) floor [7], as generally only one device can mirror its screen at a time and often the mirroring takes over the majority of the display.

By adding a secondary window with a QR code that allowed all users to mirror their mobile screens – whether they showed material on the device or used the live camera – we found that mobility allowed users to contribute to the video call from their place in the room but also they were able to move around the room and even beyond. Individual work for preparing material to be was carried out in parallel to the overall discussion around the shared display. This enabled a more fluid interleaving of individual, subgroup, and full group sharing activities.

Most importantly, in terms of accountability and intelligibility, we found that participants sometimes negotiated among themselves as to who would take the floor. This could be just-in-time (**Error! Reference source not found.**) or organized serially (bidding for a place such as 'you go first and then I'll go') (**Error! Reference source not found.**). Further, since

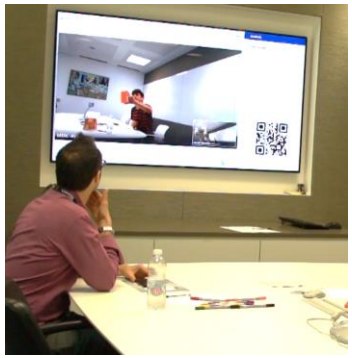

Figure 6: A local participant watches a remote participant acquire the QR code to establish mirroring, and thus knows to wait before establishing his own mirror.

acquiring the QR code on the camera was often a visible action requiring a clear pointing of the smartphone at the shared display, negotiations about upcoming likely sharing needs could be initiated by the same attention to embodied actions that we use to understand gestural onset [6]. However, this useful seamfulness worked best for local users. Remote users' bids for mirroring would sometimes be missed unless local users were watching the screen as the remote user acquired the QR code (Figure 6). Seamfulness, then, must be carefully designed take advantage of local conditions while also recognizing the asymmetry of remote access.

Conclusions

The increasing capabilities of web applications, web media stream standards such as ORTC, and IoT connection standards such as AllJoyn hold promise for the end of host-centric architectures and a bright future for realistic and robust cross-platform cross-device ad hoc adaptability of video-mediated communication systems. The design of such complex systems will need to need to foreground accountability and intelligibility to balance simplicity with seamfulness.

References

1. Edwards, W.K., Newman, M.W., Sedivy, J.Z., and Smith, T.F., 2009. Experiences with recombinant computing: Exploring ad hoc interoperability in evolving digital networks. *ACM Trans. Comput.-Hum. Interact. 16*, 1, 1-44. DOI= http://dx.doi.org/10.1145/1502800.1502803.
2. Finn, K., Sellen, A., and Wilbur, S., 1997. Video-mediated communication Erlbaum, Mahwah, NJ.
3. Gaver, W.W., Sellen, A., Heath, C., and Luff, P. One is not enough: multiple views in a media space. In *Proc. Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems 1993*, ACM (1993), 335-341.
4. Grinter, R.E., Edwards, W.K., Chetty, M., Poole, E.S., Sung, J.-Y., Yang, J., Crabtree, A., Tolmie, P., Rodden, T., Greenhalgh, C., and Benford, S., 2009. The ins and outs of home networking: The case for useful and usable domestic networking. *ACM Trans. Comput.-Hum. Interact. 16*, 2, 1-28. DOI= http://dx.doi.org/10.1145/1534903.1534905.
5. Harrison, S., 2009. Media space 20+ years of mediated life Springer-Verlag, London.
6. Mondada, L., 2007. Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies 9*, 2 (April 1, 2007), 194-225. DOI= http://dx.doi.org/10.1177/1461445607075346.
7. Müller-Tomfelde, C. and O'hara, K., 2010. Horizontal Surfaces in Media Space Assemblies. In *Tabletops: Horizontal Interactive Displays*, C. MÜLLER-TOMFELDE Ed. Springer, London, 435-456.
8. Neustaedter, C. and Judge, T.K. Peek-A-Boo: the design of a mobile family media space. In *Proc. Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing - Adjunct 2010*, ACM (2010), 449-450.
9. Sørensen, H., O'hara, K., Gosset, P., and Kjeldskov, J., 2015. Wireless Smartphone Mirroring in Video Calls. In *Human-Computer Interaction – INTERACT 2015*, J. ABASCAL, S. BARBOSA, M. FETTER, T. GROSS, P. PALANQUE and M. WINCKLER Eds. Springer International Publishing, 410-417. DOI= http://dx.doi.org/10.1007/978-3-319-22698-9_28.
10. Suchman, L.A., 1987. *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press, Cambridge, MA.