Does Voice Coordination Have to be 'Rocket Science'?

Paul M. Aoki

Intel Research
2150 Shattuck Ave., Ste. 1300
Berkeley, CA 94704-1347 USA
aoki@acm.org

Introduction

Imagine sitting down at a desk for an eight-hour shift, slipping a headset onto one ear, and finding four separate conversations going on at once. How would you approach such an environment? What practices would evolve to enable people to make sense of this while working and making highly technical decisions for hours on end?

As it happens, these are not hypothetical questions – in particular, they relate to the Voice Loops system in use in NASA's "Mission Control" at Johnson Space Center, described in a fascinating paper in CSCW 1996 (Watts *et al*. 1996). Further, they highlight the kind of issues that the Voice Loops paper continues to raise for HCI in general. Access to voice communication is spreading rapidly, particularly through the worldwide adoption of mobile telephony – in 2006, the GSM Association reported its two billionth subscription, with 82% of the second billion coming from emerging economies. Yet for most users, the actual *experience* of voice communication has changed very little since the 19th Century invention of the telephone: a simple, additive, monaural voice circuit. This model is fine for conversation between two people, but as anybody who has ever participated in an audio conference can attest, it hardly lends itself to use in group conversation. In practice, it's quite difficult to understand anything if multiple people speak at once, slowing down talk drastically. In spite of its highly-specialized setting, the Voice Loops paper gives hints of a way forward for the rest of us, one that is still largely unexplored.

Overview

The paper has two general areas of interest. First, it is one of a few papers in the HCI and CSCW literature that illustrate how complex command and control environments work. Second, it discusses how multiple multi-party voice communication flows arise in such an environment, and how users successfully manage these flows for hours on end.

The Role of Voice Loops in Mission Control

If one simply wants some "local color" about the management of space missions, it might be better to read one of the several first-hand accounts written by former NASA flight controllers. The Voice Loops paper focuses on the webs of highly concurrent communication that arise, and the resultant monitoring of many activities and conversations at once. From the abstract:

> We describe how voice loops support the coordination of activities and cognitive processes in event-driven domains like space shuttle mission control…how the loops help flight controllers synchronize their activities and integrate information, and how they facilitate directed communication and support the negotiation of interruptions.

In particular, the paper describes the physical and organizational layout of Mission Control and how Voice Loops is overlaid on this layout. Briefly: the main "front" room contains the Flight Director, who has overall responsibility for the mission, and a supporting team of specialized flight controllers, who sit at computer telemetry consoles and are responsible for their respective subsystems. In separate "back" rooms, teams of subsystem specialists support their respective flight controllers. This hierarchy is mirrored in Voice Loops, with the Flight Director talking to the flight controllers over a flight director loop and flight controllers talking to their respective support teams on dedicated front-to-back support loops. Communication within ad hoc groups of flight controllers is supported by a conferencing loop. Finally, a special loop is used for communication with the spacecraft. Hence, individuals monitor four separate loops. Separation into multiple loops provides a structure for information to be aggregated as it flows "upward" and monitored by concerned parties at all levels.

The seminal work of Heath and Luff (1992) on the London Underground Line Control Rooms brings across some of the nature of this coordination "in the small," and the work of Hughes and colleagues (e.g., (Harper and Hughes 1992)) on air traffic control centers gives a somewhat better feel for the technical work of control rooms. However, the Voice Loops paper gives a good sense for the pervasive, critical role that voice communication plays in the coordination of the multiple, interlocking teams.

Managing Attention in Voice Loops

While most of the paper describes the organizational aspects relating to Voice Loops, an extremely important section at the end discusses how individuals interact with the system. As previously noted, controllers routinely monitor four loops – up to four separate, simultaneous conversations – at once, and they do this for eight to twelve hour shifts. The paper explains how people can manage many simultaneous conversations given a few interface primitives for managing them.

> [W]e suggest factors like attentional cues, implicit protocols, and the structure and features of the loops, which might govern the success of voice loops in the mission control domain.

The authors allude to psychology papers on the subject of human attention, then plunge into a discussion of factors that allow the flight controllers to manage the "cacophony" of multi-party talk. Some are protocols specific to the MCC environment, such as specialized ways of speaking ("response on demand" and "coded language") and patterns of loop use ("which loops are monitored" and "functionally separate loops"); such protocols and patterns reduce overall traffic on the loops but do not explain how multiple loops are navigated. However, the authors do also talk about cues and tools: the "internal cues" and "external cues" that prime and trigger listeners' attention, and the ability to use "different volume levels" and to "tailor the set of loops" to separate and prioritize loops when allocating attention. Through these cues and tools, the system provides means for augmenting human capabilities for selective attention.

Impact

I've claimed that Voice Loops, as an illustration of how a system can augment human selective attention abilities, suggests a way forward for the design of voice coordination systems. But given that the paper itself talks about "experienced practitioners in space shuttle mission control," is it clear that these kinds of cues and tools can be applied in other environments?

Consider other multi-party voice applications. Audio conferencing for focused communication tasks has received a great deal of attention in the networking and multimedia literature, and media spaces as open environments for awareness (Stults 1986) have been much more prominent in the HCI literature. (For example, to place the Voice Loops paper in a historical context, consider that CSCW 1996 also saw the first report of the Thunderwire audio space (Hindus *et al*. 1996), a simple audio-only communication space from Interval Research that descended from the Xerox PARC and EuroPARC line of research on media spaces.) In nearly all of this research, the "solution" to the awkwardness of synchronous multi-party interaction has been increased media quality – higher audio bit-rates to improve immediacy, spatialized audio to ease speaker separation, greater video fidelity to capture non-verbal cues. But such solutions come at the cost of bulky and immersive equipment – cameras, multiple microphones, stereo headsets, and so on.

When developing a research agenda in mobile, multi-party audio communication, I thought back to the Voice Loops paper. Because NASA inherited many cultural and institutional aspects from the U.S. military, Voice Loops naturally shares common ideas with military command and control environments such as the Naval Tactical Data System (Boslaugh 1999), in use since the 1960s. I had experience using NTDS as a Navy officer and knew that environments like that of Voice Loops have been in use for many years by a much wider range of people than seasoned, highly-educated NASA flight controllers. Having seen ordinary sailors – most without post-secondary education or relevant previous work experience – work in these environments for hours at a time, it seemed plausible that one could extend media spaces to enable highly spontaneous, concurrent, multi-party social interaction (the kind seen in sociable dinner table conversation) through mobile audio, without the fancy headsets and cameras. Drawing on some design fieldwork of lightweight mobile audio communication (Woodruff and Aoki 2004), we prototyped a system for multi-party voice conferencing that recognized "who is talking to whom" based on an analysis of participants' turn-taking behavior as determined through audio processing and machine learning (Aoki *et al*. 2003); when the system recognized a change in conversational configuration, it adjusted the volume levels for each participant so that the people in "their" conversation were louder and others were quieter.

The "cacophony" of multiple conversations is something that we manage in social environments all the time. One might think that we must make remote communication more and more like face-to-face communication for this to be practical for ordinary users. The paradigm suggested by the Voice Loops paper – that human selective attention can be augmented through careful design and interface primitives – suggests an alternative for increasing our ability to hold spontaneous, multi-party conversation remotely. It suggests that we should aim to build tools to augment and enhance attention – tools that need not fully replicate human perception and understanding, but (as in the example above) might simply recognize human behaviors sufficiently well to enable mundane conversational practices.

References

Aoki, P.M., M. Romaine, M.H. Szymanski, J.D. Thornton, D. Wilson and A. Woodruff. 2003. *The Mad Hatter's Cocktail Party: A Mobile Social Audio Space Supporting Multiple Simultaneous Conversations.* In *Proceedings of the ACM SIGCHI Conf. on Human Factors in Computing Systems (CHI 2003)*, 425-432. New York: ACM.

Boslaugh, D.L. 1999. *When Computers Went to Sea: The Digitization of the United States Navy*. Los Alamitos, CA: IEEE CS Press.

Harper, R.H.R. and J.A. Hughes. 1992. *"What a F-ing system! Send 'em all to the same place and then expect us to stop 'em hitting": Making Technology Work in Air Traffic Control.* In *Technology in Working Order: Studies of Work, Interaction and Technology*, G. Button (ed.), 127-144. London: Routledge.

Heath, C. and P. Luff. 1992. *Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. Computer Supported Cooperative Work*, 1 (1-2), 69-94.

Hindus, D., M.S. Ackerman, S. Mainwaring and B. Starr. 1996. *Thunderwire: A Field Study of an Audio-Only Media Space*. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 1996)*, 238-247. New York: ACM.

Stults, R. 1986. *Media Space*. Systems Concepts Lab Technical Report. Palo Alto, CA: Xerox PARC.

Watts, J.C., D.D. Woods, J.M. Corban, E.S. Patterson, R.L. Kerr and L.C. Hicks. 1996. *Voice Loops as Cooperative Aids in Space Shuttle Mission Control.* In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 1996)*, 48-56. New York: ACM.

Woodruff, A. and P.M. Aoki. 2004. *Push-to-Talk Social Talk. Computer Supported Cooperative Work*, 13 (5-6), 409-441.