Assembling Collaboration: Informing the Design of Interaction Spaces

Paul Luff, Menisha Patel, Christian Heath and Hideaki Kuzuoka

Submission to RoLSI – 9th December 2013

Abstract:

From the very earliest investigations into conversation analysis, there has been a close relationship between the technology of conversation and the technologies for conversation. The symmetrical quality of conversations through the phone provided an easily accessible resource to begin to study sequences of talk from the participants’ perspective. Over the past 20 years or so researchers have in a number of ways, considered how studies of talk and interaction might in turn inform the design of new technologies. This has involved a number of different kinds of intervention drawing on substantive, methodological and conceptual resources from conversation analysis. In this paper, we discuss a ‘technical intervention’ where studies of visual and material conduct drawing from conversation analysis, informed the design and development of a prototype system: a system that sought to support rich kinds of interaction when participants were distant from one another. We conclude by briefly discussing some of the challenges and opportunities that arise when drawing from studies of social interaction to inform the design and assessment of new technologies.

Introduction

From its beginning conversation analysis (CA) has had a very close relationship to technology (Sacks 1984 [1971], Sacks 1992). The simple ability to record phone conversations made available invaluable materials for analysing everyday conversation. The apparent limitations of the phone call, just involving vocal conduct, provided analysts with resources that resonated with those available to the participants. More recently, innovations in phone technology, such as mobile or cell phones, have facilitated analysis of distinctive phenomena such as how place, location and features of the environment feature in everyday talk (Schegloff 2002, Weilenmann and Larsson 2002). There is a further way in which the concerns of conversation analysts have been entwined with those of technology. For at least 20 years designers of technology have sought inspiration from CA for suggestions on how to design and develop new interaction technologies. Whether these are ideas on how to design an interface with a computer system or suggestions for methods to assess and evaluate proposals and prototypes, conversation analysis has been seen as an invaluable resource for considering different forms of ‘interaction’ with and through technology (Luff, Gilbert et al. 1990).
In this paper we will consider the relationship between a particular kind of innovative technology and studies of interaction. These ‘media spaces’ are audio-visual environments through which participants who may be many miles apart, can work and interact (Harrison 2009). Media spaces provide more capabilities than video telephones or other kinds of simple video-mediated communication. The early examples were envisaged as being more like a shared working space than a transitory call. Hence, they typically provided additional resources with which participants could share materials to support their collaboration. However, tests with early versions of the technology were not encouraging. Perhaps because the developers were preoccupied with a face-to-face model of interaction informed by studies of informal communication, support for everyday work activities was poorly integrated with the support for communication (Heath and Luff 1992). Analyses that considered the details of how talk and visual conduct were produced and recognised suggested some interesting properties of interactions through media spaces. Unlike phone conversations, the environment of action of the two participants was not symmetric and the ways an action was produced by a participant was not visible in the same way to a recipient. The impact of a gesture, for example, was diminished in the remote environment (Heath and Luff 1992). In this paper we will consider an intervention that drew from this prior sequential analysis of video-mediated conduct, as well as from other studies of naturally occurring interactions, to inform an innovative media space called Agora. We will briefly discuss some of the features of this system and how it was informed by studies of interaction in everyday workplaces. We will focus, however on how we drew on sequential analyses of conduct in an assessment of this prototype technology and how this intervention led to a reflection back on studies of everyday interaction.

Background

In conventional media spaces and video-mediated interaction even simple actions, such as pointing to a feature on a document can prove curiously problematic. Put simply, a participant cannot ‘reach’ into the remote domain and point at an object within it. The production of conduct is divorced from the environment of action (Hindmarsh, Fraser et al. 1998). In order to give a motivating example let us consider a fragment of action from a workplace setting. Three designers of a new museum space are sitting around the corner of a desk and discussing the requirements set by their client on how much space is needed for each visitor.

**Fragment 1**

Larry: at one point five square metres per person
Phillip: including

that?

Larry: yeas which is...

As Larry states how much space is allocated (5 square meters), Phillip starts to move his right hand holding his pen towards the document in front of Larry. As Phillip hand arrives above the document Larry looks down towards the document. Phillip then moves his pen down sharply onto the document saying ‘including that’. Larry confirms what is included and goes onto discuss problems involved with this specification.

Phillip’s identification of a feature in the document is a collaborative accomplishment, finely co-ordinated with Larry’s conduct. The reference is accomplished through a sequence of talk, bodily conduct and material action (Hindmarsh and Heath 2000, Goodwin 2003). Although an apparently simple referential activity, this is the kind of conduct that is difficult to accomplish in the fragmented environment of a media space: not only are the domain of the participants and the ecology around them remote from another, but any gestures made through
such technology lose much of their performative impact (Heath and Luff 1992). Put crudely, with video-mediated systems it is impossible to ‘reach into’ a remote domain and shape your actions from moment-to-moment in the light of the conduct of a co-participant.

Examples of such referential activities drawn from audio-visual material recordings in workplaces motivated the design of an enhancement to media spaces developed in collaboration with a team of Japanese engineers. This system, called Agora, uses a combination of projection and video techniques to support collaborative activities over and around documents in a remote space. In order to support distributed interaction the Agora system aims to allow participants similar resources within ongoing interaction to those that would be accessible if they were co-present. As well as providing audio access, it offers a series of interrelated views that enable remote participants both to see and hear one another, access and share paper and digital documents, and point to and gesture over documents both in their domain and their co-participants. So, for example, the hands of a remote participant are projected into the local domain to appear over an object of interest (see Figure 1).

![Figure 1: A view of Agora showing a remote participant pointing to a physical document in the local environment. A projected image of the remote participant’s right hand can be seen over a paper plan in the local document space and on the shared screen.](image)

Studies of workplace domains suggested ways in which the spaces and devices could be configured in Agora. For example, considering how participants reached over a corner of a desk to collaborate over documents suggested the order in which the different components would be placed and the direction any movement might appear.
Data

As with any technologies in development, it is infeasible to deploy the technology in a setting for use by participants as part of their everyday work. Some other way is needed to assess prototype technologies. To explore the extent to which Agora supported seemingly simple gestures, in particular pointing and referencing to features of documents, we organised a series of what are known as ‘quasi-naturalistic experiments’. In quasi-naturalistic experiments tasks are given to the participants but these are open-ended and require the minimum of intervention or training. Compared to more constrained evaluation methods these experiments tend to be open-ended and tasks take some time (between 20 and 45 minutes) with little external interference.

For Agora, the design of the tasks drew from workplaces studies, principally those that had been undertaken in planning and design settings. They were developed to encourage participants to use a wide variety of documents including maps, photographs and textual documents. Amongst other things, the experiments were designed to examine whether participants could discuss details of the documents with their colleagues. We wanted to see whether and how, through the various areas and spaces of the system, they talked about, referenced and in other ways pointed to documents and details of those objects. The tasks demanded changing alignment and shifting reference between objects but did not specifically prescribe how they might be accomplished or organised. To encourage discussion we gave the participants slightly different instructions and collections of documents.

We carried out the experiments with 26 pairs of participants (16 Japanese and 10 English speaking pairs). Each participant used a similar Agora system but these were located in different buildings about 200m away from one other. We collected materials from 5 cameras (2 face to face views, the shared screen and 2 wide angles views of the participants in relation to the Agora system). The materials in this paper, for ease of explanation, focus on the interactions between the English-speaking participants.

Given the space available we can only give an illustrative example of the materials (see (Luff, Kuzuoka et al. 2009) for a more extensive analysis of conduct through Agora).

Analysis

The tasks involved the use of a range of different kinds of documents of various sizes and qualities. Just as in meetings or general office work, the participants had to arrange the documents in their workspaces, select documents to work on and refer to features of those they wished to talk about.

In the following fragment Andrea and Becky are discussing their proposal to build a few new cycle paths. The images for the fragment come from a view in Becky’s room and show the image of Andrea on the large projected screen of Agora. Becky
has suggested one place where a path could be developed and Andrea suggests another.

**Fragment 2 Transcript 1**

2.1

Andrea: so you think

2.2

they could build

2.3

another () small one here?

As she asks Becky ‘so you think they could build another small one here?’ Andrea raises her arm (image 2.1). She then points to a photograph that is displayed on the screen in front of her (image 2.2). As Andrea says ‘build another’ her hand reaches across to the photo and Becky reorients to her own screen where the photograph appears (image 2.3). Through the Agora system a projected image of Andrea’s hand appears over the photograph. Although in different locales, both participants are oriented to the same feature on the photograph. Andrea goes onto animate her suggestion, shaping her hand to indicate the possible width of the path and then moving it left to indicate its extent.
Through the Agora system Andrea manages to identify a detail on a document to a remote colleague, and also animate a proposal for another ‘small’ path. Agora provides a resource through which Andrea’s gestures can be displayed over the feature in question, so her talk can be co-ordinated with her visual conduct and tied to features in the environment. Becky seems to recognise this, going on to discuss in detail why she does not think Andrea’s proposal would work.

Agora not only appears to support participants so that they can point and refer to the details of documents in a remote domain, but also provides resources through which they can animate their discussions with a range of fine-grained gestures. It may be that the quality of the images supports this. But it is also obvious that the projection of an image of a hand provides for a variety of ways of annotating the environment, and allows for these animations to be transformed in the light of the ongoing conduct of a co-participant. In the brief fragment above Andrea first secures alignment to the
document before animating her proposal by shaping and reshaping her hand. The various screens provide not only Becky with access to Andrea’s visual conduct but also Andrea with resources to assess Becky’s participation in the collaborative activity. There is a symmetry to the resources that are provided and this seems to support the collaborative production of referential activities.

However, it should be noted that in its design the Agora system has a few anomalies. In order to provide common access to both the environment of another and their conduct a number of different kinds of devices needed to be provided. So, for example the physical object can only be in one of the rooms. In order to replicate the kinds of alignment found in workplace settings this meant providing multiple views of that object, one image being presented on a screen, another on the desk. The participants could then choose to refer to either the physical object or a digital image of it. This also meant that the participant’s conduct could be displayed in different locales often at the same time. Indeed, because of the ways in which the technology was configured it was possible to have several images of a hand (or fragments of it) being visible at any time (see Figure 1). However, when reviewing the materials collected this multiplication of images did not seem to be problematic for the co-participants.

Although the individual components of Agora seem to facilitate collaboration, it appears that it is the way the different resources can be combined together that supports the participants to accomplish coherent actions. A referential action, such as when a participant identifies a feature on a colleague’s document is itself an emergent, collaborative activity. So its initiation, the orientation of the body and the initial movement of an arm is visible in the large life-sized image of the co-participant, even if this is not the principal focus of attention at the time. This initial movement can both secure a realignment from a remote co-participant and project the trajectory of the forthcoming activity. So, as the arm and hand reaches out towards an object not visible on the large display, the co-participant can tie this view with the emerging image of a hand on the shared display. The screens are arranged to resonate with how trajectories of conduct are accomplished in workplace settings. These resources are also available to the participant performing the activity and so from moment-to-moment they can assess how their own actions are being made sense of. They can, in the course of a gesture, for example, gear that production in the light of the ongoing conduct of their co-participant, changing its pacing, reshaping or extending it. So, once they have secured a realignment to a detail of an object, a participant can animate it in some way, and that animation can be tailored to the emerging response of the co-participant.

Unlike conventional video-mediated systems Agora provides participants with a combination of views. Drawing from studies of visual and vocal conduct in everyday settings the views and the capabilities they afforded were configured in a way that it was envisaged would facilitate collaboration. Analysis of the data revealed that although there were some subtle transformations in the way conduct was presented (Luff et al, 2006), the system seemed to invoke few of the problems found in earlier media spaces. Indeed, some unintended features of the systems seemed to support the production and recognition of conduct. For example, the cameras for the large life-size projections were positioned so they captured the edge of the remote desk. This meant that even before a participant lifted their hands from the desk their preliminary movements were visible to a colleague. Co-participants seemed sensitive to such fine details and would begin to configure their own conduct in the light of these subtle
movements. From the materials gathered the system seemed to support the fluid transition between different kinds of collaborative activities over documents. The visual document coupled with the participants’ talk provided a resource for displaying trajectories of action and a means for co-participants to monitor moment-by-moment the prospective activities of a colleague and then to shape their next actions accordingly.

**Intervention**

It is a common requirement for most approaches to technological design to consider the perspective and circumstances of the potential ‘user’. However, how this is accomplished has been the source of much debate (Carroll 1991). When computer systems were principally intended for the individual, it seemed that psychological orientations, particularly that of cognitive psychology would not only offer methods but also potential guidelines for design (Card, Moran et al. 1983). When attention turned towards collaborative systems then it seemed that resources and methods for design, analysis and assessment might more readily emerge from the social sciences (Galegher and Kraut 1990). In recent years a number of approaches have been developed that suggest ways in which social scientists can support the development of innovative technologies. For good reasons, these suggestions typically are not in terms of simple interventions in the design process, say by specifying a list of ‘requirements’ from a study or a set of ‘design guidelines’, rather their proposers typically recommend richer forms of engagement with developers (Dourish 2006, Crabtree, Rouncefield et al. 2012).

In the development of Agora, the ‘interventions’ were made in a range of distinctive ways and involved close collaboration between the social scientists and the engineers and computer scientists designing the system. First, there was a corpus of video-based studies of earlier media spaces. Drawing from ethnomethodology and conversation analysis these studies revealed some of the problems when participants undertook simple activities through this technology. Perturbations in speech, restarts in the production of an activity and apparent delays in the production of concerted action revealed deficiencies in the ways these media spaces supported collaborative activities. When accompanied by analysis of activities in naturally occurring domains, these suggested additional capabilities that could be offered by a media space and also how this could be configured. Although the analysis was critical in informing such design discussions, the collections of short fragments of conduct on which the analysis was invaluable when helping to warrant particular design choices.

Second, prior studies also helped shape how the technology could be assessed. A number of approaches have been used to assess such collaborative systems, including a number of predefined tasks involving different kinds of puzzles, manipulation of objects or assembling constructions. From these, measures are made of, for example, the time taken to accomplish a task, the number of errors that occur or the prevalence of different categories of referential behaviour (Fussell, Kraut et al. 2000)). The outcomes of such experiments have frequently been ambivalent and tend not to be concerned with the qualities of collaborative action when mediated through technology. Prior studies of collaborative activities of everyday settings suggested ways
in which to design tasks that resonated with problems, issues and circumstances that were ‘quasi-naturalistic’.

Third, the materials gathered from these quasi-naturalistic experiments could be subjected to detailed analysis. As the Agora technology can be considered ‘multi-modal’, to analyse the conduct through it seemed to require multi-modal analysis. Short fragments of conduct could be analysed to reveal how talk, visual conduct and the features in the environment were co-ordinated to produce concerted action or in some cases which seemed to be problematic. Indeed, data session analysis sessions with the designers not only revealed potential problems with the technology but also suggested possible solutions. For example, a session where data from a pilot experiment were discussed revealed problems with the visibility of some projected gestures and suggested simple enhancements to the technology.

As is typical in prototype development these interventions took place in a number of short cycles where a design was proposed, assessed and data were analysed. In all, the project lasted about a year in three iterative cycles, each phase informing the next.

**Evaluation**

In the mid-1980s media spaces were at the heart of the agenda to support collaborative work. There was a hope that this technology would become an undemanding yet invaluable resource for office work and workplace communication and the benefits so afforded would justify its expense. Unfortunately at the time this hope was not borne out. Perhaps this is not so surprising. The early developers of media spaces remained preoccupied with a face-to-face model of interaction. Recently, there has been a renewed interest in video-mediated communication. On the one hand there are ubiquitous domestic applications for video-mediated interaction such as Skype and Apple’s Facetime. On the other expensive infrastructures with multiple, high resolution displays have been developed (e.g. HP’s Halo and Cisco’s Telepresence systems. Hence, many of the original promises of media spaces have recently been rehearsed. However, neither these ubiquitous nor the exotic developments provide resources that smoothly integrate the use of objects within a face-to-face interaction.

Digital, and in some cases, mechanical solutions to this problem, have tended to prove clumsy, restrict the range and flexibility of the participants’ referential actions and demanding, in many cases, an explicitly orientation to the problem of securing satisfactory alignment and involvement. Such technical solutions involving cursors and pointers rarely enable the subtle, progressive shaping of the action with regard to the other, or the co-participant to orient to an emerging projecting action by a colleague. Breaking apart actions from utterances, activities from the environment in which they are accomplished and too broad demarcations of the conduct of co-participants, as many technological devices have done, tends to obscure the very means by which sequential actions are accomplished.

In the light of difficulties, of earlier media spaces (Heath and Luff, 1992; Hindmarsh, et al., 1998), Agora proved surprisingly successful. This is despite participants having limited access to the remote environment, for example they
cannot grasp and manipulate objects in the remote space. However, our analysis suggests that participants are able to produce, recognise and coordinate quite complex, material-focused actions, with others through Agora. They create interesting and innovative solutions given the problems posed. They are also able to design their actions so that they are sensible to a co-participant and can see at a glance, how the other is participating, from their standpoint, within the developing and highly contingent course of an activity. Despite the complex array of scenes and views, and flexible location of documents and resources that are provided by Agora, participants were, perhaps surprisingly, able to produce sequentially coherent multi-modal activities, co-ordinating their talk, visual conduct and their manipulation of objects in a physically distributed environment.

The Agora system did however, have certain obvious drawbacks. It consists of a complex configuration of cameras, projectors, filters, monitors and screens. Although the system was only intended to be experimental it was somewhat cumbersome, and it would need significant re-design before a system that could be developed to be deployed in an organisational setting (cf. Luff, Heath et al. 2006). It also was designed to support only two people collaborating together and requires careful consideration of the arrangement of spaces before investigating whether it can be extended for use by more people. Indeed, the study of Agora informed the development and configuration of another system, called t-Room, that had more sophisticated capabilities to support distributed collaboration. T-room is also an augmented media space but where four (or more) people can work over both paper and electronic documents. The analysis of Agora not only helped configure t-Room to support coherent action between participants, but also suggested ways to assess and evaluate its use ((Luff, Jirotka et al. 2013)

**Discussion**

Media spaces may seem a curious technology. However, they can be seen as an attempt to extend simpler forms of mediated interaction such as telephone systems, which as many conversation analytic studies reveal, facilitate the production and recognition of sequences of social action. Difficulties when interacting through video-mediated systems and in early media spaces seem to be due to the asymmetric nature of the resources available to the co-participants. The way an action is produced is transformed when mediated through these technologies, it is not seen by the geographically dispersed participants. When considering the design of these spaces it seems worth considering everyday settings and the ways in which participants in the course of their activities, establish mutual orientations to features in the environment and invoke and animate them. Even what seems to be a simple referential activity, as when one person points to a feature in a document to another, is produced collaboratively through an emergent sequence of co-ordinated actions, through both talk and visual conduct. By fragmenting these resources for concerted action, mediating technologies have the potential for disrupting the sequential production of activities. As they connect geographically-dispersed environments all media spaces fragment the environment in some way. Agora appears to do this whilst preserving the resources for coherent sequential conduct.
The configuration and detailed design of Agora was informed from video-based studies of social interaction and work in everyday settings: the design aiming to support the kind of collaboration participants engage in when sitting together at a desk. Through use of cameras, filters and projection technology it was possible to develop a configuration through which collaborative activities could be accomplished. But rather than providing a single locus of action the design involved the multiplication of resources. The participants managed to assemble a coherence to the activities produced in the distributed space, they made it ‘home in the world’ (Sacks 1992). Indeed, these experiments reveal the need to pay attention to the moment-to-moment production of conduct. The participants drew on how their talk and visual conduct were finely co-ordinated. Although the system may be considered to be multi-modal in that it provides access to the talk, bodily conduct and the visual environment of a colleague it is through the interweaving of these resources that they produce coherent actions. The analysis of such sequences of multi-modal actions provides a novel way of providing qualitative assessments of innovative technologies, suggesting where particular features may prove problematic or when they facilitate concerted action.

The experiments with Agora are one example where studies of social interaction can have substantive, methodological and conceptual contributions to technological developments. However, such technical interventions also raise some distinctive challenges. With earlier media spaces and video-mediated systems it was possible to rely on cases when conduct seemed to be problematic to help evaluate a technology such as when breakdown required repair, reformulation or repetition. When collaborative behaviour through technology seems less problematic, more fluid and better co-ordinated, at least at first glance, then it can be harder to break apart. This casts in sharp light our current understandings of everyday action and interaction; how participants make sense of another’s conduct within a local environment; and how their colleagues produce the means for them to assemble that coherence. It becomes apparent what practical subtleties and social niceties abound even when someone is only trying to make a simple point. Ironically, what studies of advanced media spaces may crucially reveal, is not so much the inadequacies of the technological solutions so far developed, but more how little we still understand about the moment-to-moment accomplishment of everyday work and interaction.

References


