

Sotto Voce: Facilitating Social Learning in a Historic House

Margaret H. Szymanski¹, Paul M. Aoki⁴, Rebecca E. Grinter², Amy Hurst³,
James D. Thornton¹ & Allison Woodruff⁴

¹*Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304-1314, USA (E-mail: szymansk@parc.com)*; ²*College of Computing, Georgia Institute of Technology, 85 Fifth St. NW, Atlanta, GA 30332-0760, USA (E-mail: beki@cc.gatech.edu)*; ³*Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3891, USA (E-mail: akhurst@cs.cmu.edu)*; ⁴*Intel Research, 2150 Shattuck Ave., Suite 1300, Berkeley, CA 94704-1347, USA (E-mail: aoki@acm.org; woodruff@acm.org)*

Abstract. This study examines visitors' use of two different electronic guidebook prototypes, the second an iteration of the first, that were developed to support social interaction between companions as they tour a historic house. Three studies were conducted in which paired visitors' social interactions were video- and audio-recorded for analysis. Using conversation analysis, the data from the use of prototype 1 and prototype 2 were compared. It was found that audio delivery methods were consequential to the ways in which visitors structurally organized their social activity. Further, the availability of structural opportunities for social interaction between visitors has implications for the ways in which the learning process occurs in museum settings.

Key words: informal learning, museum, historic house, electronic guidebook, conversation analysis

1. Introduction

Learning is fundamentally a social process; people learn as they interact and accomplish their daily activities in the social world around them. Museums and cultural heritage sites are, in some ways, ideal locations in which to examine how this kind of social, often informal, learning occurs through social interaction (Falk and Dierking 2000). According to Leinhardt et al. (2002, p. ix), "People do not come to museums to talk, but they often do talk" – and indeed, a shared, interactive experience with companions is often as important to them as learning, particularly for infrequent visitors (Falk and Dierking 1992; Hood 1983). Despite this, many common educational tools employed by museums tend to reduce people's opportunities to interact with their companions. For example, docent-led tours can turn visitors into a passive audience, and audio tours often isolate visitors into experiential "bubbles" (Martin 2000). To be effective, museum tools

must consider the social context of their visitors' experience (e.g., opportunities for them to interact; Falk and Dierking 1992) as well as the quality of the visitors' activity as they tour.

We report here on a series of studies of visitors' use of two different electronic guidebook prototypes, the second an iteration of the first, that were developed to support social interaction between companions. The studies described here are of interest for several reasons. While there are many studies that consider visitors' talk in some way (e.g., McManus 1989; Allen 2002), vom Lehn et al. (2002) observe that extremely few provide detailed, principled analyses of the structure of social interaction between visitors (Hensel 1987; Hemmings et al. 2000; vom Lehn et al. 2001). Of these, even fewer involve a collaborative project between people making technological interventions in the museum and social analysts (Hindmarsh et al. 2005; see also Brown et al. 2003). Finally, the research here may be unique in its degree of technologist/analyst collaboration – the detailed analysis of the use of the first prototype concretely influenced the design of the second, and this enabled us to examine the effects on social interaction and opportunities for informal learning that resulted from these design changes. Hence, this work should be relevant for (among others) technologists working on learning technologies, social scientists interested in visitor interaction, and designers involved with technological interventions in museums and related institutions.

Too often, the design of educational artifacts in museums gives interaction-with-artifact primacy over interaction-with-companions:

The problem arises with the term interactivity. It suggests active participation, human action creatively articulated not only with regard to an object, artefact or system but in response to an active, potentially intelligent and intentional agent. Unfortunately interactivity is conflated with human social interaction. However, 'interactives' are rarely designed to support or enhance social interaction...The fact that visitors are seldom on their own and that the object, artefact or system may well be used in interaction with others is not infrequently disregarded...With the development of more technically sophisticated 'interactives', when the presence of others is taken into account, their participation is often limited to the role of spectator or witness, an accompanying visitor(s) who, it is believed, will watch their friends or partners and then engage in the particular activity itself. (Heath and vom Lehn 2002, p. 11)

Our design of electronic guidebooks for museums was motivated by the goal of enhancing social interaction. By creating opportunities to interact socially, visitors were also provided with opportunities to engage in learning activity.

Vygotsky's (1986) theory of how learning happens has been instrumental in grounding research on cognition in the social world (Bruner 1986, 1990; Goodwin and Goodwin 1992; Gutierrez 1995; Lave and Wenger 1991; Moll 1990; Resnick et al. 1991; Wertsch 1991). Because Vygotsky was interested in

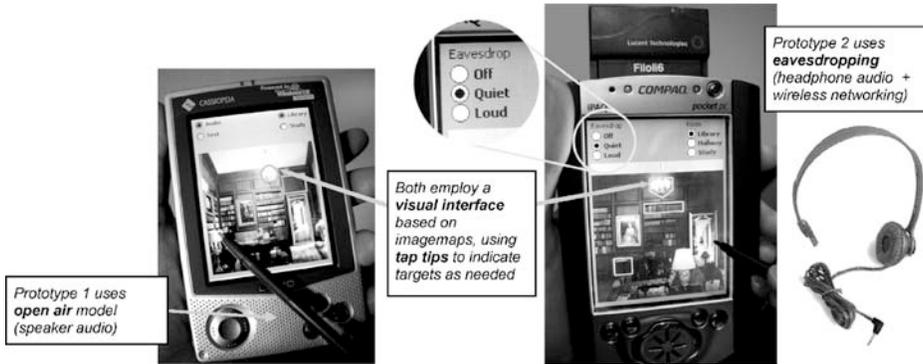


Figure 1. Comparison of Prototype 1 (left) and Prototype 2 (right).

studying how learning happens rather than simply assessing what people had already learned, he recommended analyzing people as they engage in activity. He claimed that, in contrast to traditional experiments, observing people's interactions in the social world would enable him to see "hidden processes," such as the process by which a solution is reached during problem-solving (Moll 1990, p.4). In designing learning activity according to Vygotsky's approach, the aim is to create social contexts in which people are able to use language and manipulate tools to create meaning for themselves (Tharp and Gallimore 1988). For Vygotsky, the primary learning tool that mediates all thought is language, or more specifically people's use of language. For this reason, conversation has taken a prominent role in Vygotskian-based analyses, for he maintained that through social interaction the social becomes psychological in a process of internalization (Tharp and Gallimore 1988, p.29). As Dudley-Marling and Searle (1991) explain, "thinking is an internal dialogue, an internalization of dialogues we have had with others.... – we learn to think by participating in dialogues" (p.60).¹

This paper aims to accomplish three goals: (1) detail our experiences in the iterative design of two electronic guidebook prototypes intended for use in a historic home, (2) describe the patterns of social interaction found through the analysis of visitors' activity while using the guidebooks, and (3) examine visitors' dialogues to explore how technology designed to support social interaction can create opportunities to learn.

2. The electronic guidebook prototypes

Our challenge has been to create an educational museum guidebook that facilitates social interaction. Over a 2-year period, two prototype handheld electronic guidebooks were developed (Figure 1). The two prototypes each have two main components: (1) a visual interface (which is the same in each prototype), and (2) an information delivery mechanism (which is different in the two prototypes). We describe each prototype in turn.²

2.1. Prototype 1

At the beginning of the project, we performed a task analysis and developed a design for a visual interface (Aoki and Woodruff 2000). Individual visitors obtain information about objects in their environment using a visual interface. This helps visitors maintain the flow of their visual task (looking at the room and its contents), which tends to reduce demands on user attention. The interface resembles a set of Web browser imagemap; at a given time, the visitor sees a single photographic imagemap that depicts one wall of a room in the historic house. Visitors change the view perspective (i.e., display a different imagemap) by pressing a hardware button. When visitors tap on an imagemap target, the guidebook provides a description for that object. Many, but not all, of the objects visible on the screen are targets; to help visitors identify targets, the guidebook displays *tap tips* (Aoki et al. 2001) – transient target outlines that appear when the user taps and fails to “hit” a target.

The first prototype offered visitors a choice of information delivery modes: text descriptions, audio descriptions played through headphones, or audio descriptions played through speakers (Woodruff et al. 2001); users of this prototype predominantly chose audio played through speakers in the *open air*.

The audio descriptions were designed to be of limited length in order to prevent the audio from taking over the visitors’ touring activity. For example, the descriptions in Prototype 1 are typically two or three sentences (40 words) long, with the audio duration ranging from 3 to 23 s.³

2.2. Prototype 2

The main difference between Prototype 1 and Prototype 2 lies in their information delivery mechanism. While useful in early prototyping and controlled study environments, audio played through speakers is not a feasible solution in a public space due to noise disturbance issues. The design changes in Prototype 2 are based on three key factors: headsets that do not fully occlude the ears, a careful audio design with properties that differ from those of open air, and an abstraction for audio sharing which we call *eavesdropping* (Aoki et al. 2002). Specifically, devices are paired, with each pair communicating over a wireless local area network. Each visitor in a pair always hears the content they select themselves, and additionally, each visitor has a volume control for determining how loudly they hear content from their companion. The volume can be set to “Off,” “Quiet,” or “Loud” (“Loud” being the same volume as clips selected on one’s own device). A priority model addresses overlapping clips; if visitors have selected a clip themselves, they can always hear it. When they are not listening to a clip themselves, they hear other content from their companion’s guidebook if (1) they have the volume control set to listen to their companion and (2) their companion is playing a clip. The intimate, often directed, nature of the resulting shared audio

Table 1. Summary of studies.

Study	Audio sharing mechanism	Participants	
		Pre-recruited	Public
1	Open air	14	
2	Eavesdropping	12	
3			47

context has led us to call the system *Sotto Voce*, an Italian phrase meaning “under (the) voice.”

3. Materials and methods

To understand visitors’ interactions while using the prototypes, we conducted three studies at Filoli, a Georgian Revival historic house located in Woodside, California (<http://www.filoli.org/>). For Study 1 and Study 2, we ran the studies on days on which Filoli was normally closed to the public. Predetermined visitors, largely recruited from colleagues and friends, privately toured the house. For Study 3, we ran the study on a day on which Filoli was open to the public. Participants were recruited on site. To allow a meaningful comparison across these studies, the data collection procedure, detailed below, remained essentially the same (Table 1).

3.1. Studies one and two

Study 1 examined visitors’ interactions with Prototype 1 (“Open air”), while Study 2 examined visitors’ interactions using Prototype 2 (“Eavesdropping”). In Study 1, seven pairs of visitors participated, and in Study 2, six pairs participated. These pairs were comprised of people with prior social relationships, e.g., grandmother/granddaughter or friend/friend pairs. Each pair of visitors was observed during a private tour of the house. At the beginning of the tour, each visitor was fitted with a wireless microphone.

The tour consisted of three distinct phases. In the first phase, the visitors toured eight rooms using the house’s existing paper guidebook. During this phase, a member of the research team escorted the visitors to answer questions. The visitors’ comments and conversation were recorded using the wireless microphones. In the second phase, the visitors toured several rooms using the electronic guidebook; two rooms in Study 1 and three rooms in Study 2. The researcher distributed guidebooks to the visitors and then gave brief instructions on the operation of the guidebook. The visitors were allowed to move through the rooms without constraints, i.e., they were not instructed to remain together or to interact.

Visitors typically spent about 20–25 min using the electronic guidebooks. In addition to capturing the visitors' conversation using the wireless microphones, we videotaped the visitors using a combination of handheld and fixed cameras and the visitors' use of the guidebooks was logged by the device. The third phase consisted of a semi-structured interview conducted by two members of the research team. The interviews lasted about 30 min.

3.2. Study three

In Study 3, we observed 20 pairs, one group of three, and one group of four using Prototype 2. These pairs and larger groups were comprised of visitors who had come to Filoli together, e.g., mother/daughter or friend/friend pairs. The majority of visitors had not previously used a handheld computing device. The visitors covered a wide range of ages: the youngest visitors were in the "18–29" age range,⁴ and seven visitors who used the guidebook were "over 70."

Visitors to the house were recruited at the entrance to the Library, the first room discussed in the guidebook. After signing consent forms, visitors were fitted with a wireless microphone, given guidebooks, and trained in their use. Next, they visited the three rooms for which the guidebook had content. When they finished using the guidebooks, they participated in a semi-structured interview. As before, the visitors' conversation was recorded using the wireless microphones and their guidebook use was logged by the device, but the visitors were videotaped by fixed cameras (all visitors to the house were notified that videotaping was in progress). Visitors typically spent about 15 min using the electronic guidebooks. Their participation in the study took approximately 30–45 min; no time limits were imposed during any portion of the procedure.

3.3. Conversation analysis

To analyze the visitors' interactions with the electronic guidebook, we applied the methods and findings from the field of conversation analysis (e.g. Heritage 1984a). The goal of a conversation analytic study is to describe the recurrent practices people use to accomplish their social activity. The fundamental unit of analysis is a participant's social action; a series of social actions occurring in interaction produce recognizable sequences or practices. Talk is the most highly used form of social action, but gesture, gaze, posture and the use of tools are other common forms.

We analyze visitors' interactions with the electronic guidebooks in a two step process. First, after identifying an interaction of interest, we produce a detailed, turn-by-turn transcript of the participants' vocal and non-vocal actions to aid in a sequential analysis. Here, we look for the ways in which the participants use the changing nature of the action context as a resource to observe others' actions and



Figure 2. Composite video – guidebook data merged with participants’ video.

to act in order to produce, accomplish and demonstrate their knowledge of the setting and the activity in-progress. Second, we comparatively analyze each excerpt to reveal abstract features that *generalize* across a collection of similar excerpts. The excerpts contained in this paper are representative of a larger collection of excerpts that contain the phenomenon of interest.

A conversation analytic approach is productive in examining learning activity and opportunities to learn that are socially situated. Prior studies have revealed the organization of fundamental features of social action such as turn-taking (Sacks et al. 1974), repair (Schegloff 1992) and sequence (Schegloff and Sacks 1973); these structures are the building blocks for social activity including situated learning. In addition, conversation analytic studies have applied these findings to describe how learning and social cognition occurs (e.g., Goldberg 1975; Lerner 1995; Schegloff 1992).

Table 2. Key elements of transcript notation.

Elements	
X: talk ((action))	Visitor X is speaking or acting
X-PDA: talk	Visitor X’s guidebook is playing
°soft° °°whisper°°	Speech at reduced volume
o:h oo::	Elongated pronunciation
<i>emphasis</i>	Stressed speech
(n) (.)	Pause of n seconds duration; micropause
@talk@	Utterance said laughingly
some[talk here]	Alignment of overlapping speech or actions
[more talk]	
some talk=	Latched speech or actions (no gap between them)
=more talk	

3.4. The data

For each visit, we created a composite video of visitors' interactions and their guidebook screen activity (see Figure 2). Analysis required access not only to the participants' actions, but the technological context of their actions as well. Specifically, the analyst needed to know what the visitors were seeing and hearing from the guidebook at all times.

The visitors' comments and conversation were recorded using individual wireless microphones, the visitors' actions were videotaped, and the visitors' use of the guidebooks was logged by the device. Each frame of the composite video therefore contained three or more video tracks: two tracks with device displays and at least one of the visitors' social action (cutting between the two and four tracks of the visitors available for each visit). In addition, the composite video contained four audio tracks: two of the device audio and two from the wireless microphones. Precise track synchronization was necessary because temporal alignment of events is a critical part of a sequential analysis; for example, it matters whether a visitor's comment is delivered during or after the playback of a guidebook description.

A total of approximately 11 h of composite video data were viewed for analysis. From these data, two collections of excerpts featuring visitors' activity during the playing of a single audio description were transcribed. The collection for Prototype 1 contains more than 65 excerpts, and the collection for Prototype 2 contains more than 70 excerpts. The transcription conventions are Jefferson's as described in Atkinson and Heritage (1984: ix–xvi) with additional notations to indicate when audio content is being played from the electronic guidebook. The most common notational elements used in this paper are summarized in Table 2.

4. Analysis: use of prototype 1

We discuss the analytic findings from our studies of the two prototypes sequentially in Sections 4 and 5 to show how our qualitative analysis of visitors' use of prototype 1 informed the design of prototype 2. In this section, we describe the basic structures by which visitors typically organized their interactions when using Prototype 1. We first discuss three methods by which the visitors incorporated the guidebook's object descriptions into their own interactions. In summarizing, we then relate these methods to corresponding learning resources.

4.1. Listening through open air: visitor behavior

While Prototype 1 offered visitors several delivery mode options (text descriptions, audio descriptions played through headphones, or audio descriptions played through speakers), our study revealed that predominantly visitors preferred to listen to their descriptions played through speakers in the open air.

Furthermore, they often structured their listening as a collaborative activity, preferring to listen to the same descriptions together with their companion. Listening together, or *shared listening*,⁵ provided a way for visitors to integrate the guidebook's information with opportunities for social interaction.

4.1.1. *Listening as a collaborative activity*

To achieve shared listening, participants negotiate and agree on the object for which to play a description. While the audio description plays, they gather close around the guidebook and listen. Then, when the audio description finishes playing, they engage in conversational interaction about its content. Excerpt 1 is representative of the shared listening activity with Prototype 1's open air audio.

Excerpt 1.

- 1 A: okay, so here's over here, looking this way.
 2 ((A takes step towards fireplace wall))
 3 A: we can see about the heh heh deer head.
 4 ((A selects description; A and S look at deer head))
 5 A-PDA: *this deer was shot by the Bourns' son-in-law at the Irish*
 6 *estate the Bourns purchased as a wedding gift for their*
 7 *daughter.*
 8 S: oh,
 9 A: an Irish esta:te.
 10 S: wouldn't you want that for a wedding gift?
 11 A: eh hm.

In Excerpt 1, two visitors coordinate to listen to the description of a deer head hanging on the wall of the study. As the transcript begins, A facilitates shared listening activity with S by verbalizing her synchronization between the guidebook's perspective and her physical perspective of the room. She then identifies an object of interest, after having tapped to reveal the objects containing descriptions on her target wall. A's companion, S, non-verbally coordinates with A by moving with her towards the selected object and by refraining from selecting her own description. After listening to the description together, S and A discuss their impressions of the information they have just received. S evaluates the description's newsworthy character with "oh" in line 8. Then in the three turn exchange that follows, A and S produce a sequence that encapsulates the newsworthy information: A topicalizes an Irish estate in line 9; S produces a rhetorical offer in line 10 (paralleling the one that was given in the description) which solicits A's sentiments had she been given an estate for a wedding gift; and A agrees with S that it would be nice to receive such a wedding gift.

4.1.2. *The Guidebook as a “participant” in storytelling*

After analyzing a collection of shared listening episodes, it became evident that the structure of the visitors’ interactions to set up for the shared listening parallels that of the organization of mundane conversational storytelling activity. Stories in conversation traditionally adhere to the following three-part structure: preface, telling, response (Sacks 1974). First, there is a *preface* for the story in which the participants negotiate a place for an extended turn-at-talk for the storyteller. In this negotiation, the storyteller has an opportunity to confirm that prospective recipients have not already heard the story and that it is of interest to them. In the case of the electronic guidebook, during the preface phase, the visitors select an object of interest and collaboratively create a place for its description to be played. Second, the *telling* of the story occurs; during the telling, recipients produce continuers (Schegloff 1982) that show the storyteller they are paying attention and opting to forego a turn-at-talk until the completion of the story. Third, following the story, recipients participate in a *response* phase in which appropriate comments and reactions to the story are made. These three storytelling phases are paralleled in visitor practices of guidebook use as seen in Excerpt 2.

Excerpt 2.

preface	1	J:	I want to know about that ((points to PDA/object))
	2	G:	about the painting?
	3	J:	yeah, ((points to the painting in the room))
	4	G:	yeah, okay, ((points at the painting in the room))
	5		(1.0) ((J selects painting on PDA))
telling	6	J-PDA:	<i>this is a portrait of Mrs. Roth painted by her</i>
	7		<i>friend Lloyd Sexton, a well-known Hawaiian</i>
	8		<i>painter. In the photograph from which Sexton</i>
	9		<i>created the picture, she was holding daffodils.</i>
	10		<i>Sexton replaced the daffodils, which Mrs. Roth</i>
	11		<i>did not care for, with Queen Elizabeth roses of</i>
	12		<i>which she was very fond.</i>
response	13	G:	((laughs)) diju- diju get that? they had a picture
	14		with daffodils so they (.) made a painting and
	15		put roses in her hand, ((laughs)) that’s neat,
	16		((laughs)) I personally prefer daffodils.

In Excerpt 2, G responds to the audio description as she would in mundane conversational interaction; she laughs. But when her young granddaughter, J,

does not laugh with her, G pursues her laughter by recapitulating what is laughable – that the painting, based upon a photograph, was selectively altered. After more laughter, G produces an evaluative comment, “that’s neat,” and a reflective comment showing her flower preference had she been Mrs Roth.

The storytelling structure of the visitors’ actions while using the guidebook treats the audio description as if it were a storyteller, albeit not an animate one. The paired visitors’ actions make a place to fit the audio content into their own interaction; that is, the audio, controlled by J, is being positioned within the sequence of turns between G and J as if it were a third-party storyteller. G and J, and all the paired visitors we observed engage in shared listening activity, spontaneously adopted this type of three-phase storytelling structure to organize their collaborative interactions with the guidebook.

The episodic nature of this storytelling structure naturally produced an activity structure for the visit as a whole in which conversational *engagement* (Goodwin 1981) varied over time. Paired visitors entered a state of engagement at the beginning of a given storytelling sequence; levels of engagement generally rose and then fell over the course of a given sequence; and visitors then had the option of dis-engaging (resulting in independent activity), remaining engaged in shared activity, or maintaining a nascent engagement in expectation of subsequent re-engagement (Szymanski 1999).

4.1.3. *Visitors’ overlapping talk as “continuers” and “byplay”*

In Excerpt 1 and Excerpt 2, the visitors refrained from commenting while the audio description was playing. This may be due to several factors. First, the storytelling structure specifies that recipients typically forego taking a substantial turn-at-talk until the storytelling has been completed. Second, since the audio description plays to completion, visitors’ comments would likely end up in overlap with the audio and hinder the visitors’ ability to hear the information.

Not all visitors refrained from making comments during the telling, however. The fact that the audio was recorded by a human reading the text descriptions aloud gives the visitors a rich set of resources for constructing and positioning such comments. In mundane conversation, interactants competing for a place to take a turn-at-talk monitor the current speaker’s turn-in-progress for its possible completion; participants draw on many factors (including syntax as well as the complex, dynamic prosody of natural speech) to project such a possible completion, or transition relevance place (TRP; Sacks et al. 1974). With the guidebook, since the descriptions are human voice recordings, visitors can use the same vocal resources they have with any other human interactant to monitor for a TRP and position their comments for these places. This is the case in Excerpt 3.

Excerpt 3.

- 1 S: (0.4) there, floor again, that's floor again.
 2 S-PDA: *the floor is made of oak wood carved with gouging*
 3 *planes. [after the gouges were made, stain was*
 4 A: [yeah
 5 S: [that's, that's interesting,
 6 S-PDA: *applied and puddled in the holes.*
 7 A: wow.
 8 S: hm, [that's interesting.
 9 S-PDA: [*this technique was popular in the early*
 10 *nineteen hundreds when Filoli was built.*
 11 *the effect wears away with use, as can be seen in*
 12 *door to the room.*
 13 S: eEhm uh huh, see here it's darker there, kind of
 14 interesting.
 15 A: (points towards corner) by you.
 16 S: heh.

In Excerpt 3, A and S take turns-at-talk *as* they listen to the audio description. The design and the positioning of their turn-at-talk enable them to do this side-by-side with their shared listening activity. They design their turns-at-talk to be short, single phrase utterances resembling continuers (e.g., yeah, wow, that's interesting). And to avoid overlapping their talk with the audio, they position the production of these utterances to occur at TRPs in the description, where they can project pauses will occur. In lines 4–5 and again in lines 7–8, A and S respond to the audio together; this makes their shared listening activity flow seamlessly into their post-description responsive interaction.

Visitors did not generally make lengthy comments while the guidebook was delivering a description, i.e., they did not generally interrupt it (by stopping the audio) or speak in sustained overlap with it. Visitor tolerance may have been improved by the fact that the descriptions are quite short, the longest in Prototype 1 being 23 s. When visitors do make comments longer than a single word or short phrase amidst the playing of the audio description, analogous to byplay in conversation between three or more people (Goffman 1981), these turns-at-talk are designed to minimize their interference with the description's audio. In Excerpt 4, as the audio description is playing, L identifies something funny in line 10. Her companion, S makes explicit the laughable in line 11. By this time the audio has been playing in overlap with their talk for several seconds. Nevertheless, L takes another turn to confirm S's revelation, albeit with a turn design that is oriented to the overlapping vocal context.

Excerpt 4.

- 1 L-PDA: *these eighteenth century porcelain figures are*
 2 *feng-huang birds, known in the west as phoenix*
 3 *birds.*
- 4 L: oh
- 5 S: [the phoenix,
- 6 L-PDA: [*the chinese believe the feng-huang is the wife of*
 7 *the dragon and alights on the Earth only in times*
 8 *of absolute peace. according to*
 9 [*legend, there was a [period of perfect peace]*
- 10 L: [heh | |]
- 11 S: [they're never coming=]
- 12 L-PDA: [*in the world about*]
- 13 L: [=that's what- @right, that's what-@] ((nods yes))
- 14 L-PDA: *twenty six hundred B C.*
- 15 L: oh, okay,=
- 16 L-PDA: =*at this time, the feng-huang came to Earth and*
 17 *was documented in art, so we now know what the*
 18 *bride of the dragon looks like.*
- 19 L: [oh,
- 20 S: [oh, (.) never knew that, (0.4) okay,
- 21 °let's look at this one, let's see°

In lines 4–5, L and S respond to the audio description with a continuer, “oh,” and a topicalization of new information, “the phoenix.” After the next utterance in the audio description, L responds with a laughter particle in line 10. S responds to L’s laugh particle in line 11, and makes explicit the cause of L’s laughter: since there will never be a time of absolute peace, the phoenix birds are “never coming.” At this point, the audio description has continued to play for several seconds, so when L follows up in line 13 to agree with the formulation S has given for her laughter, she orients to the audio (as speakers would in mundane conversation) with disfluencies in her talk (Schegloff 2000). First she cuts off her initial utterance, then agrees, then attempts to restart the initial utterance before cutting off again and producing a non-vocal nod. Coincidentally, L aborts her turn just as new information is about to be presented in the audio description; her ability to project the sentence-in-progress of the audio description from its syntax is a resource (Lerner 1991).

4.2. Listening through open air: resources for learning

Our findings on the interactional organization of shared listening has implications for social leaning. Specifically, we describe three key phenomena relating to Vygotsky’s constructivist learning, each corresponding to one of the findings above.

First, the coordination that facilitates the shared listening activity naturally creates learning opportunities. For example, by sharing opinions and thoughts about the audio descriptions' information, the visitors can integrate new information with their own experience. In Excerpt 1, S and A construct a mini-scenario about being given an estate for a wedding gift; this conversation confirms their understanding of the description's content and provides them with an opportunity to reflect on the lifestyle of the home's original owners. In addition, the topics raised by the guidebook elicit visitors' conversation about related experiences; for example, later in the tour, S, who has never been married, mentions how her mother recently gave her a very expensive set of china she was saving for her wedding. In Vygotskian terms, the visitors' orientation to the coordinated activity of listening and responding creates a *zone of proximal development* – a context in which peer interaction facilitates the internalization and reinforcement of the guidebook information (Tharp and Gallimore 1988, p.30). This peer interaction is made possible by the visitors' ability to hear each others' audio descriptions and coordinate their actions – something that is not possible when visitors use conventional standalone audio tour units.

Second, the visitors' organization of their interaction as a storytelling structure *scaffolds* their internal learning processes by placing listening and responding in the context of a single, engaged interaction (Tharp and Gallimore 1988, p.33). In Excerpt 2 and Excerpt 3, we see how G and S have been able to position their turns as responsive to rather than re-engaging with their respective companions. The predictability of the storytelling structure enables them to manage their participation in the activity's structure with less interactional work. This enables them to focus more on the audio description's content and on interactions with their companion. By contrast, standalone audio tours produce a more demanding series of interactions, each necessitating interactional re-engagement after the playing of long, unsynchronized audio descriptions.

Third, the practice of byplay interactions during the audio descriptions enables visitors to co-construct an understanding by positioning brief, relevant utterances as the description proceeds. Visitors' comments produced amidst the audio description contribute to the creation of a social context primed for learning. By making comments as the audio is playing, the visitors can co-construct an understanding of the audio description. As the audio plays, visitors' understanding of the information is developing and changing. For example, in Excerpt 4, L shows a change of state (Heritage 1984b) in her thinking in line 15 and alerts S to it with "oh, okay." By verbally marking this moment of resolution, L clarifies her understanding of the description and sets up for conversation about the description's content, rather than having to clarify its meaning, in the response phase. Clearly, the usual audio tour practice of silent listening to unsynchronized descriptions would not provide this resource.

5. Analysis: comparative use of prototype 2

Retaining our goal of enhancing social interaction, we set out to modify the design of Prototype 1 by capitalizing on the social practices visitors demonstrated while using it. To enhance visitors' opportunities for social interaction, Prototype 2 offered visitors the option to receive descriptions by audio using an eavesdropping feature. As previously mentioned in Section 2, this feature was motivated by the need to allow collaborative listening by multiple groups in a given location; when both members of a pair of visitors select the eavesdropping feature on their guidebook, they are in a state of *mutual eavesdropping* and are both able to overhear the other's descriptions when they are not playing one of their own.

In this section, we first compare the behavior of the pairs who chose to use mutual eavesdropping with Prototype 2 to that of similarly engaged pairs who used open air audio with Prototype 1. We then show how the effect of these aspects can be identified in the visitors' learning-related behavior.

5.1. From open air to eavesdropping: changes in visitor behavior

Similarities and differences were found between the behavior of the pairs who chose to use mutual eavesdropping and those that engaged in shared listening with Prototype 1. At a high level, visitors organized their interactions along similar lines: listening as a collaborative activity, the storytelling structure, co-construction of understanding through talk. However, important differences were evident as well. In this subsection, we discuss two specific differences: the structure of the visitors' interactions and their physical mobility.

5.1.1. *Changed activity structure*

Visitor activity was structured very differently with eavesdropped audio than with open air audio. The new structure had a lower coordination cost, demanding less attention. The decreased attention burden was reflected in the visitors' interactions.

In all of the studies described in this paper, the single overall structure that pervaded the interactions was storytelling; as part of this organization, visitors created a conversational role for the audio descriptions, i.e., they treated the guidebook as if it were a "third party" taking an extended conversational turn. With open air audio, visitor interactions tended to focus on choosing individual objects and coordinating with their companion to listen to the descriptions. This setup, repeated for each sequence, focused more attention on coordination activity than seems necessary or desirable. However, the open air audio did afford the opportunity to participate in shared responses to the "story," motivating the visitors to begin setup for another sequence. By contrast, participation in mutual

eavesdropping created an ongoing assumption that the couple would continue in the shared activity. This supposition of continuing shared activity meant that setup tended to be cursory – the playing of the audio does the work of soliciting the attention of the companion to share in the listening.

In Excerpt 5 (Figure 3), F is walking away from J following their shared listening to a painting. As she approaches a cabinet on the other wall, she begins to eavesdrop on a description that J has selected about the door surrounds. This pulls her back into the shared listening activity and prompts her to realign physically with her companion.

Excerpt 5.

1 F: I wish they'd put the worth of some of these things,
 2 I'd just be cur[ious, ((starts walking to U's left))
 3 U: [take it all to uhm (.) what's that r-
 4 (.) antiques roadshow,
 5 (0.4)
 6 F: oheh yeah. ((continues walking towards cabinet))
 7 (0.3) ((U selects Library Entrance))
 8 U-PDA: *both [the library and drawing room are entered*
 9 U: [here, ((points to doorway))
 10 [(F walks back to U))
 11 U-PDA: *through doors [surrounded by architectural*
 12 *[features like Greek columns called aedicules.*
 13 [(U points to doorway))
 14 U: does yours [highlight when mine's (.)/(looks at F-PDA)) no.
 15 U-PDA: [when the house was used for entertaining, these
 16 elaborate door surrounds communicated the fact that only
 17 privileged people were invited to enter these rooms.
 18 U: ((in English accent)) pull out the dummy (.) board, heh,
 19 F: eh heh,
 20 U: ((said laughingly)) only privelaged [pehhople heh heh heh
 21 F: [heh heh heh heh
 22 U: get the dummyboard,

It is instructive to compare the preface in Excerpt 2 (lines 1–5) to that in Excerpt 5, where U does not coordinate verbally with F prior to selecting the description. In fact, we see from the image labeled “Line 6” in Figure 3 that non-verbal coordination is not occurring, either, because F’s back is turned – though U does point to the target object once the audio begins playing in line 9 and again in line 13 when F is available to see the gesture, attempting to fill a deictic gap left by the audio-only shared context enabled by the technology (Heath and Hindmarsh 2000, pp. 96–100). This reduced preface results in modification to

the storytelling structure found with Prototype 1. With the eavesdropping feature, visitors did not necessarily solicit consensus on object selection or recruit the attention of their companion to engage in the shared listening. Instead, each individual was afforded a greater space within which to explore their environment and reorganize for a change in activity (Goodwin 1981, p. 106). That is, the reduction in conversational interaction to coordinate the shared listening activity does not result in an impoverished interactional environment for the visitors, because the activity structure (predictable storytelling) and the guidebook (eavesdropping feature) compensate. The playing of the audio description does the work of pulling in their companion to engage in shared listening.

Without having to repeatedly manage the coordination of shared listening activity, visitors have the opportunity to focus more on the descriptions themselves and to integrate the information they were receiving across the museum visit. In the previous study, responses were generally limited to receipt tokens (e.g., that’s interesting) or the repetition of a description’s content (e.g., the phoenix); extended, multi-turn conversations rarely occurred. In the mutual eavesdropping case, all of the couples engaged in response phase conversations that were more substantive than those seen in the previous study. We will develop this idea further in Section 5.2, but we see an example in Excerpt 5, in which the audio content serves as the springboard for such a conversation. Following the audio description of the library doors in Excerpt 5 (lines 8–17), U and F reflect on the fact that “only privileged people” entered these doors (lines 18–22); U play-acts as such a “privileged” person by calling for the “dummy-board,” an object in the library that they had previously heard described:

This type of flat board painted with a figure is called a “dummyboard” or a “silent companion.” Dummyboards were popular in the 17th and 18th centuries and often depict the lady of the house, children, or pets.

We can reformulate this discussion in terms of interactional engagement (Goodwin 1981; Szymanski 1999), as we did in Section 4.1. The activity structure that was generated by Prototype 1 featured episodic social interactions; the scope of the activity was the playing of a single audio description. Visitors had to do considerable interactional work to locate and agree upon an object of



Figure 3. Visitors interact during movement and audio descriptions (line numbers refer to Excerpt 5).

interest, and then at the completion of each audio description, a lapse in turn-by-turn talk was implicated. With Prototype 1, social interaction between the visitors was vulnerable to dis-engagement (Figure 4, top). Prototype 2 created an enhanced socio-interactive environment, because the eavesdropping feature resulted in extended responses and an increased supposition of continued activity, reducing the visitors' vulnerability to dis-engagement (Figure 4, bottom). Almost counter-intuitively, by reducing the need to interactively orchestrate a collaborative museum tour, the visitors focused more on the guidebook's and the home's contents, resulting in an enhanced social experience.

It is also instructive to compare this with other situations in which a continuous state of incipient talk (Sacks et al. 1974) exists, particularly ones in which there is a notion of a shared activity. For example, CSCW research often highlights the role of awareness in copresent work environments such as railway control stations (Heath and Luff 1992). In such environments, where explicit coordination might detract from the activities at hand, workers monitor each others' actions – much as F remains aware of U's selection of a new description in Excerpt 5, and in contrast to the explicit discussions between G and J in Excerpt 2.

5.1.2. Increased mobility

Compared to visitors using Prototype 1, visitors using mutual eavesdropping were noticeably more mobile during periods of engagement. This mobility was reflected in several ways.

With Prototype 1, the open air audio was played at a low volume, so any movement that changed the relative position of the visitors could cause significant sound attenuation due to distance or blockage (e.g., due to interposed obstacles – even changes in body orientation could cause the audio to be blocked). As a result, couples tended to remain close together and stationary while sharing audio descriptions. If either visitor attempted to move away while the description was playing, visitors had to coordinate their positions to maintain a shared audio context. In some cases, one visitor would attempt to move away while the description was still playing, e.g., to investigate the object being

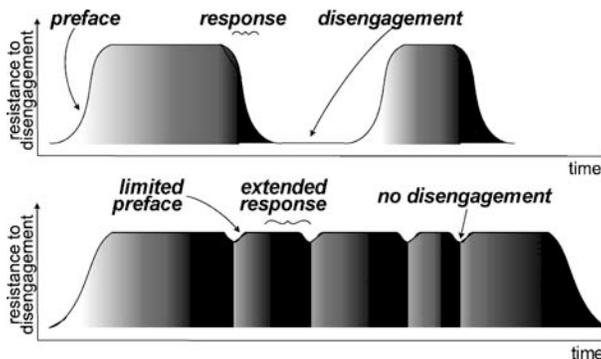


Figure 4. Activity structure: open air (*top*) vs. mutual eavesdropping (*bottom*).

described. Whenever this happened there was quick resolution: either the visitor holding the guidebook playing the audio would quickly return to his or her companion, or the companion would begin to follow the audio-playing guidebook. Some visitors displayed an awareness of their companion's needs. For example (Figure 5), while playing a description, a parent (W) walked around his child (V) so that he could more clearly see the object being described; while moving, he held his electronic guidebook to his child's ear so his listening would not be interrupted.

With the mutual eavesdropping of Prototype 2, visitors were less constrained. Because movement did not attenuate the audio, visitors could separate from each other physically while listening to descriptions and remaining engaged. We illustrate this with a series of images of the wife and husband seen in Figure 6 (D and T, respectively). In Figure 6, both visitors are listening to a description of the marble staircase. While both are examining the staircase, they have each chosen different vantage points. However, this positioning does not compromise their social connection: when the audio description reveals that only the first four steps are actually solid marble, the male visitor looks to his companion and she laughs, even though they are not standing together.

Of particular interest here is the behavior where one person would opt to *follow*, to eavesdrop on all of the descriptions selected by their companion (Grinter et al. 2002). This is the case with the husband–wife pair in Figure 6. As a follower, one is free from having to navigate through the guidebook's virtual walls to select objects of interest that correspond to the physical walls of the room. In addition, the follower is able to position himself or herself to see the aspects of the room's objects just at the moment in which they are mentioned. In Figure 7, T's wife has selected the Chinese chair. The audio describes how the chair is meant to be used in summer because its marble seat and back stay cool during warm weather; this prompts T to touch the seat.

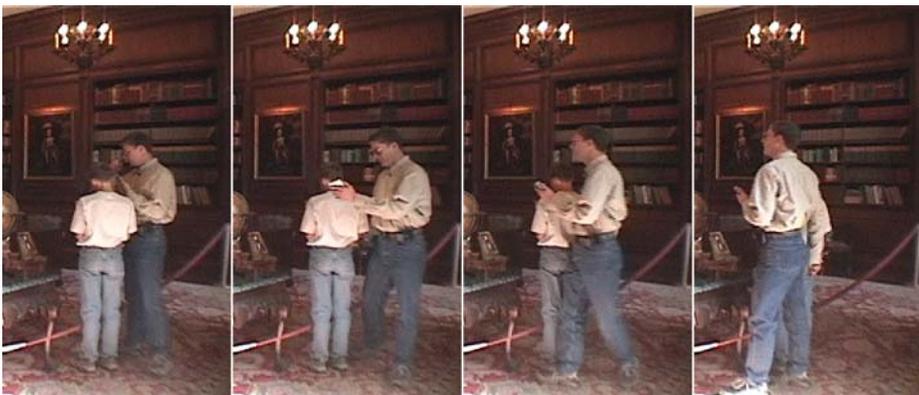


Figure 5. Managing the attenuation of audio in open air (Study 1). W moves from one side of V to the other to get a better view of the object being described. As W moves, he brings the guidebook near V's ear so V's listening will not be interrupted.



Figure 6. Shared reaction to a description heard via eavesdropping.

However, a follower is not precluded from having input about the descriptions being selected to play. As the follower moves about the space, the leader may take note of the objects in which the follower is interested by noticing where he or she is looking or moving. This is the case in Figure 8. D has selected the last three objects in succession and T has responded by approaching the objects accordingly. When T approaches the Dutch cabinet which is located on a different wall in the guidebook, D quickly looks up as T moves towards the cabinet, and she finds and selects it in the guidebook. Then, as the description begins to play, she starts moving towards the cabinet so she can see the smaller items it contains.

While we have discussed D and T's form of following here at some length, the increased mobility resulting from use of mutual eavesdropping took many forms. We observed several common behaviors that rarely, if ever, occurred with open air audio. For example, visitors would walk together while a description was playing, e.g., to approach the object being described. In other cases, a single visitor would walk closer to the object currently being described while their companion remained



Figure 7. Exploring an object whose description was heard via eavesdropping.

Excerpt 6.

- 1 V-PDA: *many of the top shelves contain false books. they are*
 2 *lighter than normal books, so they reduce the stress on the*
 3 *bookcases. many are made of greeting cards, clothing,*
 4 *fabric, et cetera.*
 5 W: eh hah, that's a riot. ((W looks at V and smiles))
 6 (0.2) they're just for looks.

While V and W do share a response, the substance is limited to a single paraphrase of the audio description, analogous to “text echo” of exhibit labels (McManus 1989; though possibly more affective, because of the audio delivery). In another instance, a description of how the original owners celebrated Christmas in the study, with the room holding the entire family including children, grandchildren and the dog prompted a visitor to share, “Speaking of holding the entire family, we’re up to a count of 30 something for our Thanksgiving dinner.” In a fourth case, following the description of charcoal sketches that were made of the previous owners of Filoli, a child volunteered that her own charcoal sketch had been done by an artist at the mall.

By contrast, the activity structure that emerged from visitors’ use of mutual eavesdropping showed that they were orienting to a larger focal activity, the museum experience; this could be seen in the integrated and detail-rich nature of their responses to the audio descriptions. We describe three broad types of such responses below.

First, visitors using Prototype 2 made comments in which they envisioned the house as it was lived in. In Section 4.1, visitor responses reflected fairly immediate reactions (Would you like that..., I like..., it's darker than...). Contrast this with Excerpt 7, in which two visitors are prompted to see the British room on which the Filoli library was originally modeled.

Excerpt 7.

- 1 L-PDA: *...all of the architectural features of this room, including*
 2 *the walnut paneling, are modeled on an 18th century British*
 3 *library. in the original library, each of the outlined panels*
 4 *would have contained framed pictures.*
 5 L: really.
 6 J: yeah.
 7 L: that's a lot of pictures. ((points at wall and sweeps arm
 8 across walls))
 9 J: that's a lot of pictures. ((nods “yes”))
 10 (.)
 11 J: that would've been very cluttered.

In Excerpt 7, L shows disbelief that each panel once contained framed pictures in line 5. J concurs with L’s understanding of the just-heard description,

reinforcing her quantitative inference that becomes explicit in line 7, “that’s a lot of pictures.” L’s gesturing at the many empty wall panels adds physical, spatial and visual elements to the experience and connects their vision of the “original” library with their actual surroundings. J first agrees with L’s evaluation of the quantity of pictures in line 9 and then, following a micropause in line 10, produces a qualitative assessment indicating that she achieved a visualization of the room as it might have been.

Not only is the original house envisioned and experienced, but so too are the lives of the home’s original owners. In Excerpt 8, two visitors bring to life a description about how guests would enter the home for large parties.

Excerpt 8.

- 1 D-PDA: *this door was the party entrance to the house.*
 2 *[during very large parties, the entry*
 3 *[(D walks towards North door)]*
 4 *courtyard was covered with a tent*
 5 *[which blocked the usual entrance to the house.*
 6 *[(T looks towards North Door)]*
 7 D-PDA: *at those times, guests would follow a decorated*
 8 *pathway and enter through this door.*
 9 D: *((whispers)) that must’ve been elegant,*
 10 T: *yeah,*
 11 D: *everybody in their (0.2)*
 12 *[monkey suits and top hats an’*
 13 T: *[it also keeps*
 14 *(.)*
 15 D: *flowing dresses ((spreads arms wide))*
 16 *an’ their (.) just hhoo:: cool.*
 17 T: *it also keeps them out of the private parts I think,*
 18 D: *o:h, well, that’s true too,*

Second, visitors established complex relationships between objects, incorporating information across audio descriptions rather than simply echoing the information most recently played. Such relationships might span objects encountered in different contexts. For example, recall that in Excerpt 5, U’s reference to “pull out the dummyboard” referred to an object explained in a description heard in the Filoli library, visited previously. Relationships might span objects linked in more specific ways, albeit ones not laid out explicitly in the descriptions. For example, a series of portraits prompted one visitor to construct the family tree, saying, “Okay, so that’s his wife, and that’s his mother, right?” Finally, relationships could be self-generated and quite abstract. T (the husband from Figure 6), upon hearing a description about a secret cabinet in the library,

constructed a category of “secret cabinets” that occur in this house; in Excerpt 9, he shows that he is alert to instances of this category as he moves from room to room.

Excerpt 9.

1 T: ah, more secret cabinets.
 2 (0.4)
 3 T: I like that a lot about this house. ((walks into the
 4 bar closet))

Interactions displaying this kind of orientation – i.e., at the granularity of a thematic collection rather than a single object – almost never occurred with open air audio.

It is worth noting here that the Prototype 2 excerpts show a number of interesting and unprompted *learner articulation* behaviors (Koschmann and LeBaron 2002). Visitors in Excerpt 7 and Excerpt 8 elaborate their scenarios more than those in Section 4.1. In Excerpt 8, D’s situated gestures show a strong visual connection with the party and guests described in the audio description. Perhaps most convincingly, the complex, visitor-established relationships articulated in the previous paragraph illustrate the internalization of the audio content.

Third, unlike open air audio visitors, mutually eavesdropping visitors discussed objects that were not described in the guidebook. Excerpt 10, in which J teaches L about a plant, occurred immediately after they finished their response to a description.

Excerpt 10.

1 J: okay, your- your test for the day, what’s that one?
 2 ((points to plant))
 3 (0.2)
 4 J: the plant.
 5 (0.4) ((L leans in to look))
 6 L: morning glory. eh heh heh heh, I don’t know,
 7 what is it?
 8 J: I think it’s a mandevilla vine, but I’m not sure.
 9 L: God, I can’t believe you know that.

With mutual eavesdropping, visitors also frequently branched off into sequences that were not directly related to the content of the description. For example, they might point out a specific physical feature of the object that was not mentioned in the description.

Both of the behavioral changes resulting from use of mutually eavesdropped audio, which we described in Section 5.1, had an impact on social interaction. The primary factor was the new activity structure, which allowed more space for reflection and for visitors to initiate new conversational sequences that were not

structured around the audio descriptions. Increased mobility constituted a secondary factor. In all studies, visitors would start descriptions while they were far away from objects. As mentioned above, visitors were unlikely to walk toward the object while the description was playing with open air audio. However, with eavesdropped audio, they were more likely to approach the object; being close to the object when the description ended gave them more opportunities to observe and discuss its specific features.

5.2.2. *Expanded resources for physical exploration*

With mutually eavesdropped audio, the examination of objects was more frequently occasioned by their presence in the *room* rather than their presence in the *guidebook*. Once visitors began to examine an object, they might discuss it or play a description of it if one were available.

This implicit shift in emphasis from the guidebook to the room as the impetus for exploration is important because it shifts the visitor's role. It is broadly accepted that learning is enhanced by enabling visitors to navigate the museum without leading them through it (Falk and Dierking 2000). However, even "free choice" navigation can be constrained by, e.g., which objects have descriptive content associated with them. Visitor behavior indicates that use of mutual eavesdropping increased the guidebook's utility as a *reference* (an adjunct to the room) as opposed to an *inventory* (a directed guide to the room).

In the study using open air audio, examination of objects often began with objects contained in the guidebook and proceeded by spatial locality. That is, visitors tended to switch the visual interface to a given wall and then look at the objects in the guidebook that interested them on that wall. Object choice was based on targets seen in the visual interface or on short-term memory of such targets.

In the eavesdropping studies, the next object to examine was not always chosen based on availability in the guidebook. We know that some of these examinations were prompted by the room rather than the guidebook because the objects were not described in the guidebook. In other cases, the visitors spoke their thoughts aloud – which was entirely self-prompted since none of the studies involved a speak-aloud protocol. Instead, visitors would encounter objects in their field of view, e.g., objects that were near an object they had just examined, or they would deliberately examine sequences of objects they perceived as being related.

For example, in Excerpt 9, T walks into a new room, notices the bar closet and actually walks into it. *After* this, his companion D finds the description in the guidebook and plays it. Note that because the sound does not attenuate, the visitors can listen to the description together while T stands inside the tiny closet and D stands outside.

While the same resources were available with open air audio, they were used much more frequently in the mutual eavesdropping case due to the changed

activity structure and the increased mobility in the room. Specifically, the mutually eavesdropped audio was more conducive to sequences that were not directly responsive to guidebook content; visitors were generally more open to external triggers with the new activity structure. Visitors acted in a manner more consistent with “Let’s see what’s here in the room” than with “Let’s see what’s here in the guidebook.” Further, visitors had more attention to give to the room due to the reduced attentional demands and wandered more in the room due to increased mobility, so they were more likely to encounter and investigate objects.

6. Conclusion

As social beings, people naturally learn in the course of their daily lives, as they interact with others, as they explore and navigate their environment, and as they engage in productive activity. So the learning challenge is not how to get people to learn, but how to facilitate productive learning by creating enhanced opportunities for people to interact with others, explore their surroundings and engage in activity. This learning challenge is perhaps more difficult in museum settings where the creation of a social learning context requires a delicate balance between the design of learning tools, visitors’ opportunity to interact and the accessibility of the museum’s objects and environment.

In an attempt to address this challenge, this paper has described how our team designed two prototypes, the second an iteration of the first, aimed at creating a context for social learning in a historic home setting. In the most common use case for Prototype 1, open air audio, visitors engaged in a series of shared listening activities around the audio descriptions. The organization of the shared listening activity involved orienting to the guidebook’s descriptions as if they were produced by a conversational participant, paralleling the three-part storytelling structure found in mundane conversational interaction. The familiar storytelling activity structure naturally created a social learning context because it supports social interaction and collaborative activity. The development of Prototype 2 drew upon the findings of the visitors’ preference to engage in the shared listening activity with Prototype 1. The eavesdropping feature added to Prototype 2 essentially built-in the shared listening activity, so visitors could reduce the interactional work needed to coordinate and instead spend more time discussing the guidebook’s content and exploring the room. As a result, visitors’ responsive comments were more content-rich and their exploration of the room and its objects was enhanced.

The key lessons and contributions of the work we have presented here can perhaps be summarized as follows:

First, we have made an explicit case for the facilitation of social interaction and social learning in informal learning sites such as museums and historic houses. The results of the analysis have indicated that technological artifacts – in this

case, electronic guidebooks – can be designed in a manner that supports social learning rather than hindering it.

Second, we have demonstrated how detailed analysis of social interaction can help to isolate critical issues relating to both visitor behavior and design. This kind of analysis revealed phenomena such as the spontaneous emergence of the storytelling structure and the significant differences in activity structure (engagement) that were seen in the respective studies. By examining visitor activity in the context of such phenomena, we have been able to think about the kinds of behaviors facilitated by a particular design and the effects of our design changes.

Third, we have made an implicit case for the incorporation of analyses of this kind into the design process. Our approach to the design of the guidebook prototypes was inductive. We analyzed the data from study one in order to make decisions about how to enhance visitors' learning opportunities with Prototype 2. Our analytic point of departure was visitors' social interactions in the activity setting, a Vygotskian notion that "incorporates cognitive and motoric action itself (activity), as well as the external, environmental, and objective features of the occasion (settings)" (Tharp and Gallimore 1988, p.72). Just as CSCW research has long helped to inform the design of workplace coordination and communication systems (Heath and Hindmarsh 2000, Heath and Luff 1992), our understanding of how people naturally interact with each other as they are using the guidebook to tour a historic house enabled us to build upon their natural practices and to enhance them.

The systems we designed and deployed at Filoli were far from perfect, and eavesdropping is only a first step in the facilitation of social interaction; a great deal of additional research – both technological and social – remains to be done. What we have pointed out is a way forward rather than an ultimate solution: that the ability to facilitate social learning of this kind lies in the ability to provide appropriate social resources, not just information or data, and to provide these resources so that visitors can naturally "fit" them into the fundamental organizational structures of social interaction.

Acknowledgements

We are deeply grateful to Tom Rogers and Anne Taylor of Filoli Center for their assistance with this project. We also thank Tom for his perceptive comments on the design of the prototype; Bob Moore, Mark Newman, Morgan Price, Peter Putz, Terkel Skaarup, Ian Smith, Michael Smith, Erik Vinkhuyzen, and Marilyn Whalen for helpful insights; Maribeth Back for assistance in recording the Study 1 audio clips; and Shane Nye for editing the Study 2 audio and video data. Amy Hurst performed this work during two internships from the College of Computing, Georgia Institute of Technology. This work was funded by the Computer Science Laboratory of the Palo Alto Research Center.

Notes

1. This is not to say that learning simply equates with language use. However, learning occurs in part through social interaction, and we can observe learning processes by observing aspects of social interaction such as “learner articulation” (Koschmann and LeBaron 2002).
2. Since our focus in this paper is on social interaction rather than the technological aspects of the guidebook itself, we limit our system descriptions to a level of detail sufficient to understand the study findings. Schwarzer (2001) provides substantial background information about electronic guidebooks.
3. For example, the 20 s description of a portrait of the Duchess of Richmond and Lennox reads:

This 17th century portrait shows the Duchess of Richmond and Lennox, about whom Pepys said in his diary, ‘Never had a woman more beauty nor less wit.’ The portrait was done by Sir Peter Lely, who was the first of the great English portrait painters. Lely created the distinctive look of British portraiture, including the three-quarter pose and the emphasis on beautiful clothing.

4. While several children participated in the first and second studies, public visitors from the ages of approximately 5–17 are quite rare at Filoli unless they are visiting with a school group.
5. We use “shared listening” to describe the organization of the visitors’ activity. It does not imply that the visitors focused their attention exclusively on the audio descriptions during the entire visit, nor that we can assert that one companion in a pair heard every sound heard by the other. However, as suggested by the excerpts in Section 4, a shared audio context was widely achieved, and Section 5.1 gives further detail about how this came about.

References

- Allen, S. (2002): Looking for Learning in Visitor Talk: A Methodological Exploration. In G. Leinhardt, K. Crowley and K. Knutson (eds): *Learning Conversations in Museums*. Mahwah, NJ: Lawrence Erlbaum, pp. 259–303.
- Aoki, P.M. and A. Woodruff (2000): Improving Electronic Guidebook Interfaces Using a Task-Oriented Design Approach. In *Proc. ACM Conference on Designing Interactive Systems*. New York: ACM, pp. 319–325.
- Aoki, P.M., A. Hurst and A. Woodruff (2001): Tap Tips: Lightweight Discovery of Touchscreen Targets. In *Extended Abstracts, ACM SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 237–238.
- Aoki, P.M., R.E. Grinter, A. Hurst, M.H. Szymanski, J.D. Thornton and A. Woodruff (2002): *Sotto Voce*: Exploring the Interplay of Conversation and Mobile Audio Spaces. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 431–438.
- Atkinson, J.M. and J. Heritage (eds.) (1984): *Structures of Social Action*. Cambridge: Cambridge University Press.
- Brown, B., I. MacColl, M. Chalmers, A. Galani, C. Randell and A. Steed (2003): Lessons from the Lighthouse: Collaboration in a Shared Mixed Reality System. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 577–584.
- Bruner, J. (1986): *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.
- Bruner, J. (1990): *Acts of Meaning*. Cambridge, MA: Harvard University Press.
- Dudley-Marling, C. and D. Searle (1991): *When Students Have Time to Talk: Creating Contexts for Learning Language*. Portsmouth, NH: Heinemann.
- Falk, J.H. and L.D. Dierking (1992): *The Museum Experience*. Washington, DC: Whalesback.
- Falk, J.H. and L.D. Dierking (2000): *Learning From Museums*. Walnut Creek, CA: Altamira.

- Goffman, E. (1981): Footing. In E. Goffman (ed): *Forms of Talk*. Philadelphia, PA: University of Pennsylvania Press, pp. 124–159.
- Goldberg, J. (1975): A System for the Transfer of Instructions in Natural Settings. *Semiotica*, vol. 14.3, pp. 269–296.
- Goodwin, C. (1981): Notes on the Organization of Engagement. In C. Goodwin (ed): *Conversational Organization: Interaction Between Speakers and Hearers*. New York: Academic, pp. 95–125.
- Goodwin, C. and M.H. Goodwin (1992): Context, Activity, and Participation. In P. Auer and A. Di Luzio (eds): *Contextualization of Language*. Amsterdam: John Benjamins, pp. 77–99.
- Grinter, R.E., P.M. Aoki, A. Hurst, M.H. Szymanski, J.D. Thornton and A. Woodruff (2002): Revisiting the Visit: Understanding How Technology Can Shape the Museum Visit. In *Proc. ACM Conf. on Computer Supported Cooperative Work*. New York: ACM, pp. 146–155.
- Gutierrez, K. (1995): Unpackaging Academic Discourse. *Discourse Processes*, vol. 19, no. 1, pp. 21–37.
- Heath, C., and J. Hindmarsh (2000): Configuring Action in Objects: From Mutual Space to Media Space. *Mind, Culture, and Activity*, vol. 7, no. 1&2, pp. 81–104.
- Heath, C., and P. Luff (1992): Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work*, vol. 1, no. 1. pp. 24–48.
- Heath, C. and D. vom Lehn (2002): Misconstruing Interactivity. In M. Hinton (ed): *Interactive Learning in Museums of Art and Design*. London: Victoria & Albert Museum.
- Hein, G. E. (1995): The Constructivist Museum. *Journal of Education in Museums*, vol. 16, pp. 21–23.
- Hemmings, T., D. Randall, L. Marr and D. Francis (2000): Talk, Task and Closure: Situated Learning and the Use of an ‘Interactive’ Museum Artefact. In S. Hester and D. Francis (eds): *Local Educational Order: Ethnomethodological Studies of Knowledge in Action*. Amsterdam: John Benjamin.
- Hensel, K.A. (1987): *Families in Museums: Interactions and Conversations at Displays*. Ed.D. dissertation, Columbia University Teacher’s College, New York.
- Heritage, J.C. (1984a): *Garfinkel and Ethnomethodology*. Cambridge: Polity.
- Heritage, J.C. (1984b): A change-of-state token and aspects of its sequential placement. In J.M. Atkinson and J. Heritage (eds): *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, pp. 299–345.
- Hindmarsh, J., C. Heath, D. vom Lehn and J. Cleverly (2005): Creating Assemblies in Public Environments: Social Interaction, Interactive Exhibits and CSCW. *Computer Supported Cooperative Work*, vol. 14, no. 1, pp. 1–41.
- Hood, M.G. (1983). Staying Away: Why People Choose not to Visit Museums. *Museum News*, vol. 61, no. 4, pp. 50–57.
- Koschmann, T., and C. LeBaron (2002): Learner Articulation as Interactional Jointing: Studying Gesture Conversationally. *Cognition and Instruction*, vol. 20, no. 2, pp. 249–282.
- Leinhardt, G., K. Crowley and K. Knutson (eds.) (2002): *Learning Conversations in Museums*. Mahwah, NJ: Lawrence Erlbaum.
- Lerner, G.H. (1991): On the Syntax of Sentences-in-Progress. *Language in Society*, vol. 20, pp. 441–458.
- Lerner, G. H. (1995): Turn design and the organization of participation in instructional activities. *Discourse Processes*, vol. 19, no. 1, pp. 111–131.
- Lave, J. and E. Wenger (1991): *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- McManus, P.M. (1989): Oh, Yes They Do! How Museum Visitors Read Labels and Interact with Exhibit Texts. *Curator*, vol. 32, no. 3, pp. 174–189.
- Martin, D. (2000): Audio Guides. *Museum Practice*, vol. 5, no. 1, pp. 71–81.
- Moll, L.C. (ed.) (1990): *Vygotsky and Education: Instructional Implications and Applications of Sociohistorical Psychology*. Cambridge: Cambridge University Press.

- Resnick, L.B., J.M. Levine and S.D. Teasley (eds.) (1991): *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association.
- Russell, T. (1994): The Enquiring Visitor: Usable Learning Theory for Museum Contexts. *Journal of Education in Museums*, vol. 15, pp. 19–21.
- Sacks, H. (1974): An Analysis of the Course of a Joke's Telling in Conversation. In R. Bauman and J. Sherzer (eds.): *Explorations in the Ethnography of Speaking*. Cambridge: Cambridge University Press, pp. 337–353.
- Sacks, H., E.A. Schegloff and G. Jefferson (1974): A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, vol. 50, pp. 696–735.
- Schegloff, E.A. (1982): Discourse as an Interactional Achievement: Some Uses of 'uh huh' and Other Things That Come Between Sentences. In D. Tannen (ed): *Analyzing Discourse: Text and Talk*. Washington, DC: Georgetown University Press, pp. 71–93.
- Schegloff, E.A. (1992): Repair After Next Turn: The Last Structurally Provided Place for the Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, vol. 95, pp. 1295–1345.
- Schegloff, E.A. (2000): Overlapping Talk and the Organization of Turn-Taking for Conversation. *Language in Society*, vol. 29, pp. 1–6.
- Schegloff, E. A. and H. Sacks (1973): Opening Up Closings. *Semiotica*, vol. 8, pp. 289–327.
- Schwarzer, M. (2001): Art & Gadgetry: The Future of the Museum Visit. *Museum News*, July/August, pp. 36–41.
- Szymanski, M.H. (1999): Re-engaging and Dis-engaging Talk in Activity. *Language in Society*, vol. 28, no. 1, pp. 1–23.
- Tharp, R.G., and R. Gallimore, (1988): *Rousing Minds to Life: Teaching, Learning, and Schooling in Social Context*. Cambridge: Cambridge University Press.
- vom Lehn, D., C. Heath and J. Hindmarsh (2001): Exhibiting Interaction: Conduct and Collaboration in Museums and Galleries. *Symbolic Interaction*, vol. 24, no. 2, pp. 189–216.
- vom Lehn, D., C. Heath and J. Hindmarsh (2002): Video Based Field Studies in Museums and Galleries. *Visitor Studies Today*, Fall/Winter, pp. 15–23.
- Vygotsky, L.S. (1986): *Thought and Language*. Cambridge, MA: MIT.
- Wertsch, J.V. (1991): *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press.
- Woodruff, A., M.H. Szymanski, P.M. Aoki and A. Hurst (2001): The Conversational Role of Electronic Guidebooks. In *Proc. International Conference on Ubiquitous Computing*. Berlin: Springer, pp. 187–208.