Multimodal Interaction Design in Collocated Mobile Phone Use

Abstract
In the context of the Social and Spatial Interactions (SSI) platform, we explore how multimodal interaction design (input and output) can augment and improve the experience of collocated, collaborative activities using mobile phones. Based largely on our prototype evaluations, we reflect on and discuss three design issues: 1) Haptic feedback design 2) Natural interaction vs. usability 3) Social Signal Processing (SSP) for supporting and evaluating collaborative interactions.

Keywords
Collocated Interaction, Multimodal Interaction, Mobile Devices, Haptic Feedback, Usability, Evaluation, Social Signal Processing

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

General Terms
Design, Experimentation, Human Factors

Introduction
While mobile devices were originally conceived for personal use, sensor and short-range communication technologies now offer possibilities to explore shared use of mobile phones. In this paradigm shift, collocated users can engage in collaborative activities using their personal
The Social and Spatial Interactions (SSI) platform extends the current individual use of these devices to support shared collocated interactions with mobile phones, that range from brainstorming (MindMap prototype [4]), tangible photo-sharing (Pass-them-around prototype [3]), and currently under investigation, multi-player gaming in public places. The main research question the platform addresses is whether people are willing to share their personal mobile devices and engage in collaborative interactions.

One principle of the SSI platform we are currently investigating is how to improve the user experience (UX) of collocated, collaborative interactions using multimodal input and output. Reflection on findings from our user evaluation of Pass-them-around raised 3 relevant issues concerning multimodal interaction design: 1) Haptic feedback design 2) Natural interaction vs. usability 3) Collaboration support and evaluation through Social Signal Processing (SSP).

The rest of this paper is structured as follows: First, we provide an overview of the principles behind SSI. Second, we briefly describe the Pass-Them-Around prototype (Fig. 1) and its interaction techniques built based on the SSI concept. Third, the three multimodal interaction design issues are presented and discussed. Finally, we conclude and speculate over future work.

**SSI Principles**

Principles behind SSI are: 1) **Social**: Supporting joint multiuser interactions by encouraging people to share their devices to reach a goal. 2) **Spatial**: Relative positions of phones tracked with respect to each other on a flat surface are used to create interactions. 3) **Tangible**: Using the mobile phone as a physical interface for manipulating data by performing simple actions (e.g., move, sort, group, join, spin, stack, etc.). 4) **Multimodal**: Currently, use of touchscreen and device movement gestures as main user input modes. Multimodal feedback (visual, haptic, auditory) is provided during interactions.

While each of the built prototypes (MindMap and Pass-Them-Around) was explicitly designed to embody the social, spatial, and tangible aspects of SSI, the multimodal aspect was not researched in depth.

**Pass-Them-Around Prototype**

The idea behind Pass-Them-Around [3] is to allow a small group of collocated people to first individually browse their collections of photos and then collectively share them as a group on any table surface. Instead of taking established digital photo sharing practices as a starting point (e.g., duplicating images, automatic slideshow mode, etc.), we tried to replicate conventional photo sharing practices (e.g., passing paper photos around, face-to-face communication, huddling, etc.) and see how technology could better support those practices.

Some supported interaction techniques are:

- **Browsing Individual Collections**: Tilting device horizontally on either side (Fig. 2a).
- **Sharing Individual Photos (Throwing)**: Image is flicked in direction of proximate device.
- **Sharing Collection (Sequential)**: Tilting device vertically towards table center starts sharing (Fig. 2b). Single intense vibration indicates gesture was performed.
- **Sharing Collection (Photo Pointing)**: Audience can create a copy of currently viewed photo on the other devices.
• **Sharing Collection (Joint Display):** All four devices are connected to create a single larger tiled display (Fig. 3).

• **Seeing Overview:** Picking up the device from the table shows which devices are connected and the currently displayed photos on each.

### Multimodal Interaction Design Issues

#### 1. Haptic Feedback Design

In designing the interactions supported by our prototypes, we faced many design decisions concerning haptic (or vibrotactile) feedback and how this feedback affects interaction. In addition to choosing when in the interaction space haptic feedback was given, we had to also decide on: onset of feedback, duration of feedback, intensity, and feedback repetition. While inclusion of haptic feedback was largely driven by necessity (e.g., notification that a photo was received), designing the parameters of the haptic signal was largely based on intuition, lacking any rigorous design foundations.

Research on haptic feedback design has been primarily concerned with ergonomic settings that take perceptual detection and identification under varying cognitive load as precedent, and only recently with other domains like light weight communication between distributed groups or mobile navigation [5]. We are interested in how to make use of the richness of haptic signal parameters (e.g., rhythm, amplitude, waveform) to structure and support collaborative turn-taking in collocated interactions [1] (e.g., silent signaling to only one user in the group that a photo has been sent in her direction). Moreover, could this rich parameter set be used to expand the perceptual and semantic bandwidth of users during collocated interactions (e.g., increased group activity awareness)?

#### 2. Natural Interaction vs. Usability

In our user evaluations, especially with Pass-them-Around, that the natural interaction we wished to support sometimes came at the expense of usability. During sequential photo sharing, about half of the participants had trouble following the discussion of photos that were not currently displayed on their devices: “The picture on my screen is not interesting because I am not hearing the comments about it. (…) You see the picture that was commented three pictures ago.” [P12].

While photo pointing (see 'Photo Pointing') partially solved the sequential sharing problem, this introduced another usability issue: few participants (4/20) picked the phone up to have a closer view of the photo, which interrupted photo pointing, and instead showed an overview of devices (see ‘Seeing Overview’). Even if picking the device up did provide a close up, this still brings to question whether the designed function should have a ‘natural’ instantiation at the cost of a fluid UX.

The foregoing scenario illustrates the need for a natural interaction equivalent of an “undo” button. Should such a function be made available? If so, should this be based on natural, multimodal interaction (e.g., [6]), or traditional interface buttons? Here we are interested in the potential for true multimodal interaction (i.e., using both gestures and speech) to make interactions with the prototypes both natural and usable. For example, error handling would improve if the system relies on both gestural and speech input before command execution [6], as well as possibly a speech-based ‘undo’ button (cf., Fig. 4 as an illustrative example of multimodal ‘undo’). This also raises the more general question: can multimodal interaction serve as an independent design framework for handling usability issues?
3. Social Signal Processing (SSP) to Support and Evaluate Collaborative Interactions

Much of our interpersonal and social communication happens through non-verbal cues comprising sensory, motor, and vocal (re)actions. For example, eye gaze, iconic body and hand gestures, and prosody provide seamless and effective roles in structuring interpersonal interactions. SSP is a computational framework that seeks to automatically analyze and infer speaker attitude and intention as conveyed through the amplitude and frequency of prosodic and gestural activities [7]. These behaviors are typically detected and recognized through computer vision techniques or speech and audio signal processing, but methods such as proximity sensing or accelerometer readings are also used.

Given the potential of SSP to detect and recognize non-verbal behavior automatically, we are interested in two main application areas: supporting collaborative interactions amongst collocated users by increasing awareness (cf., [2]) of group behavior, and as a supplementary evaluation tool to understand collocated collaborative interactions. For the first use, we are interested in how to detect and recognize communication between participants that occurs outside the sensing scope of the system (e.g., interruptions, backchannels, prosodic cues). This information would be displayed in real-time to increase awareness of group interaction, thus providing a more fluid collaborative experience.

For the second use, there seems to be high potential for automatic analysis of non-verbal cues during social interactions such as when using Pass-Them-Around. These automatically detected behaviors can be correlated with human observations, which further strengthen research findings. Relevant questions are: which automatically detected behaviors are useful in understanding collocated interactions? Can these automatic analyses be used to better understand the structure of collocated, collaborative interaction?

Conclusions

We have presented and discussed three multimodal interaction design issues we consider worth exploring in the context of SSI. Despite that each issue requires different skill sets and methods, we believe each brings an interesting and relevant direction for future work.

References